

Kehl Dániel – Dr. Sipos Béla

Excel parancsfájlok felhasználása a statisztikai elemzésekben

(Oktatási segédlet)

**Pécsi Tudományegyetem
Közgazdaságtudományi Kar
Pécs, 2011.**

Írta:

Dr. Sipos Béla egyetemi tanár, PTE KTK

Az Excel parancs fájlokat programozta:

Kehl Dániel egyetemi tanársegéd, PTE KTK

ELŐSZÓ	4
<hr/>	
BEVEZETÉS, AZ EXCEL BEÁLLÍTÁSAI, AZ EXCEL PARANCSFÁJLOK HASZNÁLATA SORÁN	7
<hr/>	
1 EGYSZERŰ ADAT-ELEMZÉSEK: VISZONYSZÁMOK SZÁMÍTÁSA ÉS GRAFIKUS ÁBRÁZOLÁS	14
<hr/>	
1.1 A DINAMIKUS VISZONYSZÁMOK PARANCSFÁJL MŰKÖDÉSE	17
GYAKORLÓ FELADATOK. (DINAMIKUS VISZONYSZÁMOK.XLS)	18
1.2 ÁBRÁK KÉSZÍTÉSE EXCEL PARANCSFÁJL MŰKÖDÉSE	19
GYAKORLÓ FELADATOK. (ÁBRÁK KÉSZÍTÉSE.XLS)	19
1.3 AZ ORSZÁGONKÉNTI KORFA PROGNÓZIS KÉSZÍTÉSE 2050-IG EXCEL PARANCSFÁJL MŰKÖDÉSE	21
GYAKORLÓ FELADATOK. (ORSZÁGONKÉNTI KORFA PROGNÓZIS KÉSZÍTÉSE 2050-IG.XLS)	22
1.4 NEMZETKÖZI ÖSSZEHOSONLÍTÁSOK EXCEL PARANCSFÁJLOK FELHASZNÁLÁSÁVAL	22
GYAKORLÓ FELADATOK. (NEMZETKÖZI ÖSSZEHOSONLÍTÁSOK EXCEL PARANCSFÁJLOK)	24
<hr/>	
2 ELEMÍ MŰVELETEK A VÁLTOZÓKKAL ÉS EMPIRIKUS ELOSZLÁSOK ELEMZÉSE	24
<hr/>	
2.1 SZÁMLÁLÁS, RANGSOROLÁS, ÖSSZEGZÉS	24
2.2 KÖZÉPÉRTÉKEK ÉS KVANTILISEK	25
2.3 SZÓRÓDÁSI MÉRŐSZÁMOK	27
2.4 AZ ELEMÍ MŰVELETEK PARANCSFÁJL MŰKÖDÉSE	27
2.5 EMPIRIKUS ELOSZLÁSOK ELEMZÉSE EXCEL PARANCSFÁJL MŰKÖDÉSE	29
GYAKORLÓ FELADATOK. (ELEMÍMŰVELETEK.XLS ÉS EMPIRIKUSELOSZLÁSOK ELEMZÉSE.XLS)	37
<hr/>	
3 AZ IDŐSOROK ELEMZÉSI MÓDSZEREI	38
<hr/>	
3.1 A DEKOMPOZÍCIÓS IDŐSORMODELLEK	39
3.1.1 AZ IDŐSOROK ÖSSZETEVŐI ÉS KAPCSOLÓDÁSI MÓDJAI	39
3.1.2 A TREND VAGY A HOSSZÚ TÁVÚ ALAPIRÁNYZAT BECSLÉSI MÓDSZEREI	41
3.1.3 A SZABÁLYOS RÖVID TÁVÚ (SZEZONÁLIS) INGADOZÁS	51
3.1.4 A CIKLIKUS (PERIODIKUS) MOZGÁS MODELLEZÉSE.*	52
3.2 AZ ELŐREJELZÉSEK HIBÁINAK A MÉRÉSE . (A HIBAKÉPLETEK EXCEL PARANCSFÁJL MŰKÖDÉSE)	59
3.3 TRENDSEZON-HIBASZÁMÍTÁS PARANCSFÁJL MŰKÖDÉSE	61
GYAKORLÓ FELADATOK. KONJUNKTÚRA CIKLUSOK MODELLEZÉSE, A TRENDSEZON - HIBASZÁMÍTÁS EXCEL PARANCSFÁJL MŰKÖDÉSE)	63
3.4 A TELÍTŐDÉSI, A LOGISZTIKUS (S-ALAKÚ)- ÉS ÉLETGÖRBE TRENDFÜGGVÉNYEK BECSLÉSE EXCEL PARANCSFÁJLLAL	64
3.4.1 INFLEXIÓS PONTTAL NEM RENDELKEZŐ TELÍTŐDÉSI GÖRBÉK	65
3.4.2 EGY INFLEXIÓS PONTTAL RENDELKEZŐ TRENDFÜGGVÉNYEK	66
3.4.3 KÉT INFLEXIÓS PONTTAL RENDELKEZŐ TRENDFÜGGVÉNYEK	75
3.4.4 A LOGISZTIKUS TRENDEK BECSLÉSE EXCEL PARANCSFÁJL MŰKÖDÉSE	78
GYAKORLÓ FELADATOK. DEKOMPOZÍCIÓS IDŐSORMODELLEK	79
3.5 NAIV ELŐREJELZÉSI TECHNIKÁK. (A NAIVMÓDSZER-PARANCSFÁJL MŰKÖDÉSE.)	81
GYAKORLÓ FELADATOK. (NAIVMÓDSZER.XLS)	85
3.6 AZ EXPONENCIÁLIS KIEGYENLÍTÉS MÓDSZERE (SIMIT.XLS ÉS EXPS FOR WINDOWS)*	85
3.6.1 A SIMIT.XLS PARANCSFÁJL MŰKÖDÉSE.	85
GYAKORLÓ FELADATOK. (SIMIT.XLS)	89
3.6.2 AZ EXPS FOR WINDOWS SZOFTVER MŰKÖDÉSE	89
<hr/>	
3.7 A SABL-MÓDSZER (SZOFTVER) FELHASZNÁLÁSA ADATELŐKÉSZÍTÉSRE, A TREND ÉS A PERIODIKUS HULLÁMZÁS SZÉTVÁLASZTÁSÁRA .*	98
<hr/>	
GYAKORLÓ FELADATOK A SABL-SZOFTVER ALKALMAZÁSÁRA*	106

3.8.1 AZ ARIMA MODELLEZÉS LÉPÉSEI.	109
3.8.2 AZ ARIMA MODELL AZONOSÍTÁSA	122
3.8.3 AZ ARIMA MODELLEK BECSLÉSE	124
3.8.4 EXCEL-PARANCSFÁJLOK AZ ARIMA MODELLEZÉS TÉMAKÖRÉBŐL	127
3.8.5 SPEKTRÁLANALÍZIS.XLS PARANCSFÁJL MŰKÖDÉSE.	141
3.8.6 R+ INTERNETEN ELÉRHETŐ: FREE STATISTICS SOFTWARE (CALCULATOR)	143

4. A KORRELÁCIÓ- ÉS REGRESSZIÓSZÁMÍTÁS

147

4.1 A REGRESSZIÓ.XLS PARANCSFÁJL MŰKÖDÉSE	149
4.1.1 AZ ADAT MUNKALAP	149
4.1.2 A MÁTRIX MUNKALAP	153
4.1.3 A MARADÉK MUNKALAP	156
4.1.4 A MULTIKOLLINEARITÁS MUNKALAP	156
4.1.5 AZ AUTOKORRELÁCIÓ MUNKALAP	164
4.1.6 A HOMOSZKEDASZTICITÁS MUNKALAP	166
4.2 GYAKORLATI ALKALMAZÁSOK BEMUTATÁSA IDŐSOROS ÉS KERESZTMETSZETI ADATOK ALAPJÁN	168
4.3 COCHRANE-ORCUTT ITERÁCIÓS ELJÁRÁS, A COTRANSZFORMÁCIÓ.XLS PARANCSFÁJL MŰKÖDÉSE*	176
4.4 A SZROETER-HARRISON-KING-FÉLE PRÓBA. (SZROETERESZ.XLS PARANCSFÁJL MŰKÖDÉSE) ÉS A GOLDFELD-QUANDT-PRÓBA (GOLDFELD-QUANDT-PRÓBA.XLS PARANCSFÁJL MŰKÖDÉSE)*	178
4.4.1 A SZROETER-HARRISON-KING-FÉLE PRÓBA	178
4.4.2 A GOLDFELD-QUANDT-PRÓBA	181
4.5 A REGRESSZIÓS EGYÜTTHATÓK ÖSSZEFÜGGÉSEI (AZ ÚTELEMZÉS)	184
4.6 KÉSLELTETETT REGRESSZIÓS MODELLEK. (KÉSLELTETETTMÁTRIX.XLS PARANCSFÁJL MŰKÖDÉSE)*	185
4.6.1 A KÉSLELTETÉS MODELLJEINEK RÖVID TÖRTÉNETE	186
4.6.2 A FORDÍTOTT V-KÉSLELTETÉSŰ MODELLEK.	188
4.6.3 KOYCK MÓDSZEREI	189
4.6.4 ALMON-FÉLE POLINOM ELOSZLÁSÚ OSZTOTT KÉSLELTETÉSŰ MODELLEK	193
4.7 A HATVÁNYKITEVŐS, COBB-DOUGLAS TERMELÉSI FÜGGVÉNY (A TERMELÉSI FÜGGVÉNY ÁTLAG ÉS HATÁRMUTATÓI.XLS PARANCSFÁJL MŰKÖDÉSE)*	195
GYAKORLÓ FELADATOK: C-D-TERMELÉSI FÜGGVÉNY ÁTLAG ÉS HATÁRMUTATÓI EXCEL PARANCSFÁJL.XLS ALKALMAZÁSA. NEM LINEÁRIS, DE LINEARIZÁLHATÓ REGRESSZIÓS FÜGGVÉNYEK BECSLÉSE REGRESSZIO.XLS EXCEL PARANCSFÁJLLAL*	206
4.8 A CES-FÜGGVÉNY BECSLÉSE. (CES1.XLS, CES2.XLS CES3.XLS)*	207
GYAKORLÓ FELADATOK CES1.XLS, CES2.XLS ÉS CES3.XLS*	211
4.9 LOGISZTIKUS REGRESSZIÓS FÜGGVÉNYEK*	211
4.10 A SZTOCHASZTIKUS KAPCSOLAT ELEMZÉSE, AZ ASSZOCIÁCIÓS EGYÜTTHATÓK EXCEL PARANCSFÁJL MŰKÖDÉSE	213
4.11 KENDALL-FÉLE RANGKONKORDANCIA-MUTATÓ	221

FÜGGELÉK

223

F.1 INTERNETES INGYENES SZOFTVEREK ÉS ADATBÁZISOK	223
F.2 A MATRIX.XLS PARANCSFÁJL MŰKÖDÉSE	226
F.3 TUDOMÁNYTÖRTÉNETI ÖSSZEFOGLALÓ	228
F.4. TÁBLÁZATOK	233
F.4. A GÖRÖG BETŰK	243

FELHASZNÁLT IRODALOM

244

Előszó

A valós méretű statisztikai modellek megoldása kézi számításokkal általában nem, vagy csak nehezen végezhető el, a számítógépes feldolgozás lehetősége azonban új utakat nyitott meg a statisztika tudományában is. Napjainkban a számolási igény – a személyi számítógépek megjelenése és elterjedése miatt – már nem jelent különösebb akadályt, a számítások megkönnyítésére több matematikai-statisztikai és ökonometriai szoftvert is megalkottak. Ezeknek a programoknak az oktatás és a gyakorlati felhasználás szempontjából azonban több hiányossága is van. Az eladásra szánt programcsomagok¹ általában „fekete dobozként” működnek, azaz a felhasználó nem látja, azt, hogy mi történik a háttérben, a bevitt input és az értelmezendő output jelenik meg csupán. A hivatkozott, legtöbbször az Amerikai Egyesült Államokban kiadott szakkönyvek a hallgatók számára nehezen beszerezhetőek és drágák. Az ilyen szoftverekkel kapcsolatos további gond az is, hogy folyamatosan újabb verziók jelennek meg, ami széleskörű alkalmazásuk lehetőségét megnehezíti. Drágítja a felhasználásukat továbbá, hogy az éves licenccij kifizetésén túl a gépszám függvényében gyakorta külön díjat kell fizetni. A felsőoktatásban sok esetben a szoftverek csak az egyetemi/főiskolai számítógépeken érhetőek el, a hallgatók otthoni számítógépükre legálisan nem telepíthetik azokat. A különböző szoftverek emellett különböző felhasználói felülettel rendelkeznek. A „preferált” csomag kiválasztása így meglehetősen önkényes. A különböző formátumok miatt a programcsomagok közötti váltás némely esetben gondokat okoz. Az interneten található, ingyenesen letölthető ökonometriai programcsomagok, mint például az egyik legismertebb és legelterjedtebb gretl (Gnu Regression, Econometrics and Time-series Library, <http://gretl.sourceforge.net/>), igen sokoldalú szolgáltatást nyújtanak, de az elméleti háttér feldolgozásához a megadott angol nyelvű szakirodalmat² is be kell szerezni és el kell sajátítani. A jelenleg legnépszerűbb irodai programcsomag a Microsoft Office Windows változata 1990-ben jelent meg. A Microsoft Office³ és ezen belül az MS Excel⁴ világviszonylatban és Magyarországon is széleskörűen alkalmazott szoftver. Egyrészt ez a tény indokolja az MS Excel (továbbiakban Excel) alkalmazását, továbbá az is, hogy az előzőekben ismertetett problémákat részben ki lehet küszöbölni. Az Excel sok statisztikai műveletet képes elvégezni, de az alapfunkciók segítségével felépíthetők a bonyolultabb statisztikai és ökonometriai módszerek is a függvények segítségével. Az Excellel ilyen módon a széles értelemben vett modellezést is taníthatjuk a hallgatóknak. További előny, hogy a módszerek, a felhasznált képletek megjelennek, azok alakíthatók, az adott feladat megoldásához testre szabhatók, láthatóvá, és megérthetővé válnak a részeredmények és a mellékszámítások. Az Excel – a speciális statisztikai szoftverekhez hasonlóan, de messze nem olyan részletességgel – a statisztika módszertanának nagy részét felöleli beépített modulja (Analysis ToolPak) segítségével, de jó néhány apróbb hiba (pl. rossz, vagy félreérthető magyarra fordítás) és hiányosság is a sajátja. Az említett félrefordításoknál nagyobb hibák is megfigyelhetők, melyek az Excel korábbi verzióiban csakúgy megtalálhatók voltak, mint a legújabbakban. Az Excel a főként a következtetési statisztikában oly fontos eloszlások esetén némely speciális esetben hibás, nagyban félrevezető értékeket szolgáltat. A témakör bőséges irodalommal rendelkezik, itt csak utalunk Knüsel⁵ illetve McCullough és Wilson⁶ vagy az Excel legújabb kiadásával kapcsolatban Yalta⁷ munkáira, melyekből az érdeklődő olvasó kimerítő „hibalistát” meríthet. Az említett hibák azért is bosszantóak, mert több éve ismertek. Hasonló problémák más szoftverek esetén is előfordultak, de valamennyit a lehető leggyorsabban javították, míg az Excel esetében ez a jelentős tudományos visszhang ellenére sem történt meg. Ennek megfelelően az Excelt tudományos felhasználásra nem, oktatásra azonban ajánlják a szerzők. Az Excel kétségtelen és messze legfontosabb előnye ugyanakkor, hogy az Office csomag elterjedése miatt szinte mindenhol megtalálható. Általános elérhetősége egyben azt is jelenti, hogy akár mikro- és kisvállalatok – amelyek a drága, és folyamatosan friss verziókkal jelentkező

¹ Pl.: BMDP, SPSS, SAS, STATISTICA, MINISTAT, MINITAB, EVIEWS, stb.

² Hill R. C., Griffiths W. E., Lim G. C. [2008].

³ Ld.: Baczoni Pál [2007], Bártfai Barnabás [2002].

⁴ Az Excel Windows változata 1987-ben jelent meg.

⁵ Ld.: Knüsel, [1998], [2002], [2005].

⁶ Ld.: McCullough-Willson, [1999], [2002].

⁷ Ld.: Yalta, [2008]. A felsorolt hibákat az Excel 2007 sem küszöbölte ki. Pl. mi a valószínűsége, hogy 1000-szer feldobunk egy érmét, és abból maximum 1 fej lesz. Nyilván a való életben nem sok értelme van ennek a valószínűségnek, tudományos munkák esetén azonban lehet jelentősége.

szoftvereket nem képesek megvásárolni – elemzési eszköztárát is erősítheti. Megemlítjük továbbá azt a fontos tény, hogy a statisztika oktatásában ma már Magyarországon, a nagyobb egyetemeken és főiskolákon az Excel, mint táblázatkezelő szoftver elterjedt, főként könnyű elérhetősége okán. Az első könyv e témakörben Magyarországon Rappai Gábor: Üzleti statisztika Excellel⁸ c. műve volt. Ismereteink szerint csak az Excel alapszolgáltatásainak használata terjedt el az oktatásban és az üzleti életben Magyarországon⁹, pedig – mint arról már szó esett – az Excel ennél többre képes, lehet batch file-okat, kötegelt parancsállományokat (a továbbiakban parancsfájlokat, illetve programokat) készíteni. Internetes keresés¹⁰, és a rendelkezésünkre álló szakkönyvek feldolgozása alapján¹¹ megállapítottuk, hogy az USA-ban igen elterjedtek a parancsfájlok, bár legtöbbször csak korlátozott szolgáltatásokat nyújtanak. A további szolgáltatásokat külön meg kell fizetni, azokat a könyvekhez mellékelt CD-k nem tartalmazzák. Rátérve az alkalmazási lehetőségekre, véleményünk szerint az adatelemzés öt szintje oldható meg az Excellel: Az *első szint* az, amikor a Függvény beszúrása varázslót (ikont) használjuk, tehát beépített statisztikai, matematikai és trigonometriai, mátrix, adatbázis, stb. függvényeket alkalmazunk. A *második szint*, amikor az Eszközök - Adatelemzés¹² menüpont szolgáltatásait (pl. korrelációanalízis, regresszió) használjuk. A *harmadik szint*, amikor magunk írunk konkrét adatsorhoz vagy adatsorokhoz képleteket, mivel nem minden feladathoz áll rendelkezésre megírt függvény. A *negyedik szint* az, amikor parancsfájlokat készítünk – vagyis a harmadik szintet általánosítjuk – aminek felhasználásával az általunk megadott adatbázis terjedelméig (ez az adatbázisok sajátosságainak¹³ függvényében 25 - 10000 megfigyelés) új adatbázisok felhasználásával korlátlan számban számításokat végezhetünk a programozott képletek, illetve függvények alkalmazásával. Gyakran igen sok számítást kell elvégezni. Eben az esetben az idővel való takarékos gazdálkodás a cél, mert gyakran a harmadik szintnél egy feladatsor számításainak elvégzése több óra, vagy több nap, amit a parancsfájlok felhasználásával egy perc alatt el lehet végezni. Az *ötödik szint* az, amikor a feladat a hagyományos módon nem oldható meg. Erre példa a CES termelési függvény, ahol a változók száma több mint a rendelkezésre álló egyenletek száma. A feladat a legjobban illeszkedő függvény paramétereinek a megkeresése¹⁴. A logisztikus és egyéb speciális trendfüggvények esetében a függvényeket nem lehet lineárisra transzformálni, a cél megkeresni azokat a paramétereket, amelyek mellett az illesztés a legpontosabb¹⁵. A logisztikus regressziós függvények sem linearizálhatók, de iterációs eljárással, a paraméterek változtatásával a paraméterek becsülhetők, meghatározható egy olyan függvény, ahol a többszörös determinációs együttható a legnagyobb. Az Excel a Visual Basic for Applications (VBA) felhasználásával programozható, így ezek a feladatok egy iterációs eljárással megoldhatók. A negyedik és ötödik szint további előnye az, hogy szakértői értékelésre is felhasználhatók, vagyis javaslatot lehet tenni a különböző modellek elfogadására vagy elutasítására, továbbá kiküszöbölhetőek az Excel fordítási és tartalmi hiányosságai. Éppen ezért, és az eddig felsoroltak miatt gondoltuk úgy, hogy érdemes lenne olyan Excel alkalmazásokat létrehozni, melyek megkönnyítik a tanultak elsajátítását, dinamikusak, a felhasznált képletek „könnyen leolvashatók”, megkönnyítik a feladatmegoldást, és didaktikusak. A hallgatóknak lehetőségük nyílik a nagy mennyiségű számítási folyamat „mögé nézni”. További nagy előnye a következőkben ismertetett módszernek az, hogy az érdeklődő hallgatók – amennyiben valamilyen speciális módszer alkalmazására van szükségük, a bemutatott programok alapján, vagy azok módosításával – elkészíthetik saját, testhezálló Excel fájljaikat is. A munkalapokat egységes szerkezetben építettük fel. A változtatható, illetve megadható vagy megadandó adatokat sárga mezők jelölik, az eredményeket pedig egységes struktúrában, illetve szóhasználatával kívántuk megjeleníteni. A megértéshez szükséges végeredmények, és az egyes cellák számításához használt képletek valamennyi cella esetén láthatóak. Természetesen a képletek, függvények olvasásához alapvető táblázatkezelési ismeretek elengedhetetlenek, ezzel a számítás menete követhetővé válik. Szintén nagyon fontos, hogy egyetlen cella, vagy vezérlőelem (Checkbox, legördülő menü stb.) megváltoztatása az eredmények azonnali változását vonja maga után, és mindezt –

⁸ Rappai Gábor [2001].

⁹ Pl. Balázsne Mócsai Andrea-Csetényi Arthur [2003], Jánosa András [2005].

¹⁰ Ld.: pl. statistiXL, ami 30 napos ingyenes változat, utána meg kell venni, regresszió-, faktor- és klaszter-analízist is számol, elérhetősége az interneten: <http://www.statistixl.com/> (1.8 verzió: 2009 december 31)

¹¹ Evans James R. [2007], Aczel, Amir D. [2002], Berenson, Mark L. – Levine David M. – Krehbiel Timothy C. [2006].

¹² Az Eszközök – Bővítménykezelő – Data Analysis Toolpak bejelölése után.

¹³ Pl. hisztogram 25, korfa 101, elemi műveletek 5 ezer stb.

¹⁴ Ld.: ces1.xls

¹⁵ Ld.: Kehl Dániel – Dr. Sipos Béla [2009]. és logisztikusregresszio.xls

hála a gyors számítási sebességnek – azonnal elérhetjük. Rappai Gábor az informatikai támogatottsággal és az Excel felhasználásával kapcsolatban a következőket írta: „meggyőződésem szerint a legszélesebb körben rendelkezésre álló támogatóeszköz használata a legindokoltabb”¹⁶. A modernizáció jelentőségére hívja fel a figyelmet Kovács Péter tanulmánya is¹⁷, aki a Szegedi Tudományegyetemen bevezetett tanterven keresztül mutatja be a szegedi modellt, ami szintén erősen támaszkodik az Excelre. Úgy gondoljuk, hogy az általunk felvázolt, Excel alapú oktatás az egyik, természetesen nem kizárólagos irány lehet a jövőben. Rappai Gábor dékán javaslatára 2006-ban kezdtük meg a fejlesztő munkát. Az általunk írt oktatási segédlet „felhasználóbarát” stílusban íródott, csak annyi matematikai képletet tartalmaz, ami az Excel parancsfájlok megértéséhez és a feldolgozáshoz, az eredmények értelmezéséhez feltétlenül szükséges és széleskörű hazai és nemzetközi adatbázist dolgoz fel, ami a szakmai megértést elősegíti. Az oktatási segédlet függelékében felhívjuk a figyelmet arra és bemutatjuk, hogy hogyan lehet a feldolgozott adatsorokat az interneten megkeresni és letölteni. Oktatási segédletünkben azokat az Excel parancsfájlokat mutatjuk be, melyek elkészítését feladatul tűztük ki, és amelyek felölelik a statisztika illetve ökonometria három fontos területét; 1. egyszerű elemzések: viszonyszámok számítása, grafikonok készítése, empirikus eloszlások elemzése és elemi statisztikai műveletek; 2. dekompozíciós és sztochasztikus idősorlemzés fontosabb statisztikai módszerei; 3. korreláció- és regressziószámítás, sztochasztikus kapcsolatok elemzése és egyes speciális alkalmazások: pl. késleltetett regressziós modellek, CES-függvények, logisztikus regressziós függvények. Az oktatási segédlet megértéséhez szükséges elméleti háttér nagy része megtalálható a Pintér József – Rappai Gábor (szerkesztő) [2007]: Statisztika.¹⁸ c. BSC tankönyvben, amire az Excel parancsfájlok kidolgozása során támaszkodtunk, ezért csak az Excel parancsfájlok megértéséhez feltétlenül szükséges elméleti ismereteket és képleteket ismertetjük. Az oktatási segédletet a BSC és ezen felül az MSC, MBA és PHD képzésben is használjuk, géptermes gyakorlati oktatás keretében. Azokat az Excel parancsfájlokat, amelyek nem képezik a BSC alapképzés tananyagát, a könyvben *-gal jelöljük. Munkánk során értékes segítséget kaptunk Hunyadi László emeritus egyetemi tanártól (Corvinus Egyetem) és Rédey Katalin nyugdíjas egyetemi adjunktustól (Pécsi Tudományegyetem, Közgazdaságtudományi Kar). Segítségüket ezúton is köszönjük. Az Excel alkalmazásának témájával bővebben foglalkozó kézikönyv (Kehl Dániel – Dr. Sipos Béla: Excel parancsfájlok felhasználása a statisztikai elemzésekben) és az Excel parancsfájlok (BSC.zip, MSC.zip és SABL.zip) a PTE honlapján a KTK-GMI Publikációk - Sipos Béla internet címen letölthetők.^{19 20}

A Szerzők

¹⁶ Rappai Gábor [2008]: 840.

¹⁷ Kovács Péter [2008b]

¹⁸ Pintér József – Rappai Gábor (szerkesztő) [2007].

¹⁹ <http://www.gmi.ktk.pte.hu/index.php?mid=33#SiposB>

²⁰ Kiss Tibor-Sipos Béla: EXPS for Windows. 1998 szoftver 12 módszerrel végzi el a becslést. Ld.: Kiss Tibor – Sipos Béla [2000].

Bevezetés, az Excel beállításai, az Excel parancsfájlok használata során

Office 2003:

Módosítások: Eszközök – Beállítások – Biztonság – Makró védelem - közepesre állítani. A megnyitáskor a Makró használatát engedélyezni kell. Javasoljuk, a felhasználóknak, hogy az eredeti fájlt őrizzék meg és más néven lementett fájllal dolgozzanak.

Eszközök – Bővítménykezelő - Analysis Tool Pak, bejelölni. Érdemes a többi Bővítménykezelőt is bejelölni.(x)

Csak ez az Excel fájl legyen megnyitva, a módosítások után le kell menteni és be kell zárni a fájlt és újra meg kell nyitni.

Új adatbázis bevitele: a sárga mezőben lévő adatok cserélhetők, általában a munkalapon található adatok törlése gomb segítségével. Az üres sárga mezőbe az új adatok másolását a következőképpen kell elvégezni: először törölni kell az adatokat, az új adatbázist kijelölni - másolni- és irányított beillesztéssel belül - értéket választani. Az új adatbázist a felhasználónak értelemszerűen elő kell készítenie vagy be kell írni az üres sárga mezőbe.

A logisztikus trendek és regressziós függvények, ARIMA valamint a simit (exponenciális simítás) esetében még a következő beállításokra van szükség:

Eszközök – Bővítménykezelő - Solver beikszelni, vagy ha be van jelölve kieszedni a bejelölést, kilépni, belépni és újra bejelölni a Solvert. (A többi bővítményt is célszerű bejelölni)

Továbbá:

Eszközök – Makró - Visual Basic Editor - Tools (felül) - References - Solver legyen bejelölve. Megoldható úgy is, hogy Alt+F11 Tools, Preferences, és ki kell jelölni (pipa jel) a SOLVER feliratot, ha nem volt bejelölve. Ha nem működik a program újra: Eszközök – Makró - Visual Basic Editor - Tools (felül) - References – Solvert bejelölni.

Office 2007:

Makrók beállítása: Kattintson a Microsoft Office gombra, majd a *Programnév* beállításai gombra (Excel Options), ahol a *Programnév* az éppen használt alkalmazás neve, például Az Excel beállításai (alul található). Kattintson az Adatvédelmi központ (Trust Center) elemre, majd Az Adatvédelmi központ beállításai (Trust Center Settings) gombra végül a Makróbeállítások (Macro Settings) elemre. Kattintson a kívánt beállításra, a választás: Az összes makró engedélyezése. (Enable all macros) Minden választás után Ok.

Kattintson a Microsoft Office gombra, majd a *Programnév* beállításai gombra, (Excel Options) ahol a *Programnév* az éppen használt alkalmazás neve, például Az Excel beállításai (alul található). Kattintson a Bővítmények (Add-Ins) gombra, majd válassza a Kezelés – Excel bővítményeket alul, az ugrást választva a Bővítményeket bejelölheti (Manage: Excel Add-Ins, Go, megjelenik: Analysis Tool Pak, érdemes a többit is bejelölni.). A Bővítményeket az Excel installálja. A Bővítmények megjelennek: Adatok - Adatelemzés ikonnál. (Data – Data Analysis)

A logisztikus trendek és logisztikus regressziós függvények valamint a simit (exponenciális simítás) ARIMA, esetében még a következő beállításokra van szükség:

Eszközök – Bővítménykezelő - Solver beikszelni, vagy ha be van jelölve kieszedni a bejelölést, kilépni, lementeni, belépni és újra bejelölni a Solvert. (A többi bővítményt is célszerű bejelölni) Továbbá:

Eszközök – Makró - Visual Basic Editor - Tools (felül) - References - Solver legyen bejelölve. Megoldható úgy is, hogy Alt+F11 Tools, Preferences, és ki kell jelölni (pipa jel) a SOLVER feliratot, ha nem volt bejelölve.

A másik elérési lehetőség: Kattintson a Microsoft Office gombra, majd a *Programnév* beállításai gombra, (Excel Options) ahol a *Programnév* az éppen használt alkalmazás neve, például Az Excel beállításai (alul található). Jelölje be: Fejlesztőeszközök lap megjelenítése a szalagon. Megjelenik a Fejlesztő eszközök szalag, azon belül Visual Basic -Tools-References-Solvert be kell jelölni. Kilépés után mindig menteni kell.

Körkörös hivatkozás esetén, ha iterációt végez az Excel, az Excel által javasolt módosítás: Az Excel beállításai – Képletek - Közelítés engedélyezése.

Office 2010:

Beállítások: Fájl - Beállítások. Az Excel beállításai, innen azonos a beállítások módosítása az Office 2007-ben leírtakkal. Kattintson az Adatvédelmi központ (Trust Center) elemre, majd Az Adatvédelmi központ beállításai (Trust Center Settings) gombra végül a Makróbeállítások (Macro Settings) elemre.

Kattintson a kívánt beállításra, a választás: Az összes makró engedélyezése. (Enable all macros) Minden választás után Ok.

Kattintson a Bővítmények (Add-Ins) gombra, majd válassza a Kezelés – Excel bővítvényeket alul, az ug-rást választva a Bővítvényeket bejelölheti (Manage: Excel Add-Ins, Go, megjelenik: Analysis Tool Pak, érdemes a többi is bejelölni.). A Bővítvényeket az Excel installálja. A Bővítvények megjelennek: Ada-tok - Adatelemzés ikonnál. (Data – Data Analysis)

Körkörös hivatkozás esetén, ha iterációt végez az Excel, az Excel által javasolt módosítás: Az Excel beál-lításai – Képletek - Közelítés engedélyezése.

A logisztikus trendek és logisztikus regressziós függvények valamint a simit (exponenciális simítás) ARIMA, esetében még a következő beállításokra van szükség:

Eszközök – Bővítvénykezelő - Solver beikszelni, vagy ha be van jelölve kieszedni a bejelölést, kilépni, lementeni, belépni és újra bejelölni a Solvert. (A többi bővítvényt is célszerű bejelölni) Továbbá:

Eszközök – Makró - Visual Basic Editor - Tools (felül) - References - Solver legyen bejelölve. Megold-ható úgy is, hogy Alt+F11 Tools, Preferences, és ki kell jelölni (pipa jel) a SOLVER feliratot, ha nem volt bejelölve.

Az Excel beállításai. Menüszalag testreszabása-Fő lapok-Jelölje be: Fejlesztőeszközök. Ok. Megjelenik a Fejlesztő eszközök szalag, azon belül Visual Basic-Tools-References-Solvert be kell jelölni. Kilépés után mindig menteni kell.

Technikai tudnivalók az Office 2010, 2007 (Office 2003) használata során.

Logisztikustrendek.xls először a 12 munkalapon (Bertalanffytól Hubbertig) kell törölni az adatokat, majd a 12 munkalapba beilleszteni az új adatokat. Ezt úgy is meg lehet egyszerűbben és gyorsabban csinálni, hogy az első munkalapon (Bertalanffy) az összes munkalapot kijelöljük az egér műveletekkel (bal gomb majd jobb gomb és a menüsorból, a minden munkalapot kijelölt válasszuk) Az utolsó „Ciklus” munkala-pot nem szabad kijelölni, mert az adatok behívása programozva van, ezért itt alkalmazzuk a Shift+kattintást, az utolsó munkalapon, ahol még változtatni akarunk, vagyis a Ciklus munkalap előtt a Hubbertnél, így kiválasztottuk az összes munkalapot az utolsó, „Ciklus” kivételével, majd lehet törölni, illetve beilleszteni. A Ciklus munkalap színe ekkor pirosra vált, a kijelölt 12 munkalap sárga színe pedig fehérre vált.

A törlést a Delete billentyűvel végezzük a sárga mezőben található adatok kijelölése után, tehát ne hasz-náljuk az „adatok törlése” vezérlőelemet, ami mindegyik munkalapon megtalálható. A törlés után megint mindegyik munkalapot kijelöljük, kivéve az utolsó „ciklus” munkalapot, az előbb leírtak szerint, majd vá-lasztjuk az új adatállományt (idő és adatsor), az utasítások: kijelöl, másol majd irányított beillesztés – ér-ték és az adatokat az első munkalapba (Bertalanffy) beillesztjük. A másolás és beillesztés (másik Excel fájlból, nem a logisztikustrendek.xls parancsfájlból történjen a másolás.) előtt a kijelölést az előbb leírtak szerint (Shift+kattintás a Hubert munkalapon) meg kell csinálni. Másolni egy másik Excel fájlból kell az adatokat, mert csak akkor illeszti be mindegyik munkalapra a kiválasztott adatsort.

Fontosabb alapismeretek: táblázat: sorok-oszlopok. Cella, aktív cella. A cellák téglalap alakú halmaza tar-tomány. Munkalapok és munkafüzetek. A munkafüzetnek fájl nevet adunk. A munkalap mérete Excel 2007-ben 16384 oszlop és 1048576 sor. Képlet (pl. = A2/B2) másolása fogantyúval. A relatív cellahivat-kozás esetében a képlet másolásakor a másolás irányának megfelelően módosul a képlet. Abszolút cella-hivatkozásnál a cella címe a másolásakor nem változik, ekkor a táblázatkezelő a cella tényleges helyét tá-rolja. A sor és oszlopkoordináta elé \$ jelet kell tennünk, vagy az abszolút cellahivatkozás érdekében az F4 funkcióbillentyűt használhatjuk. (pl.: =A1*\$E\$1 az E1 az abszolút hivatkozás.)

A cella tartalma lehet: szöveg, szám, képlet.

Cella tartalmának javítása F2 funkcióbillentyűvel is történhet.

Minden függvénynél: Súgó a függvényről, leírást, képleteket és mintapéldát ad.

Cella, ábra különböző részei stb. kijelöl (klikkelés) bal egérgomb, majd jobb egérgomb, felajánlja a vál-toztatási lehetőségeket, amit az adott helyen választani lehet.

Egyéb technikai tudnivalók:

Interneten amerikai adatok esetében, ha szövegfájlról van szó, beillesztés előtt a Regionális beállításoknál a nyelvet és az országot magyarról amerikaira (USA) kell átállítani, majd az adatok beillesztése (beillesz-tés vagy irányított beillesztés-szöveg opciót választva) után vissza lehet állítani a regionális beállításokat.

Példa: <http://www.measuringworth.com/datasets/interestrates/result.php>

Excel 2007 segítség a felhasználónak:

<http://office.microsoft.com/hu-hu/excel-help/ujdonsagok-a-microsoft-office-excel-2007-programban-HA010073873.aspx>

Diagram készítés:

<http://office.microsoft.com/hu-hu/excel-help/diagram-keszítése-HP001233728.aspx?CTT=3>

Adatok beillesztése pdf fájlból:

Lementeni kell, majd megnyitni a pdf fájlt, utána lehet másolni.

Adatok kijelölése-másolás (Ctrl+C) beillesztés a jegyzetömbbe (notepad) lementeni *.txt formátumba.

Excel 2007- Adatok- Szövegből megnyitni a *.txt fájlt, kiválasztani a 28592 Közép európai ISO, - tovább – szóközt választani –megjelenik az adatokat elválasztó vonal- beillesztés.

Szöveg fájlok beillesztése Excelbe:

Pl.

<http://www.measuringworth.com/datasets/DJA/result.php>

DATA SETS

Daily DJA

Kijelölés kezdő év: 1896 Okt. 7. Végdátum megadja: 2011 Febr. 28.

Az adatok kijelölése: Utolsó jobb oldali adatot megkeresni és Shift+jobb egér gombra klikkelni. Beillesztés, Ctrl+F csere, pontot vesszőre. (kezdőlap-keresés és kijelölés-csere.)

Fontosabb matematikai és trigonometrikus függvények

Függvény	Leírás
ABS	Egy szám abszolút értékét adja eredményül.
KOMBINÁCIÓK	Adott számú objektum összes lehetséges kombinációinak számát számítja ki.
KITEVŐ	Az e adott kitevőjű hatványát adja eredményül.
FAKT	Egy szám faktoriálisát számítja ki.
LN	Egy szám természetes logaritmusát számítja ki.
LOG	Egy szám adott alapú logaritmusát számítja ki.
LOG10	Egy szám 10-es alapú logaritmusát számítja ki.
MDETERM	Egy tömb mátrix-determinánsát számítja ki.
INVERZ.MÁTRIX	Egy tömb mátrix inverzét adja eredményül.
MSZORZAT	Két tömb mátrix-szorzatát adja meg.
MARADÉK	Egy szám osztási maradékát adja eredményül.
HATVÁNY	Egy szám adott kitevőjű hatványát számítja ki.
SZORZAT	Argumentumai szorzatát számítja ki.
QUOTIENT	Egy hányados egész részét adja eredményül.
RADIÁN	Fokot radiánná alakít át.
VÉL	Egy 0 és 1 közötti véletlen számot ad eredményül.
RANDBETWEEN	Megadott számok közé eső véletlen számot állít elő.
KEREKÍTÉS	Egy számot adott számú számjegyre kerekít.
KEREKÍTÉS.LE	Egy számot lefelé, a nulla felé kerekít.
KEREKÍTÉS.FEL	Egy számot felfelé, a nullától távolabbra kerekít.
SERIESSUM	Hatványsor összegét adja eredményül.
ELŐJEL	Egy szám előjelét adja meg.
SIN	Egy szög szinuszt számítja ki.
GYÖK	Egy szám pozitív négyzetgyökét számítja ki.

RÉSZÖSSZEG	Lista vagy adatbázis részösszegét adja eredményül.
SZUM	Összeadja az argumentumlistájában lévő számokat.
SZUMHA	A megadott feltételeknek eleget tevő cellákban található értékeket adja össze.
SZUMHATÖBB	Több megadott feltételnek eleget tevő tartománycellák összegét adja eredményül.
SZORZATÖSSZEG	A megfelelő tömbelemek szorzatának összegét számítja ki.
NÉGYZETÖSSZEG	Argumentumai négyzetének összegét számítja ki.
SZUMX2BÖLY2	Két tömb megfelelő elemei négyzetének különbségét összegzi.
SZUMX2MEGY2	Két tömb megfelelő elemei négyzetének összegét összegzi.
SZUMXBÖLY2	Két tömb megfelelő elemei különbségének négyzetösszegét számítja ki.
TAN	Egy szám tangensét számítja ki.
CSONK	Egy számot egészszé csonkít.

Fontosabb statisztikai függvények

Függvény	Leírás
ÁTL.ELTÉRÉS	Az adatpontoknak átlaguktól való átlagos abszolút eltérését számítja ki.
ÁTLAG	Argumentumai átlagát számítja ki.
ÁTLAGA	Argumentumai átlagát számítja ki (beleértve a számokat, szöveget és logikai értékeket).
ÁTLAGHA	A megadott feltételnek eleget tevő tartomány celláinak átlagát (számtani közepét) adja eredményül.
ÁTLAGHATÖBB	A megadott feltételeknek eleget tevő cellák átlagát (számtani közepét) adja eredményül.
BÉTA.ELOSZLÁS	A béta-eloszlás függvényt számítja ki.
INVERZ.BÉTA	Adott béta-eloszláshoz kiszámítja a béta eloszlásfüggvény inverzét.
BINOM.ELOSZLÁS	A diszkrét binomiális eloszlás valószínűségértékét számítja ki.
KHI.ELOSZLÁS	A khi-négyzet-eloszlás egyszélű valószínűségértékét számítja ki.
INVERZ.KHI	A khi-négyzet-eloszlás egyszélű valószínűségértékének inverzét számítja ki.
KHI.PRÓBA	Függetlenségvizsgálatot hajt végre.
MEGBÍZHATÓSÁG	Egy statisztikai sokaság várható értékének megbízhatósági intervallumát adja eredményül.
KORREL	Két adathalmaz korrelációs együtthatóját számítja ki.
DARAB	Megszámolja, hogy argumentumlistájában hány szám található.
DARAB2	Megszámolja, hogy argumentumlistájában hány érték található.
DARABÜRES	Egy tartományban összeszámolja az üres cellákat.
DARABTELI	Egy tartományban összeszámolja azokat a cellákat, amelyek eleget tesznek a megadott feltételnek.
DARABHATÖBB	Egy tartományban összeszámolja azokat a cellákat, amelyek eleget tesznek több feltételnek.
KOVAR	A kovarianciát, azaz a páronkénti eltérések szorzatának átlagát számítja ki.
KRITBINOM	Azt a legkisebb számot adja eredményül, amelyre a binomiális eloszlásfüggvény értéke nem kisebb egy adott határértéknél.
SQ	Az átlagtól való eltérések négyzetének összegét számítja ki.
EXP.ELOSZLÁS	Az exponenciális eloszlás értékét számítja ki.
F.ELOSZLÁS	Az F-eloszlás értékét számítja ki.

INVERZ.F	Az F-eloszlás inverzének értékét számítja ki.
ELŐREJELZÉS	Az ismert értékek alapján lineáris regresszióval becsült értéket ad eredményül.
GYAKORISÁG	A gyakorisági vagy empirikus eloszlás értékét függőleges tömbként adja eredményül.
F.PRÓBA	Az F-próba értékét adja eredményül.
GAMMA.ELOSZLÁS	A gamma-eloszlás értékét számítja ki.
INVERZ.GAMMA	A gamma-eloszlás eloszlásfüggvénye inverzének értékét számítja ki.
GAMMALN	A gamma-függvény természetes logaritmusát számítja ki.
MÉRTANI.KÖZÉP	Argumentumai mértani középértékét számítja ki.
NÖV	Exponenciális regresszió alapján ad becslést.
HARM.KÖZÉP	Argumentumai harmonikus átlagát számítja ki.
HIPERGEOM.ELOSZLÁS	A hipergeometriai eloszlás értékét számítja ki.
METSZ	A regressziós egyenes y tengellyel való metszéspontját határozza meg.
CSÚCSOSSÁG	Egy adathalmaz csúcosságát számítja ki.
NAGY	Egy adathalmaz k-adik legnagyobb elemét adja eredményül.
LIN.ILL	A legkisebb négyzetek módszerével az adatokra illesztett egyenes paramétereit határozza meg.
LOG.ILL	Az adatokra illesztett exponenciális görbe paramétereit határozza meg.
INVERZ.LOG.ELOSZLÁS	A lognormális eloszlás inverzét számítja ki.
LOG.ELOSZLÁS	A lognormális eloszlásfüggvény értékét számítja ki.
MAX	Az argumentumai között szereplő legnagyobb számot adja meg.
MAX2	Az argumentumai között szereplő legnagyobb számot adja meg (beleértve a számokat, szöveget és logikai értékeket).
MEDIÁN	Adott számhalmaz mediánját számítja ki.
MIN	Az argumentumai között szereplő legkisebb számot adja meg
MIN2	Az argumentumai között szereplő legkisebb számot adja meg, beleértve a számokat, szöveget és logikai értékeket.
MÓDUSZ	Egy adathalmazból kiválasztja a leggyakrabban előforduló számot.
NORM.ELOSZL	A normális eloszlás értékét számítja ki.
INVERZ.NORM	A normális eloszlás eloszlásfüggvénye inverzének értékét számítja ki.
STNORMELOSZL	A standard normális eloszlás eloszlásfüggvényének értékét számítja ki.
INVERZ.STNORM	A standard normális eloszlás eloszlásfüggvénye inverzének értékét számítja ki.
PEARSON	A Pearson-féle korrelációs együtthatót számítja ki.
PERCENTILIS	Egy tartományban található értékek k-adik percentilisét, azaz százalékosztályát adja eredményül.
SZÁZALÉKRANG	Egy értéknek egy adathalmazon belül vett százalékos rangját (elhelyezkedését) számítja ki.
VARIÁCIÓK	Adott számú objektum k-ad osztályú ismétlés nélküli variációinak számát számítja ki.
POISSON	A Poisson-eloszlás értékét számítja ki.
VALÓSZÍNŰSÉG	Annak valószínűségét számítja ki, hogy adott értékek két határérték közé esnek.
KVARTILIS	Egy adathalmaz kvartilisét (negyedszintjét) számítja ki.

SORSZÁM	Kiszámítja, hogy egy szám hányadik egy számsorozatban.
RNÉGYZET	Kiszámítja a Pearson-féle szorzatmomentum korrelációs együtthatójának négyzetét.
FERDESÉG	Egy eloszlás ferdeségét határozza meg.
MEREDEKSÉG	Egy lineáris regressziós egyenes meredekségét számítja ki.
KICSI	Egy adathalmaz k-adik legkisebb elemét adja meg.
NORMALIZÁLÁS	Normalizált értéket ad eredményül.
SZÓRÁS	Egy statisztikai sokaság mintájából kiszámítja annak szórását.
SZÓRÁSA	Egy statisztikai sokaság mintájából kiszámítja annak szórását (beleértve a számokat, szöveget és logikai értékeket).
SZÓRÁSP	Egy statisztikai sokaság egészéből kiszámítja annak szórását.
SZÓRÁSPA	Egy statisztikai sokaság egészéből kiszámítja annak szórását (beleértve számokat, szöveget és logikai értékeket).
STHIBAYX	Egy regresszió esetén az egyes x-értékek alapján meghatározott y-értékek standard hibáját számítja ki.
T.ELOSZLÁS	A Student-féle t-eloszlás értékét számítja ki.
INVERZ.T	A Student-féle t-eloszlás inverzét számítja ki.
TREND	Lineáris trend értékeit számítja ki,
RÉSZÁTLAG	Egy adathalmaz középső részének átlagát számítja ki.
T.PRÓBA	A Student-féle t-próbához tartozó valószínűséget számítja ki.
VAR	Minta alapján becslést ad a varianciára.
VARA	Minta alapján becslést ad a varianciára (beleértve számokat, szöveget és logikai értékeket).
VARP	Egy statisztikai sokaság varianciáját számítja ki.
VARPA	Egy statisztikai sokaság varianciáját számítja ki (beleértve számokat, szöveget és logikai értékeket).
WEIBULL	A Weibull-féle eloszlás értékét számítja ki.
Z.PRÓBA	Az egyszélű z-próbával kapott valószínűségértéket számítja ki.

Fontosabb adatbázis-kezelő függvények

Függvény	Leírás
AB.ÁTLAG	A kijelölt adatbáziselemek átlagát számítja ki.
AB.DARAB	Megszámolja, hogy az adatbázisban hány cella tartalmaz számokat.
AB.DARAB2	Megszámolja az adatbázisban lévő nem üres cellákat.
AB.MEZŐ	Egy adatbázisból egyetlen olyan rekordot ad vissza, amely megfelel a megadott feltételeknek.
AB.MAX	A kiválasztott adatbáziselemek közül a legnagyobb értéket adja eredményül.
AB.MIN	A kijelölt adatbáziselemek közül a legkisebb értéket adja eredményül.
AB.SZORZAT	Az adatbázis megadott feltételeknek eleget tevő rekordjaira összesorozza a megadott mezőben található számértékeket, és eredményül ezt a szorzatot adja.
AB.SZÓRÁS	A kijelölt adatbáziselemek egy mintája alapján megbecsüli a szórást.
AB.SZÓRÁS2	A kijelölt adatbáziselemek teljes sokasága alapján kiszámítja a szórást.
AB.SZUM	Összeadja a feltételnek megfelelő adatbázisrekordok mezőoszlopában a számokat.
AB.VAR	A kijelölt adatbáziselemek mintája alapján becslést ad a szórásnégyzetre.
AB.VAR2	A kijelölt adatbáziselemek teljes sokasága alapján kiszámítja a szórásnégyzetet

Az Analysis ToolPak betöltése.

Az Analysis ToolPak betöltését követően elérhetővé válik az Adatelemzés parancs az Adatok lap Elemzés csoportjában.

Elérhetővé válik többek között: **Adatok - Adatelemzés**

Variancia - analízis. Egy- és kéttényezős.

Korreláció - analízis.

Kovariancia - analízis.

Leíró statisztika.

Exponenciális simítás.

Mozgóátlag.

Regresszió (felhasználható trendszámításra $x=t$).

Mozgóátlagolás.

Diagramok készítése Excel 2007.²¹

Diagramok használatával grafikus formátumban jeleníthetők meg a numerikus adatsorok, így könnyebben értelmezhetők a nagy mennyiségű adatok, valamint a különböző adatsorok közötti kapcsolatok.

1. Az Excel alkalmazásban történő diagramkészítés első lépése a numerikus adatok munkalapon való fel-tüntetése. Ha más munkalapon megvannak az adatok, akkor Másolás CTRL+C beillesztés CTRL+V. Ki-vágás CTRL+X. Kezdőlap – **Vágólap menüszalagon** is megtalálhatók a felsorolt műveletek.

2. Adatok kijelölése:

Kijelölhetők egérműveletekkel vagy:

Adott sor vagy oszlop celláit úgy is kijelölheti, hogy az első cella kijelölését követően a CTRL+SHIFT+NYÍL (sorok esetén a JOBB vagy BAL, oszlopok esetén a FEL vagy LE nyílbillentyű) billentyűkombinációt használja.

Megjegyzés: Adatokat tartalmazó sor vagy oszlop esetén a CTRL+SHIFT+NYÍL billentyűkombináció a legutolsó használatban lévő celláig bővíti a kijelölést, a billentyűkombináció másodszori lenyomása pedig a teljes sort illetve oszlopot jelöli ki

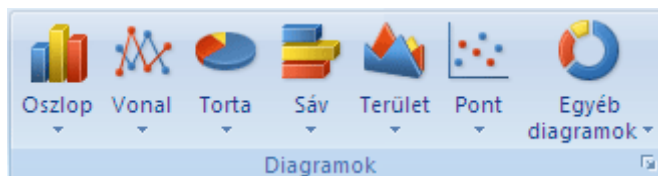
3. Diagram készítése.

Ezután az adatok diagramba emeléséhez válasszon diagramtípust: Nyissa meg a **Beszúrás menüszalagot** és a **Diagramok elemcsoportban** válassza ki a diagramtípust.

Hajtsa végre a megfelelő műveletet a Beszúrás lap Diagram csoportjában: Jelölje ki az ábrázolandó ada-tokat, majd:

Jelöljön ki egy diagramtípust, majd kattintson a diagram használni kívánt altípusára.

Ha látni szeretné az összes elérhető diagramtípust, a Diagram beszúrása párbeszédpanel megjelenítéséhez kattintson valamelyik diagramtípusra, majd a Minden diagramtípus parancsra; a nyílakra kattintva gör-gebbe végig a használható diagramtípusokat és -altípusokat, majd kattintson arra, amelyiket alkalmazni szeretné.



Tipp Ha egy diagramtípus vagy -altípus fölé viszi az egérmutatót, megjelenik a diagramtípus nevét mu-tató elemleírás. A felhasználható diagramtípusokról a Diagramtípusok gyűjteménye című témakörben ol-vashat.

A diagramot kijelölve változtatni lehet a diagramtípusokon és -altípusokon.

²¹ <http://office.microsoft.com/hu-hu/excel-help/diagram-keszítése-HP001233728.aspx?CTT=3>

4. Diagrammeszközök - Módosítási lehetőségek -Elrendezés.
Be lehet írni pótlólag a diagramcímet, a tengelycímekeket stb.

Trendek számítása diagramra illesztett trendvonalal.

Ha meglévő adatokhoz szeretne előre jelezni trendet, létrehozhat egy trendvonalat a diagramban. Ha például van egy diagram az Excelben, amely az év első néhány hónapjának értékesítési adatait ábrázolja, hozzáadhat egy olyan trendvonalat a diagramhoz, amely mutatja az értékesítés általános trendjét (növekvő, csökkenő vagy stagnáló), vagy előrejelzi az elkövetkező hónapok trendjét.

Ez az eljárás azt feltételezi, hogy a diagram már korábban elkészült meglévő adatok alapján. Ellenkező esetben a Diagram készítése című témakörben olvashat erről bővebben.

Kattintson a diagramra.

Jelölje ki azt az adatsort, amelyhez trendvonalat vagy mozgó átlagot szeretne illeszteni.

Kattintson az Elrendezés lap Elemzés csoportjában a Trendvonal gombra, majd válassza ki a regressziós trendvonal vagy mozgó átlag kívánt típusát.

A beállítások szerkesztéséhez és a regressziós trendvonal vagy mozgó átlag formázásához kattintson a jobb gombbal a trendvonalra, és válassza a helyi menü Trendvonal formázása parancsát.

Adja meg a kívánt beállításokat, vonalakat és effektusokat.

Ha a Polinomiális típust választotta, írja be a Fokszám mezőbe a független változó legmagasabb hatványkitevőjét.

Ha a Mozgóátlag típust választotta, írja be a Periódus mezőbe a mozgó átlag kiszámításához használt időszakok számát.

Megjegyzések

Az Alapul szolgáló csoportban láthatja az összes olyan adatsort, amelyeknél használhatja a trendvonalakat. Ha trendvonalat szeretne használni egy sorban, kattintson a nevére, és válassza ki a megfelelő beállításokat.

Ha mozgó átlagot illeszt egy pontdiagramhoz, a mozgó átlag értéke a diagrambeli x értékek sorrendjétől függ. Előfordulhat, hogy a kívánt eredmény eléréséhez meg kell adnia az x értékek sorrendjét, mielőtt a mozgó átlag számításába fogna.

1 Egyszerű adat-elemzések: viszonyszámok számítása és grafikus ábrázolás

Az egyszerű adatelemzések tárgyalása előtt, röviden ismertetjük azokat az alapvető statisztikai fogalmakat, amelyeket a továbbiak során használni fogunk. **Statisztikai sokaságnak** nevezzük a statisztikai megfigyelés tárgyát képező egyedek összességét. A sokaság egyedei lehetnek élőlények, szervezetek, tárgyak, események, képzett egységek, stb. Amennyiben a sokaság egy adott időpontra (ezt az időpontot szokás ún. *eszmei időpontnak* is nevezni) vonatkozó állapotát vizsgáljuk, álló sokaságról beszélünk. A mozgó sokaság folyamatot fejez ki, ebből következően *időtartamra* értelmezhető. **Statisztikai ismérveknek** nevezük a sokaság egyedeire vonatkozó tulajdonságokat, jellemzőket. A sokaság egységei (egyedei) az ismérvek hordozói. Az ismérv lehetséges kimenetelei (változatai) az **ismérvváltozatok**. Az ismérveknek négy típusát különböztethetjük meg:

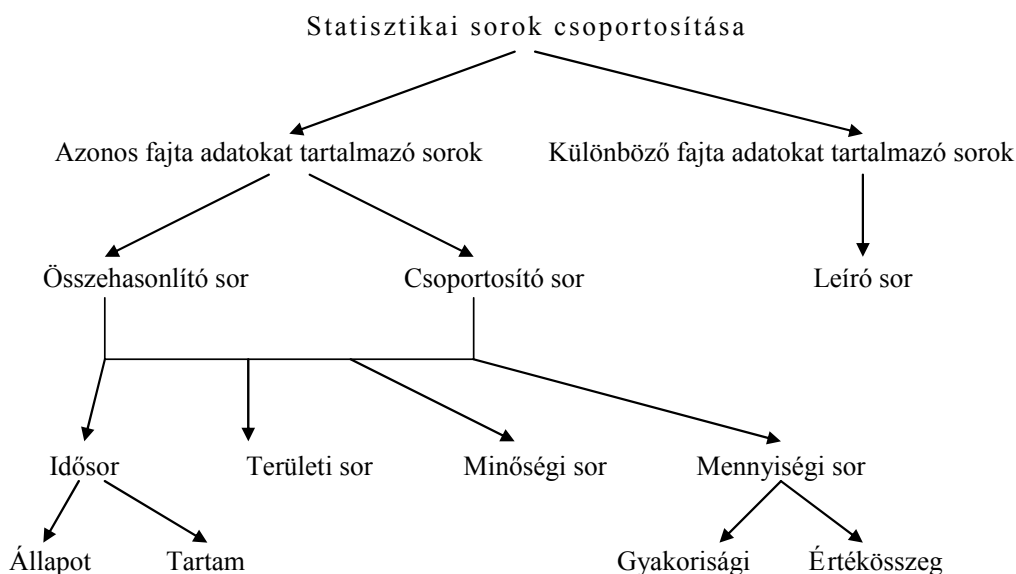
1. Az **időbeli ismérv** a sokaság egységeire nézve valamilyen időbeli elhatárolást ad.
2. A **területi ismérv** a sokaság egységeire nézve valamilyen földrajzi elhatárolást ad.
3. A **minőségi ismérv** a sokaság egységeire jellemző verbálisan leírható tulajdonság.
4. A **menyiségi ismérv** azon tulajdonságokat jelenti, melyek számadatokkal leírhatóak, és valamilyen mérés vagy számlálás eredményei.

A **csoportosítás**, avagy **osztályozás** a statisztikai sokaságnak valamely ismérv szerinti tagolása, rendszerezése. A csoportosítás lényegében azt is jelenti, hogy a statisztikai sokaságot minőségileg különböző részekre, csoportokra bontjuk, és így tanulmányozzuk szerkezetét, felépítését. A csoportosítás a gyakorlatban úgy történik, hogy a csoportképző ismérv alapján az ismérv változatainak megfelelően a sokaság egyes tagjait a konkrét ismérvváltozatokhoz rendeljük. A **csoportképző ismérvek** a sokaság lényeges tulajdonságait tükrözik, ezek alapján lehetőség nyílik a sokaságon belül az alapvető különbségek, eltérések feltárására, elemzésére. A csoportosításhoz felhasznált csoportképző ismérv változatai sok esetben adot-

tak: pl. nem, kor, beosztás és szakképzettség. A statisztikai nomenklatúrák fontos csoportképző ismérvek: pl. a tevékenységek azonosítása a TEÁOR-on (Tevékenységek Egységes Ágazati Osztályozási Rendszere), a termékeké az ITJ-n (Ipari Termékek Jegyzéke), a METJ-n (Mezőgazdasági, Erdészeti Termékek Jegyzéke), az ÉJ-n (Építményjegyzék), a SZATJ-n (Számítástechnikai Alkalmazási Termékek Jegyzéke) a szolgáltatásoké pedig a Szolgáltatások Jegyzékén (SZJ) alapul. Ismert nomenklatúra továbbá a Foglalkozások Egységes Osztályozási Rendszere (FEOR).

Fontos szempont a csoportosítás során, hogy az adatok *egyértelműen besorolható* legyenek. Ez annyit jelent, hogy valamennyi egyed egy és csak egy csoportba kerülhet. Egy statisztikai sokaság egyidejűleg több ismerv szerinti csoportosítását **kombinatív csoportosításnak** hívjuk. A statisztikai adatok feldolgozásának – a csoportosítás mellett – gyakorta alkalmazott másik elemi módszere az összehasonlítás.

Az **összehasonlítás** statisztikai adatok egymás mellé rendelését jelenti elemzési célból. Az összehasonlítással a mindennapi életünkben gyakran találkozunk, és szinte semmilyen megállapítást nem teszünk nélküle. *Összehasonlíthatónak* tekintjük azokat az adatokat, amelyek csak olyan tényezők miatt térnek el egymástól, amelyeknek a szerepét éppen kutatjuk. A statisztikai adatok valamilyen ismerv szerinti felsorolását **statisztikai sornak** nevezzük. A sorok csoportosítás eredményeként, vagy összehasonlítás céljából állíthatók elő. Az azonos fajta adatokból álló statisztikai sorok – amelyek általában csoportosító vagy összehasonlító sorok – az ismérvek típusai szerint is osztályozhatók. Így beszélhetünk időbeli, minőségi, mennyiségi és területi statisztikai sorokról. A különböző fajta, de egymással összefüggő adatokat tartalmazó sort leíró sornak nevezzük. Az álló sokaság (stock) adatait tartalmazó idősor az ún. **állapot idősor**, melynek jellemzője, hogy minden megfigyelt adata egy eszmei időponthoz tartozik, valamint ez a sor csak *összehasonlító* jellegű statisztikai sor lehet. A mozgó sokaságot tartam idősorral tudjuk szemléltetni. A **tartam idősor** sajátossága, hogy a statisztikai adatok a sokaság flow-jellegéből adódóan mindig időtartamhoz kötődnek, adatai így akár összesíthetők is. A statisztikai sorok változatait az alábbiakban foglaljuk össze:



A statisztikai alapfogalmak áttekintése után röviden tárgyaljuk a viszonzyszámok számítását és a grafikus ábrázolást.

Viszonzyszámnak nevezzük két egymással kapcsolatban álló statisztikai adat hányadosát.

A viszonzyszám képlete:

$$V = \frac{A}{B}$$

ahol:

A – a viszonyítás tárgya,

B – a viszonyítási alapja, más néven bázisa.

A viszonzyszámok legfontosabb fajtái:

- intenzitási,
- megoszlási,
- koordinációs,

- dinamikus viszonyszám.

Az **intenzitási viszonyszám** általában különböző, de egymással kapcsolatban álló statisztikai adatok hányadosa, ebből következően a mértékegysége a számláló és a nevező mértékegységéből képződik.

Megoszlási, illetve **koordinációs viszonyszámok**at a sokaság csoportosítását követően számíthatunk. Az előbbieket egy részsokaságot hasonlítanak az egészhez, az utóbbiak két részsokaságot viszonyítanak egymáshoz. A viszonyítás eredményét vagy ún. együttthatós formában, vagy százalékos formában szokás megadni.

A **dinamikus viszonyszámok** két időszak vagy időpont adatainak hányadosai, melyeket általában százalékos formában adunk meg. A viszonyítás alapját képező időpontot, időszakot bázisidőszaknak, míg a viszonyítás tárgyát tárgyidőszaknak szokták nevezni. Amennyiben kettőnél több időszak vagy időpont adataival rendelkezünk a viszonyítás alapja lehet állandó vagy változó; ezen utóbbi esetben általában a megelőző időszak (időpont) adatát tekintjük viszonyítási alapnak. Az első esetben bázisviszonyszámokat, a második esetben láncviszonyszámokat számítunk.

A bázisviszonyszám képlete, ha az idősor első megfigyelését tekintjük bázisnak akkor:

$$B_t = \frac{y_T}{y_1}$$

A láncviszonyszám képlete:

$$L_t = \frac{y_T}{y_{T-1}}$$

Az időbeli összehasonlításokra a bázis- és láncviszonyszámok egyaránt alkalmasak. Míg a bázisviszonyszámok a fejlődés (változás) relatív mérésére, addig a láncviszonyszámok a fejlődés (változás) ütemének számszerűsítésére szolgálnak.

Időbeli összehasonlításokra az ún. **differentia-képzést** is használhatjuk. Ebben az esetben két szomszédos időszak vagy időpont adatának a különbségét képezzük, melyet elsőrendű differenciának nevezünk. Képlete:

$$D_t = y_T - y_{T-1}$$

Az idősor adatainak szigorúan kötött a felsorolási rendje, mely egyben azt is jelzi, hogy a szomszédos adatok különbségeinek és hányadosainak számításánál, mindig a későbbi adatból vonjuk ki korábbi, illetve a későbbi adatot osztjuk a korábbival.

Az időbeli összehasonlításokra – amennyiben kettőnél több időszak vagy időpont adatát ismerjük – gyakran használjuk az átlagos abszolút és relatív változás mutatóit is.

Az időszakról időszakra, illetve időpontról időpontra történő változások (a D_t és L_t értékek) átlagos értékét kiszámítva jutunk az előbb említett mutatószámokhoz. Az **átlagos abszolút változás** mutatója – melyet azokban az esetekben alkalmazunk, ha feltételezhető, hogy a változások a vizsgált időszakban abszolút nagyságukat tekintve állandóságot mutatnak – az alábbi képlettel határozható meg:

$$\bar{D} = \frac{(y_2 - y_1) + (y_3 - y_2) + \dots + (y_T - y_{T-1})}{T - 1} = \frac{y_T - y_1}{T - 1}$$

Ha az egymást követő megfigyelések hányadosai mutatnak viszonylagos állandóságot, akkor az **átlagos relatív változás** mutatóját célszerű kiszámítani:

$$\bar{L} = \sqrt[T-1]{\frac{y_2}{y_1} \times \frac{y_3}{y_2} \times \dots \times \frac{y_T}{y_{T-1}}} = \sqrt[T-1]{\frac{y_T}{y_1}}$$

A **grafikus ábra** az elemzések és közlések fontos eszköze.²² A grafikus ábrák felhívják a figyelmet a statisztikai adatok által reprezentált jelenségek alapvető jellemzőire, a főbb arányokra, tendenciákra, összefüggésekre. Az ábrázolás célja lehet a jelenségek közötti kapcsolatok vizsgálata, a leíró célú alkalmazás, a döntés előkészítés alátámasztása, az elemzések eredményeiről történő tájékoztatás, közlés. A grafikus ábrázolás lényege az összehasonlítás, ezért az arányokat érzékelteti és nem az abszolút nagyságokat.

A grafikus ábrázolással szemben támasztott követelmények:

²² Ld.: Hunyadi László [2002]

- A legmegfelelőbb ábrázolási módot kell kiválasztani, amit a vizsgált jelenséget bemutató adatok jellege, illetve a jelenségek között lévő kapcsolat természete dönt el.
- A kiválasztott grafikus ábra legyen egyszerű, arányos, áttekinthető és kifejező.
- Minden grafikon két részből álljon: ábra és magyarázó jelölések (cím, skála, jelmagyarázat, alkalmazott mértékegységek, forrás).

A grafikus ábrák lehetnek a teljesség igénye nélkül:

Egyszerűbb statisztikai ábrák (diagramok):

- Oszlop (téglalap)-,
- vonal-,
- kör-,
- szalag-,
- XY, vagy pontdiagram.

Összetett (kifejezetten statisztikai műveletek eredményeként keletkező) diagramok:

Gyakorisági sorok elemzésére szolgáló ábrák:

- hisztogram,
- gyakorisági poligon,

Speciális szalagdiagram az ún. korfa, mely a demográfiai elemzésekben használatos.

1.1 A dinamikus viszonyszámok parancsfájl működése

A dinamikus viszonyszámok.xls fájl alkalmazása során bevihető az időváltozó, ezen kívül egyszerre 12 különböző adatsor, maximum 2000 adat. Az összehasonlítás műveletének két alapvető módja elvégezhető: az összehasonlítandó adatokból történő hányados-, illetve különbségképzés. Ki kell választani a számlálót és a nevezőt, ezt követően a program kiszámolja a hányados értékét. Ha a számláló és a nevező hányadosa kis számérték, akkor a szorzótényező (pl. 10, 100, 1000) megadásával a hányadost beszorozza. Ha százalékos, illetve ezrelékes formát alkalmazunk akkor a szorzótényező 100 illetve 1000. Ha erre nincs szükség akkor a szorzótényező értéke alapértelmezésben 1. Ha nem akarunk hányados képezni, akkor a hányados képzés-nevező –menüben a „nincs osztás”-t válasszuk. A különbség képzésnél, ha az közgazdaságilag értelmezhető, ki kell jelölni a kisebbítendő és a kivonandó adatsort. A program mindkét esetben elkészíti az ábrát, továbbá kiszámítja a bázis- és láncviszonyszámokat. A különbségképzés (elsőfokú differencia) oszlop a szomszédos adatok különbségeit számítja ki.

Ha az átlagos abszolút és átlagos relatív változás mutatóit akarjuk kiszámítani, akkor az átlagos változás munkalapon meg kell adni a T , az y_T és az y_1 értékeket és a program kiszámítja az átlagos abszolút és relatív változás mutatóit.

Példa a dinamikus viszonyszámok.xls alkalmazására, a munkatermelékenység változásának elemzése.

Vezessük be az alábbi jelöléseket:

- Y = előállított termék értéke millió Ft-ban, (millió Ft.).
- X_{11} = a fizikai dolgozók ledolgozott (munkahelyen eltöltött illetve fizetett) munkaideje órában. (óra)
- X_{12} = a fizikai dolgozók átlagos állományi létszáma főben. (fő)
- X_{13} = a foglalkoztatottak átlagos állományi (teljes: fizikai + nem fizikai) létszáma főben. (fő)

A rendelkezésünkre álló vállalati adatsor²³: 1996-2007 éves adatok.

A munkaügyi tényezők hatását az alábbi egyenlet mutatja, ahol a munkatermelékenységet a termelés és a foglalkoztatottak átlagos állományi létszámának hányadosaként határoztuk meg, (millió Ft/fő):

$$\frac{Y}{X_{13}} = \frac{Y}{X_{11}} \frac{X_{11}}{X_{12}} \frac{X_{12}}{X_{13}}$$

A fenti összefüggés esetén vizsgálható:

A munkaidő egy órájára jutó termelés nagysága (millió Ft/óra):

²³ Feltételezett adatok.

$$\frac{Y}{X_{11}}$$

A munkaidő és a fizikai dolgozók átlagos állományi létszámának az aránya (óra/fő):

$$\frac{X_{11}}{X_{12}}$$

A fizikai dolgozók átlagos állományi létszáma egy főjére jutó termelés nagysága (millió Ft/fő):

$$\frac{Y}{X_{12}}$$

A fizikai dolgozók átlagos állományi létszámának és a foglalkoztatottak átlagos állományi létszámának az aránya*100 (%)

$$\frac{X_{12}}{X_{13}}$$

A számítások végeredményének bemutatása a dinamikus viszonzszámok.xls fájl felhasználásával.

1-1. tábla: A termelékenység (Y/X₁₃) változása 1996-2007 között

Idő	Bázisviszonzszámok	Láncviszonzszámok
1996	1,000	-
1997	1,018	1,018
1998	0,992	0,975
1999	1,003	1,010
2000	1,050	1,047
2001	1,123	1,069
2002	1,222	1,088
2003	1,172	0,960
2004	1,215	1,037
2005	1,256	1,033
2006	1,297	1,032
2007	1,298	1,001

Átlagos abszolút változás
0,232
Átlagos relatív változás
1,024

Az 1-1. táblában lévő adatok:

Y = előállított termék értéke millió Ft.-ban, (millió Ft.).

X₁₃ = a foglalkoztatottak átlagos állományi (teljes: fizikai + nem fizikai) létszáma főben. (fő)

A rendelkezésünkre álló vállalati adatsor: 1996-2007 éves adatok.

Gyakorló feladatok. (dinamikus viszonzszámok.xls)

1. Elemezze az egy főre jutó GDP alakulását 2000-es \$-ban Magyarország és az EU országok (27 ország) adatai alapján az 1969-2007 közötti adatok felhasználásával.
2. Elemezze az egy főre jutó GDP alakulását 2000-es \$-ban az USA-ban 1790-2006 közötti adatok felhasználásával A számláló: Reál GDP 2000-es \$-on (milliárd \$), a nevező: Népesség 1000 fő, a szorzótényező 1000.
3. A termelékenység1996-2007.xls fájl felhasználásával végezze el a számításokat az előzőekben bemutatott képletek felhasználásával.

4. Elemezze az egy főre jutó GDP alakulását 2005-es \$-ban Magyarország és az EU országok (27 ország) adatai alapján az 1969-2008 közötti adatok felhasználásával.²⁴

1.2 Ábrák készítése Excel parancsfájl működése

Az ábrák készítése parancsfájl segítségével a következő ábrákat nyerhetjük: vonal/oszlop/ hisztogram/kör/szalag/XY/korfa. Mindegyik ábrának megfelel egy munkalap. A kiválasztott ábrának megfelelő munkalapot megnyitva és az ábrázolandó adatsort bemásolva, az ennek megfelelő ábrát a program szolgáltatja. Az **oszlop- és vonal-diagramot** általában idősorok ábrázolására használjuk. Az idősorok esetén a grafikonra kattintva, lehetőség van a trendvonal felvételére (lineáris, féllogaritmikus, polinomiális, hatvány, exponenciális és mozgóátlagos trend). Tartamidősorok esetén a vízszintes tengelyen intervallumok szerepelnek, a jelenséget pedig célszerű ezen intervallumok fölé rajzolt téglalapokkal (oszlopokkal), tehát oszlopdiagrammal bemutatni. Állapot idősorok esetén az időbeli ismerv értékei egy-egy időponthoz tartoznak, ezért célszerű ábrázolásuk egy-egy pont, az egyes pontokat egyenesekkel össze is lehet kötni.²⁵

A gyakorisági sorok ábrázolására a **hisztogramot** tartalmazza a program. A hisztogram munkalapba az ábrázolandó osztályközös gyakorisági sort bemásolva, az ábrát a program szolgáltatja. Az osztályközös gyakorisági sor képzése az elemi műveletek parancsfájlban található. Hisztogramnak nevezzük azt a grafikus ábrát, amely olyan - a derékszögű koordináta rendszerben hézag nélküli oszlopdiagramot jelent, ahol az oszlopok alapját az osztályközök hossza, a magasságát pedig a gyakoriságok adják. A hisztogram oszlopainak **területe** arányos a gyakoriságokkal, ezért az egyenlő hosszúságú osztályközök esetén az ábrázolás nem okoz gondot. Eltérő hosszúságú osztályközök esetén – mivel a hosszabb osztályközhez aránytalanul nagyobb gyakoriság tartozna – ezért módosítani kell a gyakoriságokat.

A **kördiagram** a minőségi (területi) sorok ábrázolásának általános eszköze. A minőségi ismerv szerinti megoszlás eredményét a kördiagramban megjelenő megoszlási viszonyszámok segítségével szemlélteti. A kördiagramot általában az adatok relatív gyakoriságának ábrázolására használják. A teljes kör jelképezi a 100%-ot, és az egyes adatok relatív gyakoriságát ábrázoló körcikkhez tartozó középponti szög arányos a relatív gyakorisággal. Természetesen a kördiagram akkor mutatja jól a megoszlásokat, ha az ismerv kevés változattal rendelkezik. Ha a minőségi (területi) ismerv változatainak a száma nagy, akkor a **szalagdiagram** az ábrázolás javasolt módszere.

Pontdiagramot (XY ábrát) két egymással összefüggésben lévő mennyiségi ismerv értékeinek ábrázolására használjuk.

A **korfa** olyan speciális szalagdiagram, amelynek egyik oldalán a férfiak, a másik oldalán a nők számának megfelelő hosszúságú vízszintes sávok mutatják az adott életkorú népesség számát vagy %-os megoszlását korcsoportonként. Az alkalmazott, az adatszolgáltatásban szokásos korcsoportok: 0-4, 5-9, 10-14, ... ,85-89, 90-94, 95-99, 100+ év. A népesedési helyzet vizsgálatának egyszerű, ugyanakkor szemléletes és látványos eszköze a **korfa**.

Gyakorló feladatok. (ábrák készítése.xls)

1. Magyarországi hosszú idősorok ábrázolása és elemzése. A KSH minden évben - 2001 óta - közli a hosszú idősorokat Excel formátumban is tartalmazó Statisztikai Évkönyv CD mellékletét. Ezek az idősorok 1960-tól tartalmazzák folyamatosan évenként mért adatokat. Ábrázolja és értékelje a 2006. évi Statisztikai Évkönyv CD mellékletében található hosszú (1960-2006) adatsorokat szakmai bontásban. A szakmai területeken belül többféle mutató időbeli alakulása vizsgálható:
 - 1.1. Népesség, népmozgalom mutatói.
 - 1.2. A háztartások jövedelme és fogyasztása, lakásépítés.
 - 1.3. Egy főre jutó élelmiszer- és tápanyagfogyasztás.
 - 1.4. Társadalombiztosítás, szociális ellátás.
 - 1.5. Egészségügy.
 - 1.6. Oktatás

²⁴ <http://www.ers.usda.gov/Data/Macroeconomics/>

²⁵ Hunyadi László [2002]: 29.

- 1.7. Kultúra.
- 1.8. Bűnözés.
- 1.9. Gazdasági aktivitás, bruttó hazai termék (GDP), beruházás.
- 1.10. A mezőgazdaság főbb mutatói.
- 1.11. Növények összes termése [ezer tonna].
- 1.12. Állattenyésztés.
- 1.13. Ipar.
- 1.14. Kereskedelem, turizmus.
- 1.15. Szállítás.
- 1.16. Posta és távközlés.
- 1.17. Az MNB interneten elérhető adatsorai²⁶ alapján ábrázolja és elemesse a rendszerváltást követően:

A fogyasztói árindex alakulását 1993-tól havi- és éves átlagos bontásban, piaci javak szerint is.

A külkereskedelmi termékforgalom havi alakulását áru főcsoportonként 1996-tól havi bontásban.

A fedezetlen bankközi forint kihelyezések havi átlagkamatlábainak alakulását 2000-től havi bontásban.

A devizában fennálló nettó adósság alakulását 1995 és 2008 között negyedéves bontásban.

2. Készítse el a világ tíz legnépesebb országára, a világra és Magyarországra a korfákat a 2000. 2025. és 2050. év adatai alapján. Értékelje a korfákból levonható következtetéseket.
3. Az amerikai elnök részére készített 2008. évi jelentés²⁷ Excel formátumban közreadott adatsorai közül ábrázolja és értékelje a következőket:
 - 3.1. B-2. A GDP (Real gross domestic product), az export és az import, a háztartások (Personal consumption expenditures, total) fogyasztása és a kormányzati fogyasztás (Government consumption expenditures and gross investment, total) alakulása 2000-es \$-ban. 1959-2007.
 - 3.2. B-35. A munkanélküliség rátájának alakulása a polgári területen dolgozóknál. Az agrár és a nem agrár területen dolgozó aktív népesség számának alakulása.28 (Unemployment rate, civilian workers Civilian population and labor force) 1929-2007. A hiányzó adatokat (1930-1931, 1934-1938) becsülje meg egyenletes növekedési ütemet feltételezve.
 - 3.3. B-36. A munkanélküliek számának alakulása összesen és nemek szerint. (Civilian employment and unemployment by sex and age). 1960-2007.
 - 3.4. B-42. A munkanélküliség ráta alakulása összesen és fő csoportok (nem, kor: 16-19, és 20- évek, származás, fehér és nem fehér) szerint. (Civilian unemployment rate). 1960-2007.
 - 3.5. B-51. A teljes ipari termelési index (2002=100%) alakulása. (Industrial production indexes, major industry divisions). 1959-2007.
 - 3.6. B-60. A fogyasztói árindex alakulása (1982-84=100) az összes termékre, az élelemre (food), a vasúti szállításra (transportation), orvosi ellátásra (medical care) vonatkozóan. (Consumer price indexes for major expenditure classes). 1960-2007.
 - 3.7. B-77. Fogyasztói teljes hiteltartozások (total consumer credit, millió \$-ban.) alakulása. (Consumer credit outstanding). 1959-2007.
 - 3.8. B-97. A farmok teljes jövedelmének alakulása milliárd \$-ban. (Farm income). 1945-2007.
4. Az MNB folyamatosan közli a forint napi árfolyamát a különböző valutákhoz képest²⁹.
 - 4.1. A 2007. év napi árfolyamait ábrázolja és elemesse a Ft/USD (USD= USA dollár), a Ft/CAD (CAD= kanadai dollár) és a Ft/EUR (EUR= euró) esetében.
 - 4.2. Ábrázolja³⁰ 1990. jan. 1. és 2008. ápr. 4. között a Ft/USD (USD= USA dollár) és a Ft/CAD (CAD= kanadai dollár) napi árfolyamának a változását.

A grafikonok az MNB honlapjáról is lekérhetők, különböző bontásban³¹.
5. Kördiagram készítése

²⁶ http://www.mnb.hu/engine.aspx?page=mnbhu_statisztikai_idosorok az adatokat a KSH szolgáltatja.

²⁷ Internetes elérés: Economic Report of the President. Statistical Tables. 2008:

<http://www.gpoaccess.gov/eop/download.html> A 2010-es adatok is letölthetők:

<http://www.gpoaccess.gov/eop/tables09.html>

²⁸ 1947-ig a 14 évnél idősebb, utána a 16 évnél idősebb népességre vonatkozó adatok.

²⁹ <http://www.mnb.hu/engine.aspx?page=arfolyamlekerdezes>

³⁰ Árfolyam1949-2007.xls

³¹ <http://www.mnb.hu/engine.aspx?page=arfolyamlekerdezes>

- 5.1. Foglalkoztatottak számának alakulása 2000 és 2006 között gazdasági ágak szerint Magyarországon.
- 5.2. Iskolai végzettség változása nemenként 1960 és 2006 között.

1.3 Az országonkénti korfa prognózis készítése 2050-ig Excel parancsfájl működése

Ezzel a parancsfájllal elemezhető a munkalapon szereplő 25 ország és a világ népességének adatállománya korcsoport és nemhez való tartozás szerint. Az egyes országok népességszáma általában az 1990-es évektől a 2050-ig előrejelzett értékekkel együtt évente rendelkezésre áll. Az adatokat az USA Népszámlálási Hivatala (US Census Bureau) hozza rendszeresen nyilvánosságra.^{32,33} A vizsgálni kívánt ország korfáit 2050-ig megrajzolja a program.

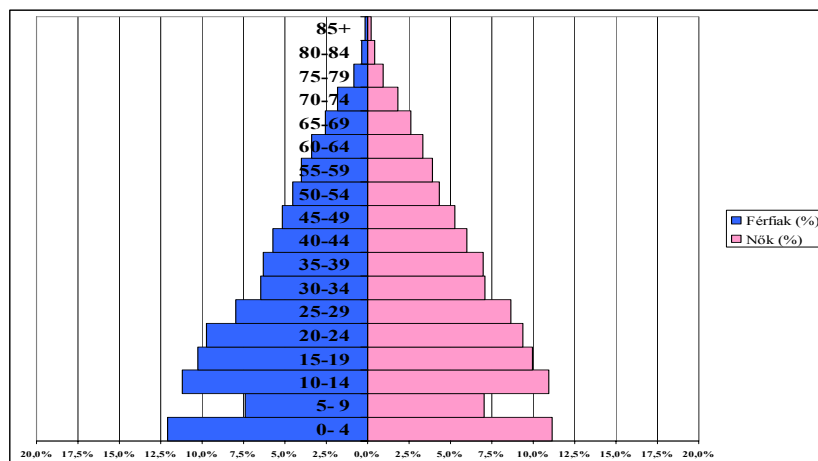
Az Excel parancsfájl működése:

1. Az első „ábra” nevű munkalapon 25 ország és a világ népessége közül választhatunk, a következő munkalapokon megtaláljuk az egyes országok népességi adatait. A választható országok, feltüntetve a rendelkezésre álló adatok kezdő évét:
 - Afganisztán (AFG) 1979-
 - Amerikai Egyesült Államok (USA) 1950-
 - Ausztrália (AUS) 1986-
 - Ausztria (AUT) 1991-
 - Banglades (BAN) 1991-
 - Belgium (BEL) 1989-
 - Bosznia-Hercegovina (BHV) 1991-
 - Brazília (BRA) 1971-
 - Bulgária (BUL) 1993-
 - Csehország (CZE) 1991-
 - Franciaország (FRA) 1990-
 - India (IND) 1991-
 - Indonézia (INA) 1980-
 - Japán (JPN) 1990-
 - Kanada (CAN) 1991-
 - Kína (CHN) 1990-
 - Lengyelország (POL) 1989-
 - Magyarország (HUN) 1924-, hiányoznak az 1942-1946 évek.
 - Egyesült Királyság (GBR) 1991-
 - Németország (GER) 1991-
 - Nigéria (NGR) 1953-
 - Oroszország (RUS) 1989-
 - Pakisztán (PAK) 1981-
 - Románia (ROM) 1992-
 - Szlovénia (SLO) 1991-
 - Világ (VILÁG) 1996-
2. A csúszda mozgatásával (egy kattintás egy év előre) nyomon követhetjük a korfa (férfi % és nő %) változását. Az évszámok 1950-2050 között változnak évi bontásban. Meg kell nézni melyik évtől vannak korfa adatok (pl. Magyarország 1924-2050) és a csúszdát az évszámhoz kell mozgatni. Az előtte lévő éveknél mivel nincs adat a korfa üres marad. Baloldalt látható az évszám, jobboldalt felül pedig a vizsgált ország vagy a világ adott, tehát éppen vizsgált évéhez tartozó korfa adatai, férfi, nő és összesen (mindkét nemre vonatkozóan) bontásban.

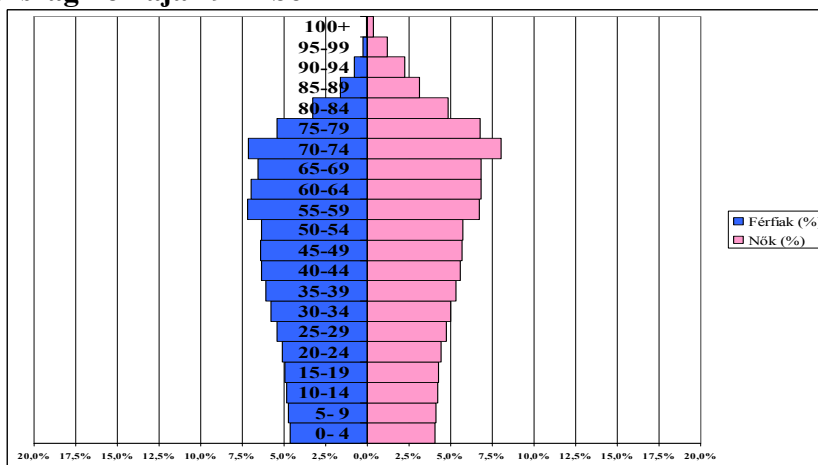
³² <http://www.census.gov/ipc/www/idb/summaries.html>

³³ 2009 10 24 után az adatszolgáltatás egyszerűbbé vált: Country ki kell választani az országot, Ctrl lenyomásával az összes év kiválasztható, utána Population Pyramids-t választva, mindegyik országra elkészíti a korfát és az utolsó korfa után az adatok letölthetők Excel/CSV formátumban.

Az alábbi 1-1. ábra mutatja Magyarország korfáját³⁴ 1924-ben és a 2050-re várható korfát a 1-2. ábra tartalmazza, látható, hogy a piramisból (1924) egy fordított piramis (2050) lesz, ami komoly veszélyeket prognosztizál.



1-1. ábra: Magyarország korfája 1924-ben



1-2. ábra: Magyarország prognosztizált korfája 2050-ben

Gyakorló feladatok. (országokénti korfa prognózis készítése 2050-ig.xls)

A csúszda mozgatásával vizsgálja meg a korfa változását a felsorolt országokban és értékelje a 2025 és 2050 évekre vonatkozó prognózis becsléseit.

Amerikai Egyesült Államok (USA) 1950-2050

Brazília (BRA) 1971-2050

Csehország (CZE) 1991-2050

Franciaország (FRA) 1990-2050

India (IND) 1991-2050

Kína (CHN) 1990-2050

Lengyelország (POL) 1989-2050

Románia (ROM) 1992-2050

Szlovénia (SLO) 1991-2050

Világ (VILÁG) 1996 - 2050

1.4 Nemzetközi összehasonlítások Excel parancsfájlok felhasználásával

A GDP-re, a népességszámra, a fogyasztói árindexre, a \$ árfolyamára vonatkozó hosszú idősorok országoként és régióként (216-231 ország és régió) 1969-2008 között álltak először rendelkezésre.³⁵ Az adatokat évente frissítik. Ennek alapján 7 Excel parancsfájlt készítettünk el. Parancsfájlonként először a ren-

³⁴ Az adatok forrása: <http://www.nepszamlalas.hu/>

³⁵ Az adatok forrása: <http://www.ers.usda.gov/Data/Macroeconomics/>

delkezésre álló adatok jelennek meg, az ezt követő 1. munkalapon ezekből két országot illetve régiót lehet kiválasztani és ennek alapján a program elkészíti azt a grafikon, amely ábrázolja 1969 és 2008 között a vizsgált mutató alakulását. A 2. munkalapon 10 országot, illetve régiót lehet kiválasztani a vizsgált mutató szerinti összehasonlításra. Az időtengely (a csúszda) mozgatójával nyomon követhetők a változások 1969 és 2008 között. A rendelkezésre álló adatok extrapolációt is tartalmaznak a 2009 és 2020/2030 közötti időszakra. Ezért a prognózisokat is meg lehet tekinteni 2009 és 2020 illetve 2030³⁶ között. Az idősorok és a prognosztizált időszak hosszát a parancsfájlok neve tartalmazza.

A letöltött adatok alapján az alábbi parancsfájlok kerültek kidolgozásra:

- *A gdpegyfőreeső1969-2008ésprognózis2030.xls*

Ez a parancsfájl az egy főre jutó GDP reál értékeket (\$/fő, 2005-ös \$-ban) tartalmazza országonként és régiókként (228 ország, a világ illetve régió).

- *A gdp1969-2008 ésprognózis2030.xls*

Ez a parancsfájl a GDP reál értékeket (milliárd \$, 2005-ös \$-ban) tartalmazza országonként és régiókként (230 ország, a világ illetve régió).

- *A népesség1969-2008 ésprognózis2030.xls*

Ez a parancsfájl 228 ország, a világ illetve régió esetében közli a népesség alakulását főben.

- *A gdpdefláció1969-2009 ésprognózis2020.xls*

Ez a parancsfájl a 227 ország illetve régió esetében közli a GDP deflációs indexének alakulását %-ban 2005-ös \$-ban.

- *A gdprészesedése1969-2008 ésprognózis2030.xls*

Ez a parancsfájl a A GDP reál értékek világtermelésből (world=100 %) való részesedést tartalmazza 2005-ös \$-ban országonként és régiókként (228 ország illetve régió).

- *A fogyasztóiárindex1969-2008ésprognózis2020.xls*

Ez a parancsfájl a 231 ország, a világ illetve régió esetében közli a fogyasztói árindex (CPI Consumer price index 2005=100%) alakulását %-ban.

- *A \$árfolyama1970-2008és prognózis2020.xls*

Ez a parancsfájl 216 ország illetve régió esetében közli a nemzeti valuták \$ árfolyamát (US =1) reálértéken 1970 és 2008 között. A névleges \$ árfolyamokat a fogyasztói árindexszel (CPI Consumer price index 2005 = 100 %) korrigálták.

A nemzetközi összehasonlítások Excel parancsfájlt 2011 januárjában a 2010 dec. 27-i adatok ismeretében frissítettük. Az új Excel parancsfájlok, amit közreadunk:

A történeti adatok (Historical Data Files) 1969-2010 között állnak rendelkezésre. 190 ország és 34 régió adatait közölték. Az adatokat évente frissítik. Ennek alapján 7 Excel parancsfájlt készítettünk el. Az első munkalap a vizsgált adatokat tartalmazza, az 1. munkalapon két országot illetve régiót lehet kiválasztani és a grafikon ábrázolja 1969 és 2010 között a vizsgált mutató alakulását. A 2. munkalapon 10 ország illetve régió vizsgált adatát lehet kiválasztani és az időtengely mozgatójával nyomon követhetők a változások 1969 és 2010 között. Az adatbázisok első rátekintésre áttekinthetetlen halmazt képeznek, hiszen több mint 8000 adatot tartalmaznak. A grafikus ábrázolás az összefüggések gyors áttekintését biztosítja. Az adatok extrapolációját (Baseline Data Files) is közlik a 2011 2030 közötti időszakra. Ezért a prognózisokat is meg lehet tekinteni 2011 és 2030 között, az éveket itt *-gal jelöltük.

Kidolgozók: Oxfordi Gazdasági Előrejelzések Hivatala (Global Insight, Project Link), Világbank (World Bank World Development Indicators), Nemzetközi Valutaalap (IMF, International Financial Statistics), Népszámlálási Hivatal nemzetközi adatbázisa, USA (Census Bureau International Population Database). A letöltött adatok alapján az alábbi parancsfájlok kidolgozásra kerültek:

Gdpegyfőreeső1969-2010ésprognózis2030.xls

Az egy főre jutó GDP reál értékeket (\$/fő) tartalmazza 2005-ös \$-ban országonként és régiókként (228 ország, a világ illetve régió). 1969 és 2010 között található a tényadatok és 2011 és 2030 között a prognosztizált becslés adatok. Az adatokat az első GDPperfü1969-2030 munkalap tartalmazza. Az 1GDPperfü1969-2030 munkalapon ki lehet választani két országot illetve régiót és a grafikon ábrázolja

³⁶ Az extrapolált időszak éveit *-gal jelöltük.

1969 és 2010 között a reál GDP/fő (\$/fő) alakulását, továbbá a 2011 és 2030 közötti prognózist, ahol a prognosztizált időszak éveit *-val jelöltük. A 2 GDPperffő1969-2030 munkalapon 10 ország illetve régió adatát lehet kiválasztani és az időtengely mozgatójával láthatók a változások 1969 és 2010 között, továbbá a 2011 és 2030 közötti prognózist nyomon lehet követni. A többi Excel parancsfájl hasonlóan működik, ezért az 1. és 2. munkalap tartalmának ismertetésétől a következőkben eltekintünk.

A gdp1969-2010 és prognózis2030.xls parancsfájl működése.

A GDP reál értékeket (milliárd \$) tartalmazza 2005-ös \$-ban országonként és régiókként (230 ország, a világ illetve régió). 1969 és 2010 között található a tényadatok és 2011 és 2030 között a prognosztizált becslés adatok, ahol a prognosztizált időszak éveit *-val jelöltük. Az adatokat a GDP munkalap tartalmazza.

A népesség1969-2010 és prognózis2030.xls parancsfájl működése.

228 ország, a világ illetve régió esetében közli a népesség alakulását főben. 1969 és 2010 között található a tényadatok és 2011 és 2030 között a prognosztizált becslés adatok, ahol a prognosztizált időszak éveit *-val jelöltük. Az adatokat a népesség munkalap tartalmazza.

A gdpdefláció1969-2010 és prognózis2030.xls parancsfájl működése.

227 ország illetve régió esetében közli a GDP deflációs indexének alakulását %-ban 2005-ös \$-ban. 1969 és 2010 között található a tényadatok és 2011 és 2030 között a prognosztizált becslés adatok, ahol a prognosztizált időszak éveit *-val jelöltük. Az adatokat a gdpdefláció munkalap tartalmazza.

A gdpárszesedése1969-2010 és prognózis2030.xls parancsfájl működése.

A GDP reál értékek világtermelésből (world=1) való részesedést tartalmazza 2005-ös \$-ban országonként és régiókként (228 ország illetve régió). 1969 és 2010 között található a tényadatok és 2011 és 2030 között a prognosztizált becslés adatok, ahol a prognosztizált időszak éveit *-val jelöltük. Az adatokat a gdpárszesedés munkalap tartalmazza.

A fogyasztóiárindex1969-2009 és prognózis2030.xls parancsfájl működése.

231 ország, a világ illetve régió esetében közli a fogyasztói árindex (CPI Consumer price index 2005=100%) alakulását %-ban. 1969 és 2009 között található a tényadatok és 2010 és 2030 között a prognosztizált becslés adatok, ahol a prognosztizált időszak éveit *-val jelöltük. Az adatokat a CPI-index munkalap tartalmazza.

A \$árfolyama1970-2010 és prognózis2030.xls parancsfájl működése.

222 ország illetve régió esetében közli a nemzeti valuták \$ árfolyamát (US =1) reálértéken 1970 és 2010 között. A névleges \$ árfolyamokat a fogyasztói árindexszel (CPI Consumer price index 2005 = 100 %) korrigálták. A 2011 és 2030 között található a prognosztizált becslés adatok, ahol a prognosztizált időszak éveit *-val jelöltük. Az adatokat a \$árfolyama munkalap tartalmazza.

Gyakorló feladatok. (Nemzetközi összehasonlítások Excel parancsfájlok)

A Gdpegyfőreeső1969-2010 és prognózis2030.xls, a fogyasztóiárindex1969-2009 és prognózis 2030.xls parancsfájl, a \$árfolyama1970-2010 és prognózis2030.xls alapján milyen lesz Magyarország helyzete: 2011-ben és 2030-ban. Viszonyítási alapok: Világ, USA, EU 27 ország, Új EU országok, környező országok: Ausztria, Csehország, Szlovákia, Románia, Ukrajna, Szlovénia és Horvátország.

2 Elemi műveletek a változókkal és empirikus eloszlások elemzése

A sokaság mennyiségi ismérvi szerinti megfigyelése során a sokaság minden egyes egyedére vonatkozóan egy számszerű adattal rendelkezünk. Ezek a mennyiségi ismérvtételek jelentik a megfigyelt adatbázist, melynek nagysága esetünkben maximum 5000 lehet. Ezt az adathalmazt rendezni, rendszerezni szükséges, hogy a vizsgált jelenségről általános, tömör jellemzést tudjunk adni. Ezt a célt szolgálja a kidolgozott parancsfájl. Az adatok-számítások munkalap az aktuális adatbázisra jellemző számlálás, rangsorolás, összegzés eredményeit jeleníti meg, továbbá középértékeket és kvantiliseket számít és meghatározza a szórási különböző mérőszámait. A parancsfájl működésének a leírása előtt röviden ismertetjük az általunk használt statisztikai ismeretanyagot.

2.1 Számlálás, rangsorolás, összegzés

A legegyszerűbb statisztikai művelet a **számlálás** vagyis annak meghatározása, hogy az adott változó szempontjából hány megfigyeléssel rendelkezünk. A számlálás végeredményét általában n -nel jelöljük, így már szemléltetni tudjuk a teljes adathalmazt:

$$X_1, X_2, \dots, X_n$$

illetve az adathalmaz általános elemét: x_i -t.

Rangsornak nevezzük a változóértékek növekvő, vagy csökkenő sorrendben történő felsorolását. A statisztikai módszertan rangsoron - főszabályként – emelkedő rangsort ért. A rangsorba rendezett értékeket, annak érdekében, hogy megkülönböztessük az egyszerű lajstromtól (felsorolástól) általában az indexértékek megkülönböztetésével jelöljük:

$$X_{(1)}, X_{(2)}, \dots, X_{(n)}$$

A mindennapi életben, így a statisztikai elemzésekben is kitüntetett szerepe van a legkisebb, illetve a legnagyobb ismérvértéknek, ezeket – akárcsak a köznyelvben – minimumnak, illetve maximumnak hívjuk, és

$$X_{\min} = X_{(1)}$$

$$X_{\max} = X_{(n)}$$

szimbólumokkal jelöljük.

A rangsorba rendezés során az eredeti adathalmaz számértékeihez ún. rangszámokat rendelünk. **Rangszám**nak nevezzük azt a pozitív egész számot, amely megmutatja, hogy egy konkrét adat hányadik az adathalmaz emelkedő rangsorában, vagyis

$$R_i = k, \text{ ha } x_i = x_{(k)}$$

Könnyen belátható, hogy a minimális érték rangszáma 1; a maximális pedig n , a rangszámok pedig a természetes számokkal egyenlők 1-től n -ig.

Ha egy-egy változóérték többször is előfordul a lajstromban, akkor valamennyi azonos értékhez azt az azonos rangszámot rendeljük, amely a sorban következik, de a következő (nagyobb) értékhez annnyival nagyobb rangszámot adunk, ahányszor előtte előfordult az azonosság. A másik megoldás az, hogy az azonos értékekhez rendelt rangszám nem a sorban következő, hanem egy képzett szám, melyet úgy képezzünk, hogy az azonos értékekhez rendelt rangszámok összege akkora legyen, mintha az értékek különböznenek. (pl. ha a 2-ik és a 3-ik érték azonos, akkor 2,5 és 2,5 rangszámot kapnak)

A legegyszerűbb statisztikai műveletek közé soroljuk az **összegzés (szummázás)** műveletét is. Összegzésnek nevezzük azt a folyamatot, mely során összeadjuk az adatbázisban szereplő változó értékeit, vagyis képezzük a változó értékösszegét:

$$x' = \sum_{i=1}^n x_i$$

2.2 Középértékek és kvantilisek

Középértéknek, az azonos fajta számszerű értékek tömegének közös jellemzőjét nevezzük. A középérték egyetlen értékkel tömören jellemzi a sokaságot a vizsgált mennyiségi ismérv szerint.

A középértékekkel szemben támasztott két legfontosabb követelmény:

- egyértelműen számíthatók és könnyen értelmezhetők legyenek
- közepes és tipikus értékek legyenek

A közepesség azt jelenti, hogy ne egy szélső mennyiségi ismérvérték legyen a közös jellemző, hanem valamilyen közepen elhelyezkedő, míg a tipikusság olyan értéket jelent, ami a sokaságban sokszor fordul elő. Középértékként többféle statisztikai jellemző ismeretes, amelyek természetesen nem egyformán felelnek meg a fenti követelményeknek.

A középértékek fajtái:

- Számított középértékek (átlagok), a számtani-, a harmonikus-, a mértani (geometriai)- és a négyzetes (kvadratikus) átlag
- Helyzeti középértékek, a módusz és a medián

A számított középértékeket az előforduló valamennyi érték felhasználásával, matematikai képlet, formula segítségével számítjuk. Az átlagok közül általánosan a számtani átlag használatos, a többi átlagot csak speciális esetekben használjuk.

A helyzeti középértékeket az előforduló értékek közül választjuk ki, az értékek elhelyezkedési rendje szerint. A mediánt ugyanúgy általánosan használjuk, mint a számtani átlagot. A módusz nem minden esetben határozható meg egyértelműen, így alkalmazása egyértelműen a gyakorisági sorok elemzéséhez köthető.

A **számtani átlag** az a szám, melyet az átlagolandó értékek helyébe téve, azok összege változatlan marad. Képlete szerint, a mennyiségi ismerv előforduló értékeinek összegét, az értékek számával osztjuk:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Az átlagot mindig az x_i értékek nagyságrendjében és mértékegységében kapjuk meg. A számítási módból látható, hogy a számtani átlagot akkor célszerű alkalmazni, ha az átlagolandó értékek összege ($\sum x$) értelmezhető. Mint a legegyszerűbben számítható és értelmezhető átlagot, számos más esetben is használni tudjuk. A számtani átlag nevezetes tulajdonságai közül kettőt emelünk ki:

1. Az átlagolandó értékeknek a számtani átlagtól mért algebrai összege zérus,

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

2. Az átlagolandó értékeknek a számtani átlagtól mért eltérés-négyzetösszege minimális,

$$\sum_{i=1}^n (x_i - \bar{x})^2 \rightarrow \min$$

A számtani átlag egyértelműen számítható és jól értelmezhető, közepes helyet foglal el az előforduló értékek között, de nem biztos, hogy tipikus érték. Számított jellegéből adódóan ugyanis, lehet, hogy ilyen érték nem is fordul elő.

A **medián** a rangsorba rendezett (x_i) mennyiségi ismervértékek közül a középső érték. Olyan jellemzője a sokaságnak a mennyiségi ismerv szerint, amelyiknél ugyanannyi kisebb, mint nagyobb érték fordul elő. A szó szoros értelmében közepes érték. A **medián** értékének megállapításához először az n számú x_i mennyiségi ismervértéket rangsorba rendezzük, és megkeressük a középen elhelyezkedő értéket.

A medián páratlan elemszámú adathalmaz esetén:

$$Me = x_{\left(\frac{n+1}{2}\right)}$$

A medián páros számú adat esetén a medián nem esik egybe egy konkrét megfigyeléssel, így ilyenkor, konvencionálisan a

$$Me = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

képlettel határozható meg.

A medián egyik fontos tulajdonsága, hogy minden ismervérték mediánnal történő helyettesítésekor elkövetett hibák abszolút értékben számított összege minimális lesz:

$$\sum_{i=1}^n |x_i - a| \rightarrow \min, \text{ ha } a = Me$$

Egy statisztikai sokaságban, ha az adatokat növekvő sorrendben rendeztük, megkereshetjük azt az ismervértéket (osztópontot) amelynél az ismérvek fele, negyede, tizede, százada stb. kisebb, a többi pedig nagyobb értékű. A **kvantilis**ek tehát olyan osztópontok, amelyek a rangsorba rendezett számszerű ismervértékek 2,3,4,...,k-ad részét jellemzik. Definíciónk szerint a j-edik kvantilis az a változóérték, amelynél az összes előforduló érték j/k-ad ($j=1,2,\dots,k-1$) része kisebb, illetve $1-(j/k)$ -ad része nagyobb. Vagyis

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(i)} \leq q_j^{(k)} \leq x_{(i+1)} \leq \dots \leq x_{(n)}$$

$$\frac{i}{n} = \frac{j}{k}$$

ahol $q_j^{(k)}$ a j-edik k-ad rendű kvantilis.

Az osztópontokat – a mediánál megismert módon – könnyen meg lehet határozni, a rangsorba rendezett sokaság megfelelő értékének kiválasztásával, illetve a két szomszédos érték átlagolásával.

A **kvartilisek** számításánál, az n számú rangsorolt értéket négy ($k=4$) egyenlő részre osztjuk, és az így nyert három ($k-1=3$) osztópontot, alsó (Q_1), középső (Q_2) és felső (Q_3) kvartilisnek nevezzük. A középső kvartilis egyben a medián ($Q_2=Me$). Az alsó kvartilis az az érték, amelynél az előforduló értékek egynegyede kisebb, háromnegyede nagyobb, míg a felső kvartilis értékénél az értékek háromnegyede kisebb, és egynegyede nagyobb. A nevezetes kvartilisek ($k=2, 3, 4, 5, 10, 100$) közül, a már említett kvartiliseken ($k=4$) kívül, a **deciliseket** ($k=10$) alkalmazzuk az Excel parancsfájlból.

2.3 Szóródási mérőszámok

A középértékek számításánál abból indultunk ki, hogy a mennyiségi ismérvek általában igen sokféle értéket vehetnek fel, és a célunk az, hogy egy olyan közös jellemzőt keressünk, amellyel az egyedi értékek helyettesíthetők. Azt, hogy e célunkat hogyan sikerül elérni, nagymértékben függ attól is, hogy a közös jellemző „mögött lévő” értékek mennyire különbözőek. Lehet, hogy körülbelül hasonló nagyságú, egymástól kevéssé eltérő értéket átlagolunk, de előfordulhat, hogy igen jelentős különbségeket sikerül kiegyenlíteni a középérték-számítással. A statisztikai elemzésekben ezt úgy mondjuk, hogy a sokaság vizsgált mennyiségi ismérve szerint kevésbé, vagy jobban szóródik. **A szóródás a mennyiségi ismérvértékek különbözőségét jelenti.** A szóródás ismert mutatószámai közül a következőket alkalmazzuk: a terjedelem, a szórás, a variancia (szórásnégyzet) és a relatív szórás.

A **terjedelem (range)**: – az előforduló legnagyobb és legkisebb érték különbségeként – a mennyiségi ismérvértékek rangsorából könnyen megállapítható,

$$T = x_{\max} - x_{\min}$$

Kijelöli annak az intervallumnak a nagyságát, amelyben az értékek előfordulnak.

A szóródás leggyakrabban alkalmazott mérőszáma a szórás. A **szórásnak** nevezzük az átlagolandó értékek számtani átlagtól való eltérésének négyzetes átlagát.

A **szórás képlete**:

$$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

A szórás az alábbi intervallumban vehet fel értékeket:

$$0 \leq \sigma \leq \sqrt{n-1} \times \bar{x}$$

A szórás négyzetét **varianciának** (σ^2) hívjuk. Önálló tartalommal bír, bizonyos statisztikai eljárásokban nagyon fontos szerepet tölt be.

Képlete:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

A **relatív szórás** mutatója a szórásnak a számtani átlaghoz viszonyított arányával fejezi ki a szóródást, amelyet százalékos formában is megadhatunk:

$$V_x = \frac{\sigma_x}{\bar{x}}$$

A relatív szórás mutatójának jelentőségét a különböző nagyságrendű, – és sokszor különböző mértékegységekkel is mért – átlagokkal és szórásokkal jellemzett sokaságok, összehasonlítása adja. Határai:

$$0 \leq V_x \leq \sqrt{n-1}$$

2.4 Az elemi műveletek parancsfájl működése

Az elemi műveletek parancsfájl két munkalapból áll. Az adatok-számítások munkalap az aktuális adatbázison elvégzi az elemi műveletek címszó alatt összefoglalt statisztikai elemzéseket. A segítség munkalap a parancsfájl használatához szükséges tudnivalókat tartalmazza, amiket a következőkben ismertetünk. A szemléltető példa: egy benzinkút egyik kútoszlopánál egy adott időszak – 4 óra – alatt kiszolgált benzin mennyisége literben, amit az események sorrendjében jegyeztek fel, ld. x_i sárga mezőben lévő oszlopot.

Az adatbázis maximum 5000 értéket tartalmazhat.

A felhasznált Excel Adatok - számítások munkalap eredményei és magyarázatok:

Az új adatbázis bevitele után a munkalap az alábbi eredményeket jeleníti meg.

Növekvő oszlop: az x_i adatsor növekvő sorrendben.

Csökkenő oszlop: az x_i adatsor csökkenő sorrendben.

A sorba rendezést a program az AZ↓ Rendezés növekvő ikon és a ZA↓ Rendezés csökkenő ikon használatával is elvégzi.

Rang oszlop: az x_i adatsor rangszámai, amelyek az Excel beépített függvénye felhasználásával készülnek: SORSZÁM (szám; hiv; sorrend)

Alapstatisztikák³⁷:

Megfigyelések száma: az adatbázisban szereplő megfigyelések száma. A mintapéldában $n = 60$.

Összeg = x_i adatsor összege, =SZUM(xi) = 2259 liter.

$$x' = \sum_{i=1}^n x_i = 27 + 5 + 27 + \dots + 28 = 2259$$

Minimum: a legkisebb előforduló érték az adatsorban: $x_{\min} = x_{(1)} = \text{MIN}(xi) = 5$ liter

Maximum: legnagyobb előforduló érték az adatsorban: $x_{\max} = x_{(n)} = \text{MAX}(xi) = 70$ liter

Középértékek:

Számtani átlag: az x_i értékek számtani átlaga

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{27 + 5 + 27 + \dots + 28}{60} = 37,65$$

=ÁTLAG(xi) = 37,65 liter

Medián: a növekvő rangsorba rendezett x_i értékek középső értéke.

$$Me = \frac{x_{\left(\frac{60}{2}\right)} + x_{\left(\frac{60}{2}+1\right)}}{2} = \frac{36 + 36}{2} = 36 \text{ liter.}$$

=MEDIÁN(xi) = 36 liter.

Kvantilisek:

Kvartilisek: a négy részre osztott növekvő rangsorba rendezett sokaság osztópontjai.

Alsó kvartilis (Q_1): =KVARTILIS(xi;1) = 27,75 liter

Medián (Q_2): =KVARTILIS(xi;2) = 36,00 liter

Felső kvartilis (Q_3): =KVARTILIS(xi;3) = 48,50 liter

Decilisek: a tíz részre osztott növekvő rangsorba rendezett sokaság osztópontjai.

Első decilis: =PERCENTILIS(xi;0,1) = 18,00 liter

Második decilis: =PERCENTILIS(xi;0,2) = 26,40 liter

Harmadik decilis: =PERCENTILIS(xi;0,3) = 28,00 liter

Negyedik decilis: =PERCENTILIS(xi;0,4) = 33,00 liter

Ötödik decilis: =PERCENTILIS(xi;0,5) = 36,00 liter

Hatodik decilis: =PERCENTILIS(xi;0,6) = 40,40 liter

Hetedik decilis: =PERCENTILIS(xi;0,7) = 48,00 liter

Nyolcadik decilis: =PERCENTILIS(xi;0,8) = 51,00 liter

Kilencedik decilis: =PERCENTILIS(xi;0,9) = 60,00 liter

Szóródás mérőszámai:

A szóródás terjedelme: a legnagyobb és a legkisebb x_i érték különbsége.

$$T = x_{\max} - x_{\min}$$

³⁷ A továbbiakban ahol használtuk a függvény beszúrása parancsot ott utalunk erre.

A szóródás terjedelme: $T = 70 - 5 = 65$ liter.

A szórás: az x_i értékeknek a az átlaguktól mért különbségeinek a négyzetes átlaga.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{60} [(27-37,65)^2 + (5-37,65)^2 + \dots + (28-37,65)^2]} = 15,60 \text{ liter}$$

$$= \text{SZÓRÁSP}(x_i) = 15,60 \text{ liter}$$

A variancia (szórásnégyzet): a szórás négyzete.

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{1}{60} [(27-37,65)^2 + (5-37,65)^2 + \dots + (28-37,65)^2] = 243,49$$

$$= \text{VARP}(x_i) = 243,49$$

Relatív szórás: a szórás a számtani átlag százalékában.

$$V = \frac{\sigma}{\bar{x}} = \frac{15,60}{37,65} = 0,4145 \rightarrow 41,45\%$$

2.5 Empirikus eloszlások elemzése Excel parancsfájl működése

Az empirikus eloszlások elemzése parancsfájl öt munkalapból áll. Az első, az „adatok-számítások” munkalap magában foglalja az elemi műveletek a változókkal parancsfájl ugyanezen elnevezésű munkalapját, kibővítve a gyakorisági sorok képzésével és ezek elemzésére szolgáló további mutatószámokkal. Az 1. 2. 3. munkalap az osztályközös gyakorisági sorok képzését, az ezekből nyerhető becült mutatószámokat, a gyakorisági sor hisztogramját, és a koncentráció vizsgálatát tartalmazza. A segítség munkalap a parancsfájl használatával kapcsolatos tudnivalókat tartalmazza.

A gyakorisági sor

A **gyakorisági sor** felsorolja a mennyiségi ismerv előforduló különböző értékeit, és mindegyikükhöz hozzárendeli az előfordulásuk számát, azaz a gyakoriságukat. A gyakorisági sor, **csoportosító sor**, azaz a sokaság megoszlását mutatja a vizsgált mennyiségi ismerv szerint. A gyakoriságok összege a sokaság elemszámát adja meg. A gyakorisági sor képzését az előforduló értékek, - általában növekvő - rangsorából kiindulva a legegyszerűbb elvégezni. Ha kevés számú különböző érték fordul elő a sokaságban, - mint például a gyermekek száma, a keresők száma a családokban, stb. - akkor a gyakorisági sor képzése könnyen elvégezhető. Az így képzett sort egyszerű gyakorisági sornak nevezzük. Ha nagyon sokféle előforduló értékkel találkozunk, akkor az információ tömörítés érdekében nem célravezető a különböző értékek felsorolása. Ilyenkor az értékekből intervallumokat, ún. osztályközöket képezünk és az egyes osztályközökhöz tartozó gyakoriságokat állapítjuk meg. Az ilyen sort, **osztályközös gyakorisági sornak** nevezzük. Az osztályközös gyakorisági sor képzésének kulcskérdése, az osztályközök számának és hosszának a meghatározása. Azonos hosszúságú osztályközök esetére sokféle osztályköz-meghatározási módot ismer az irodalom. A gyakorlatban elterjedt módszerek közül itt csupán kettőt említünk meg.

Az osztályközök hossza (h) meghatározható az alábbi módon:

$$h = \frac{X_{\max} - X_{\min}}{r}$$

ahol r az osztályközök száma és

$$r = \sqrt{\sum f_j} + 1$$

Az osztályközök számát meghatározhatjuk az alábbi képlet alkalmazásával is:

$$r = 1 + 3,3 \lg(n)$$

mely alapján az osztályközök hossza az előzőek szerint állapítható meg.

Természetesen a fenti módszerek automatikus alkalmazása nem garantálja egyértelműen a jó megoldást. Soha nem szabad szem előtt téveszteni a vizsgált sokaság sajátosságait, ugyanis a szakmai ismeret, a szubjektív vélemények figyelembe vétele jobb eredményt hozhat, mint az algoritmusok mechanikus alkalmazása. Az osztályközök hosszának meghatározása után az **osztályközös gyakorisági sor** az alábbi séma szerint képezhető:

$$x_j^{\text{felső}} = x_{j+1}^{\text{alsó}} = x_{\min} + j \times h$$

Természetesen – gyakorlati okokból – törekszünk könnyen kezelhető osztályok (kerek számok az osztályhatárok, azonos hosszúságúak az osztályközök) megállapítására. Az előforduló értékeknek az osztályközökbe történő egyértelmű besorolása érdekében, az egymást követő osztályközök alsó és felső határát meg kell különböztetni egymástól.

2-1. tábla: Az osztályközös gyakorisági sor sémája

Ismérvváltozatok	Gyakoriság
$x_1^{\text{alsó}} - x_1^{\text{felső}}$	f_1
$x_2^{\text{alsó}} - x_2^{\text{felső}}$	f_2
\vdots	\vdots
$x_r^{\text{alsó}} - x_r^{\text{felső}}$	f_r
Összesen	$n = \sum_{j=1}^r f_j$

Az osztályközös gyakorisági sorok esetén sokszor élünk az ún. nyitott osztályok alkalmazásának lehetőségével, vagyis a legalsó, illetve a legfelső osztályt „nyitva hagyjuk”, ezáltal lehetővé téve a kiugró (extrem) értékek besorolását.

A gyakorisági sorokból további mennyiségi sorokat származtathatunk. A **relatív gyakorisági sor** a tényleges gyakoriságok helyett, az azokból számított megoszlási viszonyszámokat tartalmazza, melyeket **relatív gyakoriságoknak** nevezünk, melyek összege 1. Az **értékösszeg-sor** a változóértékek (osztályközök) felsorolása mellett, az ezekhez tartozó értékek összegét tartalmazza. Osztályközös gyakorisági sor esetén **becsült értékösszeg-sor** állítható elő az osztályközepek és a gyakoriságok szorzata alapján. Az értékösszegekből számított megoszlási viszonyszámok segítségével **relatív értékösszeg-sor** képezhető. Az előbbi sorok mindegyikén elvégezhető a halmozott összeadás (kumulálás) művelete, amely jelenthet kumulálást (felfelé kumulálást) és lefelé kumulálást. Ez azt jelenti, hogy a növekvő értékek felé, illetve a csökkenő értékek felé történik a halmozott összeadás. Így nyerjük a **kumulált sorokat**. A gyakorisági sor alapvető ábrázolási módszere a **hisztogram**.

Középérték és szóródás számítás gyakorisági sorokból

A gyakorisági sorokból a **számtani átlagot** és a **szórást** súlyozott formában számíthatjuk ki. Súlyként mindkét esetben a gyakoriságok (f_i -k), illetve a relatív gyakoriságok (g_i -k) szerepelnek.

A súlyozott számtani átlag képlete:

$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{\sum_{i=1}^k f_i} = \frac{\sum_{i=1}^k f_i x_i}{n}$$

$$\bar{x} = \sum_{i=1}^k g_i x_i \text{ mivel } \sum_{i=1}^k g_i = 1$$

Ahol

x_i = az előforduló változó-értékek, illetve az osztályközepek, ezek az átlagolandó értékek

A fenti képlet alapján látható, hogy a súlyozott számtani átlag nagyságát két tényező határozza meg:

1. az átlagolandó értékek (abszolút) nagysága,
2. a súlyok viszonylagos nagysága, más szóval a súlyarányok.

A **szórás** súlyozott formulája:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n f_i (x_i - \bar{x})^2}{\sum_{i=1}^n f_i}} = \sqrt{\sum_{i=1}^n g_i (x_i - \bar{x})^2}$$

Mindkét helyzeti középértéknek a módusznak és a mediánnak is jelentős szerepe van a gyakorisági sorok elemzésében.

Az egyszerű gyakorisági sorban – ahol a mennyiségi ismérvet diszkrét számértékek jellemzik – a módusz meghatározása nem okoz különösebb nehézséget, mivel a legnagyobb gyakorisággal rendelkező ismérvértéket kell kiválasztani. Komplikáltabb a számítás, ha az adatok osztályközös gyakorisági sorba vannak rendezve. A módusz, mint tudjuk, a leggyakrabban előforduló ismérvérték. Osztályközös gyakorisági sor, valamint folytonos mennyiségi ismérveket tartalmazó sorok esetén a definíciót kissé általánosabban fogalmazzuk meg. Ilyen esetben a **módusz**³⁸ az az érték, amely körül az előforduló értékek legjobban sűrűsödnek, ahol a gyakorisági görbének *maximума* van. Ez a meghatározás egyben azt is jelenti, hogy a módusz egzakt meghatározására osztályközös sorok esetén nincs mód, értékét csak közelítő számítással tudjuk meghatározni.

A módusz meghatározása két lépésben történik:

1. Ki kell választani az ún. **modális osztályközt**, amelyben a módusz található. Ez egyenlő hosszúságú osztályközök esetén az az osztályköz, amelyhez a legnagyobb gyakoriság tartozik. Nem egyenlő osztályközök esetén azonban ezt meg kell előznie a gyakoriságok korrekciójának, hiszen a modális osztályköz azonosításakor figyelembe kell venni azt a tényt, hogy eltérő hosszúságú osztályközök esetén egyenletességet feltételezve pl. kétszer olyan hosszú osztályközhöz kétszer akkora gyakoriságnak kellene tartozni. Így – egységnyi osztályköz-hosszra vetítve a tényleges gyakoriságot – az osztályköz hosszából számított egyenértékessel a tényleges gyakoriságot korrigálni kell.

$$\text{Mindez képletbe rendezve : } f_j^* = f_j \frac{b^*}{x_j^{\text{felső}} - x_j^{\text{alsó}}}$$

ahol b^* az egyenértékesnek tekintett osztályköz hosszúsága, f_j^* a j-edik osztályhoz tartozó korrigált gyakoriság.

2. A kiválasztott modális osztályköz birtokában alkalmazható az alábbi képlet:

$$Mo = x_{mo}^{\text{alsó}} + \frac{(f_{mo}^* - f_{mo-1}^*)}{(f_{mo}^* - f_{mo-1}^*) + (f_{mo}^* - f_{mo+1}^*)} (x_{mo}^{\text{felső}} - x_{mo}^{\text{alsó}})$$

ahol: f_{mo}^* a modális osztályköz korrigált gyakorisága,

f_{mo-1}^* a modális osztályköz előtti osztályköz korrigált gyakorisága,

f_{mo+1}^* a modális osztályközt követő osztályköz korrigált gyakorisága,

$x_{mo}^{\text{alsó}}$, $x_{mo}^{\text{felső}}$ a modális osztályköz alsó és felső határai.

Egyszerű gyakorisági sorok esetén a **medián** kiválasztása a rangsor alapján könnyen elvégezhető, a medián sorszámának megfelelő x_i érték megállapításával. Osztályközös gyakorisági sor esetén csak valamilyen közelítő eljárással lehetséges.

A **medián** meghatározására az alábbi algoritmust szokás használni:

$$Me = x_{me}^{\text{alsó}} + \frac{n/2 - f'_{me-1}}{f_{me}} (x_{me}^{\text{felső}} - x_{me}^{\text{alsó}})$$

ahol: $x_{me}^{\text{alsó}}$ a mediánt magában foglaló osztályköz alsó határa,

$x_{me}^{\text{felső}}$ a mediánt magában foglaló osztályköz felső határa,

f'_{me-1} a mediánt megelőző osztályköz felfelé kumulált gyakorisága,

f_{me} a mediánt tartalmazó osztályköz gyakorisága,

$n/2$ a medián sorszáma.

A származtatott mutatószámok meghatározása körében van kitüntetett szerepe a momentumoknak. Egy kvantitatív változó **r-ed rendű momentuma**, a bevett statisztikai meghatározás szerint, a változóértékek r-edik hatványainak átlagával egyenlő, vagyis súlyozott esetben:

³⁸ Pintér József - Rappai Gábor [2007]: 128-131.

$$m_r = \frac{\sum_{i=1}^k f_i x_i^r}{n}$$

Ha a $r=1$, az ún. elsőrendű momentum a számtani átlag, $m_1 = \bar{x}$.

Definiálható valamely tetszőlegesen megválasztott „a” értékre vonatkozó r -ed rendű momentum is, ha $a = \bar{x}$, akkor ez az **r -ed rendű centrális momentum** súlyozott esetben:

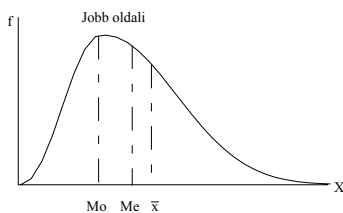
$$m_r(c) = \frac{\sum_{i=1}^k f_i (x_i - \bar{x})^r}{n}$$

Az első rendű centrális momentum, $m_1(c) = 0$, mivel az átlagtól mért eltérések algebrai összege 0. A másod rendű centrális momentum megegyezik a szórásnégyzettel. $m_2(c) = \sigma^2$.

Empirikus eloszlástípusok. Az aszimmetria és a ferdeség mérésére szolgáló mutatók

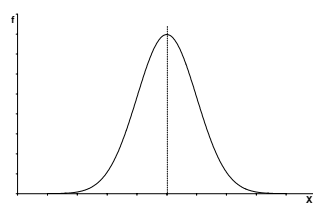
A **gyakorisági eloszlások elemzése** a gyakorisági görbe alakjának vizsgálatát jelenti. Az empirikus eloszlást jellemző grafikus ábrát, a hisztogramot, illetve gyakorisági poligont, összehasonlítjuk a normális eloszlás szimmetrikus gyakorisági görbéjével. A gyakorisági poligon felrajzolása során az osztályközepéknél felmért gyakoriságok pontjait (ezek a hisztogramok oszlopközepének felelnek meg) összekötjük. Az ún. egymódusú eloszlás esetében, azt vizsgáljuk, hogy az empirikus eloszlásunk **szimmetrikusnak** tekinthető-e, vagy a görbe, valamelyik széle felé jobban elnyúlik. Ez utóbbi esetben **jobb, vagy bal oldali aszimmetriáról** beszélünk. A gyakorisági görbe alakjáról a grafikus ábra és a középértékek nagyságrendje már tájékoztat. Szimmetrikus tekinthető empirikus eloszlások esetén a számtani átlag, a medián és a módusz értéke közel azonos.

Jobb oldali aszimmetriánál az említett három középérték közül a módusz értéke a legkisebb, bal oldali aszimmetriánál pedig a számtani átlagé. Jobb oldali aszimmetriájú eloszlás esetén, a görbe a csúcspontját valamilyen alacsonyabb x_i értéknél veszi fel, a magasabb x_i értékek felé haladva a gyakoriságok egyre kisebbek lesznek, a görbe hosszan elnyúlik. **Bal oldali aszimmetriánál** természetesen fordított a helyzet, amint az alábbi ábra mutatja.



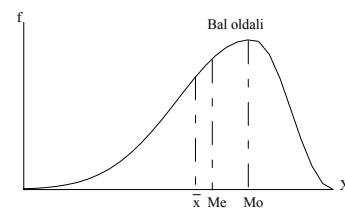
$$Mo < Me < \bar{x}$$

Jobboldali aszimmetria



$$Mo = Me = \bar{x}$$

Szimmetrikus eloszlás



$$Mo > Me > \bar{x}$$

Baloldali aszimmetria

2-1. ábra: Szimmetrikus és aszimmetrikus eloszlások

Az aszimmetria mérésére használt mutatószámok közös jellemzője, hogy dimenzió nélküliek, értékük nulla, ha az eloszlás szimmetrikus, és előjelük jelzi a jobb- és baloldali aszimmetriát.

A Pearson-féle A mutató:

$$A = \frac{\bar{x} - Mo}{\sigma}$$

A számtani átlag és a módusz értékének azonossága esetén a mutató értéke 0, ami jelzi a szimmetriát. A mutató pozitív előjele a *jobb oldali*, negatív előjele a *bal oldali* aszimmetriát mutatja. A mutatónak *nincs* felső korlátja, tehát a mutatószám értéke nem utal közvetlenül az aszimmetria mértékére.

Az aszimmetria mérésére szolgáló F-mutató:

$$F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$$

Az F-mutató számítása azon alapul, hogy szimmetrikus eloszlásnál a kvartilisek egyenlő távolságra vannak egymástól, ekkor 0 a mutató értéke. A jobb és bal oldali aszimmetriát az F mutatószám előjele ugyanúgy jelzi, mint az A mutató, abszolút értéke azonban maximum 1 lehet. Ez az extrém aszimmetriát jeleznék, amikor is a medián értéke valamelyik szélső kvartilisével (Q_3 vagy Q_1) egyezik meg.

A ferdeség mérésére szolgáló **S mutató**, amely meghatározható a centrális momentumok alapján:

$$S = \frac{m_3^2(c)}{m_2^3(c)}$$

A S -mutató értéke 0, ha az eloszlás szimmetrikus, pozitív előjel esetén jobb oldali, míg negatív előjel esetén bal oldali aszimmetriára következtethetünk. Hasonlóan az A-mutatóhoz ez a mutatószám sem korlátos.

A gyakorisági eloszlások vizsgálata kiterjed az empirikus eloszlások további alak-vizsgálatára, a csúcsosság-lapultság elemzésére is. **Csúcsosságon** az eloszlást jellemző görbe meredekségét értjük a módusz környezetében. A csúcsosság mérőszámai is a normális eloszláshoz viszonyítva vizsgálják az eloszlás relatív *lapultságát*, vagy erőteljes meredekségét. Tárgyi értelmét tekintve a szimmetrikus, de lapult eloszlás az egyenletes eloszlás felé közelít, míg a csúcsos eloszlás jelzi az átlag körüli tömörülés erőteljes voltát.

A csúcsosság mutatószáma:

$$K = \frac{m_4(c)}{m_2^2(c)}$$

A K-mutató értéke 3, normális eloszlás esetén. Ennél kisebb érték esetén lapultnak, nagyobb érték esetén csúcsosnak tekinthetjük az eloszlást.

A koncentráció vizsgálata

A gyakorisági sorok adatai alapján vizsgálható a sokaság x változó szerinti koncentrációja. Az előzőekben bemutatott empirikus eloszlások vizsgálatának nem szerves része a koncentráció elemzése. Azért tárgyaljuk együtt, mert ugyanabból az adatbázisból – a gyakorisági sorokból – mindkét vizsgálat elvégezhető. Koncentráció vizsgálatot akkor végzünk, ha értelmezhető a sokaság x változó szerinti koncentrációja, vagyis a gyakorisági sor mellett képzett értékösszeg-sor adatainak tárgyi értelme van. Pl. a népesség jövedelem szerinti koncentrációját vizsgálva, a jövedelem eloszlása mellett a megszerzett jövedelmek összege, az összjövedelemből való részesedés is értelmezhető. Beszélhetünk a gazdálkodási egységeknek a termelési érték, a létszám, a beruházás, a forgalom, az árbevétel és a nyereség szerinti koncentrációjáról; a népesség jövedelem, vagyon szerinti koncentrációjáról.

Koncentráción a gazdasági-társadalmi jelenségekben megfigyelhető *tömörüléseket*, *összpontosulásokat* értjük. A koncentráció fogalma mind a gazdasági folyamatokat, mind azok eredményeként létrejött állapotokat jellemzi. Koncentrációnak nevezzük azt a jelenséget, amikor a sokasághoz tartozó értékösszeg jelentős része a sokaság kevés egységére összpontosul. A koncentráció jelenségét a mennyiségi sorok alapján jellemezhetjük (mérhetjük) úgy, hogy egy adott X ismerv gyakorisági és értékösszeg eloszlását hasonlítjuk össze. Ha a teljes értékösszeg megoszlása *nem egyenletes*, azaz a teljes értékösszeg nagy része a sokaság egységeinek kis részénél összpontosul, **relatív koncentrációról** beszélünk. Az egyenletes eloszlás az az elméleti határeset, amely a koncentráció teljes hiányát jeleznék. A másik elméleti határeset – a leg-erősebb fokú – az abszolút koncentráció, amikor a teljes értékösszeg egyetlen sokasági egységhez tartozik. A relatív koncentráció elemzése elvégezhető a **koncentrációs tábla** alapján, a relatív gyakorisági és relatív értékösszeg sor összehasonlításával. A koncentráció meglétét az jelzi, hogy az alacsonyabb x_i értékeknél a relatív gyakoriságok rendre nagyobbak a relatív értékösszegek értékeinél. A nagyobb x_i értékeknél fordított helyzetet tapasztalunk.

A koncentráció ábrázolására és elemzésére szolgáló speciális grafikus ábrát, megalkotójáról³⁹ Lorenz-görbének hívjuk. A Lorenz-görbe egységoldalú négyzetben elhelyezett ábra, amely a kumulált relatív

³⁹ Max Otto Lorenz (1880-1962) amerikai statisztikus 1905-ben publikálta először az általa elsősorban jövedelem-egyenlőtlenségek illusztrálására javasolt ábrát.

gyakoriságok (g'_i) függvényében ábrázolja a kumulált relatív értékösszegeket (z'_i).⁴⁰ Amennyiben az egységeknek az értékösszegeből való részesedése egyforma lenne, a kumulált relatív gyakoriságok és a kumulált relatív értékösszegek rendre megegyeznek ($g_i = z_i$), a görbe a négyzet átlójával egybeesne. Ha a görbe közel van az átlóhoz, akkor gyenge a koncentráció. Ha a relatív gyakoriságok és relatív értékösszegek igen jelentősen eltérnek egymástól, a görbe a tengelyekhez áll közel, ez mutatja az erős koncentrációt. A görbe és az átló által bezárt terület tehát, a koncentráció relatív nagyságát jellemzi. Ezt a területarányt közelíti az ún. **Gini-féle koncentrációs arányszám**.

$$K = \frac{\sum_{j=1}^n \sum_{i=1}^n f_i f_j |x_i - x_j|}{2\bar{x}n(n-1)}$$

A fenti képlettel meghatározott koncentrációs arányszám értéke egyenlő a Lorenz-görbe és az átló által bezárt területnek és a tengelyek és az átló által bezárt háromszög területének az arányával. A mutatószám határai: $0 \leq K \leq 1$.

Az empirikus eloszlások elemzése parancsfájl öt munkalapból áll. Az adatok-számítások munkalap részben megegyezik az elemi műveletek parancsfájl azonos nevű munkalapjával. A munkalap további részében az egyszerű gyakorisági sor képzését és az empirikus eloszlások jellemzőit közli a parancsfájl. Az 1. és 2. munkalap az adatok és számítások munkalapon szereplő adatbázisból osztályközös gyakorisági sorokat képez és hisztogramot készít. A 3. munkalap az empirikus eloszlások elemzésére szolgál abban az esetben, amikor az egyedi adatok nem állnak rendelkezésre, és így a kiinduló adatbázis egy osztályközös gyakorisági sor. A működést bemutató szemléltető példa: egy benzinkút egyik kútoszlopánál egy adott időszak – 4 óra – alatt kiszolgált benzin mennyisége literben, amit az események sorrendjében jegyeztek fel, ld. xi sárga mezőben lévő oszlopot.

Az adatbázis maximum 5000 értéket tartalmazhat.

Momentumok (r-ed rendű, súlyozás nélküli): az x_i változó értékeinek r-edik hatványából számított átlagok az r=1,2,3, és 4 esetekben

Momentumok	
Elsőrendű	37,65
Másodrendű	1661,016667
Harmadrendű	81014,25
Negyedrendű	4241460,617

Centrális momentumok: az \bar{x} értékre vonatkozó r-ed rendű momentumok, r=1,2,3, és 4

Centrális momentumok	
Elsőrendű	1,30266E-15
Másodrendű	243,4941667
Harmadrendű	141,86175
Negyedrendű	139775,4399

Empirikus eloszlások jellemzői

A mutató (aszimmetria): az aszimmetria Pearson-féle mérőszáma

$$A = \frac{\bar{x} - Mo}{\sigma} = \frac{37,65 - 36}{15,6} = 0,11$$

F mutató (aszimmetria): az aszimmetria kvartilisekből számított mérőszáma

$$F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)} = 0,20$$

S mutató (aszimmetria), amely meghatározható a centrális momentumok alapján:

⁴⁰ Hunyadi László – Vita László [2008] I. 120.

$$S = \frac{m_3(c)}{m_2(c)} = 0,037$$

Az Excel által alkalmazott képlet:

$$S' = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^3 = \frac{60}{(60-1)(60-2)} \frac{[(27-37,65)^2 + (5-37,65)^2 + (27-37,65)^2 + \dots + (28-37,65)^2]^3}{15,60^3} = 0,038$$

$$=FERDESÉG(xi) = 0,038$$

K mutató (csúcsosság):

$$K = \frac{m_4(c)}{m_2(c)} - 3 = -0,64$$

K' mutató (csúcsosság): az Excel által alkalmazott képlet:

$$K' = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{i=1}^n \left[\frac{x_i - \bar{x}}{\sigma} \right]^4}{\sigma^4} - \frac{3(n-1)^2}{(n-2)(n-3)} = \frac{60(60+1)}{(60-1)(60-2)(60-3)} \frac{[(27-37,65)^2 + (5-37,65)^2 + \dots + (28-37,65)^2]^4}{15,60^4} - \frac{3(60-1)^2}{(60-2)(60-3)} = -0,59$$

$$=CSÚCSOSSÁG(xi) = -0,59$$

Az 1. munkalap – az adatok-számítások munkalapon szereplő x_i adatsor, illetve az egyszerű gyakorisági sor alapján – olyan osztályközös gyakorisági sort képez, ahol az első és az utolsó osztályköz nyitott. Az 1. munkalapon minden számítás automatikus, az adatok-számítások munkalapon meghatározott egyszerű gyakorisági sorból (x_i, f_i) képezi az osztályközös gyakorisági sort. Az 1. munkalapon csak az első osztályköz (a példában: xa) és az utolsó osztályköz (a példában: xf) határát lehet módosítani, amit a sárga szín is jelez. A 2. munkalapon ugyanezt végzi el, azzal a különbséggel, hogy az osztályközök száma és az osztályközök felső határa is tetszés szerint változtatható (a sárga szín jelzése szerint), így akár többszöri próbálkozással lehet a tényleges eloszlást legjobban leíró gyakorisági sort megkeresni. Ehhez ad információt a tényleges értékösszeg-sor megjelenítése, a becsült sorok mellett.

Az osztályközös gyakorisági sorból képezhető további sorok mindegyikét közli a parancsfájl, így az értékösszeg, relatív és kumulált sorokat is.

A következő jelöléseket használja a program:

- fi' = a (felfelé) kumulált gyakoriságok
- fi'' = a lefelé kumulált gyakoriságok
- gi = a relatív gyakoriságok
- gi' = a (felfelé) kumulált relatív gyakoriságok
- gi'' = a lefelé kumulált relatív gyakoriságok
- si = a tényleges értékösszeg
- si' = a felfelé kumulált tényleges értékösszeg
- si'' = a lefelé kumulált tényleges értékösszeg
- zi = a relatív tényleges értékösszeg (%)
- zi' = a felfelé kumulált relatív tényleges értékösszeg (%)

- z_i'' = a lefelé kumulált relatív tényleges értékösszeg (%)
- si_becs = a becslt értékösszeg [osztályközép (xi)*(fi)]
- si'_becs = a felfelé kumulált becslt értékösszeg [osztályközép (xi)*(fi)]
- si''_becs = a lefelé kumulált becslt értékösszeg [osztályközép (xi)*(fi)]
- zi_becs = a relatív becslt értékösszeg (%)
- zi'_becs = a felfelé kumulált relatív becslt értékösszeg (%)
- zi''_becs = a lefelé kumulált relatív becslt értékösszeg (%)

A program megadja – a különböző sorokat tartalmazó táblában – az f_i^* -gal jelölt korrigált gyakoriságokat, amelyek a módusz becsléséhez és a hisztogram szerkesztéséhez szükségesek.

Az 1. és 2. munkalap záró része a koncentráció elemzésére megadja a koncentrációs táblát, a Lorenz-görbét és a Gini-mutatót, a munkalapokon szereplő adatbázis alapján. A számításokat kétféle módon tudja elvégezni a program, a tényleges, illetve a becslt értékösszeg-sor adatai alapján, melyek közül választhatunk a legördülő menüsor használatával. Az 1. és 2. munkalapon szereplő bemutató példa nem alkalmas a koncentráció elemzésének bemutatására, mivel itt az értékösszeg-sor adatainak (a vásárolt összes benzin mennyiségének) nincs információ tartalma, nehezen értelmezhető a vásárlóknak a vásárolt benzinmennyiség szerinti koncentrációja. Itt újra hangsúlyozzuk, hogy, ha empirikus eloszlások elemzése a célunk, – és adatbázisunk alapján nem értelmezhető a koncentráció – akkor a program által megadott, a koncentrációra vonatkozó számítási eredményeket figyelmen kívül kell hagynunk.

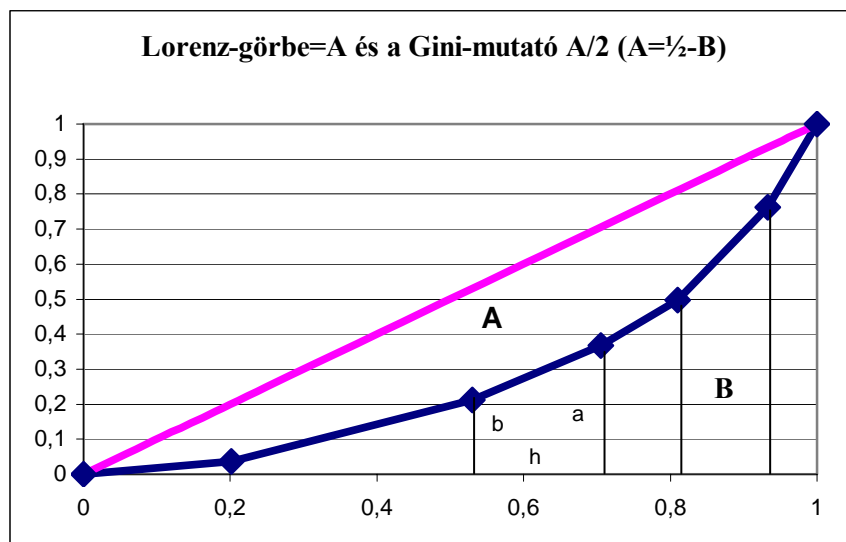
A 3. munkalap – amint már utaltunk rá – az empirikus eloszlások elemzésére szolgál abban az esetben, amikor az egyedi adatok nem állnak rendelkezésre, és így a kiinduló adatbázis egy osztályközös gyakorisági sor. Az adatok bevitelét az osztályközök száma sárga mezőben kitöltésével kell kezdeni. Ha az utolsó osztályköz felső határa nyitott, akkor az osztályközök száma eggyel több, mint a beírandó felső határok. Az osztályközök bevitelére az osztályközök felső határa (Felső oszlop), míg a gyakoriságok bevitelére az f_i oszlopa szolgál a sárga mezőben. Ha ún. nyitott osztályos a gyakorisági sorunk, akkor az x_a és x_f sárga mezőbe beírhatjuk az általunk választott alsó és felső osztályköz határt. Amennyiben az említett sárga cellákat nem töltjük ki, a program mechanikusan elvégzi az osztályközepek számítását. A kiinduló adatbázisunk tartalmazhatja az értékösszeg-sort is (si oszlop), ebben az esetben az adatainkat a sárga cellákban rögzíthetjük és így a további számítások a tényleges értékösszegekből történnek. Ha a tényleges értékösszegek nem ismeretesek, akkor a program az osztályközepek segítségével becslt értékösszegeket számít.

A 3. munkalapon ugyanazok az elemzési eredmények jelennek meg, mint az 1. és 2. munkalapon. Részletesebb magyarázatot csak a koncentráció elemzése igényel. A koncentrációs tábla adatai alapján összehasonlíthatók a kumulált relatív gyakoriságok a kumulált relatív értékösszegekkel. Ezen adatok alapján a program megrajzolja a Lorenz-görbét és közli a kiszámított Gini-mutató értékét.

Az Excel felhasználásával meghatározhatjuk a Lorenz-görbe és az átló által bezárt területet (**A**) úgy is, hogy meghatározzuk a Lorenz-görbén kívüli terület nagyságát (**B**), felhasználva azt, hogy a trapéz területe:

$$T = \frac{(a + b)h}{2}$$

A trapéz két alapját a-val és b-vel, magasságát h-val jelöltük. Összegezzük a területek nagyságát, ami az alábbi ábrában a trapéz területe. A **B** terület ismeretében a Lorenz-görbe: ($A = \frac{1}{2}B$) a Gini mutató pedig ($A/2$)



2-2. ábra: A Lorenz-görbe és a Gini-mutató becslése grafikus módon

A 3. munkalapon a bemutató példa az APEH honlapján elérhető⁴¹ SzJA bevételek Magyarországon 2004-ben, jövedelemsávok szerint.

Gyakorló feladatok. (elemiműveletek.xls és empirikuseloszlásokelemzése.xls)

1. A Demográfiai Évkönyv 2005. és 2006. KSH, Budapest, közli Excel formátumban (A_3_1_2 2205.xls [3.1.2. A NÉPESSÉG SZÁMA VÁROSONKÉNT]) a magyarországi városok népességét városok közötti bontásban az 1970 és 2007 közötti (1970, 1980, 1990, 2001, 2002, 2003, 2004, 2005, 2006, 2007) időszakban. Vizsgálja meg a városi népesség koncentrációjának a változását Budapest nélkül. Az adatbázis az eredeti formájában nem alkalmas a számítások elvégzésére. A Gyakoriságok függvény felhasználásával csoportosíthatjuk a városokat a népességszám alapján. A népességszámvárosokéntrendezve 1970-2006.xls fájlban nyomon követhető a programozás a Gyakoriságok függvény felhasználásával. (Gyakoriságok: a gyakorisági vagy empirikus eloszlás értékét függőleges tömbként adja eredményül. A gyakorisági eloszlás adott értékalmazból és adott számú osztálynál [intervallumnál] az egyes intervallumokban előforduló értékek számát méri.)

A megoldáshoz segítség: 1970-ben a Gini-mutató 0,518, 2006-ban 0,512 volt.

2. A jövedelemkoncentráció változása Magyarországon az SzJA bevételek alapján. Rendelkezésre álló adatok: 2004, 2005 és 2006. Végezze el a számításokat az elemiműveletek.xls fájl segítségével, és értékelje a kapott adatokat.

3. Egy kereskedelmi ágazat vállalkozásainak éves árbevétel adatait az alábbi táblázat tartalmazza árbevétel-sávok szerint. Végezze el a koncentráció elemzést elemiműveletek.xls fájlal és értékelje a kapott adatokat.

4. Egy wellness hotel bárjának előző napi forgalmáról (az euróban kibocsátott számlák alapján) rendelkezésünkre álló adatsor (számhalmaz). Értelmezze az elemi műveletek (elemiműveletek.xls fájlal) számításainak (közéértékek, szóródási mutatók, az empirikus eloszlás alakjának vizsgálata) a kapott adatokat.

5. Három társasház 132 lakásának havi átlagos vízfogyasztása (m^3 /lakás), amit a Pécsi Vízmű az előző egy éves fogyasztás alapján állapított meg, mint havi átlagmennyiséget a következőképpen alakult. Értelmezze az elemi műveletek (elemiműveletek.xls fájlal) számításainak (közéértékek, szóródási mutatók, az empirikus eloszlás alakjának vizsgálata) az adatait.

6. Ipoly Erdő Zrt. Kemence WS 3600 állomás napi (01, 09, 13, 19 órakor és közép) hőmérsékleti menete. 2007 március. Értelmezze az elemi műveletek (elemiműveletek.xls fájlal) számításainak (közéértékek, szóródási mutatók, az empirikus eloszlás alakjának vizsgálata) az adatait.

7. Egy kereskedelmi bankban egy adott héten (5 munkanap) lekötött (rövid, 1, 2, 3, 4 illetve 6 hónapos futamidejű) betétek értékei, (eFt), növekvő sorrendben 257 adatot tartalmaz.

Értelmezze az elemi műveletek (elemiműveletek.xls fájlal) számításainak (közéértékek, szóródási mutatók, az empirikus eloszlás alakjának vizsgálata, koncentráció, Gini-mutató) az adatait.

⁴¹ Lásd www.apoh.hu.

8. Egy iparvállalat dolgozóinak havi bruttó átlagkeresetére vonatkozóan az alábbi adatokkal rendelkezünk. Értelmezze az elemi műveletek (elemiműveletek.xls fájljal) számításainak (közéértékek, szóródási mutatók, az empirikus eloszlás alakjának vizsgálata, koncentráció, Gini-mutató) az adatait.

3 Az idősorok elemzési módszerei

Az idősor-elemzési módszereket alapvetően két célból használjuk:

- a jelenségek időbeli alakulásának elemzésére múltbeli adatok alapján;
- illetve a jelenségek jövőbeni alakulására vonatkozó előrejelzések (prognózisok) készítésére.

Az idősoros adatbázisok sajátossága, hogy nemcsak az adatok összessége, hanem azok sorrendje is érdekel bennünket. Ez egyben azt is jelenti, hogy szemben a többi statisztikai sorral, idősorok esetén az adatok felsorolásának sorrendje szigorúan kötött, mindig a legkorábbi időpontból (időszakból) haladunk a jelenig. Az idősorok rendkívül fontosak az elemzési és prognosztikai tevékenységben, mivel a múltat ábrázolják és ebből bizonyos következtetések a jövőre vonatkozóan levonhatók. Az idősorok tehát a prognóziskészítés alapját képezik, ahhoz ugyanis, hogy valamely jelenség vagy folyamat jövőbeli alakulásáról helyes képet kapjunk, ismerni kell a prognózistárgy múltbeli alakulását is. Az idősorok alapján történő előrejelzések azon a feltételezésen alapulnak, hogy a történeti sémák folytatódnak a jövőben is. A prognóziskészítés tipikus módszerének tekinthető tehát az idősor-kutatáson alapuló előrejelzés, mert számítási alapelve megfelel a prognosztizálás alapelveinek: a múlt és jelen ismert paraméterei alapján előrevetíteni az összefüggést a jövőbe. Az idősorokban megfigyelt tendencia prognosztizálása azon a feltételezésen nyugszik, hogy a jövőben is ugyanolyan irányú és mértékű változások következnek be, mint amilyenek jellemzőek voltak a vizsgált múltbeli időszakban.

Az idősor adatai általában egymást egyenlő időközzel követő időpontokra illetve időszakokra vonatkoznak. Ezek az időpontok illetve időszakok a gyakorlati tapasztalatok szerint lehetnek:

percek, félórák, órák, napok, hetek, hónapok, negyedévek, évek, évtizedek. A statisztikai hivatalok által közölt adatok általában évenkénti, negyedévenkénti és havonkénti megfigyeléseket tartalmaznak. Rövidebb időszakokra (pl. napi, óránkénti, félóránkénti) vonatkozó adatgyűjtést például a tőzsdeindexek esetében végeznek.

Az idősor mindig konkrét megfigyelés eredménye, vagyis az elméleti idősorok egyetlen realizációja. Végeredményben az idősor egyetlen realizációja, a tapasztalati idősor alapján kell a vizsgált jelenséget illetve folyamatot befolyásoló, illetve meghatározó törvényszerűségeket feltárni.

Az időben lejátszódó folyamatok mindegyike sztochasztikus folyamatként definiálható, mely valószínűségi változók sorozataként jelenik meg. Ezt **elméleti idősoroknak** nevezzük és az alábbi módon jelöljük:

$$Y_{-M}, Y_{-M+1}, \dots, Y_t, \dots, Y_M$$

Az Y_t ($-M \leq t \leq M$) valószínűségi változók mindegyikére vonatkozóan egy megfigyeléssel rendelkezünk, ez a modellezés adatbázisát jelentő tapasztalati idősor.

A realizált idősori értékeket **tapasztalati idősoroknak**, egyszerűen **idősoroknak** nevezzük és a továbbiakban y_t -vel ($t = 1, 2, \dots, N$) jelöljük⁴².

$$y_1, y_2, \dots, y_t, \dots, y_N$$

Az idősorelemzés legegyszerűbb, de egyben legtöbbször alkalmazott módszerei a **grafikus ábrázolás** és a **dinamikus (bázis és lánc) viszonyszámok** számítása. Ezek ismertetése az 1. fejezetben megtalálható.⁴³ Az egyszerűbb elemzési eszközök további csoportját alkotják az idősor-elemzés területén alkalmazható átlagok. Az idősor elemzés esetében ritkán használunk átlagszámítást, mivel nem az információ tömörítés a cél, hanem az időben lejátszódó folyamatok leírása. Speciális esetekben szükség lehet egy idősor átlagos értékére. Hasonlóan, a grafikus ábrázolás során az álló és mozgó sokaságok közötti megkülönböztetéshez, az idősori átlagszámításnál különböző átlagfajtákat használunk állapot- és tartam-idősor esetén.

Ha egy **tartam-idősor** átlagos értékére vagyunk kíváncsiak, – mivel az idősor adatai összegezhetőek – alkalmazható az **egyszerű számtani átlag**:

⁴² A t a latin tempus szóból származik.

⁴³ Ld. Dinamikus-viszonyszámok.xls parancsfájlt.

$$\bar{y} = \frac{\sum_{t=1}^N y_t}{N}$$

Az **állapot idősor** adatainak átlagolására az úgynevezett **kronologikus átlagot** szokás használni:

$$\bar{y}_t^k = \frac{y_1 + y_2 + \dots + y_{N-1} + y_N}{N-1}$$

Az idősorelemzésnek két fő megközelítési módja ismert:

1. A **determinisztikus idősorelemzés**⁴⁴ abból a feltételezésből indul ki, hogy az idősorok alakulását egy viszonylag hosszú távú növekedési/csökkenési pálya határozza meg, amely körül tartósan ható szabályos hullámmozgások (pl. szezonálitás) mutathatók ki. Ezeket eseti, egyedi eltérítő hatások – maradékként – a véletlen tényezőt jelenítik meg.
2. A **sztochasztikus idősorelemzés**⁴⁵ (pl. Box – Jenkins, ARIMA stb. modellek) abból a feltételezésből indul ki, hogy az aktuális időszaki értékeket korábban realizálódott értékei és a véletlen hatás alakítja ki, a determinisztikus modellezés feltételezte hosszú távú tendencia befolyásoló szerepe ebben a megkülönböztetésben közvetlenül nem jelenik meg.

Az Excel parancsfájlokat néhány determinisztikus idősorelemzési módszerre készítettük el.

3.1 A dekompozíciós idősormodellek

A dekompozíciós idősorelemzés statisztikai módszerei azt feltételezik, hogy a vizsgált idősorok általában négy fő, egymástól független és elkülöníthető tényezőből tevődnek össze. Ezek a következők: a trend vagy hosszú távú alapirányzat, az ettől szabályos (általában havi vagy negyedéves) ingadozásokkal eltérő rövid távú (szezonális) ingadozást leíró összetevő, a szabálytalan hosszabb távú ingadozásokat leíró ciklikus összetevő (konjunktúraciklus) és a véletlen változó. A következőkben ezen tényezők kimutatására szolgáló módszereket mutatjuk be.

3.1.1 Az idősorok összetevői és kapcsolódási módjai

1. A trend vagy a hosszú távú alapirányzat

A trend az idősorban tartósan – az ingadozásokon keresztül – hosszú távon érvényesülő tendencia, amely az idősor alakulásának fő irányát jelenti. A trendet több, a vizsgált jelenség alakulását alapvetően meghatározó tényező alakítja. Ha az adott időintervallumon becsült tendenciát extrapolálással ki akarjuk terjeszteni a vizsgált intervallum határain kívülre, ezt csak azzal a feltételezéssel tehetjük, hogy ott is érvényesül ez a stabilitás. Az idősorokban érvényesülő trendek meghatározására kidolgozott módszerek többsége jelentős adatmennyiség ismeretében is csak rövid illetve középtávú prognózisok készítésére használható. Hosszabb távú prognosztizálásra való felhasználásukat korlátozza az a tény, hogy hosszabb távlatban ritkán feltételezhető a vizsgált jelenségre ható tényezők körének és irányának változatlansága. Ugyanakkor, ha hosszú, 100-200 éves, vagy ennél hosszabb idősorokkal rendelkezünk, akkor a hosszú távú, ún. megatrendeket is azonosítani tudjuk, ami megalapozottabb trend extrapolációt eredményezhet. A megatrendek „nem az egyik napról a másikra alakulnak ki és enyésznek el. Ezek az átfogó társadalmi, gazdasági, politikai és műszaki változások lassan bontakoznak ki, de ha egyszer már kialakultak, akkor befolyást gyakorolnak ránk egy ideig, hét-tíz évig, vagy még tovább. Megvan a hatókörük és alakító erejük egy-egy évtized változásspektrumának a meghatározásához.”⁴⁶

2. A szabályos rövid távú (szezonális) ingadozást leíró összetevő

A szezonális vagy idényszerű ingadozás állandó periódushosszú, olyan hullámmozgás, melynek a periódushossza egy évnél rövidebb időszak.

3. A szabálytalan hosszabb távú ingadozásokat leíró ciklikus összetevő (konjunktúraciklus)

⁴⁴ A latin determinatio (be)határolás szóból ered.

⁴⁵ Sztochasztikus görög szó, jelentése statisztikai valószínűség alapuló.

⁴⁶ Naisbitt, John - Aburdene Patricia [1991] 11-12.

A konjunktúraciklus a trend körüli ingadozást jelenti. Általában változó periódushosszú és amplitúdójú ingadozás, a legtöbb esetben különböző – egy évnél hosszabb – periódusú (pl. 3, 9, 18, 54 éves) hullámzás. Ezen ciklusok jelenlétét csak hosszabb (legalább 15-100 éves) idősorok alapján lehet kimutatni.

4. A zavaró hatásokat leíró véletlen változók,

A véletlen ingadozás az idősorban kimutatható szabálytalan mozgás, ami nem mutat semmiféle törvényszerűséget, amelyekről többnyire csak azt feltételezzük, hogy várható értékük additív kapcsolatnál 0, illetve multiplikatív kapcsolat esetében 1. A tapasztalati idősorok adatai általában eltérnek a trend, a ciklus és a szezonális alapján várt értékektől. Az eltérést a szabálytalan, rövid távon ható véletlen ingadozás okozza.

Kiegészítésként meg kell említeni a **strukturális törést**. Strukturális töréseknek nevezzük az olyan egy-szeri, jelentős tendenciaváltozásokat, melyek oly számottevően befolyásolják az adott időszakban a jelenség alakulását, hogy külön vizsgálatot igényelnek. Strukturális törés megléte esetén fontos cél, a létrehozó ok vagy okok feltárása, a hatás vagy hatások tovagyrűzésének, esetleges „elhalásának” elemzése. Ha a strukturális törések száma és jelentősége nagy, akkor a dekompozíciós idősorelemzés módszereinek hatékonysága megkérdőjelezhető.

Tételezzük fel, hogy az idősorban mind a négy tényező megjelenik. Az idősor összetevői: additívan (összszegszerűen) vagy multiplikatívan (szorzatszerűen) kapcsolódhatnak egymáshoz.

Additív modell esetén feltételezzük, hogy az idősor megfigyelt értékei, a trend, a szezon, a konjunktúra – ciklus és a véletlen komponens értékeinek összegeként állítható elő.

Ennek alapján ha kapcsolódás módja additív:

$$y_{ij} = \hat{y}_{ij} + s_j^* + c_l^* + v_{ij}^*$$

ahol:

y_{ij} = a megfigyelt idősor értéke, az i -edik periódus j -edik szezonjában

\hat{y}_{ij} = a trendérték, az i -edik periódus j -edik szezonjában

s_j^* = a j -edik szezonban a szezonális eltérés

c_l^* = az l -edik becslött (*) konjunktúra – ciklus, aminek a periódusa különböző (3-60 év) lehet.

v_{ij}^* = a véletlen hatás, az i -edik periódus j -edik szezonjában

$i = 1, 2, \dots, n$ = a periódusok (pl. évek) száma

$j = 1, 2, \dots, m$ = a perióduson belüli időszakok, azaz a szezonok (pl. a hónapok, a negyedévek) száma

$l = 1, 2, \dots, o$ = a konjunktúra ciklusok periódusainak⁴⁷ a száma.

Az újonnan bevezetett jelölésekhez néhány megjegyzést fűzünk. Idősorunk általános elemének jelölésére eddig az y_t ($t=1, \dots, N$) módot használtuk és eszerint, idősorunk N számú adatból áll. Ha az idősorban szezon-komponenst is megkülönböztetünk, a bevezetett új jelölés szerint az idősor általános eleme: y_{ij} ($i=1, 2, \dots, n; j=1, 2, \dots, m$), és az idősor adatainak száma: $n \cdot m = N$. Az állandó szezonális ingadozás eltérítő hatása standard, a vizsgált $t=1, \dots, N$ időtartam alatt. Ez azt jelenti tehát, hogy a vizsgált összes periódusban – példaként az elmúlt öt évben -, negyedéves adatokkal leírt szezonális ingadozás esetén négy szezon-tényezővel, havi adatok esetén tizenkét szezon-tényezővel jellemezzük a szezon-ingadozást. Ezért kell tehát az s_j ($j=1, 2, \dots, m$) jelölést a szezon-komponensre bevezetni (negyedéves adatoknál $m=4$, havi adatoknál $m=12$), és ilyenkor a periódusok (pl. évek) jelölésére az $i=1, 2, \dots, n$ módot használni.

Multiplikatív modell esetén feltételezzük, hogy az idősor megfigyelt értékei, a trend, a szezon, a konjunktúra – ciklus és a véletlen komponens értékeinek szorzataként (jele: *) állítható elő.

Ennek alapján ha kapcsolódás módja multiplikatív:

$$y_{ij} = \hat{y}_{ij} * s_j * c_l * v_{ij}$$

ahol, a már ismert jelölések mellett,

⁴⁷ A periódus az az időköz ami alatt a ciklus átlagosan ismétlődik.

s_j = a j-edik szezonhoz tartozó szezonális komponens, a szezonindex
 c_l = az l-edik konjunktúra ciklushoz tartozó konjunktúraindex.

Összefoglalóan megállapítható, hogy a szezonális és a konjunktúra ciklusok eltérítő hatása a megfelelő szezonoknál és konjunktúra ciklusoknál az additív modellben **abszolút** állandóságot, a multiplikatív modellben pedig a trendhez mért **relatív** állandóságot mutat.

3.1.2 A trend vagy a hosszú távú alapirányzat becslési módszerei

A trendszámítás kétféle módszerét ismertetjük: a mozgóátlagolású és az analitikus trendszámítást. Az analitikus trendszámítás keretében a lineáris és a lineárisra visszavezethető trendekkel foglalkozunk először.⁴⁸ A lineárisra vissza nem vezethető trendek egy részét a telítődési, a logisztikus és az életgörbe trendek becslése során ismertetjük.⁴⁹

3.1.2.1 Mozgóátlagolású trendszámítás

A mozgóátlagolás alapgondolata, hogy a trendet az eredeti adatsor dinamikus átlagaként állítjuk elő. A gyakorlatban igen elterjedt trendszámítási módszer, mert egyszerű és gyorsan számítható. Hátránya viszont, hogy a kiegyenlített sor rövidebb, tehát kevesebb adatot tartalmaz, mint az eredeti, így a nagyon rövid idősor esetében szinte lehetetlen a trendet e módszerrel egyértelműen jelezni. A **mozgóátlagolású trendszámítás** során az idősor értékeiből – általunk választott tagszámú (jele: k) – átlagokat számítunk úgy, hogy az idősor elejéről indulva – az átlagolandó értékek közül az elsőt elhagyva, és az utolsó értéket követőt hozzávéve – az eljárást addig folytatjuk, míg az utolsó adatot is felhasználtuk. Minden kiszámított átlagot az átlaggal jellemzett időszak közepéhez rendeljük. Az így nyert mozgó átlagok sora az alapirányzat értékeit, azaz a mozgó átlagolású trendértékeket adja, melyek száma kevesebb, a megfigyelt idősor adatainál. Például $k=3$ tagú mozgó átlagolásnál 2 adattal rövidül az idősorunk, az első és az utolsó megfigyelt időszakhoz nem kapunk trendértéket. A nagyobb tagszámú átlagolás, jobban kiszűri a véletlen hatást, de több adattal rövidül a trendértékek sora. Ha az idősorban szezonhatás van, a mozgó átlag tagszámát úgy választjuk meg, hogy átfogjon legalább egy, vagy több teljes periódust.

Például negyedéves szezonális esetén 4, 8, 12, ... tagú, havi szezonális esetén 12, 24, 36, ... tagú átlagokat kell választani. A trendértékek sora ekkor 4, 8, 12, ... , illetve 12, 24, 36, ... adattal rövidül. A megfelelő tagszám-választás a szezonhatás kiszűrését célozza.

Ha k páros, akkor az összeget szolgáltató időszak közepe két átlagolt érték közé esik. Ez esetben egy ismételt $k=2$ -es mozgóátlagolást, más néven centrírozást végzünk, így a kétszeri eltolódás miatt az y_t és \hat{y}_t értékek már egymásnak megfelelőek.

Az előzőek alapján a mozgó átlagolású trendszámítás lépései:

1. A periódus alapján eldöntjük, hogy hány tagú mozgóátlagot számítunk. (tagszám= k)
2. Kiszámítjuk az első k adat egyszerű számtani átlagát. Ezt az értéket az átlag által lefedett időszak közepéhez rendeljük. Ez páratlan k esetén a $(k+1)/2$ -dik időszak, páros k esetén a $k/2$ és a $(k/2)+1$ -edik időszak közé rendeljük. Ez utóbbi esetben ahhoz, hogy az eredeti idősor időszakaihoz rendelhessünk adatot, szükség van a centrírozásra.
3. Ezután elhagyjuk az első adatot, és helyette a $k+1$ -edik adattal bezárólag számítjuk ki a k -tagú egyszerű számtani átlagot, és az így adódó időszak közepéhez rendeljük az átlagot.
4. A 3. lépést ismétljük az elemzendő idősor utolsó adatáig.

3.1.2.2 Analitikus trendszámítás

A lineáris és a lineárisra visszavezethető trendfüggvények esetében a legkisebb négyzetek módszerét alkalmazzuk, vagyis olyan függvényt keresünk, amely esetében a megfigyelt és a modell által számított értékek közötti eltérés négyzetösszege minimális.⁵⁰

⁴⁸ Ld. Trendszezon-hibaszámítás.xls Excel parancsfájl.

⁴⁹ Ld. Logisztikustrendekbecslése.xls Excel parancsfájl.

⁵⁰ Hunyadi László – Vita László [2008] I. 272-273.

A legjobb közelítést tehát az a függvény adja, melyiknél a reziduumokból $(y_i - \hat{y}_i)$ számított eltérésnégyzetösszeg a legkisebb:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min$$

A feladat megoldását a klasszikus legkisebb négyzetek módszerének nevezik.⁵¹

A következőkben bemutatásra kerülő lineáris és lineárisra visszavezethető trend függvénytípusok logaritmikus transzformációkkal, illetve új változók bevezetésével linearizálhatóak, és klasszikus legkisebb négyzetek módszerével a paramétereik becsülhetők.

A következőkben néhány vizsgálati szempontot sorolunk fel:

I. A lineáris és a lineárisra visszavezethető trendfüggvények a függvényt leíró egyenlet alapján vizsgálhatók:

- monotonitás szempontjából:
 1. növekvők: pl. lineáris, parabolikus, fél-logaritmikus, hatvány alakú és parabolikus (ha az együtthatók pozitívak), exponenciális, ha $b_1 > 1$
 2. monoton csökkenők: pl. hiperbolikus, parabolikus (ha az együtthatók pl. pozitívak, de a legnagyobb fokszámú időtényezőnél a paraméter negatív), hatvány alakú ha $b_1 < 0$ és exponenciális, ha $0 < b_1 < 1$
 3. egyes függvények: pl. parabolikus trendek, (az időváltozó, a t együtthatói között található pozitív és negatív is) ahol a monoton növekedés monoton csökkenésbe vagy fordítva a monoton csökkenés monoton növekedésbe megy át.
- telítődés szempontjából:
 1. telítődési függvények: pl. féllogaritmikus, hiperbolikus, ahol a függvény egy végső határérték felé tart,
 2. telítődés nélküli függvények: pl. lineáris, parabolikus, hatvány alakú, exponenciális, ahol a növekedésnek vagy a csökkenésnek nincsenek korlátai.
- inflexiós pont szempontjából:
 1. a függvények rendelkezhetnek inflexiós ponttal: pl. harmadfokú parabolikus
 2. a függvényeknek nincs inflexiós pontjuk: pl. lineáris, féllogaritmikus, hatvány alakú, exponenciális.

Egy függvény inflexiós pontján azt értjük, hogy ebben a pontban az érintő átmetszi a görbét. Az inflexiós pont létezésének szükséges feltétele – ha a függvénynek az inflexiós pontban a harmadrendű deriváltja is létezik – az, hogy a függvény második deriváltja ebben a pontban nulla legyen, az elégséges feltétel pedig az, hogy a harmadik derivált az inflexiós pontban ne legyen egyenlő nullával. Természetesen feltételezzük, hogy a függvény az inflexiós pont környezetében háromszor differenciálható. Az inflexiós pont létezésének szükséges és elégséges feltétele az is, ha a második derivált a zérus pontjában előjelet vált, ami azt mutatja, hogy a konvex (konkáv) ívet konkáv (konvex) követi.

- a függvények nevezetes pontjait is vizsgálhatjuk, pl. melyek a trend kezdeti, (az értelmezési tartomány $t=0$ pontjában mekkora az y érték) és határfeltételei, van-e a függvénynek maximuma (pl. a parabolikus trend, ha a paraméterek előjele különböző) vagy nincs (a legtöbb vizsgált függvény esetében, pl. lineáris, féllogaritmikus, hatvány alakú, exponenciális). Vizsgálható továbbá, ha $t \rightarrow \infty$, akkor mekkora az \hat{y} érték.

II. A függvényt közelítő egyenes meredekségéből, az úgy nevezett deriváltból következtethetünk:

$$f'(t) > 0$$

$$f'(t) < 0$$

⁵¹ Statisztika. [2007] Szerk: Pintér József – Rappai Gábor. 164.

- a függvény növekedésének irányára azaz, hogy monoton növekvő (a t időtengely mindegyik pontjában az első derivált $f'(t)$ nem negatív) vagy monoton csökkenő, (a t időtengely mindegyik pontjában az első derivált $f'(t)$ nem pozitív):
- a növekedés mértékére (gyorsan változik-e a függvény vagy lassan). A differencia hányados az idősorok trend vizsgálatánál a következő képlettel közelíthető, felhasználva az analízisben tanult összefüggést:

$$\alpha_t = \frac{dy}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Delta y}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{y(t + \Delta t) - y(t)}{\Delta t} \approx (y_t - y_{t-1}) / (t - [t-1]) = (y_t - y_{t-1})$$

- esetleges szélsőértékére (van-e abban a pontban a függvénynek maximuma vagy minimuma), az alábbi összefüggés szerint, ahol a második derivált t szerint $f''(t)$:

$$f'(t_{\max}) = 0$$

$$f''(t_{\max}) < 0$$

$$f'(t_{\min}) = 0$$

$$f''(t_{\min}) > 0$$

Lineáris (lin - lin⁵²) trendszámítás^{53 54 55}

Az idősorban a változás tendenciája egyenes vonallal jól leírható, ha a szomszédos időszakok közötti abszolút változás (növekedés vagy csökkenés) viszonylag állandó az időben.

A függvényt leíró formula:

$$\hat{y}_t = b_0 + b_1 t$$

Ahol: $b_1 \neq 0$ mert ha $b_1=0$ akkor a trendfüggvény konstans: b_0 .

Regressziós modellnél⁵⁶ (pl. Excel) a

Bemeneti Y tartomány: y_t

Bemeneti X tartomány: t (pl. $t = 1, 2, 3, \dots, N$)⁵⁷⁵⁸

A b_0 és b_1 a lineáris trendfüggvény ismeretlen paraméterei. A b_1 becült paraméter megmutatja azt, hogy a vizsgált időszakban a vizsgált jelenség időegységenként (pl. a megfigyeléseknek megfelelően: évenként, hónaponként stb.) átlagosan hány egységnyivel változott. A lineáris trend t szerinti deriváltja ugyanis b_1 . A növekedés (ha $b_1 > 0$) illetve csökkenés (ha $b_1 < 0$) átlagos abszolút értéke illetve iránytangense időegységre vetítve tehát állandó. A b_1 paraméter jelentése megegyezik a \bar{D} (átlagos abszolút változás) mérőszám jelentésével. A két mutató abban különbözik egymástól, hogy a b_1 meghatározásánál, a legkisebb négyzetek módszerének felhasználásával, a megfigyelt adatokhoz legjobban illeszkedő egyenest választjuk ki. A \bar{d} mutató esetében viszont az egyenes meredekségét (iránytangensét) az első és az utolsó adat alapján határozzuk meg. Az időszakonkénti változások átlagára ezért a b_1 megbízhatóbb becslést ad, mint a \bar{D} mutató. A b_0 becült paraméter a tengelymetszet és ezt az értéket akkor veszi fel a trend, ha $t=0$.

Ha $t = 0$ akkor: $\hat{y}_{t=0} = b_0$

⁵² A lin-lin: y_t lin=lineáris, a t =lin=lineáris ld.: Ramanathan Ramu [2003]: 519-521.

⁵³ Az α_t , β_t , ϵ_t becslésére a lineáris és lineárisra visszavezethető trendek mindegyikére Excel parancsfájlokat dolgoztunk ki, ld. Trendtípus(neve=lineáris, féllogaritmikus stb.)ciklus.xls fájlokat, Lehet változtatni a mozgó átlag tag-számát is.

⁵⁴ Kerékgyártó Györgyné – Mundruczó György [1995]: 460-463.

⁵⁵ Hunyadi László – Vita László [2002]: 518-519.

⁵⁶ A regressziós modelleket a 4. fejezetben tárgyaljuk.

⁵⁷ A t időváltozó egymástól egyenlő távolságra lévő értékek sorozata.

⁵⁸ A gyakorlati számításokban $t=1, 2, \dots, n$

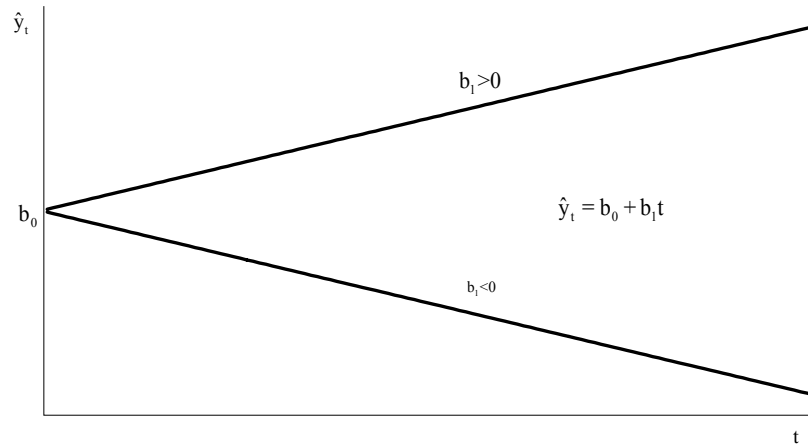
Azt feltételezzük, hogy a múltbeli folyamatok folytatódnak a jövőben. A lineáris trend az extrapolációnál állandónak veszi az átlagos abszolút növekményt, ami hosszabb távon ritkán teljesülő feltétel, mivel a lineáris trendfüggvény a végtelenbe tart, ha az idő is a végtelenbe tart.

Ha: $t \rightarrow +\infty$ és ha $b_1 > 0$

$$\hat{y}_t (t \rightarrow +\infty) \rightarrow +\infty$$

Ha: $t \rightarrow +\infty$ és ha $b_1 < 0$

$$\hat{y}_t (t \rightarrow +\infty) \rightarrow -\infty$$



3-1. ábra: A lineáris trend

Féllogaritmikus (szemi-logaritmikus, féllogaritmusos, lin-log) trend

Előfordul a gyakorlatban, különösen a hosszabb távú extrapolációnál olyan összefüggés, amelyben a megfigyelt idősor (y_t) természetes értéke és az időváltozó (t) logaritmusa között írható fel egy lineáris modell. A féllogaritmikus trend az extrapolációnál sok esetben jobb eredményt ad, mint a lineáris trend, mert nem tekinti az átlagos abszolút növekményt állandónak, hanem feltételezi, hogy a növekmény nagysága az időben előrehaladva csökken.

A függvényt leíró formula:

$$\hat{y}_t = b_0 + b_1 \ln t$$

Ahol: $b_1 \neq 0$ mert ha $b_1 = 0$ akkor a trendfüggvény konstans b_0 .

Regressziós modellnél (pl. EXCEL) a

Bemeneti Y tartomány: y_t

Bemeneti X tartomány: $\ln t$ (pl. $t = 1, 2, 3, \dots, N$)

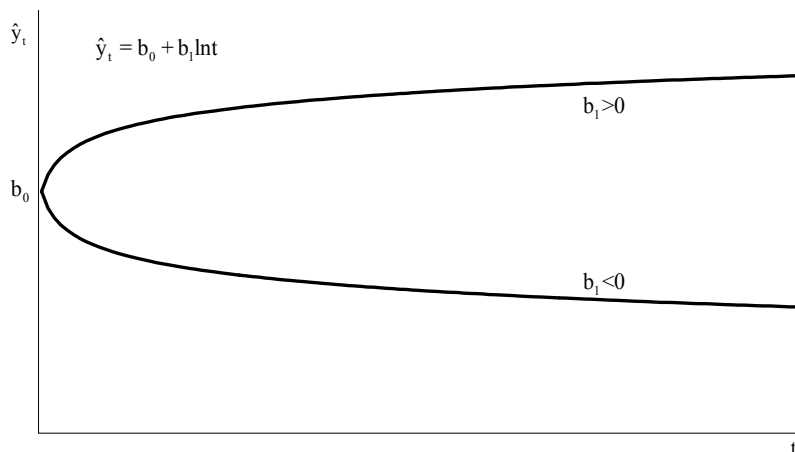
A növekedés (ha $b_1 > 0$) illetve csökkenés (ha $b_1 < 0$) átlagos abszolút értéke időegységre vetítve csökkenő, ugyanis, pl.: $\ln 1 = 0$, $\ln 2 = 0,693$, $\ln 3 = 1,098, \dots, \ln 10 = 2,302, \dots, \ln 100 = 4,605, \dots, \ln 1000 = 6,907$.

Ha

$t=1$

akkor:

$$\hat{y}_{t=1} = b_0$$



3-2. ábra: A féllogaritmikus trend

Másodfokú polinomiális (másodfokú parabolikus, kvadratikus) trend

Sok esetben használhatjuk elemzésre és előrejelzésre a polinomiális trendet, amely általában feltételezi, hogy a nemlineáris folyamatok alakulásában fordulópont van, pl. az idősorban tendenciaváltás tapasztalható, vagyis az idősor növekedésből, hullámhegyből csökkenésbe, hullámvölgybe – vagy fordítva - megy át, akár ismétlődően is. Értelemszerű, hogy a fokszám növelése egyre jobb illesztést ad, de megállapítható, hogy a 3 fokszámnál magasabb fokszám alkalmazása már igen nehezen indokolható. A trendszámításnál az elfogadott gyakorlat szerint a polinom fokszáma 2 vagy 3 lehet. A polinom fokszámának növekedésével ugyan nő az illesztés pontossága, de egyre inkább elveszítjük a valódi tartós irányzatot és a polinom „követi” a szezonális, a ciklikus és a véletlen komponenseket is. A másodfokú polinomiális trend azt feltételezi, hogy az idősorban maximum egy fordulópont (egy hullámhegy és egy hullámvölgy) van.

A függvényt leíró formula:

$$\hat{y}_t = b_0 + b_1 t + b_2 t^2$$

Ahol: $b_2 \neq 0$ mert ha $b_2 = 0$ akkor a trend lineáris.

Ha $b_2 = 0$ és $b_1 = 0$ akkor pedig a trendfüggvény konstans b_0 .

Regressziós modellnél (EXCEL) a

Bemeneti Y tartomány: y_t

Bemeneti X tartomány: t t^2 (pl. $t = 1, 2, 3, \dots, N$)

Legyen: $b_0 > 0, b_1 > 0, b_2 < 0$

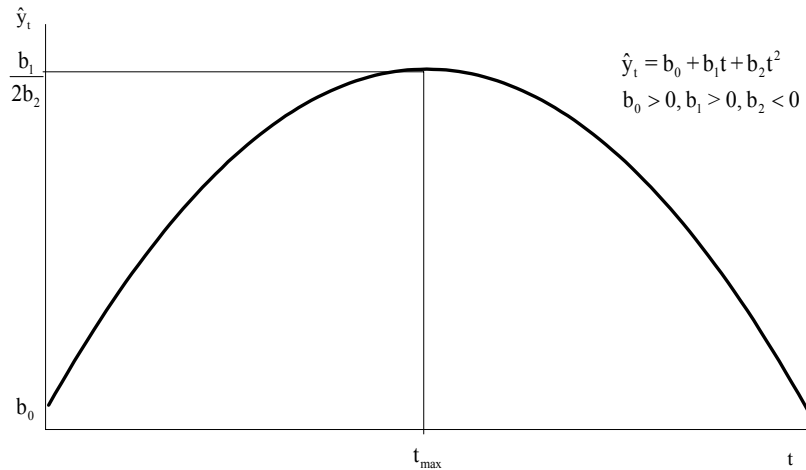
$\hat{y}_t = b_0 + b_1 t + b_2 t^2$ tehát a függvény fordított U (\cap) alakú.

Akkor a függvény maximuma:

$$\frac{d\hat{y}}{dt} = b_1 - 2b_2 t = 0$$

$$t_{\max} = \frac{b_1}{2b_2}$$

A t_{\max} helyen a második derivált t szerint negatív ($-2b_2$), tehát a függvénynek maximuma van.



3-3. ábra: A másodfokú polinomiális trend

Legyen: $b_0 > 0, b_1 < 0, b_2 > 0$

$\hat{y}_t = b_0 + b_1t + b_2t^2$ tehát a függvény U alakú.

Akkor a függvény minimuma:

$$\frac{d\hat{y}}{dt} = -b_1 + 2b_2t = 0$$

$$t_{\min} = \frac{b_1}{2b_2}$$

A t_{\min} helyen a második derivált t szerint pozitív ($2b_2$), tehát a függvénynek minimuma van.

Harmadfokú polinomiális (harmadfokú parabolikus) trend

A harmadfokú polinomiális trend azt feltételezi, hogy az idősorban maximum egy –vagy két fordulópont (egy vagy két hullámhegy és egy vagy két hullámvölgy) van.

A függvényt leíró formula:

$$\hat{y}_t = b_0 + b_1t + b_2t^2 + b_3t^3$$

Ahol: $b_3 \neq 0, b_2 \neq 0$ mert ha $b_3=0, b_2=0$ akkor a trend lineáris. Értelmszerűen $b_1 \neq 0$.

Regressziós modellnél (pl. EXCEL) a

Bemeneti Y tartomány: y_t

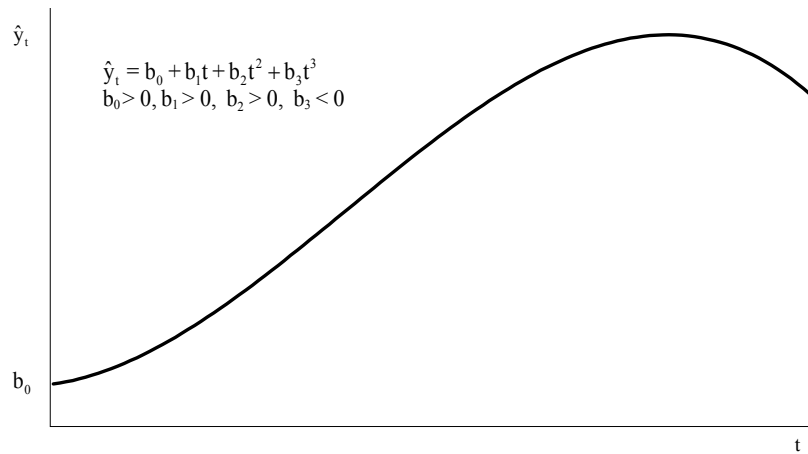
Bemeneti X tartomány: $t \ t^2 \ t^3$ (pl. $t = 1, 2, 3, \dots, n$)

Legyen:

$b_0 > 0, b_1 > 0, b_2 > 0, b_3 < 0$

Akkor:

$$\hat{y}_t = b_0 + b_1t + b_2t^2 + b_3t^3$$



3-4. ábra: A harmadfokú polinomiális trend

Hatvány alakú (log - log) trend

A függvényt leíró formula:

$$\hat{y}_t = b_0 t^{b_1}$$

$$\ln \hat{y}_t = \ln b_0 + b_1 \ln t$$

Regressziós modellnél (pl. EXCEL) a

Bemeneti Y tartomány: $\ln y_t$

Bemeneti X tartomány: $\ln t$ (pl. $t = 1, 2, 3, \dots, n$)

A hatványalakú trend ábráján látható, hogy a leírt tendencia a b_1 nagyságától függ.

A b_1 lehet:

$$b_1 > 1$$

$$0 < b_1 < 1$$

$$b_1 = 1$$

$$b_1 < 0$$

$b_1 = 0$ akkor:

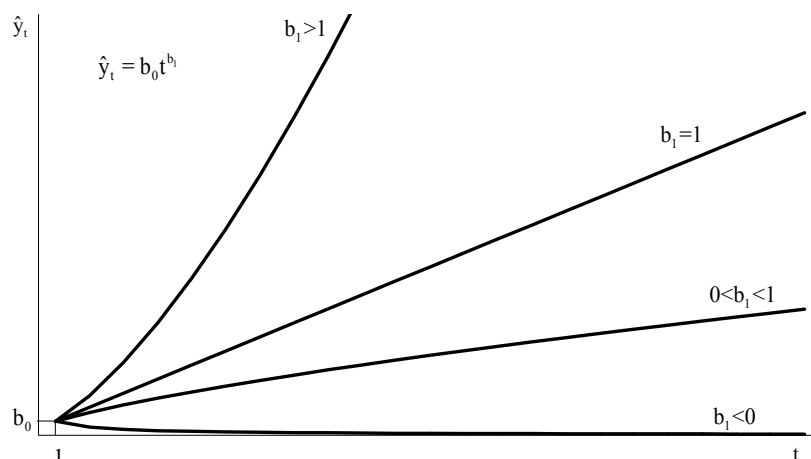
$$\hat{y}_t = b_0$$

Ha: $t \rightarrow +\infty$ és ha $b_1 > 0$

$$\hat{y}_t (t \rightarrow +\infty) \rightarrow +\infty$$

Ha: $t \rightarrow +\infty$ és ha $b_1 < 0$

$$\hat{y}_t (t \rightarrow +\infty) \rightarrow 0$$



3-5. ábra: A hatvány alakú trend

Exponenciális (log - lin) trend

Az exponenciális trendnél a relatív változások (a láncviszonyyszámok) mutatnak viszonylagos állandóságot (stabilitást). Általában a közép és hosszú távú gazdasági és társadalmi folyamatok jellemzésének alapmodellje. Akkor alkalmazzuk, ha feltételezhető, hogy egységnyi időváltozás hatására a folyamat változása relatíve állandó, azaz a vizsgált időszakban a megfigyelések az előző értékhez képest rendre megközelítően azonos százalékos növekedést vagy csökkenést mutatnak.

A függvényt leíró formulák:

$$\hat{y}_t = b_0 b_1^t$$

$$\ln \hat{y}_t = \ln b_0 + t \ln b_1$$

Az Excel diagram – trend által számolt exponenciális trend formulája:

$$\hat{y}_t = b_0 e^{ct}$$

$$\hat{y}_t = b_0 b_1^t$$

$$b_1 = e^c$$

A függvény helyettesítési értéke a $t=0$ helyen (y_0):

$$\hat{y}_{t=0} = b_0$$

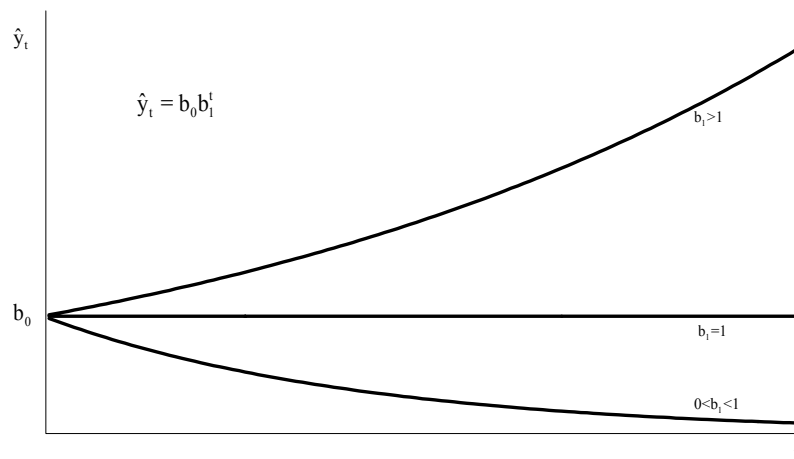
Regressziós modellnél (pl. EXCEL) a

Bemeneti Y tartomány: $\ln y_t$

Bemeneti X tartomány: t (pl. $t = 1, 2, 3, \dots, n$)

A szomszédos értékek hányadosa, tehát a növekedés átlagos üteme állandó.

$$\frac{\hat{y}_t}{\hat{y}_{t-1}} = \frac{b_0 b_1^t}{b_0 b_1^{t-1}} = b_1$$



3-6. ábra: Az exponenciális trend

A duplázódás/felezési idő számítása

Az extrapolációnál figyelembe kell vennünk, hogy nem biztos az, hogy a múltat jól leíró trend a jövőben is igaznak bizonyul. Pl. a fejlődés hosszabb távon [50-200 év] nem írható le lineáris vagy exponenciális trenddel, mivel érvényesülnek az évszázados trendek és hosszú ciklusok. Rövidebb távon [7-30 év], ha a fejlődés exponenciális, meghatározható a duplázódási idő⁵⁹:

$$\hat{y}_t = b_0 (1 + p)^t$$

Ahol:

⁵⁹ Korán Imre [1978]: 22-23.

b_0 = a bázisérték,

p = az éves növekedési/csökkenési ütem,

$b_1 = 1 + p$,

így a **duplázódási idő (t)** a

$$2b_0 = b_0 (1 + p)^t$$

formulából becsülhető, ha $p > 0$ illetve $b_1 > 1$, akkor a duplázódási idő:

$$\ln(2) = t \ln(1 + p)$$

$$t = \ln(2) / \ln[1 + p].$$

A **felezési idő** ha $-1 < p < 0$ illetve $0 < b_1 < 1$

$$0,5b_0 = b_0 (1 + p)^t$$

$$t = \ln(0,5) / \ln(1 + p)$$

Példa⁶⁰:

ha $p = 0,05$ akkor $b_1 = 1,05$ a duplázódási idő: $t = 14,2$

ha $p = 0,1$ akkor $b_1 = 1,10$ a duplázódási idő: $t = 7,27$

Az éves növekedési ütem kétszeresére nőtt, aminek eredményeképpen a duplázódási idő közel felére csökkent.

Példa:

ha $p = -0,05$ akkor $b_1 = 0,95$ a felezési idő: $t = 13,51$

ha $p = -0,1$ akkor $b_1 = 0,90$ a felezési idő: $t = 6,57$

Az éves növekedési ütem kétszeresére csökkent, aminek eredményeképpen a felezési idő közel felére csökkent.

Hiperbolikus trend

Gyakran előfordul, hogy az idősor aszimptotikusan közelít egy adott értéket. Ekkor trendfüggvényként valamelyik bemutatásra kerülő hiperbolikus függvény alkalmazható. Az önköltséget, az árak alakulását jellemző folyamatok gyakran modellezhetők e módon.

A függvényt leíró formula:

$$\hat{y}_t = b_0 + b_1 \frac{1}{t} = b_0 + b_1 t^{-1}$$

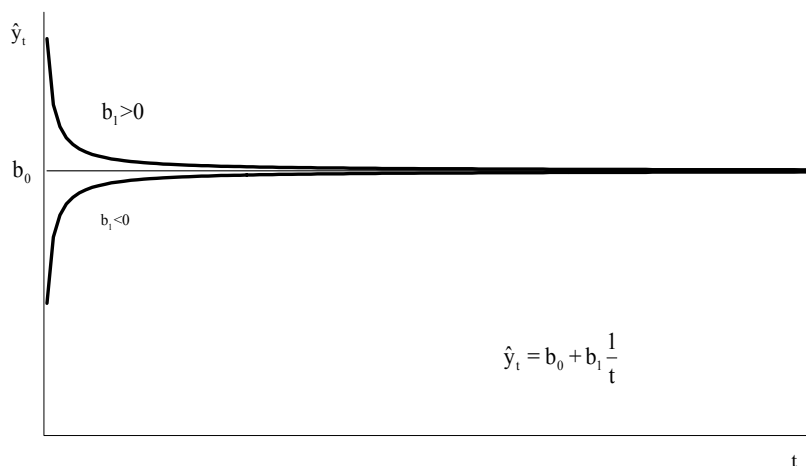
Regressziós modellnél (pl. EXCEL) a

Bemeneti Y tartomány: y_t

Bemeneti X tartomány: $1/t$ (pl. $t = 1, 2, 3, \dots, n$)

Ha $t = 1$ akkor $\hat{y}_{t=1} = b_0 + b_1$

Ha $t \rightarrow \infty$ akkor $\hat{y}_{t \rightarrow \infty} \rightarrow b_0$



3-7. ábra: A hiperbolikus trend

⁶⁰ A trendszezeszt.xls parancsfájl felhasználásával.

Elsőfokú hiperbolikus trend

A függvényt leíró formula, a lineáris trend reciproka:

$$\hat{y}_t = \frac{1}{b_0 + b_1 t} = (b_0 + b_1 t)^{-1}$$

$$\frac{1}{\hat{y}_t} = b_0 + b_1 t$$

Ha $t = 0$ akkor $\hat{y}_{t=0} = \frac{1}{b_0}$

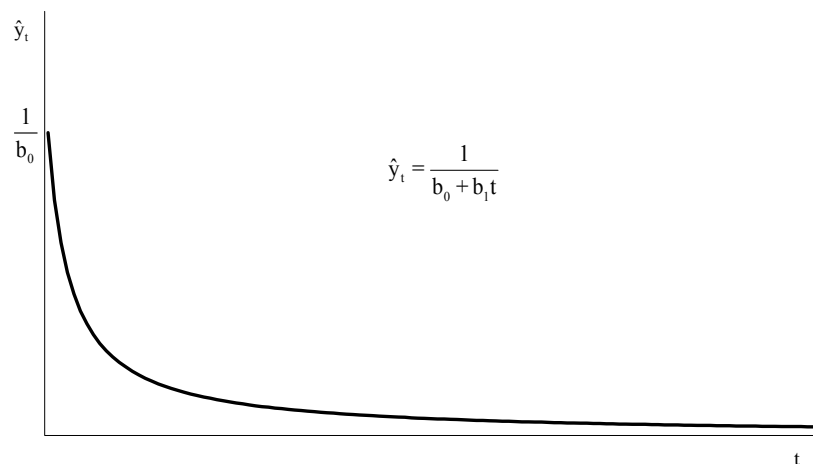
Ha $t = 1$ akkor $\hat{y}_{t=1} = \frac{1}{b_0 + b_1}$

Ha $t \rightarrow \infty$ akkor $\hat{y}_{t \rightarrow \infty} \rightarrow 0$

Regressziós modellnél (pl. EXCEL) a

Bemeneti Y tartomány: $1/y_t$

Bemeneti X tartomány: t (pl. $t = 1, 2, 3, \dots, n$)



3-8. ábra: Az elsőfokú hiperbolikus trend

Másodfokú hiperbolikus trend

A függvényt leíró formula, a másodfokú parabolikus trend reciproka:

$$\hat{y}_t = \frac{1}{b_0 + b_1 t + b_2 t^2} = (b_0 + b_1 t + b_2 t^2)^{-1}$$

$$\frac{1}{\hat{y}_t} = b_0 + b_1 t + b_2 t^2$$

Ha $t = 0$ akkor $\hat{y}_{t=0} = \frac{1}{b_0}$

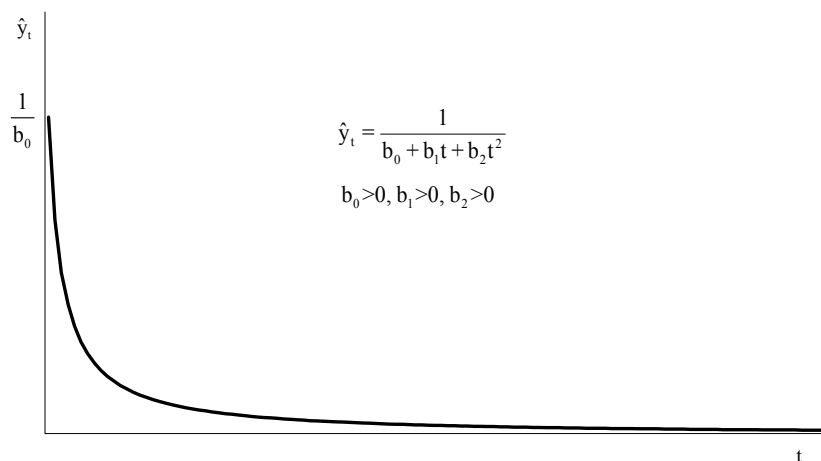
Ha $t = 1$ akkor $\hat{y}_{t=1} = \frac{1}{b_0 + b_1 + b_2}$

Ha $t \rightarrow \infty$ akkor $\hat{y}_{t \rightarrow \infty} \rightarrow 0$

Regressziós modellnél (pl. EXCEL) a

Bemeneti Y tartomány: $1/y_t$

Bemeneti X tartomány: t t^2 (pl. $t = 1, 2, 3, \dots, n$)



3-9. ábra: A másodfokú hiperbolikus trend

3.1.3 A szabályos rövid távú (szezónális) ingadozás

Ha a szezonális hullámmozgás kitérései, amplitúdói abszolút értelemben vagy relatív (a trendhez viszonyítva) értelemben állandóságot mutatnak, akkor állandó szezonálisról beszélünk.

Ha a periódus (i) hossza az év, ezen belül a szezonkénti (j) hossza lehet pl. 4 negyedév, 12 hónap, 52 hét, 365 nap, 230-252 munkanap, 250-252 tőzsdenap.

Ha a periódus (i) hossza a hónap ezen belül a szezonkénti (j) hossza lehet 4 hét, 28-31 nap.

Ha a periódus (i) hossza a hét ezen belül a szezonkénti (j) hossza lehet 7 nap, 5 munkanap.

Ha a periódus (i) hossza a nap ezen belül a szezonkénti (j) hossza lehet 24 óra.

Az **additív modell** esetén tehát azt tapasztaljuk, hogy a különböző periódusok azonos szezonjában, a trendtől mért eltérések nagysága megközelítőleg ugyanakkora. Mivel a szezonális hullámmozgást az alapirányzathoz képest jelentkező szisztematikus pozitív és negatív eltéréseként definiáltuk, elvárható követelmény, hogy egy teljes perióduson belül kiegyenlítsék egymást. Ezért additív modell esetén a szezonális eltérésekre vonatkozó követelmény úgy írható fel, hogy

$$\sum_{j=1}^m s_j^* = 0.$$

A véletlen komponensre hasonló követelmény írható fel, eszerint

$$\sum_{t=1}^N v_t^* = \sum_{i=1}^n \sum_{j=1}^m v_{ij}^* = 0$$

vagyis „megköveteljük”, hogy a véletlen tag ne eredményezzen szisztematikus eltérést az alapirányzathoz képest.

A **multiplikatív modell** logikájának megfelelően, a szezonindexekre vonatkozó követelmény úgy írható fel, hogy:

$$\frac{\sum_{j=1}^m s_j}{m} = 1$$

A véletlen komponensre hasonló követelmény írható fel, eszerint

$$\sum_{t=1}^N v_t = \sum_{i=1}^n \sum_{j=1}^m v_{ij} = 1$$

Összefoglalóan megállapítható, hogy a szezonális eltérítő hatása a megfelelő szezonoknál additív modellben abszolút állandóságot, multiplikatív modellben a trendhez mért relatív állandóságot mutat.

A szezonkénti meghatározása két lépésben történik:

- a trendhatás leválasztásával, és
- a véletlen hatás kiszűrésével

A szezonális hatás számszerűsítését multiplikatív és additív modell esetén együtt mutatjuk be és most eltekintünk a konjunktúra ciklusok modellezésétől. Kiindulva a modellekből

$$y_{ij} = \hat{y}_{ij} * s_j * v_{ij}$$

$$y_{ij} = \hat{y}_{ij} + s_j^* + v_{ij}^*$$

először a már előzetesen számszerűsített trendhatást szűrjük ki a megfigyelt idősorunkból. Ez a két alapmodellnek megfelelően az alábbiak szerint történik:

1. A megfigyelt és trend-értékek hányadosai, illetve különbségei, feltételezéseink szerint már csak a szezon- és véletlen hatást tartalmazzák.
2. A második lépésben a véletlen hatás kiszűrését kívánjuk elvégezni, az előbbieken nyert hányadosok, illetve különbségek, azonos szezonra eső értékeinek átlagolásával.

A két alap-modellből így – a szezonok számának megfelelően m számú – nyers szezonindexekhez az Excel parancsfájlból a jele: Index, illetve nyers szezonális eltérésekhez a jele: Eltérés jutunk.

$$s_j = \frac{1}{n} \sum_{i=1}^n \frac{y_{ij}}{\hat{y}_{ij}}$$

$$s_j^* = \frac{1}{n} \sum_{i=1}^n (y_{ij} - \hat{y}_{ij})$$

A szezonkomponensekre megfogalmazott követelményteljesülését vizsgálni kell mindkét modell esetén. Ha a követelmény nem teljesül, a nyers szezon-tényezőket korrigálni kell. Korrekciós tényező a két modellben:

$$\bar{s} = \frac{\sum_{j=1}^m s_j}{m}$$

$$\bar{s}^* = \frac{\sum_{j=1}^m s_j^*}{m}$$

ami nem más, mint az m számú nyers szezonindex, illetve -eltérés egyszerű számtani átlaga.

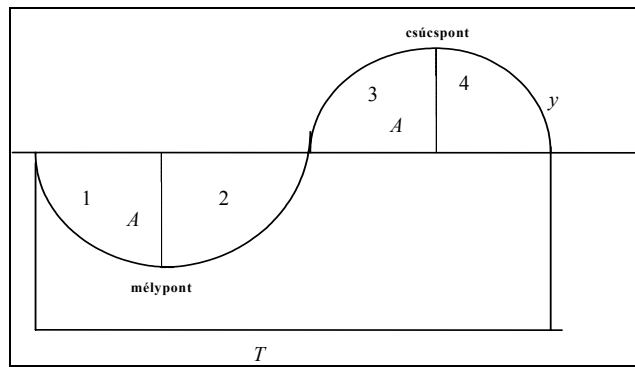
A tisztított szezonindexeket (A jele: Korr. Index) úgy nyerjük, hogy a nyers szezonindexeket a korrekciós tényezőjükkel rendre elosztjuk. Hasonlóan nyerjük a tisztított szezonális eltéréseket (A jele: Korr. Elt.), a nyers szezonális eltérésekből rendre levonva korrekciós tényezőjüket. Az alap-modellekből számszerűsített szezonkomponens értelmezése: bármely periódus j -edik szezonjában a szezonális hatást a trendhez képest módosítja az időszori értékeket. A módosítás a multiplikatív modellben s_j -szeres növelést, illetve csökkentést jelent; a szezonindexeket százalékban kifejezve, a 100% feletti rész a százalékos növelést, a 100%-nál kisebb szezonindexek esetén a 100%-ra kiegészítő érték a százalékban kifejezett csökkenés mértékét adja meg. A módosítás additív modellben az s_j^* -nek megfelelő mértékű növelését, vagy csökkentését jelenti az idősor értékeinek, az s_j^* előjelétől függően, az idősor adatainak nagyságrendjében és mértékegységében.

3.1.4 A ciklikus (periodikus) mozgás modellezése.*

A ciklikus (periodikus) mozgás sémáját az alábbi ábra mutatja. Ez az ábra a konjunktúraciklus elméleti alapját, a harmonikus rezgőmozgást mutatja be, és a fizikából ismert harmonikus rezgés modelljére épül, melyben az

1. szakasz a pangás, (depresszió),
2. szakasz a megélénkülés, (expanzió),
3. szakasz a fellendülés, (prosperitás),
4. szakasz pedig a válság (recesszió vagy hanyatlás) időszaka.

Ha csak egy fellendülő és egy visszaeső fázist különböztetnek meg, gyakran a következő terminológiát használják: felszálló ág és leszálló ág, alacsonyabb fordulópont vagy hullámvölgy (mélypont), fellendülés vagy növekedés, magasabb fordulópont (csúcs), és visszaesés vagy csökkenés. A gyakorlati tapasztalatok szerint a ciklusok periódusa és amplitúdója változik.



3-10. ábra: A ciklikus (periodikus) mozgás sémája

A hullámhossz egy teljes hullámnak a hossza, pl. a csúcsponttól a csúcspontig, vagy a mélyponttól a mélypontig:

$$\lambda = \frac{v}{f} = vT$$

Ez a megközelítés tehát a Newton-féle „akció egyenlő reakció”, illetve „hatás egyenlő ellenhatás” elvből indul ki, vagyis azt feltételezi, hogy a gazdasági életben – éppen úgy, mint a fizika hullámjelenségeiben – az egyensúlyi helyzetből való kilengést az abba való visszatérés jelensége követi, majdnem mechanikus módon. Ez a fizikai modell természetesen elméleti, és így egy ideális megvalósulást ír le, a gyakorlatban a ciklus képe eltér a fenti szabályos mintától.

Ahol:

y = kitérés, az egyensúlyi helyzettől mért távolság,

A = amplitúdó, a nyugalmi helyzettől mért legnagyobb kitérés (mélypont, illetve csúcspont),

T = periódus (rezgésidő),

ω = körfrekvencia, a frekvencia 2π -szerese,

f = frekvencia (gyakoriság), a rezgések számának és időtartamának hányadosa, amely megadja az egységnyi idő alatt történt rezgések számát:

$$f = \frac{1}{T} = \frac{\omega}{2\pi}$$

v = sebesség:

$$v = A\omega \cos(\omega t) = A\left(\frac{2\pi}{T}\right) \cos\left(\frac{2\pi}{T} t\right)$$

A nemzetközi szakirodalom a következő öt konjunktúra-ciklust feltételezi és különbözteti meg:⁶¹

1. a 3–5 éves leltár (készlet) vagy Kitchin-ciklus;
2. a 7–11 éves állandó befektetési (gépi beruházási) vagy Juglar-ciklus;
3. a 15–25 éves építési vagy Kuznets-ciklus;
4. a 45–60 éves hosszú vagy Kondratyev-ciklus;
5. a 100 évnél hosszabb évszázados vagy szekuláris trendek

A ciklusok periódusa tehát duplázódhat, pl. egy Kondratyev ciklus [57 év] tartalmazhat 6 Juglart [9,5 év] és egy Juglar tartalmaz 3 Kitchint [3,16 év]. Ha például a Kondratyev ciklus hosszát [periódusát] átlagosan 54 évnak vesszük, és a Kuznets ciklust 18 évesnek állítjuk, a Juglar ciklust 9 évesnek, a Kitchin ciklust 4,5 évesnek vesszük, akkor a kapcsolat teljesen tiszta:

1 Kondratyev ciklus = 3 Kuznets ciklus = 6 Juglar = 12 Kitchin.

Egyszerű technikai eljárásokkal a ciklusokat részmozgásokra oszthatjuk, egyiket-másikat kiszűrhetjük a vizsgálni kívánt mozgás kimutatása érdekében. A trend a ciklus kiküszöbölésével felfedhető (például mozgóátlagolással, grafikus becsléssel, vagy a szokásos legkisebb négyzetek módszerének alkalmazásával). Kondratyev vizsgálati módszerének az a lényege, hogy az árakat egyszerű statisztikai indexszel ábrázolja, egyes pénzügyi (kamatrátá, bérék), vegyes jellegű (külkereskedelmi forgalom), illetve tisztán naturális sorok esetében a trendtől való eltérés számítási módszerét alkalmazza. Az utóbbiaknál (külkeres-

⁶¹ Országonként, vizsgált mutatóként és időszakonként jelentősek a különbségek a periódusok alakulásában.

kedelem és termelés, valamint fogyasztás) mindig egy főre jutó adatokat használ, és a legkisebb négyzetek módszerével számított trendtől való eltéréseket vizsgálja úgy, hogy 9 éves mozgóátlagolással megpróbálja kiszűrni a rövidebb ciklusú mozgásokat. Ennek a megközelítésnek az a lényege, hogy ne az egyedi, hanem a hosszútávon érvényesülő összetett hatásokat ragadjuk meg, és ezáltal tudjunk a múltban érvényesült, és részben a jövőre is várható tendenciák segítségével hozzátenni valamit a rövid távú szakmai elemzésekhez. A történelem folyamán a modernkori Európában a Kondratyev - ciklusok a következők szerint alakultak⁶²⁻⁶³:

A Kondratyev-ciklusok alakulása az elmúlt két évszázadban

Felszálló ág kezdete	Csúcspont	Leszálló ág vége	Periódus
1790	[1815]	1850	60
1850	[1875]	1896	46
1896	[1929]	1945	49
1945	[1973]	1996 ?	51
1996	[2020] ?		

Az évszázados trendek alakulása ugyanakkor a következő volt:

Az évszázados trendek alakulása

Felszálló ág kezdete	Csúcspont	Leszálló ág vége	Periódus
1250	[1350]	1510	260
1510	[1650]	1740	230
1740	[1817]	1896	156
1896	[1973]	2030 ?	134 ?

Láthattuk, hogy a ciklusok időtartama (periódusa) duplázódik. Ugyanakkor a különböző időtartamú ciklusok egyidejűek, keverednek, mozgásukkal csökkentik, vagy növelik az egész hullámmás amplitúdóját. Ha például az évszázados trend felszálló ága találkozik a Kondratyev ciklus leszálló ágával, akkor ez a válságot mérsékli, ellenkező esetben erősíti. Itt is érvényesül a fizikából ismert interferencia jelensége, illetve törvénye. Egyszerű technikai eljárásokkal a ciklusokat részmozgásokra oszthatjuk, egyiket-másikat kiszűrhetjük a vizsgálni kívánt mozgás kimutatása érdekében.

A hosszú ciklus kimutatása:

Az idősor elemei a **hosszú ciklus** vizsgálatánál a következők lehetnek:

1. Évszázados trend
2. 3-9-27 éves Kitchin-, Juglar-, és a Kuznets- ciklusok
3. Kondratyev ciklus
4. Véletlen hatás

Az idősor hossza legalább 100 év, ebben az esetben egy évszázados trend és két hosszú-ciklus mutatható ki. A hosszú ciklusok vizsgálatánál először az eredeti idősor 9 vagy 27 tagú mozgóátlagát vesszük. A 9 tagú mozgóátlagolással a 3 éves periódusú Kitchin- és a 9 éves periódusú Juglar-, illetve 27 tagú mozgóátlagolásnál a Kuznets-ciklust is kiküszöböljük. A rövidebb ciklusok természetesen – az előzőekben ismertettek szerint – különböző periódusúak lehetnek, pl. 4-8-16 évesek, ebben az esetben 16 tagú⁶⁴ mozgóátlagolással küszöbölhető ki a Kitchin, a Juglar és a Kuznets-ciklus. A mozgóátlagolással a véletlen hatást is kiszűrjük. A mozgóátlagolás csak akkor küszöböli ki a periodikus hullámmást, ha a mozgóátlag tagszáma a periódus hosszával vagy egészszámú többszörösével egyenlő. Ha nagyobb tagszámot választunk, akkor a hullámmás ellentétes lesz a ciklussal, ha kisebb tagszámot választunk, mint a ciklus periódusa, akkor csak tompítjuk a hullámmást. A grafikus ábrát ezért alaposan kell elemezni, hogy a megfelelő mozgóátlag tagszámot kiválasszuk. Ezt követi a trend kiküszöbölése. Additív kapcsolat esetén a trendet az eredeti idősorból kivonjuk, multiplikatív kapcsolat esetén az eredeti idősort a trenddel osztjuk. Eljárha-

⁶² A felszálló ág kezdete – [csúcspont] – a leszálló ág vége és a következő felszálló ág kezdete.

⁶³ Ld. pl. Braudel F. [1972], Braudel F. [2004].

⁶⁴ $A \sin(t+n2\pi) = \sin(t)$ ahol: $n=1,2,..$ és a $\sin(t)$ függvény periódusának a tartalma: $2\pi=360^\circ$ ugyanis $\sin(t+2\pi) = \sin(t)$.

tunk úgy is, hogy először a trendet küszöböljük ki, s ezt követi a mozgóátlagolás. A két eljárás [sorrend], azonos eredményre vezet.

Feltételeztük, hogy a rendelkezésre álló empirikus idősorban (y_i $i = 1, 2, \dots, N$ éves adatokkal dolgozunk, így nincs szezonális hatás az idősorban) a következő tényezők különböztethetők meg:

1. \hat{y}_i = az évszázados trendérték
2. c_1^* = a becsült (*) konjunktúra – ciklus
3. k_m^* = a Kondratyev féle hosszú ciklus
4. v_{ij}^* = a becsült (*) véletlen hatás.

Ahol:

- $i = 1, 2, \dots, N$ = az évek száma, legalább 100 év
- $m = 1, 2, \dots, p$ = a Kondratyev féle hosszú ciklus periódusa (pl. 45-50 év)
- $l = 1, 2, \dots, o$ = a rövidebb konjunktúra ciklusok periódusa (3-27 év)

A * index additív kapcsolatot jelez, ha nincs * index, akkor a kapcsolat multiplikatív.

A kapcsolódás módja lehet additív:

$$y_i = \hat{y}_i + c_1^* + k_m^* + v_{ij}^*$$

Ebből a hosszú periódusú ciklus:

$$k_m^* = y_i - (\hat{y}_i + c_1^* + v_{ij}^*)$$

A kapcsolódás módja lehet multiplikatív:

$$y_i = \hat{y}_i * c_1 * k_m * v_{ij}$$

Ebből a hosszú periódusú ciklus:

$$k_m = y_i / (\hat{y}_i * c_1 * v_{ij})$$

A rövid (Kitchin) ciklus kimutatása:

Az idősor elemei a rövid ciklus vizsgálatánál a következők lehetnek:

1. Trend, ami lehet valamely hosszú ciklus fel, vagy leszálló ága
2. 3-5 éves Kitchin ciklus
3. Szezonális hullámmás
4. Véletlen hatás

Ha csak rövidebb ciklusokat vizsgálunk, akkor a szezonhatást küszöböljük ki mozgóátlagolással, s ezután a trendhatást. Havi adatoknál 12 tagú, negyedéves adatoknál 4 tagú mozgóátlagolást alkalmazunk. A korábbi jelöléseket alkalmazva:

A kapcsolódás módja lehet additív:

$$y_{ij} = \hat{y}_{ij} + s_j^* + c_1^* + v_{ij}^*$$

Ebből a rövid periódusú ciklus:

$$c_1^* = y_{ij} - (\hat{y}_{ij} + s_j^* + v_{ij}^*)$$

A kapcsolódás módja lehet multiplikatív:

$$y_{ij} = \hat{y}_{ij} * s_j * c_1 * v_{ij}$$

Ebből a rövid periódusú ciklus:

$$c_1 = y_{ij} / (\hat{y}_{ij} * s_j * v_{ij})$$

Az elemzés és a prognóziskészítés során át kell tekintenünk azokat a módszereket, amelyek megadják a jövőre vonatkozó adatokat és információkat. A trendfüggvény típusok esetében például ismernünk kell a

függvények sajátosságait és kiválasztásuk módszereit. Vizsgálunk kell többek között: a trend függvényt leíró formulát, annak sajátosságait⁶⁵.

Additív kapcsolat esetén a trendet az eredeti idősorból kivonjuk, $(y_{ij} - \hat{y}_{ij})$ multiplikatív kapcsolat esetén az eredeti idősort a trenddel osztjuk (y_{ij} / \hat{y}_{ij}) , így kiküszöböljük a trendhatást. Ezt követően, ha havi adatokkal rendelkezünk, 12 tagú mozgóátlagát vesszük a különbségeknek (additív kapcsolat) vagy a hányadosoknak (multiplikatív kapcsolat). Így mivel havi adatokkal rendelkezünk kiküszöböltük a szezonális hullámzást és a véletlen hatást is. A kapott adatsor illetve grafikus ábra a konjunktúra ciklust mutatja. Eljárhatunk úgy is, hogy először mozgóátlagolással (pl. napi adatoknál, ha 252 munkanap van egy évben, akkor a mozgóátlag tagszáma 252, havi adatoknál 12, negyedéves adatoknál pedig 4) a szezonális hullámzást és a véletlen hatást küszöböljük ki, majd ezt követően szűrjük ki a mozgóátlagolású adatsorból a trendhatást, az előzőekben ismertetett módon. Azt, hogy a trend és a szezonhatás kapcsolódása additív, vagy multiplikatív egyszerű módszerrel megállapíthatjuk: ha a trend növekvő (pl. a lineáris trend b_1 együtthatója pozitív vagy az exponenciális trend b_1 együtthatója nagyobb, mint 1) és additív a kapcsolat, akkor a komponensek összegződnek, tehát az amplitúdók nem változnak, nem függnak a trend nagyságától. Multiplikatív kapcsolat esetén viszont a tényezők szorzódnak és a növekvő trend miatt az amplitúdók nőnek.

Csökkenő trendnél (pl. a lineáris trend b_1 együtthatója negatív vagy az exponenciális trend b_1 együtthatója kisebb, mint 1, de nagyobb mint 0) additív kapcsolatot jelez, ha nincs változás az amplitúdókban, mivel összegződnek a tényezők. A multiplikatív kapcsolatot viszont az jelzi, hogyha az amplitúdók az időben előrehaladva csökkennek, tehát csillapodó a rezgés.

A ciklusok periódusának becslése deduktív módszerrel⁶⁶:

A periódust két tényező időtartama határozza meg.

1. A gesztációs idő: a ciklus hosszát meghatározó egyik idő, az az idő, ami alatt a ciklust hordozó objektum alkalmassá válik feladata betöltésére (adaptációs időnek is nevezik).
2. Az élettartam: a ciklus hosszát meghatározó másik idő, az az idő, ami alatt a ciklust hordozó objektum alkalmas a feladata betöltésére.

A ciklus periódusát (T) tehát a hordozó objektum gesztációs (kihordási) ideje, (jele, g) és élettartama, működési ideje (jele, b) alapján becsülhetjük az alábbi összefüggés szerint:

$$T = 2\pi\sqrt{g \times b}$$

Az ember esetében g (terhesség ideje) 0,75 év, b (várható élettartam) 72 év, akkor a periódus:

$$T = 2\pi\sqrt{0,75 \times 72} \approx 46 \text{ év, ez egy Kondratyev – ciklus.}$$

Építkezés esetében, ha az építkezés ideje 4 hónap, tehát $g=4/12=0,33$ év, az épület élettartama, $b=30$ év, akkor a periódus:

$$T = 2\pi\sqrt{0,33 \times 30} \approx 20 \text{ év, ami egy Kuznetz – ciklus.}$$

Gépkocsi esetében $g=0,083$ év, $b=5$ év:

$$T = 2\pi\sqrt{0,083 \times 5} \approx 4 \text{ év, ami egy Kitchin – ciklus.}$$

Kulturális ciklus esetében a tanulás ideje: $g=25$ év, 70 éves korig végzett szellemi munka esetében: $70-25=45$ $b=45$

$$T = 2\pi\sqrt{25 \times 45} \approx 210 \text{ év, ami egy kulturális ciklusnak felel meg.}$$

A tőzsdeindikátorok.xls parancsfájl működése. (A tőzsdeindexek átlagolása, indikátorok. ⁶⁷)

A mozgóátlag az egyik legismertebb és legkönnyebben használható indikátor. Kisimítja az árfolyam hullámzásait és könnyebbé teszi a trendek felismerést, mely nagy segítség a gyorsan mozgó részvények ese-

⁶⁵ Ld. Haustein H. D. [1972] 48-54. 352-360.

⁶⁶ Ld.: Bródy András [1983]: A lassuló idő.

⁶⁷ Forrás: [http://www.investopedia.com/articles/trading/10/simple-exponential-moving-averages-compare .asp](http://www.investopedia.com/articles/trading/10/simple-exponential-moving-averages-compare.asp) (2011. január 17.)

tében. A tőzsdeindexek átlagolására két módszert mutatunk be, amelyekre Excel parancsfájlt dolgoztunk ki⁶⁸:

Az egyszerű (aritmetikai) mozgóátlag (Simple Moving Average, SMA): egy adatsor (A) egyszerű számtani átlagát jelenti. A formula:

$$SMA_{(j)} = \frac{1}{N} \sum_{i=j-N}^j A_i$$

Ahol:

$SMA_{(j)}$ a j időpontban számított mozgóátlag,

N a mozgóátlag periódusainak a száma,

A_i a tényleges érték az i időpontban.

Az egyszerű mozgóátlagot képezhetjük a tőzsdeindexek nyitó, minimum és maximum árakból is, de legtöbb esetben a záróárakat használjuk. Az egyszerű mozgóátlag képletében a legújabb és a legrégebbi adat is azonos súllyal szerepel, pedig a frissebb adatok az előrejelzések során nyilvánvalóan fontosabbak. A probléma egyik megoldása az volt, hogy a frissebb adatokat fokozottabban vették figyelembe. Az exponenciális mozgóátlag (EMA) ugyanis egy speciális súlyozott mozgóátlag, ahol a régebbi adatok súlya exponenciálisan csökken.

Az Excel 2007 által használt képlet a mozgóátlagolás (SMA) esetében, az Excel súgója (F_1) alapján:

„Ez az eljárás a becült időszak értékeit úgy számítja ki, hogy a megelőző időszak adatait megadott számú periódusonként átlagolja. A mozgóátlagolás módszerével olyan részletek derülhetnek ki a trendről, amelyek a meglévő adatok egyszerű átlagolásával elmosódnak. Ez a módszer értékesítési adatok, raktárkészlet adatok vagy egyéb változó adatok előrejelzésére használható. A becült értékeket az alábbi képlet alapján számítja ki:

$$F_{(t+1)} = \frac{1}{N} \sum_{j=1}^N A_{t-j+1}$$

ahol:

N a mozgóátlag periódusainak száma

A_j a tényleges érték a j időpontban

F_j a becült érték a j időpontban”

A két képlet azonos, a különbség, hogy az Excellé fentről lefelé, a másik letről felfelé adja össze ugyanazokat a számokat. Az egyik képlet a szummázási határookban, az Excellé az indexben oldja meg a végigfutást.

A mozgóátlagolás célja, a fő tendencia kimutatása; az átlagolás tompítja a véletlen szerepét; a dinamikus átlag kifejezésre juttatja a fő tendenciát; probléma, hogy megrövidíti az idősort, a trendet hosszabb távra nem tudjuk extrapolálni, továbbá ha a mozgóátlag tagszáma kisebb mint a periódus, akkor, a hullámzást nem küszöböli ki teljesen, csak az erősege csökken, ha pedig a mozgóátlag tagszáma nagyobb mint a periódus, akkor a hullámzás az átlagolás után ellentétes irányú lehet, és az eredeti idősor hullámhegyével szemben hullámvölgy, illetve hullámvölgyével szemben hullámhegy állhat. Az elkövetett hiba annál nagyobb, minél erősebb a véletlen ingadozás és minél kevésbé szabályos a periodikus ingadozás. A mozgóátlagolás a régebbi adatoknak is ugyanolyan súlyt ad, mint az újabbaknak, viszont ismeretes, hogy az extrapoláció szempontjából az új adatok fontosabb szerepet játszanak mint a régiek.

A tőzsdei gyakorlatban a mozgóátlag számítása nem centrikus, hanem visszatekintő, azaz a mozgóátlagokat nem úgy számítják, hogy a kiválasztott adat (záró árfolyam) környezetében végzik az átlagolást, hanem a kiválasztott adat előtti adatokra számítják az értékeket. Ennek megfelelően csak az idősor elején vesznek el mozgóátlagok, ami a technikai elemzésben nem jelent problémát, hiszen a kulcstényezőt az idősor végén található adatok jelentik. A mozgóátlagok így késő (lagging) indikátorok. Az idősor rövidülése: ha N, a mozgóátlag tagszáma páratlan, akkor az idősor (N-1) egységgel, ha páros akkor N egységgel (a középre igazítás, a centrírozás miatt) csökken. A heti tőzsdénapiak száma általában 5, így legtöbbször 5 egész számú többszörösét választják tagszámmak.

⁶⁸ Forrás: <http://betbulls.hu/cms/indikatorok/reszletes/MA> (2011. január 17.)

Exponenciális mozgóátlag (Exponential Moving Average, EMA): a technikai elemzők gyakran használják azért, hogy csökkenteni tudják az egyszerű mozgóátlag lemaradását. Az EMA esetében az időben hozzánk közelebb álló árfolyamok nagyobb, míg a régebbi adatok kisebb súllyal szerepelnek a kalkulációba, a kapott értéket összeadjuk, majd a súlyok összegével elosztjuk. Az egyszerű mozgóátlaggal szemben az exponenciális mozgóátlag érzékenyebb az újabb adatokra, ezért gyorsabban mutatja a trendfordulókat.

Például, egy 20 napos exponenciális mozgóátlag az új árak legfeljebb 9,524 %-at súlyozza [$2/(20+1)*100=9,524\%$], míg egy 200 napos EMA az új árak legfeljebb 0,995%-át [$2/(200+1)*100=0,995\%$] súlyozza. Fontos megjegyezni tehát, hogy az EMA több súlyozást hajt végre az új árakon, mint az SMA. Az exponenciális mozgóátlag elterjedt formája a periódus alapú EMA, ahol a paraméter az EMA időtartamát reprezentálja. Az EMA számítása során alkalmazott képlet, amit a tőzsdeindikátorok.xls parancsfájl kidolgozása során használtuk:

$$EMA_{(jelenlegi)} = [\hat{A}r_{(jelenlegi)} - EMA_{(előző)}] * X + EMA_{(előző)}$$

Ez a képlet átrendezhető az alábbi formában:

$$EMA_{(jelenlegi)} = [\hat{A}r_{(jelenlegi)} * X] + [EMA_{(előző)}] * (1 - X)$$

A periódus alapú EMA esetében a súly értéke $X=2/(1+N)$, ahol az N a periódusok száma. Minden egyes előző záróár adatát felhasználjuk az EMA számításánál. A régebbi adatok befolyása az idő múlásával csökken, de soha sem szűnik meg. Némi bonyodalom csak az első adatpont számításával kapcsolatban van, hiszen ott még nem áll rendelkezésre előző napi érték, emiatt az adatsor első elemének egy egyszerű mozgóátlagot kell számolni. Feltétel, hogy $N \geq 2$ és egész szám, de az N kisebb mint a megfigyelések száma, figyelembe véve azt, hogy az idősor tagjainak száma (N-1) illetve N értékkel csökken.

Vezessük be az alábbi jelöléseket:

$$EMA_{(jelenlegi)} = EMA_t$$

$$EMA_{(előző)} = EMA_{t-1}$$

$\hat{A}r_0$ = az első SMA, ahol a mozgó átlag tagszáma N.

$$\hat{A}r_{(jelenlegi)} = \hat{A}r_t$$

$$X = \alpha$$

$$N > 2 \text{ és } 0 < \alpha < 1$$

Akkor:

$$EMA_t = \alpha \hat{A}r_t + (1 - \alpha) EMA_{t-1}$$

Ebből következik, hogy

$$EMA_0 = \hat{A}r_0$$

$$EMA_1 = \alpha \hat{A}r_1 + (1 - \alpha) \hat{A}r_0$$

$$EMA_2 = \alpha \hat{A}r_2 + (1 - \alpha) EMA_1 = \alpha \hat{A}r_2 + (1 - \alpha) [\alpha \hat{A}r_1 + (1 - \alpha) \hat{A}r_0] = \alpha \hat{A}r_2 + \alpha(1 - \alpha) \hat{A}r_1 + (1 - \alpha)^2 \hat{A}r_0$$

$$EMA_3 = \alpha \hat{A}r_3 + (1 - \alpha) EMA_2 = \alpha \hat{A}r_3 + (1 - \alpha) [\alpha \hat{A}r_2 + \alpha(1 - \alpha) \hat{A}r_1 + (1 - \alpha)^2 \hat{A}r_0] =$$

$$= \alpha \hat{A}r_3 + \alpha(1 - \alpha) \hat{A}r_2 + \alpha(1 - \alpha)^2 \hat{A}r_1 + (1 - \alpha)^3 \hat{A}r_0$$

$$\alpha [\hat{A}r_t + (1 - \alpha) \hat{A}r_{t-1} + (1 - \alpha)^2 \hat{A}r_{t-2} + (1 - \alpha)^3 \hat{A}r_{t-3} + (1 - \alpha)^4 \hat{A}r_{t-4} + \dots + (1 - \alpha)^{t-1} \hat{A}r_{t-(t-1)}] + (1 - \alpha)^t \hat{A}r_0$$

$$EMA_t = [\sum_{i=0}^{t-1} \alpha(1 - \alpha)^i \hat{A}r_{t-i}] + (1 - \alpha)^t \hat{A}r_0$$

Látható, hogy az új érték kialakításában a legutolsó ($\hat{A}r_{t-0} = \hat{A}r_t$) tehát t időpontban megfigyelt adat α súllyal, az azt megelőző, (t-1) időpontban megfigyelt ($\hat{A}r_{t-1}$) adat $(1-\alpha)$, míg az i-edik időszakot megelőző adat ($\hat{A}r_{t-(i-1)}$), $(1-\alpha)^i$ súllyal járul hozzá az EMA_t érték kiszámításához. Az idősor legrégebbi adatának ($\hat{A}r_{t-t} = \hat{A}r_0$) súlya pedig $(1-\alpha)^t$. A két szomszédos EMA adat között a különbség $(1-\alpha)$ – szoros. Az idősor a mozgó átlag tagszámának (N) megfelelően rövidül. Ha N páros, a rövidülés N adat, ha páratlan a rövidülés (N-1). Az exponenciális mozgóátlag is a tőzsdei trendfordulók meghatározásában lehet a befektetők segítségére. Használata megegyezik az egyszerű mozgóátlagével, de nagyobb súllyal veszi figyelembe a friss árfolyamadatokat. Vannak időszakok, amikor az exponenciális mozgóátlag figyelése nagyobb

nyereséget eredményezhet, de más időszakokban az egyszerű mozgóátlag ad jobb jelzéseket. Nem lehet általánosságban kijelenteni, hogy az egyik módszer megbízhatóbb lenne a másiknál.

A vételi és az eladási jelzés szűrésére sok módszert kidolgoztak. Ezek közül az a legjobb, amikor két mozgóátlagot (például a 200 naposat és az 50 naposat) ábrázolunk.

Az indikátor vételi jelzést ad a következő két esetben:

1. a hosszú mozgóátlag emelkedik és a rövid mozgóátlag alulról fölfelé metszi a hosszú átlagot
2. a hosszú mozgóátlag csökkenésből emelkedésbe vált és a rövid mozgóátlag a hosszú fölött helyezkedik el

Az indikátor eladási jelzést ad, amikor:

1. a hosszú mozgóátlag csökken és a rövid mozgóátlag fölülről lefelé metszi a hosszú átlagot
2. a hosszú mozgóátlag emelkedésből csökkenésbe vált és a rövid mozgóátlag a hosszú alatt helyezkedik el

Minél érzékenyebb egy indikátor, annál több jelzést fog adni. Ezek a jelzések időben érkehetnek, de a meg növekedett érzékenységgel a hibás jelzések száma is megnőtt. Minél érzéketlenebb egy indikátor, annál kevesebb jelzést produkál. Azonban ez a kevesebb jelzés sokkal megbízhatóbb is. Habár ezek a jelzések néha késve érkeznek. A mozgóátlagok esetében hasonló a dilemma. Rövidebb mozgóátlagok sokkal érzékenyebbek és több jelzést generálnak. Az EMA, mely tulajdonképpen érzékenyebb, mint az SMA, több jelzést generál. Azonban a hibás jelzések száma is megnő. A hosszabb mozgóátlagok lassabban mozognak, és kevesebb jelzést adnak. Ezek a jelzések sokkal megbízhatóbbak, de késve következnek be. Minden befektetőnek tapasztalatot kell szereznie a különböző mozgóátlagok használatában, hogy megtalálja a középutat az érzékenység és a megbízhatóság között. Számos befektető rövidtávon az exponenciális mozgóátlagot használja, hogy gyorsan tudjon reagálni az árváltozásokra. A befektetők másik fele viszont jobban preferálja az egyszerű mozgóátlagot, hogy hosszabb távú trendeket találjon.

A részvények kiválasztása után a következő feladat a mozgóátlag periódusnak és típusának megválasztása. Minél mozgékonyabb egy részvény, annál több finomítást igényel, vagyis hosszabb mozgóátlagot érdemes használni. Azoknál a részvényeknél, melyek nem mutatnak erős trendet, szintén a nagyobb mozgóátlagok használata célszerű. Nincsen meghatározott paraméter, de a legismertebbek a 20, 50, 89, 150 és a 200 napos, valamint a 10, 30 és 40 hetes mozgóátlagok. A rövidtávon kereskedők 2-3 hétben gondolkodnak, és 21 napos mozgóátlagot használnak, míg a hosszabb távban gondolkodók a 3-4 hónapos trendeket preferálják, mihez 50 hetes mozgóátlagot használnak.

A szakemberek többsége a következő intervallumokat ajánlja a helyes érték megválasztásához:

- rövidtáv: 5-20 nap, legtöbbször 20 nap,
- középtáv: 21-84 nap (4-12 hét), legtöbbször 50 nap,
- hosszútáv: 84-200 nap (13-40 hét), legtöbbször 200 nap.

Adatok forrása:

BUX:

<http://www.mnb.hu/Statisztika/statisztikai-adatok-informaciok/adatok-idosorok>

http://www.bet.hu/magyar_egyeb/dinportl/nonrealtimehistdata

Dow-Jones-index

<http://uk.finance.yahoo.com/q/hp?s=%5EDJI>

Regionális beállításokat előtte USA-ra váltani, utána vissza Magyarra

3.2 Az előrejelzések hibáinak a mérése (A hibaképletek Excel parancsfájl működése)

A prognózisok hibaképleteit az alábbiakban foglaljuk össze.^{69 70 71}

Tételezzük fel, hogy n időszak adata áll rendelkezésre a t kiindulási időpontban és m időszakra készítünk előrejelzést. A megfigyelt értékeket jelöljük X -szel, az előrejelzett értékeket pedig F -fel. (forecast = prognózis) A prognózist akkor tudjuk ellenőrizni, ha a prognosztizált időpont vagy időszak bekövetkezett és így van tényadatunk. Az alkalmazott modell lehet bármely prognóziskészítési módszer. Tételezzük fel, hogy az idősor (t) hossza, amire tényleges és prognosztizált adattal rendelkezünk: $t = 1, 2, \dots, i, \dots, T$. A megfigyelt értékeket jelöljük X -szel, az előrejelzett értékeket pedig F -fel. Ha ex-post ellenőrizzük az előrejelzés hibáját, akkor az idősort két részre osztjuk, az első időszak a becslési-, a második a tesztidőszak. A becslési időszak lehet az idősor fele, vagy annál nagyobb érték. A becslési időszak adatai alapján előrejelzéseket készítünk a teszt időszakra vonatkozóan és így a prognózis hibáját mérni tudjuk, mivel tényleges adatokkal is rendelkezünk. Ex ante előrejelzés esetén csak akkor tudjuk ellenőrizni a prognózis hibáját, ha az előrejelzési időszak bekövetkezik. Pl. havi adatok esetében a következő hónapban, negyedéves adatok esetében a következő negyedévben, stb. Az előrejelzések hibáinak mérésére az alábbi mutatókat használtuk.

Az alkalmazott jelölések:

A hiba ($e = \text{Error}$):

$$e_i = X_i - F_i$$

F_i = előrejelzett vagy illesztett értékek,

X_i = megfigyelt érték.

Az egyszerűség miatt a tesztperiódus időszakot (amikor tényleges és prognosztizált értékekkel is rendelkezünk) jelöljük így:

$$i = 1, 2, \dots, T.$$

1. Átlagos hiba. [ME=Mean Error]

$$ME = \frac{\sum_{i=1}^T e_i}{T}$$

Az a kedvező ha a hiba mértéke 0 körüli érték. Fölébecslés esetén a hiba negatív, alábecslés esetén pozitív. Az előjel váltások miatt az átlagos hiba mutató a hiba valóságos mértékéről nem tájékoztat. Ezt a problémát úgy lehet kiküszöbölni, hogy a hiba abszolút értékével vagy négyzetével számolunk. A másik probléma az, hogy a különböző egységben mért (pl. kg, Ft, \$ stb.) prognózisokat nem lehet összehasonlítani, ezért célszerű relatív hibát is számítani.

2. Átlagos abszolút hiba. [MAE = MEAN ABSOLUTE ERROR⁷²]

$$MAE = \frac{\sum_{i=1}^T |e_i|}{T}$$

3. Átlagos négyzetes hiba. [MSE = MEAN SQUARE ERROR⁷³]:

$$MSE = \frac{\sum_{i=1}^T e_i^2}{T}$$

4. A hiba szórása [SDE = STANDARD DEVIATION OF ERRORS⁷⁴]:

⁶⁹ A trendszezon-hibasámítás.xls parancsfájlnál a felsorolt hibaképleteket programoztuk be.

⁷⁰ Farnum Nicholas R., - Stanton LaVerne W. [1989]: 22-31.

⁷¹ S. Makridakis-S. C. Wheelwright- V. Mcgee [1983]: 43-54.

⁷² Használják a nemzetközi szakirodalomban a Mean Absolute Deviation (MAD) kifejezést is.

⁷³ Használatos még ASE – Average Squared Error

⁷⁴ Használják a nemzetközi szakirodalomban a Root Mean Square Error (RMSE) mutatót is, amikor az MSE mutató négyzetösszegét veszik.

$$SDE = \sqrt{\frac{\sum_{i=1}^T e_i^2}{T-1}}$$

5. Átlagos relatív [%-os] hiba [MPE = MEAN PERCENTAGE ERROR]

Relatív [%-os] hiba [$PE_i = \text{PERCENTAGE ERROR}$]:

$$PE_i = \frac{X_i - F_i}{X_i} \times 100$$

$$MPE = \frac{\sum_{i=1}^T PE_i}{T}$$

Az előjel váltások miatt az átlagos relatív hiba mutató a hiba valóságos mértékéről nem tájékoztat. Ezt a problémát úgy lehet kiküszöbölni, hogy a hiba abszolút értékével vagy négyzetével számolunk.

6. Átlagos relatív [%-os] abszolút hiba [MAPE = MEAN ABSOLUTE PERCENTAGE ERROR]:

$$MAPE = \frac{\sum_{i=1}^T |PE_i|}{T}$$

MAPE 20%-nál kisebb értéke az elfogadható.

7. Theil féle U-statisztika. [Theil's U-Statistic⁷⁵]:

$$U = \sqrt{\frac{\sum_{i=1}^{T-1} \left(\frac{F_{i+1} - X_{i+1}}{X_i} \right)^2}{\sum_{i=1}^{T-1} \left(\frac{X_{i+1} - X_i}{X_i} \right)^2}}$$

Ha $U = 0$, akkor $F_i = X_i$, vagyis az előrejelzés megegyezik a valósággal. Különben U értéke 0-tól különbözik. A nevezőben az a relatív négyzetes hiba van feltüntetve, amikor az előrejelzett érték az utolsó tényadattal egyenlő, ezt naiv előrejelzésnek hívjuk. A számláló az idősorkutatási módszerrel elkészített előrejelzés relatív négyzetes hibáját tartalmazza. Ha ez utóbbi jobb előrejelzést adott mint a naiv előrejelzés, akkor a számláló értéke (a hiba) kisebb mint a nevező értéke, tehát az U -statisztika értéke egynél kisebb. Értelemszerűen, ha a naiv és az idősorkutatási modellel készített előrejelzés azonos értéket ad, akkor a számláló és nevező megegyezik, így az U -statisztika értéke eggyel egyenlő. Ha a naiv előrejelzés ad jobb eredményt, akkor az U -statisztika értéke egynél nagyobb.

8. MBA. [McLaughlin Batting Averages]

$$MBA = [4 - U] \times 100$$

- $U=0$ esetén, $MBA = 400$
- $U=1$ esetén, $MBA = 300$
- $U>1$ esetén, $MBA < 300$

A Theil- féle U statisztika egy másik képlete az MBA mutató.

3.3 Trendszezon-hibaszámítás parancsfájl működése

A program lehetővé teszi a korábban ismertetett 9 féle lineáris illetve lineárisra visszavezethető trend vizsgálatát. Megadható a mozgó átlag tagszáma, ahol a korlát csak az idősor hossza. Vizsgálhatjuk továbbá az adatbázis függvényében a szezonális hatást és a konjunktúra ciklusokat.

A trendek megbízhatóságának ellenőrzése.

⁷⁵ Theil, H. [1961].

A becslési időszak megadásával a rendelkezésre álló idősort két részre bontjuk, becslési és teszt időszakra. A becslési időszak felhasználásával a trendbecslést végezzük el, a teszt időszak adatai alapján a becslések pontosságát lehet ellenőrizni, a megadott hibaképletek felhasználásával. Az idősor hosszát felismeri a program, a sárga mezőbe (Ebből becslésre felhasznált:) be kell írni a becslésre felhasznált idősor hosszát, ami az idősor fele vagy annál nagyobb érték. Ha tesztelni kívánjuk a trendeket, akkor a becslésre felhasznált idősor hosszának legnagyobb értéke az idősor hossza mínusz egy. Az extrapoláció hosszát is meg lehet adni. Az ábra felett lehet választani a trendtípusok között. A trendek paramétereit és a többszörös determinációs együtthatót (R^2), valamint az eredeti adatok és a trend ábráját közli a program. A 9 trend extrapolációja és a hibaképletek számításai számszerűen is megtalálhatók az eredeti adatokat (y) követő oszlopokban. Célszerű azonos becslési időszak felhasználásával mind a 9 trendtípusra elvégezni a számításokat és kiválasztani azt a trendtípust, amelyik esetében a legkisebb a hiba, pl. a MAPE. Ugyancsak célszerű a becslési időszak hosszát változtatni. Ki lehet választani a legjobban illeszkedő becslést, vagyis a múlt időszak alapján legjobb előrejelzést adó trendtípus és információt kapunk a trend - extrapolációk megbízhatóságáról is. A program tehát érzékenységvizsgálatokat is végez. Ha a tesztperiódus kezdetének változtatásával a hibaképletek alapján kiválasztott "legjobb trend" típusa is változik, akkor az idősor nem stabil, az idősorban vizsgált trend tendenciája változik. Az elemzés így bizonytalan, és az előrejelzés sem lesz megbízható. Hasonló módon a magas hibaértékek (pl. a MAPE 10 % - nál nagyobb) is a bizonytalan becslést jelzik. A becslési időszak növelésével általában a hibák is csökkennek.

A szezonálitás vizsgálata.

A Periódus megadásánál meg kell adni az adatbázis függvényében a sárga mezőben a periódus számát, aminek értéke 2-24 lehet, ha nem adunk adatot, akkor nem számol a szezonálitással. Ha fél éves adatokkal dolgozunk, akkor a periódus értéke 2, negyedéves adatoknál 4, havi adatoknál 12. Megadandó a Kezdő periódus sorszáma is, lehetséges értékei: 1-12, pl. ha havi adatokkal dolgozunk és az első évben, csak a 7-ik hónaptól vannak adataink, akkor a beírandó érték 7. Ha viszont teljes az idősor akkor a Kezdő periódus sorszáma 1. (0 értéket nem szabad megadni) A program kiszámítja a szezonális eltéréseket és a szezonindexeket. Elvégzi a korrekciót is és ábrázolja a kiválasztott trendet és szezon modellt (additív vagy multiplikatív). Ha nincs szezonálitás az idősorban akkor választani lehet a szezonálitás nélkül opciót (pl. éves adatsorok esetén). A program ábrázolja a véletlen összetevőt is a kiválasztott trend-szezon modell alapján.

Konjunktúra-ciklusok elemzése

Az Adatok-bevitele munkalapon kiválaszthatjuk a 9 trend közül azt, amit vizsgálni kívánunk (természetesen egyenként mind a 9 trendfüggvény konjunktúra ciklusait vizsgálhatjuk és összehasonlíthatjuk) és meg kell adni a mozgóátlag tagszámát. A mozgóátlag tagszáma az adatbázistól függ. Rövid ciklusok vizsgálata esetében a szezonálitást kűszöböljük ki, ha havi adatok állnak rendelkezésre a mozgóátlag tagszáma 12, negyedéves adatok esetében 4. (Megjegyezzük, hogy a szezonális hullámozás kiszűrése után, ha ezt követően kiszűrjük a trendhatást, akkor a Kitchin-féle rövid ciklus becslését kapjuk meg.) Ha a hosszú ciklusokat vizsgáljuk éves adatokkal, akkor általában 8 vagy 9 tagú mozgóátlagot választunk a rövidebb periódusú (pl. 4-8 vagy 3-9 éves) ciklusok kiszűrésére. (Megjegyezzük, hogy a rövidebb ciklusok kiszűrése után, ha ezt követően kiszűrjük a trendhatást, akkor a Kondratyev-féle hosszú ciklus becslését kapjuk meg.) Vizsgálni lehet az idősorokat abból a szempontból is, hogy hogyan reagálnak a mozgóátlag tagszámának megváltoztatására. Választhatunk az additív és a multiplikatív modell között. Az adatok bevitele után a program ábrázolja az eredeti adatokat és a trendet és a második ábrában a ciklust. Kiszámítja multiplikatív kapcsolatot feltételezve az eredeti idősor és a trend hányadosát (y_i / \hat{y}_i) valamint additív kapcsolatot feltételezve az eredeti idősor és a trend különbségét ($y_i - \hat{y}_i$). Így mindkét feltételezés mellett kikűszöböli a trendhatást. A ciklikus mozgásokat a trendtől megtisztított idősoroknál a felhasználó által megadott mozgóátlag tagszám alapján mozgóátlagolással kűszöböli ki és számítja, valamint ábrázolja.

A rövid ciklusok átlagos periódushosszának becslésére **a ciklusfordulópontokszámítása Excel parancsfájl** használható:

A számításokat először pl. a trend – szezon - hibaszámítás Excel parancsfájllal végezhetjük el. Kiválasztjuk az 5 legjobb modellt a trendfüggvények alapján. Először az eredeti sort és az időváltozót másoljuk

be, utána az eredeti adatoknak a trendértékektől való eltérését (additív modell) vagy hányadosát (multiplikatív modell). Zárójelben a kiválasztott trendek - illesztési pontosságát jelző - többszörös determinációs együtthatóit (R^2) is feltüntettük, amit a trend – szezon - hibaszámítás Excel parancsfájl kiszámított.

$$R^2 = 1 - \frac{\sum_{t=1}^{t=N} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{t=N} (y_t - \bar{y}_t)^2}$$

Mivel azonos adatoknál a megfigyelések száma (t) megegyezik, ezért az átlagtól való eltérésnégyzetösszeg azonos, így a R^2 kizárólag az eredeti és becsült adatok eltérés négyzetösszegétől függ.

Egy futtatásnál összesen 5 változóval dolgozhatunk.

A mintafeladat:

Évek: 1800-2007. A megfigyelések száma $N=208$

x_1 = a réz (copper) árak 2000-es \$-ban, \$/libra (1 libra=0,454 gramm).

x_2 = eredeti adatok-lineáris trend becsült értékei ($R^2=0,647$)

x_3 = eredeti adatok-másodfokú parabolikus trend becsült értékei ($R^2=0,649$)

x_4 = eredeti adatok-harmadfokú parabolikus trend becsült értékei ($R^2=0,676$)

x_5 = eredeti adatok- exponenciális trend becsült értékei ($R^2=0,635$)

A program megkeresi az alsó és felső fordulópontokat, a felső fordulópontot (amikor a növekedés után csökkenés következik) + jellel, az alsó fordulópontot (amikor csökkenés után növekedés következik) – jellel jelöli. A hullámhossz egy teljes hullámnak a hossza, ami becsülhető a csúcsponttól a csúcspontig (+tól a következő +ig), vagy a mélyponttól a mélypontig (-tól a következő -ig). A program megkeresi a legelső + illetve – értéket és az eredmények munkalapon ezeket a fordulópontokat bejelöli. A következő táblázat sorszámokkal látja el a fordulópontokat, pl. az első + érték kapja az egy értéket, a ciklushosszát innentől számítjuk, a következő megfigyelések addig kapnak 1 értéket, amíg a program meg nem találja a következő + jellel jelölt fordulópontot, innen a jelölés 2. Ennek alapján megállapítható hány fordulópont van az idősorban. A következő táblázatok közlik a fordulópontok átlagos távolságát időegységben (példánkban években) majd a kétféle fordulópont átlagát, egész számra kerekítve. A rövidebb ciklusok kiküszöbölésére ezt az értéket vagy egész számú többszörösét használjuk mozgó átlag tagszámként.

A trend – szezon - hibaszámítás Excel parancsfájl felhasználásával, most már megadhatjuk a mozgó átlag tagszámát, ami példánkban 4 vagy 8 vagy 12 stb. lehet. A kézikönyv tartalmazza a különböző ciklusok kiküszöbölésének a módszerét. Példánkban az analitikus trendet kivontuk az idősorból, mert additív trendet feltételeztünk, a rövidebb ciklusokat és a véletlent pedig 4 tagú mozgóátlagolással küszöböljük ki.

A megfelelő modell kiválasztása. Trend-szezonteszt Excel parancsfájl működése (trendszezonteszt.xls)

A program működése: Módosíthatók a trendek paraméterei és a szezon erősege (ahol a nagyobb szám nagyobb amplitúdójú szezonkomponenst generál). A következő trendek modellezhetők: lineáris, féllogaritmikus, másodfokú parabolikus, harmadfokú parabolikus, exponenciális, elsőfokú hiperbolikus és S-alakú logisztikus. A program elkészíti a trendek és a szezonkomponensek ábráját az additív és a multiplikatív modell esetben, továbbá csak a trendet, ha feltételezzük, hogy nincs szezonális hatás. A program célja az, hogy felismerjük azt, hogy egy idősor grafikus ábrája alapján milyen trendtípusok és szezonális (periodikus) kapcsolatok (additív vagy multiplikatív) jöhetnek számításba. A trend paramétereinek és a szezon erősege változtatásával nyomon követhetők a grafikus ábrák változásai.

Gyakorló feladatok. Konjunktúra ciklusok modellezése, a trendszezon - hibaszámítás Excel parancsfájl működése)

Mutassa ki a konjunktúra ciklusokat a magyar mezőgazdaság sorai alapján:

A MAPE alapján a 9 analitikus trend közül melyik adta a legjobb közelítést. Alkalmazza a 9 tagú mozgó átlagot a rövidebb ciklusok kiküszöbölésére.

- Cukorrépa (kg/fő) 1920-2006
- Árpa (kg/fő) 1876-2006
- Burgonya (kg/fő) 1870-2006
- Búza (kg/fő) 1876-2006
- Kukorica (kg/fő) 1870-2006
- Rozs (kg/fő) 1920-2006
- Zab (kg/fő) 1921-2006
- Sertés állomány (db/ezer fő) 1870-2006
- Szarvasmarha állomány (db/ezer fő) 1870-2006
- Ló állomány (db/ezer fő) 1904-2006
- Mezőgazdaságban dolgozó aktív keresők aránya (%)

3.4 A telítődési, a logisztikus (S-alakú)- és életgörbe trendfüggvények becslése Excel parancsfájllal

A telítődési, a logisztikus és az életgörbe trendfüggvények olyan folyamatok, jelenségek leírására alkalmasak, amelyeknek a növekedése korlátos. A tartós fogyasztási cikkek (például tévé, rádió, telefon, autó stb.) forgalmának alakulása – a piac korlátozottsága miatt – telítődési trendet követ, mert van egy szint, ami fölé a kereslet nem emelkedik. Az ipari termékek életgörbéje is hasonló tendenciát mutat az első felszálló szakaszban, az eltérés azonban az, hogy a csúcspont elérése után leszálló szakasz következik, előbb csökken és végül megszűnik a gyártás és a forgalom. Gyakran előfordul, hogy az első felszálló szakaszban a fejlődés három szakaszát különíthetjük el: 1. a kísérletezés stádiuma, amit a gyártás beindítása, a lassú növekedés jellemez; 2. a „nagy felfutás” időszaka, nő a kereslet, a gyártás ezért tömegszerűvé válik; 3. a piac telítődése, amikor már csak az elhasználódás pótlására van lehetőség.⁷⁶ A piaci érettség, a telítődés szakaszát a kereslet és a gyártás hanyatlása (többnyire új, korszerűbb termék jelenik meg a piacon), és végül a gyártás megszüntetése követi. A termékéletgörbe konkrét alakjának meghatározása a tervezés időszakában nem egyszerű feladat. Az élettartam a terméktől függően lehet néhány hónap (a divat által erősen befolyásolt termékek), néhány év (informatikai eszközök) vagy több évtized (mezőgazdasági termények). A telítődési függvényeket (elsősorban a Gompertz- és Johnson-függvényeket) a demográfusok és a biztosítási szakemberek a népesedési és túlélési folyamatok leírására és közelítésére használják. Az inflexiós pont jelzi a vizsgált jelenség fejlődésében bekövetkező jelentősebb változást és annak várható időpontját is. Az inflexiós pont kifejezi, hogy a fejlődés „hajtóerői” kifulladásra várható, hogy a fejlődés jellege is megváltozik és lelassul. Tulajdonképpen a fejlődés egyik kritikus pontja éppen az inflexiós pontnál van. A telítődési függvények monoton növekvő függvények, ahol az időváltozó (t) növekedésével a növekedési értékek a nullához tartanak. Más megfogalmazásban a függvényértékek K telítődési paraméterhez, szaturációs szinthez, vagyis egy konstans értékhez tartanak, ha az időváltozó a végtelenbe tart:

$$\lim_{t \rightarrow \infty} f(t) = K$$

A feladat megkeresni azokat a paramétereket, amelyek mellett az illesztés a legpontosabb. A tapasztalatok szerint az egyszerűbb függvényformákkal (lineáris, hatványkitevős, exponenciális, parabolikus stb.) szemben a bonyolultabb telítődési függvények alkalmazása lényegesen számításigényesebb, ugyanakkor kiküszöbölik az egyszerűbb függvények azon hibáját, hogy a növekedésnek (vagy csökkenésnek) nincs felső vagy alsó korlátja.⁷⁷ E probléma megoldására Excel parancsfájlt dolgoztunk ki. A számításigényesség mint probléma megszüntethető a bemutatásra kerülő parancsfájl alkalmazásával. A trendek ismertetésénél figyelembe vettük Descartes⁷⁸ módszertani szabályait, ami szerint az egyszerűtől kell haladni a bonyolult felé, vagyis figyelembe kell venni azt, hogy az összetett módszerek általában speciális esetként tartalmazzák az egyszerűbbeket. A másik fontos szabály a felsorolás elve, vagyis a teljességre kell törekedni.⁷⁹ Descartes azt is hangsúlyozza, hogy minden módszernek elméleti következményei vannak. A Földön hosszú távon gyakorlatilag minden gazdasági-társadalmi-demográfiai folyamat korlátozott térben zajló növekedési folyamat és ennek következménye az, hogy igen gyakran tapasztaljuk, hogy léteznek te-

⁷⁶ Theiss [1958] 199–200.

⁷⁷ Ld.: Freschl [1982], Herman–Varga [1983], Herman [1985], Valkovics [2001], Hunyadi [2004].

⁷⁸ Descartes [1961] 214–215.

⁷⁹ A teljességre természetesen csak törekedni lehet, számos logisztikus függvényt nem tudunk bemutatni, például: Korf-függvény (Liao–Podrázský–Liu [2003] 545.) Weibull-és Béta-függvények (Xinyou Yin et al. [2003] 362. és 369.), Causton- és Venus-függvény (Colin [1999] 715.).

lítódési pontok.⁸⁰ A Descartes-i módszertani szabályokat figyelembe véve először a legegyszerűbb telítődési függvényeket ismertetjük, amelyek inflexiós ponttal nem rendelkeznek. A bemutatásra kerülő következő hét S-alakú trendfüggvénynek egy inflexiós pontja van. Az életgörbe- és a Hubbert-féle trendfüggvény két inflexiós ponttal rendelkezik. Az életgörbe trenddel többek között a termékéletgörbék alakulását lehet modellezni, ahol a növekedési szakaszt egy csökkenő szakasz követi. A logisztikus függvények régóta foglalkoztatják az idősor modellezés kutatóit. A XIX. század első felében Gompertz és Verhulst munkásságát lehet kiemelni. A logisztikus függvény a XX. század első felében az ökonometriai modellezés egyik fontos eszköze volt és számos modell került kidolgozásra. Népszerűsége bizonyos területeken az elmúlt évszázadban sem csökkent, különösen a piaci és demográfiai folyamatok gyakori jellemzője, hogy egy ideig gyors ütemben nőnek, majd később érvényesülnek a növekedés korlátai, csökken a növekedési ütem. A folyamat jellegétől függően a növekedés bizonyos idő után a nullához tart, illetve az is elképzelhető, hogy a tendencia megfordul.

3.4.1 Inflexiós ponttal nem rendelkező telítődési görbék

A következőkben a Mitscherlich-,⁸¹ a Bertalanffy-,⁸² az egyszerűen modifikált exponenciális⁸³, a Törnquist 1., valamint Törnquist 2. függvényeket mutatjuk be röviden. (Lásd a 3.1. táblát) E függvények közös jellemzője, hogy inflexiós ponttal nem rendelkeznek, értelmezési tartományukon ($0-\infty$ intervallumon) konkáv módon viselkednek. Ez utóbbi tulajdonság akkor teljesül, ha a függvény kétszer differenciálható az időváltozó szerint és a második derivált negatív. Az ismertetésre kerülő telítődési függvények második deriváltja mindenütt negatív. Ha K azonos, akkor $Kb=a$, tehát a Bertalanffy függvény és az egyszerűen modifikált exponenciális függvény azonos.

Az inflexiós ponttal nem rendelkező telítődési függvények sematikus ábrája a következőképpen néz ki, ld. 3-11 ábrát. Tipikus alkalmazási területük a jelentős technológiai újítással gyártott termékek gyors elterjedése az adott szegmensben, amikor termékváltás következik be: például ilyen volt a rugós órát felváltó kvarcóra, a fekete-fehér tévét váltó színes tévé vagy a közelmúltban a plazma- és LCD-tévék megjelenése a piacon.

3-1. tábla: Inflexiós ponttal nem rendelkező telítődési görbék főbb jellemzői

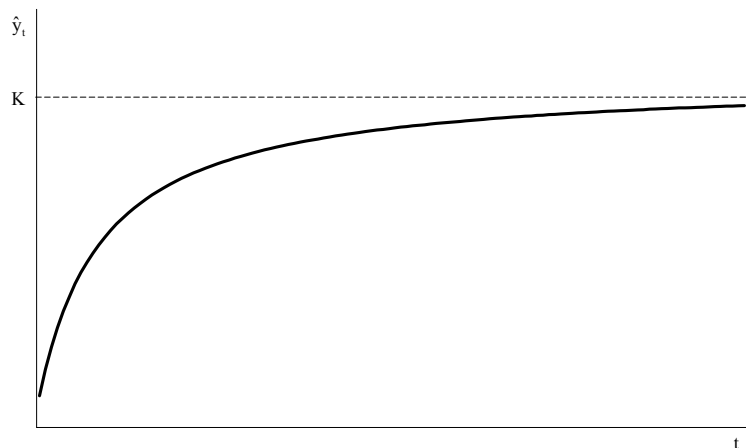
Függvény	Formula	\hat{y}_0	$\frac{d^2\hat{y}_t}{dt^2}$	$\lim_{t \rightarrow \infty} \hat{y}_t$
Mitscherlich	$\hat{y}_t = K(1 - e^{-rt})$	0	$-Kr^2e^{-rt}$	K
Bertalanffy	$\hat{y}_t = K(1 - be^{-rt})$	$K(1-b)$	$-Kbr^2e^{-rt}$	K
Egyszerűen modifikált exponenciális függvény	$\hat{y}_t = K - ae^{-rt}$	$K - a$	$-ar^2e^{-rt}$	K
Törnquist 1.	$\hat{y}_t = \frac{Kt}{t+a}$	0	$\frac{-2aK}{(a+t)^3}$	K
Törnquist 2. ($a < b$)	$\hat{y}_t = \frac{K(t+a)}{t+b}$	$\frac{Ka}{b}$	$\frac{2K(a-b)}{(b+t)^3}$	K

⁸⁰ Ld.: Fokasz [2006] 19–51.

⁸¹ Eilhard Alfred Mitscherlich (1874–1956) német agronómus. Ld.: Mitscherlich [1919] 167–182.

⁸² Ludwig von Bertalanffy (1901–1972) osztrák születésű biológus, a rendszerelmélet megalkotója. Ld.: Bertalanffy [1938] 181–213.

⁸³ Ld.: Kotz et al. [2006] 14. kötet. 8727–8728.



3-11. ábra: Az inflexiós ponttal nem rendelkező telítődési függvény

Mitscherlich trendfüggvény

$$\hat{y}_t = K(1 - e^{-rt})$$

Bertalanffy trendfüggvény

$$\hat{y}_t = K(1 - be^{-rt})$$

Egyszerűen modifikált exponenciális függvény

$$\hat{y}_t = K - ae^{-rt}$$

Törnquist 1. trendfüggvény

$$\hat{y}_t = \frac{Kt}{t + a}$$

Törnquist 2. trendfüggvény

$$\hat{y}_t = \frac{K(t + a)}{(t + b)}$$

3.4.2 Egy inflexiós ponttal rendelkező trendfüggvények

A logisztikus trendfüggvények kezdetben konvex, később konkáv függvénygörbét írják le. Ezeket a logisztikus trendfüggvényeket, alakjuk miatt S-alakú függvényeknek is hívják a szakirodalomban. A következőkben a logisztikus, a késleltetett logisztikus, a négyzetesen logisztikus, a Gompertz-féle, a 63 százalékos, a Johnson-féle és az általánosított Richards-féle trendfüggvényeket mutatjuk be.

A logisztikus trendfüggvény

Az egyik első logisztikus trendfüggvényt (népesség növekedési modell) *Verhulst*⁸⁴ 1838-ban publikálta:

$$\hat{y}_t = \frac{K}{1 + e^{-c(t-m)}} = \frac{Ke^{ct}}{e^{cm} + e^{ct}}$$

ahol

- K – a telítettségi szint, $K > 0$;
- m – az inflexiós pont helyét adja meg, $m > 0$;
- c – a növekedési sebességet jellemző paraméter, ha $c > 0$, akkor logisztikus növekedésről, ha $c < 0$, akkor logisztikus csökkenésről van szó.

A függvény inflexiós pontjának koordinátái:

⁸⁴ Pierre-Francois Verhulst belga matematikus, statisztikus (1804–1849). Ld.: Verhulst, P. F. [1838]: 113–121.

$$\frac{d^2 \hat{y}_t}{dt^2} = \frac{Kc^2 e^{c(m+t)} (e^{cm} - e^{ct})}{(e^{cm} + e^{ct})^3} = 0,$$

$$e^{ct} = e^{cm},$$

$$t_w = m,$$

$$\hat{y}_{t_w} = \frac{K}{2}.$$

A harmadik derivált ebben a pontban nem egyenlő 0-val, mivel:

$$\left. \frac{d^3 \hat{y}_t}{dt^3} \right|_{t_w} = 0,$$

$$e^{2cm} - 4e^{2cm} + e^{2cm} = 0,$$

$$-2e^{2cm} \neq 0.$$

A függvény helyettesítési értéke a $t=0$ helyen és a telítődési szint:

$$\hat{y}_0 = \frac{K}{(1 + e^{cm})},$$

$$\lim_{t \rightarrow \infty} \hat{y}_t = K.$$

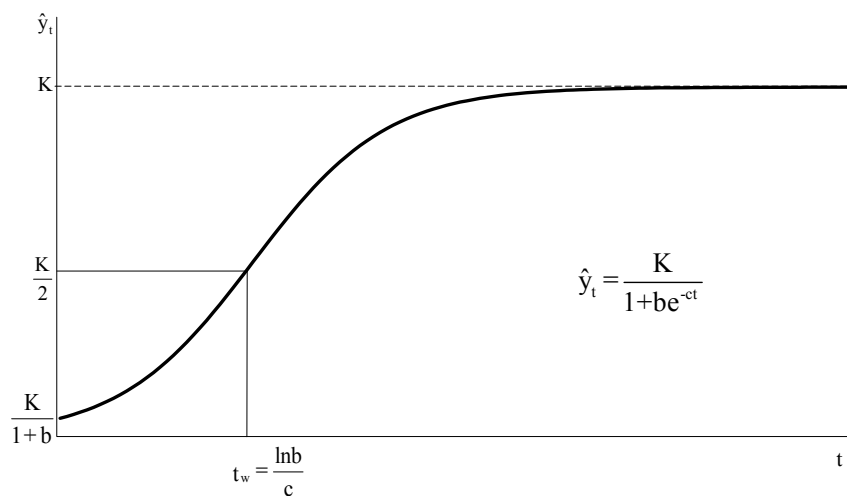
Napjainkban logisztikus trendfüggvény névvel az egyik legelterjedtebben alkalmazott, Pearl–Reed-féle⁸⁵
⁸⁶ logisztikus telítődési függvényt illeti az irodalom A függvény szimmetrikus az inflexiós pontra, ahol a görbe a telítettségi szint felét ($K/2$) éri el. A telítődés annak a következménye, hogy a görbe a konvex szakaszból konkávba megy át, azaz a vizsgált jelenségben fordulópont (minőségi változás) következett be.

A függvényt leíró formula:

$$\hat{y}_t = \frac{K}{1 + be^{-ct}} = \frac{Ke^{ct}}{e^{ct} + b},$$

ahol:

- K – a telítettségi szint, ha $K > 0$;
- b – a helyzetparaméter, ha $b > 0$;
- c – növekedési sebességet jellemző paraméter, ha $c > 0$ logisztikus növekedésről, ha $c < 0$, akkor logisztikus csökkenésről van szó.



3-12. ábra: A Pearl-Reed-féle logisztikus trendfüggvény

A függvény inflexiós pontjának koordinátái:

⁸⁵ Raymond Pearl (1879–1940) amerikai biológus, Lowell J. Reed (1886–1966) matematikus, biostatistikus.

⁸⁶ Ld.: Pearl-Reed [1920] 275–288. és Farnum–Stanton [1989] 189–191.

$$\frac{d^2 \hat{y}_t}{dt^2} = \frac{bc^2 Ke^{ct} (b - e^{ct})}{(e^{ct} + b)^3} = 0,$$

$$t_w = \frac{\ln b}{c},$$

$$\hat{y}_{t_w} = \frac{K}{2}.$$

A harmadik derivált ebben a pontban nem egyenlő 0-val, hiszen $b > 0$, tehát ebben a pontban inflexiós pont van:

$$\left. \frac{d^3 \hat{y}_t}{dt^3} \right|_{t_w} = 0,$$

$$e^{2 \ln b} - 4be^{\ln b} + b^2 = 0,$$

$$b = 0.$$

A függvény helyettesítési értéke a $t = 0$ helyen és a telítődési szint:

$$\hat{y}_0 = \frac{K}{1+b},$$

$$\lim_{t \rightarrow \infty} \hat{y}_t = K.$$

Könnyen belátható az, hogy a Verhulst és a Pearl-Reed-féle függvények lényegében azonosak:

$$\frac{Ke^{ct}}{e^{cm} + e^{ct}} = \frac{Ke^{c't}}{e^{c't} + b'}$$

Késleltetett logisztikus trendfüggvény

Az előzőekben bemutatott logisztikus trendfüggvénynek az inflexiós pontra való szimmetriája sok esetben modellezési szempontból nem helytálló. Ha a kezdeti növekedés gyorsabb ütemű, és a görbe az inflexiós pontot a telítettségi szint felénél korábban éri el, akkor használhatjuk a késleltetett logisztikus trendfüggvényt. Az inflexiós pont után a késleltetett logisztikus trendfüggvény konkáv szakasza hosszabb és elnyújtottabb, mint a logisztikus trend hasonló konkáv szakasza, tehát késleltetett hatás érvényesül a telítődési szint elérésében. A 3-13. ábra a késleltetett logisztikus trendfüggvény alakját és nevezetes pontjait mutatja be.

A függvényt leíró formula:

$$\hat{y}_t = \frac{K}{1 + \left(\frac{T}{t}\right)^a},$$

Ahol

- T -a helyzetparaméter, ha $T > 0$;
- a – növekedési sebességet jellemző paraméter, ha $a > 1$.

Amennyiben $t = T$, úgy:

$$\hat{y}_T = \frac{K}{2}.$$

A függvény inflexiós pontjának koordinátái

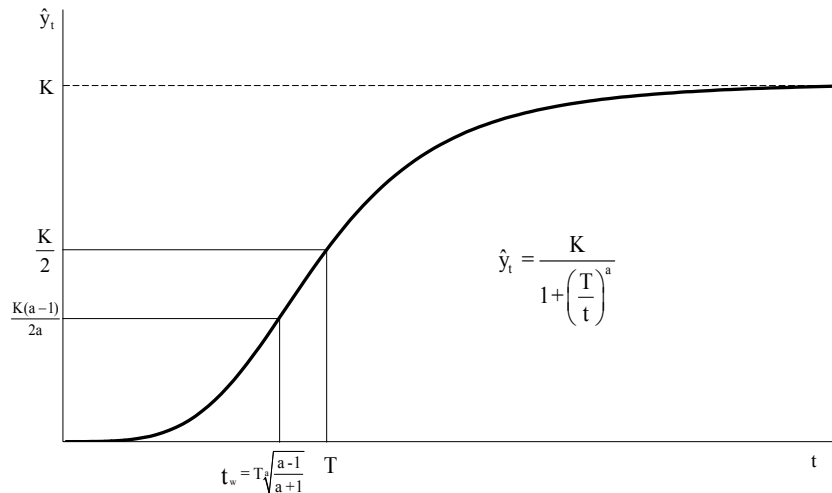
$$\frac{d^2 \hat{y}_t}{dt^2} = \frac{aK \left[(a-1) \left(\frac{T}{t}\right)^a - a - 1 \right]}{t^2 \left[\left(\frac{T}{t}\right)^a + 1 \right]^3} \left(\frac{T}{t}\right)^a = 0,$$

$$t_w = T \sqrt{\frac{a-1}{a+1}},$$

$$\hat{y}_{t_w} = \frac{K(a-1)}{2a}.$$

A telítődési szint:

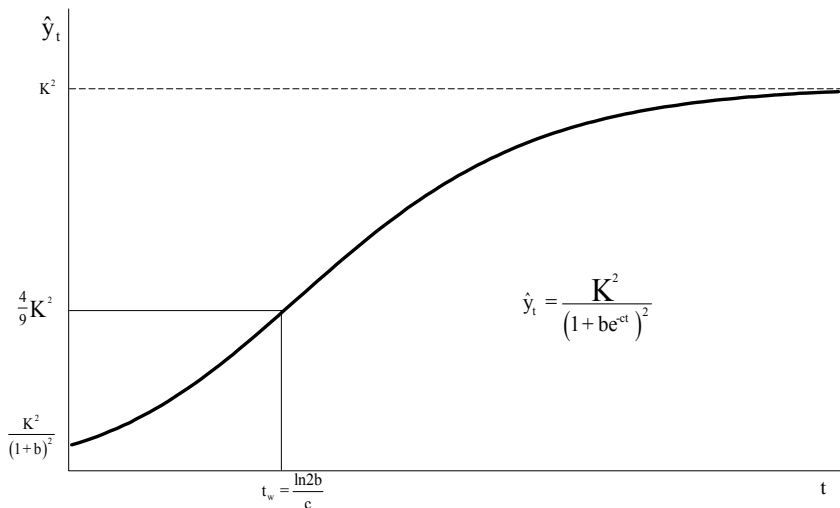
$$\lim_{t \rightarrow \infty} \hat{y}_t = K.$$



3-13. ábra: A késleltetett logisztikus trendfüggvény

Négyzetesen logisztikus trendfüggvény

A 3-14. ábra mutatja be a négyzetesen logisztikus függvény alakját és nevezetes pontjait.



3-14. ábra: A négyzetesen logisztikus trendfüggvény

A négyzetesen logisztikus függvény, mint a neve mutatja, a logisztikus trendfüggvény négyzete. A függvényt leíró formula:

$$\hat{y}_t = \left(\frac{K}{1 + be^{-ct}} \right)^2 = \frac{K^2 e^{2ct}}{(e^{ct} + b)^2},$$

ahol

- b – a helyzetparaméter, ha $b > 0$;
- c – a növekedési sebességet jellemző paraméter, ha $c > 0$ logisztikus növekedésről, ha $c < 0$, akkor logisztikus csökkenésről van szó.

A függvény inflexiós pontjának koordinátái:

$$\frac{d^2 \hat{y}_t}{dt^2} = \frac{2bc^2 K^2 e^{2ct} (2b - e^{ct})}{(e^{ct} + b)^4} = 0,$$

$$t_w = \frac{\ln 2b}{c},$$

$$\hat{y}_{t_w} = \frac{4}{9} K^2.$$

A harmadik derivált ebben a pontban nem egyenlő 0-val, hiszen $b > 0$, tehát ebben a pontban inflexiós pont van:

$$\left. \frac{d^3 \hat{y}_t}{dt^3} \right|_{t_w} = 0,$$

$$e^{2 \ln 2b} - 7be^{\ln 2b} + 4b^2 = 0,$$

$$4b^2 - 14b^2 + 4b^2 = -6b^2 = 0,$$

$$b = 0.$$

Az inflexiós ponttal jellemzett irányváltás tehát a négyzetesen logisztikus függvény esetében később következik be, mint a logisztikus trendfüggvényénél ugyanis:

$$\frac{\ln b}{c} < \frac{\ln 2b}{c}.$$

A függvény helyettesítési értéke a $t = 0$ helyen és a telítődési szint:

$$\hat{y}_0 = \frac{K^2}{(1+b)^2},$$

$$\lim_{t \rightarrow \infty} \hat{y}_t = K^2.$$

Gompertz-függvény

*Benjamin Gompertz*⁸⁷ már a XIX. század elején felfedezte azt a halandósági törvényt, amit az állatokon végzett vizsgálatok is megerősítenek. Az emberi halandósági ráta a nemi érettség elérése idején a legkisebb, utána exponenciálisan emelkedik. Az idő ebben az esetben az életkor. *Valkovics Emil*⁸⁸ például az eredeti Gompertz-formula megfelelő átalakításával újradefiniálta a halandósági tábla függvényeit, és példákkal szemléltette a Gompertz-függvény megnövekedett felhasználási lehetőségeit a demográfia egyes területein.

A Gompertz-függvény eredeti alakja⁸⁹:

$$\hat{y} = Kb^{c^t},$$

$$\ln \hat{y} = \ln K + c^t \ln b.$$

A Gompertz-függvény az előbbi függvény továbbfejlesztése alapján egy kettős exponenciális függvény (a kitevőben is egy exponenciális kifejezés szerepel). Az eredeti modell és levezetése alapján a nemzetközileg leginkább elterjedt és elfogadott forma jelenleg a következő:

$$\hat{y}_t = Ke^{-be^{-ct}},$$

$$\ln \hat{y}_t = \ln K + (-be^{-ct}).$$

Ahol

- c – a növekedési sebességet jellemző paraméter, $c > 0$;
- b – a helyzetparaméter, $b > 0$.

⁸⁷ Benjamin Gompertz [1779–1865] biztosítási matematikus. Ld.: (Gompertz [1825].

⁸⁸ Ld.: Valkovics Emil [2001] 121–141

⁸⁹ Kotz [2006] 14: kötet, 8727–8728.

Az 3-15 ábra mutatja be a függvény alakját és nevezetes pontjait.

A Gompertz-függvény a logisztikus függvéynél meredekebben emelkedik a fejlődési szakaszban és így hamarabb éri el a telítettségi szintet.

A függvény inflexiós pontjának koordinátái:

$$\frac{d^2 \hat{y}_t}{dt^2} = e^{-be^{-ct}} (c^2 b^2 K e^{-2ct} - c^2 b K e^{-ct}) = 0$$

$$t_w = \frac{\ln b}{c},$$

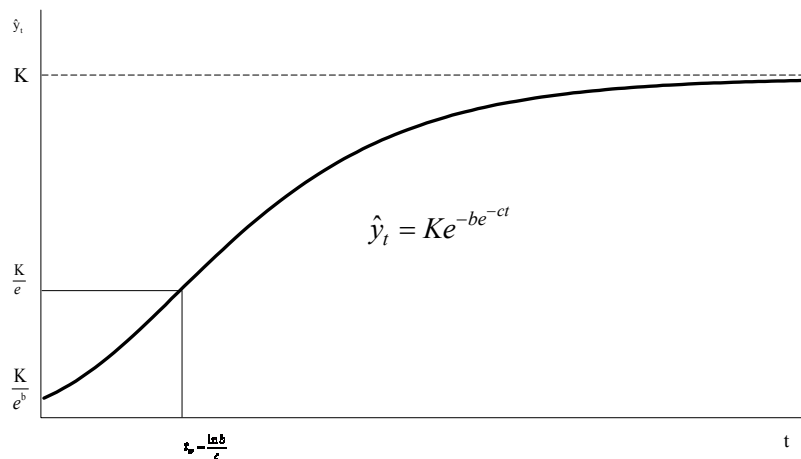
$$\hat{y}_{t_w} = \frac{K}{e}.$$

A harmadik derivált ebben a pontban nem egyenlő 0-val, tehát ebben a pontban inflexiós pont van:

$$\left. \frac{d^3 \hat{y}_t}{dt^3} \right|_{t_w} = 0,$$

$$e^{-\ln b} - 3be^{-2\ln b} + b^2 e^{-3\ln b} = 0,$$

$$b = \pm \infty.$$



3-15. ábra: A Gompertz-féle trendfüggvény

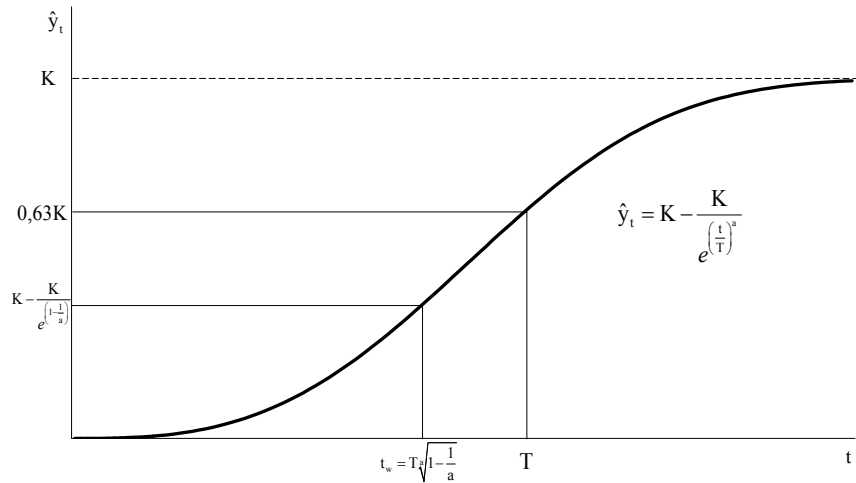
A függvény helyettesítési értéke a $t = 0$ helyen és a telítődési szint:

$$\hat{y}_0 = K e^{-b} = \frac{K}{e^b},$$

$$\lim_{t \rightarrow \infty} \hat{y}_t = K.$$

A 63 százalékos trendfüggvény

A 63 százalékos függvény egyik nevezetes pontjáról kapta a nevét: a függvény a $t = T$ időpontban éri el a telítődési szint 63 százalékát. A 3-16. ábra mutatja a függvény alakját és nevezetes pontjait.



3-16. ábra: A 63 százalékos trendfüggvény

A függvényt leíró formula:

$$\hat{y}_t = K - \frac{K}{e^{\left(\frac{t}{T}\right)^a}},$$

ahol

- T – a helyzetparaméter, ha $T > 0$;
- a – a növekedési sebességet jellemző paraméter, ha $a > 1$.

A függvény nevét adó pont, $t=T$ esetén a függvény a telítődési érték 63%-át veszi fel.

$$\hat{y}_T = K - \frac{K}{e} = K \left(1 - \frac{1}{e}\right) = 0,632K.$$

A függvény inflexiós pontjának koordinátái:

$$\frac{d^2 \hat{y}_t}{dt^2} = -e^{-\left(\frac{t}{T}\right)^a} \left[\frac{Ka^2 \left(\frac{t}{T}\right)^{2a}}{t^2} + \frac{Ka(1-a) \left(\frac{t}{T}\right)^a}{t^2} \right] = 0,$$

$$t_w = T \sqrt[1-a]{1 - \frac{1}{e}},$$

$$\hat{y}_{t_w} = K - \frac{K}{e^{\left(1 - \frac{1}{e}\right)^a}}.$$

A harmadik derivált ebben a pontban csak akkor egyenlő 0-val, ha $a = 1$, viszont az $a > 1$, tehát az inflexiós pont létezésének elégséges feltételét is bizonyítottuk:

$$\left. \frac{d^3 \hat{y}_t}{dt^3} \right|_{t_w} = 0,$$

$$a^2 \left(1 - \frac{1}{a}\right)^3 + 3a(1-a) \left(1 - \frac{1}{a}\right)^2 + (a-1)(a-2) \left(1 - \frac{1}{a}\right) = 0.$$

A két említett nevezetes pont $(\hat{y}_T, \hat{y}_{t_w})$, azaz a 63%-os és az inflexiós pont koordinátái alapján jól látható, hogy az a paraméter növekedése esetén a két pont egyre közelebb kerül egymáshoz, azaz $\lim_{a \rightarrow \infty} t_w = T$ és

$$\lim_{a \rightarrow \infty} \hat{y}_{t_w} = \hat{y}_T.$$

A függvény helyettesítési értéke a $t = 0$ helyen és a telítődési szint:

$$\hat{y}_0 = 0,$$

$$\lim_{t \rightarrow \infty} \hat{y}_t = K.$$

Johnson-trendfüggvény

A Johnson-görbe a logisztikus függvényénél gyorsabban emelkedik és nem szimmetrikus, azaz az inflexiós pont rövidebb idő alatt érhető el, tehát az inflexiós pont és a telítettségi szint közötti szakasz hosszabb. A gyors növekedést lényegesen lassúbb telítődési (érett növekedési) szakasz követi, mint a logisztikus függvény esetében. A 3-17. ábra mutatja be a Johnson-trendfüggvény alakját és nevezetes pontjait. Az inflexiós pont korán, a $b/2-c$ időpontban bekövetkezik, amit egy elnyújtott, hosszabb konkáv szakasz követ.

A függvényt leíró formula:

$$\hat{y}_t = e^{K - \frac{b}{c+t}},$$

$$\ln \hat{y}_t = K - \frac{b}{c+t},$$

ahol

- b – a helyzetparaméter, ha $b > 0$;
- c – a növekedési sebességet jellemző paraméter, ha $c > 0$.

A függvény inflexiós pontjának koordinátái:

$$\frac{d^2 \hat{y}_t}{dt^2} = \frac{ae^{K - [a/(b+t)]} [a - 2(b+t)]}{(b+t)^4} = 0,$$

$$\frac{d^2 \hat{y}_t}{dt^2} = \frac{be^{K - [b/(c+t)]} [b - 2(c+t)]}{(c+t)^4} = 0,$$

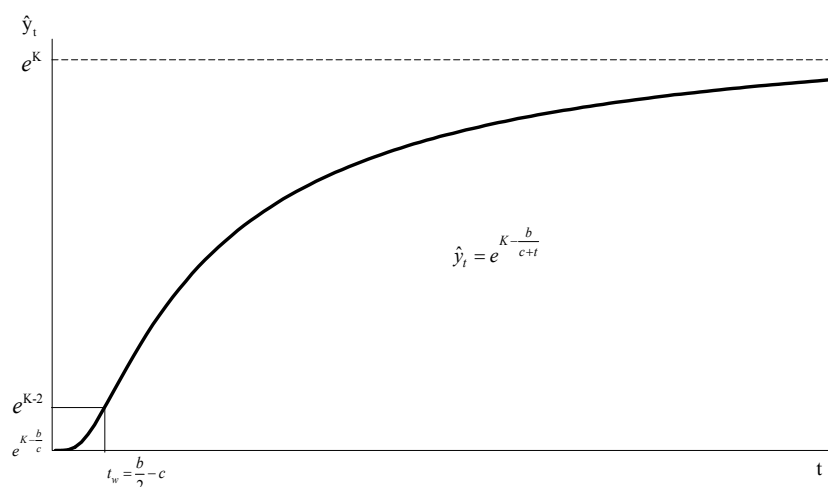
$$t_w = \frac{b}{2} - c,$$

$$\hat{y}_{t_w} = e^{K-2}.$$

A harmadik derivált ebben a pontban nem egyenlő 0-val, hiszen $b > 0$, tehát ebben a pontban inflexiós pont van:

$$\left. \frac{d^3 \hat{y}_t}{dt^3} \right|_{t_w} = 0,$$

$$b^2 \neq 0.$$



3-17. ábra: A Johnson-trendfüggvény

A függvény helyettesítési értéke a $t = 0$ helyen és a telítődési szint:

$$\hat{y}_0 = e^{\frac{K-b}{c}},$$

$$\lim_{t \rightarrow \infty} \hat{y}_t = e^K.$$

Az általánosított Richards-féle logisztikus trendfüggvény

Richards⁹⁰ kiegészítette egy v paraméterrel a logisztikus Verhulst - függvényt, ezzel az inflexiós pontban aszimmetrikussá téve azt:

$$\hat{y}_t = \frac{K}{(1 + ve^{-c(t-m)})^{1/v}},$$

ahol

- v – szabályozza az inflexiós pontban felvett függvényértéket, $v > 0$;
- c – növekedési sebességet jellemző paraméter, $c > 0$;
- m – az inflexiós pont, $m > 0$.

A függvény inflexiós pontjai:

$$t_w = m,$$

$$\hat{y}_{t_w} = \frac{K}{(1+v)^{1/v}}.$$

A függvény helyettesítési értéke a $t = 0$ helyen és a telítődési szint:

$$\hat{y}_0 = \frac{K}{(1 + ve^{cm})^{1/v}},$$

$$\lim_{t \rightarrow \infty} \hat{y}_t = K.$$

Az általánosított Richards-féle logisztikus trendnek a hazai és a nemzetközi szakirodalomban jelenleg széles körben elfogadott képletében feloldották azt a feltételezést, hogy a függvény nem rendelkezik alsó aszimptotával.⁹¹ Ez az öt paraméterrel rendelkező függvény egy olyan S-alakú görbét határoz meg, amely pályáját az A alsó korlát és a K telítődési szint között futja be.

A függvényt leíró formula

$$\hat{y}_t = A + \frac{(K - A)}{(1 + ve^{-c(t-m)})^{1/v}}.$$

A függvény inflexiós pontjának koordinátái:

$$\frac{d^2 \hat{y}_t}{dt^2} = c^2 e^{c(m+t/v)} (K - A) (e^{cm} - e^{ct}) (ve^{cm} - e^{ct})^{-(2v+1)/v} = 0,$$

$$t_w = m,$$

$$\hat{y}_{t_w} = A + \frac{(K - A)}{(1+v)^{1/v}}.$$

A harmadik derivált ebben a pontban nem egyenlő 0-val, mivel $v > 0$, tehát ebben a pontban inflexiós pont van:

$$\left. \frac{d^3 \hat{y}_t}{dt^3} \right|_{t_w} = 0,$$

$$v = -1.$$

A függvény helyettesítési értéke a $t = 0$ helyen és a telítődési szint:

⁹⁰ Ld.: Richards ([1959] 290–301. és Xinyou Yin et al. [2003] 361–371.

⁹¹ Ld.: Pella–Tomlinson [1969] 421-496. Colin [1999] 713-723. és Fokasz [2006] 29.

$$\hat{y}_0 = A + \frac{(K - A)}{(1 + ve^{em})^{1/v}},$$

$$\lim_{t \rightarrow \infty} \hat{y}_t = K.$$

Ha $A = 0$ és $v = 1$, akkor az általánosított Richards-féle trendfüggvény logisztikus függvénné alakítható:

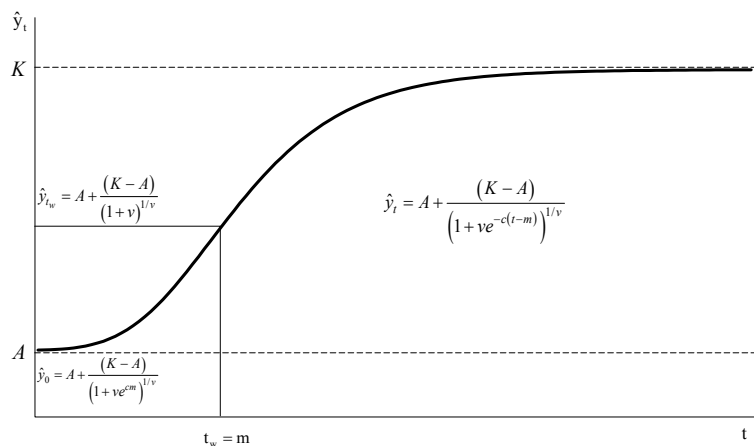
$$\hat{y}_t = \frac{K}{1 + e^{-c(t-m)}}.$$

Ha $A = 0$, $m = 0$ és $v \rightarrow 0$, akkor a Richards-függvény Gompertz-függvénné alakul át⁹², mert $1 + x \approx e^x$, ha $x \rightarrow 0$:

$$\hat{y}_t = Ke^{-e^{-c(t-m)}}.$$

Ha $A = 0$, $m = 0$ és $v = -1$ a Richards-függvény Mitscherlich-trendfüggvénné alakul:

$$\hat{y}_t = \frac{K}{(1 - e^{-c(t-0)})^{-1}} = K(1 - e^{-ct})$$



3-18. ábra: Az általánosított Richards-féle trendfüggvény

3.4.3 Két inflexió ponttal rendelkező trendfüggvények

Az egy inflexió ponttal rendelkező telítődési görbék bemutatása után a két inflexió ponttal rendelkező trendek közül az életgörbe⁹³ és Hubbert-függvényeket ismertetjük.

Életgörbe trendfüggvény

Az életgörbe trendfüggvény a termékéletgörbe alakulását mutatja, nevét is innen kapta, mivel a termék piaci forgalmának (volumenének) alakulását ábrázolja az idő függvényében. A következő – a marketing szakirodalmában ismert – szakaszokat lehet megkülönböztetni: keletkezés, bevezetés, növekedés, érettség (telítődés, itt éri el a forgalom a maximális értéket), tehát eddig egy S-alakú trenddel leírható a folyamat alakulása, s ezt követően történik a változás, a telítődést követi a hanyatlás. A termék kereslete egy bizonyos idő után drasztikusan csökkenhet. Dönteni kell a termék gyártásának leállításáról, a piacról való kivonásról. A 3-19 ábra mutatja be az életgörbe függvény alakját és nevezetes pontjait. A vállalatgazdasági szakemberek⁹⁴ általánosan elfogadják a termékéletgörbék leírását a következő trendfüggvénnyel

$$\hat{y}_t = ae^{-\omega^2(t-\tau)^2} = \frac{a}{e^{\omega^2(t-\tau)^2}},$$

ahol a termék életgörbe alakulásának megfelelően:

⁹² Ld.: Xinyou et al. [2003] 362.

⁹³ Hausteine [1972] az életgörbe trendet ökológiai függvénynek nevezi.

⁹⁴ Ld.: Korán [1978] 123–124., Kotler–Keller [2006] 189. és Iványi [1984] 28–29.

- \hat{y}_t – a termékből a t -edik évben értékesített mennyiség becsült értéke;
- a – az éves termelési értékesítési volumenek várható maximuma;
- ω – a görbe alakját, az inflexiós pontok helyét meghatározó alakparaméter;
- τ – a maximális értékesítés várható időpontja.

A függvény maximuma, mivel ebben a pontban az első derivált nulla és a második derivált a τ pontban negatív -1 :

$$\frac{d\hat{y}_t}{dt} = 2a\omega^2 (\tau - t) e^{-\omega^2(t-\tau)^2} = 0,$$

$$t_{\max} = \tau,$$

$$\hat{y}_{t_{\max}} = a.$$

A függvény inflexiós pontjai:

$$\frac{d^2\hat{y}_t}{dt^2} = 2a\omega^2 e^{-\omega^2(t-\tau)^2} [2\omega^2 (t-\tau)^2 - 1] = 0,$$

$$t_{w_1;w_2} = \tau \pm \frac{1}{\sqrt{2}\omega},$$

$$\hat{y}_{t_{w_1;w_2}} = \frac{a}{\sqrt{e}}.$$

Bizonyítható, hogy a harmadik derivált ezekben a pontokban nem egyenlő 0-val.

A függvény helyettesítési értéke a $t = 0$ helyen és a telítődési szint:

$$\hat{y}_0 = \frac{a}{e^{(\omega\tau)^2}},$$

$$\lim_{t \rightarrow \infty} \hat{y}_t = 0.$$

Az életgörbe trendfüggvény és a normális eloszlás sűrűségfüggvénye közötti összefüggést az alábbiakban mutatjuk be⁹⁵.

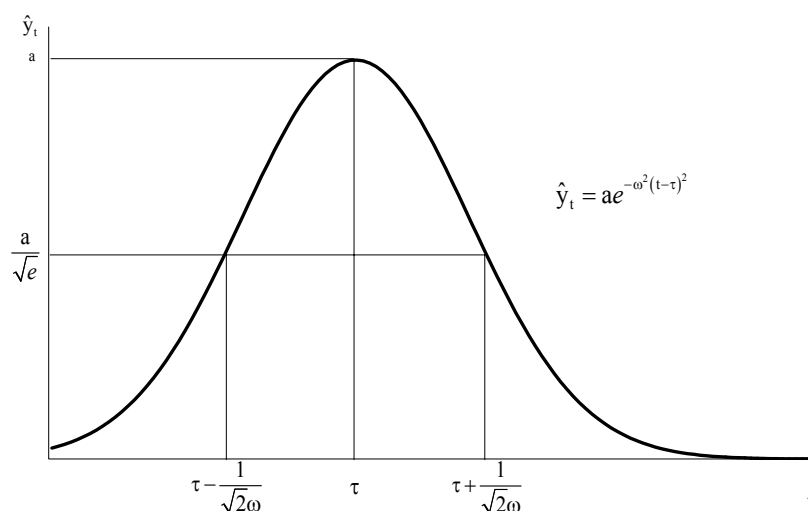
Amennyiben

$$a = \frac{1}{\sqrt{2\pi\sigma^2}}, \quad \omega = \frac{1}{\sqrt{2\sigma^2}},$$

úgy

$$\hat{y}_t = a e^{-\omega^2(t-\tau)^2} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(t-\tau)^2}$$

a τ várható értékű, σ^2 varianciájú normális eloszlás sűrűségfüggvénye.



3-19. ábra: Az életgörbe trendfüggvény

⁹⁵ Haustein [1972] 186–187.

Hubbert-trendfüggvény

A Hubbert-féle trendfüggvény alapján olajhozamcsúcsnak nevezzük a kőolaj kitermelésének időbeli tetőzését. Az olajhozamcsúcs az ún. Hubbert-féle csúcsmélelet alapján számítható, amelyet Hubbert, a Shell Oil Kutatólaboratórium geofizikusa 1956-ban alkotott meg. Az elmélet az Egyesült Államok kőolaj-kitermelésének maximumát 1965–1970 időszakra becsülte. Ezzel mindössze egy évet tévedett, az Egyesült Államok kitermelési csúcsa 1971-ben volt. A földgáz, a kőszén, a vasérc, valamint más nyersanyagok rendelkezésre álló mennyisége – hasonlóan a kőolajhoz – véges, így termelésük hosszú távon a Hubbert-függvénnyel prognosztizálható. *Hubbert*⁹⁶ matematikai modelljét egy olajmező várható élettartamának modellezésére dolgozta ki. A modell alkalmazható egyes területek vagy akár az egész Föld készleteire is.⁹⁷ A 3-20 ábra mutatja be a Hubbert-trendfüggvény alakját és nevezetes pontjait.

A függvényt leíró formula⁹⁸

$$\hat{y}_t = \frac{bUe^{b(t-\tau)}}{(1+e^{b(t-\tau)})^2} = \frac{bUe^{b(t+\tau)}}{(e^{bt} + e^{b\tau})^2},$$

ahol

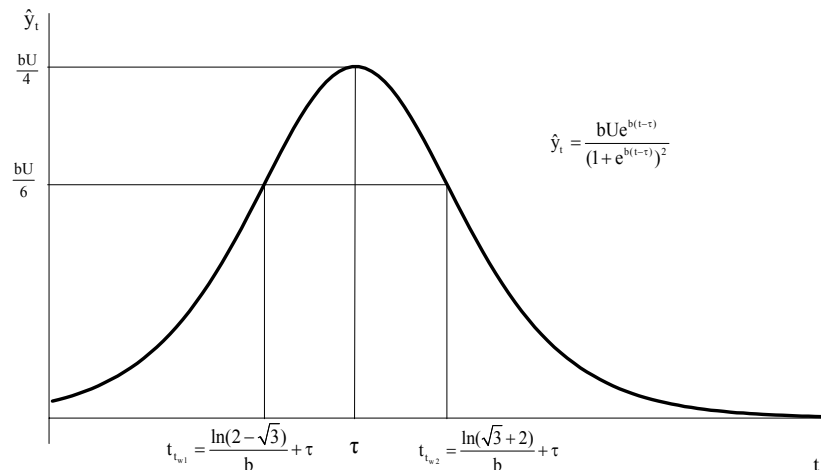
- b – a meredekséget kifejező, vagyis az emelkedést (a felszálló ágat) és az ereszkedést (leszálló ágat) leíró, az inflexiós pontokat meghatározó alakparaméter;
- U – a teljes Hubbert-trend becsült értékeinek összege,⁹⁹ az időtengelyen;
- τ – az az időpont, ahol a görbének csúcspontja van: $\hat{y}_\tau = \frac{bU}{4}$.

A függvény maximuma, mivel ebben a pontban az első derivált nulla és a második derivált a τ pontban negatív ($-2e^{2b\tau}$):

$$\frac{d\hat{y}_t}{dt} = \frac{b^2 U e^{b(t+\tau)} (e^{b\tau} - e^{bt})}{(e^{b\tau} + e^{bt})^3} = 0,$$

$$t_{\max} = \tau,$$

$$\hat{y}_{t_{\max}} = bU/4$$



3-20. ábra: A Hubbert-trendfüggvény

A függvény inflexiós pontjai:

⁹⁶ Marion King Hubbert (1903–1989). Kutatási eredményei: www.hubbertain.com/hubbertain/Bibliography.htm. (Elérés dátuma: 2009. január 28.)

⁹⁷ Hubbert 1956-ban megjelent tanulmányában azt prognosztizálta (www.hubbertain.com/hubbertain/1956/1956.pdf), hogy például a texasi kőolaj- és földgáztermelés 1973-ban eléri a csúcstermelést, majd ezt követően a termelés csökkenni fog és 2050-ben meg fog szűnni. A Hubbert-prognózist napjainkig igazolta az idő.

⁹⁸ Laherrère [2000]

⁹⁹ Ultimate recovery of crude oil: the total resource available (a kőolaj utolsó kitermelése, vagyis az összes elérhető készlet).

$$\frac{d^2 \hat{y}_t}{dt^2} = 0,$$

$$e^{2bt} - 4e^{b(t+\tau)} + e^{2b\tau} = 0$$

$$t_{w_1, w_2} = \tau \pm \frac{\ln(2 + \sqrt{3})}{b} \approx \tau \pm \frac{1,317}{b},$$

$$\hat{y}_{t_{w_1, w_2}} = \frac{bU}{6}.$$

A függvény helyettesítési értéke a $t = 0$ helyen és a telítődési szint:

$$\hat{y}_0 = \frac{bUe^{-b\tau}}{(1 + e^{-b\tau})^2},$$

$$\lim_{t \rightarrow \infty} \hat{y}_t = 0.$$

3.4.4 A logisztikus trendek becslése Excel parancsfájl működése

A logisztikustrendbecslése.xls Excel parancsfájl a bemutatott tizenkét logisztikus trendfüggvény illesztését könnyíti meg a felhasználók számára. Tizenkét olyan munkalapot tartalmaz, melyek a különböző függvényformák nevei alapján kerültek elnevezésre, valamint egy Ciklus nevű munkalapot, mely a különböző illesztett trendek alapján a rövid és hosszú távú ciklusok vizsgálatát teszi lehetővé. A fájlban a cellák színezése jelentőséggel bír: a halványsárga cellák szabadon változtathatók, a zöld cellák az egyes paraméterek javasolt kezdeti értékeit adják meg, míg a fehér cellák számítási (részeredményeket tartalmaznak). A színezés alapján látható, hogy a fájl maximálisan 1000 hosszúságú idősor feldolgozására képes. Az egyes munkalapok között nincs összefüggés, azaz amennyiben több trendet kíván a felhasználó illeszteni ugyanarra az adatsorra, úgy az adatokat valamennyi kiválasztott lapra be kell másolnia. Új adatbevitel esetén mind a 12 munkalapon törölni kell az adatokat, az adatok törlése ikonra kattintva, utána pedig bemásolni az új adatállományt. A fájl az adatok beillesztését követően azonnal ábrázolja az idősort, valamint az aktuálisan bevitt paraméterek alapján a trendfüggvényt is. Az Idő oszlop kitöltése opcionális, amennyiben kitöltésre kerül, úgy az ábrázolás esetén ezt figyelembe veszi a fájl az időtengely felirataként. A javasolt kezdeti paraméterek az egyes függvények nevezetes pontjai (maximálisan felvett érték, első időszak megfigyelésének értéke, inflexiós pont stb.) alapján kerültek meghatározásra. A Pearl–Reed-féle logisztikus függvény példáján keresztül mindez azt jelenti, hogy a K paraméter javasolt értéke az idősorban található maximális értékkel egyezik meg ($\tilde{K} = y_{\max}$). A b paraméter kezdő értékének meghatározása az

$$\hat{y}_0 = \frac{K}{1+b} \approx y_1$$

összefüggés alapján

$$\tilde{b} \approx \frac{\tilde{K}}{y_1} - 1$$

módon történik, ahol \sim -mal az adott paraméter közelítő, kezdeti értékét jelöljük. Az inflexiós pont nyújt segítséget a harmadik paraméter közelítő meghatározásához. A fájl megkeresi az inflexiós pontban felvett $K/2$ függvényértékhez legközelebb eső, de annál kisebb tényleges idősori értéket, amiből t_w feltételezett értéke következik. Ekkor

$$\tilde{c} = \frac{\ln \tilde{b}}{\tilde{t}_w}$$

alapján következtethetünk c közelítő értékére. A többi trend esetén az ajánlott kezdőértékek teljesen hasonló módon kerültek meghatározásra. Az induló paraméterek ilyen módon történő meghatározása azt okozhatja, hogy amennyiben olyan jelenséget vizsgálunk, amely már „lefutott”, és az illeszteni kívánt függvény alkalmas, úgy a kezdő paraméterek megadásával jól illeszkedő függvényt kapunk. Amennyiben egy telítődési, vagy életciklus-folyamat elején tartó jelenséget vizsgálunk, úgy a feltüntetett paraméterek helyett szakértői, elemzői tapasztalatra kell támaszkodni. Az induló paraméterek természetesen nem min-

den esetben adnak tökéletes javaslatot, így lehetőség van a paraméterek kézi vezérlésére is. Valamennyi munkalap tartalmaz olyan parancsgombokat, melyek a paraméterek finomhangolását végzik el (Opt. mind). A parancsgombok az Excel beépített Solver funkcióját hívják meg, a célfüggvény pedig az R^2 maximalizálása az egyes paraméterek iteratív változtatásával. Lehetőség van arra is, hogy a Solver a (kézzel, szakértői becslés alapján beállított) telítődési paraméter értékén ne változtasson, ekkor csupán a többi paraméter nagyságát fogja a program meghatározni (Opt. K nélkül). Az Excel beépített Solver csomagja nem képes minden esetben globális optimumot találni,¹⁰⁰ így érdemes az illesztést több különböző, kézzel beállított indulóértékkel elvégezni. Főként az életgörbe és Hubbert függvények esetén problémát okozhat a paraméterek nagyságrendjének jelentős eltérése. A Solver ebben az esetben a túlságosan nagy paramétert nem mozditja el kezdeti értékéről. A megoldás az eredeti adatsor dimenziójának változtatása (pl. 1000-rel való osztás). A fájl

$$R^2 = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y}_t)^2} = 1 - \frac{\text{SSE}}{\text{SST}}$$

módon számít, ahol SSE (Sum of Squared Errors) a reziduumok négyzetösszege; SST (Sum of Squares Total) a teljes eltérés négyzetösszeg.

A kielégítőnek ítélt paraméterek megtalálása után az extrapoláció beállításával lehetőség van a trend mechanikus kiterjesztésére (a megfigyelések száma az extrapolációval együtt sem lépheti át az 1000 darabot). A **Ciklus munkalapon** az illesztett trendek további vizsgálatára (reziduumok ábrázolása, mozgóátlagolása), és összehasonlítására nyílik lehetőség. A Ciklus munkalap valamennyi esetben az adott függvény munkalapján beállított paraméterezés alapján dolgozik.

A ciklus munkalapon kiválaszthatjuk a 12 trend közül azt, amit vizsgálni kívánunk (természetesen egyenként mind a 12 trendfüggvény konjunktúra ciklusait vizsgálhatjuk és összehasonlíthatjuk) és meg kell adni a mozgóátlag tagszámát. A mozgóátlag tagszáma az adatbázistól függ. Rövid ciklusok vizsgálata esetében a szezonaritást kűszöböljük ki, ha havi adatok állnak rendelkezésre a mozgóátlag tagszáma 12, negyedéves adatok esetében 4. (Megjegyezzük, hogy a szezonális hullámzások kiszűrése után, ha ezt követően kiszűrjük a trendhatást, akkor a Kitchin-féle rövid ciklus becslését kapjuk meg.)

Ha a hosszú ciklusokat vizsgáljuk éves adatokkal, akkor általában 8 vagy 9 tagú mozgóátlagot választunk a rövidebb periódusú (pl. 4-8 vagy 3-9 éves) ciklusok kiszűrése érdekében. (Megjegyezzük, hogy a rövidebb ciklusok kiszűrése után, ha ezt követően kiszűrjük a trendhatást, akkor a Kondratyev – féle hosszú ciklus becslését kapjuk meg.) Vizsgálni lehet az idősorokat abból a szempontból is, hogy hogyan reagálnak a mozgóátlag tagszámának megváltoztatására. Választhatunk az additív és a multiplikatív modell között. Az adatok bevitele után a program ábrázolja az eredeti adatokat és a trendet és a második ábrában a ciklust. Kiszámítja multiplikatív kapcsolatot feltételezve az eredeti idősor és a trend hányadosát (y_{ij} / \hat{y}_{ij}) valamint additív kapcsolatot feltételezve az eredeti idősor és a trend különbségét ($y_{ij} - \hat{y}_{ij}$). Így mindkét feltételezés mellett kikűszöböli a trendhatást. A ciklikus mozgásokat a trendtől megtisztított idősoroknál a felhasználó által megadott mozgóátlag tagszám alapján mozgóátlagolással kűszöböli ki és számítja valamint ábrázolja.

Ciklus munkalapon: Válassza ki a trend típusát, a mozgóátlag tagszámát (ha 1 akkor nincs mozgóátlagolás) a trend és a periodikus hullámzás kapcsolódás módját: additív vagy multiplikatív. A sárga kockákban levő számok cserélhetők. Ha a mozgóátlag tagszáma 1, akkor az eredeti adatok és a trend különbségével illetve hányadosával számol a program.

Ha az Idő oszlopba bemásoljuk a tényleges időt, pl. éveket, akkor az ábra X tengelye ezt veszi figyelembe, ha az Idő oszlop üresen marad, akkor a sorszámok jelennek meg az X tengelyen. Ha a két tengelyt egyenként egérművelettel kijelöljük (bal gomb) akkor a jobb gomb lenyomásával a skála mértéke változtatható. Az adatok bevitele után a program közli a trend paramétereket, az illesztés pontosságát, az $R^2 - t$ és elkészíti a trend ábrát. Közli továbbá az SSE és SST értékeket is, amiből az $R^2 - t$ számítja.

Gyakorló feladatok. Dekompozíciós idősormodellek

¹⁰⁰ A piacon elérhető globális optimumot meghatározó, Excelbe beépülő Solverek is, ám ezek nem ingyenesek.

1. Az IBM¹⁰¹ értékesítésének és nyereségének prognosztizálása, a trendszezon-hibaszámítás.xls Excel parancsfájl felhasználásával.

Válassza ki a két legjobban illeszkedő trendfüggvényt a MAPE hibaképlet alapján, ha:

- az IBM árbevételét prognosztizálja: becslési időszak 1954-1979, teszt időszak 1980-1984:
- az IBM árbevételét prognosztizálja: becslési időszak 1954-1984, teszt időszak 1985-1990:
- az IBM nyereségét prognosztizálja: becslési időszak 1954-1979, teszt időszak 1980-1984:
- az IBM nyereségét prognosztizálja: becslési időszak 1954-1984, teszt időszak 1985-1990:

Ábrázolja:

- az IBM értékesítési árbevételének tényleges alakulását 1985 és 2007 között.
- az IBM nyereségének tényleges alakulását 1985 és 2007 között.
- az IBM létszámának tényleges alakulását 1985 és 2007 között.

Milyen következtetéseket von le az ábrák alapján.

2. Grafikusan ábrázolja Anglia (UK) rendelkezésre álló hosszú idősorait, elemezze a hosszútávú tendenciákat, az évszázados trendeket, és végezzen trendextrapolációt a legjobban illeszkedő trend alapján!

2.1. Az alábbi idősorok esetében: a becslési időszak 1830-2000, teszt időszak: 2001-2006, trendextrapoláció: 2007-2011.¹⁰²

1830-2006-ig rendelkezésre álló idősorok:

- Névleges GDP (millió font)
- Reál GDP (2003-as millió font)
- GDP deflációs index (index 2003=100%)
- Néesség (ezer fő)
- Nominál GDP/fő (forgalomban lévő font)
- Reál GDP/fő

2.2. Az alábbi idősorok esetében: a becslési időszak: 1264-2000, a teszt időszak: 2001-2006, trendextrapoláció: 2007-2011.¹⁰³

1264-2006-ig rendelkezésre álló idősorok:

- Kiskereskedelmi árindex (1913 = 100 %)
- Átlagos nominális nyereség (1913 = 100 %)
- Átlagos reálnyeresség (1913 = 100 %)

2.3. Az alábbi idősor esetében: a becslési időszak: 1257-2000, a teszt időszak: 2001-2006, trendextrapoláció: 2007-2011.¹⁰⁴

Egy uncia színarany év végi ára angol fontban.

2.4. Az alábbi idősor esetében: a becslési időszak: 1265-2000, a teszt időszak: 2001-2006, trendextrapoláció: 2007-2011.¹⁰⁵

A fogyasztói árindex alakulása az előző évhez viszonyítva %-ban.

3. Ábrázolja és elemezze a BUX-index napi záró értékének alakulását¹⁰⁶ 1991.01 és 2008.04.04 között.

4. Mutassa ki a trendet és a szezonindexeket Magyarország energiafelhasználása 1999-2007 (PJ) idősora (negyedéves adatok) alapján.

5. Az USA lakossági teljes villamosenergia fogyasztás (milliárd kwóra/hónap) idősora 1973 január és 2007 november között rendelkezésünkre áll.¹⁰⁷ Mutassa ki a szezonális hatást. Melyik analitikus trend illesztés adja a legjobb eredményt. A trend és a szezonkomponens között multiplikatív- vagy additív –e a kapcsolat.

¹⁰¹ Az adatok forrása: http://www-03.ibm.com/ibm/history/history/history_intro.html

¹⁰² <http://www.measuringworth.com/datasets/ukgdp/result.php>.

¹⁰³ Az adatok forrása: <http://www.measuringworth.com/datasets/ukearnpci/result.php>

¹⁰⁴ Az adatok forrása: <http://www.measuringworth.com/datasets/ukearnpci/result.php> Regionális beállításokat módosítani kell.

¹⁰⁵ Az adatok forrása: <http://www.measuringworth.com/inflation/#> Regionális beállításokat módosítani kell..

¹⁰⁶ Az MNB közli 1991.01.02 – től folyamatosan a BUX-index napi átlagos indexének alakulását (1991. január 2.=1000) Excel fájlban. http://www.mnb.hu/engine.aspx?page=mnbhu_statistikak

¹⁰⁷ <http://www.eia.doe.gov/emeu/mer/elect.html>

6. Grafikusan ábrázolja az Amerikai Egyesült Államok (USA) rendelkezésre álló hosszú idősorait, elemezze a hosszútávú tendenciákat, az évszázados trendeket, és végezzen trendextrapolációt a legjobban illeszkedő trend alapján Az USA alábbi hosszú idősoraival rendelkezünk.

6.1 Rövid lejáratú kamatláb (%) 1831-2006 között.¹⁰⁸

6.2 Fogyasztói árindex (CPI = Consumer Price Index, CPI =1982-84=100) 1774-2007 között.¹⁰⁹

6.3 A szinarany \$/uncia év végi ár alakulása 1786 és 2006 között.¹¹⁰

6.4 Ábrázolja és elemezze a Dow-Jones index napi záróértékének az alakulását 1885.02.17 és 2007.04.03 (33739 adat!) között.

6.5 Egy főre jutó aranytermelés 1860-2005 között.

6.6 Egy főre jutó kőolajtermelés 1866-2005

6.7 Egy főre jutó kőszénttermelés 1860-2005

6.8 Egy főre jutó ólomtermelés 1830-2005

6.9 Egy főre jutó acéltermelés 1867-2005

6.10 Egy főre jutó vasérctermelés 1880-2004

6.11 Egy főre jutó alumíniumtermelés 1896-2004

6.12 USA kőolajtermelése naponta 1900-2007

7. Grafikusan ábrázolja a világ összesen rendelkezésre álló hosszú idősorait, elemezze a hosszútávú tendenciákat, az évszázados trendeket, és végezzen trendextrapolációt a legjobban illeszkedő trend alapján A világ összesen alábbi hosszú idősoraival rendelkezünk.

7.1 Egy főre jutó aranytermelés 1876-2005 között.

7.2 Egy főre jutó ezüsttermelés 1870-2005 között

7.3 Egy főre jutó kőolajtermelés 1860-2005

7.4 Egy főre jutó kőszénttermelés 1860-2005

7.5 Egy főre jutó barnaszénttermelés 1890-2005

7.6 Egy főre jutó acéltermelés 1890-2005

3.5 Naiv előrejelzési technikák. (A naivmódszer-parancsfájl működése.)

A naiv szabályok^{111 112 113} egyszerű, de potenciálisan hatékony idősor előrejelzési módszerek. Szabályok abban az értelemben, hogy előre meghatározottak, így nincs szükség paraméterértékek becslésére, mint pl. a dekompozíciós vagy az exponenciális simítási módszereknél. A naivítás azt a tényt jelenti, hogy az idősor bármely naiv előrejelzésének alapja az idősor önmaga. Az idősort használjuk önmaga előrejelzésére, azaz az idősor történeti értékeit használjuk ugyanazon idősor jövőbeni értékeinek képzésére vagy létrehozására. A naiv szabályok egyszerűségükben viszonylag alacsony költségű előrejelzési módszerek, de ha egy módszer hatékony, nem kell hezitálni az alkalmazásával egyszerűsége vagy naivitása miatt. A naiv szabályok hatékonyabbak rövidtávú, mint hosszú távú előrejelzésre. Amint a prognózistáv hosszabbodik, a naiv előrejelzés valószínűleg kevésbé pontos, és nagyobb az ilyen előrejelzéseken alapuló döntések velejáró kockázata. Minden idősor változást mutat az egyes megfigyelések és a következők között az idősor teljes hosszán. Bármely idősor feltehetően egy vagy több típusú hullámzást tartalmaz, amint korábban bemutattuk: a trendet, a konjunktúra ciklust, a szezonális hullámzást és a véletlent. Az elemzési módszer megpróbálja figyelembe venni az idősorban levő hullámzás minden típusát. A következőkben leírt egyszerű naiv szabályok készíthetők úgy, hogy figyelembe vegyék az idősorban levő trend és szezonális tényezőket, de a naiv szabályok ritkán képesek figyelembe venni a sorban levő ciklikus viselkedést. Az egyszerű, naiv szabályoknak négy csoportja van:

1. azok, amelyek sem a trendet, sem a szezonalitást nem veszik figyelembe (alapértelmezett előrejelzési szabályok),

¹⁰⁸ <http://www.measuringworth.com/datasets/interestrates/result.php>

¹⁰⁹ <http://www.measuringworth.com/uspci/>

¹¹⁰ <http://www.measuringworth.com/datasets/gold/result.php>

¹¹¹ Farnum Nicholas R., - Stanton LaVerne W. [1989]: 15-17. 105-108.

¹¹² <http://facweb.furman.edu/~dstanford/forecast/h3.htm>

¹¹³ Theiss Ede: szerk. [1958]: 220-221.

2. azok, amelyek a trendtényezőt figyelembe veszik, de felteszik, hogy a szezonális nem szignifikáns,
3. azok, amelyek figyelembe veszik a szezonális tényezőt, de felteszik, hogy a trend nem szignifikáns,
4. azok, amelyek megpróbálják figyelembe venni az idősor trend és szezonális komponenseit is.

Racionális emberek gyakran hoznak döntéseket anélkül, hogy először bármilyen előrejelzési módszerrel foglalkoznának. Amikor így tesznek, azt feltételezik, hogy a jövőbeli állapot hasonló lesz a jelenhez vagy a közelmúlthoz. Az alapértelmezett előrejelzés megfelelő lehet a mindennapi élet sok minimális következménnyel járó döntésével kapcsolatban. Ezért az alapértelmezett előrejelzés nem szükségszerűen irracionális. Formalizálni lehet az alapértelmezett előrejelzési módszert, amely a következő formában írható:

1.1 Szabály:

$$\hat{y}_{t+1} = y_t$$

ahol:

- t = az az időpont, melyen a prognózis alapul, rendszerint a legutóbbi elérhető megfigyelés,
- \hat{y}_{t+1} = a következő megfigyelés előrejelzése a sorban.

A szabály általánosítható i prognózistávval, átírva a következőképpen:

1.2 Szabály:

$$\hat{y}_{t+i} = y_t$$

Az 1.2 szabály értelme az, hogy a sor állapota néhány, i , periódusban a jövőben várhatóan hasonló a megfigyeléshez, melyen az előrejelzés alapul. Ezek az előrejelzések azonban sem a trendet, sem a szezonálitást nem veszik figyelembe.

Az 1.1 és 1.2 szabály olyan naiv és egyszerű, hogy kételkedni lehet formalizálásuk célszerűségében. Három célt szolgálnak: (a) feltárják a szimbolikus jelölések használatát a legegyszerűbb formában; (b) kiindulópontként szolgálnak a következő szabályok kifejlesztéséhez; és (c) összehasonlítási alapszabályt hoznak létre, melyhez a többi előrejelzési szabály teljesítményének hatékonysága hasonlítható.

2. Szabályok, amelyek figyelembe veszik a trendtényezőt.

A trend a hosszú távú változás jelensége egy gyűjtött adatsorban általában azonos irányban az idősor teljes hosszában érvényesülhet. A trend jelenléte lehet, hogy nem felismerhető az idősorban néhány egymást követő megfigyelésből, különösen ha egyéb típusú hullámozás (szezonális, ciklikus vagy tiszta véletlen) is van benne. Az idősor ábrázolása feltárhatja a trend jelenlétét. Felfelé emelkedő trend növekedési jelenséget mutathat; lefelé lejtő irány csökkenést jelezhet.

A szabályok 2. osztályában, melyet a következőkben áttekintünk, az a feltevés, hogy nincs más típusú (pl. szezonális és ciklikus) szignifikáns tényező a sorban, mint a trend. A 2. osztályban használt előrejelzési módszer létrehoz egy trendkiigazítási tényezőt, melyet a megfigyelt sorhoz kapcsol, amely az előrejelzés alapját képezi.

A legegyszerűbb módszer, amely megpróbálja a változást figyelembe venni egy előrejelzési szabályban, a legutóbbi két megfigyelés közötti abszolút változást, a különbségképzést (elsőfokú differenciát) ($y_t - y_{t-1}$) hozzáadja a sor legutóbbi megfigyeléséhez y_t -hez, annak érdekében, hogy létrehozza a sor következő y_{t+1} értékének előrejelzését. Ezt a prognózis módszert naiv lineáris trendnek is hívhatjuk. Ha az elemző több, periódus értékét kívánja előrejelezni, a legutóbbi megfigyelésen túl, mindössze meg kell szorozni a számított változást i -vel, melyet ezután prognózistávnak nevezünk. Ez algebrailag a következőképpen írható:

2.1 Szabály:

$$\hat{y}_{t+i} = y_t + i(y_t - y_{t-1})$$

Ez nagyon leegyszerűsített trenddel foglalkozó módszer, mivel csak az adatok legutóbbi változásának információját tartalmazza. Ha ezt a szabályt összehasonlítjuk a többivel, úgy tekinthető, kielégítő eredményt adhat. Tegyük fel, okunk van azt feltételezni, hogy a vizsgált idősorban a növekedési ütem állandó, ezért a relatív változás sokkal kifejezőbb mint az abszolút változás. A 2.2 szabály a 2.1 szabály mó-

dosítása, vagyis a prognosztizált értéket a legutóbbi relatív változás y_t/y_{t-1} hozzákapcsolásával képezi figyelembe véve a legutóbbi megfigyelést:

2.2 Szabály:

$$\hat{y}_{t+i} = y_t (y_t/y_{t-1})^i$$

Ez az alkalmazás multiplikatív műveletet tartalmaz, a 2.1 szabály additív művelete helyett. Ha az i prognózistáv egy periódusnál nagyobb, a relatív változás tényezőt i -szer kell alkalmazni a legutóbbi megfigyelésre, ezért alkalmazzuk az i hatványkitevőt, melyre a relatív változást emeljük. A 2.2 szabály trenddel kapcsolatos fogalmi nehézségei azonosak a 2.1 szabályéival.

2.3 Szabály:

A 2.3 szabály képviseli az erőfeszítést a hosszú távú változás képletbe foglalása problémájának megoldására. A legutóbbi abszolút változás használata helyett kiigazítási tényezőként, 2.3 az összes egymást követő megfigyelés növekmény mediánját használja a sorban, és ezáltal alkalmazza a teljes adatsorra kiterjedő információt. A számtani átlag nagysága kizárólag az idősor legrégebbi ($k=2$) és legújabb (m) adatától függ, és ha ez a két adat eltér az általános tendenciától, akkor az átlagos növekmény nem ad pontos értéket. A medián, mint helyzeti középérték alkalmazása ezért indokolt, ugyanis a medián – mint korábban már bemutattuk - egy biztosan közepes, meglehetősen robusztus (azaz viszonylag érzéketlen a kiugró értékekre) középértéknek bizonyult. A medián robusztus tulajdonságát könnyű megérteni, ha arra gondolunk, hogy a rangsorba rendezett adatok szélső értékei nagyságát nem befolyásolják. A medián, a szó legszorosabb értelmében közepes érték, a mennyiségi ismérvértékek közül az az érték, amelynél ugyanannyi kisebb, mint nagyobb érték fordul elő.

Meghatározása igen egyszerű, mivel értéke a rangsorba rendezett ismérvértékek középső tagja, tehát, ha az elsőrendű differenciákat jelöljük:

$$\alpha_k = (y_k - y_{k-1})$$

- m = a megfigyelések száma az idősorban.
- $k=2, \dots, m$

- páratlan elemszámú adathalmaz esetén:

$$Me_\alpha = \alpha_{k\left(\frac{(m-1)+1}{2}\right)}$$

- páros számú adat esetén a medián nem esik egybe egy konkrét megfigyeléssel, így ilyenkor, konvencionálisan a

$$Me_\alpha = \frac{\alpha_{k\left(\frac{(m-1)+1}{2}\right)} + \alpha_{k\left(\frac{(m-1)+1}{2}+1\right)}}{2}$$

képlettel határozható meg.

A szabály algebrai formulája:

$$\hat{y}_{t+i} = y_t + i(Me_\alpha)$$

2.4 Szabály

A 2.3 szabályhoz hasonlóan a 2.4 szabály a teljes adatsorra kiterjedő információt használja. A 2.4 szabály a 2.2 módosítása úgy, hogy a periódusról periódusra számított relatív változás mediánját használja a teljes sorra (a legutóbbi egyperiódusú relatív változás helyett) trendkiigazítási tényezőként. Az átlagos növekményi változást úgy számítjuk, hogy az egymást követő egyperiódusú növekményi változások mediánját határozzuk meg. Az előrejelzés képlete: a legutolsó megfigyelt értéket meg kell szoroznunk a számított trendkiigazítási tényező (Me_α) i -ik hatványával. Mivel a 2.3 szabály a teljes adatsorra kiterjedő információt használja, helyesebben tekinthető trendet becslőnek, mint a 2.1 szabály.

$$\beta_k = (y_k/y_{k-1})$$

- m = a megfigyelések száma az idősorban.
- $k=2, \dots, m$

- páratlan elemszámú adathalmaz esetén:

$$Me_{\beta} = \beta_k \left(\frac{(m-1)+1}{2} \right)$$

- páros számú adat esetén a medián nem esik egybe egy konkrét megfigyeléssel, így ilyenkor, konvencionálisan a

$$Me_{\beta} = \frac{\beta_k \left(\frac{(m-1)+1}{2} \right) + \beta_k \left(\frac{(m-1)+1}{2} \right)}{2}$$

képlettel határozható meg.

A szabály algebrai formulája:

$$\hat{y}_{t+i} = y_t \left(Me_{\beta} \right)^i$$

Ennek a négy egyszerű naiv szabálynak az a célja, hogy figyelembe vegye a trendváltozást az adatsorban; ez nem az egyetlen lehetséges mód a trendváltozás figyelembevételére, csak a legegyszerűbb.

3. Szabályok, melyek figyelembe veszik a szezonális tényezőt.

A szezonális kapcsolódhat a mezőgazdasági munkákhoz, a szezonális időjárás változásokhoz, szokásokhoz és hagyományhoz, vallási vagy világi ünnepekhez. Fontos megjegyezni, hogy az egyik idősor szezonális sémája lehet, hogy hasonlít, lehet, hogy nem más idősorok szezonális sémájához.

3.1 Szabály.

A 3.1 szabály szemlélteti a legegyszerűbb módszert a próbálkozásra hogy figyelembe vegyük a szezonálisitást az idősorban:

havi adatok esetében:

$$\hat{y}_{t+i} = y_{t+i-12}$$

negyedéves adatok esetében:

$$\hat{y}_{t+i} = y_{t+i-4}$$

Ez a szabály nem tesz erőfeszítést a trend figyelembevételére. Alapfeltevése hasonló az alapértelmezett szabályokéihoz, kis kiigazítással: a sor értéke az előrejelzett hónapban (negyedévben) valószínűleg azonos lesz az előző év azonos hónapjával (negyedévével). Ha t a legutóbbi megfigyelési érték hónapja és i a prognózistáv, a $t+i$ periódus előrejelzését az előző év megfelelő hónapjának (negyedévének) megtalálásával (visszaszámolunk a sorban) kapjuk a $(t+i-12)$ illetve $(t+i-4)$ periódusnál. Ez a viszonylag egyszerű módszere a szezonálisnak a legutóbbi tizenkét hónap (négy negyedév) információit használja, és valóban elhagy minden korábbi információt.

4. Szabályok, melyek figyelembe veszik a trendet és a szezonálisitást is.

A 3.1 szabályban alkalmazott módszer, lehetővé teszi a 2.3 és 2.4 szabály trendkiigazítási tényezőjének módosítását. Ezeket a szabályokat tehát átírhatjuk a következőképpen.

4.1 Szabály:

$$\hat{y}_{t+i} = y_{t+i-12} + i \left(Me_{\alpha} \right)$$

Negyedéves adatoknál értelemszerűen a 12 helyett 4-vel számolunk:

$$\hat{y}_{t+i} = y_{t+i-4} + i \left(Me_{\alpha} \right)$$

4.2 Szabály.

Havi adatoknál:

$$\hat{y}_{t+i} = y_{t+i-12} \times \left(Me_{\beta} \right)^i$$

Negyedéves adatoknál értelemszerűen a 12 helyett 4-vel számolunk:

$$\hat{y}_{t+i} = y_{t+i-4} \times (Me_{\beta})^i$$

A naivmódszer-parancsfájl működése. (naivmodszert.xls)

Az i =az előrejelzés periódusa, a $szezon$ = a szezon hossza (maximum=12)

A naiv módszerek megbízhatóságának ellenőrzése.

A becslési időszak megadásával a rendelkezésre álló idősort két részre bontjuk, becslési és teszt időszakra. A becslési időszak felhasználásával a becslést végezzük el, a teszt időszak adatai alapján a becslések pontosságát lehet ellenőrizni, a MAPE hibaképlet felhasználásával. A MAPE %-os értékeit a naiv módszerek szabályai (1.1, 1.2,.....4.2) szerint közli a program. Az idősor hosszát felismeri a program, a sárga mezőbe (Ebből becslésre felhasznált:) be kell írni a becslésre felhasznált idősor hosszát, ami az idősor fele vagy annál nagyobb érték. Ha tesztelni kívánjuk a naiv módszereket, akkor a becslésre felhasznált idősor hosszának legnagyobb értéke az idősor hossza mínusz egy. Az idősort és a naiv módszerekkel elvégzett becslések ábráit az Ábra1 és Ábra2 munkafüzetben közli a program. Az Ábra1 azoknak a szabályoknak az ábráit rajzolja meg, ahol nincs szezonáltság (1.1...2.4), az Ábra2 pedig azokat az ábrákat közli, ahol van szezonáltság (3.1, 4.1, 4.2). A hibaképletek munkalapon egy gördülő menüben megtalálható az egyes naiv prognózis módszerek hibái¹¹⁴, azokban az esetekben, ha nincs szezonáltság (1.1.....2.4). Az autokorreláció eredményét is közli a program.

Gyakorló feladatok. (naivmodszert.xls)

Az USA lakossági teljes villamosenergia fogyasztás (milliárd kWh/hónap) idősora 1973 január és 2007 november között rendelkezésünkre áll.¹¹⁵ Végezze el a számításokat a naivmodszert.xls parancsfájllal.

3.6 Az exponenciális kiegyenlítés módszere (simit.xls és EXPS for Windows)*

3.6.1 A simit.xls parancsfájl működése.

Az exponenciális kiegyenlítés^{116,117} a számtani illetve a mozgó átlagolás továbbfejlesztett változata. Kiküszöböli a mozgó átlagolás azon hibáját, hogy az minden adatot azonos súllyal vesz figyelembe. A legújabb adatok általában nagyobb szerepet játszanak a jövőbeli adatok alakulásában, mint a régebbi adatok. Ezért a legfrissebb adatoknak a megelőzőnél relatíve nagyobb súlyt kell adnunk. Előnye, hogy egyszerűen kezelhető, nem kíván nagyobb matematikai elmélyülést. Az alkalmazhatóság feltételeit is meg kell említeni: megfelelő hosszúságú idősor [20-90] szükséges a kiegyenlítéshez.

Az alapképlet szerint a legfrissebb érték α -súllyal, míg az összes korábban megfigyelt érték $(1-\alpha)$ súllyal járul hozzá a becsült érték kialakításához, így az α becslése alapvető fontosságú.

Az előrejelzés:

$$F_{t+1} = \alpha X_t + (1-\alpha) F_t$$

ahol:

- F = az exponenciális kiegyenlítéssel nyert érték
- X = megfigyelt érték
- t = időszak
- α = reakcióparaméter, $0 \leq \alpha \leq 1$

Előrejelzés időszaka: m .

¹¹⁴ Ld. Az előrejelzések hibáinak a mérése.

¹¹⁵ <http://www.eia.doe.gov/emeu/mer/elect.html>

¹¹⁶ S. Makridakis-S. C. Wheelwright- V. Mcgee [1983] 84-123. felhasználásával.

¹¹⁷ Ld.: Kiss Tibor – Sipos Béla: ExpS for Windows 1995. szoftver 12 exponenciális simítási módszerrel végez számításokat iterációs eljárással. A szoftver a PTE KTK géptermeiben használható. A simit.xls ugyanezen műveleteket végzi el, de csak 4 simítási módszer becslését dolgoztuk ki. Az ExpS for Windows leírását és alkalmazását lásd: Kiss Tibor - Sipos Béla - Szentmiklósi Miklós [1995] és Kiss Tibor – Sipos Béla [2000].

A kezdő simított értéket meg kell adni (inicializálás) ami általában, az idősor első értéke. Az első simított érték lehet továbbá az első néhány adat valamilyen átlaga, vagy egyéb becsléssel nyert érték.

Legyen:

$$F_0 = X_0$$

Ez utóbbi képletet a számítások során célszerű használni. Ennek alapján, ha:

$$F_0 = X_0$$

$$F_1 = \alpha X_1 + (1-\alpha)X_0$$

$$F_2 = \alpha X_2 + (1-\alpha)F_1 = \alpha X_2 + (1-\alpha)[\alpha X_1 + (1-\alpha)X_0] = \alpha X_2 + \alpha(1-\alpha)X_1 + (1-\alpha)^2 X_0$$

$$F_3 = \alpha X_3 + (1-\alpha)F_2 = \alpha X_3 + (1-\alpha)[\alpha X_2 + \alpha(1-\alpha)X_1 + (1-\alpha)^2 X_0] =$$

$$= \alpha X_3 + \alpha(1-\alpha)X_2 + \alpha(1-\alpha)^2 X_1 + (1-\alpha)^3 X_0$$

$$F_t = \left[\sum_{i=0}^{t-1} \alpha(1-\alpha)^i X_{t-i} \right] + (1-\alpha)^t X_0$$

Látható, hogy az új érték kialakításában a legutolsó ($X_{t,0} = X_t$) tehát t időpontban megfigyelt adat α súllyal, az azt megelőző, $(t-1)$ időpontban megfigyelt (X_{t-1}) adat $(1-\alpha)$, míg az i -edik időszakot megelőző adat ($X_{t-(t-1)}$), $(1-\alpha)^i$ súllyal járul hozzá az F_t érték kiszámításához. Az idősor legrégebbi adatának ($X_{t-t} = X_0$) súlya pedig $(1-\alpha)^t$. Ha $\alpha=0$ akkor a legrégebbi adat súlya 1.

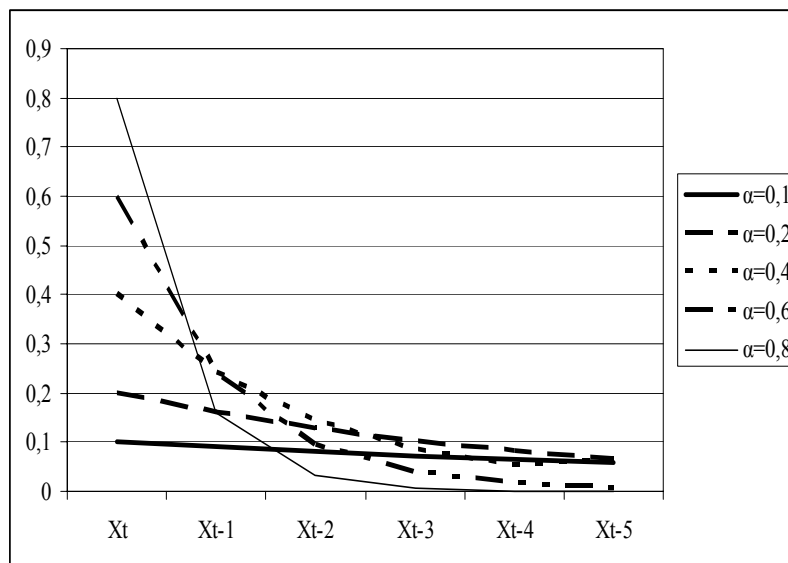
A súlyok geometriai sorozat szerint csökkennek, ahol a két szomszédos tag hányadosa: $q=(1-\alpha)$ és a sorozat első tagja $q_0=\alpha$ -val. A súlyok összege így 1-t ad, vagyis a várható értéket határoztuk meg. A geometriai sorozat összegképletét (s) felhasználva ugyanis:

$$s = q_0 \left(\frac{q^n - 1}{q - 1} \right) = \alpha \frac{(1-\alpha)^t - 1}{(1-\alpha) - 1} + (1-\alpha)^t = 1$$

A súlyok alakulását mutatja az alábbi 3-2. tábla és 3-21. ábra, ha $\alpha = 0,1, 0,2, 0,4, 0,6, 0,8$.

3-2. tábla: Az α súlyok változása

Súlyok	$\alpha=0,1$	$\alpha=0,2$	$\alpha=0,4$	$\alpha=0,6$	$\alpha=0,8$
X_t	0,1	0,2	0,4	0,6	0,8
X_{t-1}	0,09	0,16	0,24	0,24	0,16
X_{t-2}	0,081	0,128	0,144	0,096	0,032
X_{t-3}	0,0729	0,1024	0,0864	0,0384	0,0064
X_{t-4}	0,06561	0,08192	0,05184	0,01536	0,00128
X_{t-5}	0,059049	0,065536	0,065536	0,006144	0,000256



3-21. ábra: Az α súlyok változása

Ha nagy α -t választunk [pl. $\alpha = 0,8$], akkor az utolsó év adatai nagy súlyt kapnak [az utolsó évé 0,8, előtte levőé $0,8[1-0,8] = 0,16$ majd: 0,032, 0,0064, 0,00128 stb.]. Ha a véletlen erősen befolyásolja ezek nagyságát, az előrejelzés kevésbé megbízható lesz. Ha kis reakcióparamétert választunk [pl.: $\alpha = 0,2$], akkor a régebbi adatoknak is nagyobb szerepe lesz (a súlyok: 0,2, 0,2 $[1-0,2]=0,16$, majd: 0,128, 0,1024 stb.), a véletlen hatásokat jobban kiküszöbölhetjük. A legjobb reakcióparamétert kísérletezéssel lehet meghatározni felhasználva a teszt időszak adta lehetőségeket. A megfigyelt időszakot két részre bontjuk, az első időszak alapján becsülünk illetve a második teszt időszakra elvégzett becsülés alapján (felhasználva a hibaképleteket pl. a MAPE-t) választjuk ki a legjobb módszert, α -t kezdőértéket stb.

A következő exponenciális simítási módszereket programoztuk:

Az első módszer: SES¹¹⁸ normál exponenciális simítás, amelynek alapegyenlete,

$$F_{t+1} = \alpha X_t + (1 - \alpha) F_t$$

A becsléshez az adatok rendelkezésre állnak, kivéve az első becslést:

$$F_1 = \alpha X_0 + (1 - \alpha) F_0$$

Az egyik megoldás az, hogy

$$F_1 = X_1$$

A másik módszer, hogy a felhasználó adja meg az első értéket, pl. az első néhány adat átlagát veszi.

A normál exponenciális simítás (SES) módszerét akkor alkalmazzuk, ha sem trend, sem szezonális hatás nincs az idősorban.

A második módszer: kétszeres simítás, a Brown egyparaméteres lineáris módszere. Kétszeres exponenciális kiegyenlítés, az egyszer már kiegyenlített értékeket, S_t^1 még egyszer kiegyenlítjük S_t^2 , mivel lineáris trendet feltételezünk az idősorban:

$$S_t^1 = \alpha X_t + (1 - \alpha) S_{t-1}^1$$

$$S_t^2 = \alpha S_{t-1}^1 + (1 - \alpha) S_{t-1}^2$$

$$a_t = 2S_t^1 - S_t^2$$

$$b_t = \frac{\alpha}{1 - \alpha} (S_t^1 - S_t^2)$$

$$F_{t+m} = a_t + b_t m$$

A Brown egyparaméteres lineáris módszerét akkor használjuk, ha lineáris trendhatás van az idősorban, viszont szezonális hatás nincs az idősorban.

A harmadik módszer: háromszoros simítás, Brown egyparaméteres kvadratikusan módszere. Háromszoros exponenciális kiegyenlítés, az egyszer már kiegyenlített értékeket, S_t^1 még egyszer kiegyenlítjük S_t^2 , majd a kétszeresen kiegyenlített idősort még egyszer, tehát harmadszor is S_t^3 kiegyenlítjük, mivel másodfokú parabolikus (kvadratikusan) trendet feltételezünk az idősorban:

¹¹⁸ SES: Single Exponential Smoothing.

$$\begin{aligned}
S_t^1 &= \alpha X_t + (1 - \alpha)S_{t-1}^1 \\
S_t^2 &= \alpha S_t^1 + (1 - \alpha)S_{t-1}^2 \\
S_t^3 &= \alpha S_t^2 + (1 - \alpha)S_{t-1}^3 \\
a_t &= 3S_t^1 - 3S_t^2 + S_t^3 \\
b_t &= \frac{\alpha}{2(1 - \alpha)^2} [(6 - 5\alpha)S_t^1 - (10 - 8\alpha)S_t^2 + (4 - 3\alpha)S_t^3] \\
c_t &= \frac{\alpha^2}{(1 - \alpha)^2} (S_t^1 - 2S_t^2 + S_t^3)
\end{aligned}$$

$$F_{t+m} = a_t + b_t m + \frac{1}{2} c_t m^2$$

A Brown egyparaméteres kvadratikus módszerét akkor használjuk, ha másodfokú parabolikus trendhatás van az idősrőban, viszont szezonális hatás nincs az idősrőban.

A negyedik módszer: ARRSSES¹¹⁹ adaptív reagálású egyszerű exponenciális simítási módszer. Ennél a módszernél α_t értéke változik periódusról peridusra, amint az adatséma [minta, pattern] változik. A módszer alapegyenlete

$$F_{t+1} = \alpha_t X_t + (1 - \alpha_t) F_t$$

Ahol:

$$\begin{aligned}
\alpha_{t+1} &= \frac{|E_t|}{M_t} \\
E_t &= \beta e_t + (1 - \beta) E_{t-1} \\
M_t &= \beta |e_t| + (1 - \beta) M_{t-1} \\
e_t &= X_t - F_t
\end{aligned}$$

Az α és β 0 és 1 közé esik, e_t a hiba E_t a simítási hiba M_t az abszolút simítási hiba értéke.

Az adaptív reagálású egyszerű exponenciális simítási módszerét akkor alkalmazzuk, ha több száz vagy több ezer hasonló adatsort kell feldolgozni és az α érték automatikusan változik, ahogy változik az adat séma.

Tanácsok a gyakorlati alkalmazáshoz.

Az igazán megalapozott rövidtávú előrejelzés készítéséhez az idősrő hossza lehetőleg legyen legalább hat év, ha negyedéves vagy havi adatokkal rendelkezünk. Az első három év adata lehet a számítási időszak, a második három év adata pedig a tesztperiódus. Ebben az esetben ha van szezonális az idősrőban, akkor a szezonális becslésére három a tesztelésére szintén három év áll rendelkezésre. Ha negyedéves adatokkal dolgozunk, akkor ez azt jelenti, hogy $6 \cdot 4 = 24$ megfigyelt adatra van szükség, ha pedig havi adatokat használnánk, akkor $6 \cdot 12 = 72$.¹²⁰ Ez utóbbi esetben a tesztperiódus kezdete $72/2 + 1$, azaz 37, míg a negyedéves adatok esetében a tesztperiódus kezdete $24/2 + 1$, azaz 13. időpontban történhet. Keresést kérve, a simítási paraméterek értékét az Excel program úgy számítja ki, hogy megkeresi azt a paraméterkombinációt, ahol a hiba (MAPE) a legkisebb.

Érzékenységvizsgálatok.

Az alkalmazott módszerek stabilitását ellenőrizni kell. Ezt úgy végezhetjük el, hogy a tesztperiódus kezdetét változtatjuk, pl. az idősrő $3/4$ -nél határozzuk meg, vagy az utolsó év adatait tekintjük tesztperiódusnak. Negyedéves adatok esetén az előző példát folytatva a tesztperiódus kezdete az alábbi lehet. $6 \cdot 4 = 24$ megfigyelés esetén a háromnegyed-idősrő utáni első negyedév a 19. negyedév; az utolsó évtől kezdődő tesztperiódus esetén a 21. negyedév. Ha ugyanazt a módszert választja ki legjobbnak a program, és a paraméterek sincsenek nagyon távol egymástól (az eltérés a 10-20%-ot nem haladja meg) akkor az idősrő

¹¹⁹ ARRSSES: Adaptive Response Rate Single Exponential Smoothing.

¹²⁰ A megadott adatszám a minimálisan elvárható, annál jobb, minél több adat áll rendelkezésre.

stabilnak tekinthető és a módszer elemzésre és feltételezhetően előrejelzésre is hatékonyan alkalmazható. Ha a tesztperiódus kezdetének változtatásával a MAPE statisztika alapján kiválasztott „legjobb” módszer típusa is változik, akkor az idősor nem stabil, az idősorban vizsgált komponensek (trend, szezonális) tendenciája változik.

Gyakorló feladatok. (simit.xls)

Az USA lakossági teljes villamosenergia fogyasztás (milliárd kWh/hónap) idősora 1973 január és 2010 november között rendelkezésünkre áll. Végezze el a számításokat a simit.xls parancsfájllal.

3.6.2 Az EXPS for WINDOWS szoftver működése

Windows Változat^{121 122}

INSTALL.EXE Object Vision program installálása után az ExpsW könyvtárat be kell másolni, pl:

C:/VISIONR/EXPS

C:/EXPS

D://EXPS

Expswm.ovd	- Főprogram (Enter és indul a program)
Expswm.exe	- Segédprogram
ener.dta	- Példaállomány

Az új adatokat célszerű a megfelelő C:/VISIONR/EXPS (C:/EXPS D:/EXPS) könyvtárba másolni.

Adatállomány:

Elő kell készíteni azt az adatállományt, amely tartalmazza a vizsgálandó adatsort. Ennek egyszerű szöveges adatállománynak (ASCII) kell lennie, csaknem tetszőleges formátumban, szóközzel elválasztva egymástól. Bármely szövegszerkesztővel, illetve a legtöbb adatbázis kezelővel előállítható ilyen állomány a már meglévő adatokból is. Amennyiben az eddigi adatállomány Mátrix formában van megadva, ahol a változók az egyes oszlopok, az adatsorok akkor is elemezhetők a program segítségével. Az adatsor neve maximum 8 karakter legyen, kiterjesztése célszerűen *.dta.

Az oszlopnak ne legyen neve az első cellában, csak az idősort használjuk. Mentsük le módosított mentéssel: Formázott szöveg (szóközzel tagolt, kiterjesztés: *.prn) formátumban. Input adatállomány neve: *.prn vagy *.dta beírva felismeri az adatállományt, ellenőrizni lehet a Bemenő adatokra klikkelve. A vesszőket cseréljük ki pontokra a TOTAL COMMANDER vagy szövegszerkesztő használatával.

Ha a 12-ik módszernél kiakad a program, adjunk kezdő értéket, pl. az idősor első adatát induló értéknek.

Irodalom:

1. Kiss Tibor - Sipos Béla - Szentmiklósi Miklós [1995]: Az üzleti ciklus modellezése és prognosztizálása EXPS- programmal. Statisztikai Szemle. A Központi Statisztikai Hivatal folyóirata. 73. évf. 8 - 9. sz. 681 - 698. old.

http://www.ksh.hu/statszemle_archive/viewer.html?ev=1995&szam=08-09&old=59&lap=18

2. Béla Sipos- Tibor Kiss [2000]: EXPS for Windows, a software application. Hungarian Statistical Review. 78. Volume. Special Number. 146 - 164. old.

Az előrejelzés nemzetközi és hazai irodalmában számos, többé kevésbé kifinomult, matematika-igényes statisztikai módszer létezik. Az exponenciális simítás még mindig nagyon népszerű. Világszerte ismert és hatékony, különösen rövid távú előrejelzésekre. Más módszerekkel összehasonlítva - mint például a Box-Jenkins modellek - gyakran jobbnak bizonyulnak. Gyakorlati használhatóságukat a viszonylag egyszerű

¹²¹ Kiss Tibor-Sipos Béla: EXPS JPTE. 1998.

¹²² Spyros Makridakis - Steven C. Wheelwright - Victor E. McGee: Forecasting. Methods and Applications. 2. Ed. John Wiley & Sons, New York - Chichester - Brisbane - Toronto - Singapore, 1983. felhasználásával.

matematikai képletek és a számítógépes feldolgozáshoz, az adatok tárolásához biztosított kedvező feltételek segítik. További előny, hogy, kevés az elméleti feltevés és a frisebb adatoknak nagyobb súlyt lehet adni. Az induló paraméterek döntő fontosságúak lehetnek. Gyakran jelentős mértékben változtathatják az eredményeket mind pozitív, mind negatív irányban. A számítógépek rövid idő alatt sok számításra képesek elvégezni, ezért lehetségessé válik egy optimális megoldás kikísérletezése. Egy megfelelő iterációs eljárás segítségével ki lehet választani a legjobban illeszkedő becslést, azonban meg kell adni a lehetőségét tesztölges paraméterek megadásának is. Az ExpS, exponenciális simítási program kísérletet tesz arra, hogy mindét problémára megoldást adjon. Egy megfelelő iterációs eljárás segítségével lehetővé válik tizenkét módszer közül a legjobbnak a kiválasztása. Az adott módszeren belül ezután lehetővé válik a legjobb paraméter-együttes meghatározása.

A program futtatása:

A program futtatása a következő parancssorral történik:

exps Fnev

ahol az "Fnev" az adatbázis neve

Számos lehetőség van a program paraméterezésére.

[/o nev] [/k KE] [/a Alfa] [/lIL][e] [/t típus][p periódus]

[/m periódus] [/1 1. paraméter] [/2 2. paraméter] [/f oszlop]

ahol :

/o nev = Eredmények, alapértelmezés: *.out

/k KE = Kezdő érték (F1), máskülönben kiszámolja

/a Alfa = Alfa, máskülönben kiszámolja

/l IL = Az U statisztika javulásának határa, %-ban
alapértelmezés:0.01

/h IH = Az iterációk számának felső határa.
alapértelmezés: 30

/e Az egyedüli output a képernyőre érkezik.
alapértelmezés: Output állomány.

/p A teszt periódus kezdete
alapértelmezés: 10, ha az esetek száma nem kevesebb mint 11.

/1 Az első, nem alfa paraméter értéke. (szezonális). p2
alapértelmezés: 0.1, 0.15, 0.2

/2 A második, nem alfa paraméter értéke. (trend). p1
alapértelmezés: 0.1, 0.15, 0.2

/m Az előrejelzett értékek száma
alapértelmezés: 1, vagy a szezonbeli periódusok száma

/f Az adatbázis neve, ha az mátrixalakban van (szöveges file)
oszlop = a változó sorszáma a mátrixon belül
alapértelmezés: Csak egy adatsor van

/s A paraméterek skálája (szezonális és trend), amelyek részt vesznek az iterációban.

1 : (alapértelmezés) : 0.1, 0.15, 0.2

2 : 19-es skála, 0.05-től 0.95-ig.

3 : 190-es skála, 0.005-től 0.95-ig.

/c A szezonális feltételezett esetszámok (0, 4, 5, 7, 12, 24)
alapértelmezés : Automatikus keresés

/t A simítás típusa

A simítás típusait az alábbi Táblázat tartalmazza

Az exponenciális simítás alkalmazott típusai

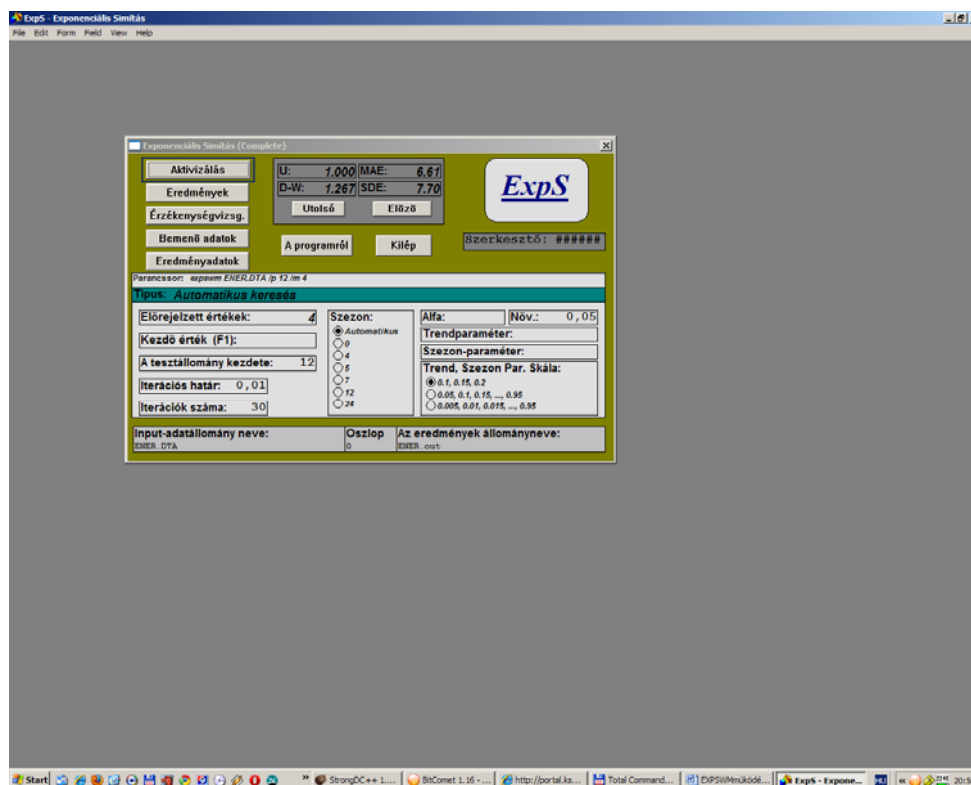
T- 1 Egyszerű exponenciális simítás

T- 2 Szezonális - additív, Trend nincs

T- 3 Szezonális - multiplikatív, Trend nincs

- T- 4 Szezonalitás - nincs, Trend additív (Holt módszere)
- T- 5 Szezonalitás - additív, Trend additív
- T- 6 Szezonalitás - multiplikatív, Trend additív (Winters módszere)
- T- 7 Szezonalitás - nincs, Trend multiplikatív
- T- 8 Szezonalitás - additív, Trend multiplikatív
- T- 9 Szezonalitás - multiplikatív, Trend multiplikatív
- T-10 Adaptív Reagálású Módszer (ARRSES)
- T-11 Brown Egy-Paraméteres Lineáris módszere
- T-12 Brown Egy-Paraméteres Quadratikus módszere

Az ExpS program főképernyője



A főképernyő alsó részén vannak a megfelelő paraméterek, melyeket ki lehet tölteni. Az “*Input adatállomány neve*” mezőt feltétlenül ki kell tölteni, mivel ezzel történik meg az adatállományra való hivatkozás. A program ezután fogja megtalálni a fenti “*Bemenő adatok*” gombbal az input adatokat. Ez egy ellenőrzési lehetőség is arra, hogy az adatállomány nevét jól adtuk - e meg. Az *Eredmények állományneve* mező automatikusan kitöltődik az állománynév alapján, amely “.out” végződést kap. Ez a mező átírható, módosítható, amennyiben az adott adatsor több eredményét is szeretnénk lementeni. Az eredmények szövegfájlban lementhetők.

Az *Aktivizálás* gomb elindítja a segédprogramot, amely a beállított paramétereknek megfelelően létrehozza az eredményeket.

Ezek az eredmények az *Eredmények* gombra való rákattintással tekinthetők meg. Az *Érzékenységvizsgálat* rész gyorsabb munkát tesz lehetővé a megfelelő modell kialakítása során.

Ezután lehet állítani a megfelelő paramétereket.

A *Szezon* mezőnél a képernyő középső, alsó részén lehet állítani a szezonitás esetszámát (a mozgóátlag tagszámát). *Automatikus keresés* esetén 4-től 24-ig a feltüntetett értékeknek megfelelően keresi az optimális mozgóátlag tagszámot. Amennyiben tudjuk hogy nincs az adatsorban szezonitás, úgy ez az érték 0-ra

állítható. Ha nem akarunk kísérletezni a szezon tagszámmal, akkor azonnal beállítható egy konkrét értékre, pl. negyedéves adatok esetén 4-re.

A *Kezdő érték*, *Alfa*, *Szezon* és *Trend paramétere*knél amennyiben nem adunk meg konkrét értéket, úgy a program iterál egyet. Az *Alfa* utáni *Növ.* (növekmény) mező tartalmazza azt az értéket, hogy az alfa-érték milyen lépésközzel menjen 0-tól 1-ig. Az alapértelmezés 0,05.

A *Trend*, *Szezonparaméter-skála* azt jelzi, hogy a trend- és szezon-paraméterek milyen értékeket vesznek fel az iteráció során, ha nem adunk meg konkrét értéket a trend és szezon paramétereire.

Az eredményeknél *par1* (*p1*) a trend paraméter, *par2* (*p2*) szezonparaméter értéke.

A képernyő közepén lévő sor nem más, mint az a parancssor, melyet a DOS változatban kellene kiadni, hogy a megfelelő paraméterezést hajtsuk végre.

A *Szerkesztő* mező tartalmazza a szövegszerkesztő nevét, amellyel az eredményeket illetve az adatokat előhívjuk. Amennyiben nagy az adatállomány és az alapértelmezés szerinti *Notepad* nem alkalmas erre a célra, úgy váltsunk szerkesztőt, pl. használjuk a *WINWORD*-t. A bemenő adatok szerkesztésénél vigyázzunk, hogy az adatok mentése mindig szöveges állományba történjen!

A Windows változat jellegzetessége, hogy a felhasználó követni tudja a négy, talán legfontosabb statisztikának (*U* statisztika, *D-W* statisztika, *MAE* és *SDE*, lásd az *Elmélet* részt) alakulását. Az *Utolsó* és az *Előző* gomb mindig cseréli az utóbbi (legutolsó modell) és az előző (az előző modell) eredményeket, hogy lehessen követni a változást. Amennyiben előzőleg más adatsorra kerestünk modellt, úgy az *Előző* gomb annak az eredményeit fogja mutatni!

A felhasznált módszerek algoritmusai:

Az egyes módszerek ($t=1, 2, \dots, 12$) egyenletei a következők:

Az exponenciális kiegyenlítés általános alakja:

$$S_t = \alpha P + (1 - \alpha)Q$$

Ahol:

Q = a trend

P = a szezonális tényező

Q és P a trend és a szezonális típusa szerint változik.

Ezt mutatja az alábbi táblázat:

A szezonális és a trend összefüggései.

	Szezonális nincs	Szezonális additív	Szezonális multiplikatív
Trend nincs	$P=X_t$ $Q=S_{t-1}$	$P=X_t - C_{t-L}$ $Q=S_{t-1}$	$P=X_t / D_{t-L}$ $Q=S_{t-1}$
Trend additív	$P=X_t$ $Q=S_{t-1} + A_{t-1}$	$P=X_t - C_{t-L}$ $Q=S_{t-1} + A_{t-1}$	$P=X_t / D_{t-L}$ $Q=S_{t-1} + A_{t-1}$
Trend multiplikatív	$P=X_t$ $Q=S_{t-1} B_{t-1}$	$P=X_t - C_{t-L}$ $Q=S_{t-1} B_{t-1}$	$P=X_t / D_{t-L}$ $Q=S_{t-1} B_{t-1}$

ahol:

X_t = megfigyelt (tényleges) adat

S_t = simított adat:

$$S_t = \alpha P + (1 - \alpha)Q$$

A Q (trend) és a P (szezonális) helyébe az alábbi egyenleteket lehet helyettesíteni:

Additív trend:

$$A_t = \beta(S_t - S_{t-1}) + (1 - \beta)A_{t-1}$$

Multiplikatív trend:

$$B_t = \gamma(S_t / S_{t-1}) + (1 - \gamma)B_{t-1}$$

Additív szezonális:

$$C_t = \delta(X_t - S_t) + (1 - \delta)C_{t-L}$$

Multiplikatív szezonalitás:

$$D_t = \theta(X_t/S_t) + (1-\theta)D_{t-L}$$

L = a szezonaritás periódusának a hossza, pl. 4 negyedéves adatok esetében, míg havi adatoknál 12.

Az $\alpha, \beta, \gamma, \delta, \theta$ (alfa, béta, gamma, delta, théta) paraméterek 0 és 1 közé esnek.

A következő táblázat a prognózist (F_{t+m}) mutatja m időszakra előre

A prognózis képletei.

Trend	Szezonaritás		
	nincs	additív	multiplikatív
nincs	$F_{t+m} = S_t$	$F_{t+m} = S_t + C_{t-L+m}$	$F_{t+m} = S_t D_{t-L+m}$
additív	$F_{t+m} = S_t + mA_t$	$F_{t+m} = S_t + mA_t + C_{t-L+m}$	$F_{t+m} = (S_t + mA_t)D_{t-L+m}$
multiplikatív	$F_{t+m} = S_t B_t^m$	$F_{t+m} = S_t B_t^m + C_{t-L+m}$	$F_{t+m} = S_t D_{t-L+m} B_t^m$

Az előző két táblázat jelöléseit felhasználva, a P és Q helyébe behelyettesítve az egyenleteket 9 simítási eljárás állítható elő. Ezen kívül még 3 módszer alkalmazását ismertetjük.

Az első módszer: normál exponenciális simítás, SES.¹²³

$$F_{t+1} = \alpha X_t + (1-\alpha)F_t$$

A becsléshez az adatok rendelkezésre állnak, kivéve az első becslést:

$$F_1 = \alpha X_0 + (1-\alpha)F_0$$

Az egyik megoldás az, hogy

$$F_1 = X_1$$

vagy a táblázat szerinti formában:

$$S_t = \alpha X_t + (1-\alpha)S_{t-1}$$

A másik módszer, hogy a felhasználó adja meg az első értéket, pl. az első néhány adat átlagát veszi. A harmadik megoldás: az ExpsW program megkeresi a legjobb becslést adó kezdő értéket, a hibastatisztikák (U-statisztika) alapján. A normál exponenciális simítás (SES) módszerét akkor alkalmazzuk, ha sem trend, sem szezonális hatás nincs az idősorban.

Prognózis m időszakra előre:

$$F_{t+m} = S_t$$

A második módszer: additív szezonaritás, trend nincs:

$$S_t = \alpha(X_t - C_{t-L}) + (1-\alpha)S_{t-1}$$

$$C_t = \delta(X_t - S_t) + (1-\delta)C_{t-L}$$

Előrejelzés m időszakra előre:

$$F_{t+m} = S_t + C_{t-L+m}$$

A harmadik módszer: multiplikatív szezonaritás, trend nincs:

$$S_t = \alpha(X_t/D_{t-L}) + (1-\alpha)S_{t-1}$$

$$D_t = \theta(X_t/S_t) + (1-\theta)D_{t-L}$$

¹²³ SES: Single Exponential Smoothing.

Prognózis m periódusra előre:

$$F_{t+m} = S_t D_{t-L+m}$$

A negyedik módszer: szezonalitás nincs, trend additív (Holt módszere):

$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + A_{t-1})$$

$$A_t = \beta(S_t - S_{t-1}) + (1 - \beta)A_{t-1}$$

Előrejelzés m periódusra előre:

$$F_{t+m} = S_t + mA_t$$

Látható, hogy az eljárás azonos Holt módszerével, aki az additív trendet (A_t) b_t -vel jelölte és a β paramétert γ -val jelölte.

Holt, lineáris kétparaméteres módszere két reakcióparamétert [α és γ] alkalmaz, additív lineáris trendet a megfigyelési időszakra:

$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + b_{t-1})$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}$$

$$F_{t+m} = S_t + b_t m$$

A b_t meghatározására a lehetőségek:

$$b_1 = x_2 - x_1$$

$$b_1 = \frac{(x_2 - x_1) + (x_3 - x_2) + (x_4 - x_3)}{3}$$

Az ötödik módszer: szezonalitás additív, trend additív:

$$S_t = \alpha(X_t - C_{t-L}) + (1 - \alpha)(S_{t-1} + A_{t-1})$$

$$A_t = \beta(S_t - S_{t-1}) + (1 - \beta)A_{t-1}$$

$$C_t = \delta(X_t - S_t) + (1 - \delta)C_{t-L}$$

A prognózis m periódusra előre:

$$F_{t+m} = S_t + mA_t + C_{t-L+m}$$

A hatodik módszer: szezonalitás multiplikatív, trend additív (Winters módszere):

$$S_t = \alpha \frac{X_t}{I_{t-L}} + (1 - \alpha)(S_{t-1} + b_{t-1})$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}$$

$$I_t = \beta \frac{X_t}{S_t} + (1 - \beta)I_{t-L}$$

$$F_{t+m} = (S_t + b_t m)I_{t-L+m}$$

ahol b_t a trend, I a szezonalitást kiigazító faktor.

A becslés a táblázat alapján:

$$S_t = \alpha(X_t/D_{t-L}) + (1 - \alpha)(S_{t-1} + A_{t-1})$$

$$A_t = \beta(S_t - S_{t-1}) + (1 - \beta)A_{t-1}$$

$$D_t = \theta(X_t/S_t) + (1 - \theta)D_{t-L}$$

Előrejelzés m periódusra előre:

$$F_{t+m} = (S_t + mA_t)D_{t-L+m}$$

A módszer megegyezik Winters eljárásával. aki az additív trendet (A_t) b_t - vel jelölte és a szezonális D_t jelölése helyett az I_t jelölést alkalmazta.

A hetedik módszer: szezonális nincs, trend multiplikatív:

$$S_t = \alpha X_{t-1} + (1-\alpha)S_{t-1}B_{t-1}$$

$$B_t = \gamma(S_t/S_{t-1}) + (1-\gamma)B_{t-1}$$

A prognózis m periódusra előre:

$$F_{t+m} = S_t B_t^m$$

A nyolcadik módszer: szezonális additív, trend multiplikatív:

$$S_t = \alpha(X_t - C_{t-L}) + (1-\alpha)S_{t-1}B_{t-1}$$

$$B_t = \gamma(S_t/S_{t-1}) + (1-\gamma)B_{t-1}$$

$$C_t = \delta(X_t - S_t) + (1-\delta)C_{t-L}$$

Előrejelzés m periódusra előre:

$$F_{t+m} = S_t B_t^m + C_{t-L+m}$$

A kilencedik módszer (t=9) szezonális multiplikatív, trend multiplikatív:

$$S_t = \alpha(X_t/D_{t-L}) + (1-\alpha)S_t B_{t-1}$$

$$B_t = \gamma(S_t/S_{t-1}) + (1-\gamma)B_{t-1}$$

$$D_t = \theta(X_t/S_t) + (1-\theta)D_{t-L}$$

A prognózis m periódusra előre:

$$F_{t+m} = S_t D_{t-L+m} B_t^m$$

Ezzel a trend és szezonális típusok szerint különböző simítási módszereket áttekintettük.

A tízedik módszer: Adaptív Reagálású Egyszerű Exponenciális Simítási Módszer (ARRSES¹²⁴).

Ennél a módszernél α_t értéke változik periódusról peridusra, amint az adatséma [minta, pattern] változik.

A módszer alapegyenlete:

$$F_{t+1} = \alpha_t X_t + (1-\alpha_t)F_t$$

Ahol:

$$\alpha_{t+1} = \left| \frac{E_t}{M_t} \right|$$

$$E_t = \beta e_t + (1-\beta)E_{t-1}$$

$$M_t = \beta |e_t| + (1-\beta)M_{t-1}$$

$$e_t = X_t - F_t$$

¹²⁴ Adaptive-response-rate single exponential smoothing: ARRSES.

Az α és β 0 és 1 közé esik a $||$ az abszolút érték jele, e_t a hiba E_t a simítási hiba M_t az abszolút simítási hiba értéke.

A 11 módszer: a Brown egyparaméteres lineáris módszere.

Kétszeres exponenciális kiegyenlítés, az egyszer már kiegyenlített értékeket, S_t^1 még egyszer kiegyenlítjük S_t^2 , mivel lineáris trendet feltételezünk az idősorban:

$$S_t^1 = \alpha X_t + (1 - \alpha)S_{t-1}^1$$

$$S_t^2 = \alpha S_{t-1}^1 + (1 - \alpha)S_{t-1}^2$$

$$a_t = 2S_t^1 - S_t^2$$

$$b_t = \frac{\alpha}{1 - \alpha}(S_t^1 - S_t^2)$$

$$F_{t+m} = a_t + b_t m$$

A Brown egyparaméteres lineáris módszerét akkor használjuk, ha lineáris trendhatás van az idősorban, viszont szezonális hatás nincs az idősorban.

A 12. módszer: Brown egyparaméteres kvadratikus módszere.

Háromszoros exponenciális kiegyenlítés, az egyszer már kiegyenlített értékeket, S_t^1 még egyszer kiegyenlítjük S_t^2 , majd a kétszeresen kiegyenlített idősort még egyszer, tehát háromszor is S_t^3 kiegyenlítjük, mivel másodfokú parabolikus (kvadratikus) trendet feltételezünk az idősorban:

$$S_t^1 = \alpha X_t + (1 - \alpha)S_{t-1}^1$$

$$S_t^2 = \alpha S_{t-1}^1 + (1 - \alpha)S_{t-1}^2$$

$$S_t^3 = \alpha S_{t-1}^2 + (1 - \alpha)S_{t-1}^3$$

$$a_t = 3S_t^1 - 3S_t^2 + S_t^3$$

$$b_t = \frac{\alpha}{2(1 - \alpha)^2} [(6 - 5\alpha)S_t^1 - (10 - 8\alpha)S_t^2 + (4 - 3\alpha)S_t^3]$$

$$c_t = \frac{\alpha^2}{(1 - \alpha)^2} (S_t^1 - 2S_t^2 + S_t^3)$$

$$F_{t+m} = a_t + b_t m + 1/2 c_t m^2$$

A Brown egyparaméteres kvadratikus módszerét akkor használjuk, ha másodfokú parabolikus trendhatás van az idősorban, viszont szezonális hatás nincs az idősorban.

A kezdő értékek megadása, az inicializálás:

Az 1. és 10. módszernél:

$$F_1 = X_1$$

A 11. módszernél:

$$S_t^2 = S_t^1 = X_1$$

$$a_1 = X_1$$

$$b_1 = \frac{(X_2 - X_1) + (X_4 - X_3)}{2}$$

A 4. módszernél:

$$S_1 = X_1$$

$$b_1 = \frac{(X_2 - X_1) + (X_4 - X_3)}{2}$$

A 12. módszernél:

$$S_1^3 = S_1^2 = S_1^1 = X_1$$

$$a_1 = X_1$$

$$b_1 = \frac{(X_2 - X_1) + (X_3 - X_2) + (X_4 - X_3)}{3}$$

$$c_1 = \frac{(X_3 - X_1)}{2}$$

A szezonalitást tartalmazó modellek (Winters és 2. 3. 5. 6. 8. 9.) esetében:

$$S_{L+1} = X_{L+1}$$

Ahol:

L= a szezonaritás hossza (negyedéves adatok: L=4, havi adatok: L=12, tőzsdenapok: L=252)

$$I_1 = X_1 / \bar{X}$$

$$I_2 = X_2 / \bar{X}$$

$$I_3 = X_3 / \bar{X}$$

.

.

.

$$I_L = X_L / \bar{X}$$

Ahol:

$$\bar{X} = \sum_{i=1}^L \frac{X_i}{L}$$

$$b_{L+1} = \frac{(X_{L+1} - X_1) + (X_{L+2} - X_2) + (X_{L+3} - X_3)}{3(L)}$$

Tanácsok a gyakorlati alkalmazáshoz

Az igazán megalapozott rövidtávú előrejelzés készítéséhez az idősor hossza lehetőleg legyen hat év. Az első három év adata lehet a számítási időszak, a második három év adata pedig a tesztperiódus. Ebben az esetben a szezonaritás beclésére három a tesztelésére szintén három év áll rendelkezésre. Ha negyedéves adatokkal dolgozunk, akkor ez azt jelenti, hogy $6 \cdot 4 = 24$ megfigyelt adatra lenne szükség, ha pedig havi adatokat használnánk, akkor $6 \cdot 12 = 72$ megfigyelt adat az optimális. Ez utóbbi esetben a tesztperiódus kezdete $72/2 + 1$, azaz 37, míg a negyedéves adatok esetében a tesztperiódus kezdete $24/2 + 1$, azaz 13. időpontban történhet.

Az alkalmazott módszer mind a 12 esetben azt feltételezi, hogy a kimutatott összefüggés (trend, szezonaritás) stabil a megfigyelési és az előrejelzési időszakban. Ha pl. multiplikatív trendet és szezonaritást mutat ki az eljárás, illetve az ilyen modell beclésére alkalmas 9. módszer adja a legjobb közelítést, akkor ez azt jelenti, hogy az említett összefüggést a teljes megfigyelési és előrejelzési időszakra állandónak tekintjük. Ha tegyük fel ez a kapcsolat a tesztperiódus második felében megváltozik, pl. a multiplikatív trendkapcsolat additívra változik, akkor az előrejelzés bizonytalanává válik.

Érzékenységvizsgálatok.

Az előzőekben elmondottak következménye az, hogy az alkalmazott módszerek stabilitását ellenőrizni kell. Ezt úgy végezhetjük el, hogy a tesztperiódus kezdetét változtatjuk, pl. az idősor $\frac{3}{4}$ -nél határozzuk meg, vagy az utolsó év adatait tekintjük tesztperiódusnak.

Negyedéves adatok esetén az előző példát folytatva a tesztperiódus kezdete az alábbi lehet. $6 \cdot 4 = 24$ megfigyelés esetén a háromnegyed-időszak utáni első negyedév a 19. negyedév; az utolsó évtől kezdődő tesztperiódus esetén a 21. negyedév.

Automatikus keresést kérve, az alfa, a kezdő érték, a p_1 és p_2 szezon illetve trend paraméterek értékét meg nem adva újra futtatjuk a feladatokat. Ha ugyanazt a módszert választja ki legjobbnak a program, és a kezdő paraméterek sincsenek nagyon távol egymástól (az eltérés a 10-20 %-ot nem haladja meg) akkor az idősor stabilnak tekinthető és a módszer elemzésre és feltételezhetően előrejelzésre is hatékonyan alkalmazható. Ha a tesztperiódus kezdetének változtatásával az U-statisztika alapján kiválasztott "legjobb" módszer típusa is változik, akkor az idősor nem stabil, az idősorban vizsgált komponensek (trend, szezonális) tendenciája változik. Az elemzés így bizonytalan, és az előrejelzés sem lesz megbízható. Ez esetben megfelelő előrejelzési módszer lehet a CENSUS II program, amely alkalmas a változó szezonális és trend követésére - havi adatok esetén.

Az *Érzékenységvizsgálat* gombot lenyomva megkapjuk a program által készített érzékenységvizsgálatot. A legjobban illeszkedő módszerre számítja ki a 12 simítási eljárás közül a Theil féle U - statisztikát.

Első tesztperiódus kezdete: idősor fele.

Második tesztperiódus kezdete: idősor kétharmada.

Harmadik tesztperiódus kezdete: idősor négyötöde.

Ha a nincs nagy különbség a háromféle módon számolt U-statisztikák között, akkor a becslés (a legjobb paraméterkombináció és simítási módszer) stabil.

3.7 A SABL-módszer (szoftver) felhasználása adatelőkészítésre, a trend és a periodikus hullámzás szétválasztására *

A SABL^{125 126} eljárást a Bell Laboratórium munkatárai 1979-ben publikálták¹²⁷. A SABL eljárás a szezonális illetve periodikus idősorok simítását végzi el. Additív komponensekre bontja az eredeti vagy a transzformált adatokat: trend, szezonális [periodikus] és irreguláris [fehér zaj] összetevőket különböztet meg, rezisztens [ellenállóképes] lineáris vagy nemlineáris simítási módszerek alkalmazásával. A rezisztens fogalma ebben az összefüggésben azt a tulajdonságot jelenti, hogy a módszer nem érzékeny néhány adat kiugróan nagy eltérése [az úgynevezett outlier] által okozott erős zavaróhatásra, torzításra. Azokat az értékeket tekintjük szélsőséges, extrém értékeknek, outlier-eknek, amelyek nagyon távol vannak az eloszlás közepétől, jelentősen különböznek a többi értékektől. Elkülönítésük mind elemzési, mind előrejelzési illetve modellalkotási szempontból fontos feladat. Feltárásukat a grafikus ábrázolás is segíti. A szezonális illetve periodikus illesztés célja általában a múltbeli és a jelenbeli adatok szezonális illetve periodikus hullámzásának meghatározása előrejelzési célból. Fontos, például árpolitikai és beruházás-politikai célból, hogy az üzleti vállalat tisztában legyen azzal, változik-e az üzleti aktivitás adott időszakban és a

¹²⁵ SABL: Seasonal Adjustment – Bell Laboratories.

¹²⁶ Kiss Tibor – Kruzslíc Ferenc - Sipos Béla - Szentmihályi Miklós [1997]: 844 - 863.

¹²⁷ Cleveland, W. S.-Dunn, D. M.-Terpenning, I. J. [1979].

hullámzás kisebb, nagyobb, illetve szezonális jellegű-e. Az építőanyagipari, a divat, a mezőgazdasági termékeket feldolgozó, idegenforgalmi stb. vállalkozások esetében nagy jelentősége van, mert a szezonális erőteljesen érvényesül. A szezonális meghatározásának célja az ilyen fluktuáció eltávolítása az idő-sorból az alapul szolgáló trendhatás azonosítása érdekében. Számos módszer áll rendelkezésre a szezonális komponens azonosítására, a legtöbb azonban érzékeny a fent említett extrém értékek torzító hatására. A szezonális illesztés egyik célja olyan szezonális tényező elérése, amely stabil, azaz nem változik az időszakok folyamán. A SABL néven ismert eljárás¹²⁸ lényegét tekintve simítási módszereket alkalmaz, amihez mozgómediánokat, rezisztens mediánokat és átlagokat, valamint duplanégyzet becsléseket használ. A SABL módszer bármely idősorra alkalmazható amely tartalmaz periodikus ingadozást és trendet. A módszer alkalmazásához legalább három periódus megfigyelt adataira van szükség, napi, heti, havi, negyedéves, féléves vagy éves bontásban. Éves adatsorokat az évszázados trendnél rövidebb (pl. 3, 9, 18, 60 éves) konjunktúraciklusok kimutatására alkalmazunk. A SABL felhasználásával előállított trend és szezon (periodikus) összetevők felhasználásával megalapozottabb prognózisokat készíthetünk, mivel az outlierok zavaró hatását kiszűrtük. A kiszűrés illetve a becslés 45 lépésben, egy iterációs eljárás eredménye.

Az idősorok SABL dekompozícióján a következő felbontást értjük:

$$Y = T + S + I$$

ahol:

T képviseli a hosszú távú trendet,

S jelöli a szezonális komponenst, értelemszerűen a konjunkturális komponenst is, tehát a periodikus hullámzást,

I az irreguláris részt, azaz a fehér zaj.

Általában **Y** alatt az eredeti idősor adatait értjük, vagy annak transzformált [pl. ln-transzformáció] alakját. A SABL módszer egy iteratív eljárás. Minden iterációs lépésben újra számítjuk, finomítjuk a **T**, **S**, **I** értékeket. A lépések során kialakult adatsorok kölcsönhatásban vannak egymással.

A módszer lényegéből fakadóan rendelkezik az alábbi tulajdonságokkal:

A **T**, **S** és **I** komponensek közötti adatáthatás minimális. Extrém vagy szokatlan adatok nem befolyásolják a **T** és **S** értékek meghatározását illetve becslését. Az ilyen szokatlan adatok hatása csak az **I** komponensben tükröződik. A **T** és **S** becslése amennyire lehetséges érzékenyen reagál az idősor szerkezetének változására. A módszer csak akkor használható, ha legalább 3 periódus adatai ismertek. Sok helyen a zérussal való osztás elkerülése érdekében és a transzformációs szabály alkalmazhatósága miatt ki kell kötni, hogy az eredeti adatok csak pozitívak lehetnek.

A SABL eljárás első lépése, hogy az eredeti adatsort transzformációnak vetjük alá. A transzformáció célja, hogy: minimalizálja a szezonális amplitúdójának függését a trendkomponens szintjétől, egyszerűbbé tegye a szezonális illesztést, lényegében az idősor komponenseit additívvá alakítsa.

Az Y^p adatsor transzformáltját **D[0]**-al jelöljük, a transzformációs függvényt egy **p** paraméter értékének megadásával választhatjuk ki. A **p** paraméter az idősor auditivitásának mértékét mutatja, szokásos értékei: -1 ; $-0,5$; $-0,25$; 0 ; $0,25$; $0,5$; 1 ; A transzformációs függvény az alábbi:

$$D[0] = Y^p \quad p > 0$$

$$D[0] = \lg Y \quad p = 0$$

$$D[0] = -Y^p \quad p < 0$$

A **p** paraméter értékének kiválasztásához meg kell határozni az idősor komponensei: **Y**, **T**, és **S** közötti kapcsolat típusát. Additív kapcsolatot esetén az $Y = T + S$ összefüggés érvényes, a multiplikatív kapcsolat az $Y = T \cdot S$ formulával jellemezhető. A multiplikatív kapcsolat logaritmus segítségével additívvá alakítható, azaz $p=0$ választásával. Ekkor ugyanis az $\lg Y = \lg T + \lg S$ formához jutunk. $p=1$ esetén az adatsor nem változik, $p=0,5$ esetén az eredeti sor négyzetgyökét, $p=0,33$ esetén a köbgyökét kapjuk. Tehát ha $p=0,33$ akkor harmadfokú, ha $p=0,5$ akkor másodfokú parabolikus trendet feltételezünk. Ugyanis: az **S** és **I** komponens várható értéke éves szinten 0.

¹²⁸ Levenbach, H. - Cleary, J. P. [1982]: 248-274.

Ha $p=0,5$

$$Y^2 = (T + S + I)^2 = T^2$$

$$Y = (Y^2)^{0,5} = ((T + S + I)^2)^{0,5} = (T^2)^{0,5} = T$$

Ha $p=0,33$

$$Y^3 = (T + S + I)^3 = T^3$$

$$Y = (Y^3)^{0,33} = ((T + S + I)^3)^{0,33} = (T^3)^{0,33} = T$$

Ha az eredeti adatok [pl. relatív növekedési ütem] negatív, akkor a p is negatív ($p < 0$), így a $D[0]$ pozitív lesz.

Először simítási eljárást alkalmazunk az első rezisztens simított trend meghatározásának érdekében. A simított trendet kivonjuk az adatokból: $[D-T] = [S+I]$ alapján így kapjuk az $[S+I]$ sort. Ezután az $[S+I]$ csökkenő súlyozású [tapered] mozgómediánjainak és mozgóátlagainak számítása következik, így nyerjük a kezdeti rezisztens szezonális komponensét. Ezután a kapott kezdeti szezonális komponens a [már transzformált] eredeti adatsorból történő kivonása után a második simított trendet határozzuk meg. Ezt követi az irreguláris komponens számítása: $[D-T-S=I]$.

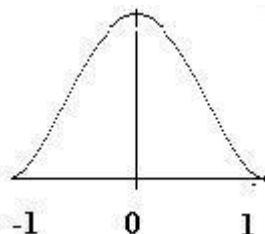
A SABL eljárás vázlata ¹²⁹

A csökkenő súlyozású [tapered] mozgóátlag és mozgómedián számítása során használt súlyok meghatározásához a súlyok olyan sorozatára van szükségünk, ahol a súlyok az adott időszaktól [hónaptól, évtől, naptól, stb. attól függően, hogy milyen adatokkal dolgozunk] való távolság arányában csökkennek. Ilyen tulajdonsággal rendelkezik a duplanégyzet [bisquare] függvény:

$$B(u) = (1 - u^2)^2 \quad |u| \leq 1$$

$$B(u) = 0 \quad |u| > 1$$

A $B[u]$ függvény ábrája a következőképpen néz ki:



3-22. ábra: Duplanégyzet függvény

Tételezzük fel, hogy havi adatokkal rendelkezünk és kis trendsimítási paraméter (nts) használunk, aminek az értéke: $(12/2)+1=7$. Legyen tehát „hét - periódusú – ablak” a duplanégyzet súlyok $[W(t)]$ meghatározásához.

$$W(t) = B[t/(T+1)] = B(u)$$

$$u = t/(T+1)$$

$$t = -T, \dots, 0, \dots, T$$

Példánkban: $t = -3, -2, -1, 0, 1, 2, 3$

$$T+1=4$$

A súlyok $W(t)$:

$$W(-3)=B(-3/4)=[1-(-3/4)^2]^2 = 0,191$$

$$W(-2)=B(-2/4)=[1-(-2/4)^2]^2 = 0,563$$

¹²⁹ A SABL programozás lépéseinek részletes ismertetésétől annak bonyolultsága miatt eltekintünk. Ld.: Levenbach, H. - Cleary, J. P. [1982]: 215-220. és 248-274.

$$W(-1)=B(-1/4)=[1-(-1/4)^2]^2 = 0,879$$

$$W(0)=B(0)=[1-0^2]^2 = 1,000$$

$$W(1)=B(1/4)=[1-(1/4)^2]^2 = 0,879$$

$$W(2)=B(2/4)=[1-(2/4)^2]^2 = 0,563$$

$$W(3)=B(3/4)=[1-(3/4)^2]^2 = 0,191$$

A többi súly 0, pl.: $T=5$, mert $u = 5/4 > 1$

Ha $|u|>1$ akkor $B(u)=0$

A csökkenő súlyozású (tapered) átlag számítását mutatja a 3-3. táblatábla.

3-3. tábla: Csökkenő súlyozású (tapered) átlag számítása április hónapban

Hónap	Adat	Súly	Súlyozott =Adat*Súly	érték	Csökkenő súlyozású átlag=24,972/4,266
Jan.	5,831	0,191	1,114		
Febr.	5,84	0,563	3,288		
Márc.	5,849	0,879	5,141		
Ápr.	5,856	1	5,856		5,854
Máj.	5,861	0,879	5,152		
Jun.	5,863	0,563	3,301		
Júl.	5,864	0,191	1,120		
Összesen		4,266	24,972		

Az egyes lépések végrehajtásához a következő adatok szükségesek:

- **ndt** = az idősor hossza adatokban számolva;
- **ndp** = egy szezon hossza adatszámban mérve, szokásos értékei: [ha hónapokkal dolgozunk]: 4 [negyedéves], 12 [éves],), ha hetekkel dolgozunk: 52 (éves szinten). Ha éves adatokkal dolgozunk akkor a ciklus hosszát kell megadni: pl.: 3, 9, 18, 54.

• **nss** = a szezonális rész simítási faktora, minél nagyobbat választunk, annál simább adatokat kapunk a számítás során. Szokásos értékei: $[ndt/ndp]+1$ [kis], $2[ndt/ndp]+3$ [közepes], $4[ndt/ndp]+7$ [nagy], ahol a szögletes zárójelek az egészrész függvényét jelentik.

Pl.: Ha a havi adatok idősora 68 adatból áll ($ndt=68$), akkor: a nss = a szezonális rész simítási faktora $[68/12]+1=5+1=6$ (kis), $2[68/12]+3= 13$ (közepes), $4[68/12]+7=27$ (nagy).

Ha negyedéves adataink vannak és az adatsor 40 adatból áll ($ndt=40$) akkor: a nss = a szezonális rész simítási faktora $[40/4]+1=10+1=11$ (kis), $2[40/4]+3= 23$ (közepes), $4[40/4]+7=47$ (nagy).

- **nts** = a trend rész számításakor használt simítás foka. Itt is a nagyobb szám nagyobb simítást eredményez. Szokásos értékei: $[ndp/2]+1$ [kis], $ndp+3$ [közepes], $2ndp+7$ [nagy], a nagyobb simítási érték mindig az előző kétszeresénél eggyel nagyobb.

Pl.: Havi adatoknál, ahol a periódus hossza (ndp) 12, a kis trendsimítási paraméter (nts): $(12/2)+1=7$, a közepes trendsimítási paraméter (nts): $12+3=15$, a nagy trendsimítási paraméter (nts): $(2*12)+7=31$.

Ha negyedéves adataink vannak, ahol a periódus hossza (ndp) = 4, a kis trendsimítási paraméter (nts): $(4/2)+1=3$, a közepes trendsimítási paraméter (nts): $4+3=7$, a nagy trendsimítási paraméter (nts): $(2*4)+7=15$.

A SABL szoftver felismeri az adatállomány hosszát és megadja a trend és szezon komponens közepes simítási fokát.

- **ndd** = a számítás közben kieső adatok pótlására szolgáló becslés tartományának hossza: lehetséges értékei: **1, 2, ... $ndt/[2*ndp]$** ;

var = ezt a paramétert nem kötelező megadni, de ha megadjuk, akkor az iterációs eljárás végeztével a végeredményben az outlierok irreguláris értékeit korrigálni lehet. Minden olyan irreguláris adat vágásra kerül, ahol:

$$|irreguláris_adat| \geq var * irreguláris_sor_szórása$$

azaz, ha az irreguláris adat nagyobb, mint az irreguláris sor szórásának var-szorosa, akkor azzal az értékkel helyettesítjük. A Var megadásával az outlierek hatását mérsékelni lehet.

A Var szokásos értékei: 2 és 3.

Az eljárás lépéseit foglaljuk össze a következő táblában.

A SABL tehát elsődlegesen az adatállomány transzformációját végzi el, eredeti formájában közvetlen előrejelzésre nem alkalmas. A helyesen transzformált adatok additív kapcsolatú trend és periódus [szezonzón vagy konjunktúra] komponenseket tartalmaznak, az irreguláris részt viszont nem. A SABL által transzformált sor felhasználható a különböző időszorkutatási módszerekkel [pl. trend-extrapoláció, exponenciális simítás, stb.] történő becslésekhez. A módszer alkalmazására számítógépes program készült¹³⁰, amely a következőképpen működik:

3-4. tábla: A SABL dekompozíciós folyamat¹³¹

INPUT	MŰVELET	OUTPUT
Y=Eredeti adatok	Adatok transzformálása	$D[0] = Y^p \quad p > 0$ $D[0] = \lg Y \quad p = 0$ $D[0] = -Y^p \quad p < 0$
D[0]=transzformált adatok	Első trend simítás	T[1]=D[0] adatok első trend simítása
D[0]-T[1]=nyers szezonális	Szezonális trend simítása	S[1]=D[0]-T[1] simított szezonális
S[1]	Szezonális trend simítása és eltávolítása	S[2]=S[1] - S[1] trend simítása
S[2]	Elveszett szezonális komponensek becslése	S[3]=a kiegészített szezonális értékek hozzáadása a sor elejéhez és végéhez
D[0]-S[3]=nyers trend	Trend simítás	T[2]=D[0]-S[3] trend simítása
D[0]-T[2]=nyers szezonális	Szezonális trend simítása	S[4]=D[0]-T[2] simított szezonális
S[4]	Szezonális trend simítása és eltávolítása	S[5]=S[4] - S[4] trend simítása
D[0]-S[5]=nyers trend	Trend simítás	T[n]=D[0]-S[5] trend simítása
D[0]-T[n]=nyers szezonális	Szezonális trend simítása	S[6]=D[0]-T[n] simított szezonális
S[6]	Szezonális trend simítása és eltávolítása	S[n]=S[6] - S[6] trend simítása
D[0], T[n], S[n]	Irreguláris komponens számítása,	I[n]=D[0]-T[n]-S[n]
D[0], T[n], S[n], I[n]	grafikus ábrák készítése (T, S, I, T+S)	grafikus ábrák értelmezése és diagnosztizálása

3-4. tábla folytatása: SABL szoftver működése: megadandó paraméterek

Paraméter neve	Jelentése	Lehetséges értéke
in	az idősort tartalmazó adatfájl neve, amit a SABL felismer, egy oszloban adatsor, egérművelettel megkereshető, Excel fájlnál, módosított mentés: formázott szöveg szóközzel tagolt, *.prn, majd a , cseréje .-ra a Total Commanderben: *.txt vagy *.dat. Meg kell adni.	*.dat *.txt

¹³⁰ Kruzsliz Ferenc-Kiss Tibor-Sipos Béla: SABL Decomposition of Time Series ©1997 PTE Version 2.1.

¹³¹ Cleveland, W. S.-Dunn, D. M.-Terpenning, I. J.[1979] Table 1, p. 206.

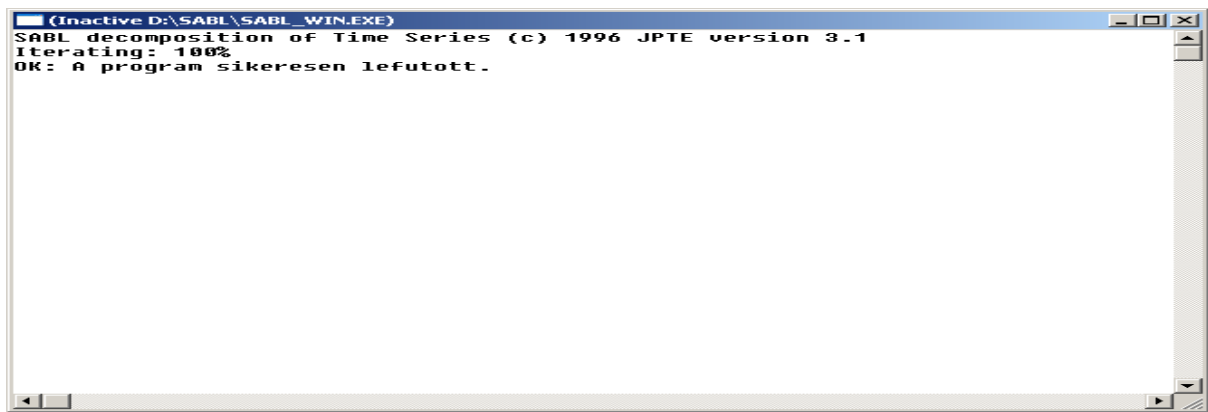
sbl, out	a fájl neve, amelybe az eredményeket tesszük, egér-művelettel megkereshető. Meg kell adni.	*.sbl *.out
ndp	az idősor egy ciklusának hossza (havi adatoknál 12, negyedéves adatoknál 4, hosszú ciklusoknál 9 stb.), amit be kell írni. Meg kell adni.	12, 4, 9 stb.
nts	a trend simítás nagysága, lehet változtani: kis, közepes és nagy, alapértelmezés közepes, ndp megadása után dupla kattintás.	Közepest megadja
nss	a szezonális simítás nagysága, lehet változtani, kis, közepes és nagy, alapértelmezés közepes, ndp megadása után dupla kattintás.	Közepest megadja
ndd	az iterációban kieső adatok helyreállításához használandó regressziós adatok száma (max. a periódusok számának fele, tehát az idősor hossza osztva a periódus hosszával)	1..[ndt/ (2·ndp)]

Az "in" "out" és "ndp" paraméter megadása kötelező, az összes többi nem, mivel a szoftver a többi paramétert becsüli az nts és nss értékekre közepes simítási értéket ad.

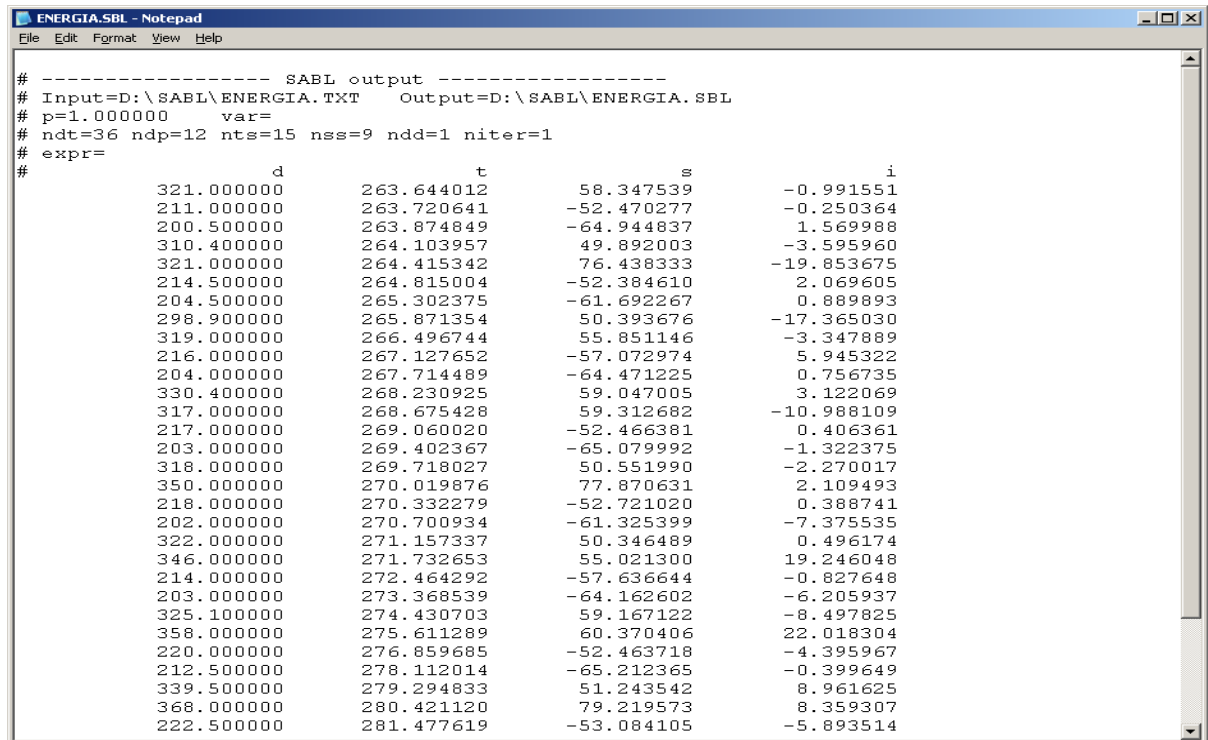
ndt	az idősor hossza (az adatok száma), a program kiszámolja, nem kell megadni	1...16384
niter	az algoritmus során végrehajtandó iterációs lépések száma	0...65535
p	a be- és kimeneti adatok transzformációs paramétere: Meg kell adni, 0, 0.5, 0.33 stb.	bármely valós szám: 1 az alap eset
var	irreguláris rész levágása, az alábbi összefüggés szerint $ \text{irreguláris adat} \geq \text{var} \cdot \text{szórás (irreguláris sor)}$ Alapeset: 1.	pozitív valós szám: 2, 3
Eredmény	az Eredmény munkalapban a számítások paramétereit és eredményeit láthatók: D : eredeti adatok T : trend komponens S : szezonális komponens I : irreguláris rész A D, T, S, I adatsorokat Excelbe másoljuk, a tizeses .-t cseréljük ,-re és az adatokat pl. T, S, I, T+S ábrázolhatjuk elemezhetjük.	D, T, S, I

A SABL SHELL.EXE szoftver alkalmazásának bemutatása.

Számolás után:

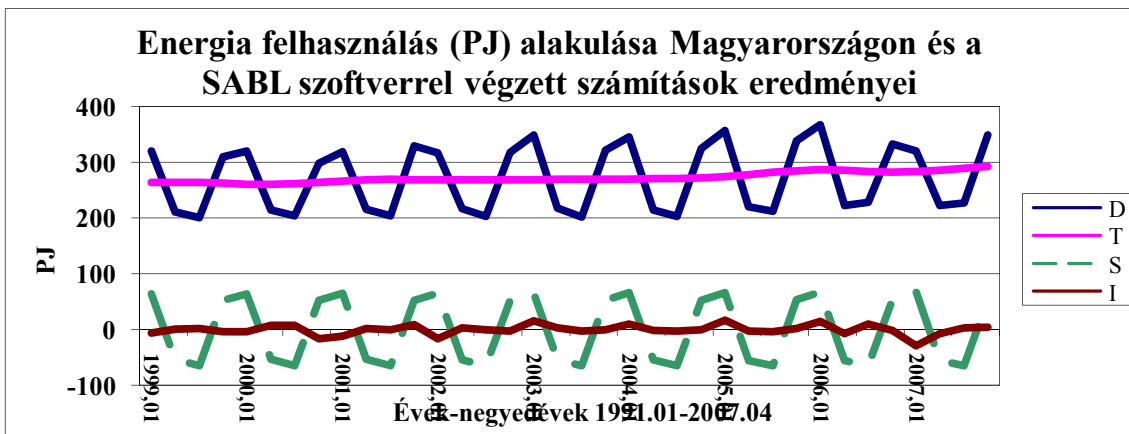


Az Eredmény: d = eredeti idősor, t = SABL trend, s = SABL szezonális, i = irreguláris rész.



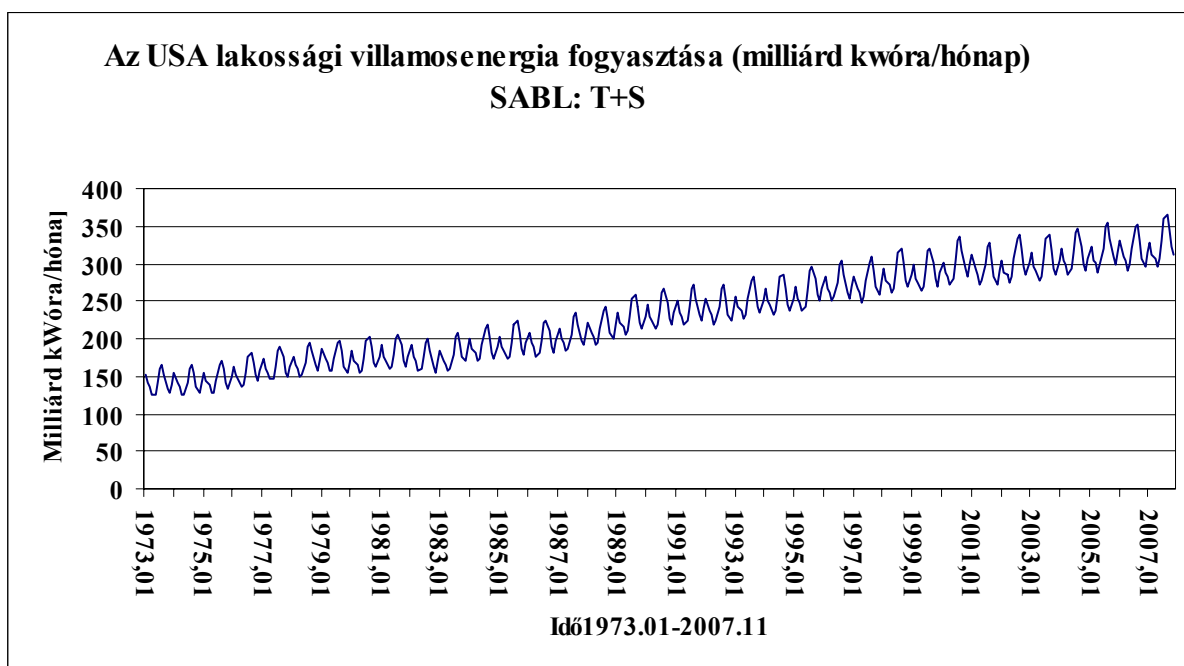
A SABL szoftvert a főkönyvtárba kell másolni: pl. C:/SABL vagy: D:/SABL.

A program a SABL SHELL.EXE fájljal indul. Előtte az adatokat elő kell készíteni. Az Excel fájl csak egy adatsort tartalmazzon, azt amit majd a SABL-lal le kívánunk futtatni. Az oszlopoknak ne legyen neve az első cellában, csak az idősorot használjuk. Mentsük le módosított mentéssel: Formázott szöveg (szóközzel tagolt, kiterjesztés: *.prn) formátumban. Windows Commanderrel szerkesszük, a vesszőket cseréljük ki pontokra. Használhatjuk a jegyzettömböt (notepad-ot) is ahol megnyitjuk a *.prn fájlt és a vesszőket cseréljük ki pontokra és lementjük a szöveg állományt: *.txt. Ezt a fájlt a SABL felismeri, majd az output fájlnek nevet kell adni: pl. *.sbl és a periódus hosszát (ndp) módosítani kell az adatsor ismeretében, pl. havi adatok 12, negyedéves adatok: 4, stb. Ezt követően a szezon és a trend simítása cellákra klikkelve a közepes simítás fokát kiszámolja, amitől el lehet térni. Ha a közepes simítási fok felét vesszük akkor kis, ha dupláját akkor nagy simítási fokkal számolunk. Erőteljes irreguláris hullámmás és gyakori kiugró értékek esetében célszerű a nagy simítási fokot alkalmazni. A többi paramétert [Pl.: Adatok transzformációja (p=), Kiugró értékek vágása (var=)] az előzőekben leírtak szerint lehet megadni. Ha lefutott a program az Eredmény munkalapban megjelenik a D, T, S és I oszlopvektor, amit Excelbe lehet másolni és a tizedes pontokat ki kell cserélni vesszőkre. Az adatok Excelben (pl.: T, S, I, T+S) ábrázolhatók és értékelhetők. A számításokat a Magyarország energiafelhasználása, negyedéves adatok, 1999-2007 (PJ) alapján mutatjuk be, ahol: **D**: eredeti adatok, **T**: trend komponens, **S**: szezonális komponens és **I**: irreguláris rész. Grafikusan ábrázolva az eredeti adatokat (D) és a SABL trendet (T), a szezonális (S) és az irreguláris részt (I):



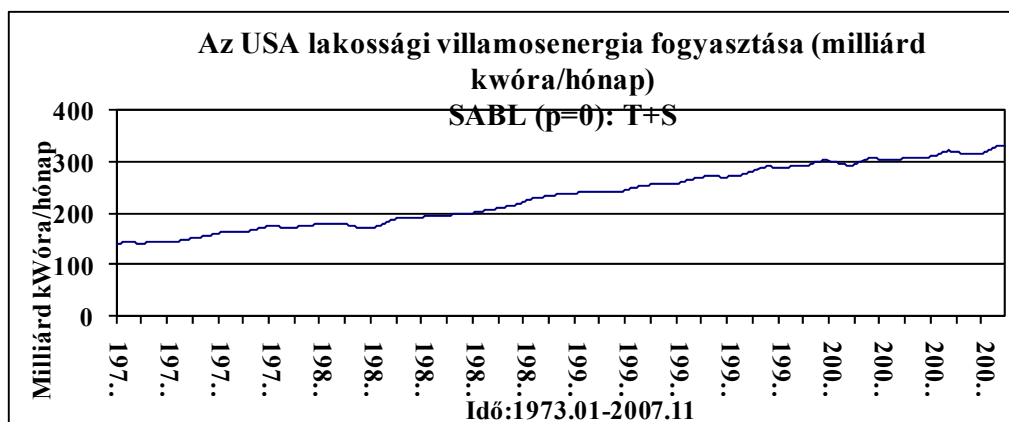
3-23. ábra: Energia felhasználás alakulása Magyarországon

Az USA energiafelhasználása¹³² esetén a számításokat elvégezve a T+S az alábbi ábrát mutatja:



3-24. ábra: Az USA villamos energia fogyasztása

Ha az adatok transzformációja $s=0$, akkor a trend és szezon tényező közötti multiplikatív kapcsolatot additívvá alakítjuk.



3-25. ábra: Az USA villamos energia fogyasztása. A SABL trend

¹³² Az adatsort a szezonális vizsgálatánál már bemutattuk.

Gyakorló feladatok a SABL-szoftver alkalmazására*

1. A SABL szoftver felhasználásával válassza szét a trend (T), a periódikus (S) és az irreguláris (I) tényezőket és ábrázolja a T+S összeget az Amerikai Egyesült Államok (USA) rendelkezésre álló hosszú időszakai alapján. A periódus hosszának vegyen 9 és 18 évet.

1.1 Egy főre jutó aranytermelés 1860-2005 között.

1.2 Egy főre jutó kőolajtermelés 1866-2005

1.3 Egy főre jutó kőszéntermelés 1860-2005

1.4 Egy főre jutó ólomtermelés 1830-2005

1.5 Egy főre jutó acéltermelés 1867-2005

1.6 Egy főre jutó vasérctermelés 1880-2004

3.8 Az ARIMA modellezés menete

George E. Box és Gwilym M. Jenkins népszerűsítette az autoregresszív/mozgóátlag modellek alkalmazását idősor prognosztizálási feladatokra. Miközben ezt a módszert eredetileg az 1930-as években fejlesztették ki, nem volt széleskörűen ismert, amíg Box és Jenkins nem publikálta részletes leírását könyv formában 1970-ben¹³³. A Box és Jenkins által ajánlott általános módszer, ARIMA modellek alkalmazása idősorelemzésre, prognosztizálásra és ellenőrzésre az idősorelemzés Box-Jenkins módszertanaként lett ismert. Az ARIMA (AUTOREGRESSIVE-INTEGRATED-MOVING-AVERAGE=AUTOREGRESSZÍV INTEGRÁLT MOZGÓÁTLAG) modellezés menetét a következőkben foglaljuk össze.¹³⁴

A sztochasztikus idősori modellek *integrált autoregresszív és mozgóátlag* (rövidítve ARIMA) modellcsaládjának elnevezésében, az AR az autoregresszív, az MA a mozgóátlag jelzőre, az I betű (INTEGRATED) pedig az összegzésre utal. Az *autoregresszív (AR) modell*, az idősor jelenlegi értékét, saját előző értékeinek függvényében fejezi ki, természetesen, mint sztochasztikus modell, kiegészülve a véletlen ingadozást reprezentáló változóval. Az autoregresszió a regresszió olyan formája, melyben az eredményváltozó más magyarázó változók helyett saját különböző késleltetésű múltbeli értékeihez kapcsolódik. Statisztikai szempontból tehát egyváltozós idősorelemzést végzünk.

ARIMA (p, 0,0):

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

vagy:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

ahol: p az autoregresszivitás rendjét jelöli.

A *mozgóátlag (MA) modell* az idősor jelenlegi értékét, a jelenlegi és a múltbeli véletlen változók függvényében fejezi ki.

ARIMA (0,0,q)

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

ahol: q a mozgóátlag folyamat rendjét jelöli.

(théta= θ , ε =epszilon, ϕ)

Mozgóátlag. Két különböző jelentése van ennek a kifejezésnek. Először: idősoroknál a k- tagú mozgóátlagot k egymást követő megfigyelési érték átlagaként definiálhatjuk. Ezt felhasználhatjuk simításra vagy előrejelzésre. Másodsor: a Box - Jenkins modellezésben az MA a mozgóátlag rövidítése az ARIMA-ban, és az jelenti, hogy az idősor értékét a t időpontban befolyásolja a jelenlegi hibatag és a múltbeli hibatagok súlyozott kombinációja.

¹³³ Box, G.E.P. - Jenkins, G.M. [1970]: Time Series Analysis. Forecasting and Control. Holden-Day, San Francisco, CA.

¹³⁴ Herman-Pintér-Rappai-Rédey: Statisztika II. [1994] Pécs. 8.6. Sztochasztikus idősori modellek. felhasználásával és kiegészítésével.

ARMA modell Az ilyen típusú időseries modell formája lehet autoregresszív (AR) vagy mozgóátlag (MA) vagy a kettő kombinációja (ARMA, vagy más néven vegyes modell).

A *vegyes (ARMA) modell* az időseries jelenlegi értékét, saját előző értékeinek, és a jelenlegi, illetve a múltbeli véletlen változók függvényében fejezi ki.

ARIMA(p,0,q):

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Az *autoregresszív integrált mozgóátlag (ARIMA)* modell, a differencia- vagy különbszet- képzéssel stacionáriussá transzformált, ún. d-ed rendű *integrált* [I(d)] időseriesokra felírt ARMA modell. Ha például az időseries első differenciái (az $Y_t - Y_{t-1}$ értékek) stacionáriusak, az eredeti időseries elsőrendű integrált [I(1)].

Az ARIMA modellezés kiindulópontja annak megállapítása, hogy a vizsgálni kívánt időseriesunk stacionárius-e, illetve, ha nem, akkor az, hogy alkalmas transzformációval stacionáriussá tehető-e. Ezzel eldöntöttük azt, hogy az adott időserieshoz illeszthető-e ARIMA modell, ha igen milyen (d) dimenzióval (fokkal) rendelkezik. A következő kérdés annak megválaszolása, hogy milyen típusú ARMA modell illesztésével próbálkozunk, illetve, milyen legyen az autoregresszivitás (p) és/vagy, a mozgóátlagolás (q) rendje. Erre a kérdésre a választ a tapasztalati, vagy a transzformált időseries ACF és PACF értékei (autokorrelációs- és parciális autokorrelációs együtthatók) alapján adjuk meg. A modellezés ezen fázisát, modell azonosításnak (identifikációnak) nevezi a szakirodalom.

Ezután a modellezés lépései alapvetően megfelelnek a már ismert lineáris regressziós modellezésnek. A választott modell paraméterbecslése után a modell ellenőrzése következik. A modell ellenőrzése során vizsgáljuk azt, hogy paraméterei szignifikánsak-e, illetve véletlen változók fehér zaj folyamatot követnek-e. Speciálisan az ARMA modelleknek van stacionaritási (az időseries jellemzői időben állandóak, azaz függetlenek a t időváltozótól) és invertibilitási (azaz a becslt paraméterek abszolút értéke kisebb mint egy) feltétele is, melyek a modell paramétereinek értékére vonatkozó megszorításokként jelennek meg. Ezután döntünk arról, hogy felhasználható-e az illesztett modell elemzésre, előrejelzésre, vagy más modell választásával kell próbálkoznunk. A modellkészítés menetét illusztrálja az alábbi folyamatábra.

Az ARIMA modellek dimenzióit a következő módon adjuk meg: ARIMA (p, d, q). A gyakorlati alkalmazások szerint az időseriesok nagy része jól közelíthető olyan modellekkel, melyeknél az autoregresszivitás és a mozgóátlag folyamat rendje (p és q), illetve a differenciaképzés foka (d) alacsony. Általában mindhárom dimenzió - a (p), a (d), és a (q) is 0, vagy 1, vagy 2 értéket vesz fel. Olyan időseriesok elemzésére, melyek szezonális ingadozást is tartalmaznak, az ún. *szezonális ARIMA modellek* alkalmasak, melyekkel pl. havi adatsorok (s=12), vagy negyedéves adatsorok (s = 4) elemezhetők. A szezonális modellek általános jelölése: ARIMA (p, d, q) (P, D, Q)_s.

A paraméterek függvényében az időseries hossza rövidül, a kieső adatok száma=p+d+q+s(P+D+Q)

Az ARIMA modellezés *feltevése* szerint tehát, az időseriesok által reprezentált sztochasztikus folyamatok, viszonylag egyszerű lineáris modellekkel leírhatók. A modellezési technika kidolgozásának *alapvető célja*, megbízható rövidtávú előrejelzések készítése.

A modellezés során a megfigyelt időseriesban tapasztalható belső összefüggések alapján következtetünk a sztochasztikus folyamat jellegzetességeire. E jellemzők határozzák meg a választott modell típusát. A modell paramétereinek segítségével - melyeket a megfigyelt időseriesból becslünk, - leírható az a "szisztéma", aminek alapján az időseries jelen időszaki értéke előállítható, az időseries múltbeli értékeinek és/vagy az elmúlt időszakban realizálódott véletlen eltéréseknek a lineáris kombinációjaként. A modellek alkalmazásának végső soron az a célja, hogy választott megbízhatósági szint mellett az időseries jövőbeni értékeire intervallumbecslést tudjunk adni. Az időseriesokban tapasztalható belső összefüggések megállapítása az időseriesok korrelációs struktúrájának feltárását jelenti, ez indokolja a modellkészítés nagy adatigényét. Általában 100-120 elemű megfigyelt időseriesra van minimálisan szükség, mivel célszerű minél nagyobb készletességig elmenni, ami azt jelent, hogy általában a $K = 20-25$. Az általánosan alkalmazott összefüggés $K < (n/4)$ alapján 25 időszakkal történő készletetéshez legalább 100 elemű időseriesra van szükségünk. Nem szezonális, tehát éves adatok esetében a K lehet kisebb, pl. 12 (ez 48 éves adatot igényel), de szezonalitást tartalmazó, pl. havi adatoknál a tendencia feltárásához $3 \cdot 12 = 36$ havi készletetésre is szükség lehet, itt az $n > 144$ havi adat. A hosszú időtávot átfogó, összehasonlítható adatsor megfigyelése sokszor megnehezíti a modellezést, amihez hozzájárul még az alkalmas, lehetőleg ingyenes, szoftverek beszerzése. (pl. JMulti, [<http://www.jmulti.de/>] Gretl, [<http://gretl.sourceforge.net/>] R+ interneten működő szoftver)

R+ interneten elérhető: Free Statistics Software (Calculator).

Adatcsere után paramétereket be kell állítani és számol, hibás paraméterezés esetében üzenetet ad, Excel-be is lehet exportálni, a vesszőket ki kell cserélni, ha az adatokkal további számításokat végzünk az Excelben, pl. diagram készítése.

(Partial) Autocorrelation Function - Free Statistics Software (Calculator):

http://www.wessa.net/rwasp_autocorrelation.wasp#output

ACF és PACF számítása és tesztelése.

ARIMA Backward Selection - Free Statistics Software (Calculator)

http://www.wessa.net/rwasp_arimabackwardselection.wasp#output

ARIMA paramétereinek becslése, backward eliminációs módszer alkalmazása.

ARIMA Forecasting - Free Statistics Software (Calculator) :

http://www.wessa.net/rwasp_arimaforecasting.wasp

Előrejelzés és tesztelés (teszt peridus megadása).

Az elméleti idősort jelentő sztochasztikus folyamat jellemzői (várható értéke, szórásnégyzete és autokorrelációs együtthatói) azonban, csak akkor becsülhetők a tapasztalati idősből, ha ezek a jellemzők *időben állandóak*, azaz függetlenek a t időváltozótól. Az ilyen tulajdonságokkal rendelkező idősorokat *stacionárius idősoroknak* nevezzük.

Stacionárius idősor az $Y_1, Y_2, \dots, Y_t, \dots, Y_T$ elméleti idősor ha :

$$1. E(Y_t) = \mu$$

$$2. \text{Var}(Y_t) = \sigma^2$$

$$3. \rho_k = \frac{\text{Cov}(Y_t, Y_{t-k})}{\sigma^2} \quad k = 1, 2, \dots$$

$$\text{ahol: } \text{Cov}(Y_t, Y_{t-k}) = E[(Y_t - \mu)(Y_{t-k} - \mu)] \text{ és}$$

$$\sigma^2 = \sigma_t \times \sigma_{t-k} \text{ mivel } \sigma_t = \sigma_{t-k} = \sigma \quad (\mu = m\ddot{u})$$

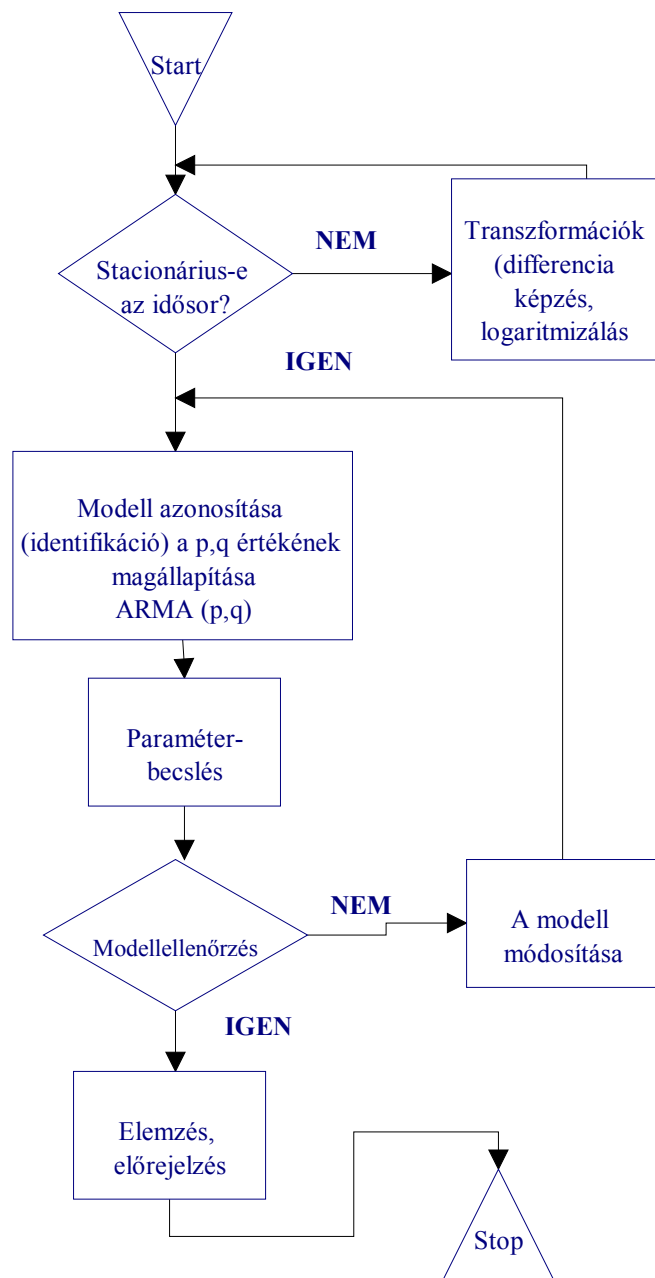
A stacionárius idősorok *nem tartalmazzak időtrendhatást*, az idősor értékei egy állandó átlagos szint körül ingadoznak, állandó szórással. Az állandó szórás azt jelenti, hogy az *ingadozások intenzitása időben nem változik* (nem növekszik vagy csökken). Ezenkívül a stacionárius idősorokra jellemző az, hogy az *autokorrelációs együtthatói* (ρ_k) *időben állandóak*, nem függenek t -től, csak a változók egymás közötti távolságától, k -től. Amennyiben idősorunk nem stacionárius folyamatból származik, alkalmasan megválasztott transzformációval stacionerré próbáljuk alakítani. Ha ez nem sikerül, a folyamatra ARIMA modellek nem illeszthetők.

1. A stacionaritás biztosítása

Az ARIMA modellezés kiindulópontja annak megállapítása, hogy a vizsgálni kívánt idősorunk stacionárius-e, illetve, ha nem, akkor az, hogy alkalmas transzformációval stacionáriussá tehető-e. Ezzel eldöntöttük azt, hogy az adott idősorhoz illeszthető-e ARIMA modell, ha igen milyen (d = DIFFERENCE, DIFFERENCIA) dimenzióval (fokkal) rendelkezik.

1.a Stacionárius idősor. Az előzőekben definiáltuk a stacionárius idősor kritériumait. Ennek alapján a sor stacionárius, ha a statisztikai jellemzői függetlenek a konkrét időperiódustól a megfigyelés folyamán.

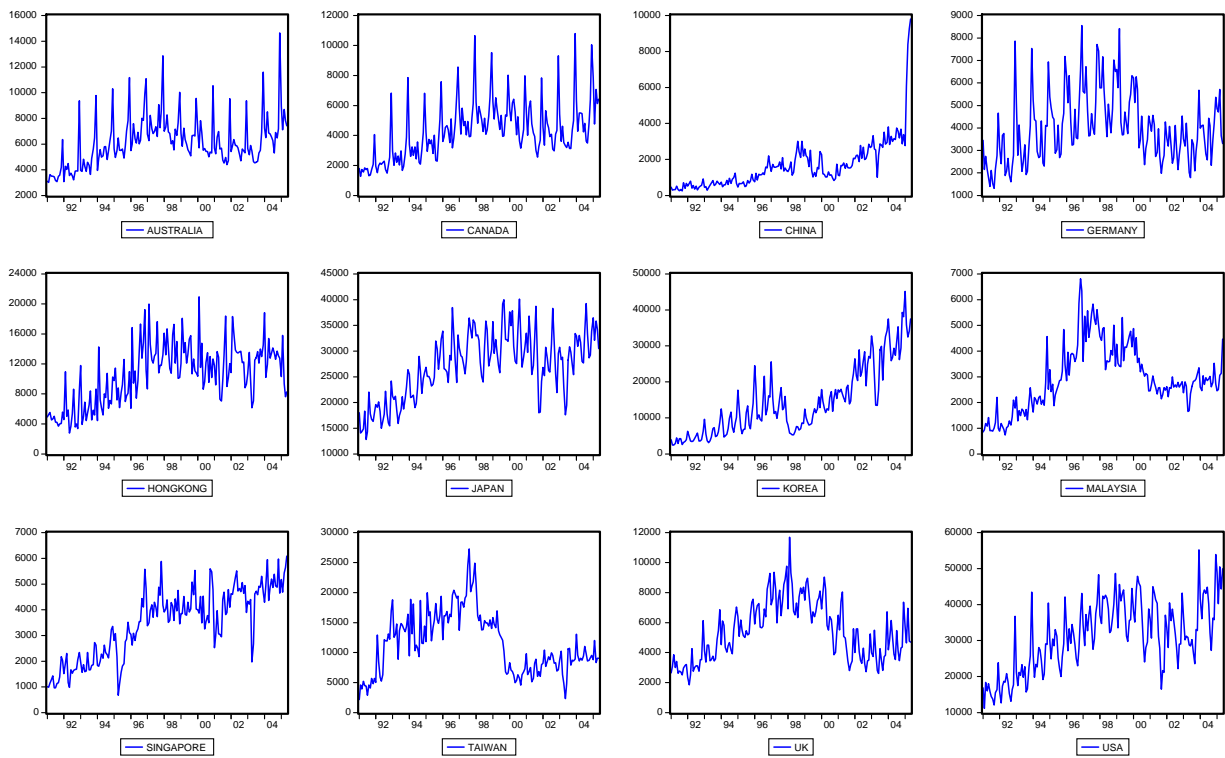
1.b Nem stacionárius idősor. A tapasztalati idősor nem stacionárius jelleget mutat, ha az alapjául szolgáló folyamat átlaga és/vagy varianciája nem konstans. A gyakorlatban az idősor ábrájának tanulmányozása, illetve az idősor autokorrelációinak sorozata felhasználható annak megállapítására, hogy az egyik, vagy mindkét feltétel fennáll-e. Az első lépés tehát stacionaritás biztosítása.



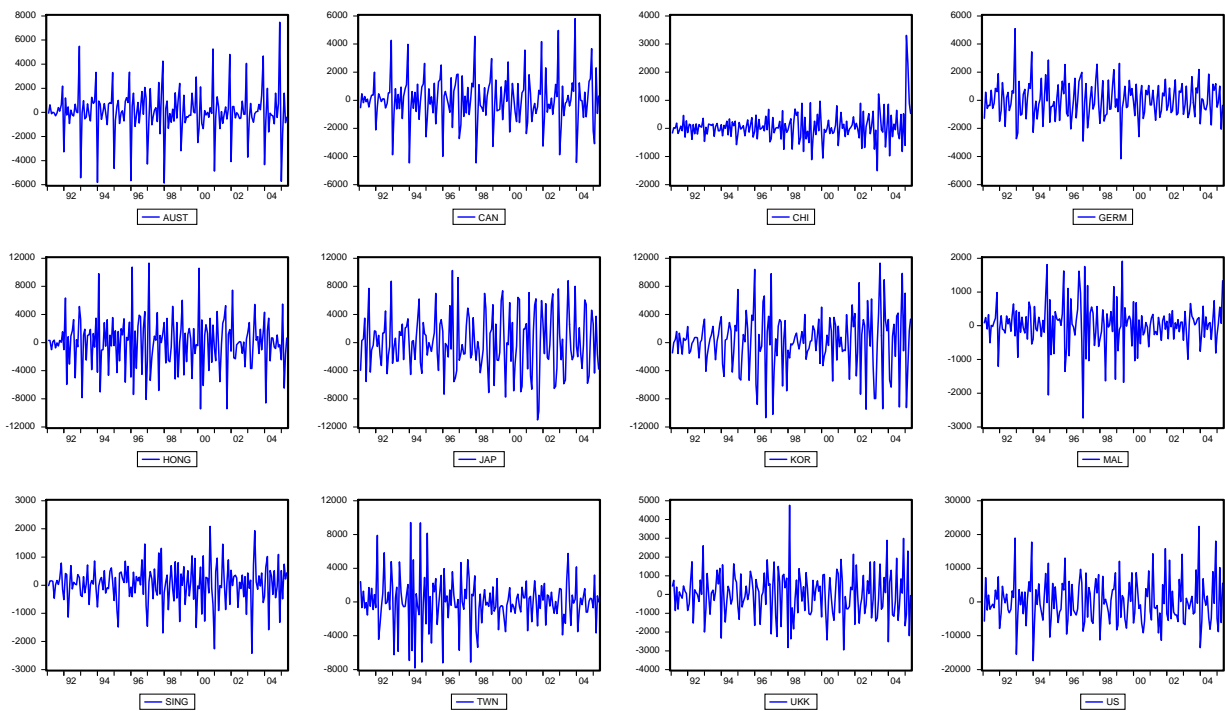
3.8.1 Az ARIMA modellezés lépései.

Nem stacionárius idősorok¹³⁵:

¹³⁵ Forrás: <http://www.google.com/search?q=forecasting+error+definition+filetype:ppt&hl=en&client=netscape-pp&rls=com.netscape:en-US&prmd=ivns&ei=QUJKTtGLLYzOswan9JmUB w&start=40&sa=N> Koreából érkező turisták száma, havi bontásban 1992-2004.



Stacionárius idősorok, (differencia képzéssel):



I A stacionaritás biztosítása éves adatok esetében, nem szezonális differencia képzés.¹³⁶

Általában megállapítható, hogy a gazdasági, társadalmi idősorok többsége jelentős fejlődést mutat, általában jellemző rájuk az emelkedő, vagy csökkenő tendencia. Ilyen esetekben mondhatjuk, hogy az idősor várható értékében nem stacionárius, és ezért képezzük a sor első differenciáját, azaz képezzük az idősor $(Y_t - Y_{t-1})$ értékeit. Az első differenciák, az eredeti sorból számított *változásokként* értelmezhetők.

$$\Delta Y_t = Y_t - Y_{t-1}$$

Vegyük az alábbi lineáris időtrendet:

$$\hat{Y}_t = b_0 + b_1 t$$

Az időtrend első differenciája:

¹³⁶ Ramanathan Ramu [2003]: Bevezetés az ökonometriába alkalmazásokkal. Panem. 537-540.

$$\frac{dy}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Delta Y}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{Y(t + \Delta t) - Y(t)}{\Delta t} \approx (Y_t - Y_{t-1}) / (t - [t-1]) = (Y_t - Y_{t-1})$$

A lineáris trend t-szerinti deriváltja:

$$\frac{d\hat{Y}}{dt} = b_1$$

Mivel tehát a közelítőleg lineáris időtrend esetén az egymást követő tényadatok között konstans értékű a különbség, vagyis a növekmény vagy csökkenés, így az egységintervallum szerinti differenciaképzéssel a időtrendhatást tekintve konstans értéken tartható a folyamat.

Amennyiben az első differenciák nem stacionáriusak, akkor másodszor is differenciálni kell, mégpedig az első differenciákat $[(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})]$. Ilyenkor az idősor másodrendű integrált $[I(2)]$. Az elsőfokú differencia-sornak n-1, a másodfokú differencia-sornak n-2, a míg a tizenketted fokú differencia-sornak [tizenkettedrendű integrált $I(12)$] pedig n-12 adata lesz.

$$\Delta Y_t = (Y_t - Y_{t-1})$$

$$\Delta^2 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

A magasabb fokú differencia-képzés szükség estén tovább folytatható.

$$\Delta^3 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) - (Y_{t-2} - Y_{t-3}) = Y_t - 2Y_{t-1} + Y_{t-3}$$

$$\Delta^4 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) - (Y_{t-2} - Y_{t-3}) - (Y_{t-3} - Y_{t-4}) = Y_t - 2Y_{t-1} + Y_{t-4}$$

.

.

.

$$\Delta^{12} Y_t = Y_t - 2Y_{t-1} + Y_{t-12}$$

.

.

$$\Delta^{24} Y_t = Y_t - 2Y_{t-1} + Y_{t-24}$$

Ha az idősor exponenciális időtrenddel rendelkezik, vagyis az idősor állandó %-os ütemben, exponenciálisan nő, logaritmusos lineáris időtrendet tartalmaz, ami már differenciálható.

$$\hat{Y}_t = b_0 b_1^t$$

$$\ln \hat{Y}_t = \ln b_0 + t \ln b_1$$

A másodfokú polinomiális időtrend eliminálható kétszeri- ($\Delta^2 Y_t$), a harmadfokú polinomiális időtrend pedig háromszori ($\Delta^3 Y_t$) differenciálással.

$$\hat{Y}_t = b_0 + b_1 t + b_2 t^2$$

$$\frac{d^2 \hat{Y}}{dt} = (b_1 + 2b_2 t)' = 2b_2$$

$$\hat{Y}_t = b_0 + b_1 t + b_2 t^2 + b_3 t^3$$

$$\frac{d^3 \hat{Y}}{dt} = (b_1 + 2b_2 t + 3b_3 t^2)' = (2b_2 + 6b_3 t)' = 6b_3$$

Trend-stacionaritás esetében: $\hat{Y}_t - \hat{Y}_{trend}$ Az időtrend lehet hiperbolikus, hatványkitevős stb. számítása trend-szezon-hiba.xls parancsfájllal is történhet. Az időtrend-stacionaritás azt jelenti, hogy a trendtől való eltérések stacionáriusak. Kiszámítása esetében, beillesztés után az I.1 0 nincs transzformációt lehet választani.

A gyakorlati alkalmazásokban a nem szezonális differenciaképzésnél a differenciaképzés foka (degree of non-seasonal differencing=d) legtöbbször d=0,1,2.

Ha az az idősor stacionárius, akkor nullad rendű integrált $[I(0)]$ sornak is nevezzük.

A Box-Cox transzformáció¹³⁷:

$$Y(\lambda) = \begin{cases} (Y^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \ln(Y), & \lambda = 0 \end{cases} \quad Y > 0$$

λ (Lambda) = 1 nincs transzformáció

λ (Lambda) $\rightarrow 0$ log-transzformáció, ugyanis az egyik nevezetes határérték:

$$\lim_{x \rightarrow 0} \frac{a^x - 1}{x} = \ln(a)$$

esetünkben $a=Y$ és $x= \lambda$

$$\lim_{\lambda \rightarrow 0} \frac{(Y^\lambda - 1)}{\lambda} = \ln(Y)$$

λ Lambda = (-2, -1,9.....+1,9,+2)

$$-2 \leq \lambda \leq 2$$

A Box-Cox transzformáció után az adatok visszatranszformálása:

$$Y = (Y(\lambda)\lambda + 1)^{1/\lambda}$$

A Box-Cox transzformáció részletes leírása:

Y-t a kijelölt hatványra emeli és a megadott képlettel számol:

$$Y(\lambda) = (Y^\lambda - 1)/\lambda$$

$$\lambda = 1 \rightarrow Y$$

$$Y(1) = Y - 1$$

$$\lambda = 0 \rightarrow \ln Y$$

$$\lambda = 2 \rightarrow Y^2$$

$$Y(2) = (Y^2 - 1)/2$$

$$\lambda = \frac{1}{2} \rightarrow Y^{\frac{1}{2}} = \sqrt{Y}$$

$$Y(1/2) = (Y^{1/2} - 1)/(1/2)$$

$$\lambda = -1 \rightarrow Y^{-1} = \frac{1}{Y}$$

$$Y(-1) = (Y^{-1} - 1)/(-1)$$

A mintapéldák adatbázisát és a részletes számításokat a Box-Cox-transzformációk.xls fájl tartalmazza. A véletlen tényezőt (e) véletlenszám generátorral állítottuk elő, nagysága a -1 és +1 intervallumban ingadozik. A számításokat az R ARIMA Forecasting - Free Statistics Software (Calculator) felhasználásával végeztük. Internetes elérés:

http://www.wessa.net/rwasp_autocorrelation.wasp#output

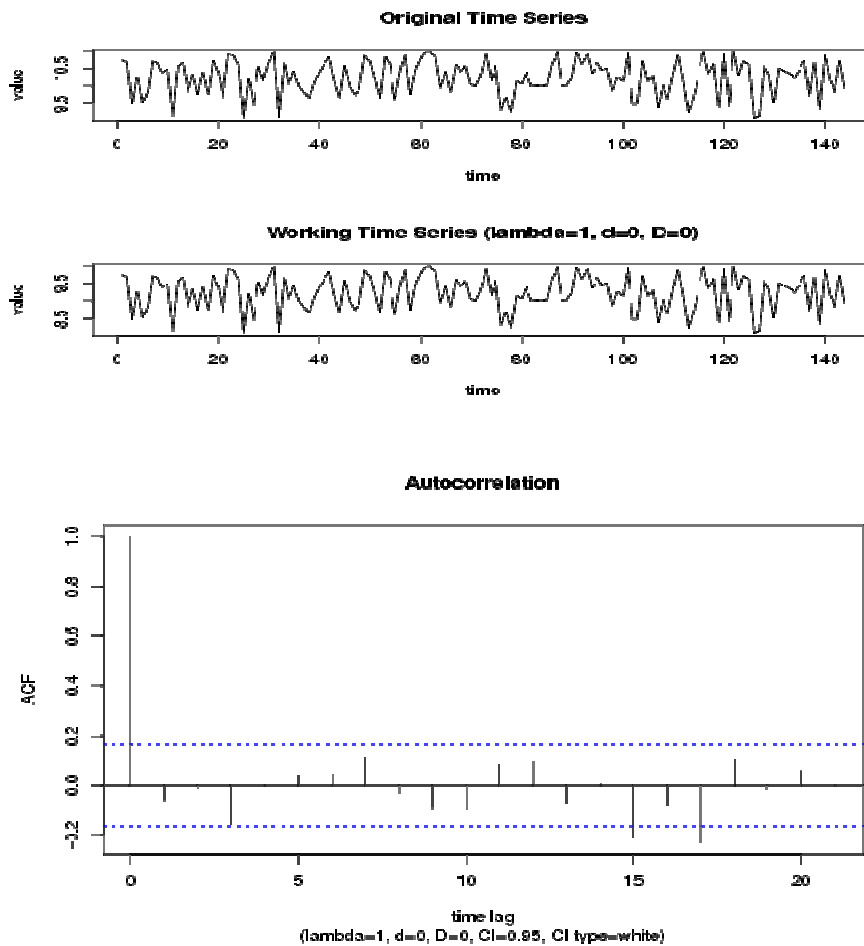
1. Nincs transzformáció ($\lambda=1$), ha az idősor a stacionaritási feltételeknek (az előzőek alapján az egyes változók várható értéke (μ), varianciája (σ^2), valamint a különböző időpontokhoz tartozó változók (Y_t, Y_{t-k}) kapcsolatát kifejező (auto)kovariancia időben állandó) eleget tesz.

Legyen $Y_t = 10 + e$

$$Y(1) = (10 - 1) + u$$

A számítások eredményei grafikusán:

¹³⁷ Ramanathan Ramu [2003]: Bevezetés az ökonometriába alkalmazásokkal. Panem. 281. ¹³⁸ Chan N. H. [2002]: Time series. 41.



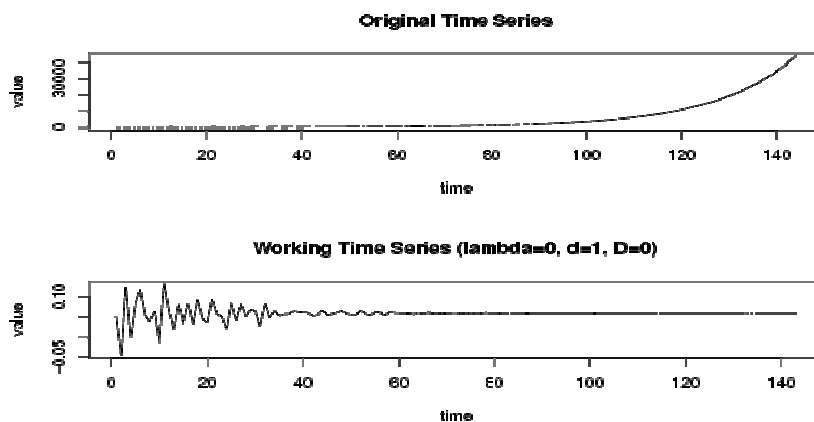
2. Az ln-transzformációt ($\lambda=0$) akkor használjuk, ha hibatényező szórása (σ) az Y növekedésével szintén nő (pl. exponenciális időtrend, amikor az átlagos növekedés üteme állandó, mint példánkban 6 %/év vagy ha növekvő az időtrend és multiplikatív szezonális, tehát az időben előrehaladva az amplitudó nő) vagy ha a hiba (ϵ_t) eloszlása jobbra ferdült, (jobbaldali asszimetria, balra hosszán elnyúló eloszlás)

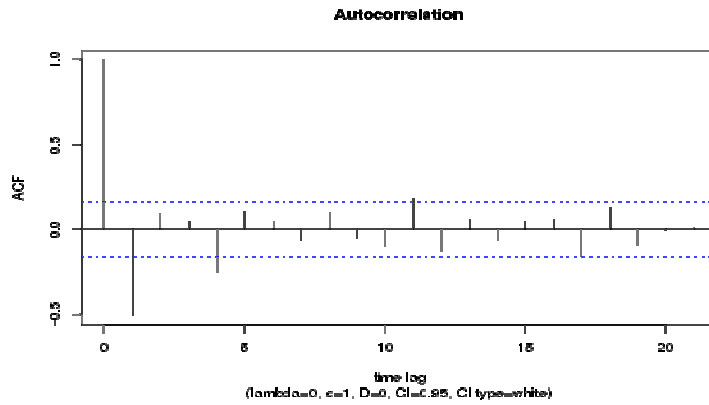
Legyen:

$$Y_t = 10 * 1,06^t * e^{\epsilon}$$

$$\ln(Y_t) = \ln(10) + t * \ln(1,06) + \epsilon$$

Ebben az esetben ln-transzformációt ($\lambda=0$) használtunk, így a relative (%-os) növekedés abszolúttá ($\ln(1,06)$) vált és elsőfokú differencia ($d=1$) alkalmazásával stacionáriussá alakítottuk a sort.



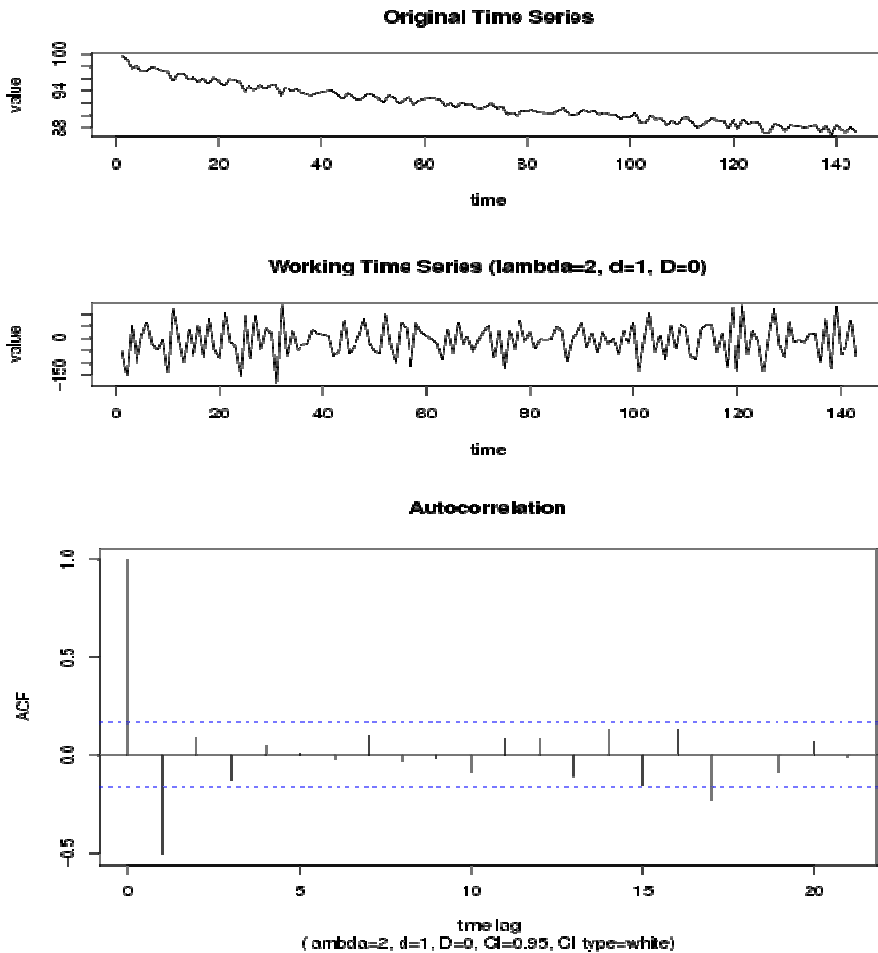


3. Négyzetes transzformációt ($\lambda=2$) akkor használjuk, ha a hibatenyező varianciája (σ^2) arányos a várható értékkel (μ) vagy ha a hiba (ε_t) eloszlása balra ferdült, (baloldali asszimetria, jobbra hosszán elnyúló eloszlás) Ha az adatsor gyökös formát követ, akkor a $\lambda=2$ transzformációval linearizálható az idősor és $d=1$ differenciálással stacionáriussá tehető.

$$Y_t = 1,05 * \sqrt{t}$$

$$(Y_t)^2 = 1,05^2 * t$$

$$Y(2) = (1,05^2 - 1)/2$$



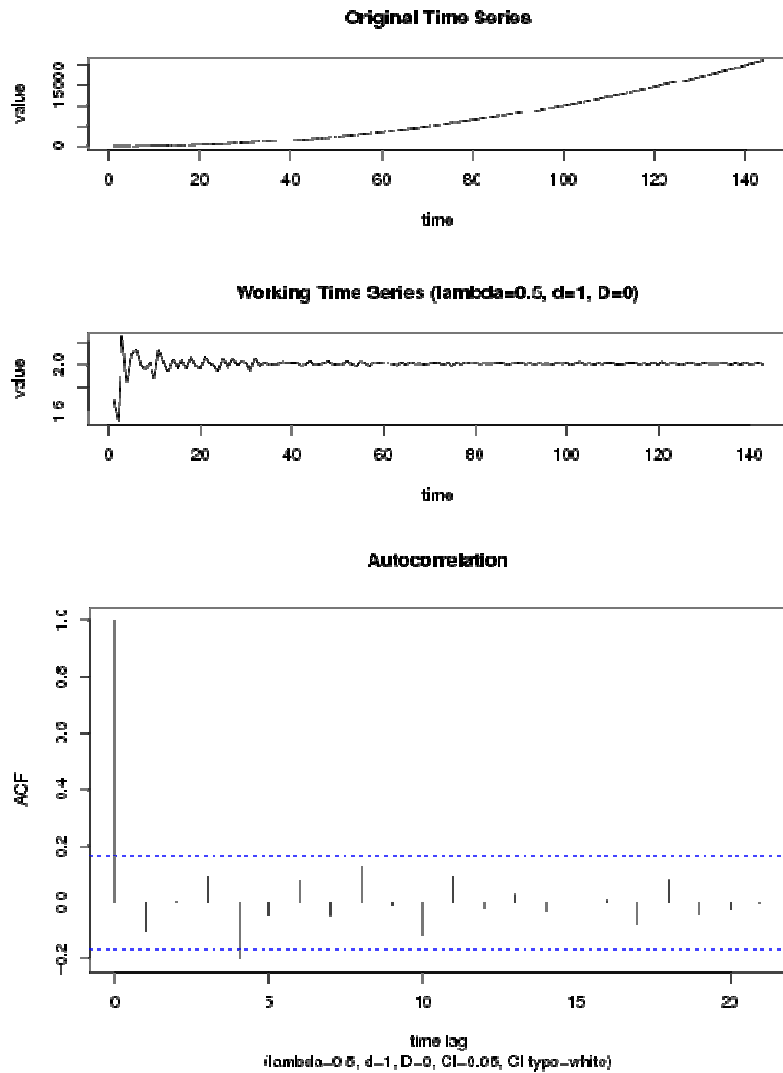
4. A négyzetgyökös transzformációt ($\lambda=1/2$) akkor használjuk, ha a hibatenyező varianciája arányos a várható értékkel. Ezt használjuk másodfokú parabolikus időtrendnél.

A mintapélda:

$$Y_t = 1,01 * t^2$$

$$(Y_t)^{1/2} = \sqrt{1,01} * t$$

$$Y(1/2) = (1,01^{1/2} - 1)/(1/2)$$



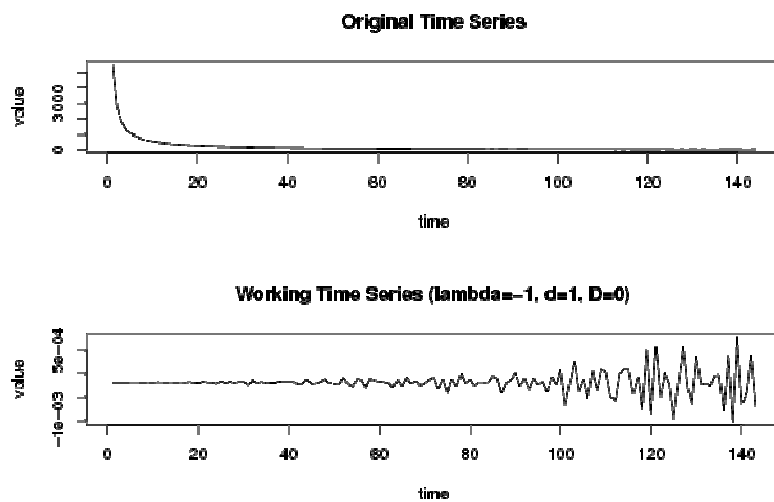
5. A reciprok transzformációt ($\lambda=-1$) akkor használjuk, ha a hibatenyező varianciája (σ^2) csökken, amikor a változó (Y) értéke csökken. Ezt használjuk hiperbolikus időtrendek linearizálásánál.

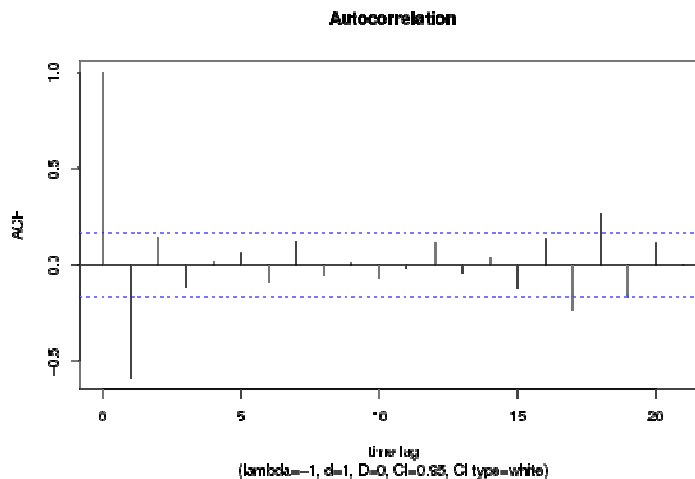
A mintapélda:

$$Y_t = 5500 * \frac{1}{t}$$

$$(Y_t)^{-1} = \frac{1}{5500} * t$$

$$Y(-1) = ([1/5500]^{-1} - 1)/(-1)$$





II A stacionaritás biztosítása havi adatok esetében, szezonális differencia képzés.

Egy további eset, amikor gyakran előfordul a stacionaritás hiánya: a szezonális. A periódusidő szerinti differenciálással a szezonális időeltolás mellett tapasztalható hatások, a periodikus mozgások szűrhetők ki. Negyedéves és havi idősorokban a stacionaritás hiánya gyakran eltüntethető a megfelelő differenciálással.

Negyedéves adatoknál:

$$\Delta^4 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) - (Y_{t-2} - Y_{t-3}) - (Y_{t-3} - Y_{t-4}) = Y_t - 2Y_{t-1} + Y_{t-4}$$

Havi adatoknál:

$$\Delta^{12} Y_t = Y_t - 2Y_{t-1} + Y_{t-12}$$

További probléma lehet, hogy az ingadozás intenzitása állandó-e vagy nem, ha nem ln-transzformációt lehet alkalmazni.

Ha szükséges, mindkét transzformáció elvégezhető, először a logaritmizálás illetve Box-Cox transzformáció, majd a transzformált adatok differencia-képzése. Az ARIMA számítások elvégzése után vissza kell transzformálni az adatokat. A szoftverek, pl. az ARIMA.xls vagy a GRETL elvégzi a differenciálást és a becslésnél visszaalakítja az adatokat eredeti formátumukba, de pl. az ln transzformált adatokat nem, mivel ezekkel számoltunk. Ilyenkor a becsült értékeket e hatványára kell emelni (Excelben kitévő (Y becsült)) Először a szezonális, esetünkben a havi differenciálást kell elvégezni, majd ezt követi ha szükséges az éves adatoknál már bemutatott differenciálás.

Szezonális differencia képzés:

$D=1$ ($y_t - y_{t-12}$), az idősor 12 adattal rövidül.

$D=2$ ($y_t - y_{t-24}$), az idősor 24 adattal rövidül.

Nem szezonális differencia képzés:

$d=1$ ($y_t - y_{t-1}$), az idősor 1 adattal rövidül.

$d=2$ ($y_t - y_{t-2}$), az idősor 2 adattal rövidül.

3.8.1 Az autokorrelációs- és parciális autokorrelációs együtthatók és az ACF és PACF értékek becslése és tesztelése

Az időben lejátszódó folyamatok mindegyike sztochasztikus folyamatként definiálható, mely valószínűségi változók sorozataként jelenik meg. Ezt *elméleti idősornak* nevezzük.

$Y_1, Y_2, \dots, Y_t, \dots, Y_T,$

Az Y_t ($t = 1, 2, \dots, T$) valószínűségi változók mindegyikére vonatkozóan egy megfigyeléssel rendelkezünk, ez a modellezés adatbázisát jelentő *tapasztalati idősor*,

$y_1, y_2, \dots, y_t, \dots, y_n$

melyet a sztochasztikus folyamatból vett n elemű mintának tekintünk.

Mind az elméleti idősort alkotó valószínűségi változóknak, mind a tapasztalati idősor különböző időpontokhoz (időtartamokhoz) tartozó megfigyelt értékeinek a felsorolása kötött. Az időszori sztochasztikus modellezés ezt, az adatok sorrendiségében rejlő információt használja fel az idősor jövőbeni értékeinek becslésére.

A megfigyelések sorrendjében rejlő információ leírásával, a tapasztalati idősorban lévő „szisztéma” megállapításával az elméleti idősor jellegzetességeire kívánunk következtetni, azaz arra a sztochasztikus folyamatra, amelyből a mintánk származik.

Az egymást követő megfigyelések között fennálló összefüggések megállapítása az idősorok korrelációs struktúrájának leírását jelenti, mely az autokorrelációs és a parciális autokorrelációs együtthatók számításával történik.

A mintából az **autokorrelációs együtthatók** becslése k késleltetéssel, a következőképpen történik:

$$r_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad k = 1, 2, \dots, K$$

A k késleltetés különböző értékeihez ($k = 1, 2, 3, \dots, K$) rendelt autokorrelációs együtthatók, az autokorrelációs függvényt alkotják:

k	1	2	3	...	K
r_1	r_2	r_3		r_K

Az autokorrelációs együtthatók becslött értékei, az Y idősor k időegységgel késleltetett értékei közötti lineáris korrelációs kapcsolat szorosságát mérik. Az r_1 az egymást követő, az r_2 , az egymástól két időegységre lévő értékek közötti kapcsolat intenzitását jelenti, stb. Az r_k együtthatók a késleltetés függvényében ($k = 1, 2, \dots, K$), az *autokorrelációs függvényt*, rövidítve az ACF-et (Autocorrelation function) alkotják. Az autokorrelációs függvény értékeit mátrixba foglalhatjuk:

$$\mathbf{R}_k = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{k-1} \\ r_1 & 1 & r_1 & \cdots & r_{k-2} \\ r_2 & r_1 & 1 & \cdots & r_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \cdots & 1 \end{bmatrix}$$

Az autokorrelációs együtthatók esetében tesztelhetjük, hogy vajon van-e kapcsolat az y_t és az y_{t-k} között. Hipotézisrendszerünk:

$$H_0 : r_{y_t y_{t+k}} = 0$$

$$H_1 : r_{y_t y_{t+k}} \neq 0$$

A nullhipotézis értelmében az y_t és az y_{t-k} változók között nincs szignifikáns autokorreláció, ennek elvetése az autokorrelációs kapcsolat szignifikáns voltát igazolja.

A becslött autokorrelációs együtthatóra épülő próbafüggvényünk:

$$|t| = \frac{r_{y_t y_{t+k}} \sqrt{n-2}}{\sqrt{1-r_{y_t y_{t+k}}^2}}$$

A nullhipotézis teljesülése esetén $(n-2)$ szabadságfokú kétoldali t -eloszlást követ.

A kapcsolat nem szignifikáns, 5%-os szignifikancia-szinten, tehát a H_0 -hipotézist elfogadjuk, ha:

$$|t| = \frac{r_{y_t y_{t+k}} \sqrt{n-2}}{\sqrt{1-r_{y_t y_{t+k}}^2}} < t_{0,025(n-2)}$$

A kapcsolat szignifikáns, 5%-os szignifikancia-szinten, tehát a H_1 -alternatív hipotézist fogadjuk el, ha:

$$|t| = \frac{r_{y_t, y_{t-k}} \sqrt{n-2}}{\sqrt{1-r_{y_t, y_{t-k}}^2}} > t_{0,025(n-2)}$$

Az idősor stacionaritásának vizsgálata történhet a tapasztalati idősorból számított ACF (autokorrelációs függvény= autocorrelation function) alapján, amennyiben a k késleltetés különböző értékeihez (k= 1,2,3,...,K) rendelt autokorrelációs együtthatók értékei lassan csökkennek, vagy majdnem lineárisan, ez indokolja a differenciaképzést. A megfelelő fokú differenciák elérését az autokorrelációs együtthatók gyors csökkenése jelzi. Ha az autokorrelációs együtthatók értékei a szezonális komponens hatásának megfelelően hullámoznak, akkor a szezonhatást először ki kell szűrni.

A parciális korrelációs együtthatók becslése:

1. Módszer. (Cramer szabály)

$$\hat{\Phi}_{kk} = \frac{|\mathbf{R}_k^*|}{|\mathbf{R}_k|}$$

Ahol:

K=1,2,...,K

\mathbf{R}_k^* úgy kapjuk meg, hogy az \mathbf{R}_k mátrix utolsó sorát (ld. az alábbi mátrixot), vagy oszlopát az (r_1, r_2, \dots, r_k) sorral (vektorral) helyettesítjük.

$$\mathbf{R}_k^* = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{k-1} \\ r_1 & 1 & r_1 & \cdots & r_{k-2} \\ r_2 & r_1 & 1 & \cdots & r_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_1 & r_2 & r_3 & \cdots & r_k \end{bmatrix}$$

A korrelációs együtthatók tehát, az \mathbf{R}_k^* és \mathbf{R}_k mátrix determinánsának hányadosaként határozhatók meg.

A k késleltetés különböző értékeihez (k= 1,2,3,...,K) rendelt autokorrelációs együtthatók, a parciális autokorrelációs függvényt alkotják:

k	1	2	3	...	K
$\hat{\Phi}_{kk}$	$\hat{\Phi}_{11}$	$\hat{\Phi}_{22}$	$\hat{\Phi}_{33}$	$\hat{\Phi}_{KK}$

A *parciális* autokorrelációs együtthatók ($\hat{\Phi}_k$) az idősor k időegységgel késleltetett értékei közötti lineáris korrelációs kapcsolat szorosságát méri úgy, hogy a közbenső, 1, 2, ... k-1 késleltetések hatását kiszűrjük. A $\hat{\Phi}_k$ együtthatók a késleltetés függvényében (k = 1, 2, ..., K), a *parciális autokorrelációs függvény*t, rövidítve a PACF-et (Partial autocorrelation function) alkotják.

Speciálisan:

$$\text{Ha } k=1 \quad \hat{\Phi}_{11} = r_{y_t, y_{t-1}} = r_1$$

$$\text{Ha } k=2 \quad \hat{\Phi}_{22} = r_{y_t, y_{t-2} | y_{t-1}} = \frac{|\mathbf{R}_2^*|}{|\mathbf{R}_2|} = \frac{r_2 - r_1^2}{1 - r_1^2}$$

Ahol: $\mathbf{R}_2 =$

1	r_1
r_1	1

és $\mathbf{R}_2^* =$

1	r_1
r_1	r_2

A determináns a diagonálisok szorzatának különbsége, mivel a mátrix 2*2-es.

2. Módszer. (Durbin - féle rekurzív eljárás)

Ha k=1

$$\hat{\Phi}_{11} = r_1$$

ha: k > 1

$$\hat{\Phi}_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \hat{\Phi}_{k-1,j} * r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\Phi}_{k-1,j} * r_j}$$

$$\hat{\Phi}_{k,j} = \hat{\Phi}_{k-1,j} - \hat{\Phi}_{k,k} \hat{\Phi}_{k-1,k-j}$$

$$j = 1, 2, \dots, k-1$$

Például:

$$\hat{\Phi}_{22} = \frac{r_2 - \hat{\Phi}_{11} * r_1}{1 - \hat{\Phi}_{11} * r_1} = \frac{r_2 - r_1^2}{1 - r_1^2}$$

ha: k = 2

$$\hat{\Phi}_{11} = r_1$$

$$\hat{\Phi}_{33} = \frac{r_3 - (\hat{\Phi}_{21} * r_2 + \hat{\Phi}_{22} * r_1)}{1 - (\hat{\Phi}_{21} * r_1 + \hat{\Phi}_{22} * r_2)}$$

ha: k = 3

$$\hat{\Phi}_{21} = \hat{\Phi}_{11} - \hat{\Phi}_{22} \hat{\Phi}_{11} = r_1 - \hat{\Phi}_{22} * r_1$$

Látható, hogy ez az eljárás a parciális autokorrelációs együtthatókat, az alacsonyabb rendű folyamatra már kiszámított parciális autokorrelációs együtthatók segítségével állítja elő.

3. Módszer. (A Yule-Walker egyenletekkel történő becslés¹³⁸.)

$$\hat{\Phi} = \mathbf{R}^{-1} \mathbf{r} = \begin{bmatrix} 1 & r_1 & r_2 & \dots & r_{k-1} \\ r_1 & 1 & r_1 & \dots & r_{k-2} \\ r_2 & r_1 & 1 & \dots & r_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_1 \\ \cdot \\ \cdot \\ r_k \end{bmatrix}$$

i=1,2,3...k késleltetett értékekre külön-külön ki kell számolni a szorzatokat ($\mathbf{R}^{-1} \mathbf{r}$) és a parciális autokorrelációs együttható (k=1,2...K) a számított vektor $\hat{\Phi}$ utolsó eleme lesz.

Az Excelben mind a három módszer programozható, de a harmadik a legkényelmesebb és ezért ezt használtuk a parciális autokorrelációs együtthatók becslésére.

A tapasztalati idősor ACF és PACF értékei alapján azonosítható a sztochasztikus folyamat típusa, amely egyben kijelöli a választandó modell típusát. Az EXCEL rendelkezik determinánst meghatározó programmal.

Ha egy stacionárius idősor ACF és PACF értékei nem különböznek szignifikánsan nullától, az idősorban nem található „szisztéma”, az idősorunk egy olyan egyszerű véletlen folyamatként modellezhető, melyben az idősor értékei egy konstans várható érték (μ) körül véletlenszerűen ingadoznak.

$$Y_t = \mu + \varepsilon_t$$

ε =epszilon

Ha $\mu = 0$, akkor a véletlen folyamatot ún. *fehér zajnak* (white noise) nevezzük, mely a legegyszerűbb sztochasztikus folyamat.

$$Y_t = \varepsilon_t$$

A fehér zaj folyamat jellemzői:

1. $E(\varepsilon_t) = 0$
2. $\text{Var}(\varepsilon_t) = \sigma^2$
3. $\rho_k = 0 \quad k = 1, 2, \dots$

A fenti három követelményt teljesítő változó „teljesen” véletlen folyamatot követ, és a modellezésben ilyen értelemben használjuk az ε_t véletlen változót. A fehér zaj folyamat tehát, nulla várható értékű, állandó szórású és korrelálatlan változókból álló stacioner sztochasztikus folyamat. Feltételezzük továbbá, hogy az idősor adatai normális eloszlást követnek.

Az ACF és PACF értékek szignifikanciájának tesztelését a fehér zaj folyamat r_k autokorrelációs együtthatóinak ismert mintaeloszlása alapján végezhetjük el. Elméletileg a „teljesen” véletlen ε_t , fehér zajt követő változók minden autokorrelációs együtthatójának nullának kellene lennie. Véges mintából becsülve, nem számíthatunk arra, hogy minden ACF és PACF érték zérus lesz. Bizonyítható, hogy a fehér zaj folyamat autokorrelációs együtthatóinak mintaeloszlása nulla várható értékű és $(1/\sqrt{n})$ szórású normális eloszlást követ. Ezért az ACF és PACF értékeknek, a késleltetés függvényében készített grafikus ábráján, – melyet korrelogramnak nevezünk –, a 95 %-os valószínűségi szinthez tartozó $\pm 1,96/\sqrt{n}$ hibahatárt is fel szokták tüntetni. A gyakorlati számításokban általában a $\pm 2/\sqrt{n}$ képlettel számolnak. A korrelogram így közvetlenül alkalmassá válik az autokorrelációk zérus voltára vonatkozó nullhipotézis tesztelésére, 5 %-os szignifikancia szinten. A $\pm 2/\sqrt{n}$ hibahatár (+2Se: -2Se:) által meghatározott sávon belüli autokorrelációs együtthatók 5 %-os szignifikancia szinten, zérusnak tekinthetők. A sávon kívüli, azaz nullától szignifikánsan különböző autokorrelációs együtthatók, „szisztéma” jelenlétére utalnak az idősorban, tehát meg kell keresni az alkalmas ARMA modellt.

Ha az idősor első differencia sora egy nem nulla várható érték körül állandó varianciával ingadozik, és az ACF és PACF értékei nem különböznek szignifikánsan a zérustól, akkor a az idősor ún. *véletlen bolyongási folyamatot* követ:

$$Y_t - Y_{t-1} = \mu + \varepsilon_t \quad \text{vagy átalakítva} \quad Y_t = Y_{t-1} + \mu + \varepsilon_t$$

A fenti modell a véletlen bolyongási folyamat azon változata, mely a lineáris trend sztochasztikus megfelelője. E szerint az idősor értéke egyik időszakról a másikra egy állandó-, és egy véletlen értékkel változik. A véletlen bolyongási modell magasabb fokú differenciákra is felírható.

A stacionárius, vagy azzá transzformált idősorok ACF és PACF értékei általában tartalmaznak nullától szignifikánsan különböző értékeket, és ez az idősorban valamilyen „szisztéma” jelenlétére utal. A „szisztéma” megkeresése úgy történik, hogy a különböző típusú ARMA folyamatok közül kiválasztjuk azokat, amelyeknek az ismert elméleti ACF és PACF sémájára leginkább hasonlítanak a vizsgált idősor tapasztalati autokorrelációs és parciális autokorrelációs együtthatói. A kiválasztott ARMA folyamatoknak megfelelő modellek lesznek azok, melyek illesztésével megpróbálkozunk.

A különböző típusú ARMA folyamatok elméleti ACF és PACF értékeinek alakulását rendszerezve, - a gyakorlati alkalmazásokhoz könnyen használható formában -, a szakirodalom és a számítógépes szoftverleírások is tartalmazzák. A normalitást általában a Jarque-Bera féle teszttel ellenőrzik.

Autokorrelált hibák Ha a modell hibatagjai autokorrelációt mutatnak, ez azt jelzi, hogy a modell nem távolított el minden sémát az adatokból. Sok hipotézis ellenőrzés van az autokorrelációs hibák tesztelésére. A Box – Pierce -, és a Ljung-Box - teszt ellenőrzi, hogy az autokorrelációk sorozata szignifikánsan különbözik-e zérusok sorozatától; a Durbin - Watson teszt csak az elsőrendű autokorrelációkat ellenőrzi a regressziós modell illesztése után.

A szoftverek általában, így az XLSTAT-TIME a „leíró statisztika” (Descriptive analysis) menüpontban számítják ki az ARIMA modellezéshez szükséges statisztikákat (ACF, PACF értékek és ábrák, korrelogramok és hibahatárok, Jarque-Bera féle teszt, Box – Pierce -, a Ljung-Box - teszt).

Az autokorrelációs együtthatók (Autocorrelation) (r_k) esetében a standard hiba (Standard error) ($s(r_k)$) a következő képlettel közelíthető:

$$s(r_k) = \sqrt{\frac{1}{n} [1 + 2(r_1^2 + r_2^2 + r_3^2 + \dots + r_{k-1}^2)]}$$

$$k = 1, 2, 3, \dots, K$$

A konfidencia-alsó és felső hibahatár (Lower bound (95%)-Upper bound (95%)) által meghatározott sáv kiszámítása 5 %-os szignifikancia szinten az alábbi képlettel történik:

$$\pm 1,96s(r_k)$$

A parciális korrelációs együtthatóknál (p_k) (Partial autocorrelation) a standard hiba (Standard error) ($s(p_k)$):

$$s(p_k) = 1/\sqrt{n}$$

A konfidencia-alsó és felső hibahatár (Lower bound (95%)-Upper bound (95%)) által meghatározott sáv kiszámítása 5 %-os szignifikancia szinten az alábbi képlettel történik:

$$\pm 1,96/\sqrt{n}$$

A $\pm 1,96/\sqrt{n}$ hibahatár, illetve $\pm 2/\sqrt{n}$ (+2Se: -2Se:) által meghatározott sávon belüli autokorrelációs együtthatók 5 %-os szignifikancia szinten, zérusnak tekinthetők.

Box - Pierce tesztstatisztika (Q-teszt).

Ez az autokorrelált hibák ún. Q-tesztje. Ha a modell hibái fehér zajt alkotnak, akkor a Box - Pierce statisztika közelítőleg χ^2 - eloszlású.

Box - Pierce tesztstatisztika kiszámítása.¹³⁹

A mintából a Q-teszt becslése a következőképpen történik:

$$Q = n \sum_{k=1}^{k=K} r_k^2$$

Ahol:

r_k = a ϵ_t reziduumok k-ad rendű autokorrelációs együtthatója,

n = megfigyelések száma,

K = a számított autokorrelációs együtthatók előre megválasztott száma, pl. 21 vagy több.

Ha a reziduumok sora fehér zaj, akkor a Q χ^2 -eloszlást követ (K-p-q) vagy (K) szabadságfokkal. Ha a Q számított értéke nagyobb, mint a χ^2 -eloszlás kritikus értéke, akkor arra a következtetésre jutunk, hogy a reziduumok nem fehér zajok. ($Q >_{K-p-q} \chi^2_{0,05}$) Fordított esetben elfogadjuk a nullhipotézist, hogy a reziduumok fehér zajok ($Q <_{K-p-q} \chi^2_{0,05}$).

Ennek alapján a hipotézis rendszer:

H_0 = a reziduumok fehér zajok,

H_1 = a reziduumok nem fehér zajok,

A H_0 -t elfogadjuk, ha:

$$Q <_{K-p-q} \chi^2_{0,05}$$

A H_1 alternatív hipotézist fogadjuk el, ha:

$$Q >_{K-p-q} \chi^2_{0,05}$$

A szignifikancia-érték (p-érték) az a legkisebb szignifikancia szint, amin a H_0 már éppen elvethető a H_1 -gyel szemben. Ha pl. 0,05-nél kisebb a p-érték akkor 5 %-os szignifikancia szinten elutasítjuk a nullhipotézist, miszerint a reziduumok fehér zajok.

Használatos a Q-teszt, másik formája is, ahol a d-fokát is figyelembe vesszük:

$$Q^1 = (n - d) \sum_{k=1}^{k=K} r_k^2$$

Ahol:

n-d, az idősor d számú differenciálása után felhasználható megfigyelések száma.

A kritikus érték: $_{K-p-q} \chi^2_{0,05}$

Döntés az előző módon.

¹³⁹ Ramanathan Ramu [2003]: Bevezetés az ökonometriába alkalmazásokkal. i. m. 542-543. és Greene, William H. [2003]: Econometric analysis. 271.

A másik teszt a Ljung - Box teszt (Q*), amely a Box – Pierce Q-teszt továbbfejlesztett változata.

A Ljung - Box portmanteau-próba (LJB vagy Q*-teszt)¹⁴⁰:

$$Q^* = n(n+2) \sum_{k=1}^{k=K} \left[\frac{r_k^2}{n-k} \right]$$

Ahol:

k = a számított autokorrelációs együtthatók előre meghatározott száma, pl. 24.

n = a megfigyelések száma.

H₀ = a reziduumok fehér zajok,

H₁ = a reziduumok nem fehér zajok,

Q* χ^2 -eloszlást követ (K-p-q) vagy (K) szabadságfokkal.

Ennek alapján a hipotézis rendszer:

A H₀-t elfogadjuk, ha:

$$Q^* <_{K-p-q} \chi^2_{0,05}$$

A H₁ alternatív hipotézist fogadjuk el, ha:

$$Q^* >_{K-p-q} \chi^2_{0,05}$$

A szignifikancia-érték (p-érték) az a legkisebb szignifikancia szint, amin a H₀ már éppen elvethető a H₁-gyel szemben. Ha pl. 0,05-nél kisebb a p-érték akkor 5 %-os szignifikancia szinten elutasítjuk a nullhipotézist, miszerint a reziduumok fehér zajok.

Használatos az LJB vagy Q*-teszt, másik formája is, ahol a d-fokát is figyelembe veszik:

$$Q^{*1} = n'(n'+2) \sum_{k=1}^{k=K} \left[\frac{r_k^2}{n'-k} \right]$$

Ahol:

n' = n-d, az idősor d számú differenciálása után felhasználható megfigyelések száma.

A tesztelési eljárás az előzőhöz hasonló módon történik.

Maddala véleménye szerint a Q-próbáknál vannak jobb eljárások, de nagy K-érték mellett használata megfelelő eredményt adhat.

3.8.2 Az ARIMA modell azonosítása

A következő kérdés annak megválaszolása, hogy milyen típusú ARMA modell illesztésével próbálkozunk, illetve, milyen legyen az autoregresszivitás (p) és/vagy, a mozgóátlagolás (q) rendje. Erre a kérdésre a választ a tapasztalati, vagy a transzformált idősor ACF és PACF értékei alapján adjuk meg. A modellezés ezen fázisát, *modell azonosításnak (identifikációnak)* nevezi a szakirodalom. A mintából becsült autokorrelációs és parciális autokorrelációs együtthatók grafikus ábrája, a korrelogram (ACF és PACF) alapján lehet a legkönnyebben az autokorrelációs együtthatók viselkedését - a késleltetés (k) függvényében - tanulmányozni. Ugyanis a r_k becsült autokorrelációs együtthatók konfidencia intervalluma alapján közvetlenül megállapíthatók a nullától szignifikánsan különböző r_k értékek. Ezek a konfidencia sávon kívül helyezkednek el.

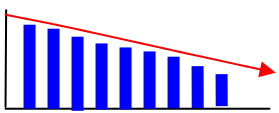

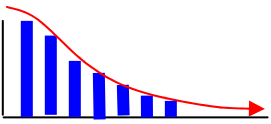

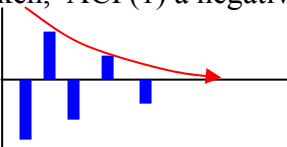
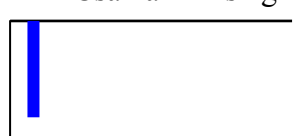
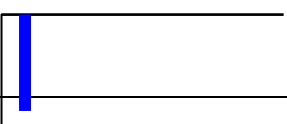
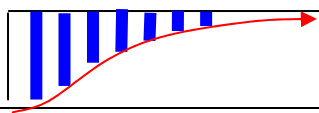
ARIMA modellek jellemzése


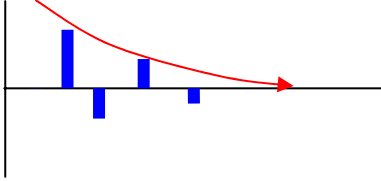
Modelltípus ARIMA (p,d,q)	Autokorrelációs együtthatók ACF (p _k)	Parciális autokorrelációs együtthatók (Φ _{kk})
(1, d, 0) AR(1)	Exponenciálisan csökken, ha ρ ₁ > 0, csillapodó szinusz görbe szerint csökken, ha ρ ₁ < 0	φ ₁ ha k = 1 0 ha k > 1
(2, d, 0) AR(2)	Exponenciálisan és/vagy csillapodó szinusz görbe szerint csökken	φ _{11} ha k = 1 φ_{22} ha k = 2 0 ha k > 2}}

¹⁴⁰ Ramanathan Ramu [2003]: Bevezetés az ökonometriába alkalmazásokkal. i. m. 542-543. és Maddala G. S. [2004]: 592-594.

(0, d, 1) MA(1), ha d = 0 vagy IMA (d, 1)	ρ_1 ha $k = 1$ 0 ha $k > 1$	Exponenciálisan, vagy csillapodó szinusz görbe szerint csökken
(0, d, 2) MA(2), ha d = 0 vagy IMA (d, 2)	ρ_1 ha $k = 1$ ρ_2 ha $k = 2$ 0 ha $k > 2$	Exponenciálisan és/vagy csillapodó szinusz görbe szerint csökken
(1, d, 1) ARMA (1, 1) ha d = 0, vagy ARIMA (1, d, 1)	Exponenciálisan, vagy csillapodó szinusz görbe szerint csökken a második értéktől kezdődően	Exponenciálisan, vagy csillapodó szinusz görbe szerint csökken a második értéktől kezdődően
(1, d, 2) ARMA (1, 2) ha d = 0 vagy ARIMA (1, d, 2)	$\rho_0 = 1$ és ρ_1 után exponenciálisan csökken	Exponenciálisan és/vagy csillapodó szinusz görbe szerint csökken
(2, d, 1) ARMA (2, 1) ha d = 0 vagy ARIMA (2, d, 1)	Exponenciálisan és/ vagy csillapodó szinusz görbe szerint csökken	$\phi_{11} = \rho_1$ és ϕ_{22} után exponenciálisan csökken
(2, d, 2) ARMA (2, 2) ha d = 0 vagy ARIMA (2, d, 2)	$\rho_0 = 1$ után exponenciálisan csökken	$\phi_{11} = \rho_1$ után exponenciálisan és/vagy csillapodó szinusz görbe szerint csökken

A modellválasztásban az alábbi táblázat is segít:

Modelltípus ARIMA (p,d,q)	Autokorrelációs együtthatók ACF (ρ_k)	Parciális autokorrelációs együtthatók (Φ_{kk})
ARIMA (0,0,0)	Nem szignifikáns –mindegyik k-ra	Nem szignifikáns –mindegyik k-ra
ARIMA (0,1,0)	Lineárisan csökken, mindegyik k- ra szignifikáns 	Csak a k=1 szignifikáns 
ARIMA (1,0,0) $1 > \Phi > 0$	Exponenciálisan csökken a k=1 és esetleg a k=2 szignifikáns 	Csak a k=1 szignifikáns 
ARIMA (1,0,0) $-1 < \Phi < 0$	Exponenciálisan kétoldalúan csök- ken, ACF(1) a negative csúcs 	Csak a k=1 szignifikáns 
ARIMA (0,0,1) $1 > \theta > 0$	Csak a k=1 szignifikáns, negative csúcs 	Exponenciálisan csökken, az első kettő k, vagy több is szignifikáns. 

ARIMA (0,0,1) $-1 < \theta < 0$	Csak a $k=1$ szignifikáns, pozitív csúcs 	Exponenciálisan kétoldalúan csökken, PACF(1) a pozitív csúcs 
---	---	--

3.8.3 Az ARIMA modellek becslése

Ezután a modellezés lépései alapvetően megfelelnek a már ismert lineáris regressziós modellezésnek. A választott *modell paraméterbecslése* után a *modell ellenőrzése* következik. A modell ellenőrzése során vizsgáljuk azt, hogy paraméterei szignifikánsak-e, illetve véletlen változók fehér zaj folyamatot követnek-e. Ezután döntünk arról, hogy felhasználható-e az illesztett modell *elemzésre, előrejelzésre*, vagy más modell választásával kell próbálkoznunk.

A modell ellenőrzésére a szokásos eljárások alkalmazhatók:

- a becsült paraméterek standard hiba számítása és szignifikancia vizsgálata (pl. t-próbával),
- az e_t tapasztalati reziduumok alapján az ε_t véletlen változók véletlen jellegének vizsgálata.

Mindezek mellett speciális tesztelési eljárások alkalmazására is sor kerül, amelyeket a számítógépes programok is tartalmaznak. Amennyiben a választott és számszerűsített modellünk megfelel mindazon feltételeknek, melyekkel az illesztett modell "jóságát" ellenőrizhetjük, a modell felhasználható elemzésre és a tulajdonképpeni legfontosabb felhasználási területére, az előrejelzések készítésére. Ha modellünk nem felel meg a fenti feltételeknek (nem szignifikánsak a paraméterei, vagy az ε_t időSORA nem véletlenszerűen alakul) a modellazonosítás fázisától újra indulva, más modell típusok alkalmazásával próbálkozhatunk. (használható a regresszio.xls parancsfájl.)

Az ARIMA modellek igen széles választékából, most csak az alacsonyabb rendű tiszta, valamint vegyes modellek legfontosabb jellemzőit ismertetjük.

Az *első-* ($p=1$) és *másodrendű* ($p=2$) *autoregresszív modell* felírható az alábbi formában:

ARIMA (1, 0, 0) vagy AR (1) modell
$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

ARIMA (2, 0, 0) vagy AR (2) modell
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

Az autoregresszív folyamat mindenkor értéke kifejezhető saját késleltetett értékeinek lineáris kombinációja és egy fehér zaj folyamat összegeként.

A stacionaritási feltétel teljesülése érdekében az autoregresszív paraméterekre speciális korlátok érvényesek. $p=1$ esetén $|\phi_1| < 1$, míg $p=2$ esetén a következő három feltételt kell kielégíteni:

$$|\phi_2| < 1 \quad \phi_2 + \phi_1 < 1 \quad \phi_2 - \phi_1 < 1$$

Általában az AR (p) folyamatok elméleti ACF értékei $p \geq 2$ esetén exponenciális csökkenés és/vagy csillapodó szinusz görbe szerint alakulnak, a ϕ_1 és ϕ_2 paraméterek előjelétől függően. Az AR (1) folyamat elméleti ACF értékei exponenciálisan csökkennek, ha ϕ_1 előjele pozitív, és csillapodó szinusz görbe szerint csökkennek, ha ϕ_1 negatív.

Az AR (p) folyamatok elméleti PACF értékei p késleltetés után zérusok, tehát $p=1$ esetén csak az első, $p=2$ esetén az első kettő parciális autokorreláció nem nulla.

A két legegyszerűbb sztochasztikus modell, nevezetesen a fehér zaj, illetve a véletlen bolyongási modell, az autoregresszív modellek speciális eseteként is felírható. Ha $\phi_1 = 0$, az Y_t értékei fehér zaj folyamatot követnek, melyet ARIMA (0,0,0) modellnek lehet tekinteni. Ha $\phi_1 = 1$, az Y_t értékei véletlen bolyongás szerint alakulnak, akkor a folyamatot ARIMA (0,1,0)-ként lehet osztályozni.

Az *első-* ($q=1$) és *másodrendű* ($q=2$) *mozgóátlag modell* felírható az alábbi formában:

ARIMA (0, 0, 1) vagy MA (1) modell

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

ARIMA (0, 0, 2) vagy MA (2) modell

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

A mozgóátlag folyamat mindenkori értéke kifejezhető különböző késleltetésű fehér zajok lineáris kombinációjaként. A negatív előjelezést konvencionálisan használják a mozgóátlag folyamatoknál.

Az invertibilitási feltétel teljesülése érdekében az mozgóátlag paraméterekre is ugyanazon speciális korlátok érvényesek, mint amelyek az autoregresszív modellek vonatkoznak. $q=1$ esetén $|\theta_1| < 1$, míg $q=2$ esetén a következő három feltételt kell kielégíteni: $|\theta_2| < 1$ $\theta_2 + \theta_1 < 1$ $\theta_2 - \theta_1 < 1$

Az ACF és PACF sémája pontosan a fordítottja annak, amit az autoregresszív folyamatoknál láttunk.

Az MA (q) folyamatok elméleti ACF értékei q késleltetés után zérusok, tehát $q=1$ esetén csak az első, $q=2$ esetén csak az első kettő autokorreláció nem nulla.

Az MA (q) folyamatok elméleti PACF értékei $q \geq 2$ esetén exponenciális csökkenés és/vagy csillapodó szinusz görbe szerint alakulnak, a θ_1 és θ_2 paraméterek előjelétől függően. Az MA (1) folyamat elméleti ACF értékei exponenciálisan csökkennek, ha θ_1 előjele pozitív, és csillapodó szinusz görbe szerint csökkennek, ha θ_1 negatív.

Az AR és MA modellek kombinálásával a modellek igen sok variációja állítható elő. Az alacsonyabb rendű *vegyes ARMA modellek* az alábbi módon írhatók fel:

$$\text{ARIMA (1, 0, 1)} \quad Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$\text{ARIMA (2, 0, 1)} \quad Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$\text{ARIMA (1, 0, 2)} \quad Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

$$\text{ARIMA (2, 0, 2)} \quad Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

Az autoregresszív mozgóátlag folyamat mindenkori értéke kifejezhető saját késleltetett értékeinek és különböző késleltetésű fehér zajok lineáris kombinációja összegeként.

Amennyiben a vegyes modellek valamelyikét az idősor differenciáira írjuk fel, ARIMA (p, d, q) modellhez jutunk. A legegyszerűbb ARIMA (1, 1, 1) modell az alábbi módon írható fel:

$$Y_t - Y_{t-1} = \phi_1 (Y_{t-1} - Y_{t-2}) + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

A paraméterekre vonatkozó megszorítások és az elméleti ACF, PACF sémák általánosan a vegyes modellekre vonatkoznak, mivel függetlenek a differenciális fokától.

A vegyes modellek paramétereire vonatkozó megszorítások megegyeznek a modellek tiszta AR és MA részeire megállapítható korlátozásokkal.

Az elméleti ACF és PACF sémák is nagyon hasonlóak a tiszta AR és MA modellekre jellemzőkhöz.

A szezonális ARIMA (p, d, q) (P, D, Q)_s modellek, a szezonális ingadozást is tartalmazó idősorokban meglévő kettős függőségi rendszer leírására két ARIMA modell építenek egymásra. Az egymás után következő időszori értékek összefüggését a modell (p, d, q) dimenzióival rendelkező része mutatja, hasonlóan a szezonális nem tartalmazó modellekhez. Az egyes évek azonos szezonjai közötti összefüggéseket a modell ún. szezonális része képviseli, (P, D, Q)_s dimenziókkal, ahol s a szezonok számát jelenti. A szezonális kezelést az egyik leggyakrabban alkalmazott ARIMA (0, 1, 1)₁₂ modell példáján mutatjuk be. Az egyenlet bal oldalán a „kétszeres” differenciaképzést úgy végezzük, hogy először a D=1 szezonális első differenciákat a különböző évek azonos hónapjainak adatai alapján számítjuk, és így s=12 adattal (az első év teljes adatsorával) rövidül az adatsorunk. Ezután újabb d=1 első differenciákat számítunk, most az egymás után következő szezonális differenciákból, így egy további adattal rövidül az adatsorunk. A „kétszeres” differenciaképzés következtében összesen (d+sD) (estünkben 13) taggal rövidül az adatsorunk. Az egyenlet jobb oldalán „kétszeres” mozgóátlag folyamatot írunk fel az ε_t véletlen változóra. Először a k=1 késleltetésnek megfelelően a θ paraméterrel, majd ebből az s=12 szezonális késleltetésű véletlen változóra Θ paraméterrel.

$$(Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13}) = \varepsilon_t - \theta \varepsilon_{t-1} - \Theta (\varepsilon_{t-12} - \theta \varepsilon_{t-13})$$

A magasabb rendű modellek, és különösen a szezonális modellek, a fenti módon már igen nehezen kezelhetők, ezért általában az ARIMA modelleket az ún. *operátor* jelölésmóddal szokás felírni. A késleltető (visszaléptető) operátort, B-t, a következőképpen használjuk: $BY_t = Y_{t-1}$

A B művelet hatása Y_t -re, az adat visszaléptetése egy periódussal. A B művelet kétszeres alkalmazása Y_t -re, két periódussal lépteti vissza az adatot: $B(BY_t) = B^2Y_t = Y_{t-2}$

Havi adatok esetén az előző év azonos hónapjának adata a B^{12} jelöléssel érhető el, $B^{12}Y_t = Y_{t-12}$.

A differencia képzés egyszerűen leírható a B operátor segítségével. Például az elsőfokú differenciaképzés a következőképpen jelölhető: $Y_t - Y_{t-1} = Y_t - BY_t = (1-B)Y_t$, ahol $(1-B)$ jelöli az első differenciát. Hasonlóan a másodfokú differenciákat (az első differenciák differenciáit) az alábbi módon jelölhetjük:

$$(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2} = (1 - 2B + B^2)Y_t = (1 - B)^2 Y_t$$

Általánosan a d-ed fokú differencia a következőképpen írható: $(1 - B)^d Y_t$.

A szezonális differenciák első differenciáinak jelölése a következő:

$$(1 - B)(1 - B^s)Y_t = (1 - B - B^s + B^{s+1})Y_t = Y_t - Y_{t-1} - Y_{t-s} + Y_{t-s-1}$$

Az ARIMA (0, 1, 1) (0, 1, 1)₁₂ modell az operátor jelölésmóddal felírva a következő:

$$(1 - B)(1 - B^{12})Y_t = (1 - \theta B)(1 - \Theta B^{12})\varepsilon_t$$

$$(1 - B)(1 - B^{12})Y_t = \theta(B)\Theta(B^{12})\varepsilon_t$$

Az általános ARIMA (p, d, q) (P, D, Q)_s modell operátorokkal:

$$\Phi_p(B)\Phi_P(B^s)(1 - B)^d(1 - B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

Néhány gyakran alkalmazott szezonális ARIMA modell elméleti ACF sémáját szakirodalmi leírás alapján közöljük Ábrahám, B. – Ledolter, J. (1986) p.

1. modell: (0, d, 0) (0, D, 1)₁₂ $Y_t = \Theta(B^{12})\varepsilon_t$

Szignifikáns ACF a ρ_{12} , azaz a k=12 késleltetésű autokorrelációs együttható.

2. modell: (0, d, 0) (1, D, 0)₁₂ $\Phi(B^{12})Y_t = \varepsilon_t$

Szignifikáns ACF a ρ_{12} , ρ_{24}, \dots , exponenciálisan, vagy csillapodó szinusz görbe szerint csökkenve.

3. modell: (0, d, 0) (1, D, 1)₁₂ $\Phi(B^{12})Y_t = \Theta(B^{12})\varepsilon_t$

A $\rho_1 = 1$, és szignifikáns ACF a ρ_{12} , ρ_{24}, \dots , ρ_{36}, \dots , exponenciálisan, vagy csillapodó szinusz görbe szerint csökkenve.

4. modell: (0, d, 1) (0, D, 1)₁₂ $Y_t = \theta(B)\Theta(B^{12})\varepsilon_t$

Szignifikáns ACF a ρ_1 , ρ_{11} , ρ_{12} , és a ρ_{13} , ($\rho_{11} = \rho_{13}$). Előjelük pozitív és negatív is lehet a modell paraméterek előjelétől függően.

5. modell: (0, d, 1) (1, D, 0)₁₂ $\Phi(B^{12})Y_t = \theta(B)\varepsilon_t$ Szignifikáns ACF a $\rho_1, \rho_{11}, \rho_{12}, \rho_{13}$ ($\rho_{13} = \rho_{11}$); $\rho_{23}, \rho_{24}, \rho_{25}$, ($\rho_{25} = \rho_{23}$); $\rho_{35}, \rho_{36}, \rho_{37}$, ($\rho_{37} = \rho_{35}$); exponenciálisan, vagy csillapodó szinusz görbe szerint csökkenve.

6. modell: (0, d, 1) (1, D, 1)₁₂ $\Phi(B^{12})Y_t = \theta(B)\Theta(B^{12})\varepsilon_t$ Szignifikáns ACF a $\rho_1, \rho_{11}, \rho_{12}$, és a ρ_{13} , ($\rho_{11} = \rho_{13}$). Egyébként az 5. modell szerint alakul.

7. modell: (0, d, 2) (0, D, 1)₁₂ $Y_t = \theta_2(B)\Theta(B^{12})\varepsilon_t$ Szignifikáns ACF a $\rho_1, \rho_2, \rho_{10}, \rho_{11}, \rho_{12}, \rho_{13}$, ($\rho_{13} = \rho_{11}$), ρ_{14} , ($\rho_{14} = \rho_{10}$);

A szezonális modellek PACF sémájáról általánosan elmondható, hogy a szezonális és nem szezonális mozgó átlagolású komponens behozza az exponenciális és csillapodó szinusz görbe szerinti csökkenést, a szezonális és nem szezonális késleltetésnél is. Az autoregresszív folyamatok PACF-je pedig véges sok értéket tartalmaz.

JMulti ingyenes, bonyolult sztochasztikus idősorkutatási módszereket (ARCH, ARIMA, VAR, VECM, stb) becselő szoftver:

<http://www.jmulti.de/>

JMulti egy nyílt forráskódú interaktív szoftver, ami az ökonometriai elemzés és a többváltozós idősorok elemzése céljából készült. Ez egy Java grafikus felhasználói felület.

Statisztikai programcsomagok összehasonlítása:

http://en.wikipedia.org/wiki/Comparison_of_statistical_packages

AdaMSoft ingyenes szoftver:

<http://sourceforge.net/projects/adamsoft/files/ADaMSoft/3.16.1/InstallADaMSoft.jar/download>

ARIMA-t becselő statisztikai szoftverek

Product	Ár
BMDP	\$1095
EViews	\$1075
GRETL	Free
JMP	\$1895
Mathematica	\$1095
Minitab	\$1395
NumXL	Lite version (Free), Professional edition (\$300)
R	Free
RATS	\$500
Sage	Free
SAS	\$6000
SHAZAM	\$1600
Stata	\$595
Statgraphics	\$1495
STATISTICA	\$695
StatPlus	\$150
SPSS	\$1599
SYSTAT	\$1299
TSP	\$1200
UNISTAT	\$895
YMulti	Free

EXCEL megoldások

<http://forecast.umkc.edu/ftppub/BDS545/>

3.8.4 EXCEL-parancsfájlok az ARIMA modellezés témaköréből.

Az ARIMA modellezés három lépésben, három Excel parancsfájl használatával történik.

Először a Stacionaritás-biztosítása.xlsm parancsfájl alkalmazásával megvizsgáljuk, hogy stacionárius-e az idősor vagy nem. Ha nem akkor a differenciálás fokának változtatásával illetve Box-Cox transzformációval kiválasztjuk azt a transzformált idősort, ami grafikusan leginkább eleget tesz a stacionaritás követelményeinek. Itt csak a grafikus ábra megszemlélésére illetve vizsgálatára van lehetőség, pl. látható, ha az eredeti vagy transzformált idősor átlaga és szórása konstans-e vagy nem. A következő lépés a tesztelés.

Másodszor a kiválasztott eredeti vagy transzformált idősort átmásoljuk az ACF-PACF-Qszámítása.xlsm Excel parancsfájl Adatok-Számítások munkalapja adatok oszlopába (B3-B1048576) Az ACF és PACF korrelogramok illetve autokorrelációs- és parciális autokorrelációs együtthatók vizsgálata és a Q* statisztikák alapján eldöntjük, hogy az átalakított idősor valóban stacionáriusnak tekinthető-e. Ha igen, akkor elvégezhetjük az ARIMA modellezést.

Harmadik lépésben az ARIMA.xls parancsfájlba bemásoljuk az eredeti vagy a Box-Cox-transzformált adatokat. A differenciált idősort nem kell használni, mert az ARIMA.xlsm parancsfájlban a differenciálás foka (amit már ismerünk az első lépés számításai alapján) beállítható, és a becslésnél ezt figyelembe veszi. A Box-Cox-transzformált adatok használata esetén a legjobb modell kiválasztása után lehetőség van az eredeti- és becsült adatok visszatranszformálására. (ARIMA.xlsm Adatok visszatranszformálása munkalapon). Az ARIMA becslés után sokoldalúan lehet ellenőrizni a modellt, pl. a becslési és főleg a teszt-időszak hibáinak (pl. MAPE) alapján, ha megosztottuk az idősort becslési és teszt időszakra, vizsgálhat-

jük a reziduumokat, hogy az ARIMA modell eleget tesz-e a lineáris regressziós modell szokásos feltételeinek. A feltételek a véletlen változóra vonatkoznak, így a modell ellenőrzése az et reziduumok véletlen jellegének a vizsgálatát jelenti. A kiválasztott ARIMA modell esetében ez azt jelenti, hogy az et véletlen változó független véletlen folyamatot, fehér zaj folyamatot követ, normális eloszlással, nulla várható értékkel és konstans szórással. Ezeket a teszteket közli az ARIMA.xlsm parancsfájl, viszont a fehér zaj tesztelésére (pl. az autokorrelációs együtthatók nem szignifikánsak) ismét az ACF-PACF-Qszámítása.xlsm parancsfájlt kell használnunk.

Stacionaritás-biztosítása.xlsm Excel parancsfájl működése.

Éves illetve egyéb (pl. napi adatok) esetében (nincs szezonáltság) Eredeti adatok és $d=1$, $d=2$ és $d=3$ számítása és grafikus ábrázolása. A másik lehetőség Box-Cox transzformáció és ebből $d=1$ és $d=2$ számítása és grafikus ábrázolása. A differencia képzés az ismert képletekkel történik:

$$d = 1 = \Delta Y_t = (Y_t - Y_{t-1})$$

$$d = 2 = \Delta^2 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

$$d = 3 = \Delta^3 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

Box-Cox transzformáció. A λ -t meg kell adni és a célszerű intervallum:

$$Y(\lambda) = \begin{cases} (Y^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \ln(Y), & \lambda = 0 \end{cases} \quad Y > 0$$

$$-2 \leq \lambda \leq 2$$

A Box-Cox transzformáció után az adatok visszatranszformálása:

$$Y = (Y(\lambda)\lambda + 1)^{1/\lambda}$$

Havi adatok esetében: szezonális és nem szezonális differencia képzés.

Szezonális differencia képzés:

$D=1$ ($y_t - y_{t-12}$), az idősor 12 adattal rövidül.

$$D = 1 = \Delta^{12} Y_t = Y_t - 2Y_{t-1} + Y_{t-12}$$

Ezután nem szezonális differencia képzés:

$d=1$ és $d=2$ számítása, az idősor további 1 illetve 2 adattal rövidül.

A másik lehetőség itt is a Box-Cox transzformáció elvégzése és utána a szezonális és nem szezonális differencia képzés.

ACF-PACF-Qszámítása.xlsm Excel parancsfájl működése.¹⁴¹

Tetszőleges méretű adatállományt fel tud dolgozni, amit az Excel megenged. (Maximum: 1048576)

Tetszőleges max k késleltetés, a gyakorlatban ez max=60

Az új adatbevitel előtt a régi adatokat törölni kell, az „Adatok törlése” ikonra kattintva. A PACF – értékeket a Yule-Walker egyenletek módszerével becsüli.

ACF, PACF ikonra kattintva számol (ACF, PACF, Q*-statisztika, p-érték) és korrelogramokat készít. (ACF és PACF munkalap). Ha pl. 0,05-nél kisebb a p-érték akkor 5 %-os szignifikancia szinten elutasítjuk a nullhipotézist, miszerint a reziduumok fehér zajok. Ha nagyobb 0,05-nél a p-érték, akkor 5 %-os szignifikancia szinten elfogadjuk a nullhipotézist, miszerint a reziduumok fehér zajok.

Az alkalmazott képletek:

Az autokorrelációs együtthatók becslése k késleltetéssel:

¹⁴¹ Forrás: Kidolgozta Lu Wang PHD 2007. UCDAVIS University of California, Department of Statistics anson.ucdavis.edu/~luwang/acf_pacf.xls Módosította és kiegészítette Kehl Dániel és Sipos Béla.

$$r_k = \frac{\sum_{t=1}^{(T-m)-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T-m} (y_t - \bar{y})^2} \quad k = 1, 2, \dots, K$$

T= az adatok száma, a minta nagysága

m= a tesztidőszak száma: m

T-m (t=1,2,...,T-m) a becslésre használt adatok száma összesen

K = a számított autokorrelációs együtthatók előre meghatározott száma, pl. 36.

A k késleltetés különböző értékeihez (k= 1,2,3,...,K) rendelt autokorrelációs együtthatók, az autokorrelációs függvényt alkotják.

Az autokorrelációs függvényt értékei mátrixba foglalva:

$$\mathbf{R}_k = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{k-1} \\ r_1 & 1 & r_1 & \cdots & r_{k-2} \\ r_2 & r_1 & 1 & \cdots & r_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \cdots & 1 \end{bmatrix}$$

A parciális autokorrelációs együtthatók becslése a Yule-Walker egyenletekkel történik:

$$\hat{\Phi} = \mathbf{R}^{-1} \mathbf{r} = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{k-1} \\ r_1 & 1 & r_1 & \cdots & r_{k-2} \\ r_2 & r_1 & 1 & \cdots & r_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_1 \\ \cdot \\ \cdot \\ r_k \end{bmatrix}$$

i=1,2,3...k késleltetett értékekre külön-külön kiszámolja a parancsfájl a szorzatokat ($\mathbf{R}^{-1} \mathbf{r}$) és a parciális autokorrelációs együttható (k=1,2...K) a számított vektor $\hat{\Phi}$ utolsó eleme lesz.

A Ljung - Box portmanteau-próba (LJB vagy Q*-teszt):

$$Q^* = (T - m)(T - m + 2) \sum_{k=1}^{k=K} \left[\frac{r_k^2}{n - k} \right]$$

Ahol:

k = a számított autokorrelációs együtthatók előre meghatározott száma, pl. 36.

A becslésre használt adatok száma összesen T-m (t=1,2,...,T-m),

H₀=a reziduumok fehér zajok,

H₁=a reziduumok nem fehér zajok,

Q* χ^2 -eloszlást követ (K) szabadságfokkal.

Ennek alapján a hipotézis rendszer:

A H₀-t elfogadjuk, ha:

$$Q^* <_K \chi^2_{0,05}$$

A H₁ alternatív hipotézist fogadjuk el, ha:

$$Q^* >_K \chi^2_{0,05}$$

A szignifikancia-érték (p-érték) az a legkisebb szignifikancia szint, amin a H₀ már éppen elvethető a H₁-gyel szemben. Ha pl. 0,05-nél kisebb a p-érték akkor 5 %-os szignifikancia szinten elutasítjuk a nullhipotézist, miszerint a reziduumok fehér zajok.

Az ACF és PACF munkalapokon a tesztelés:

A konfidencia alsó és felső hibahatár által meghatározott sáv kiszámítása 5 %-os szignifikancia szinten az alábbi képlettel történik:

$$\pm 1,96 / \sqrt{T - m}$$

A $\pm 1,96/\sqrt{T-m}$ hibahatár által meghatározott sávon belüli autokorrelációs- és parciális együtthatók 5 %-os szignifikancia szinten, zérusnak tekinthetők, különben nem.

ARIMA.xls Excel parancsfájl működése.¹⁴²

Beállítások Office 2007

Nyisson meg egy üres EXCEL fájlt.

Makrók beállítása: Kattintson a Microsoft Office gombra, majd a *Programnév* beállításai gombra (Excel Options), ahol a *Programnév* az éppen használt alkalmazás neve, például Az Excel beállításai (alul található). Kattintson az Adatvédelmi központ (Trust Center) elemre, majd Az Adatvédelmi központ beállításai (Trust Center Settings) gombra végül a Makróbeállítások (Macro Settings) elemre. Kattintson a kívánt beállításra, a választás: Az összes makró engedélyezése. (Enable all macros) Minden választás után Ok.

Kattintson a Microsoft Office gombra, majd a *Programnév* beállításai gombra, (Excel Options) ahol a *Programnév* az éppen használt alkalmazás neve, például Az Excel beállításai (alul található). Kattintson a Bővítmények (Add-Ins) gombra, majd válassza a Kezelés – Excel bővítményeket alul, az ugrást választva a Bővítményeket bejelölheti (Manage: Excel Add-Ins, Go, megjelenik: Analysis Tool Pak, érdemes a többit is bejelölni.). A Bővítményeket az Excel installálja. A Bővítmények megjelennek: Adatok - Adatelemzés ikonnál. (Data – Data Analysis)

Az ARIMA.xls esetében még a következő beállításokra van szükség:

Eszközök – Bővítmenykezelő - Solver beikszelni, vagy ha be van jelölve kizedni a bejelölést, kilépni, lementeni, belépni és újra bejelölni a Solvert. (A többi bővítményt is célszerű bejelölni) Továbbá:

Eszközök – Makró - Visual Basic Editor - Tools (felül) - References - Solver legyen bejelölve. Megoldható úgy is, hogy Alt+F11 Tools, Preferences, és ki kell jelölni (pipa jel) a SOLVER feliratot, ha nem volt bejelölve.

A másik elérési lehetőség: Kattintson a Microsoft Office gombra, majd a *Programnév* beállításai gombra, (Excel Options) ahol a *Programnév* az éppen használt alkalmazás neve, például Az Excel beállításai (alul található). Jelölje be: Fejlesztőeszközök lap megjelenítése a szalagon. Megjelenik a Fejlesztő eszközök szalag, azon belül Visual Basic-Tools-References-Solvert be kell jelölni. Kilépés után mindig menteni kell. Előfordulhat olyan hibaüzenet, hogy Eszközök – Makró - Visual Basic Editor - Tools (felül) - References – Solvermissing, kijelölést (pipa jel) ki kell szedni és a solvert be kell (pipa jel) jelölni.

Körkörös hivatkozás esetén, ha iterációt végez az Excel, az Excel által javasolt módosítás: Az Excel beállításai – Képletek - Közelítés engedélyezése.

Új adatállomány bevitele:

Az adatsor kiválasztása. A program elvégzi a differencia-képzést, de ha ln-transzformációra, vagy Box-Cox transzformációra van szükség, azt el kell végezni és a transzformált adatokkal kell a továbbiakban dolgozni. Erre szolgál a stacionaritás - biztosítása Excel parancsfájl. A megfelelő ARIMA modell (paraméterek beállítása) kiválasztására szolgál az ACF-PACF-Q* Excel parancsfájl. (Ld. ARIMA modellek jellemzése táblát) Ellenőrizni kell, hogy a transzformált adatsor stacionárius-e, mert csak ebben az esetben lehet alkalmazni az ARIMA modellt. Ha nem stacionárius az idősor, akkor gyakran valamelyik paraméter vagy paraméterek 1 vagy -1 értéket vesznek fel.

Az s érték beállítása. Először az ARIMA munkalapon a szezonok (s) számát kell beírni. Lehetőségek: 1, 4, 7, 12. Ha nincs szezonális, s=1, pl. éves adatok vagy napi tőzsdeindexek, ha negyedéves a szezonális, s=4, negyedéves adatok, ha havi a szezonális, s=12 havi adatok, s=4 és s=12 esetében a peridusok az évek, heti szezonális, s=7 napi adatok, ahol a peridusok a hetek. Ha óránkénti adatok vannak, akkor s=1 ahol a periódus a nap.

Adatbevétel: Ctrl+a,

Megfigyelések számának módosítása. ? x

¹⁴² Kidolgozta: Mongkol Temrangsitornrat (UMKC, University of Missouri-Kansas City), átalakította, lefordította és átprogramozta Kehl Dániel és Sipos Béla (PTE KTK).

A következő szöveg jelenik meg: A munkalap jelenleg hosszúságú adatsort támogat. Adja meg, hogy hány megfigyelést kíván illesztésre felhasználni! (36 megfigyelésnél többnek kell lennie!)

Be kell írni az új idősor becslésre felhasznált adatainak a számát. Ezt követően a következő szöveg jelenik meg:

A munkalap jelenleg ... előrejelzési megfigyelést támogat. Adja meg, hogy hány előrejelzési periódust kíván használni. (Legalább 1 periódusnál többnek kell lennie) Ha nem közli a korábbi beállításokat, akkor solver hiba van, s a solvert az előzőekben leírtak szerint újra be kell állítani.

A Ctrl+a tehát közli a jelenlegi becslésre felhasznált idősor hosszát, a korábbi számítás alapján és kéri az új becslésre szolgáló idősor adatainak a számát (az argumentumban, egy oszlopban szereplő cellák maximális sorszámát, ami elméletileg 1048576 sor lehet. A program akár 35 ezer adat, pl. DJA napi index 1896-2011 esetében is működik kb. 10 perc alatt megoldást ad).

Megfigyelések számának módosítása. ? x

Ezt követően kéri az előrejelzés hosszát, közölve a jelenlegi, korábbi számítás értékét, aminek csak szakmai korlátja van, ha ex-post ellenőrzést illetve ex-ante előrejelzést kérünk. Ebben az esetben az idősort becslési és teszt időszakokra bontjuk. Lehet pl. ex-post ellenőrzésnél a becslési időszak az idősor fele, (a teszt időszak akkor szintén az idősor fele) kétharmada, (a teszt időszak akkor az idősor egyharmada) négyötöde (a teszt időszak akkor az idősor egyötöde) vagy becslési időszak az idősor hossza mínusz az utolsó év, ami havi adatoknál pl. 12, negyedéves adatoknál pedig 4. Hibaképleteket akkor számol a program, ha azt megadjuk, tehát az összes bemásolt adat számának és a becslésre szolgáló adatok számának a különbsége egy vagy egynél nagyobb érték.

Ex-post ellenőrzésnél az ARIMA illesztést (becslést) végzi el a program a becslési időszak adatainak az alapján és kiszámítja az illesztett hibákat a becslési időszakra, továbbá az ARIMA modell előrejelzéseket készít, ha megadtuk a megfelelő előrejelzési időszakot, vagyis a teszt időszak adatainak a számát és mivel a tényadatokat is megadtuk a teszt időszakra, a program hibaképleteket tud számítani, amit az előrejelzett hibák néven közöl. A hibaképletek az ARIMA munkalapon a számítások elvégzése (Ctrl+b és Ctrl+k) után a számsorok után az O (Illesztett hibák) és R (Előrejelzett hibák) oszlopban jelennek meg.

Az adatok módosítása (Ctrl+a) után a régi adatsort törli a parancsfájl és be lehet másolni az új adatsort. Ha az idősort megosztjuk becslési és teszt időszakra, akkor célszerű így megadni az adatokat, pl. a teszt időszak hossza megegyezik az előrejelzés hosszával. (pl. havi adatok, az idősor hossza 456 hónap adata, az előrejelzés hossza 12 hónap, akkor az új idősor hossza kérdésre a válasz 444 hónap, és ebből számol, az idősor bevétel előrejelzés celláiba be lehet másolni a tesztidőszak 12 adatát. Az előrejelzés hosszának lehet kérni 24-t, ebből 12 adat az ex-post előrejelzés ellenőrzéséhez szükséges és 12 lesz az ex-ante előrejelzés, amikor már tényadatokkal nem rendelkezünk. A hibaképleteket a program kiszámolja az illesztett (becslési) és ebben az esetben az előrejelzett (teszt) időszakokra is.

Hibaüzenetnél Debug-t kell választani és a sárga mezőben lévő utasításnál shift 1, aposztrófot (') kell beírni és menteni. A másik lehetőség a solver beállítása az előzőekben leírtak szerint.

Célszerű először ex-post elemzést végezni és a legjobb ARIMA modell ismeretében ex-ante előrejelzéseket készíteni.

Az ARIMA munkalapon a kiinduló ARIMA modell felírása (azonosítása, identifikációja) és az ARIMA modell paramétereinek a becslése.

A felhasználóknak meg kell határozniuk a (p-d-q-C-P-D-Q-S) ARIMA változók értékét a sárga cellában (A3:H3). A változók felvehető értékei: d és D: 0-3, p: 0-12, P: 0-2, q: 0-12, Q: 0-2, C: 0-1. S: 1, 4,7,12. Hibás adat beírása esetén a program hibát jelez, ilyenkor csökkenteni kell a beírt értéket 1-gyel, és ha akkor is hibát jelez, mindaddig csökkenteni kell a beírt változó értékét, amíg hibát nem jelez.

Havi adatok esetén S 12 kell legyen, és ha az adatok negyedévesek, S 4 kell legyen, és heti adatoknál S 7 kell legyen. A szürke cellák számítási területek, így oda adatokat nem szabad írni, csak a sárga cellákba. Az ARIMA paraméter értékének beírása után, a paraméterek módosítását végrehajtó Solvert futtatni kell. (Ctrl+b). A koefficiensek értékének meghatározásához a felhasználóknak futtatniuk kell az ARIMA modell becslését végző Solvert, ami iterációkkal (feltételezéses maximum likelihood (CML) módszerével) elvégzi a számításokat. (Ctrl+k)

A változó cellák a sárga cellák, melyek tartalmazzák számítások eredményét, értelemszerűen, ha kiválasztásra kerültek, tehát nem 0 paramétert jelöltünk be.

$\theta_0 = X28$, $\phi (1-12)$ AA 28-39, $\phi_S (1-2)$ AE 28-29, $\theta (1-12)$ AI 28-39, $\theta_S (1-2)$ AM 28-29.

A θ (theta) és ϕ (phi) koefficiens korlátai, kisebbek, mint 1, de nagyobb, mint 0, kivéve a θ_0 -t (theta 0-t) amely bármely valós szám lehet.

ARIMA modellek jellemzése tábla segít a modell kiválasztásában.

Az ARIMA modell paramétereinek becslése a feltételezéses maximum likelihood (ML) (conditional maximum likelihood ~ feltételes legnagyobb esélyesség) módszerével történik. Ennek lényege, hogy a becslőfüggvény készítője ismerettel rendelkezik az alapsokaság eloszlására vonatkozóan, pontosabban ismeri az alapsokasági eloszlás típusát, de nem ismeri a konkrét alapsokasági paraméterek értékét. Ez a feltételezés az ARIMA modellek esetében az, hogy a megfigyelések normális eloszlást követnek és eleget tesznek a stacionaritási követelményeknek.¹⁴³ A kezdőértékek számításánál a feltételezéses (conditional) ML módszert használja, vagyis a kezdőértékek nullák. Például a ϕ paraméter maximum likelihood (ML) becslése az a $\hat{\phi} = \phi(Y)$ paraméterérték, amelyre $\phi \rightarrow f_{\phi}(Y)$ (tehát a likelihood-függvény) maximális. Tehát azt a paramétert választjuk becslésnek, ami mellett az Y minta bekövetkezésének „valószínűsége” a legnagyobb.

Az ARIMA mukafüzetben, a leírtak szerint:

Meg kell adni a paraméterek számát. ARIMA (p, d, q). C (konstans, van=1, nincs= 0) Ha van szezonálítás: (P, D, Q)_s

Ctrl+b módosítja az ARIMA modell paramétereit, amiket a sárga cellákba beírtunk.

Ctrl+k iterációs eljárással becsüli a paramétereket (Ctrl+b utáni paraméterek szerint) és kiszámítja a becsült paramétereket valamint a következő statisztikákat:

A rendelkezésre álló adatok száma: T

A teszt időszak száma: m

A becslési időszak száma: T-m

A becslésre használt adatok száma összesen T-m (t=1,2,...T-m), minimum 36.

Kieső adatok száma= Kieső adatok száma=p+d+s(P+D)

Felhasználható adatok száma= Az adatok száma - Kieső adatok száma = (T-m) - [p+d+s(P+D)]

Szabadságfok (df) számítása: df=(T-m)-[p+d+s(P+D)] - (p+q+P+Q)

Számítási átlag (\bar{Y}_t):

$$\bar{Y} = \frac{1}{T-m} \sum_{t=1}^{T-m} Y_t$$

SzórásP (σ_Y):

$$\sigma_Y = \sqrt{\frac{1}{T-m} \sum_{t=1}^{T-m} (Y_t - \bar{Y})^2}$$

Hibanégyzet-összeg (SSE):

$$SSE = \sum_{t=1}^{T-m} (Y_t - \hat{Y})^2$$

A reziduális szórás (s):

$$s = \sqrt{\frac{SSE}{df}} = \sqrt{\frac{\sum_{t=1}^{T-m} (Y_t - \hat{Y})^2}{(T-m)-[p+d+s(P+D)] - (p+q+P+Q)}}$$

R^2 többszörös determinációs együttható:

$$R^2 = 1 - \left(\frac{s}{\sigma_Y} \right)^2 = 1 - \frac{\sum_{t=1}^{T-m} (Y_t - \hat{Y})^2}{\frac{1}{T-m} \sum_{t=1}^{T-m} (Y_t - \bar{Y})^2}$$

Az R^2 negatív is lehet, ha $s > \sigma_Y$, de a megfelelő modell kiválasztása esetén egyhez közeli érték, különösen ha sok magyarázóváltozó szerepel a modellben.

¹⁴³ Ld. Pintér-Rappai szerk. Statisztika.PTE, KTK. 2007. 309-313.

DW statisztika. (Durbin-Watson d-próba) A próbafüggvény:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

A program, a többi szoftverhez hasonlóan közli a DW statisztika értéket, ami nem használható az AR modelleknél, ahol a magyarázóváltozók valamelyike az eredményváltozó egynél több késleltetettje, vagyis Y_{t-1} , Y_{t-2} ... alakú.¹⁴⁴ Ha a késleltetés 1, (Y_{t-1}) akkor viszont használható a DW statisztika.

Az ARIMA modellek diagnosztikai ellenőrzése.

Az R^2 nem alkalmas az illeszkedés mérésére, mert a paraméterek számának növekedésével gyakorlatilag 1-hez tart. A modell alkalmasságának vizsgálatára a legjobb módszer a mintán kívüli előrejelzés, vagyis a minta egy részét visszatartjuk (nem használjuk fel a becslés során, ez a teszt időszak), ezekre ex post előrejelzéseket készítünk, majd összehasonlítjuk az előrejelzett értékeket (\hat{Y}_t) az Y_t ismert értékeivel. A becslési időszak alapján készítjük az előrejelzéseket. A „mi lett volna ha” feltételezés mellett nagy valószínűséggel a legjobb modell kiválasztható. Természetesen feltételezzük, hogy az ex-ante időszakban jelentősebb tendencia változás pl. válság nem következik be. A mintán kívüli előrejelzéshez az adatbevitelt (Ctrl+a) kell megfelelően beállítani.

Az ARIMA parancsfájl az „Eredeti és becsült” illetve a „Reziduum” munkalapokon a számítások alapján a grafikus ábrákat is közli.

Az alkalmazott hibaképletek.

Kiszámítja a hibaképleteket a becsült ($t=1,2,\dots,T-m$) értékekre minden esetben, a becslésre szolgáló megadott adatok (Y_t) és az ARIMA modell alapján becsült adatok (\hat{Y}_t) alapján.

Kiszámítja továbbá a hibaképleteket az ex-post előrejelzett ($T, T-1, T-2,\dots,T-(m+1)$) értékekre is, ha az idősort megosztottuk becslési (cellák száma: $T-m$) és teszt (cellák száma: m) időszakra és megadtuk a teszt időszak értékeit.

Ex-ante előrejelzést készít az $M-m$ időszakra, amikor már megfigyelésekkel nem rendelkezünk: ($T+1, T+2, \dots, T+M-m$). Ellenőrzése akkor lehetséges amikor az ex-ante időszak bekövetkezik és így ex-post időszakká válik.

Az alkalmazott jelölések:

A hiba ($e = \text{Error}$):

$$e_t = Y_t - \hat{Y}_t$$

Y_t = az illesztésre (becslésre) felhasznált megfigyelt értékek ($1,2,\dots,T-m$) és az ex-post becslésre ($T, T-1, T-2,\dots,T-m-1$) használt megfigyelt adatok.

\hat{Y}_t = az illesztett (becsült) ($1,2,\dots,T-m$) és az ex-post előrejelzett ($T, T-1, T-2,\dots,T-m-1$) értékek.

Az illesztett hibák számítása a fenti jelölések felhasználásával:

$t=1,2,3,\dots,T-m$

Átlagos hiba. [ME=MEAN ERROR]

$$ME = \frac{\sum_{t=1}^{T-m} Y_t - \hat{Y}_t}{T - m}$$

A hibák (reziduumok) négyzetösszege. SSE (SUM OF SQUARED ERRORS)

$$SSE = \sum_{t=1}^{T-m} e_t^2 = \sum_{t=1}^{T-m} (Y_t - \hat{Y}_t)^2$$

¹⁴⁴ Ramanathan i. m. 406.

Relatív [%-os] hiba [PE_i = PERCENTAGE ERROR]:

$$PE_t = \frac{Y_t - \hat{Y}_t}{Y_t} \times 100$$

Relatív [%-os] abszolút hiba [PE_i = ABSOLUTE PERCENTAGE ERROR]:

$$|PE_t| = \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100$$

A hiba abszolút értéke:

$$e_t = |Y_t - \hat{Y}_t|$$

Átlagos relatív [%-os] abszolút hiba szórása, [RSE=ROOT SQUARE ERROR of MEAN ABSOLUTE PERCENTAGE ERROR]:

$$RSE = \sqrt{\frac{\sum_{t=1}^{T-m} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100}{(T-m) - p+d+s(P+D)}}$$

Átlagos abszolút eltérés [MAD=MEAN ABSOLUTE DEVIATION OF ERROR]

$$MAD = \frac{\sum_{t=1}^{T-m} |Y_t - \hat{Y}_t|}{T - m}$$

Átlagos relatív [%-os] hiba [MPE = MEAN PERCENTAGE ERROR]

$$MPE = \frac{\sum_{t=1}^{T-m} \frac{Y_t - \hat{Y}_t}{Y_t} \times 100}{T - m}$$

Átlagos relatív [%-os] abszolút hiba [MAPE = MEAN ABSOLUTE PERCENTAGE ERROR]:

$$MAPE = \frac{\sum_{t=1}^{T-m} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100}{T - m} (\%)$$

Az ex-post előrejelzett (T, T-1, T-2,...T-m-1) értékek hibaszámítása hasonló módon történik, csak az indexek különböznek: T, T-1, T-2,...T-m-1 értékeket vesznek fel.

Pl. az előrejelzett MAPE számítása a teszt periódusra:

$$MAPE = \frac{\sum_{t=T-(m+1)}^m \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100}{m} (\%)$$

Több ARIMA modell közül azt választjuk, amelyre a MAPE, a MAD, RSE hibák kisebbek.

Szelekciós kritériumok alapján történő választás.¹⁴⁵

Akaike – féle információs kritérium. (Akaike Information Criterion = AIC)

Schwartz-féle bayesi kritérium (Schwarz Bayesian Criterion = SBC) vagy bayesi információs kritérium (Bayesian (Schwartz) Information Criterion = BIC vagy SIC).

Hanan-Quinn kritérium (Hanan-Quinn Criterion = HQC)

Az ARIMA modellek összehasonlítására szolgálnak és egyszerre figyelembe veszik az illeszkedés jóságát és a becsült együtthatók számát.

A jelölések:

A becsült paraméterek száma: $\tilde{P} = p+q+P+Q$

Az adatok száma, a minta nagysága: T

A tesztidőszak száma: m

¹⁴⁵ Ramanathan i. m. 173-174..

Az illesztésre (becslésre) felhasznált adatok száma T-m
 A hibák (reziduumok) négyzetösszege.¹⁴⁶

$$SSE = \sum_{t=1}^{T-m} e_t^2 = \sum_{t=1}^{T-m} (Y_t - \hat{Y}_t)^2$$

Reziduális szórásnégyzet (variancia) (s^2)

$$s^2 = \frac{\sum_{t=1}^{T-m} (Y_t - \hat{Y}_t)^2}{(T-m)}$$

A becsült paraméterek száma: $\tilde{P} = p+q+P+Q$

Az AIC, BIC és HQC mutatók eredeti és logaritmizált formákban:

$$AIC_{(\tilde{P})} = s^2 e^{\frac{2\tilde{P}}{T-m}}$$

$$LN(AIC_{(\tilde{P})}) = LN(s^2) + \frac{2\tilde{P}}{T-m}$$

$$BIC_{(\tilde{P})} = s^2 (T-m)^{\frac{\tilde{P}}{T-m}}$$

$$LN(BIC_{(\tilde{P})}) = LN(s^2) + \frac{\tilde{P}}{T-m} LN(T-m)$$

$$HQC_{(\tilde{P})} = s^2 (LN[T-m])^{\frac{2\tilde{P}}{(T-m)}}$$

$$LN(HQC_{(\tilde{P})}) = LN(s^2) + \frac{2\tilde{P}}{T-m} LN(LN[T-m])$$

Több ARIMA modell közül azt választjuk, amelyre az AIC-, BIC- a HQC - érték a legkisebb. A szelekciós kritériumok működésük elve a következő: a kisebb hibájú modellt preferálják, de ezek közül is azt választják, amelyik a kevesebb paramétert használ. A kétféle kritérium egymásnak ellentmondó eredményre is vezethet. A magyarázó változók számának növekedését a Schwartz-féle bayesi kritérium (BICp) jobban bünteti. A szelekciós kritériumok kiválasztása arra vonatkozik, hogy mely koefficienseket használjuk (tartjuk meg) a modellünkben. Minél egyszerűbb ARIMA modellel tudunk jó prognózist készíteni, annál megbízhatóbb prognózist tudunk készíteni.

Az (SSE/T-m), egy megfigyelési idősor egységre vetítve az átlagos hiba. Így kezdődik mind három mutató. A hiba minél kisebb annál jobb, tehát ez a modell úgy fog működni, hogy minél kisebb a hiba, annál jobb a mutató, de a javulást ellensúlyozza a paraméterek számának növekedése. Ezt fejezi ki az e a (T-m) illetve $\ln(T-m)$ alapú hatvány, amit büntetőfaktornak nevezünk.

A véletlen változó, (reziduumok e_t) vizsgálata.

A modell ellenőrzése annak a vizsgálatát jelenti, hogy a becsült ARIMA modell eleget tesz-e a lineáris regressziós modell szokásos feltételeinek. A feltételek (ld. regresszio.xls parancsfájl leírását) a véletlen változóra vonatkoznak, így a modell ellenőrzése az e_t reziduumok véletlen jellegének vizsgálatát jelenti.

Ezeknek a számításoknak egy részét már elvégeztük a modell azonosítása munkafázisban, amikor azt ellenőriztük, hogy az Y_t változók illetve azok transzformációjával (Box-Cox-transzformáció és differencia képzés, d és D) nyert transzformált változók eleget tesznek-e a stacionaritás követelményének illetve, hogy a változó, a megfigyelt idősor (Y_t) hibái, vagyis az autokorrelációs együtthatók fehér zajt alkotnak-e. Ezt Q*-tesztel ellenőriztük.

Most viszont becsült ARIMA modellel rendelkezünk és ellenőriznünk kell azt, hogy a becsült reziduumok véletlen jellege igazolható-e.

A lineáris regressziós modell feltételeit az ARIMA modellre felírva: az \hat{e}_t véletlen változó független véletlen folyamatot (fehér zaj folyamatot) követ, nulla várható értékkel és konstans szórásnégyzettel. A konstans szórásnégyzet állandóságára vonatkozó feltétel teljesülését ellenőrizhetjük a reziduumok grafikus ábrája alapján. Ld. Reziduum munkalapot.

¹⁴⁶ Másik neve az SSE-nek a nemzetközi irodalomban: Residual Sum of Squares (RSS)

Az ARIMA munkalapon az O oszlopban közli a program a reziduum (e_t) értékeket, ezekből kell először autokorrelációs együtthatókat számítani pl. $k=36$ késleltetéssel.

$$r_k(\hat{e}_t) = \frac{\sum_{t=k+1}^{T-m} (\hat{e}_t - \bar{\hat{e}})(\hat{e}_{t-k} - \bar{\hat{e}})}{\sum_{t=1}^{T-m} (\hat{e}_t - \bar{\hat{e}})^2}$$

Az a $r_k(\hat{e}_t)$ standard hibájának a közelítése:

$$s[r_k(\hat{e}_t)] = \frac{1}{\sqrt{T-m}}$$

Az előzőek alapján:

$$r_k(\hat{e}_t) = \text{ACF}$$

A reziduális változó függetlensége esetén az autokorrelációs együtthatók $r_k(\hat{e}_t)$ ($k=1,2,\dots,K$) nem szignifikánsak. Célszerű, a már ismertett Ljung-Box portmanteau-próba (LJB vagy Q^* -teszt) használni az autokorrelációs együtthatók tesztelésére.

$$Q^* = (T-m)(T-m+2) \sum_{k=1}^{k=K} \left[\frac{r_k^2(\hat{e}_t)}{T-m-k} \right]$$

A reziduumok (e_t) az ARIMA.xls parancsfájl ARIMA munkalapján az O oszlopban található, ezeket a parancsfájl bemásolja a „Modell ellenőrzése munkalap” B oszlopába, a sorszámokat az A oszlopába. Az e_t reziduumokat bemásoljuk az ACF-PACF-Qszámítása Excel parancsfájlba, a „Reziduumok másolása” cellára kattintva, ami kiszámítja az autokorrelációs együtthatókat a megadott K késleltetéseknek megfelelően, a parciális autokorrelációs együtthatókat, a Q^* -statisztikákat és a p-empirikus szignifikancia értékeket. A számítások eredményét visszamásoljuk az ARIMA.xls fájl „A modell ellenőrzése” munkalapjába. A régi adatokat előtte kitöröljük.

A becsült reziduális változó átlagának $[M(e_t)]$ nullához közel kell esni. A nullhipotézis ellenőrzése:

A reziduumok (e_t) átlaga, a reziduális szórásnegyzet [variancia $s^2(e_t)$] és a reziduális változó (e_t) standard hibájának $s(e_t)$ becslése:

$$\bar{\hat{e}}_t = \frac{\sum_{t=1}^{T-m} \hat{e}_t}{T-m}$$

$$s_{\hat{e}_t}^2 = \frac{\sum_{t=1}^{T-m} \hat{e}_t^2}{(T-m) - (p+q+P+Q)}$$

$$s(\bar{\hat{e}}) = \sqrt{\frac{s_{\hat{e}_t}^2}{T-m}}$$

Az egymintás t-próba, nem követeli meg a sokasági eloszlás szórásának ismeretét, de annak normális voltát továbbra is kiköti. :

$$t = \frac{\bar{\hat{e}} - 0}{s(\bar{\hat{e}})}$$

$$H_0 : M(e_t) = 0$$

$$H_1 : M(e_t) \neq 0$$

A próbafüggvény a nullhipotézis fennállása esetén:

$$t = \frac{\bar{\hat{e}}}{s(\bar{\hat{e}})} < t_{(1-\alpha/2)(T-m-1)}$$

A próbafüggvény a nullhipotézis elutasítása esetén:

$$t = \frac{\bar{\hat{e}}}{s(\bar{\hat{e}})} > t_{(1-\alpha/2)(T-m-1)}$$

A reziduumok (e_t) Jarque-Bera féle normalitási-tesztje.

A reziduumok száma:

Becslésre felhasznált adatok száma (T-m)

Kieső adatok száma $k=p+d+s(P+D)$

Reziduumok száma $n = (T-m) - [p+d+s(P+D)]$

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4}(K-3)^2 \right)$$

Ahol:

Reziduumok száma $n = (T-m) - [p+d+s(P+D)]$

S= a ferdeség mérésére szolgáló S mutató:

$$S = \frac{n}{(n-1)(n-2)} \sum_{t=1}^n \left(\frac{e_t - \bar{e}_t}{\sigma} \right)^3$$

A S - mutató értéke 0, ha az eloszlás szimmetrikus, pozitív előjel esetén jobb oldali, míg negatív előjel esetén bal oldali aszimmetriára következtethetünk.

K = a csúcsosság mutatószáma:

$$K = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{t=1}^n [e_t - \bar{e}_t]^4}{\sigma^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

A (K-3) relatív csúcsosság mutató értéke 0, normális eloszlás esetén. Ennél kisebb érték esetén lapulnak, nagyobb érték esetén csúcsosnak tekinthetjük az eloszlást.

A szabadság fok (df) mindig az S és K paraméterek miatt = 2

A tesztstatisztika 2 szabadságfokú χ^2 -eloszlást követ. Minél magasabb az értéke, annál biztosabban (alacsonyabb szignifikancia-szint mellett) tudjuk elutasítani a nullhipotézist, miszerint a megfigyelések normális eloszlást követnek. Normális eloszlás esetén a $JB=0$, mert $S=0$ és $(K-3)=0$, tehát alacsony értéke normalitásra utal. Ha a p-érték kisebb mint 0,05 akkor 5 %-os szignifikancia szinten elutasítjuk a nullhipotézist, miszerint a megfigyelések normális eloszlást követnek.

Ennek alapján a hipotézis rendszer:

H_0 = az adatsor normális eloszlást követ,

H_1 = az adatsor nem követ normális eloszlást,

A H_0 -t elfogadjuk, ha:

$$JB <_2 \chi^2_{0,05}$$

A H_1 alternatív hipotézist fogadjuk el, ha:

$$JB >_2 \chi^2_{0,05}$$

A szignifikancia-érték (p-érték) az a legkisebb szignifikancia szint, amin a H_0 már éppen elvethető a H_1 -gyel szemben. Ha pl. 0,05-nél kisebb a p-érték akkor 5 %-os szignifikancia szinten elutasítjuk a nullhipotézist, miszerint a megfigyelések normális eloszlást követnek.

ACF-PACF-ARIMAszámítás144adattal.xlsm Excel parancsfájl működése.

Az Adatbevitel munkalapra 144 adat másolható. Ctrl+a utasítással az adatállomány bővíthető, kéri az adatállomány növekedésének a nagyságát. Pl. az új adatállomány 200, a régi 144, a bővülés, amit be kell írni. $200-144=56$. Nyomon követhető az ACF, PACF, Q^* számítása és tesztelése, az ACF és PACF korrelogramok készítése. A PACF – értékeket a Yule-Walker egyenletekkel módszerével becsüli. A számítások képletei a munkafüzetekben megtalálhatók, maguk a számítások nyomon követhetők és az elméleti részben az előzőekben leírásra kerültek. Az ARIMA munkalap 144 adattal működik, a programozás látható. Ennek a parancsfájlnak az a célja, hogy a számítások menetét lehessen látni, az ne legyen "fekete doboz". Az ACF-PACF- Q^* számítását az előzőekben bemutattuk. Az ARIMA modellezés lépéseit ismertetjük. A parancsfájllal csak a megadott ARIMA változokkal végzi el a számításokat, tehát iterációkat nem végez.

Bevezetés:

Ez a munkalap bemutatja az ARIMA módszert, amely képes kezelni heti, havi negyedéves adatokat.

Az ARIMA módszerrel a modell 3 fő komponenset tartalmaz, nevezetesen autoregressziós, differencia és mozgóátlag. Minden komponensnek két alkomponense van, azaz egyszerű és szezonális alkomponens. Az autoregressziós komponens általános formája:

$$(1 - \phi_{1S} B^{1S} - \phi_{2S} B^{2S} - \dots - \phi_{PS} B^{PS}) (1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = e_t$$

ahol p-t és P-t az ARIMA paraméter p és P értéke határozza meg

B jelöli a visszaléptető operátort és S jelöli a szezonaritást

A differencia komponens általános formája:

$$(1 - B^{DS}) (1 - B^d) Y_t = e_t$$

ahol d-t és D-t az ARIMA paraméter d és D értéke határozza meg

B jelöli a visszaléptető operátort és S jelöli a szezonaritást

A mozgóátlag komponens általános formája:

$$Y_t = (1 - \theta_{1S} B^{1S} - \theta_{2S} B^{2S} - \dots - \theta_{QS} B^{QS}) (1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q) e_t + \theta_0$$

ahol q-t és Q-t az ARIMA paraméter q és Q értéke határozza meg

B jelöli a visszaléptető operátort és S jelöli a szezonaritást

és θ_0 jelöli a konstans az ARIMA modelben (ha a C ARIMA paraméter 1)

Ezen három komponenset egyesítve az ARIMA általános formája pdqCPDQS paraméterrel az alábbi lesz

$$(1 - \phi_{1S} B^{1S} - \phi_{2S} B^{2S} - \dots - \phi_{PS} B^{PS}) (1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p) (1 - B^{DS}) (1 - B^d) Y_t = (1 - \theta_{1S} B^{1S} - \theta_{2S} B^{2S} - \dots - \theta_{QS} B^{QS}) (1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q) e_t + \theta_0$$

Hogyan használjuk a munkatáblát?

A felhasználóknak meg kell határozniuk a (p-d-q-C-P-D-Q-S) ARIMA paraméterek értékét a sárga cellában (A3:H3). A késleltetési adatok korlátozottsága miatt p és d értéke 0-36 között kell legyen. P és D értéke 0-15 között kell legyen. q és Q értéke 0-2 között kell legyen. C értéke 0 vagy 1 kell legyen. Havi adatok esetén S 12 kell legyen, és ha az adatok negyedévesek, S 4 kell legyen, és heti adatoknál S 7 kell legyen. Kérem, vegye figyelembe, a szürke cellák számítási terület

Az ARIMA paraméter értékének meghatározása után, a munkalap kiválasztja, milyen konstans és koefficiens vesz részt a számításban a meghatározott ARIMA paraméternek megfelelően. Ha a koefficienset vagy konstans kiválasztották, a mellette levő cella 1-re változik. És 0 ha nem választják ki.

A koefficiens értékének meghatározásához a felhasználóknak futtatniuk kell a megoldót. A célcellák lehetnek ME vagy SSE. Ezek az értékek a D95 illetve a D96-ban vannak. A változó cellák a sárga cellák, melyek tartalmazzák a konstans és phi-t és theta-t azaz M28 a θ_0 (theta0) konstansot, Q28:Q63 cella a $\phi_1 \dots \phi_{36}$ (phi1 ... phi36) autoregressziós koefficiens, V28:V42 a $\phi_{S1} \dots \phi_{S15}$ (phiS1 ... phiS15) szezonális autoregressziós koefficiens, Y28:Y29 tartomány a $\theta_1 \dots \theta_2$ (theta1 ... theta2) mozgóátlag koefficiens, AC28:AC29 tartomány a $\theta_{S1} \dots \theta_{S2}$ (thetaS1 ... thetaS2) szezonális mozgóátlag koefficiens. A theta és phi koefficiens korlátai, kisebb, mint 1, de nagyobb, mint 0, kivéve a theta 0, amely bármely valós szám lehet. A havi és negyedéves előrejelzési munkafüzet ARIMA modulja kissé különbözik ARIMA.xls-étől. A különbségek a következők:

1) A havi és negyedéves előrejelzési munkafüzet ARIMA modulja csak havi adatokkal tud dolgozni, mivel a Szezonális MA-t speciálisan havi adatok feldolgozására tervezték. Ugyanez érvényes a negyedéves munkafüzetre

2) Az ARIMA.xls makrója nem regenerálja a munkalapot. Egyszerűen növeli a megfigyelések számát. A havi és negyedéves előrejelzési munkafüzet ARIMA modulját regenerálták Ctrl+a megnyomásával.

Ezen munkalap Számítási Folyamata

A Számítási Folyamatot 6 fő modulra osztottuk:

1. egyszerű differenciálás modul
2. szezonális differenciálás modul
3. egyszerű autoregressziós modul

4. szezonális autoregressziós modul
5. mozgóátlag modul (egyszerű és szezonális kombinálva)
6. összegző modul

Mindegyik modul számítási eredményét továbbítjuk a másik modulba. Az egyes modulok működése

1 Egyszerű Differenciálás Modul

Ez a modul az AG - BT oszlopban található a 21. sortól kezdve. Az AG - BQ oszlop tartalmazza a késleltetett adatokat 1 késleltetéstől kezdve 36 késleltetésig. És az AD oszlop tartalmazza a δ mátrixot, amely megmondja, milyen késleltetést választunk a d ARIMA paraméter értékének megfelelően. A BS oszlop tartalmazza mátrix szorzás eredményét. Ezeket az eredményeket érvényesítjük a BT oszlopban, és DIFF_t -nek nevezzük

$$(1 - B^d) Y_t := \text{DIFF}_t$$

A δ mátrix	Az Y_t mátrix	
↙	↙	
1		
0		
:		
-1	X	$[Y_t \ Y_{t-1} \ Y_{t-2} \ \dots \ Y_{t-36}] := \text{DIFF}_t$
0		
:		
0		

2 Szezonális Differenciálás Modul

Ez a modul a BV - CN oszlopban található a 21. sortól kezdve. A BV - BQ oszlop tartalmazza a késleltetett DIFF_t (Y_t -t egyszerű differenciálás után) 1 szezonális késleltetéstől kezdve 15 szezonálisig. És az AE oszlop tartalmazza a δ_s mátrixot, amely megmondja, milyen szezonális késleltetést választunk a D ARIMA paraméter értékének megfelelően. A CM oszlop tartalmazza mátrix szorzás eredményét. Ezeket az eredményeket érvényesítjük a CN oszlopban, és SDIFF_t -nek nevezzük:

$$(1 - B^{DS}) \text{DIFF}_t := \text{SDIFF}_t$$

the δ_s matrix	the DIFF_t matrix	
↙	↙	
1		
0		
:		
-1	X	$[\text{DIFF}_t \ \text{DIFF}_{t-1S} \ \text{DIFF}_{t-2S} \ \dots \ \text{DIFF}_{t-15S}] := \text{SDIFF}_t$
0		
:		
0		

3 Egyszerű Autoregressziós Modul

Ez a modul a CP oszloptól EQ oszlopig található a 21. sortól kezdve. A CP - DZ oszlop tartalmazza a késleltetett SDIFF_t -t (Y_t -t egyszerű és szezonális differenciálás után) 1 késleltetéstől kezdve 36 késleltetésig. És az N oszlop a 28 sortól kezdve a 63 sorig tárolja az autoregressziós koefficiensek értékeit a $\phi_1 \dots \phi_{36}$ ($\phi_1 \dots \phi_{36}$) tartományban. Az M oszlop N-től balra mutatja, melyik koefficienset válasszuk a p ARIMA paraméter értékének megfelelően. Ez az M és N oszlop kombinálva alkotja az AB oszlopnál található ϕ mátrixot. Ez a ϕ mátrix mutatja, melyek az autoregressziós koefficiensek (ϕ_N) ebben a számításban. Az EB oszlop tartalmazza mátrix szorzás eredményét. Ezeket az eredményeket érvényesítjük az EC oszlopban, és AR_t -nek nevezzük

$$(1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p) \text{SDIFF}_t := \text{AR}_t$$

a ϕ mátrix	az SDIFF_t mátrix	
↙	↙	
1		
- ϕ_1		
- ϕ_2	X	$[\text{SDIFF}_t \ \text{SDIFF}_{t-1} \ \text{SDIFF}_{t-2} \ \dots \ \text{SDIFF}_{t-36}] := \text{AR}_t$
:		
- ϕ_{36}		

4 Szezonális Autoregressziós Modul

Ez a modul EE oszloptól EW oszlopig található a 21. sortól kezdve. Az EE - ET oszlop tartalmazza a késleltetett AR_t -t (Y_t egyszerű és szezonális differenciálás után) 1 késleltetéstől kezdve 15 késleltetésig. És az R oszlop a 28 sortól kezdve a 42 sorig tárolja a szezonális autoregressziós koeficiens értékei a $\phi_{1S} \dots \phi_{15S}$ ($\phi_{1S} \dots \phi_{15S}$) tartományban. A Q oszlop N-től balra mutatja, a p ARIMA paraméter értékének megfelelően melyik koefficienst választjuk. Ez a Q és R oszlop kombinálva alkotja az AC oszlopnál található ϕ_S mátrixot. Ez a ϕ mátrix mutatja, melyek az autoregressziós koeficiensek (ϕ_{NS}) ebben a számításban. Az EB oszlop tartalmazza mátrix szorzás eredményét. Ezeket az eredményeket érvényesítjük az EC, oszlopban, és SAR_t -nek nevezzük

$$(1 - \phi_{1S} B^{1S} - \phi_{2S} B^{2S} - \dots - \phi_{PS} B^{PS}) AR_t := SAR_t$$

a ϕ_S mátrix az AR_t mátrix

$$\begin{bmatrix} 1 \\ -\phi_{1S} \\ -\phi_{2S} \\ \vdots \\ -\phi_{15S} \end{bmatrix} X [AR_t \quad AR_{t-1S} \quad AR_{t-2S} \quad \dots \quad AR_{t-15S}] := SAR_t$$

5 Mozgóátlag Modul

Ez a modul az FU oszloptól FY oszlopig található a 21. sortól kezdve. Az FV oszlop számítja ki a mozgóátlag komponens szezonális alkomponens nélkül (MA). Az FW oszlop számítja ki a mozgóátlag komponens negyedéves szezonális alkomponenssel (SMA4). Az FX oszlop számítja ki a mozgóátlag komponens heti szezonális alkomponenssel (SMA7). És az FY oszlop számítja ki a mozgóátlag komponens havi szezonális alkomponenssel (SMA12)

A V oszlop a 28 sortól a 29 sorig tárolja az egyszerű mozgóátlag koeficiens θ_1 és θ_2 (theta1 és theta2) értékeit. A Z oszlop a 28 sortól kezdve a 29 sorig tárolja a szezonális mozgóátlag koeficiens θ_{1S} and θ_{2S} értékeit (theta1S and theta2S). Az U és X oszlop a V és Y oszloptól balra mutatja, a q és Q ARIMA paraméter értékének megfelelően melyik koefficienst választjuk. Ez U és V oszlop kombinálva és az X és Y oszlop kombinálva alkotja az FR illetve FS oszlopban található θ and θ_S mátrixokat. Például ha a q és Q ARIMA paraméter értéke 0 vagy üres, ez azt jelenti, a felhasználók mozgóátlag komponens nélküli ARIMA-t akarnak, akkor ez a két mátrix (θ and θ_S) 0

A J21 cella tárolja θ_0 (theta0) konstans értékét mutatja. A I21 cella a J21 cellától balra mutatja, választjuk-e ezt a koefficienst a C ARIMA paraméter értékének megfelelően

$$SAR_t = (1 - \theta_{1S} B_{1S} - \theta_{2S} B_{2S}) (1 - \theta_1 B_1 - \theta_2 B_2) e_t + \theta_0$$

$$[e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_{1S} e_{t-S} + \theta_1 \theta_{1S} e_{t-(S+1)} + \theta_2 \theta_{1S} e_{t-(S+2)} - \theta_{2S} e_{t-2S} + \theta_1 \theta_{2S} e_{t-(2S+1)} + \theta_2 \theta_{2S} e_{t-(2S+2)}]$$

a mozgóátlag komponens

$$e_t = SAR_t - [-\theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_{1S} e_{t-S} + \theta_1 \theta_{1S} e_{t-(S+1)} + \theta_2 \theta_{1S} e_{t-(S+2)} - \theta_{2S} e_{t-2S} + \theta_1 \theta_{2S} e_{t-(2S+1)} + \theta_2 \theta_{2S} e_{t-(2S+2)}] + \theta_0$$

Az MA szezonális komponense miatt, SMA4, SMA 7 és SMA12 különböző. A késleltetett hiba különbözőképpen jelenik meg. A mozgóátlag komponens első tagj ($-\theta_1 e_{t-1}$) először a 23. sorban jelenik meg, és folytatódik a következő sorban. A második tag ($-\theta_2 e_{t-2}$) először a 24. sorban jelenik meg, és folytatódik a következő sorban. A harmadik és negyedik tag, melynek szezonális alkomponense van, először a 25 és 26 sorban jelenik meg, és folytatódik a következő sorban SMA4-re, a 28 és 29 sorban SMA7-re, és 34 sorban SMA12-re

Következésképpen lesz e_t minden S-re GB - GE oszlopban But only the one that corresponds to the value of S in ARIMA parameter will be selected as the final e_t in column GF

$$e_t = SAR_t - MA_t + \theta_0$$

5 Összegző modul. A GA oszloptól a GH oszlopig tart a 21-ik sortól. Az ARIMA paraméterek alapján becsüli a hibát (e_t), amiből az Y^t becsült értéket határozza meg.

$$Y^t = Y_t - e_t \text{ (C oszlop, a kieső adatoknál nincs becsült érték.)}$$

Az A4-I4 cellákban láthatók a kiválasztott ARIMA változók. A D5-D14 cellák a számított értékeket és statisztikákat közli.

Kieső adatok száma=
 $p+d+q+s(P+D+Q)$
 Megfigyelések száma (n)
 Szabadságfok (df)
 R^2
 Számtani átlag(Y_t):
 Teljes négyzetösszeg
 Hibanégyzetösszeg (SSE)
 Korrigált reziduális szórás
 (s^*_e)
 Durbin-Watson statisztika
 Késleltetések száma 36

3.8.5 Spektrálanalízis.xls parancsfájl működése.

Az elméleti hatter részletesebb ismertetésétől eltekintünk, az megtalálható Pintér József tanulmányában.¹⁴⁷

Adatbevitel: Ctrl+a,

(Ctrl+a) a szezonok számának, az illesztéshez megadott megfigyelések számának, és az előrejelzés hosszának a változtatása.

(Ctrl+k) a Fourier modell becslése

(Ctrl+g) a grafikonok újrajárolása.

A Ctrl+ utasításokat az első munkalapon kell végrehajtani!

Ex-post ellenőrzésnél az FOURIER illesztést (becslést) végzi el a program a becslési időszak adatainak az alapján és kiszámítja az illesztett hibákat a becslési időszakra, továbbá az FOURIER modell előrejelzéseket készít, ha megadtuk a megfelelő előrejelzési időszakot, vagyis a teszt időszak adatainak a számát és mivel a tényadatokat is megadtuk a teszt időszakra, a program hibaképleteket tud számítani, amit az előrejelzett hibák néven közöl. A hibaképleteket a program kiszámolja az illesztett (becslési) és ebben az esetben az előrejelzett (teszt) időszakokra is. Célszerű először ex-post elemzést végezni és a legjobb FOURIER modell ismeretében ex-ante előrejelzéseket készíteni.

A rendelkezésre álló adatok száma: T

Szezonálitás (periódikus ingadozás) hossza = s (4 vagy nagyobb szám)

A teszt időszak száma: m

A becslési időszak száma: T-m

A becslésre használt adatok száma összesen T-m ($t=1,2,\dots,T-m$)

A számítások menete. Az alábbi számításokat végzi el a program:

Y_t = a megfigyelt adat a t-edik időpontban

$Y_t - Y_{t-s}$ a t-dik megfigyelt érték és a (t-s) időpontban megfigyelt érték különbsége.

Számtani átlag (Y_t):

$$\bar{Y} = \frac{1}{T-m} \sum_{t=1}^{T-m} Y_t$$

Trend számítása és kiszűrése:

$$T_t = a_0 + b_0 t$$

$$b_0 = \frac{\sum_{t=s+1}^{T-m} (Y_t - Y_{t-s})}{s}$$

$$a_0 = \bar{Y} - b_0 \bar{t}$$

$$\bar{t} = \frac{\sum_{i=1}^{T-m} t_i}{T-m}$$

$$\text{Trend kiszűrve} = Y_t - T_t$$

¹⁴⁷ Pintér József [2007]: A spektrálanalízisről. Statisztikai Szemle. 85. évf. 2.
http://www.ksh.hu/statszemle_archive/2007/2007_02/2007_02_130.pdf

A program kiszámítja az alábbi trigonometrikus függvényeket:

$\cos(\omega t)$, $\sin(\omega t)$, $\cos(2\omega t)$, $\sin(2\omega t)$, $\cos(3\omega t)$, $\sin(3\omega t)$, $\cos(4\omega t)$, $\sin(4\omega t)$

ahol:

f = frekvencia,

ω = körfrekvencia

A = amplitúdó, a legnagyobb kitérés

Y = a függvény általános alakja

s = a periódus (rezgésidő)

$$\omega = \frac{2\pi}{s} = \frac{2\pi f}{(T - m)}$$

$$f = \frac{(T - m)}{s}$$

$$Y = A \sin(\omega t)$$

Az amplitúdó megbecsléséhez a következő lineáris regressziószámítást kell elvégezni, konstanssal:

$$Y = \text{Trend kiszűrve} = Y_t - T_t$$

X változók: $X_1 = \cos(\omega t)$, $X_2 = \sin(\omega t)$, $X_3 = \cos(2\omega t)$, $X_4 = \sin(2\omega t)$, $X_5 = \cos(3\omega t)$, $X_6 = \sin(3\omega t)$, $X_7 = \cos(4\omega t)$, $X_8 = \sin(4\omega t)$

A kapott regressziós együtthatók: $a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8$.

$$(Y_t - T_t) = a_0 + a_1 \cos(\omega t) + a_2 \sin(\omega t) + a_3 \cos(2\omega t) + a_4 \sin(2\omega t) + a_5 \cos(3\omega t) + a_6 \sin(3\omega t) + a_7 \cos(4\omega t) + a_8 \sin(4\omega t)$$

$$H_0 : a_i = 0$$

$$H_1 : a_i \neq 0$$

$$H_0 : b_i = 0$$

$$H_1 : b_i \neq 0$$

A fenti hipotézisrendszert teszteljük.

Ha 5 %-os szignifikancia szinten elutasítjuk ($p < 0,05$) azt a H_0 hipotézist, hogy az a_i ($i=1,2,\dots,8$) paraméter egyenlő 0-val, akkor a paramétert felhasználjuk az amplitúdó számításához, ha a H_0 hipotézist fogadjuk el, ($p > 0,05$) akkor nem számítunk amplitúdót. Az A_1 amplitúdó számításához szükséges, hogy az a_1 és a b_1 is, szignifikánsan különbözzön 0-tól. Ugyanez érvényes az A_2, A_3 és A_4 amplitúdókra is. Természetesen dönthetünk úgy is, hogy a szignifikancia szint 10 %-os, akkor ha a $p < 0,1$, elutasítjuk a nulhipotézist. Az elfogadás esetén 1-t, az elutasítás esetén 0-t kell beírni a sárga cellákba.

Az amplitúdók számítása:

$$A_1 = \sqrt{a_1^2 + b_1^2}$$

$$A_2 = \sqrt{a_2^2 + b_2^2}$$

$$A_3 = \sqrt{a_3^2 + b_3^2}$$

$$A_4 = \sqrt{a_4^2 + b_4^2}$$

A becsült Fourier egyenlet:

$$\hat{Y}_t = T_t + A_1 \cos(\omega t) + A_2 \sin(\omega t) + A_3 \cos(2\omega t) + A_4 \sin(2\omega t) + A_5 \cos(3\omega t) + A_6 \sin(3\omega t) + A_7 \cos(4\omega t) + A_8 \sin(4\omega t)$$

Az illesztett és az ex-post hibaszámítás megegyezik az ARIMA.xls parancsfájlnál leírtakkal.

Kidolgozta: Mongkol Temrangsitornrat és Stephen DeLurgio (UMKC, University of Missouri-Kansas City), átalakította, lefordította és átprogramozta Kehl Dániel és Sipos Béla (PTE KTK).

forecast.umkc.edu/ftppub/BDS545/xls/Expos.xls

forecast.umkc.edu/ftppub/BDS545/xls/Fourier.xls

forecast.umkc.edu/ftppub/BDS545/xls/Winters.xls

forecast.umkc.edu/ftppub/BDS545/xls/Decomp.xls

forecast.umkc.edu/ftppub/BDS545/xls/forecast.xls

3.8.6 R+ interneten elérhető: Free Statistics Software (Calculator)

Adatcsere után (DATA törlendő, utána be lehet másolni az adatállományt, csak számokat akár 10 ezer adatot is) paramétereket be kell állítani és számol, hibás paraméterezés esetében üzenetet ad, Excelbe is lehet exportálni, a vesszőket ki kell cserélni, ha az adatokkal további számításokat végzünk az Excelben, pl. diagram készítése. Az eredményeket másolni is lehet Excelbe, ez célszerűbb, az ábrákat egyenként menteni kell utána be lehet illeszteni a dokumentumokba.

A számításokat három lépésben lehet elvégezni:

1. ACF és PACF számítása és tesztelése t-próbával. A stacionaritás biztosítása.

(Partial) Autocorrelation Function - Free Statistics Software (Calculator):

http://www.wessa.net/rwasp_autocorrelation.wasp#output

Küldje az eredményeket (Send output to:) a kiválasztott paramétereknek megfelelően:

Böngésző kék/fehér, grafikonok fehérek (Browser Blue – Charts White) alapbeállítás, maradhat.

Minta terjedelem, Sample Range: (hagyja üresen, hogy tartalmazza az összes bemásolt adatot)

A késleltetés értékének (K=Number of time lags) megadása, alapértelmezés (default) K=21, legnagyobb érték K=60, éves adatoknál célszerű ha K=21, havi adatoknál ha K=36.

A stacionaritás biztosítása paraméterezéssel, Box-Cox transzformáció, Lambda (λ) lehetséges értékei: 2 és -2 között, $\lambda=0$, logaritmus transzformáció, a többi esetben az alábbi képlet szerint számol:

$$Y(\lambda) = \begin{cases} (Y^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \log(Y), & \lambda = 0 \end{cases} \quad Y > 0$$

Nem szezonális (d=0,1,2) differenciaképzés (Degree of non-seasonal differencing, d=0,1,2) Ha d=0, az konvencionálisan azt jelenti, hogy nincs nem szezonális differenciaképzés, ha d=1, akkor az idősor első differenciáit ($y_t - y_{t-1}$) képezi, ami az idősor szomszédos adatai közötti különbséget jelenti. Az idősor egy adattal rövidül.

Ha d=2, akkor a második differenciákat $[(y_t - y_{t-1}) - (y_{t-1} - y_{t-2})]$ számítja. Az idősor két adattal rövidül.

A szezonális (D=0,1,2) differenciaképzés, (Degree of seasonal differencing (D=0,1,2)). Ha D=0 az konvencionálisan azt jelenti, hogy nincs szezonális differencia képzés, ha D=1 akkor szezonális differenciákat ($y_t - y_{t-12}$) képezünk, havi adatoknál az idősor 12 adattal rövidül, negyedéves adatoknál a differenciaképzés eredményeképpen ($y_t - y_{t-4}$) az idősor 4 adattal rövidül. Ha D=2, akkor a differenciálást $[(y_t - y_{t-12}) - (y_{t-1} - y_{t-13})]$ folytatjuk, és az idősor például havi adatok esetében 13 adattal rövidül.

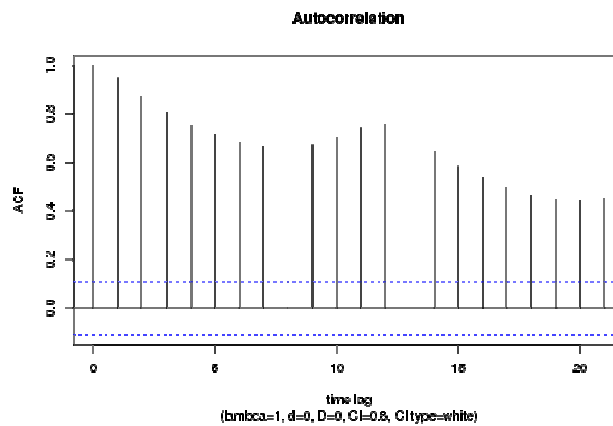
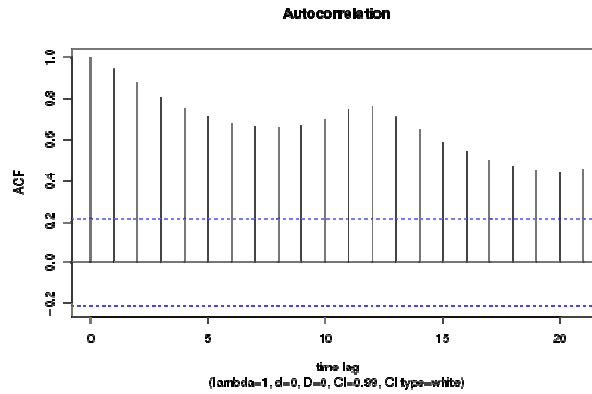
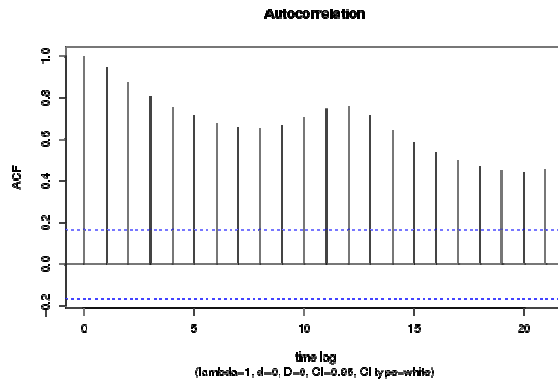
A szezonális periódusa (Seasonality) (nincs szezonális, tehát éves adatok esetében S=1, havi adatoknál S=12, negyedéves adatoknál S=4, fél éves adatok=2, maximális értéke S=12)

Lehetőségek: s=1,2,3,4,6,12.

A konfidencia intervallumot (CI type) kiszámítja két módon (The confidence interval can be computed in two different ways):

- Feltételezzük, hogy a vizsgált idősor fehér zaj folyamatot követ, (assuming a white noise time series (CI type = White Noise) Ez az alapeset.
- Feltételezzük, hogy a vizsgált idősor egy MA (k-1) folyamatot követ amikor az ACF (k) számítása ennek alapján történik (CI type = MA) (assuming that the series is a MA(k-1) process when the CI of ACF(k) is computed (CI type = MA) Ez félrevezető lehet.

A konfidencia intervallum (Confidence Interval=CI) alapesetben 0,95, tehát 5%-os a szignifikancia szint. Ezt lehet változtatni, pl. 0,99, akkor ez 1 %-os szignifikancia szint. A CI nagyságának növelésével, a konfidencia intervallum is nő. Az alapmodell esetében, ha nem alkalmazunk transzformációt, csak a CI értéket változtatjuk, az alábbi módon változik a konfidencia intervallum, ha CI=0,95, 0,99 és 0,80:



Ugyanis: ha szignifikancia - szint (α) a kétoldalú a standard normális eloszlás esetében, akkor, ha megkeressük a z értékeket a táblázatban, az alábbi konfidencia intervallumokat kapjuk:

$\alpha=0,05$, $CI=0,95$, $z=1,96$, a konfidencia intervallum: $\pm 1,96/\sqrt{n} = \pm 1,96/\sqrt{144} = \pm 0,163$

$\alpha=0,01$, $CI=0,99$, $z=2,587$, a konfidencia intervallum: $\pm 2,587/\sqrt{n} = \pm 2,587/\sqrt{144} = \pm 0,216$

$\alpha=0,2$, $CI=0,80$, $z=1,280$, a konfidencia intervallum: $\pm 1,280/\sqrt{n} = \pm 1,280/\sqrt{144} = \pm 0,107$

Ha nem akarjuk használni a Box-Cox transzformációt, akkor a logaritmus alapját be kell írunk, pl. 10, e alapunál 2.71, a tizedesvessző helyett pontot kell választani.

Kiszámít (Compute) elvégzi a számításokat.

Meghatározza a K értékeknek megfelelően az ACF és PACF értékeket, a t-statisztikákat és a p-értékeket. A nullhipotézis értelmében az y_t és az y_{t-k} változók között nincs szignifikáns autokorreláció, ennek elvetése az autokorrelációs kapcsolat szignifikáns voltát igazolja. Ha a p nagyobb, mint 0,05 akkor 5 %-os szignifikancia szinten elfogadjuk a nullhipotézist, ellenkező esetben az alternatív hipotézist fogadjuk el. A stacionaritás biztosítása szempontjából 5 %-os szignifikancia szinten az a kedvező, ha p nagyobb, mint 0,05. A p-érték az a legkisebb érték, ami mellett a nullhipotézis elutasítható. Közli a program az eredeti és

transzformált adatok ábráit és a korrelogramokat is, ahol feltünteti a választott paramétereket (lambda, d, D, CI, type). Nem közli viszont a transzformált adatokat.

2. ARIMA Backward Selection - Free Statistics Software (Calculator)

ARIMA becslés, az interneten. Backward eliminációs módszer, visszafelé történő választás.

http://www.wessa.net/rwasp_arimabackwardselection.wasp#output

Küldje az eredményeket (Send output to:) a kiválasztott paramétereknek megfelelően:

Böngésző kék/fehér, grafikonok fehérek (Browser Blue – Charts White) alapbeállítás, maradhat.

A paraméterek (Lambda, d, D, s, CI type) beállítása, az 1. ponban kipróbált azon paraméterek beállításával történik, amelyek esetén a stacionaritás legjobban biztosítható. Az ARIMA paraméterek beállításánál, ha a stacionaritás biztosított, az ACF és PACF ábrák alapján meg kell nézni, hogy milyen modellel [AR(p), MA(q), SAR(P), SMA(Q)] célszerű próbálkozni, a lehetőségeket az „ARIMA modellek jellemzése” táblázat tartalmazza. Meg kell jegyezni az 1. pont „legjobb” paramétereit, (lambda, d, D, CI, type). és a becslésnél ezeket kell használni.

A másik megoldás. Érdekes a legnagyobb megadható paraméterekkel indulni és kihagyni a nem szignifikáns paramétereket, ahol pl. a p-empirikus érték nagyobb, mint 0,05.

Minta terjedelem, Sample Range: (hagyja üresen, hogy tartalmazza az összes bemásolt adatot)

Benne legyen az átlag, Include mean? A válasz alapesetben nem, fehér zajt (white noise) feltételezünk (False), ha igen True, véletlen bolyongási folyamatot (random walk) feltételezünk.

A stacionaritás biztosítása paraméterezéssel, Box-Cox transzformáció, Lambda (λ) lehetséges értékei: 2 és -2 között, $\lambda=0$, logaritmus transzformáció. (Box-Cox lambda transformation parameter (lambda).

Nem szezonális (d=0,1,2) differenciaképzés (Degree of non-seasonal differencing, d=0,1,2)

A szezonális (D=0,1,2) differenciaképzés, (Degree of seasonal differencing (D=0,1,2).

A szezonális periódusa (Seasonality=s) s=1,2,3,4,6,12

AR(p) = p az autoregresszivitás rendjét jelöli (Maximum AR(p) order) lehet: 0,1,2,3.

Ha p=0, nem használjuk az AR modellt.

Ha p=1, akkor ARIMA (1, 0, 0) vagy AR (1) modellt becslünk:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

Ha p=2 akkor ARIMA (2, 0, 0) vagy AR (2) modellt becslünk:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

Ha p=3 akkor ARIMA (3, 0, 0) vagy AR (3) modellt becslünk:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \varepsilon_t$$

MA(q)= q a mozgóátlag folyamat rendjét jelöli. (Maximum MA(q)=order) lehet: 0,1.

Ha q=0, akkor nem használjuk az MA modellt.

Ha q=1 akkor: ARIMA (0, 0, 1) vagy MA (1) modellt becslünk:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Az AR(p) és MA(q) kombinálásával a modellek igen sok variációja állítható elő. Az alacsonyabb rendű vegyes ARMA modellek az alábbi módon írhatók fel:

$$\text{ARIMA (1, 0, 1)} \quad Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$\text{ARIMA (2, 0, 1)} \quad Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$\text{ARIMA (3, 0, 1)} \quad Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

SAR(P) P az autoregresszivitás rendjét jelöli a szezonális modell esetében (Maximum SAR(p) order) lehet: 0,1,2.

SMA(Q) Q a mozgóátlag folyamat rendjét jelöli a szezonális modell esetében. (Maximum SMA(Q)=order) lehet: 0,1.

Az eredményeket a kijelölt paraméterekkel és p-értékekkel iterációként adja meg. (ARIMA Parameter Estimation and Backward Selection) A paraméterek szignifikanciáját kell vizsgálni. Közli a program a reziduumokat (Estimated ARIMA Residuals) és az ACF és PACF ábrákat konfidencia intervallummal.

A paraméterek (p, q, P, Q) kombinálásával ellenőrizhető, hogy melyik paraméter esetében kisebb a p-érték pl. 0,05-nél (5 %-os szignifikancia szint), tehát a paraméter szignifikánsan különbözik 0-tól, és amelyik paraméternél nem különbözik szignifikánsan a becslt paraméter 0-tól, mert a p-érték nagyobb, mint 0,05, ott változtatni érdemes és ki kell hagyni a paramétert vagy módosítani kell, a megadott lehetőségeken belül. A számítások gyorsan elvégezhető, az elméletileg lehetséges változatok száma

$4*2*3*2=48$, tehát elméletileg 48 féle modell futatható le. A valós változatok száma ennél lényegesen kisebb, ha figyelembe vesszük azt, hogy különböző, stb. λ , d , D paraméterek esetében megnézzük az ACF és PACF kolleogramokat (1. lépés) és kiválasszuk azt a paraméterkombinációt, ami mellett a stacionaritás biztosítható. Ha nem biztosítható a stacionaritás, akkor az ARIMA becslést már nem célszerű elvégezni.

3. ARIMA előrejelzés, az interneten.

http://www.wessa.net/rwasp_arimaforecasting.wasp

Ugyanazon paramétereket [λ , d , D , s , AR(p), MA(q), SAR(P) SMA(Q), include MEAN, FALSE/TRUE] kell megadni, mint az 1 és 2 pontban ismertettünk, értelem szerűen a legjobb változatot kell kiválasztani, ahol az elméleti feltételek biztosítottak. Plusz adat a tesztperiódus megadása (Testing Period=TP) maximális értéke 24, lehet, célszerű havi adatoknál 12-t választani, ez egy év, negyedéves adatoknál 4-t megadni, ami szintén egy év.

A számítások elvégzése (compute) után, közli a becsült értékeket, a konfidencia intervallumot, a p -értéket, a nullhipotézist: ($H_0: Y[t] = F[t]$) és a valószínűségeket az alábbi esetekre:

$$P(F[t] > Y[t-1])$$

$$P(F[t] > Y[t-s])$$

$$P(F[t] > Y[n-TP])$$

F a becsült és Y a tényleges adat.

A teszt periódusra pl. $t=1,2,\dots,24$ megadja a hibaképletek nagyságát. $Y(t)$ az eredeti adatok, $F(t)$, a becsült adatok.

Kieső adatok száma, a korábban megadott képlettel számol a program= $p+d+q+s(P+D+Q)$

PE: relatív hiba, PERCENTAGE ERROR:

$$PE = \frac{(Y[t]-F[t])}{F[t]}$$

Négyzetes hiba: SQUARE ERROR:

$$e_i^2 = Sq, E = (Y[t]-F[t])^2$$

MSE: Átlagos négyzetes hiba: MEAN SQUARE ERROR:

$$MSE = \frac{\sum_{i=1}^t e_i^2}{t}$$

$$e_i^2 = (Y[t]-F[t])^2$$

RMSE, a hiba szórása, az átlagos négyzetes hiba gyöke, ROOT MEAN SQUARE ERROR:

$$RMSE = \sqrt{\frac{\sum_{i=1}^t e_i^2}{t}}$$

$$e_i^2 = (Y[t]-F[t])^2$$

MAPE, Átlagos relatív abszolút hiba, MEAN ABSOLUTE PERCENTAGE ERROR:

$$MAPE = \frac{\sum_{i=1}^t |PE_i|}{t}$$

$$PE_t = \frac{(Y[t]-F[t])}{F[t]}$$

Közli végül az előrejelzés grafikonját is.

Wessa P., (2009), ARIMA Forecasting (v1.0.5) in Free Statistics Software (v1.1.23-r7), Office for Research Development and Education, URL http://www.wessa.net/rwasp_arimaforecasting.wasp/

The R code is based on :

Borghers, E, and P. Wessa, Statistics - Econometrics - Forecasting, Office for Research Development and Education, <http://www.xycoon.com/>

Leírás:

http://www.stat.pitt.edu/stoffer/tsa2/R_time_series_quick_fix.htm

4. A korreláció- és regressziószámítás

A regressziós modell készítésének¹⁴⁸ első lépése a specifikáció, amin a jelenséget leíró, modellben szereplő változók, az eredmény- és a magyarázó változók kiválasztását, valamint a függvény konkrét formájának meghatározását értjük. Fontos szerepet játszik a specifikáció szakaszában az adatbázis, amelynek minősége, szerkezete nagymértékben befolyásolja a specifikáció eredményességét. A gyakorlati munkában idősoros és keresztmetszeti adatokkal dolgozhatunk, ennek a modell feltételrendszerének ellenőrzésekor lesz jelentősége. Panel adatbázisokkal jelen anyagunkban nem foglalkozunk.

A specifikáció munkafázisának lezárása után a számításokat a regresszio.xls parancsfájllal lehet elvégezni. Ennek fontosabb lépései a következők:

A regressziós paraméterek becslése a klasszikus legkisebb négyzetek módszerével, melynek feltételei:

1. A magyarázó változók nem sztochasztikusak, tehát mérési hibát nem tartalmaznak és lineárisan függetlenek (multikollinearitás hiánya), a hibatényezők (hibatagok, reziduumok) várható értéke 0, varianciájuk konstans, normális eloszlásúak és nem autokorreláltak.
2. A modell feltételrendszerének ellenőrzése. Ez a munkafázis visszahat mind a specifikációra, mind a paraméterbecslésre. Ebben a munkaszakaszban a modellező megállapítja, hogy egy adott megbízhatóság mellett mennyire fogadható el a modell. A fontosabb hipotézisellenőrzések: a regressziós modell paramétereinek globális és parciális tesztelése (a paraméterbecslés pontosságának vizsgálata, a paraméterek standard hibája, konfidencia intervalluma stb.), valamint a reziduumok vizsgálata: az autokorreláció és a homoszkedaszticitás tesztelése, és a magyarázó változók közötti kapcsolat szorossága, a multikollinearitás ellenőrzése. A próbákkal nyert információk alapján döntést lehet hozni a modell esetleges megváltoztatásáról, vagy a becslési módszer módosításáról. Ezek a döntések természetesen visszahatnak a specifikációra és indokolt esetben az egész eljárás (specifikáció, becslés, hipotézisellenőrzés) megismétlését igényelhetik.
3. A regressziós modellek felhasználása elemzésre és előrejelzésre.
4. A verifikálás, aminek során a modellt szembesítjük valósággal.

Az általunk vizsgált **regressziós modellekben** egy eredményváltozó és egy vagy több magyarázó változó van. A modell alapján megállapítható, hogy a tényezőváltozó(k) milyen módon és milyen törvényszerűség mellett fejti(k) ki hatását (hatásukat) az eredményváltozóra.

A többváltozós regressziós modell általános formája:

$$Y = f(X_1, X_2, \dots, X_k, \varepsilon)$$

ahol :

Y = az eredményváltozó,

X_j = az j-edik tényező (magyarázó) változó; (j=1,2,...,k),

ε = a véletlen változó ; maradéktag ; reziduum,

n = a megfigyelések száma, i=1,2...n

f= függvény típusa, alapesetben lineáris függvény.

Feltételek az ε véletlen változóra (hibatényezőre, reziduális változóra), az Y eredmény- és az X_j tényezőváltozókra vonatkozóan:

1. Az ε véletlen változó várható értéke nulla, vagyis a modellben szereplő változók nem hoznak létre szisztematikus hatást a hibatényezőben, a pozitív és negatív értékek kiegyenlítik egymást.
2. A véletlen változó értékei páronként nem korrelálnak egymással (idősoroknál nincs autokorreláció); ellenkező esetben autokorrelált a modell, ebben az esetben elemzésre és előrejelzésre nem használható.

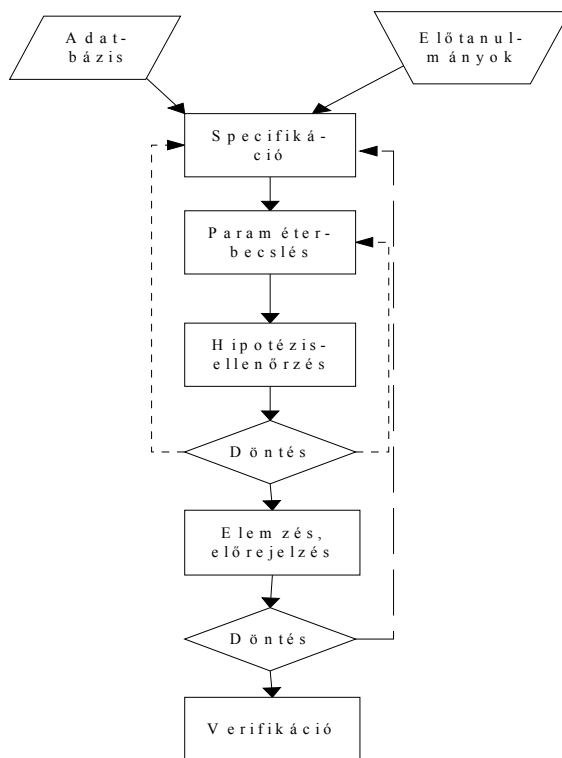
¹⁴⁸ Hajdu O. - Herman S. - Pintér J. - Rappai G. - Rédey K. [1994-95]:110-111.

3. A véletlen változó (hibatényező) szórásnégyzete (varianciája) állandó, tehát a véletlen változó homoszkedasztikus; ellenkező esetben heteroszkedasztikus a modell. Keresztmetszeti adatoknál jelent elsősorban problémát a heteroszkedaszticitás, ebben az esetben a modell elemzésre és előrejelzésre nem használható.
4. A magyarázó változók lineárisan (sztochasztikusan is) függetlenek, értékük rögzített, mérési hibát nem tartalmaznak és nem korrelálnak a hibatényezővel. Tehát: a magyarázó változók száma kisebb mint a megfigyelések száma ($k < n$), a magyarázó változók mátrixának rangja megegyezik k -val, a tényező- (magyarázó-) változók számával, azaz az $X'X$ mátrix invertálható, mivel a mátrix rangja¹⁴⁹ és rendje¹⁵⁰ egyenlő, vagyis a mátrix nem szinguláris. Ha ez a feltétel nem áll fenn, akkor a modellben a multikollinearitás jelenlétével kell számolni¹⁵¹, ami zavarja a modell helyes specifikálását és általában csökkenti a modelltől nyerhető információ minőségét. Elemzésre nem, előrejelzésre viszont felhasználható a modell, ha feltételezzük, hogy a multikollinearitás mértéke, iránya a jövőben nem változik.
5. Az eredményváltozó megfigyelt értékei (Y_i $i = 1, 2, 3, \dots, n$) normális eloszlásúak, függetlenek egymástól, érvényesül a linearitás, vagyis a regressziós paraméterek egyértelműen meghatározzák a regressziós egyenest.

A fent említett feltételek teljesülése esetén a regressziós modellt multikollinearitástól mentes **standard lineáris regressziós modellnek** tekinthetjük. Belátható, hogy ezek az igen "szigorú" feltételek, a modellépítés alapelvét képezik, amire a munka során fokozottan figyelniük kell. Mindez azt jelenti, hogy a modell konkretizálása során a feltételeket folyamatosan kell ellenőrizni, az eredményeket hipotetikus jellegűeknek kell tekinteni. Amennyiben a mintabeli adatokkal nem igazolhatók empirikusan az elméleti várakozások, a specifikáció módosítására, óvatos értelmezésre, illetve sajátos, új becslési módszer megválasztására van szükség.

A feltételek teljesülése esetén a lineáris regressziós modell paraméter-becslésére a klasszikus legkisebb négyzetek módszere (KLNМ) alkalmazható.

A modellalkotás folyamatát az alábbi ábra mutatja.



4.1 ábra: A regressziós modellezés lépései

¹⁴⁹ független oszlopvektorainak száma.

¹⁵⁰ a magyarázó változók száma

¹⁵¹ szélsőséges esete, ha a mátrix szinguláris, tehát nem invertálható.

4.1 A regresszió.xls parancsfájl működése

A program¹⁵² a bemutatásra kerülő regressziószámítást maximum 16 magyarázóváltozó és 2000 megfigyelés esetében végzi el¹⁵³. A programban megjelenő színeknek külön jelentése van. A halványsárga cellák változtathatók, itt történik meg az adatok bevitele, a kívánt szignifikancia-szint beállítása, valamint a becslés/előrejelzés adatainak megadása. A tesztek végeredményei színes számokkal jelennek meg a fájlban. A modell ellenőrzésénél háromféle szint alkalmaztunk, a modellezést zavaró eredmények piros, a megfelelő eredmények zöld, a nem egyértelmű eredmények kék színnel jelennek meg.

Tanulmányunkban az elméleti háttér részletes ismertetésétől eltekintünk (kivéve a homoszkedaszticitás tesztjeit, ahol a felsőoktatásban kevésbé ismert tesztet alkalmazunk), mert ezek az ismeretek az irodalomjegyzékben is felsorolt szak-, illetve tankönyvekben, tanulmányokban megtalálhatók, célunk csupán a szoftver bemutatása, gyakorlati, oktatási célokra való közreadása.

A fájl nyolc munkalapból áll, amelyek rendre:

- Adat
- Mátrix
- Maradék
- Multikollinearitás
- Autokorreláció
- Homoszkedaszticitás
- Idősoros adatok-példa
- Keresztmetszeti adatok-példa

Az alábbiakban a fenti munkalapok tartalmát ismertetjük, majd szemléltető példákön mutatjuk be a fájl alkalmazását. Mivel a program képes az autokorreláció és a homoszkedaszticitás tesztelésére is, ezért két példán keresztül szemléltetjük a számításokat. Az első példa idősoros adatállomány, a második pedig keresztmetszeti, az adatállományokat elhelyeztük a regresszio.xls fájlban.

4.1.1 Az Adat munkalap

A munkalap két nagyobb egységből áll. A bal oldali, sárgával jelölt terület az adatok tárolására, bevitelére szolgál, itt kell rögzíteni az aktuális adatállományt. Új adatok bevitele előtt a megjelenő mintafeladat adatállományát az adatok törlése gombra való kattintással törölhetjük. Az új adatok beillesztése a szokásos módon történhet, ám szükséges az adatok értéként való beillesztése, annak érdekében, hogy a parancsfájl formátuma megmaradjon.

A jobb oldali egység a regressziós modell alapstatisztikáit közli.

Regressziós statisztika:

- R: többszörös korrelációs együttható;
- R^2 : többszörös determinációs együttható;
- \tilde{R}^2 : korrigált determinációs együttható;
- s: modell standard hibája;
- n: megfigyelések száma.

Varianciaanalízis: a többváltozós regressziós modell varianciaanalízis táblája.

Regressziós együtthatók:

- együtthatók értéke és standard hibája, t-értékei, p-értékei és konfidencia intervallumai (tetszőleges megbízhatósági szinten);
- változók bevonásáról/kihagyásáról döntő jelölőnégyzetek.

¹⁵² Az alapötletet adta: Kiss Tibor [1988-1992]: REGAL, Szakértői rendszer többváltozós regresszió-analízisre (DOS program). Ld. még leírását: Kiss Tibor – Sipos Béla [1998]: REGAL.

¹⁵³ A magyarázóváltozókra vonatkozó korlát az Excel sajátja. A megfigyelésekre vonatkozó korlát igény szerint bővíthető, a korlátozás oka a gyors számítási sebesség megtartása.

A formátum tehát követi az Excel adatelemző menüpontja által használtat, azzal a különbséggel, hogy az egyes cellák most Excel függvényeket tartalmaznak, így az adatok megváltozásának hatása azonnal nyomon követhető az alaperedményeken. Szintén eltérés a beépített funkcióhoz képest, hogy az eredeti adatok meghagyása mellett is kihagyhatunk, illetve újra bevonhatunk változókat a paraméterek soraiban található jelölőnégyzetek segítségével.

A varianciaanalízis tábla segítségével a modell globális próbáját végezhetjük el. A hipotézisrendszerrel való döntés – didaktikai okokból – két módon is elvégezhető: tetszőlegesen beállítható szignifikancia-szinthez tartozó kritikus érték, valamint p-érték alapján is.

A gyors parciális tesztelés lehetőséget biztosít a backward eliminációs módszer alkalmazására. A módszer lényege, hogy az első lépésben olyan regressziós függvényt határozunk meg, amely az összes megfigyelt magyarázó változót tartalmazza, majd az így meghatározott regresszió függvényből kihagyjuk lépésenként azokat a változókat, amelyek nem járulnak hozzá szignifikánsan a reziduális négyzetösszeg csökkentéséhez. A változók szelektálásához a p-értékeket használjuk: ha a p-érték magasabb, mint amit megengedtünk (pl. 0,05), akkor elfogadjuk azt a nullhipotézist, hogy a regressziós paraméter nem különbözik szignifikánsan a nullától. Amennyiben több változó p-értéke is a kívántnál magasabb, úgy a legmagasabb értékkel rendelkező változót hagyjuk ki. Az eliminációt addig folytatjuk, míg valamennyi bevont paraméter szignifikáns nem lesz.

A változók szelektálását természetesen elvégezhetjük a multikollinearitás, vagy a homoszkedaszticitás parciális tesztjei, vagy szakmai ismeretek alapján is.

A felhasznált képletek:

\mathbf{R} = a korrelációs mátrix¹⁵⁴, amely négyzetes és mérete $(k+1)*(k+1)$, az első sor és oszlop az eredményváltozó, a többi sor és oszlop pedig a magyarázó változó korrelációs együtthatóit tartalmazza, a mátrix szimmetrikus, a mátrix az egyszerű, kétváltozós korrelációs együtthatókból áll, számításukat a mátrix alatt tüntettük fel, a diagonális elemek (adott változó önmagával számított korrelációja) 1-gyel egyenlők.

$$\mathbf{R} = \begin{bmatrix} 1 & r_{y1} & r_{y2} & \cdots & r_{yk} \\ r_{1y} & 1 & r_{12} & \cdots & r_{1k} \\ r_{2y} & r_{21} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{ky} & r_{k1} & r_{k2} & \cdots & 1 \end{bmatrix}$$

Az y és x_j közötti lineáris korrelációs együttható jele r_{yj} , az x_i és x_l közöttié pedig r_{jl} .

$$r_{yj} = \frac{\sum (x_j - \bar{x})(y_i - \bar{y})}{n\sigma_x \sigma_y}$$

$$r_{jl} = \frac{\sum (x_j - \bar{x})(x_l - \bar{x})}{n\sigma_{x_j} \sigma_{x_l}}$$

\mathbf{R} = a többszörös korrelációs együttható:

$$R = \sqrt{1 - \frac{1}{q_{yy}}}$$

Ahol q_{yy} a korrelációs mátrix inverzéből ($\mathbf{R}^{-1}=\mathbf{Q}$) nyerhető:

¹⁵⁴ A Mátrix munkalapon található.

$$\mathbf{R}^{-1} = \mathbf{Q} = \begin{bmatrix} q_{yy} & q_{y1} & q_{y2} & \cdots & q_{yk} \\ q_{1y} & q_{11} & q_{12} & \cdots & q_{1k} \\ q_{2y} & q_{21} & q_{22} & \cdots & q_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{ky} & q_{k1} & q_{k2} & \cdots & q_{kk} \end{bmatrix}$$

A többszörös korrelációs együttható (R) azt méri, hogy a magyarázó változók az eredményváltozóval együttesen milyen szoros kapcsolatban vannak.

R^2 = a többszörös determinációs együttható (jele: R^2) kifejezi, hogy mekkora hányadban magyarázzák meg együttesen a magyarázó változók az eredményváltozó varianciáját (szórásnégyzetét). A kapcsolatok jellegének minősítésén túl fontos szerepet tölt be a többszörös determinációs együttható a regressziós modell megítélésében. A mutató nagyobb értéke egyben azt is jelenti, hogy jobban illeszkedik a modell.

$$R^2 = 1 - \frac{1}{q_{yy}}$$

\tilde{R}^2 = a korrigált determinációs együttható: a többváltozós regressziós modellek esetében gyakran felléphet egy olyan jelenség, amely félreinformálhatja az elemzőt. Az R^2 ugyanis nagyobb magyarázó erővel bír, ha több magyarázó változó hatása szerepel benne, függetlenül attól, hogy valóban releváns hatást fejt-e ki mindegyik magyarázó változó. (Például megtévesztő lehet az R^2 alapján két modell összehasonlítása, ha az egyik három, a másik hét magyarázó változót tartalmaz.) A modellek összehasonlítása esetében a különböző számú magyarázó változóból eredő problémát próbálja feloldani az ún. korrigált vagy a szabadságfokokkal korrigált determinációs együttható (jele: \tilde{R}^2):

$$\tilde{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$$

Az s = a modell általános standard hibája, jelzi az illeszkedés jóságát, a modell annál pontosabban illeszkedik, minél kisebb az értéke. Meghatározására a regressziós paraméterek ismeretében kerülhet sor, amikor kiszámítva az y eredményváltozó becsült értékeit (\hat{y}) képezhetjük a reziduumokat ($e = y - \hat{y}$).

$$s = \sqrt{\frac{\sum (y - \hat{y})^2}{n - k - 1}} = \sqrt{\frac{\sum e^2}{n - k - 1}}$$

A regressziós modell egészének tesztelése, a globális F-próba

A **varianciaanalízis** az Adat munkalapon jelenik meg. A varianciaanalízis összefoglalja az alábbi nullhipotézis ellenőrzésére vonatkozó eredményt. Nullhipotézisünk az, hogy a magyarázó változók regressziós együtthatói mind 0-k, az alternatív hipotézis szerint létezik legalább egy 0-tól eltérő együttható.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \exists \beta_j \neq 0$$

Az ellenhipotézis elfogadása esetén azt állíthatjuk, hogy van legalább egy olyan magyarázó változó, amely szignifikáns hatással rendelkezik, tehát létezik legalább egy nullától eltérő értékű paraméter. A nullhipotézis a lineáris regresszió fennállásának tagadását jelenti és amennyiben igaz, úgy az eredményváltozó kizárólag a véletlen hatására szóródik; az alternatív hipotézis fennállása esetén a regressziós modellt elfogadhatónak ítéljük. A nullhipotézis ellenőrzését az alábbi varianciaanalízis 4-1. tábla alapján végezhetjük el.

4-1. tábla: Varianciaanalízis¹⁵⁵

Összetevő	Négyzetösszeg (SS)	Szabadság fok (df)	Szórásnégyzet becslése (MS)
Regresszió (SSR)	$S_{\hat{y}} = \sum (\hat{y}_i - \bar{y})^2$	k	$s_{\hat{y}} = S_{\hat{y}} / k$
Maradék (SSE)	$S_e = \sum (y_i - \hat{y}_i)^2$	n-k-1	$s^2 = S_e / (n - k - 1)$
Teljes (SST)	$S_y = \sum (y_i - \bar{y})^2$	n-1	$s_y = S_y / (n - 1)$

A próbafüggvény:

$$F = \frac{S_{\hat{y}} / k}{S_e / (n - k - 1)}$$

A többszörös determinációs együttható segítségével is ellenőrizhetjük a modell magyarázó erejét, az alábbi módon:

$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}$$

Ha a számított F érték nagyobb, mint a táblabeli érték $F_{0,05 [k, (n-k-1)]}$ ahol a szignifikancia-szint 5%, a számláló szabadságfoka k, a nevező (n-k-1), akkor a regressziós modellt elfogadhatónak ítéljük. Ellenkező esetben elvetjük. Az F-próbával az egész modellt teszteljük, mert arra a kérdésre keressük a választ, hogy érdemes-e a regresszió-számítást, mint elemzési módszert alkalmazni. Ha nem jó a modell, tehát esetünkben a modellünk egészében rossz, akkor a regressziós modell alkalmazását elvetjük, és egyszerűbb eljárásokkal, pl. átlagszámítással kell dolgozni. Elfogadjuk tehát a regressziós modellt pl. 5%-os szignifikancia-szinten, ha:

$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)} > F_{0,05[k, (n-k-1)]}$$

Nem fogadjuk el a regressziós modellt pl. 5%-os szignifikancia-szinten, ha:

$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)} < F_{0,05[k, (n-k-1)]}$$

Az F-számított érték nagysága az R^2 nagyságától függ, pl. ha az $R^2 = 0$, az F-érték is nulla, az R^2 növekedésével, a számított F-érték is nő, ha az $R^2 = 1$, akkor az F-érték tart a végtelenhez. ($F \rightarrow \infty$)

A varianciaanalízis tábla segítségével tehát a modell globális próbáját végezhetjük el. A hipotézisrendszerről való döntés – didaktikai okokból – két módon is elvégezhető: tetszőlegesen beállítható szignifikancia-szinthez tartozó kritikus érték, valamint p-érték alapján is, ezért mindkét eredményt közöljük..

Regressziós együtthatók:

- együtthatók értéke és standard hibája, t-értékei, p-értékei és konfidencia intervallumai (tetszőleges megbízhatósági szinten)
- változók bevonásáról/kihagyásáról döntő jelölőnégyzetek

Az alkalmazott képletek:

A regressziós paraméterek (együttható: Eható) konfidencia-intervallumait (Alsó 95%) (Felső 95%) kiszámítja a parancsfájl, alapesetben a konfidencia intervallum 95% és $t_{krit}=5\%$.

¹⁵⁵ Az eltérés-négyzetösszeg angol megfelelője (Sum of Squares) alapján SS szimbólummal jelöljük, R=Regression, E=Error, T=Total, df= degrees of freedom, MS=Mean Square.

A b_j regressziós paraméter konfidencia intervalluma $(1 - \alpha) \times 100$ valószínűségi szinten %-ban:

$$b_j \pm t_{(1-\alpha/2)(n-k-1)} s_{b_j}$$

Ahol:

- s_{b_j} = a j -ik paraméter standard hibája (St hiba),
- $t_{(1-\alpha/2)(n-k-1)}$ = a Student-féle t-próba kritikus értéke az $(n-k-1)$ szabadságfoknál, az $\alpha/2$ szignifikancia-szinten.

Az általunk előre megválasztott szignifikancia-szint alapesetben 5%-os ($p=0,05$) valószínűség (konfidencia intervallum 95%), ami változtatható a sárga cellában: pl. 1% (konfidencia intervallum 99% és $t_{krit}=1\%$, $p=0,01$), 10 % (konfidencia intervallum 90% és $t_{krit}=10\%$, $p=0,1$).

Az α szignifikancia-szint valószínűségének csökkentésével illetve a konfidencia intervallum valószínűségének növelésével az adott $(n-k-1)$ szabadságfok esetén a Student-féle t-eloszlás kritikus értékei is nagyobb számok lesznek és így a regressziós paraméterek konfidencia intervallumai is növekednek. Pl. ha az alapeset 95%-os konfidencia intervallum valószínűségét 99%-ra növeljük. Ez fordítva is igaz, az α szignifikancia-szint valószínűségének növelésével illetve a konfidencia intervallum valószínűségének csökkentésével az adott $(n-k-1)$ szabadságfok esetén a Student-féle t-eloszlás kritikus értékei is kisebb számok lesznek és így a regressziós paraméterek konfidencia intervallumai is csökkennek. Pl. az alapeset 95%-os konfidencia intervallum valószínűségét 90%-ra csökkentjük.

A Backward eliminációs módszer

A paraméterek szeparált tesztelésénél tehát a nullhipotézisünk az, hogy a j -edik ($j=1,2,\dots,k$) regressziós paraméter értéke 0, az alternatív hipotézisünk pedig az, hogy nem, azaz

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

A nullhipotézis elfogadása azt jelenti, hogy a j -edik magyarázó változó nem magyarázza az eredményváltozót, tehát a modellben való megtartása felesleges, esetleg káros.

A próbafüggvény a nullhipotézis fennállása esetén

$$t = \frac{b_j}{s_{b_j}} < t_{(1-\alpha/2)(n-k-1)}$$

A próbafüggvény a nullhipotézis elutasítása esetén

$$t = \frac{b_j}{s_{b_j}} > t_{(1-\alpha/2)(n-k-1)}$$

Ahol

- b_j = j -edik regressziós együttható becslés értéke, (együttható: Eható)
- s_{b_j} = j -edik regressziós együttható becslés standard hibája, (St hiba)

Ezt a próbát parciális t-próbának, vagy röviden csak regressziós t-próbának hívjuk. A próbát külön-külön valamennyi regressziós becslés paraméterre el kell végezni, és ennek alapján képet kapunk arról, hogy az egyes változók lényeges mértékben járulnak-e hozzá az eredményváltozó magyarázatához, vagyis az eredményváltozó reziduális varianciájának csökkentéséhez.

Az első lépésben tehát minden változót bevonunk, és ha a p -értékek (szignifikancia-szint 5%) mindegyike 0,05-nél kisebb akkor a regressziós függvényt optimálisnak tekintjük. Ha találunk olyan paramétert, ahol a p érték nagyobb, mint 0,05, amit a piros szín is jelez, akkor dönthetünk a változó kihagyásáról, ha pedig több ilyen paraméter található, akkor célszerű először azt a változót kihagyni, amelyiknél a p értéke a legnagyobb. Ezt addig folytatjuk, amíg a p értékek mindegyike 0,05-nél kisebb lesz és a modell az elméleti feltételeknek is megfelel.

4.1.2 A Mátrix munkalap

A Mátrix munkalapon a többváltozós regressziószámítással kapcsolatos mátrixok, valamint az ezekhez kapcsolódó statisztikák találhatóak meg. A mátrixok maximális mérete a magyarázó változók maximális számával összhangban van. Az alábbi mátrixok jelennek meg a munkalapon:

A B oszloptól kezdődően rendre:

- teljes korrelációs mátrix (valamennyi változóra);
- bevont korrelációs mátrix (a meghagyott magyarázó változókra, ha valamennyi magyarázó változó szerepel a végleges modellben, akkor megegyezik az előző mátrix tartalmával);
- bevont korrelációs mátrix inverze;
- determinációs együtthatók a teljes adathalmazra;
- determinációs együtthatók a bevont adathalmazra;
- bevont változók parciális korrelációit tartalmazó háromszögmátrix;
- $X^T X$ mátrix a teljes adathalmazra;
- $X^T X$ mátrix a bevont változókra;
- $(X^T X)^{-1}$ a bevont változókra.

Az U oszloptól kezdődően rendre:

- teljes korrelációs mátrixhoz tartozó t-értékek, szignifikancia alapján színezve;
- bevont korrelációs mátrixhoz tartozó t-értékek, szignifikancia alapján színezve;
- bevont magyarázó változók korrelációs mátrixának inverze;
- bevont változók parciális korrelációihoz tartozó t-érték, szignifikancia alapján színezve.

Az AO oszloptól kezdődően rendre ¹⁵⁶:

- teljes adathalmazra a sajátértékek és sajátvektorok;
- bevont adathalmazra a sajátértékek és sajátvektorok;
- a sajátértékek megoszlási és kumulált megoszlási viszonyisményai;
- főkomponenssúly-mátrix;
- főkomponenssúlyok négyzete.

A fentiekben felsorolt mátrixok közül több önmagában is fontos információkat hordoz a regresszióval kapcsolatban, néhány kiszámítása pedig a további vizsgálatok miatt szükséges. Didaktikai okokból mindegyik mátrix bemutatását szükségesnek tartottuk.

Az alkalmazott képletek:

R(teljes) az összes változó bevonása esetén közli a korrelációs mátrixot és a változók neveit. Mellette megtalálható a korrelációs együtthatók felhasználásával számított t-értékek mátrixa, sárga mezőben a szignifikancia-szint, alapeset 5%, amit lehet változtatni.

Ha a kapcsolat létéről kívánunk dönteni, az eredményváltozó (y) és bármelyik magyarázó változó (x_j) között, akkor adott szignifikancia-szinten ellenőrizhetjük, vajon van-e kapcsolat a két ismerv között. Hipotézisrendszerünk:

$$H_0 : r_{yj} = 0$$

$$H_1 : r_{yj} \neq 0$$

Hasonló módon tesztelhetjük két magyarázóváltozó (x_j és x_l) közötti kapcsolatot.

$$H_0 : r_{jl} = 0$$

$$H_1 : r_{jl} \neq 0$$

$$x_j \neq x_l$$

A nullhipotézis értelmében a két ismerv független, ennek elvetése a korrelációs kapcsolat szignifikáns voltát igazolja.

A becsült korrelációs együtthatóra épülő próbafüggvényünk:

¹⁵⁶ A 4.1.4 pontban (a multikollinearitás kiküszöbölése) és az F2 pontban (Mátrix.xls parancsfájl működése) tárgyaljuk.

$$|t| = \frac{r_{yj} \sqrt{n-2}}{\sqrt{1-r_{yj}^2}}$$

Ahol:

- n = mintaelemszám
- k = magyarázó változók száma
- r_{yj} = az y és x_j közötti kétváltozós lineáris korrelációs együtthatót jelöli.

A nullhipotézis teljesülése esetén $(n-2)$ szabadságfokú kétoldali t -eloszlást követ. Ha a számított érték nagyobb mint a Student-féle t -eloszlás táblabeli értéke, akkor adott szignifikancia-szinten a korrelációs kapcsolat szignifikáns.

A kapcsolat nem szignifikáns, 5%-os szignifikancia-szinten, tehát a H_0 -hipotézist elfogadjuk, ha:

$$|t| = \frac{r_{yj} \sqrt{n-2}}{\sqrt{1-r_{yj}^2}} < t_{0,025(n-2)}$$

A kapcsolat szignifikáns, 5%-os szignifikancia-szinten, tehát a H_1 -alternatív hipotézist fogadjuk el, ha:

$$|t| = \frac{r_{yj} \sqrt{n-2}}{\sqrt{1-r_{yj}^2}} > t_{0,025(n-2)}$$

Az R (bevont) a bevont változók esetén közli a korrelációs mátrixot és a változók neveit. Mellette megtalálható a korrelációs együtthatók felhasználásával számított t -értékek mátrixa, sárga mezőben a szignifikancia-szint, alapeset 5%, amit lehet változtatni.

Parc. korr. A parciális korrelációs együtthatók mátrixa a bevont változók esetében, mellette a parciális korrelációs együtthatók felhasználásával számított t -értékek mátrixa, sárga mezőben a szignifikancia-szint, alapeset 5%, amit lehet változtatni.

A parciális korrelációs együtthatók nullától való különbözősége, ugyanúgy, mint a korrelációs együtthatók esetében, t -próbával tesztelhető, bár a próbafüggvény némiképp módosul:

$$H_0 : r_{yj.12\dots j-1, j+1, \dots k} = 0$$

$$H_1 : r_{yj.12\dots j-1, j+1, \dots k} \neq 0$$

$$t = \frac{r_{yj.12\dots j-1, j+1, \dots k} \sqrt{n-k-1}}{\sqrt{1-r_{yj.12\dots j-1, j+1, \dots k}^2}}$$

A nullhipotézis teljesülése esetén a próbafüggvény $(n-k-1)$ szabadságfokú kétoldali t -eloszlást követ.

A H_0 nullhipotézist elfogadjuk, ha:

$$t = \frac{r_{yj.12\dots j-1, j+1, \dots k} \sqrt{n-k-1}}{\sqrt{1-r_{yj.12\dots j-1, j+1, \dots k}^2}} < t_{0,025(n-k-1)}$$

A H_1 alternatív hipotézist fogadjuk el, ha:

$$t = \frac{r_{yj.12\dots j-1, j+1, \dots k} \sqrt{n-k-1}}{\sqrt{1-r_{yj.12\dots j-1, j+1, \dots k}^2}} > t_{0,025(n-k-1)}$$

Ha tehát a számított érték nagyobb, mint a Student-féle t -eloszlás táblabeli értéke, akkor adott szignifikancia-szinten a korrelációs kapcsolat szignifikáns.

$X^T X$ (telj) és $X^T X$ (bev) a Teljes és a Bevont esetekben az adatmátrix transzponáltja szorozva az adatmátrixszal. A mátrix transzponáltjának (az eredeti mátrixot elforgatjuk, az első oszlop lesz az első sor, a második oszlop lesz a második sor stb.

$$X^T X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & & X_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix} =$$

$$= \begin{bmatrix} n & \sum_{i=1}^n X_{1i} & \sum_{i=2}^n X_{2i} & \dots & \sum_{i=k}^n X_{ki} \\ \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{1i} X_{2i} & \dots & \sum_{i=1}^n X_{1i} X_{ki} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \sum_{i=1}^n X_{ki} & \sum_{i=1}^n X_{ki} X_{1i} & \sum_{i=1}^n X_{ki} X_{2i} & \dots & \sum_{i=1}^n X_{ki}^2 \end{bmatrix}$$

$(X^T X)^{-1}$ az előző mátrix inverze a bevont változók esetében.
 $(X^T X) * (X^T X)^{-1} = E$, ahol E az egységmátrix.

A fentiekben felsorolt mátrixok közül több önmagában is fontos információkat hordoz a regresszióval kapcsolatban, néhány kiszámítása pedig a további vizsgálatok miatt szükséges. Didaktikai okokból mind-egyik mátrixok bemutatását szükségesnek tartottuk.

4.1.3 A Maradék munkalap

A Maradék munkalapon az aktuális modell empirikus maradékaiból képzett oszlopvektorok találhatóak meg, valamint lehetőség van becslés, előrejelzés elvégzésére is. A munkalapon található oszlopvektorok az alábbiak:

- y : a vizsgált eredményváltozó értékeinek vektora;
- \hat{y} : az eredményváltozó értékeinek becsült vektora, a bevont magyarázóváltozókkal történt pontbecslés;
- \hat{y}^2 : az eredményváltozó becsült értékeinek négyzete;
- e : empirikus reziduum ($e = y - \hat{y}$);
- e_{t-p} : az empirikus maradék p -vel ($p=1,2,\dots,12$) késleltetett értéke (p nagyságát az Autokorreláció munkalapon lehet megadni, jellemzően $p = 1$);
- e^2 : a maradék négyzete.

Előrejelzést (idősorok esetén), illetve pontbecslést (keresztmetszeti adatbázisok esetén) készíthetünk a H oszloptól kezdődően, a sárga mezőkbe a magyarázóváltozók kívánt értékeit kell beírni. Technikai okokból valamennyi (bevont és be nem vont) változóhoz meg kell adni értékeket, ezekből csak azokat fogja a program figyelembe venni, amelyek bevont változókhoz tartoznak. Egyszerre maximum 20 becslés, illetve előrejelzés hajtható végre. A helyesen kitöltött magyarázó változó értékekhez tartozó becsült eredményváltozó érték a H oszlopban olvasható le.

4.1.4 A Multikollinearitás munkalap

A Multikollinearitás munkalap a magyarázó változók összefüggésének problémáját vizsgálja. A multikollinearitás témaköre hatalmas irodalommal rendelkezik, a téma egyik legfrissebb és legátfogóbb

magyar nyelvű összefoglalóját adja Kovács Péter¹⁵⁷. Az elvégezhető tesztek közül nem használtuk valamennyit, csupán az oktatásban gyakran alkalmazott, általánosan elterjedt próbákat. Az alábbiakban nagyvonalakban bemutatjuk a beépített próbákat:

- a multikollinearitás globális tesztelése:
 - χ^2 – próba;
 - Kondícióindex és kondíciószám (gyökös formula);
 - Petres-féle RED mutató¹⁵⁸;
- a multikollinearitás lokalizálása:
 - parciális korrelációs együtthatók tesztelése;
 - F-próba;
 - VIF-mutató (variancia infláló faktor);
 - tolerancia mutató;
- a multikollinearitás kiküszöbölése. Főkomponens regresszió¹⁵⁹.

Az alkalmazott képletek:

A multikollinearitás globális tesztelése, a χ^2 – próba^{160*}

A multikollinearitás jelenlétére következtethetünk a magyarázó változók korrelációs mátrixának determinánsából is. Igazolható ugyanis, hogy amennyiben a magyarázó változók lineárisan függetlenek egymástól a modellt ortogonálisnak (az ortogonális mátrix kvadratikus, transzponáltja egyenlő inverzével, a mátrix és a mátrix transzponáltjának a szorzata az egységmátrixot adja és determinánsa: ± 1) tekinthetjük. Az ortogonális rendszert leíró mátrix determinánsa 1-gyel, teljes multikollinearitás esetén viszont 0-val egyenlő. Minél közelebb van a determináns nullához, annál nagyobb mérvű függőség van a magyarázó változók között.

Érvényes tehát az alábbi reláció:

$$0 \leq |\mathbf{R}_k| \leq 1$$

ahol $|\mathbf{R}_k|$ a magyarázó változók korrelációs mátrixa determinánsának abszolút értéke.

A multikollinearitás szignifikanciája az R mátrix determinánsához kapcsolódva χ^2 -próbával tesztelhető. Az így képzett próbafüggvény annak a H_0 hipotézisnek a tesztelésére szolgál, amely szerint a magyarázó változók lineárisan függetlenek.

A χ^2 -teszt tehát azt vizsgálja, hogy a változók az alapsokaságban korrelálatlanok-e (nullhipotézis), azaz azt teszteli, hogy a korrelációs mátrixnak a főátlón kívüli elemei csak véletlenül térnek-e el a nullától.

A próbafüggvény az alábbi:

$$\chi^2 = - \left[n - 1 - \frac{1}{6}(2k + 5) \right] \log |\mathbf{R}_k|$$

A fenti függvényben a magyarázó változók korrelációs mátrixa determinánsának $|\mathbf{R}_k|$ tízes alapú logaritmusával számolunk. A próbafüggvény szabadságfoka:

$$\text{sz.f.} = \frac{1}{2} k(k-1)$$

ahol:

- n = a megfigyelések száma
- k = magyarázó változók száma

¹⁵⁷ Ld.: Kovács Péter [2008a].

¹⁵⁸ Ld. Kovács-Petres-Tóth [2004, 2005]

¹⁵⁹ A számítások elvégzéséhez szükség van a mátrix.xls parancsfájltra.

¹⁶⁰ Bartlett illetve Farrar–Glauber-teszt. A χ^2 próbafüggvény kidolgozása M. S. Bartlett nevéhez fűződik. (Bartlett, M. S. [1937]: Properties of sufficiency and statistical tests. Proceedings of the Royal Statistical Society Series A 160, 268–282.) A multikollinearitás χ^2 -próban alapuló tesztjét Farrar és Glauber publikálta először. Farrar D. E. és Glauber R. R. [1967].

Adott szignifikancia-szint mellett a χ^2 táblában megkeressük a megfelelő kritikus értéket. Amennyiben a fenti próbafüggvény alapján számított érték nagyobb, mint a táblából vett, számottevőnek tarthatjuk a modellben a multikollinearitást. Ellenkező esetben, H_0 hipotézis elfogadása esetén elfogadjuk a nullhipotézist, vagyis azt, hogy a magyarázó változók lineárisan függetlenek. A program közli, hogy van, vagy nincs multikollinearitás.

Kondícióindex és kondíciószám számítása^{161 162*}

A multikollinearitás mérőszámának egy családját alkotják a tényezőváltozók korrelációs mátrixának sajátértékeire épülő mutatók. E mutatók hátránya, hogy értelmezésük szubjektív, azaz nincs egy olyan egyértelmű küszöbszám, ami már erős multikollinearitást jelez.¹⁶³ A sajátértékek meghatározásánál felhasználjuk azt az összefüggést, hogy a standardizált magyarázó változók $X'X$ mátrixa egyenlő a magyarázó változók korrelációs mátrixával. A kondíciószám (vagy állapot mutató, CN=condition number) annak mutatója, hogy milyen közel van a magyarázó változók mátrixa ($X'X$), ahhoz, hogy szinguláris legyen. A becslésre vonatkozóan, ha a becslő értékek mátrixa közel szinguláris, tehát az adatok közel kollineárisak, akkor nehéz pontos vagy precíz inverzt előállítani, és a lineáris regresszió becsült paramétereinek nagy standard hibája lesz.

A kondícióindexek¹⁶⁴ (CI – condition index) a magyarázó változók korrelációs mátrixának legnagyobb (λ_{\max}) és j-dik (λ_j) $j=1,2,\dots,k$ sajátértékei alapján határozhatóak meg:

$$CI = \sqrt{\frac{\lambda_{\max}}{\lambda_j}}$$

Ha a legkisebb sajátértéket λ_{\min} -nel jelöljük, akkor a kondíciószám (CN – condition number):

$$CN = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

azaz a legnagyobb kondícióindex neve kondíciószám.

Ha a magyarázó változók lineárisan függetlenek, valamennyi sajátérték egy, akkor a CN-mutató értéke is egyel egyenlő. Minél nagyobb a mutató nagysága, annál erősebb a multikollinearitás mértéke. A multikollinearitás mértéke, gyenge, ha a $1 < CN < 5$, zavaró, ha $5 < CN < 10$, igen zavaró, ha $CN > 10$.

A Petres-féle Red-mutató. Az adatállomány redundanciájának a mérése*

Ha a magyarázó változók között szoros kapcsolat van, akkor a nagymennyiségű adatokat tartalmazó állományok gyakran kevés információt hordoznak, számos felesleges adatot tartalmaznak, tehát redundánsak. A multikollinearitás a lineáris regressziós modellek esetén a redundancia egyik fajtájaként értelmezhető.

Mérésére a Petres-féle RED-mutatót használjuk^{165 166 167}:

$$RED(\%) = \frac{\sigma_\lambda}{\sqrt{k-1}} \times 100$$

ahol σ_λ = a magyarázó változók korrelációs mátrixa (\mathbf{R}) sajátértékeinek (λ) a szórása és k a magyarázó változók száma.

Ha minden sajátérték egy, akkor $RED(\%)=0\%$. Ez azt jelenti, hogy a saját értékek szorzata, vagyis a magyarázó változók korrelációs mátrixának a determinánsa 1-gyel egyenlő. Ebben az esetben a mátrix or-

¹⁶¹ Encyclopedia of Statistical Sciences, 16 Volume Set, [2006] 2nd Edition. 2. kötet: 1239-1240

¹⁶² A SAS programja is így számol: <http://www.ats.ucla.edu/stat/sas/webbooks/reg/chapter2/sasreg2.htm> (2010 01 14)

¹⁶³ Kovács Péter [2008]: 49.

¹⁶⁴ Encyclopedia of Statistical Sciences, 16 Volume Set, [2006] 2nd Edition. 2. kötet: 1239-1240.

¹⁶⁵ Kovács Péter – Petres Tibor – Tóth László [2004]: 598.

¹⁶⁶ Peter Kovacs – Tibor Petres – Laszlo Toth [2005]: 405-412.

¹⁶⁷ Kovács Péter [2008]: 57-58.

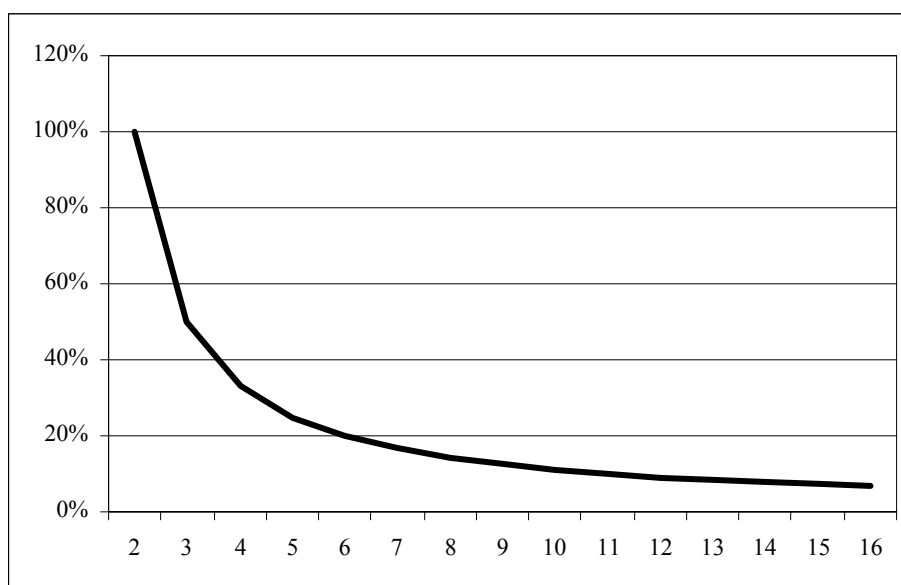
togonális, tehát nincs multikollinearitás, mert a magyarázó változók függetlenek egymástól. Amennyiben a sajátértékek távolodnak ettől az esettől, akkor a RED-mutató értéke növekszik.

A redundancia hiánya esetén a RED-mutató értéke nulla százalék, míg maximális redundancia esetén száz százalék. Ha pl. $RED(\%)=30\%$, akkor ez azt jelenti, hogy az adott méretű és minimális redundanciájú adatállományhoz képest a hasznos tartalmat hordozó adatok aránya 70%, azaz az adatok átlagos együttmozgásának a maximálishoz viszonyított mértéke pedig 30%.

A RED-mutató kifejezhető a magyarázó változók korrelációs mátrixának főátlón kívüli elemeinek négyzetes átlagaként is. Ez azt jelenti, hogy a mutató nem csak a becslőfüggvény szempontjából hasznos tartalmat hordozó adatok arányát mutatja, hanem a magyarázó változók együttmozgásának átlagos mértékét is.

A redundancia kritikus értéke, ha $k=1,2,\dots,16$ és a sajátértékek száma 1, akkor a kritikus értékek:

$$RED_{k,1} = \sqrt{\frac{1}{(k-1)(k-1)}} * 100$$



4.2. ábra: A RED-mutató kritikus értékei, mint 1 darab sajátérték zéró voltának szükséges feltétele. A k =a magyarázó változók száma ($k=2-16$)

Ha:

$RED(\%) < RED_{k,1}(\%)$ (zöld szám)

Akkor a lineáris regressziós modell illesztése után kapott becslt paraméterek szórásnégyzeteinek az összege illetve átlaga biztosan véges. (zöld szám)

Ha:

$RED(\%) > RED_{k,1}(\%)$ (piros szám)

Akkor a lineáris regressziós modell illesztése után kapott becslt paraméterek szórásnégyzeteinek az összege illetve átlaga nem biztos, hogy véges. Ezért a határpont kritikus értéként is értelmezhető. (piros szám)

A multikollinearitás lokalizálása

A korrelációs együtthatók vizsgálata

A magyarázó változók (x_j és x_i) közötti kapcsolat is tesztelhető, az a kedvező, ha nincs szignifikáns kapcsolat a két változó között, (ld. alább az első egyenlőtlenséget, zöld szám) de ha szignifikáns a kapcsolat, (ld. alább a második egyenlőtlenséget, piros szám) még ez még nem jelenti feltétlenül azt, hogy káros a multikollinearitás mértéke.

$$|t| = \frac{r_{ji} \sqrt{n-2}}{\sqrt{1-r_{ji}^2}} < t_{0,025(n-2)}$$

$$|t| = \frac{r_{ji} \sqrt{n-2}}{\sqrt{1-r_{ji}^2}} > t_{0,025(n-2)}$$

A parciális korrelációs együtthatók vizsgálata

Szignifikáns kapcsolat esetén, normális eloszlást feltételezve, káros a multikollinearitás a két magyarázó (x_j és x_i) változó között ha a parciális korrelációs együtthatók között szignifikáns a kapcsolat:

$$t = \frac{r_{x_j x_i | y_1, 2, \dots, i-1, i+1, \dots, k} \sqrt{n-k-1}}{\sqrt{1-r_{x_j x_i | y_1, 2, \dots, i-1, i+1, \dots, k}^2}} > t_{0,025(n-k-1)}$$

Ha nem szignifikáns a kapcsolat, akkor a multikollinearitás mértéke nem káros a két magyarázó (x_j és x_i) változó között ha:

$$t = \frac{r_{x_j x_i | y_1, 2, \dots, i-1, i+1, \dots, k} \sqrt{n-k-1}}{\sqrt{1-r_{x_j x_i | y_1, 2, \dots, i-1, i+1, \dots, k}^2}} < t_{0,025(n-k-1)}$$

F-próba

A magyarázó változó közötti kapcsolatokról hasznos információt nyerhetünk, ha kiszámítjuk az egyes magyarázó változóknak a többi magyarázó változóra vonatkozó többszörös determinációs együtthatóit. Ez azt jelenti, hogy a k magyarázó változót tartalmazó regressziós modellben k újabb regresszió-függvényt és többszörös determinációs együtthatót kell meghatározni. Ebben az esetben mivel: $y=x_j$ ezért $x_j=1,2,\dots,k-1$, tehát a magyarázó változók száma (k) eggyel csökken.

Regresszió-függvények:

$$X_1 = f(X_2, \dots, X_i, \dots, X_k, \varepsilon)$$

$$X_2 = f(X_1, X_3, \dots, X_i, \dots, X_k, \varepsilon)$$

$$\vdots$$

$$X_j = f(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_k, \varepsilon)$$

$$\vdots$$

$$X_k = f(X_1, \dots, X_j, \dots, X_{k-1}, \varepsilon)$$

Az egyes felsorolt függvények többszörös determinációs együtthatóit (R^2_j) megbecsüljük, majd F-próbával, a varianciaanalízisnél leírtaknak megfelelően teszteljük. Ha az F-próba alapján szignifikáns a kapcsolat az adott magyarázó változó és a többi magyarázó változó közötti modellben, akkor a multikollinearitást lokalizáltuk.

Nullhipotézisünk az, hogy a magyarázó változók regressziós együtthatói mind 0-k, vagyis nincs multikollinearitás, az alternatív hipotézis szerint létezik legalább egy 0-tól eltérő együttható vagyis van multikollinearitás.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{j-1} = \beta_{j+1} = \dots = \beta_k = 0$$

$$H_1 : \exists \beta_j \neq 0$$

Az F-próba alapján szignifikáns a multikollinearitás, ha:

$$F = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)} > F_{(0,05[k-1, n-k])}$$

Az F-próba alapján nem szignifikáns a multikollinearitás, ha:

$$F = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)} < F_{(0,05[k-,n-k])}$$

A varianciát infláló faktor (variancianövelő tényező VIF_j mutató = Variance Inflator Factor)

A magyarázó változók közötti determinációs együtthatókra épül a *varianciát infláló faktor (VIF) mutatója*. A mutatószám – melyet általában kettőnél több magyarázó változó esetén használunk – azt méri, hogy a multikollinearitás jelenléte milyen mértékben növeli a becsült paraméterek standard hibáinak a négyzetét, vagyis a varianciáját. Ez a mutató tehát azt mutatja, hogy a j-edik változó becsült együtthatójának tényleges varianciája hányszorosa annak, ami a multikollinearitás teljes hiánya esetén lenne kapható. Kiszámítása:

A varianciát infláló faktor (VIF)

$$VIF_j = \frac{1}{1 - R_j^2}$$

ahol R_j^2 – a j-edik magyarázó változó és a többi magyarázó változó közötti többszörös determinációs együttható.

A VIF mutató, könnyen meghatározható a redukált (csak a magyarázó változók közötti kapcsolatot mutató) korrelációs mátrix inverzéből, ugyanis az inverz mátrix diagonális elemei a VIF_j mutatókat adják.¹⁶⁸

Ebből határoztuk meg az R_j^2 értékeket, ugyanis:

$$R_j^2 = \frac{VIF_j - 1}{VIF_j}$$

A mutató határai:

$$1 \leq VIF_j \leq \infty$$

A β_j paraméter standard hibanégyzetének, varianciájának becslése:

$$s_{b_j}^2 = \left(\frac{s}{\sqrt{\sum (x_j - \bar{x})^2}} \right)^2 \frac{1}{(1 - R_{x_j}^2)} = \frac{s^2}{\sum (x_j - \bar{x})^2} \frac{1}{(1 - R_{x_j}^2)}$$

Ahol:

$$s^2 = \frac{\sum (y - \hat{y})^2}{n - 2}$$

Amennyiben a j-edik magyarázó változó lineárisan független a többitől, a VIF mutató értéke 1, mivel az R_j^2 értéke 0, ha viszont az R_j^2 értéke 1, akkor nem lehet értelmezni a mutatót, mert a végtelenbe tart. Ha valamelyik változó VIF mutatója 1 és 2 között van, akkor gyenge, (zöld szám), ha 2-5 között van, akkor erős, zavaró, (kék szám) ha pedig 5 felett van, akkor nagyon erős, káros (piros szám) a multikollinearitás.

A β_j paraméter standard hibanégyzetének nagysága három tényezőtől függ, akkor lesz nagyobb, ha:

1. s^2 , tehát a reziduális variancia nagyobb,
2. $\sum (x_j - \bar{x})^2$ az x_j változó átlagtól való eltérés - négyzetösszege (az x_j változó szórásnégyzete) kisebb,
3. $R_{x_j}^2$ nagyobb.

Ha az s^2 értéke kisebb és az $\sum (x_j - \bar{x})^2$ nagyobb akkor, nem lesz probléma a magas standard hibákkal, még akkor sem, ha az $R_{x_j}^2$ nagyobb lenne.

A tolerancia mutató

A VIF mutató reciprokát tolerancia mutatónak (jele T_j) hívjuk, amelynek határai:

¹⁶⁸ Ld. F2.-t a mellékletben.

$$0 \leq T_j \leq 1$$

A tolerancia mutató minél közelebb van az 1-hez, annál kevésbé zavaró a multikollinearitás. Ha valamelyik változó T_j mutatója 0,5 és 1 között van, akkor gyenge, (zöld szám), ha 0,2-0,5 között van akkor erős, zavaró, (kék szám) ha pedig 0 és 0,2 között van, akkor nagyon erős, káros (piros szám) a multikollinearitás.

A multikollinearitás kiküszöbölése. Főkomponens regresszió^{169 170 171}

A főkomponens-elemzés (PCA: Principal Components Analysis) az adatok leegyszerűsítését teszi lehetővé, a kiindulási adatmátrix dimenziójának csökkentésével. A régi magyarázó változók lineáris kombinációjával új változókat állítunk elő a sajátérték probléma megoldásával. A főkomponenselemzés mögöttes gondolata az, hogy kisszámú háttérváltozó "underlying factor" segítségével a teljes mátrixot viszonylag jól (adott hibával) reprezentálni lehet. Az új mesterséges változók korrelálatlanok (ortogonálisak, tehát egymástól lineárisan függetlenek), és csökkenő sajátértékek (eigenvalue) sorrendjében szokás sorban rakni őket. Az eljárás az eredeti, egymással szorosan korreláló k számú változót azok ugyancsak k számú főkomponensével helyettesíti, és ezek segítségével készít immáron jó tulajdonságú becsléseket. Az új regresszió az így képzett új változókra vonatkozik, így a szokásos becslési kritériumok nagy része (torzítatlanság, konzisztencia) nem értelmezhetők.

Alkalmas lehet becslésre, a főkomponens regresszió, ha:

- a kevés számú főkomponens minimális információ veszteséssel képes helyettesíteni a változókat,
- a mesterséges változók szakmailag jól értelmezhető tartalmúak,
- elsősorban nem a regressziós paramétervektorra, hanem az y becslésére vagyunk kíváncsiak

A főkomponens regresszió paramétereinek meghatározása $\mathbf{X}^T \mathbf{X}$ mátrix saját értékeinek (λ) és saját vektorainak, (faktorsúlyainak, loadings: a_{ij}) a meghatározását jelenti.

Általában különböző mértékegységű változókból állítjuk elő a mesterséges változókat, ezért a mértékegységeket ki kell küszöbölni. Ezt a standardizálás műveletével lehet biztosítani:

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}}{\sigma} \quad i=1,2,\dots,n \quad j=1,2,\dots,k$$

Az eredeti regresszió $n \times k$ méretű \mathbf{X} változó mátrixot egy $k \times k$ méretű \mathbf{A} mátrixszal egy ugyancsak $n \times k$ méretű $\mathbf{Z} = \mathbf{X}\mathbf{A}$ mátrixszá transzformáljuk. E \mathbf{Z} mátrix oszlopvektorait főkomponenseknek vagy főkomponensvektoroknak nevezik. Az \mathbf{A} mátrix tehát az $\mathbf{X}^T \mathbf{X}$ mátrix sajátvektoraiból épül fel. Az \mathbf{A} mátrix elemeit az \tilde{x}_j standardizált változók variancia-kovariancia mátrixának saját vektorai adják. A standardizált változók variancia-kovariancia mátrixa az eredeti változók korrelációs mátrixával (\mathbf{R}) azonos, így ebből a mátrixból is meghatározhatjuk a saját értékeket és saját vektorokat. Az \mathbf{A} mátrix tehát becsülhető a korrelációs mátrixból számított saját értékekhez tartozó saját vektorokkal, és ezért a program ennek alapján végzi el a számításokat. Egy-egy saját érték azt mutatja, hogy a vizsgált főkomponens az \mathbf{X} mátrix varianciájának hány %-t határozza meg. A saját értékek összege a magyarázó változók számával (k) egyezik meg. Ennek alapján a saját értékekből megoszlási illetve kumulált megoszlási viszonyszámokat képezhetünk. Általában néhány főkomponens az \mathbf{X} mátrix varianciájának igen jelentős hányadát képviselheti, ezért eljárhatunk úgy is, hogy nem a multikollinearitást okozó \mathbf{X}_j magyarázó változót zárjuk ki a modellből, hanem az alacsony saját értékekkel rendelkező főkomponenseket. Arra nincs egyértelmű szabály, hogy hány új főkomponens változót célszerű a modellben tartani. Az egyik megközelítés az lehet, hogy akkor jelentős egy főkomponens, ha a sajátértéke nagyobb mint egy illetve ha nem nagyobb mint egy, de a figyelembe vett saját értékek az összes variancia legalább 80% -át megmagyarázzák.

A számításokhoz szükséges adatokat a Mátrix munkalapon közöljük.

A főkomponensek (saját vektorok $a_{ij}, i,j=1,2,\dots,16$) és a magyarázó változók közötti összefüggés¹⁷²:

¹⁶⁹ Mundruczó György [1981]: 71-73.

¹⁷⁰ Hunyadi László [2001]: 179-181.

¹⁷¹ Petres Tibor-Tóth László [2008]: 245-246.

$$\begin{aligned} z_1 &= a_{11}x_1 + a_{21}x_2 + \dots + a_{k1}x_k \\ z_2 &= a_{12}x_1 + a_{22}x_2 + \dots + a_{k2}x_k \\ &\vdots \\ z_k &= a_{1k}x_1 + a_{2k}x_2 + \dots + a_{kk}x_k \end{aligned}$$

Majd a megtartott j számú komponensre és az y eredményváltozókra regressziós összefüggést határozunk meg:

$$\hat{y} = b_1z_1 + b_2z_2 + \dots + b_jz_j$$

A becsült b_j paraméterek segítségével az eredeti változókra transzformálhatjuk vissza a modellt:

$$\begin{aligned} \hat{y} &= b_1(a_{11}x_1 + a_{21}x_2 + \dots + a_{k1}x_k) + \\ &+ b_2(a_{12}x_1 + a_{22}x_2 + \dots + a_{k2}x_k) + \dots + \\ &+ b_j(a_{1j}x_1 + a_{2j}x_2 + \dots + a_{kj}x_k) \end{aligned}$$

Átrendezve az egyenletet:

$$\begin{aligned} \hat{y} &= (b_1a_{11} + b_2a_{12} + \dots + b_ja_{1j})x_1 + \\ &+ (b_1a_{21} + b_2a_{22} + \dots + b_ja_{2j})x_2 + \dots + \\ &+ (b_1a_{k1} + b_2a_{k2} + \dots + b_ja_{kj})x_k \end{aligned}$$

A fent leírt főkomponens transzformációval felhasználhatók a főkomponensek következő előnyös tulajdonságai:

- a főkomponensek páronként ortogonális rendszert alkotnak,
- a főkomponensek varianciáinak összege megegyezik az eredeti változók varianciáinak összegével,

$$\lambda_1 + \lambda_2 + \dots + \lambda_k = \sigma^2(x_1) + \sigma^2(x_2) + \dots + \sigma^2(x_k)$$

- a főkomponensek csökkenő varianciáik szerint vannak sorba rendezve.

$$\sigma^2(\lambda_1) \geq \sigma^2(\lambda_2) \geq \dots \geq \sigma^2(\lambda_k)$$

A főkomponens súlyok (loading változók) számítása

A főkomponens súlyok, a sajátvektorok elemeinek és a megfelelő sajátértékek négyzetgyökének szorzatai:

$$d_{ij} = a_{ij}\sqrt{\lambda_j}$$

$i, j = 1, 2, \dots, k$

A főkomponenssúlyokat tartalmazó \mathbf{D} mátrix, az ún. főkomponenssúly-mátrix, dimenziója $k \times k$ és az alábbi tulajdonságokkal rendelkezik:

- A főkomponenssúlyok abszolút értékei 1-nél nem nagyobbak.
- Az oszloponkénti négyzetösszegük a saját érték (λ_j), soronkénti négyzetösszegük pedig egy.
- A főkomponenssúlyok megadják a vizsgált magyarázóváltozók és a főkomponensek közötti lineáris korrelációs együtthatót.

A kommunalitási mutatók számítása

A főkomponens súly négyzetek felhasználásával a kommunalitási mutatók számíthatók ki:

$$d_{ij}^2 = a_{ij}^2\lambda_j$$

A \mathbf{D} mátrix k -edik sora első w darab elemeinek négyzeteit kumuláljuk, akkor a k -edik magyarázóváltozó kommunalitásához jutunk.

¹⁷² A számítások a mátrix.xls mátrixok szorzása munkalapon elvégezhető. A saját értékeket és a saját vektorokat a bevont változókra a regresszio.xls parancsfájl a mátrix munkalapon közli.

$$h_k^w = \sum_{j=1}^w d_{ij}^2$$

$$1 \leq w \leq k$$

A kumulált főkomponenssúly-négyzetek azt fejezik ki, hogy az egyes főkomponenseknek milyen jelentősége, súlya van a magyarázóváltozók varianciájában.

A főkomponensek varianciái az adatok kovariancia mátrixának sajátértékei (λ_j), az együtthatói pedig a megfelelő sajátértékhez tartozó sajátvektor együtthatói.

4.1.5 Az Autokorreláció munkalap

Amennyiben egy regressziós modellben nem teljesül a modellekkel szemben megfogalmazható azon feltevés, amely szerint a hibatenyező értékei páronként nem korrelálnak egymással, akkor a modell autokorrelált. Az autokorreláció mértékét a reziduális autokorrelációs együtthatóval mérhetjük. A p-ed rendű (p időegységgel késleltetett, a parancsfájl esetében p=1,2,...,12) elméleti autokorrelációs együtthatót az egymástól p időegységnyi távolságra álló maradéktagok korrelációs együtthatójaként becsülhetjük. A gyakorlatban az első rendű autokorrelációs együtthatót szokták tesztelni. A fájlban közöltünk több késleltetésre vonatkozó adatokat is, amire például szezonaritást mutató adatsorok esetén lehet szükség.

Az általunk használt modell:

$$e_t = \hat{\rho}_p e_{t-p} + v_t$$

Ahol:

- $\hat{\rho}^{173}$ = a p-endrendű autokorrelációs együttható becsült értéke, ami az egymástól p távolságra lévő reziduális tagok közötti korrelációs együttható
- $v_t = 0$ várható értékű, konstans szórású változó
- $t = 1, 2, \dots, n$

Képletben:

$$\hat{\rho}_p = r_{e_t; e_{t-p}} = \frac{\sum_{t=p+1}^n (e_t - \bar{e}_t)(e_{t-p} - \bar{e}_{t-p})}{\sqrt{\sum_{t=p+1}^n (e_t - \bar{e}_t)^2} \sqrt{\sum_{t=p+1}^n (e_{t-p} - \bar{e}_{t-p})^2}} \approx \frac{\sum_{t=p+1}^n e_t e_{t-p}}{\sum_{t=p+1}^n e_{t-p}^2}$$

Az Excel parancsfájlban az eredeti, és nem a közelítő p-ed rendű reziduális autokorrelációs együtthatóval számoltunk. A p-ed rendű autokorrelációs együtthatót a Student-féle t-próba felhasználásával teszteltük. A hipotézisrendszer az alábbi:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

A tesztstatisztika számítási módja:

$$t = \frac{\hat{\rho} \sqrt{n-p-1}}{\sqrt{1-\hat{\rho}^2}} \sim t_{n-p-1}$$

azaz a próbafüggvény n-p-1 szabadságfokút t-eloszlást követ, ennek megfelelően kell a kritikus értékeket meghatározni.

Ha p=1, elsőrendű reziduális autokorrelációs együtthatóról beszélünk.

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2}$$

¹⁷³ A görög ró betűvel jelöljük.

A fenti autoregresszív modellben, amennyiben a ρ autokorrelációs együtthatónak az értéke eltér nullától a regressziós modell autokorrelált, míg ellenkező esetben a regressziós modellben szereplő reziduális változó megegyezik a tiszta véletlen hatással, tehát a modell jól specifikált és autokorrelálatlan. Az autokorreláció tényének eldöntése tulajdonképpen az alábbi hipotézis tesztelésének felel meg:

$$H_0 : \rho=0$$

$$H_1 : \rho \neq 0$$

Szignifikáns autokorreláció esetén (5%-os szignifikancia-szinten, a piros szám jelzi) az alábbi feltétel teljesül:

$$|t| = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} > t_{0,025(n-2)}$$

Ha nincs szignifikáns autokorreláció (5%-os szignifikancia-szinten, a zöld szám jelzi) akkor az alábbi feltétel teljesül:

$$|t| = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} < t_{0,025(n-2)}$$

A nullhipotézis elfogadása azt jelenti, hogy a reziduális változó véletlen jellegű, a szomszédos értékek egymástól függetlenek. A fenti kétoldali próbával természetesen tesztelhető ρ pozitív és negatív értékének szignifikanciája is. Az első esetben $H_1:\rho>0$, míg a második esetben a $H_1:\rho<0$ a megfelelő alternatív hipotézis.

Az Excel parancsfájl kiszámítja és értékeli a **Durbin-Watson d-próba** alapján az autokorrelációt.

Durbin-Watson teszt: az elsőrendű D-W-statisztika értékét számítja ki és értékeli (a döntési lehetőségek:

1. pozitív autokorreláció,
2. negatív autokorreláció
3. bizonytalanság, tehát nem lehet dönteni

A próbafüggvény előállításához első lépésben ki kell számítani a tapasztalati reziduumokat (e_t), amelyeket a klasszikus legkisebb négyzetek módszerének alkalmazásával nyerünk. A próbafüggvény:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

A d-próbafüggvény értelmezéséhez hasznos segítséget ad a $\hat{\rho}$ elsőrendű autokorrelációs együttható és a **d-mutató** között - eléggé nagyszámú megfigyelés ($n>30$) esetén - felírható alábbi közelítő összefüggés.

$$d \approx 2 - 2\hat{\rho} = 2(1 - \hat{\rho})$$

Illetve:

$$\hat{\rho} \approx 1 - \frac{d}{2}$$

A fenti összefüggésből nyilvánvaló, hogy amennyiben a modell autokorrelálatlan ($\hat{\rho}=0$), a kiszámított $d=2$. Pozitív autokorreláció erősödése esetén d értéke közelít nullához, míg erősödő negatív autokorreláció esetén a 4-hez tart. A d-próba elvégzéséhez a mintából egy klasszikus legkisebb négyzetek módszerével történt becslés segítségével meghatározzuk d értékét, majd az egyenletben szereplő magyarázó változók és a megfigyelések számának, valamint a próba adott szignifikancia-szintjének megfelelően megkeressük a Durbin-Watson táblából a megfelelő értékeket¹⁷⁴. A táblázatban szereplő értékek közvetlenül a pozitív autokorreláció tesztelésére ($H_1:\rho>0$) alkalmasak, amelyeket d_L (alsó) és d_U (felső) értékekkel valósíthatunk meg. A negatív autokorreláció ($H_1:\rho<0$) tesztelése, illetve a kétoldali hipotézi-

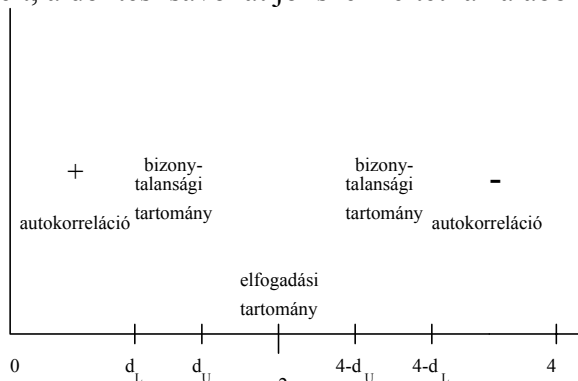
¹⁷⁴ Savin, N. E.-White K. J. [1977]: 1989-1996. alapján az 1 és 5 %-os szignifikancia szinten meghatározott kritikus értékekkel számoltunk. Ld.: továbbá Ramanathan Ramu [2003]: Durbin-Watson-próba, az 5 %-os d_L és d_U értékek egyoldali próbához. 614-617.

sek ellenőrzése újabb értékek kiszámítását igényli, amelyek d eloszlásának szimmetrikus jellege miatt nem okoznak különös gondot:

$$d'_L = 4 - d_U$$

$$d'_U = 4 - d_L$$

A próba lehetséges kimeneteleit, a döntési sávokat jól szemlélteti az alábbi ábra:



4-3. ábra: A Durbin-Watson d -próba döntési sávjai

A próba alkalmazásának viszonylagos hátránya az ún. bizonytalansági tartomány megléte, amely a gyakorlatban sok gondot okoz. Az irodalomban többféle módon igyekeztek a problémát feloldani, amelyek közül a legegyszerűbbnek tűnő megoldás az, amikor a bizonytalansági tartományt az elutasítási tartományhoz csatolják.

Amennyiben egy regressziós modellben nem teljesül a modellekkel szemben megfogalmazható azon feltétel, amely szerint a hibatényező értékei páronként nem korrelálnak egymással, tudjuk, hogy a modell autokorrelált¹⁷⁵.

Általánosságban elmondhatjuk, hogy az autokorreláció jelenléte mellett készített paraméter-, és pontbecslések ugyan torzítatlanok maradnak, de nem lesznek hatásosak. Különösen óvatosan kell kezelni az autokorrelált modellt, ha segítségével előrejelzéseket kívánunk készíteni. Autokorrelált modellek esetében az együtthatók standard hibái torzítottak, így sem a standard hibákhoz kapcsolódó próbák, sem az előrejelzésekhez kapcsolódó konfidencia intervallumok nem használhatók fel.

A program ábrázolja e_{t-p} függvényében az e_t alakulását. Az ábra alapján vizuálisan is következtethetünk az autokorreláció meglétére vagy hiányára.

4.1.6 A homoszkedaszticitás munkalap

Az idősorok esetében, mint azt az előzőekben bemutattuk, az autokorrelációt, a keresztmetszeti adatok esetében viszont a hibatényező varianciájának állandóságát szoktuk tesztelni. Ha konstans a hibatényező varianciájának várható értéke, akkor:

$$E(\varepsilon_i^2) = \sigma^2 \quad i = 1, 2, \dots, n.$$

Keresztmetszeti adatok esetén homoszkedaszticitás szempontjából is tesztelnünk kell a modelleket, hiszen elméleti feltétel, hogy a hibatényező varianciája állandó.¹⁷⁶

A nullhipotézis:

$$H_0 : \sigma_1^2 = \dots = \sigma_n^2.$$

Az alternatív hipotézis:

$$H_1 : \sigma_i^2 \neq \sigma_j^2 \quad (i \neq j),$$

ahol $i, j = 1, 2, \dots, n$.

A nullhipotézis azt fogalmazza meg, hogy a hibatényező szórásnégyzetei (varianciái) állandóak. A nullhipotézis teljesülése egyben azt is jelenti, hogy a modell homoszkedasztikus, míg az alternatív hipotézis a heteroszkedaszticitás feltételezését szimbolizálja. A heteroszkedaszticitás jelensége esetén a regresz-

¹⁷⁵ Kovács Ilona [1977]: 605.

¹⁷⁶ Pintér József [1991]: 18.

szíós együtthatók becslése torzítatlan, ugyanis továbbra is feltesszük, hogy a hibatenyező várható értéke nulla. Ugyanakkor a paraméterek varianciájára vonatkozó becslés nem lesz hatásos¹⁷⁷, a paraméterek standard hibái torzítottak lesznek, használatuk megkérdőjelezhető, a segítségükkel elvégzett próbák (pl. t- és F-próbák) és becslések félreinformálhatnak.

Globális (csoportos) Breusch-Pagan-Godfrey (BPG)- és Glejser-próba¹⁷⁸

A BPG-próba esetében a nullhipotézis megegyezik az előzőekben leírtakkal, az alternatív hipotézis pedig kissé általánosabb formában¹⁷⁹:

$$H_0 : \sigma_1^2 = \dots = \sigma_n^2$$

$$H_1 = E(\varepsilon_i^2) = \sigma^2 [h(\mathbf{Z}\alpha + v)], \text{ vagy:}$$

ahol:

- h : a reziduális változó függvénye, a függvény alakja lehet pl. lineáris, hatványkitevős, vagy exponenciális,
- \mathbf{Z} : a heteroszkedaszticitást magyarázó változók $n \times (k+1)$ típusú mátrixa,
- α : a véletlent becslő modell $(k+1) \times 1$ típusú paramétervektora,
- v : $n \times 1$ típusú, véletlen elemeket tartalmazó vektor.

A Glejser-próba esetén a teszt lényege¹⁸⁰ az, hogy lineáris regressziós kapcsolatot létesít a hibatenyező abszolút értéke és a heteroszkedaszticitást feltehetően előidéző magyarázó változók között.

$$|e_i| = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + v_i$$

A null- és - alternatív hipotézisek:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

$$H_1 : \exists \alpha_j \neq 0$$

Az F-próbával teszteljük a nullhipotézist, aminek elfogadása esetén a modell homoszkedasztikus, elutasítása esetén pedig heteroszkedasztikus.

A globális próbák az alábbiak:

- Glejser¹⁸¹-próba:

$$|e_i| = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + v_i$$

A regresszio.xls fájlban a pótlólagos regresszió többszörös determinációs együtthatóját $R^2(|e|; x)$ jelöléssel láttuk el.

- Breusch-Pagan-Godfrey (BPG)-próba:

$$e_i^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + v_i$$

A regresszio.xls fájlban a pótlólagos regresszió többszörös determinációs együtthatóját $R^2(e^2; x)$ jelöléssel láttuk el.

- Harvey¹⁸²-Godfrey¹⁸³-próba, ahol $\exp()$, az e^x exponenciális függvényt jelöli:

¹⁷⁷ Ez azt jelenti, hogy a klasszikus legkisebb négyzetek módszere (KLNМ, angolul Ordinary Least Squares: OLS) alkalmazása esetén a becslések ebben az esetben nem lesznek hatásosak, vagyis található egy másik torzítatlan lineáris becslés, aminek kisebb a varianciája, mint az KLNМ (OLS)-becslésnek. Ld.: Ramanathan Ramu [2003]: 365-366. és 397-398.

¹⁷⁸ Ld.: Glejser H. [1969]: 316-323. Godfrey, L. [1978]: 227-236. Breusch, T. S.; A. R. Pagan [1979]: 1287-1294. továbbá: Ramanathan R. [2003]: 367-369. Pintér József [1991]: 21-24. Gujarati Damodar N. [2003]: 411-412. Maddala G. S. [2004]: 244-246.

¹⁷⁹ Ld.: Pintér József [1991]: 21.

¹⁸⁰ Mundruczó György [1998]: 178.

¹⁸¹ Glejser H. [1969]: 316-323.

¹⁸² Harvey A. C. [1976]: 461-466.

¹⁸³ Godfrey, L. [1978]: 227-236.

$$\ln(e_i^2) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_p x_p + v_i$$

$$e_i^2 = \exp(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_p x_p + v_i)$$

- A **Park**¹⁸⁴-próba, ahol a függvény (h) hatványkitevős:

$$e_i^2 = \alpha_0 x_1^{\alpha_1} x_2^{\alpha_2} \dots x_p^{\alpha_p} e^{v_i}$$

$$\ln(e_i^2) = \ln \alpha_0 + \alpha_1 \ln x_1 + \alpha_2 \ln x_2 + \dots + \alpha_p \ln x_p + v_i$$

- Koenker-Bassett (KB)¹⁸⁵-próba:

$$e_i^2 = \alpha_0 + \alpha_1 \hat{y}_i^2 + v_i$$

A regresszio.xls fájlban a pótlólagos regresszió többszörös determinációs együtthatóját $R^2(e^2; \hat{y}^2)$ jelöléssel láttuk el.

A fenti képletekben:

- k = az eredeti regressziós függvényben a magyarázó változók száma;
- $i=1,2,\dots,n$ a megfigyelések száma;
- $|e_i|$ = az eredeti modell reziduális változójának abszolút értéke;
- e_i^2 = az eredeti modell reziduális változójának négyzete;
- \hat{y}_i^2 = az eredeti függvénnyel becsült eredményváltozó négyzete;
- a becsült paraméterek ($\alpha_j, j=0,1,2,\dots,k$) száma: $k+1$
- v_i : a pótlólagos regresszió reziduális változója.

A regresszió paramétereinek együttes szignifikanciája a globális F-próba segítségével mindegyik bemutatott teszt esetében vizsgálható. Ha a számított F-érték nagyobb, mint a táblabeli érték, akkor az alternatív hipotézist fogadjuk el, tehát a modell heteroszkedasztikus, ellenkező esetben homoszkedasztikus.

A heteroszkedaszticitás lokalizálása Glejser-, és Breusch-Pagan-Godfrey (BPG)-próbával.

A Glejser- és a BPG-próba lehetővé teszi a heteroszkedaszticitás lokalizálását. Amennyiben feltételezzük, hogy a magyarázó változók függvényei a reziduális változók abszolút értékei vagy a varianciái, akkor felírható magyarázó változónként egy-egy pótlólagos regressziós egyenlet.

A pótlólagos, j . magyarázó változóra vonatkozó regressziós egyenletek az alábbiak¹⁸⁶:

- Glejser-próba esetén:

$$|e_i| = \alpha_0 + \alpha_1 x_{ji} + v_i$$

- BPG-próba esetén:

$$e_i^2 = \alpha_0 + \alpha_1 x_{ji} + v_i$$

ahol x_{ji} a j -edik magyarázó változó i -edik értéke.

A regressziós együtthatót (meghatározó szerepe az α_1 együtthatónak van) a Student-féle t-próbával teszteljük, ha a számított érték nagyobb, mint a táblabeli érték, akkor az alternatív hipotézist fogadjuk el, tehát a modell heteroszkedasztikus, (piros szám jelzi) ellenkező esetben homoszkedasztikus (zöld szám jelzi).

4.2 Gyakorlati alkalmazások bemutatása idősoros és keresztmetszeti adatok alapján

¹⁸⁴ Park R. E. [1966]: 888.

¹⁸⁵ Gujarati Damodar N. [2003]

¹⁸⁶ A számításokat hatványkitevős (log-log) és exponenciális (log-lin) függvények esetében is elvégezhetjük, ha a változókat linearizáljuk. Park és Harvey-Godfrey-próba.

A regresszio.xls parancsfájl minden esetben közli az Autokorreláció és a Homoszkedaszticitás munkalapokon a számításokat. Az autokorreláció idősoros adatok esetén jelentkezik, ebben az esetben az adatok sorrendje kötött. A keresztmetszeti adatok sorrendje változtatható, ebben az esetben a homoszkedaszticitást szoktuk vizsgálni. Megjegyezzük, hogy keresztmetszeti adatoknál is előfordul, hogy a szomszédos hibatagok korrelálnak egymással, amit térbeli korrelációnak neveznek. Az autokorreláció vizsgálatánál az ökonometriai szakirodalomban¹⁸⁷ ettől eltekintenek és kizárólag az idősorok hibatagjainak vizsgálata tartozik e témakörbe. A maradékváltozó (reziduális változó) vizsgálatánál tehát lényeges kérdés, hogy idősoros vagy keresztmetszeti adatokkal dolgozunk-e. Idősoros adatbázis esetén az autokorrelációt, míg keresztmetszeti adatoknál a homoszkedaszticitást teszteljük. Ennek megfelelően két példát mutatunk be, mindkét példa valós magyarországi adatokat tartalmaz.

1. Idősoros példa.

A cementtermelés és a cementtermelést befolyásoló tényezők vizsgálata Magyarországon 1985 és 2008 között¹⁸⁸.

A regressziós modell változói:

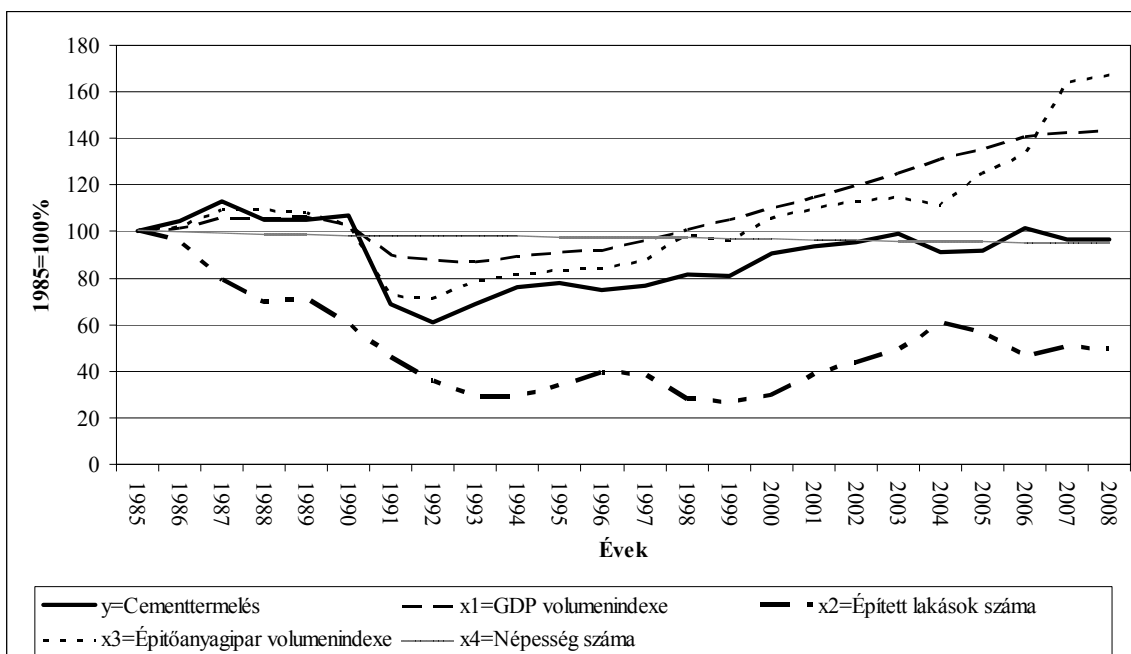
y	Cementtermelés (ezer tonna)
x ₁	GDP volumenindexe (1985=100)
x ₂	Épített lakások száma (darab)
x ₃	Építőanyagipar volumenindexe (1985=100)
x ₄	Népesség száma (ezer fő)

A rendszerváltás idején a hazai cement előállítás megközelítette a négymillió tonnát, ezt követően azonban drasztikusan visszaesett és 2000-ig közel egymillió tonnával alatta maradt a csúc sévek termelésének, majd 2001-től emelkedett ugyan a kibocsátás, de 2008-ban is közel félmillió tonnával maradt el az 1990-es szinthez képest.

A számítások megkezdése előtt célszerű az adatokat ábrázolni, hogy feltárjuk az adatok tendenciáit. A cementtermelés és a vizsgált magyarázó változók alakulását az alábbi ábra mutatja, az ábrakészítés során a vizsgált mutatók arányosságának biztosítása érdekében mindegyik mutatót (tehát a cementtermelést, az épített lakások számát és a népességszámot is) 1985-ös bázison számítottuk.

¹⁸⁷ Ld.: Maddala G. S. [2004]: 273-274. Ramanathan Ramu [2003]: 361-363. és 399-400. Gujarati Damodar N. [2003]: 401-403. és 441-443.

¹⁸⁸ Az adatok forrásai: Polt Rita [2005]: 996. Hunyadi László – Vita László [2008] II. kötet 204. Ipari és építőipari statisztikai évkönyv. KSH 1985-2005. Magyar statisztikai évkönyv. KSH. 1985-2008.



4.4. ábra: A cementtermelés és a cementtermelést befolyásoló tényezők alakulása Magyarországon 1985 és 2008 között.

Az ábrából látható, hogy a cementtermelés és a vizsgált magyarázó változók sok tekintetben hasonlóan mozognak, a mélypont a rendszerváltást követő években volt. A kivétel a népességszám alakulása. Magyarországon a vizsgált időszakban a népességszám folyamatosan csökkent, aminek mértéke 24 év alatt -5,2% volt. Eltérést mutat részben az épített lakások számának alakulása is, mert 1985 óta csökkenő tendenciát mutat, kivéve az 1995-1997 és 2000-2003 közötti időszakot.

Vizsgálhatjuk a ciklusok fordulópontjait is, az átlagos periódushossz¹⁸⁹ a cementtermelésnél 3 év, a GDP volumenindexénél 10 év, az épített lakások számánál 6 év, az építőanyagipar volumenindexénél 5 év. A népességszám esetében nem voltak fordulópontok.

A termelés elemzése és előrejelzése a regressziószámítás felhasználásával a cementipar esetében arra épült¹⁹⁰, hogy az építőanyagok, és ezen belül a cement termelése szorosan követi a GDP változását, valamint függhet az épített lakások számának, az építőanyagipar teljesítményének és a népesség számának alakulásától is. A népességszám változása és az épített lakások száma közötti kapcsolatot USA adatbázison először Kuznets modellezte, kidolgozva a róla elnevezett 15–25 éves építési ciklus elméletét¹⁹¹. Az építőanyagok és ezen belül a cement felhasználását az elmúlt években elsődlegesen az építési piac alakulása, s ezen belül az infrastruktúra- (autópályák) és a lakásépítés befolyásolta.

A számítások eredményei.

Varianciaanalízis

	df	SS	MS	F	p-érték
Regresszió	4	4845009,2	1211252,3	17,1	0,000004
Maradék	19	1347194,2	70905,0		
Összesen	23	6192203,3			

A varianciaanalízis tábla alapján a nullhipotézist elutasítjuk, tehát van legalább egy olyan magyarázó változó, amely szignifikáns hatással rendelkezik, létezik legalább egy nullától eltérő értékű regressziós paraméter.

¹⁸⁹ A ciklusfordulópontok számítása Excel parancsfájl felhasználásával.

¹⁹⁰ Polt Rita [2005]: 996-1000.

¹⁹¹ Kuznets, S. [1930].

Regressziós együtthatók						
	Eható	St hiba	t-érték	p-érték	Alsó 95%	Felső 95%
b0	-54913,67	22223,78	-2,47	0,0231	-101428,57	-8398,77
b1	49,15	20,68	2,38	0,0281	5,88	92,42
b2	-0,01	0,01	-0,89	0,3820	-0,04	0,02
b3	1,23	6,91	0,18	0,8606	-13,24	15,70
b4	5,16	2,04	2,54	0,0201	0,90	9,43

A regressziós paraméterek parciális tesztelése: a backward eliminációs módszer alkalmazása alapján először mind a négy magyarázó változót bevontuk a modellbe, majd az így meghatározott regressziófüggvényből szelektáltuk azokat a változókat, amelyek nem járulnak hozzá szignifikánsan a reziduális négyzetösszeg csökkenéséhez.¹⁹² A változók szelektálásához a p-értékeket használtuk. Ennek alapján először az $x_3 = \text{Építőanyagipar volumenindexe}$ változót, majd az $x_2 = \text{Épített lakások száma}$ magyarázó változót hagytuk ki a modellből. Meg kívánjuk jegyezni, hogy szakmailag indokolt lenne a két kihagyott változó modellben való szerepeltetése.

Regressziós együtthatók						
	Eható	St hiba	t-érték	p-érték	Alsó 95%	Felső 95%
b0	-37518,27	5884,69	-6,38	0,0000	-49756,15	-25280,39
b1	38,65	4,62	8,36	0,0000	29,03	48,27
b4	3,56	0,53	6,67	0,0000	2,45	4,67

A becslőfüggvény tehát:

$$\hat{y} = -37518,27 + 38,65x_1 + 3,56x_4$$

A multikollinearitás tesztjei:

A χ^2 globális próba alapján 5%-os szignifikancia szinten van multikollinearitás.

Khi-négyzet	8,18
Khi-szf	1
Khi krit (5%)	3,84
p-érték	0,0042

A parciális korrelációs együtthatók alapján számított t-statisztika értéke -11,66, a kritikus érték pedig 2,08, a két magyarázó változó között van multikollinearitás.

A p-értékek is a multikollinearitás létét igazolják. A VIF mutató értéke 2-5 között van, tehát erős, zavaró a multikollinearitás mértéke.

	y	R ²	F	p-érték	VIFj	Tj
R ² x	(x1)	0,584	30,86	0,0000	2,40	0,42
R ² x	(x4)	0,584	30,86	0,0000	2,40	0,42

Esetünkben a kondíciószám 2,734, azaz a mutató szerint gyenge multikollinearitást tapasztalunk a két magyarázó változó között.

A Petres-féle RED mutatót is számszerűsítettük:

Petres-féle RED		Kritikus érték (RED _{k,1})
Red(%)	76,4%	100,0%

Ha minden sajátérték egy, akkor Red(%) = 0%. Ez azt jelenti, hogy a sajátértékek szorzata, vagyis a magyarázó változók korrelációs mátrixának a determinánsa eggyel egyenlő. Ebben az esetben a mátrix ortogonális, nincs multikollinearitás, a magyarázó változók függetlenek egymástól. Amennyiben a sajátértékek távolodnak ettől az esettől, akkor a Red-mutató értéke növekszik. A maximális redundancia esetén a mutató értéke száz százalék.

Ha a számított érték a kritikusnál kisebb, akkor a lineáris regressziós modell illesztése után kapott becslt paraméterek szórásnégyzeteinek az összege illetve átlaga biztosan véges. Ellenkező esetben a lineáris regressziós modell illesztése után kapott becslt paraméterek szórásnégyzeteinek az összege illetve átlaga nem biztos, hogy véges, az adatállomány redundáns.

Esetünkben az adatállomány nem redundáns a Petres-féle RED mutató alapján.

¹⁹² Mundruczó György [1981]: 117-118.

Az autokorreláció tesztelése:

Az elsőrendű reziduális autokorrelációs együttható alapján nincs szignifikáns autokorreláció a modellben:

Autokorreláció rendje	ρ	t	t_{krit}	p-érték
1	0,342	1,707	2,074	0,1018

A népesség a vizsgált időszakban végig csökkent, a GDP volumenindexe pedig – a rendszerváltást követő éveket leszámítva – növekvő trendet mutat, ezért a két magyarázó változó együttes alkalmazása multikollinearitást okoz. Az optimális regressziós egyenes meghatározásához ezért más megoldást kereshetünk.

Szakmai indokok alapján építettünk új modellt, és azt kaptuk, hogy a modell globálisan és parciálisan is elfogadható, ha az x_2 és x_3 változókat vonjuk be a modellbe. Nyilvánvaló, hogy az épített lakások számának és az építőanyagipar volumenindexének változása (növekedése vagy csökkenése) a cementfelhasználást és így a termelést is jelentősen befolyásolja. Természetesen befolyásoló tényező a cement import volumene, de ennek vizsgálatától az adatok hiánya miatt eltekintettünk. Megállapítható továbbá, hogy a multikollinearitás mértéke és az autokorreláció nem zavaró.

A varianciaanalízis F-próbájához tartozó p-érték ebben az esetben 0,000002, tehát a nullhipotézist elutasíthatjuk.

Regressziós együtthatók						
	Eható	St hiba	t-érték	p-érték	Alsó 95%	Felső 95%
b0	1476,00	285,95	5,16	0,0000	881,33	2070,67
b2	0,0204	0,00	4,86	0,0001	0,01	0,03
b3	10,2928	2,58	4,00	0,0007	4,94	15,65

$$\hat{y} = 1476 + 0,0204x_2 + 10,2928x_3$$

A multikollinearitás mértéke ebben a modellben nem zavaró, a próbák a következők:

A χ^2 globális próba alapján 5%-os szignifikancia szinten nincs multikollinearitás.

Khi-négyzet	0,49
Khi-szf	1
Khi krit (5%)	3,84
p-érték	0,4821

A parciális korrelációs együtthatók alapján számított t-statisztika -1,77. 5%-os szignifikancia szinten a kritikus érték 2,08, tehát a két magyarázó változó között nincs multikollinearitás.

Az F-próba és a p-értékek is a multikollinearitás hiányát igazolják. A VIF mutató értéke 1-2 között van, tehát nem zavaró a hatás.

	y	R^2	F	p-érték	VIFj	Tj
R^2_x	(x2)	0,052	1,20	0,2860	1,05	0,95
R^2_x	(x3)	0,052	1,20	0,2860	1,05	0,95

A kondíciós szám esetünkben 1,26, ami gyenge multikollinearitásra utal.

A Petres-féle RED mutató:

Petres-féle RED	Kritikus érték ($RED_{k,1}$)
Red(%)	22,7%
	100,0%

A modell nem redundáns. $Red(\%) = 22,7\%$, ami azt jelenti, hogy az adott méretű és minimális redundanciájú adatállományhoz képest a hasznos tartalmat hordozó adatok aránya 77,3%, azaz az adatok átlagos együttmozgásának a maximálishoz viszonyított mértéke 22,7%.

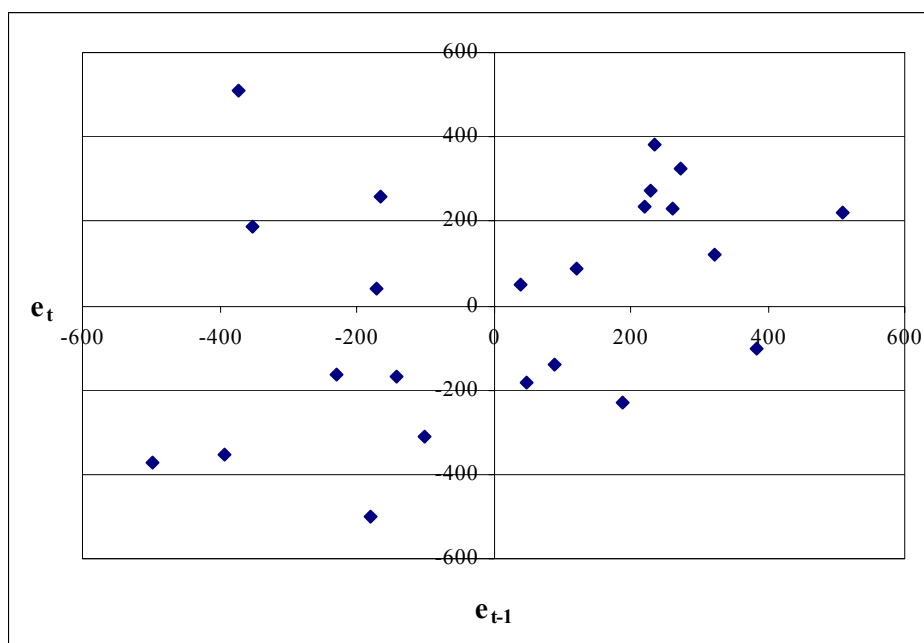
Az autokorreláció tesztelése:

Az elsőrendű reziduális autokorrelációs együttható alapján nincs szignifikáns autokorreláció a modellben:

Autokorreláció rendje	ρ	t	t_{krit}	p-érték
1	0,365	1,837	2,074	0,0797

Ezt mutatja a grafikus ábra is.

Reziduuumok ábrája.



A Durbin-Watson-féle teszt eredménye: 1,27 ami a bizonytalansági tartományba esik mindkét kérhető szignifikancia-szinten.

Durbin-Watson

D-W	1,271
dL (5%)	1,188
dU (5%)	1,546
dL' (5%)	2,454
dU' (5%)	2,812

Durbin-Watson

D-W	1,271
dL (1%)	0,959
dU (1%)	1,298
dL' (1%)	2,702
dU' (1%)	3,041

A kiválasztott modell az elméleti feltételeknek megfelel, elemzésre és előrejelzésre felhasználható.

2. Keresztmetszeti adatokon alapuló példa.

A keresztmetszeti adatok alapján történő regressziószámítást egy tapasztalati árindex modellen keresztül mutatjuk be.

Az ökonometriai modellek egyik speciális fajtája a tapasztalati (hedonikus) árindex modell¹⁹³ (hedonic price index model), amelyben egy árucikk ára a jellemzőitől függ, példa erre a gépkocsi ára és tulajdonságai közötti összefüggés. A vizsgálatba a 10 millió forintnál olcsóbb hazai forgalmazású autókat vontuk be. A gépkocsik árát nem csak saját, mérhető tulajdonságai befolyásolják, hanem minőségi tényezők is, mint például a márka, a biztonság stb.

A minta feladat:

119 autó adata, 2008-os árak, forrás: <http://www.auto2.hu/>.

y = a termék, az új autók alapárjai (ezer Ft).

x_j = a termék, az új autók tulajdonságai, az autók árát befolyásoló tényezők.

A megfigyelt magyarázó változók a következők:

4-2. tábla: Az új autók tulajdonságai

x_1	KÖBCM	hengerűrtartalom (cm ³)
x_2	TELJ	teljesítmény (LE)
x_3	NYOM	maximális nyomaték (Nm)
x_4	GYORS	0-ról 100 km/h-ra gyorsulás ideje (sec)

¹⁹³ Ramanathan Ramu [2003]: 23.

x ₅	VMAX	végsebesség (km/h)
x ₆	TÖMEG	saját tömeg (kg)
x ₇	MTÖMEG	megengedett össztömeg (kg)
x ₈	HOSSZ	hosszúság (mm)
x ₉	SZÉLES	szélesség (mm)
x ₁₀	MAGAS	magasság (mm)
x ₁₁	FOGYV	fogyasztás városban (liter/100 km)
x ₁₂	FOGYVK	fogyasztás városon kívül (liter/100 km)

A multikollinearitás kiküszöbölése. Főkomponens regresszió^{194 195 196 197}

Az autóárak és az autóárakat befolyásoló 12 magyarázó változó, az autók tulajdonságai közötti regressziós kapcsolat vizsgálata alapján az alábbi fontosabb megállapításokat tehetjük:

- A modell minden számított teszt alapján homoszkedasztikus.
- A modellben minden számított teszt alapján káros mértékű a multikollinearitás. Ennek oka az, hogy az autók tulajdonságai közül a teljesítmény erőteljesen befolyásolja a többi magyarázó változót, pl. a sebességet, a fogyasztást, a gyorsulást, a végsebességet, a tömeget, stb.
- A multikollinearitás miatt a regressziós paraméterek standard hibái nagyobbak (a VIF-mutató például 10 magyarázó változó esetében a kritikus értéknél nagyobb) és csak a b₀ és b₃ regressziós paraméter különbözik a t-próba alapján 5%-os szignifikancia szinten szignifikánsan 0-tól.
- Figyelembe véve, hogy mind a 12 magyarázó változónak a modellben való megtartása indokolt, célszerű a főkomponens-elemzést (PCA: Principal Components Analysis) elvégezni.

A regresszio.xls program közli a bevont változókra vonatkozó és a számításokhoz szükséges sajátértékeket és sajátvektorokat, továbbá a sajátértékek megoszlási és kumulált megoszlási viszonyait.

A számítások lépései:

1. A regressziószámítás elvégzése a 12 magyarázó változó bevonásával, konstans becslése nélkül. A regressziós paraméterek:

b ₁	b ₂	b ₃	b ₄	b ₅	b ₆	b ₇	b ₈	b ₉	b ₁₀	b ₁₁	b ₁₂
0,629	13,055	6,716	-10,012	-0,112	2,317	0,191	0,499	-2,280	-0,404	68,313	39,852

2. Az eredeti regresszió $n \times k$ méretű \mathbf{X} magyarázó változó mátrixot egy $k \times k$ méretű \mathbf{A} mátrixszal egy ugyancsak $n \times k$ méretű $\mathbf{Z} = \mathbf{XA}$ mátrixszá transzformáljuk. E \mathbf{Z} mátrix oszlopvektorait főkomponenseknek nevezik. Az \mathbf{A} mátrix becslhető a magyarázó változók korrelációs mátrixából számított sajátértékekhez tartozó sajátvektorokkal. A b₀-t elhagyva az \mathbf{X} magyarázó változók 119×12 típusú mátrixát kell szorozni, a magyarázó változók korrelációs mátrixából számított sajátvektorok 12×12 típusú \mathbf{A} mátrixával. A számításokat a mátrix.xls program mátrixok szorzata munkalap felhasználásával végezhetjük el. A szorzás eredményeképpen megkapjuk a 119×12 típusú \mathbf{Z} mátrixot.

3. A modellbe bevont főkomponensek kiválasztása. Az egyik megközelítés az lehet, hogy akkor jelentős egy főkomponens, ha a sajátértéke nagyobb, mint egy, illetve ha nem nagyobb egynél, de a figyelembe vett sajátértékek az összes variancia¹⁹⁸ legalább 80%-át megmagyarázzák. A számítások eredményekép-

¹⁹⁴ Mundruczó György [1981]: 71-73.

¹⁹⁵ Hunyadi László [2001]: 179-181.

¹⁹⁶ Sipos Béla [1982]:195-204.

¹⁹⁷ Petres Tibor-Tóth László [2008]: 245-246.

¹⁹⁸ A sajátértékek kumulált megoszlási viszonyzáma ebben az esetben 100%.

pen azt kaptuk, hogy az első három sajátérték nagyobb, mint egy és ez a három sajátérték kumulált megoszlási viszonyzáma alapján az összes variancia 87,84%-át megmagyarázza.

Az első három sajátérték¹⁹⁹: $\lambda_1 = 7,44$ $\lambda_2 = 1,75$ $\lambda_3 = 1,35$

A sajátvektorok (a_{ij} $i = 1, 2, \dots, 12$ $j = 1, 2, 3$):

z_1	z_2	z_3
0,348	-0,034	-0,008
0,348	-0,126	-0,030
0,280	0,273	-0,294
-0,267	0,310	0,262
0,327	-0,190	-0,254
0,326	0,260	0,084
0,308	0,332	0,093
0,322	0,127	-0,031
0,301	0,251	-0,026
0,005	0,452	0,579
0,238	-0,429	0,390
0,219	-0,362	0,524

4. A megtartott $j = 3$ számú főkomponensre és az y eredményváltozókra regressziós függvényt becsülünk:

$$\hat{y} = b_1 z_1 + b_2 z_2 + b_3 z_3$$

A számításokat a regresszio.xls programmal végezzük el, a regressziós paraméterek:

b_1	5,476
b_2	-5,215
b_3	-1,656

5. A becsült b_j paraméterek segítségével az eredeti változókra (4. egyenlet) transzformálhatjuk vissza a modellt:

A visszatranszformálás:

$$\hat{y} = (b_1 a_{1,1} + b_2 a_{1,2} + b_3 a_{1,3}) x_1 + (b_1 a_{2,1} + b_2 a_{2,2} + b_3 a_{2,3}) x_2 + \dots + (b_1 a_{12,1} + b_2 a_{12,2} + b_3 a_{12,3}) x_{12}$$

Az eredeti változókra transzformált regressziós paramétereket megkapjuk, ha a (3) mátrixot és (4) vektort összeszorozzuk:

Változók	Transzformált paraméterek
x_1	2,099
x_2	2,615
x_3	0,597
x_4	-3,513
x_5	3,204
x_6	0,290
x_7	-0,196
x_8	1,151
x_9	0,382
x_{10}	-3,290
x_{11}	2,894
x_{12}	2,218

A backward regresszióval az alábbi optimális függvényt kaptuk, ahol az autó ára két magyarázó változótól (teljesítmény és saját tömeg) függ. A becslőfüggvény:

$$\hat{y} = -2793,878 + 28,492x_2 + 3,877x_6$$

¹⁹⁹ A sajátértékek összege a magyarázó változók számával ($k=12$) egyenlő.

Összefoglalás

A regresszio.xls program felhasználása nagymértékben segíti a modellezést. Igen gyorsan ki lehet értékelni a különböző magyarázó változók kombinálása, illetve a vizsgált adatállomány változtatása (kiegészítése vagy csökkentése) esetén előálló regressziós modelleket, tehát azt, hogy az elméleti és szakmai feltételeknek melyik változat felel meg leginkább. Nem csak a modell globális és parciális tesztelésének az eredményét látjuk azonnal, hanem idősorok esetén az autokorreláció tesztjeit, keresztmetszeti adatoknál pedig a homoszkedaszticitás tesztjeit is értékelhetjük, valamint a reziduum ábrákat is elemezhetjük. Multikollinearitás esetén, ha értelemszerűen két vagy több magyarázó változó van a modellben, akkor több teszt elemzésére van lehetőség. A bevont változókat mint bemutattuk cserélhetjük a paraméterek sorában a Bevonjuk oszlopban a pipa jel beírásával vagy törlésével. A változók cseréjének hatására azonnal módosulnak a teszt eredmények és azokat gyorsan ki lehet értékelni. Különösen segíti a vizsgálatot, ha három vagy több magyarázó változóval rendelkezünk. Az összes modellvariánsok száma ismétlés nélküli kombinációk számával határozható meg, ahol k nagyságától függően a kombinációkat össze kell adnunk. Három magyarázó változó esetén a lehetséges modellvariánsok száma: 7, mert a k lehet, 1, 2 és 3. de négy változónál már 15. Az idősoros példa ismertetése arra is alkalmas volt, hogy nem lehet mechanikusan alkalmazni a backward regressziót, a szakmai ismereteket, feltételezéseket és a különböző pl. autokorreláció, multikollinearitás tesztjeinek az eredményeit is értékelni kell.

Végül megemlítjük, hogy a különböző speciális regressziós modellalkalmazásokra, figyelembe véve a téma szakirodalmát, számos Excel parancsfájlt dolgoztunk ki²⁰⁰, pl. Cochrane-Orcutt transzformáció, autokorreláció esetén, késleltetett regresszió esetén késleltetett mátrix elkészítése (12 féle lag-modell, pl. Koyck, Almon, Fisher és Alt módszerei esetén), aminek felhasználásával a becslés a regresszio.xls-sel már elvégezhető. Logisztikus regressziós függvények becslése két módszerrel, CES-függvények becslése három módszerrel, Cobb-Douglas termelési függvény paramétereinek, az átlag és határmutatóknak²⁰¹ a kiszámítása. Homoszkedaszticitás egyéb tesztjei: Goldfeld-Quandt-próba és a Szroeter-Harrison-King-féle próba.

4.3 Cochrane-Orcutt iterációs eljárás, a COtranszformáció.xls parancsfájl működése*

Szignifikáns autokorreláció esetén, az előzőek alapján a klasszikus legkisebb négyzetek hagyományos módszerével nyerhető előrebecslések félreinformálhatnak. Amennyiben az autokorreláció eredete a modellben a nem megmagyarázott részben található, a modell módosítása helyett egy iteratív paraméterbecslést célszerű elvégezni. Ez a módszer feltételezi az autokorrelációs együttható előzetes a priori ismeretét. Az autokorrelációs együttható segítségével egy egyszerű transzformációt hajtunk végre, amelynek eredményeként a "káros" hatás mértéke csökkenhet, illetve az megszűnhet. Az eljárást az alábbiakban foglalhatjuk össze:

A többváltozós regressziós modell egyenlete, idősoros adatok ($t=1,2..n$ és $j=1,2...k$) esetén:

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + \dots + b_k x_{kt} + e_t$$

A Cochrane-Orcutt^{202 203} (CORC) iterációs eljárás a regressziós modell átalakítását igényli oly módon, hogy az LNM-eljárás alkalmazható legyen. A fenti többváltozós regressziós egyenlet ($t-1$) időszakra történő átírásával kapjuk:

$$y_{t-1} = b_0 + b_1 x_{1(t-1)} + b_2 x_{2(t-1)} + \dots + b_k x_{k(t-1)} + e_{t-1}$$

A fenti egyenlet minden tagját beszorozva p -vel, majd kivonva az eredeti egyenletből kapjuk:

$$y_t - \hat{p}y_{t-1} = b_0 (1 - \hat{p}) + b_1 (x_{1t} - \hat{p}x_{1(t-1)}) + b_2 (x_{2t} - \hat{p}x_{2(t-1)}) + \dots + b_k (x_{kt} - \hat{p}x_{k(t-1)}) + v_t$$

$$y_t - \hat{p}y_{t-1} = c_0 + b_1 (x_{1t} - \hat{p}x_{1(t-1)}) + b_2 (x_{2t} - \hat{p}x_{2(t-1)}) + \dots + b_k (x_{kt} - \hat{p}x_{k(t-1)}) + v_t$$

ahol kihasználtuk, hogy az elsőrendű autokorrelációs együttható és a konstans tag:

$$e_t = \hat{p}e_{t-1} + v_t \text{ és } c_0 = b_0 (1 - \hat{p})$$

²⁰⁰ Megtalálható az MSC.zip-ben, leírásuk a kézikönyvben.

²⁰¹ Ld.: Kádás Kálmán [1944].

²⁰² Cochrane - Orcutt. [1949]: 32-61.

²⁰³ Ramu Ramanathan[2003]: 412-413.

Ezt az egyenletet átírhatjuk a következőképpen:

$$y_t^* = c_0 + b_1 x_{1t}^* + b_2 x_{2t}^* + \dots + b_k x_{kt}^* + v_t$$

ahol:

$$y_t^* = y_t - \hat{\rho} y_{t-1}, \quad x_{1t}^* = x_{1t} - \hat{\rho} x_{1(t-1)} \quad \text{és így tovább: } x_{kt}^* = x_{kt} - \hat{\rho} x_{k(t-1)}$$

$t=2^{204}, 3, \dots, n$ és $j=1, 2, \dots, k$.

A Cochrane-Orcutt – eljárás lépései:

1. lépés: Becsüljük LNM-sel az eredeti egyenletet és számítsuk ki az e_t véletlen (eltérés) - változókat.

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + \dots + b_k x_{kt} + e_t$$

2. lépés. Becsüljük az elsőrendű autokorrelációs együtthatót ($\hat{\rho}$) a fenti egyenletből, az alábbi már ismert módon:

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2}$$

3. lépés: Alakítsuk át a változókat a következőképpen:

$$y_t^* = y_t - \hat{\rho} y_{t-1}, \quad x_{1t}^* = x_{1t} - \hat{\rho} x_{1(t-1)} \quad \text{és így tovább: } x_{kt}^* = x_{kt} - \hat{\rho} x_{k(t-1)}$$

A megcsillagozott változók csak $t=2$ -től n -ig definiálhatók a $(t-1)$ -et tartalmazó tag (reziduum e_{t-1}) jelenléte miatt.

4. lépés. Magyarázzuk y_t^* -ot egy konstans [$c_0 = b_0(1-\hat{\rho})$] és $x_{1t}^*, x_{2t}^*, \dots, x_{kt}^*$, segítségével és számoljuk ki az átalakított egyenlet LNM-bebecsléseit.

5. lépés. Használjuk ezeket a becsléseket az eredeti egyenlet paramétereinek (b_j -ihez) becsléséhez és számítsuk ki az új e_t becsléseket.

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + \dots + b_k x_{kt} + e_t$$

Ezután térjünk vissza a 2. lépéshez, és az új értékekkel ismételjük meg az eljárást, amíg az alábbi leállási szabály életbe nem lép.

lépés.

6. Az iteratív eljárást akkor állíthatjuk le, ha teszt alapján autokorrelációt kiszűrtük, akkor elfogadjuk a modellt.²⁰⁵ Ha van autokorreláció akkor folytatjuk az eljárást addig, ameddig két egymást követő iteráció becsült $\hat{\rho}$ -ja közötti eltérés nem nagyobb egy előre megadott értéknél, pl. 0,001-nél. Az utolsó $\hat{\rho}$ -t használjuk az alábbi egyenletben a CORC-becslések kiszámításához.

$$y_t^* = c_0 + b_1 x_{1t}^* + b_2 x_{2t}^* + \dots + b_k x_{kt}^* + v_t$$

A konstans tag átalakítása az eredeti egyenletben:

$$c_0 = b_0(1-\hat{\rho})$$

$$b_0 = \frac{c_0}{(1-\hat{\rho})}$$

A fenti eljárást általánosított differenciák módszereként is ismerik, ami más megfogalmazásban – a klaszikus legkisebb négyzetek módszerénél általánosabb paraméterbecslési eljárásnak – az ún. az általánosított legkisebb négyzetek módszerének felel meg, és a becsléshez szükséges $\hat{\rho}$ előállítsa történik az iteratív módszerrel.

A COtranszformáció.xls parancsfájl működése:

Kéri az adatmátrixot és az előzőleg megbecsült $\hat{\rho}$ - értéket, amit a regresszió.xls fájlal megbecsülhetünk, ha az autokorreláció rendje: 1. Ezt követően az Adat munkalapba bemásolt adatokkal számolva a CO1 munkalapon kiszámítja a transzformált mátrixot. A transzformáció:

²⁰⁴ A reziduumok késleltetése (e_{t-1}) miatt a legrégebbi $t=1$ adat kiesik.

²⁰⁵ Mundruczó György [1981]: 133.

$$y_t - \hat{p}y_{t-1} = (x_{1t} - \hat{p}x_{1(t-1)})(x_{2t} - \hat{p}x_{2(t-1)}) \dots (x_{kt} - \hat{p}x_{k(t-1)})$$

Az iteráció indítása parancs a CO1 munkalapon kiszámított (transzformált) mátrixot az Adat munkalapra másolja és a CO1 munkalapon elvégzi a második CO transzformációt. Az idősor minden transzformáció után egy megfigyeléssel (a legrégebbi adattal) csökken. A transzformált adatsorral elvégezzük a regresszió számítását a regresszió.xls fájljal, és ha teszt alapján autokorrelációt kiszűrtük, akkor elfogadjuk a modellt. Ha van autokorreláció, akkor folytatjuk a számításokat. Az iterációkat akkor fejezzük be, amikor a \hat{p} paraméterek az egyik iterációról a másikra gyakorlatilag már nem változnak.

4.4 A Szroeter-Harrison-King-féle próba. (Szroeterteresz.xls parancsfáj működése) és a Goldfeld-Quandt-próba (Goldfeld-Quandt-próba.xls parancsfáj működése)*

4.4.1 A Szroeter-Harrison-King-féle próba

A heteroszkedaszticitás felismerésének a Szroeter-féle próba^{206 207} egy olyan eljárása, amely kötődik – többek között – az autokorrelációnak a d-statisztika segítségével történő teszteléséhez.

1. alternatív hipotézis.²⁰⁸

A Szroeter-féle próbánál – a szokásos lineáris regressziós modellnél – a H_0 nullhipotézist, azaz:

$$H_0 : \sigma_1^2 = \dots = \sigma_n^2$$

$$H_1 : \sigma_1^2 \leq \dots \leq \sigma_n^2$$

H_1 alternatív hipotézissel szemben ellenőrizzük, ahol legalább egy esetben teljesül az egyenlőség. Ebben az esetben egy növekvő szórású (varianciájú) alternatív hipotézist tesztelünk, de a hipotézis csökkenő sorozat esetére is felírható.²⁰⁹

A próbafüggvény az alábbi módon definiálható:

$$h = \frac{\sum_{i=1}^n h_i e_i^2}{\sum_{i=1}^n e_i^2}$$

ahol a h_i súlyszámokat az alábbi képlet segítségével határozhatjuk meg:

$$h_i = 2 \left[1 - \cos \left(i \times \pi / (n + 1) \right) \right], \quad i = 1, 2, \dots, n$$

vagy:

$$h_i = -2 \left[1 + \cos \left(i \times \pi / (n + 1) \right) \right], \quad i = 1, 2, \dots, n$$

A h_i értékekből álló sor főbb jellemzői:

$$\lim_{i \rightarrow \infty} h_i = 4 \quad \text{és} \quad \lim_{i \rightarrow 0} h_i = 0$$

2. alternatív hipotézis.²¹⁰

Ha egy csökkenő szórású (varianciájú) alternatív hipotézist tesztelünk, akkor a null-, illetve alternatív hipotézis:

$$H_0 : \sigma_1^2 = \dots = \sigma_n^2$$

$$H_1 : \sigma_1^2 \geq \dots \geq \sigma_n^2$$

A próbafüggvény az előző módon definiálható:

$$h = \frac{\sum_{i=1}^n h_i e_i^2}{\sum_{i=1}^n e_i^2}$$

²⁰⁶ Szroeter Jerzy [1978]: 1311-1327.

²⁰⁷ Pintér József [1991]: 16-36.

²⁰⁸ King Maxvel L. [1981]: 315-321.

²⁰⁹ Ebben az esetben a h_i kifejezésben a cos-függvény előtt + jelet írunk.

²¹⁰ King Maxvel L. [1981]: i. m. 316.

ahol a h_i súlyszámokat az alábbi képlet segítségével határozhatjuk meg és ahol a \cos -függvény előjele változik + lesz:

$$h_i = 2 \left[1 + \cos \left(i \times \pi / (n+1) \right) \right], \quad i = 1, 2, \dots, n$$

Választás az 1. és 2. alternatív hipotézis (teszt) közül.

Természetesen felmerül a kérdés, hogy a két alternatív hipotézis közül melyiket válasszuk. Ha ábrázoljuk a reziduális változók négyzeteit (e_i^2 $i=1,2,\dots,n$) az egyes magyarázóváltozók (x_j $j=1,2,\dots,k$) függvényében, akkor lehet következtetéseket levonni az ábrák alakulásából. Amennyiben a reziduális változók négyzetei szétnyíló (nő a reziduális változók négyzete a magyarázó változók függvényében, tehát valószínűsíthetjük az első alternatív hipotézist), vagy összezáródó (csökken a reziduális változók négyzete a magyarázó változók függvényében, tehát valószínűsíthetjük a második alternatív hipotézist) sávot mutatnak, úgy várhatóan a heteroszkedaszticitás ténye áll fent. Amennyiben az ábrán a sáv nem változik, úgy valószínűsíthetjük, hogy a nullhipotézis igaz, és homoszkedasztikus a modell.

Figyelembe véve az Excel által nyújtott lehetőségeket, mindkét alternatív hipotézissel célszerű a számításokat elvégezni. A H_0 -ról történő döntésnél (elfogadom vagy elvetem) mindig meg kell fogalmazni, hogy ez a döntés milyen alternatív hipotézissel szemben történt. Pl. ha az első esetben a nullhipotézist elfogadom, akkor lehetséges, hogy a reziduális szórásnégyzetek azonosak, de lehet az is, hogy csökkennek, ugyanis a meghatározott szignifikancia-szinten, ami általában 5%, csak azt az alternatív hipotézist utasítottuk el, hogy a reziduális szórásnégyzetek nőnek. Ha viszont a második esetben a nullhipotézist elvetjük, vagyis felételezzük, hogy heteroszkedasztikus a modell, akkor azt feltételezzük, hogy a reziduális szórásnégyzetek csökkennek. Ha mindkét alternatív hipotézist a Szroeter teszt elutasítja pl. 5%-os szignifikancia-szinten, akkor a homoszkedaszticitásra vonatkozó nullhipotézist elfogadjuk.

A Durbin-Watson táblázat felhasználása

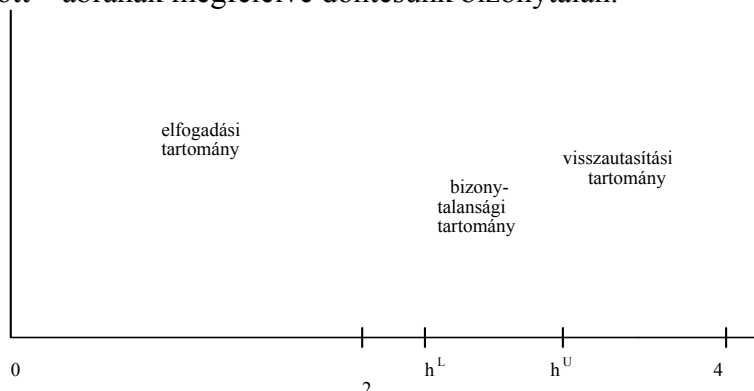
A Szroeter-próbánál, tehát előzően ábrázolni és vizsgálni kell azt, hogy a reziduális szórásnégyzetek növekvő vagy csökkenő sorozatot alkotnak-e. A próba kritikus értékeinek meghatározására többféle eljárás ismert, amelyek közül egyszerűsége miatt a Durbin-Watson táblázaton alapuló módszer érdemel első helyen említést. A két kritikus érték, amely egy bizonytalansági tartományt definiál az alábbi módon határozható meg:

$$h^L = 4 - d_{(n+1,k+1)}^U$$

$$h^U = 4 - d_{(n+1,k+1)}^L$$

ahol k a magyarázóváltozók száma.

Egyértelműen heteroszkedasztikusnak tekintjük a modellt, ha $h > h^U$, és homoszkedasztikusnak ha $h < h^L$. Amennyiben az empirikus adatok alapján kiszámított h érték a két kritikus érték közé esik, az alábbi – korábban már bemutatott – ábrának megfelelően döntésünk bizonytalan.



4-6. ábra: A heteroszkedaszticitás döntési sávjai D-W módszer alapján

A próba használatát rontja a bizonytalansági tartomány létezése. A kritikus értéket azonban a Béta-eloszlásból (β) illetve az F-eloszlás segítségével is fel lehet írni.

Becslés a Béta-eloszlás²¹¹ felhasználásával

A próbafüggvény az előzőekből ismert:

²¹¹ A Béta eloszlásról részletesebben a Wikipédiában található magyarázatok: <http://hu.wikipedia.org/wiki/B%C3%A9ta-eloszl%C3%A1s> (2009 szept. 3.)

$$h = \frac{\sum_{i=1}^n h_i e_i^2}{\sum_{i=1}^n e_i^2}$$

Ha nincs konstans a lineáris regressziós modellben ($b_0=0$), akkor a Béta-eloszlás alapján számított kritikus érték:

$$h^\alpha = 4(1-\beta^\alpha)$$

Ahol α a szignifikancia-szint, általában 5%

Ha $h > h^\alpha$, akkor a nullhipotézist elvetjük, vagyis heteroszkedasztikus a modell, ellenkező esetben ha: $h < h^\alpha$ homoszkedasztikus.

Ha van konstans a lineáris regressziós modellben²¹² ($b_0 \neq 0$), akkor a Béta-eloszlás alapján számított kritikus érték:

$$h^\alpha = 4\beta^\alpha$$

Ha $h < h^\alpha$, akkor a nullhipotézist elvetjük, vagyis heteroszkedasztikus a modell, ellenkező esetben ha: $h > h^\alpha$ homoszkedasztikus.

A következőkben közöljük a Harrison M. J. által kidolgozott Béta-eloszlás kritikus értékeit a Szroeter teszthez, 5%-os szignifikancia-szinten, ha a lineáris regressziós egyenes konstans is tartalmaz.²¹³ A Béta-eloszlás kritikus értékeit, 5 %-os szignifikancia szinten a táblázat (szroetereszt.xls parancsfáj munkafüzetében Béta 5%) a mintaelemszám ($n=8-100$) és a regressziós paraméterek függvényében közli ($m=k+1=2, 3, 4, 5, 6$).

Becslés az F-eloszlás felhasználásával

A próbafüggvény az előzőekből ismert:

$$h = \frac{\sum_{i=1}^n h_i e_i^2}{\sum_{i=1}^n e_i^2}$$

A hipotézis teszteléséhez, tehát a Béta-eloszlás használható, azonban a kritikus értékeket az F-eloszlás segítségével is fel lehet írni. A kritikus érték meghatározásának lépései:

1. Az F-eloszlás szabadságfoka (r):

$$r = \frac{3(n-k-1)(n-k+2)}{2(n-k-1)} - 1$$

2. A Béta-eloszlás kritikus értékét közvetett módon az F-eloszlásból származtatjuk:

$$\beta^\alpha = \frac{1}{1 + F_{(r,r)}^\alpha}$$

Ahol:

- β^α = a Béta-eloszlás kritikus értéke α szignifikancia-szint mellett
- $F_{(r,r)}^\alpha$ = az F-eloszlás kritikus értéke α szignifikancia-szint mellett, ahol a számláló és a nevező szabadságfoka egyaránt r .

3. Meg kell határozni a Béta-eloszlás kritikus értéke ismeretében a $h^{*\alpha}$ értéket²¹⁴, vagyis a próbafüggvényhez rendelhető kritikus értéket. Ezt az előzőekben leírtak szerint két módon tehetjük meg²¹⁵:

²¹² Ez az általánosabb és gyakoribb eset.

²¹³ M. J. Harrison [1982]: 165.

²¹⁴ A * azt jelöli, hogy az F-eloszlás felhasználásával becsültük a Béta-eloszlás kritikus értékeit, tehát nem az eredeti Béta-eloszlást használtuk fel.

²¹⁵ M. J. Harrison [1982]: i. m. 161.

3.1 Ha nincs konstans a lineáris regressziós modellben ($b_0=0$), akkor a Béta-eloszlás alapján számított kritikus érték:

$$h^{*\alpha} = 4(1 - \beta^\alpha)$$

Vagyis az előzőek alapján:

$$h^{*\alpha} = 4 \left(1 - \frac{1}{1 + F_{(r,r)}^\alpha} \right)$$

Ha $h > h^{*\alpha}$, akkor a nullhipotézist elvetjük, vagyis heteroszkedasztikus a modell, ellenkező esetben ha: $h < h^{*\alpha}$ homoszkedasztikus.

3.2 Ha van konstans a lineáris regressziós modellben²¹⁶ ($b_0 \neq 0$), akkor a Béta – eloszlás alapján számított kritikus érték:

$$h^{*\alpha} = 4\beta^\alpha$$

Vagyis az előzőek alapján:

$$h^{*\alpha} = 4 \left(\frac{1}{1 + F_{(r,r)}^\alpha} \right)$$

Ha $h < h^{*\alpha}$, akkor a nullhipotézist elvetjük, vagyis heteroszkedasztikus a modell, ellenkező esetben ha: $h > h^{*\alpha}$ homoszkedasztikus.

Először a regresszós számítást kell elvégezni, eszközök-adatelemzés-regresszió, kéri az Y vektort, utána az X mátrixot, a maradékokat be kell jelölni, majd a számítás eredményeképpen kapott maradékokat másoljuk az 1 munkalapon levő maradék oszlopba (szroeter szórás nő), először töröljük a sárga jelzésű oszlopot, utána másoljuk a reziduumot. Változtatni a sárgamezős cellákban lehet, a többi esetben a program elvégzi a számításokat.

A tesztelést (Szroeter-Harrison-King-féle próba) az előzőekben leírtak szerint végzi el és értékeli szövegesen is az eredményeket.

A Szórásnégyzetábra-Glejser-adat munkalapon közli a reziduumok (maradékok) abszolút értékeit és négyzeteit. A globális és lokális Glejser-Park próbákat az Adatelemzés - regresszós számítás felhasználásával el lehet végezni, mivel megvan az eredményváltozó (reziduum négyzete illetve abszolút értéke) és a magyarázóváltozók értékei is ismertek. Elvégezve a számításokat a globális Park próba esetében a variancia-analízis eredményei (F-próba) alapján dönthetünk arról, hogy van-e heteroszkedaszticitás vagy homoszkedasztikus-e a modell. Ha szignifikáns a kapcsolat, akkor heteroszkedasztikus a modell, ellenkező esetben homoszkedasztikus. A Glejser-Park lokális próbák alkalmazása esetében az adatállományt (magyarázóváltozók) elő kell készíteni. Pl. logaritmizálás, négyzetgyökvonás stb. az ismertetett képletek szerint. A tesztelést F-próbával, vagy t-próbával végezzük, az eredmények értelmezése hasonló, mint a globális próba esetében.

Elfogadjuk tehát a nullhipotézist, és nem létezőnek tekintjük a modellt, ha az F-statisztika számított értéke kisebb, mint egy adott *szignifikancia-szinthez* tartozó F-eloszlás táblabeli értéke. Elvetjük a nullhipotézist, vagyis a modellt létezőnek, relevánsnak tekintjük, ha a próbafüggvény értéke meghaladja a táblabeli értéket. Homoszkedasztikus a modell, ha elfogadjuk azt a nullhipotézist, hogy a paraméterek között van legalább egy nulla, vagy ha egy paraméter van, akkor az nulla. Végül is az a kedvező, ha a modellünk nem „jó”, hiszen akkor homoszkedasztikus a regressziós modell, ha a reziduum abszolút értéke illetve négyzete (vagy logaritmusai Park-próba) és a magyarázóváltozók (a magyarázóváltozók különböző csoportjai és transzformált értékei) között nincs szignifikáns kapcsolat.

4.4.2 A Goldfeld-Quandt-próba

A Goldfeld-Quandt-próba²¹⁷ azon alapszik, hogyha a reziduális változók varianciája (e_i^2) azonos a különböző megfigyelések (x_i , $i=1,2,\dots,n$) esetén, tehát ha homoszkedasztikusak, akkor ez a minta egyes részeire

²¹⁶ Ez az általánosabb és gyakoribb eset.

²¹⁷ Goldfeld S. M. – Quandt R. E. [1965]: 539-547. Id.: továbbá: Mundruczó György [1981]: 139-140. Ramanathan R. [2003]: 371-372. Pintér József [1991]: 20-21. Gujarati Damodar N. [2003]: 408-409. Maddala G. S. [2004]: 247.

is igaz. Ezért tesztelhető a reziduális változók varianciáinak egyenlősége F-próba segítségével. A tesztstatisztika a két becült variancia hányadosa. A megfigyelt értékeket (n), három részre osztjuk és a középső megfigyeléseket elhagyjuk. Ezután a két szélső adathalmazra külön-külön elvégezzük a regressziós becslést és meghatározzuk a reziduumok varianciáit. F-próbával most már tesztelhetjük azt, hogy ezek a varianciák egyenlők-e (homoszkedasztikus a modell) vagy nem (heteroszkedasztikus a modell). Feltételezzük továbbá, azt, hogy a magyarázóváltozó értékének növekedésével a reziduumok varianciái nem változnak vagy növekednek. A teszt keresztmetszeti adatok esetén alkalmazható.

A Goldfeld-Quandt-próba Excel parancsfájlt kétváltozós modellre dolgoztuk ki.

A hipotézisrendszer:

$$y = b_0 + b_1 x_j + e$$

$$H_0: E(e_i^2) = \sigma^2$$

$$H_1: E(e_i^2) = \sigma^2 x_j^2$$

A nullhipotézis azt fogalmazza meg, hogy a hibatenyező szórásnégyzete (varianciája) állandó. A nullhipotézis teljesülése egyben azt is jelenti, hogy a modell homoszkedasztikus, míg az alternatív hipotézis a heteroszkedaszticitás feltételezését szimbolizálja, amely szerint a hibatenyező varianciája arányosan változik a j -edik magyarázó változó négyzetével. A k a magyarázóváltozók száma, esetünkben 1.

A próba menete az alábbi:

1. A heteroszkedaszticitásban feltehetően meghatározó szerepet játszó x_1 magyarázó változó növekvő sorrendbe rendezett értékei szerint írjuk fel az eredményváltozót és a reziduumokat.
2. Kiválasztunk c számú közepén elhelyezkedő értéket, amelyeknek megfelelő változók megfigyelt értékeit kihagyjuk a további számításokból. (A kihagyandó megfigyelések száma önkényes, legalább 0, ebben az esetben két részre osztjuk az adatállományt és nem hagyunk ki megfigyelt értékeket, a gyakorlatban minimum a mintaelemszám (n) egyhatoda illetve maximum az egy harmada. Kisebb minta esetén, pl. ha $n=30$, akkor célszerű, hogy a c ne legyen nyolcnál nagyobb)
3. A klasszikus legkisebb négyzetek módszere segítségével az első $(n-c)/2$ és az utolsó $(n-c)/2$ megfigyeléshez külön-külön regressziós függvényt illesztünk.
4. Kiszámítjuk a két regressziós függvény reziduális négyzetösszegét és elosztjuk a

$$\frac{n-c}{2} - (k+1) = \frac{n-c-2(k+1)}{2}$$

szabadságfokkal:

$$\sigma_1^2 = \sum_{i=1}^{(n-c)/2} e_i^2 / \frac{(n-c-2k-2)}{2}$$

$$\sigma_2^2 = \sum_{i=\frac{n+c}{2}+1}^n e_i^2 / \frac{(n-c-k-2)}{2}$$

ahol i követi az x_1 növekvő sorrendjét.

Esetünkben $k=1$, tehát:

$$\sigma_1^2 = \sum_{i=1}^{(n-c)/2} e_i^2 / \frac{(n-c-4)}{2}$$

$$\sigma_2^2 = \sum_{i=\frac{n+c}{2}+1}^n e_i^2 / \frac{(n-c-4)}{2}$$

5. A próbafüggvény:

$$F = \frac{\sigma_2^2}{\sigma_1^2}$$

ahol mind a számláló mind a nevező szabadságfoka egyaránt:

$$\frac{(n-c-2k-2)}{2} = \frac{(n-c-4)}{2}$$

Amennyiben a számított F-érték nagyobb, mint egy α szignifikancia szinthez tartozó F_α érték a nullhipotézist (homoszkedasztikus a modell) elvetjük, a modellt heteroszkedasztikusnak tekintjük.

A Goldfeld-Quandt-próba Excel parancsfájl adatok törlése ikonjára kattintva az adatok törlődnek, majd az új adatbázis bemásolása után meg kell adni a c-értéket.

A mintafeladat²¹⁸:

x=GDP/fő 2005-es \$-ban 2007-ben

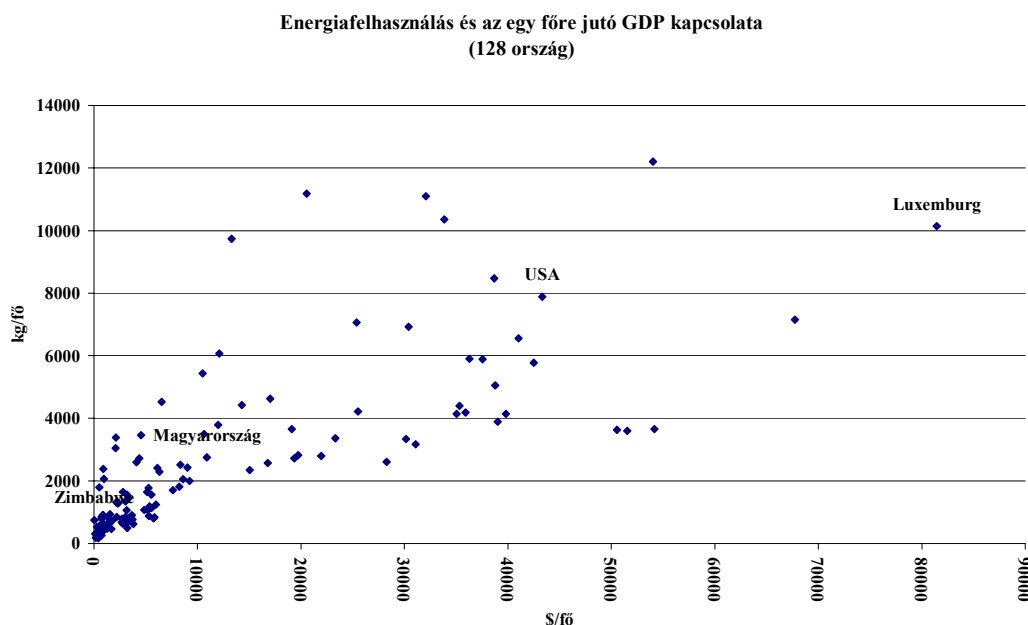
y=Egy főre jutó összes energiafelhasználás olaj kg egyenértékben kg/fő

n=128, k=1

Az adatokat x szerint növekvő sorrendbe helyeztük. ($\frac{A}{Z}$! Növekvő Excel paranccsal)

Az adatok grafikus ábráját az alábbiakban mutatjuk be.

A c érték megadása után a sárga mezőben lévő számok azt mutatják, hogy melyik megfigyeléseket hagyunk ki, a kék az első, a barna a második megfigyelt és a számítások során figyelembe vett értékek adatait mutatja, amiből a program kiszámítja a σ_1^2 és σ_2^2 értékeket. A c változtatásával igen gyorsan sok számítás elvégezhető. A mintaelemszám csökkentésével is elvégezhető a becslések, pl. megállapítható az, ha heteroszkedasztikus a modell, van-e olyan mintaelemszám amikor a modell homoszkedasztikussá válik. A program közli a nullhipotézis eredményét az F-próba alapján, továbbá a p-empirikus szignifikancia értékét. Ha p kisebb, mint 0,05, akkor 5 %-os szignifikancia szinten heteroszkedasztikus a modell, ha 0,05-nél nagyobb akkor homoszkedasztikus.



A számítások eredménye, a teljes minta esetében heteroszkedasztikus, ha a mintaelemszámot 76-ra csökkentjük, tehát csak az első 76 adattal dolgozunk, akkor homoszkedasztikus lesz a modell. Ez jelzi, hogy a

²¹⁸ Az adatok forrása:

http://earthtrends.wri.org/searchable_db/index.php?step=countries&ccID%5B%5D=0&allcountries=checkbox&the me=6&variable_ID=351&action=select_years

<http://www.ers.usda.gov/Data/Macroeconomics/>

lineáris regresszió helyett más típusú regresszióval kell becsülni, vagy a mintaelemszámot csökkenteni kell.

		c minimum 1/6	c maximum 1/3	legalább
n	128,00			
c	21,00	21,33	42,67	2,00
σ_1^2	360855,82			
σ_2^2	5759193,68			
F	15,96			
p	0,00			
Fkrit	1,59			
	0,05			
Heteroszkedasztikus modell				

		c minimum 1/6	c maximum 1/3	legalább
n	76,00			
c	13,00	12,67	25,33	2,00
σ_1^2	230282,54			
σ_2^2	427074,77			
F	1,85			
p	0,05			
Fkrit	1,86			
	0,05			
Homoszkedasztikusmodell				

4.5 A regressziós együtthatók összefüggései (Az útelemzés)

A parciális regressziós együttható a lineáris regressziós modellben szereplő többi tényezőváltozó hatását kiszűri. Ezzel szemben az egyszerű kétváltozós regressziós együttható értékében más tényezőváltozók hatása is kifejeződésre juthat. A parciális és a kétváltozós regressziós együtthatók összefüggései alapján további elemzéseket végezhetünk. Ebben használjuk fel a következő képleteket:

$$b_{y_1} = b_{y_1.2} + b_{21}b_{y_2.1}$$

$$b_{y_2} = b_{y_2.1} + b_{12}b_{y_1.2}$$

Fenti összefüggés százalékosan is kifejezhető, ha b_{y_1} -val illetve b_{y_2} -val végigosztjuk az egyenlet mindkét oldalát és szorzunk százzal.

Ahol :

b_{y_i} = az y és x_i ($i = 1, i = 2$) változókra vonatkozó egyszerű kétváltozós totális regressziós együttható; meghatározása az alábbi két függvényből történik:

$$y = b_{01} + b_{y_1}x_1 + e$$

$$y = b_{02} + b_{y_2}x_2 + e$$

A $b_{y_1.2}$ és a $b_{y_2.1}$ parciális regressziós együtthatók, meghatározása az:

$$y = b_{y_0} + b_{y_1.2}x_1 + b_{y_2.1}x_2 + e$$

függvényből történik.

A $b_{12} = x_1$ és x_2 változók egyszerű kétváltozós regressziós együtthatója (x_1 =eredményváltozó és x_2 a tényezőváltozó)

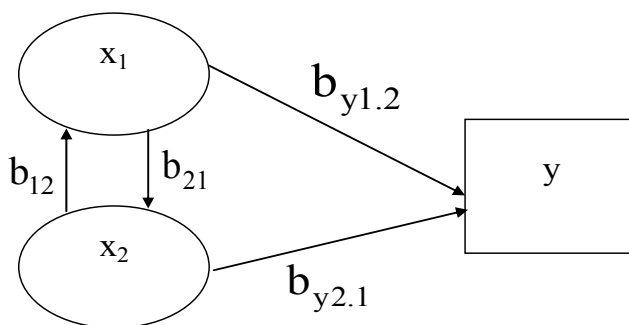
A $b_{21} = x_2$ és x_1 változók egyszerű kétváltozós regressziós együtthatója (x_2 = eredményváltozó, x_1 = tényezőváltozó) tehát:

$$x_1 = b_{10} + b_{12}x_2 + e$$

$$x_2 = b_{20} + b_{21}x_1 + e$$

A fenti két összefüggésből látható, hogy az egyszerű kétváltozós totális regressziós együttható (b_{y_i} $i = 1, 2$) két részből tevődik össze: az adott változó közvetlen hatásából ($b_{y1.2}$ és $b_{y2.1}$) és a közvetett, más változón keresztül érvényesülő hatásból. Az összefüggésből kitűnik, hogy az egyszerű kétváltozós totális, és a parciális regressziós együtthatók csak akkor egyeznek meg, ha a b_{21} (illetve b_{12}) és a $b_{y2.1}$ (illetve $b_{y1.2}$) értékek valamelyike nullával egyenlő. Ha b_{21} (illetve b_{12}) nulla, akkor a tényezőváltozók közötti kapcsolat elhanyagolható. Ha az első összefüggésben $b_{y2.1}$ nulla, akkor az x_2 tényezőváltozó bevonása a modellbe felesleges. Ugyanígy a második összefüggésben az x_1 tényezőváltozó bevonása a modellbe felesleges, ha $b_{y1.2} = 0$.

A két összefüggésből az is kiderül, hogy a legkisebb négyzetek módszerével levezetett becslőfüggvény csak akkor ad a regressziós paraméterekre torzítatlan becslést, ha a modell specifikációja során nem követünk el ún. specifikációs hibát, vagyis minden lényeges tényezőváltozót szerepeltetünk a modellben. A felsorolt regressziós modelleket le kell futtatni és a megfelelő regressziós együtthatók felhasználásával igazolhatók a fenti összefüggések.



4-7. ábra: Az útelemzés sémája

4.6 Késleltetett regressziós modellek. (Késleltetettmátrix.xls parancsfájl működése)*

A gazdasági életben gyakran tapasztaljuk, hogy egy esemény (folyamat) vagy döntés hatása csak némi késéssel, időben elhúzódva észlelhető. Tipikus példa erre a beruházás és a termelés, illetve az értékesítés kapcsolata, az import és az export (vagy belföldi) ár közötti összefüggések stb. A gazdasági jelenségek elemzésére, leírására szolgáló modellek egyik csoportja adott időpontban vagy időszakban fennálló bonyolult kölcsönkapcsolatok leírására, elemzésére szolgál, tehát egy adott (vagy feltételezett) állapotot kíván leírni. (pl. az optimalizálási feladat, egy adott időszak rögzített költség- és árviszonyai mellett keres optimális termelési programot, rögzített erőforrás és kereslet korlátok mellett). Ezeket a modelleket statikus modelleknek szokták nevezni. A modellek másik csoportja a gazdasági folyamatok mozgástörvényeit vizsgálja (pl. a gazdasági növekedés tényezőit) s arra ad választ, hogy hogyan befolyásolják az egyes jelenségek pillanatnyi állapotát saját vagy más jelenségek korábbi állapotai. Az ilyen kérdéseket vizsgáló modelleket dinamikus modelleknek hívjuk. Dinamikus összefüggés alatt azt értjük, hogy különböző időponthoz, vagy időszakhoz tartozó változók közötti összefüggéseket vizsgálunk. Statikus modellel csak stabil végállapotot lehet modellezni. Stabil viszonyok csak rövid ideig léteznek a gazdasági életben. A klasszikus newtoni fizika szerint a mozgás folytonos és így folytonos függvénnyel leírható. A newtoni fizika szerint a mozgást meghatározó erők hatásukat késleltetés nélkül fejtik ki. Pl. a rezgőmozgás esetén a nyugalmi irányba ható erő minden időpontban arányos a nyugalmi helyzettől való eltéréssel (a kitérés) távolságával. Gazdasági mozgások leírásával ez az eszköz nem hatékony, mert:

- a megfigyelés diszkrét és rögzített időpontokhoz kapcsolódik. A mérés eredménye így eltérő lehet, ha napi, ha heti, ha havi, ha negyedévi vagy ha évi adatokkal dolgozunk.

- nem végtelen kicsi, hanem véges, sőt gyakran igen hosszú reakcióidővel kell számolni.

A modern fizika (Plank, Einstein, Heisenberg stb.) az idő és mozgás fogalmát átértékelt és nem tekinti folyamatosnak sem a mozgást, sem az időt.

A késleltetés okai (akció és reakció időben szétválik):

1. A felismerési késés. A megfigyelés, regisztrálás, összegzés, feldolgozás időt igényel. (pl. tartós fogyasztási cikkek ritkábban vásárolunk)
2. Döntési késés. Időre van szükség a döntések meghozatalára és végrehajtására.
3. A technológiai késés (oka a gyártási idő).
4. A folyamatok tehetetlenségéből adódó késés.
5. Spekulációs késés, amikor pl. az eladók áremelkedésre számítanak, és ezért készleteznek.
6. Egyéb okok, pl. szervezeti késleltetés, a bürokrácia tehetetlensége és lassúsága stb.

4.6.1 A késleltetés modelljeinek rövid története

A folyamatok közötti kapcsolatok vizsgálatában korán, már az 1930-as években felmerült az a kérdés, hogy egy hatásnak milyen időbeli lefutása van, illetve az, hogy adott okok rövid és hosszú távú hatását milyen módon lehet szétválasztani. A korszak jeles képviselői például F. L. Alt és I. Fisher, akik megalapozták elméleteikkel a későbbi kutatásokat. Fisher dolgozta²¹⁹ ki 1937-ben az ún. naiv osztott késleltetésű modellt, ami a csökkenő súlyszámok (short-cut) elvén alapul, ahol a csökkenés számtani sor szerint történik. A megosztott késleltetésű modellek, az ún. DL²²⁰ modellek alaposabb kutatása az 50-es években kezdődött, elsősorban L. M. Koyck, P. D.²²¹ Cagan, M. Nerlove,²²² S. Almon²²³ és R. Solow²²⁴ kutatási eredményeit lehet kiemelni. Előtérbe került a végtelen osztású késleltetés alkalmazása, mégpedig úgy, hogy a súlyok csökkenése exponenciális módon történik. Az elmélet fejlődésével különböző modellek jöttek létre: pl. fordított V-késleltetésű, Almon-féle polinom eloszlású osztott késleltetésű modellek.

Az irodalom alapvetően két fajta modellt különböztet meg a késleltetés szempontjából:

1. egyszerű késleltetésű modellek: adott jelenség egy másik jelenség meghatározott idejű késleltetésétől függ csak, azaz
2. összetett (vagy elosztott) késleltetésű modellek²²⁵: a vizsgált jelenség a másik jelenség több (akár végtelen darabszámú) múltbeli értékétől is függ, vagyis a hatások eloszlanak az időben, azaz:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-k} + \varepsilon_t$$

Az előző két modell-egyenletben szereplő jelölések²²⁶:

- Y_t eredményváltozó, vagy magyarázott változó a t. időpontban,
- X_t magyarázó változó a t. időpontban,
- i a késleltetés mértéke egyszerű késleltetés esetén,
- k a késleltetés maximális mértéke $k \in (1, \infty)$,
- ε_t hibatermés,
- α, β_i regressziós együtthatók, paraméterek.

A β_0 regressziós együttható az X_t -hez tartozó súly és egyben parciális lineáris regressziós együttható megmutatja, hogyha X_t egy egységgel nő, akkor Y β_0 értékkel nő, a többi késleltetett magyarázóváltozó hatásának kiszűrése mellett. A β_0 -t egyidejű multiplikatornak is nevezik a nemzetközi szakirodalomban,

²¹⁹ Fisher I. [1937]

²²⁰ Distributed Lag

²²¹ Koyck, L. M. [1954]

²²² Nerlove, Marc, [1972]: 221-251

²²³ Almon, S. [1965]: 178-197.

²²⁴ Solow R. M. [1960]: 393-406.

²²⁵ Ramanathan Ramu [2003]: 452-456.

²²⁶ A késleltetett regressziós modelleknél a nagy x és y betűket használjuk a változók megnevezéséhez.

ugyanis ez nem más, mint X marginális hatása Y -ra ugyanabban t . időpontban. A β_i regressziós együtthatót i -ed rendű késleltetett multiplikátornak nevezik, mivel azt mutatja meg, hogy az X_{t-i} változó egy t előtti i időpontban bekövetkezett egységnyi növekedésének hatására mennyivel változott az Y_t értéke, vagyis a folyó időszaki eredményváltozó nagysága. Ha a gazdaság nyugalmi állapotban van, tehát más megfogalmazásban hosszú távú egyensúlyi állapot feltételezhető, akkor a regressziós paraméterek összegét hosszú távú multiplikátornak hívják, mivel megmutatja azt, hogy X változó egységnyi növekedése esetén, az Y mennyivel változik, a késleltetés időszaka alatt. Az így megbecsült kumulált hatás tehát²²⁷:

$$dY/dX = \sum_{i=0}^k \beta_i = \beta_0 + \beta_1 + \beta_2 X_{t-2} \dots + \beta_k = \beta$$

A standardizált késleltetett multiplikátor (β_i^*) megmutatja egy β_i regressziós együttható részesedését a β hosszú távú multiplikátor kumulált hatásából:

$$\beta_i^* = \frac{\beta_i}{\sum_{i=0}^k \beta_i} = \frac{\beta_i}{\beta}$$

Mivel a paraméterek nehezen becsülhetők, illetve gyakori a multikollinearitás és az autokorreláció jelensége, ezért rájuk vonatkozóan megszorításokat (ilyen lehet pl. a súlyok (β_i) geometriai sor szerinti csökkentése) kell tennünk. Ebből az ötletből, felismerésből születtek meg a fent említett osztott késleltetésű modellek.

A naiv osztott késleltetésű modellek.

*A Fisher-féle megoldás.*²²⁸

Fisher számtani haladvány szerint csökkenő súlyokat alkalmazott. Az alapján, hogy hány időszakra viszszaemlékezve számszerűsítette a hatást, beszélhetünk az általa megalkotott egyenletekről. A Fisher 1 egyenlet tehát a magyarázó változó jelenlegi, és egyel késleltetett értékét veszi figyelembe, a Fisher 2 és Fisher 3 modellek pedig egyre hosszabb késleltetést tételeznek fel.

Ennek alapján Fisher 1. modellje:

$$Y_t = \alpha + \beta_1 (2X_t + X_{t-1}) + \varepsilon_t,$$

ahol $F_1 = (2X_t + X_{t-1})$

helyettesítést alkalmazva kapjuk a következő egyenletet:

$$Y_t = \alpha_1 + \beta_1 F_1 + \varepsilon_t$$

Hasonló módon képezhető a Fisher 2. és Fisher 3. egyenlet is:

$$Y_t = \alpha_2 + \beta_2 F_2 + \varepsilon_t$$

valamint

$$Y_t = \alpha_3 + \beta_3 F_3 + \varepsilon_t$$

ahol

$$F_2 = (3X_t + 2X_{t-1} + X_{t-2}) \text{ és } F_3 = (4X_t + 3X_{t-1} + 2X_{t-2} + X_{t-3})$$

A Fisher egyenletek megoldása után az eredeti egyenletben a magyarázó változók paraméterei könnyen számszerűsíthetők.

Alt módszere.^{229 230 231}

Alt szintén több egyenletet állapított meg, melyek sorszama az előzőekhez hasonlóan a késleltetés mértékét mutatják. A különbség Alt és Fisher módszere között abban rejlik, hogy Alt nem alkalmazott megközelítést a magyarázó változók értékeire vonatkozóan. Alt először az Y_t értéket csak X_t értékével magyarázza, majd a második becslésnél az X_{t-1} -t is bevezeti, és így tovább, mindaddig, amíg az eredményül kapott regressziós együtthatónak értelme van. Számszerűsítsük a következő Alt modellt:

²²⁷ Gujarati Damodar N. [2003]: 658.

²²⁸ Fisher I. [1937]. 323-328.

²²⁹ Alt F. F. [1942]: 113-128.

²³⁰ Vető Istvánné [1980]: 28-29.

²³¹ Gujarati Damodar N. [2003]: 663-664.

$$\begin{aligned}
Y_t &= \alpha + \beta_0 X_t + \varepsilon_t \\
Y_t &= \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t \\
Y_t &= \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \varepsilon_t \\
Y_t &= \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \varepsilon_t
\end{aligned}$$

.....

Alt azt javasolta, hogy folytatni kell a késleltetést mindaddig, amíg a regressziós modell az elméleti feltételeknek eleget tesz. A további késleltetett tag bevonása feleslegessé válik például, ha a bevont regressziós paraméter nem különbözik szignifikánsan a 0-tól.

4.6.2 A fordított V-késleltetésű modellek.

A fordított V-késleltetésű modellek esetén – általában – kezdetben emelkedő, majd csökkenő súlyokat tapasztalunk, innen származik elnevezésük is. Az ún. Pascal (negatív binomiális) elosztású késleltetett modellek kidolgozása R. M. Solow²³² nevéhez fűződik. Az ilyen eloszlás speciális eseteként értelmezhető a széles körben elterjedt Koyck módszer, ami a paraméterek megfelelő megválasztásával állítható elő.

Az osztott késleltetésű modell alapegyenlete a következő:

$$Y_t = \alpha + \beta(\omega_0 X_t + \omega_1 X_{t-1} + \omega_2 X_{t-2} + \dots + \omega_i X_{t-i} + \dots) + \varepsilon_t$$

ahol $0 \leq \omega_i \leq 1$ és $\beta \sum_{i=0}^{\infty} \omega_i = 1$.

A súlyok Pascal eloszlás esetén az alábbiak:

$$\omega_i = \binom{r+i-1}{i} (1-\lambda)^r \lambda^i = \frac{(r+i-1)!}{i!(r-1)!} (1-\lambda)^r \lambda^i$$

ahol:

- ω_i a késleltetés relatív súlyrendszere,
- r az eloszlás rendje ($r \in Z^+$) pozitív egész szám: 1,2,3.....
- i a késleltetés mértéke, 0, 1, 2,.....k..... ∞
- λ a becslendő paraméter, $0 < \lambda < 1$

Figyelembe véve, hogy n elem k -ad osztályú kombinációja:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Pl. $i=2$ esetén:

$$\begin{aligned}
\omega_2 &= \binom{r+2-1}{2} (1-\lambda)^r \lambda^2 = \frac{(r+2-1)!}{2!(r-1)!} (1-\lambda)^r \lambda^2 = \\
&= \frac{(r+1)r}{2!} (1-\lambda)^r \lambda^2
\end{aligned}$$

Az eredeti egyenlet az alábbi formában írható fel:

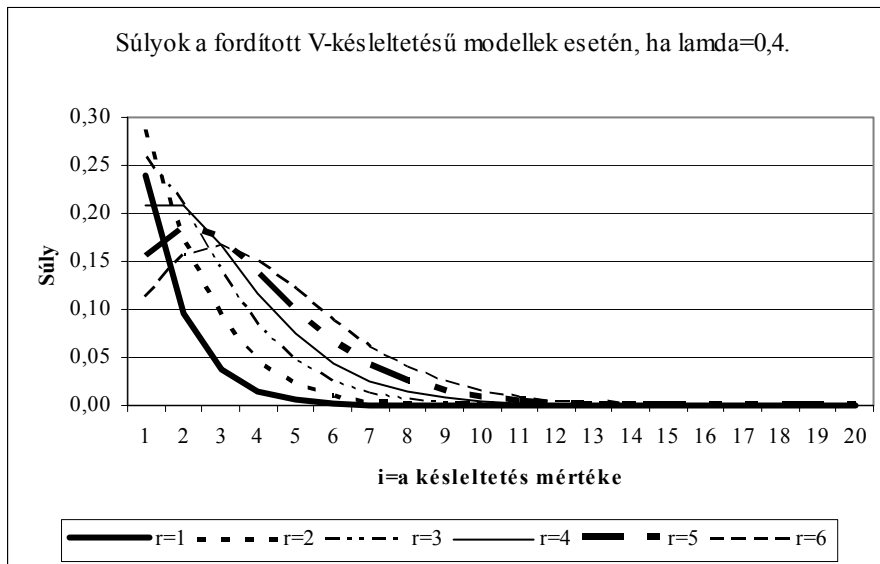
$$\begin{aligned}
Y_t &= \alpha + \beta \left\{ (1-\lambda)^r X_t + r\lambda (1-\lambda)^{r-1} X_{t-1} + \frac{r(r-1)}{2!} \lambda^2 (1-\lambda)^{r-2} X_{t-2} + \dots + \frac{r(r-1)(r-2)\dots(r-k+1)}{k!} \lambda^k (1-\lambda)^{r-k} X_{t-k} + \dots \right\} + \varepsilon_t = \\
&= \alpha + \beta (1-\lambda)^r \sum_{k=0}^{\infty} \binom{r+k-1}{k} \lambda^k X_{t-k} + \varepsilon_t
\end{aligned}$$

A súlyok alakulását mutatja az alábbi két ábra²³³, ha $\lambda=0,4$ és $\lambda=0,6$, az eloszlás rendje mindkét esetben $r = 1,2,3,4,5,6$.

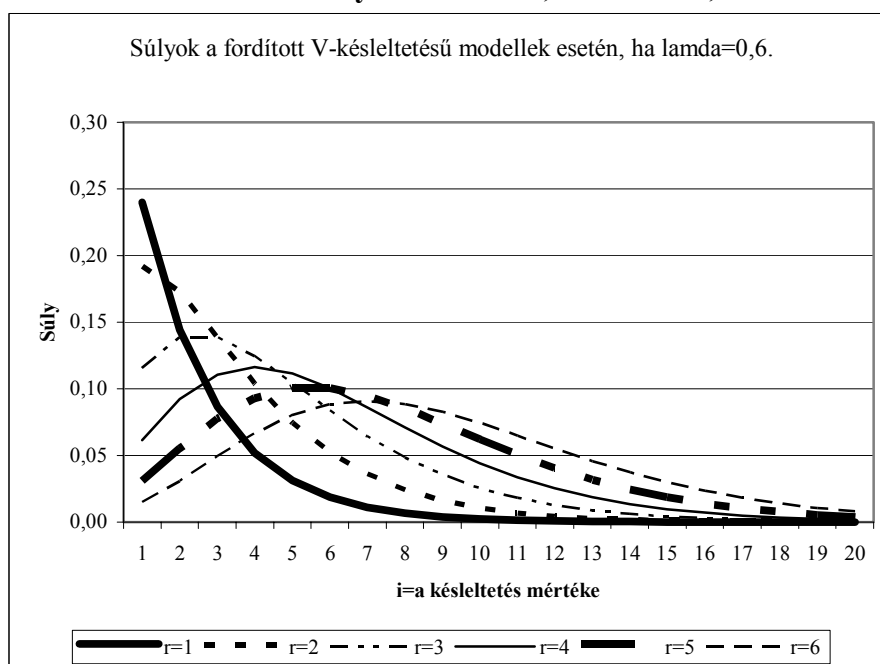
A λ növelése $\lambda=0,4$ -ről $\lambda=0,6$ -ra a következő változásokat eredményezi: $r=1$ esetben nő a késleltetés hossza, az $r=2,3,4,5,6$ eseteiben pedig a fordulópont – amikor a növekedő szakasz csökkenőbe megy át – később következik be. Az r növekedésével a fordulópont később következik be, tehát nő az emelkedő szakasz időtartama.

²³² Solow R. M.[1960].

²³³ A fordítottW.xls felhasználásával.



4.8. ábra: Súlyok alakulása, ha $\lambda = 0,4$



4.9. ábra: Súlyok alakulása, ha $\lambda = 0,6$

Ha az eloszlás rendje 1, akkor könnyen belátható, hogy

$$\omega_i = \frac{(r+i-1)!}{i!(r-1)!} (1-\lambda)^r \lambda^i = \frac{i!}{i!} (1-\lambda) \lambda^i = (1-\lambda) \lambda^i$$

Tehát geometriai késleltetésű lesz a Pascal-késleltetésű eloszlás. Ha r értéke nő akkor az emelkedő szakasz forduló pontja nagyobb késleltetésnél következik be.

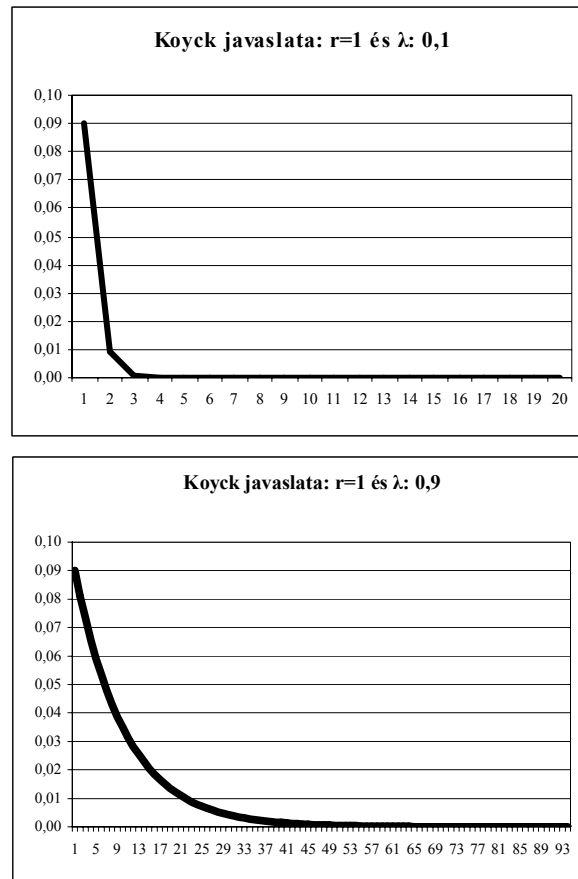
4.6.3 Koyck módszerei

Ha $r=1$, akkor tehát a Pascal-eloszlás geometriai késleltetésű eloszlásra redukálódik.²³⁴ ²³⁵ A súlyok (ω_i) alakulását $\lambda=0,1$ és $\lambda=0,9$ a következő két ábrán mutatjuk be. A két ábrán jól látható, hogy λ értéke tulajdonképpen azt mutatja meg, hogy mennyire „régire” adatoknak van még hatása a jelenre, azaz milyen gyors a felejtés. Az első ábrán, ahol $\lambda = 0,1$ a súlyok 6 késleltetés után „tűnnek el”, ekkor lesz a súly=0,00000, míg ugyanez a helyzet a $\lambda = 0,9$ esetben a 94 időszakkal késleltetett adatok esetében következik be. Általánosságban elmondhatjuk, hogy minél nagyobb egy modellben λ értéke (de $0 < \lambda < 1$), an-

²³⁴ Koyck, L. M. [1954]

²³⁵ Gujarati Damodar N. [2003]: 665-667.

nál lassabb a rendszer felejtése. Ha $\lambda \rightarrow 0$, akkor nincs késleltetés, $\omega \rightarrow 1$, ha viszont $\lambda \rightarrow 1$ akkor végtelenül nagy a késleltetés, tehát $\omega \rightarrow 0$.



4.10. ábra. Koyck javaslatai

Ha az eloszlás rendje 1, akkor, ahogy már bemutattuk:

$$\omega_i = (1-\lambda)\lambda^i \quad i=1,2,3,\dots,k.$$

Ezt az osztott késleltetésű modell alapegyenletébe visszahelyettesítve az alábbi összefüggés adódik:

$$\begin{aligned} Y_t &= \alpha + \beta_0 (\omega_0 X_t + \omega_1 X_{t-1} + \omega_2 X_{t-2} + \dots + \omega_k X_{t-k} + \dots) + \varepsilon_t = \\ &= \alpha + \beta_0 (1-\lambda) (\lambda^0 X_t + \lambda^1 X_{t-1} + \lambda^2 X_{t-2} + \dots + \lambda^k X_{t-k} + \dots) + \varepsilon_t = \\ &= \alpha + \beta_0 X_t + \beta_0 \lambda X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \dots + \beta_0 \lambda^k X_{t-k} + \dots + \varepsilon_t \end{aligned}$$

Ezt az összefüggést Y_{t-1} -re felírva, λ -val beszorozva és a két egyenletet egymásból kivonva, az alábbi modellt kapjuk, ami Koyck első módszereként ismert a szakirodalomban:

$$Y_t - \lambda Y_{t-1} = \alpha(1-\lambda) + \beta_0 X_t + (\varepsilon_t - \lambda \varepsilon_{t-1}).$$

Az egyenlet rendezése után a becslésre alkalmas függvényt kapjuk, ahol az eredményváltozó Y_t , a magyarázó változók X_t és Y_{t-1} :

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + (\varepsilon_t - \lambda \varepsilon_{t-1})$$

$$Y_t = \alpha' + \beta_0 X_t + \lambda Y_{t-1} + v_t$$

Az eredeti konstans paraméter az alábbi összefüggésből nyerhető:

$$\alpha(1-\lambda) = \alpha'$$

$$\alpha = \frac{\alpha'}{(1-\lambda)}$$

Az α β_0 λ ismeretében az eredeti Koyck késleltetett regressziós függvény felírható:

$$Y_t = \alpha + \beta_0 X_t + \beta_0 \lambda X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \dots + \beta_0 \lambda^k X_{t-k} + \dots + v_t$$

$$Y_t = \alpha + \hat{\beta}_0 X_t + \hat{\beta}_1 X_{t-1} + \hat{\beta}_2 X_{t-2} + \dots + \hat{\beta}_k X_{t-k} + \dots + v_t$$

A paraméterek közötti összefüggés:

$$\begin{aligned}\hat{\beta}_0 &= \beta_0 \\ \hat{\beta}_1 &= \beta_0 \lambda \\ \hat{\beta}_2 &= \beta_0 \lambda^2 \\ \hat{\beta}_3 &= \beta_0 \lambda^3 \\ &\vdots \\ \hat{\beta}_k &= \beta_0 \lambda^k\end{aligned}$$

Látható, hogy két egymást követő paraméter hányadosa állandó és λ -val egyenlő.

$$\lambda = \frac{\hat{\beta}_i}{\hat{\beta}_{i-1}}$$

A β paraméterek összegét felhasználva meghatározhatjuk a hosszú távú multiplikátort, tehát a kumulált hatást is, vagyis, ha pl. a magyarázó változó az üzembe helyezett beruházások összege, akkor a kumulált hatás 1, mivel a beruházások összege egy idő után üzembe helyezésre kerül. Az alábbi képletben a regressziós paraméterek összegét (kumulált hatás) úgy értelmezhetjük, hogy az X magyarázó változó egy-ségnyi változása hosszú távon mennyivel növeli átlagosan az Y eredményváltozó értékét.^{236 237}

$$\sum_{i=0}^{\infty} \beta_i = \beta_0 (1 + \lambda + \lambda^2 + \lambda^3 + \dots) = \beta_0 \left(\frac{1}{1-\lambda} \right)$$

Ugyanis:

A mértani sorozat összege legyen s , q_0 a kezdőérték (esetünkben $q_0 = \beta_0$), és q a két szomszédos tag hányadosa (esetünkben $q = \lambda$), ami állandó:

$$\begin{aligned}s &= q_0 \left(\frac{q^n - 1}{q - 1} \right) = \beta_0 \frac{\lambda^n - 1}{\lambda - 1} = \beta_0 \frac{-1}{\lambda - 1} = \beta_0 \frac{1}{1 - \lambda} = 1 \\ \beta_0 &= 1 - \lambda\end{aligned}$$

Mivel:

$$\lim_{n \rightarrow \infty} \lambda^n = 0$$

A hosszú távú (kumulált) késleltetett hatás tehát:

$$\frac{\beta_0}{1 - \lambda}$$

A kumulált késleltetés átlagos hosszát \bar{i} az alábbi képlettel határozhatjuk meg, ahol az átlagolandó érték a késleltetés hossza ($i=1, 2, \dots, \infty$) a súly pedig a regressziós paraméterek becsült értéke ($\hat{\beta}_i$)²³⁸:

$$\bar{i} = \frac{\sum_{i=0}^{\infty} i \hat{\beta}_i}{\sum_{i=0}^{\infty} \hat{\beta}_i}$$

²³⁶ Mundruczó György [1981]: 176-177.

²³⁷ Greene, William H. [2003]: 560-561.

²³⁸ Gujarati Damodar N. [2003]: 668.

A fenti képletben – az előzőek alapján – a nevezőben lévő számnak, a paraméterek összegének a közgazdasági jelentése az, hogy X tényezőváltozó egységnyi növelése k időszakon át mennyivel növeli átlagosan az Y eredményváltozó értékét.

A Koyck modell esetében a kumulált késleltetés átlagos hossza²³⁹:

$$\bar{i} = \frac{\lambda}{1-\lambda}$$

Átlagos késleltetés annak az időnek az átlagos hosszát jelenti, amíg az X magyarázó változó (egységnyi) változását átvisszük az Y függő változóra.

A Koyck modell esetében a kumulált késleltetés mediánját (Me_i) az alábbi képlettel lehet meghatározni²⁴⁰:

$$Me_i = -\frac{\log 2}{\log \lambda}$$

A medián késleltetés az az idő, ami az Y változó teljes változása első felének vagy 50%-ának X egységnyi tartós változásának követéséhez szükséges.

Az átlagos késleltetés (medián és számtani átlag) a Koyck modell esetében λ különböző értéki függvényében változik, a λ növekedésével az átlagos késleltetés időtartama is nő, amit az alábbi táblázat szemléltet. A medián és az átlag $\lambda=0,5$ értéknél egyezik meg, nagysága 1.

λ	Me_i	i átlag
0,1	0,3010	0,1111
0,2	0,4307	0,2500
0,3	0,5757	0,4286
0,4	0,7565	0,6667
0,5	1,0000	1,0000
0,6	1,3569	1,5000
0,7	1,9434	2,3333
0,8	3,1063	4,0000
0,9	6,5788	9,0000

Koyck második modellje^{241 242} abban különbözik az elsőtől, hogy az X_t és az X_{t-1} változóknak tetszőleges súlya van²⁴³ a csökkenő geometriai sorozat tulajdonságú súlyrendszer csak ezt követően kezdődik el. A könnyebb érthetőség érdekében a Koyck 1. módszerének leírásánál alkalmazott jelöléstechnikát célszerű használni. Koyck 2. módszerének geometriai késleltetésű alapkoncepciója miatt

$$Y_t = \alpha + \beta_0 X_t + \beta_1 (X_{t-1} + \lambda X_{t-2} + \lambda^2 X_{t-3} + \dots + \lambda^{k-1} X_{t-k}) + \varepsilon_t$$

alakban keressük a közelítést. Feladatunk az α , β_0 , β_1 , és a λ paraméterek meghatározása. Végezzük el a kijelölt szorzási műveletet, majd képezzük az $Y_t - \lambda Y_{t-1}$ kifejezést!

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_1 \lambda X_{t-2} + \beta_1 \lambda^2 X_{t-3} + \dots + \beta_1 \lambda^{k-1} X_{t-k} + \varepsilon_t$$

$$Y_{t-1} = \alpha + \beta_0 X_{t-1} + \beta_1 X_{t-2} + \beta_1 \lambda X_{t-3} + \beta_1 \lambda^2 X_{t-4} + \dots + \beta_1 \lambda^{k-1} X_{t-k-1} + \varepsilon_{t-1}$$

$$\lambda Y_{t-1} = \lambda \alpha + \lambda \beta_0 X_{t-1} + \lambda \beta_1 X_{t-2} + \beta_1 \lambda^2 X_{t-3} + \beta_1 \lambda^3 X_{t-4} + \dots + \beta_1 \lambda^k X_{t-k-1} + \lambda \varepsilon_{t-1}$$

²³⁹ Gujarati Damodar N. [2003]: 668.

²⁴⁰ Gujarati Damodar N. [2003]: 668.

²⁴¹ Griliches Z. : Distributed Lags: A survey. Econometrica. 1967. évi 1. sz. 16-49. p.

²⁴² Sipos Béla [1982]: 40-45.

²⁴³ Kiss Tibor [1985]: 1001-1011.

$$Y_t - \lambda Y_{t-1} = \alpha(1-\lambda) + \beta_0 X_t + (\beta_1 - \lambda\beta_0) X_{t-1} + (\varepsilon_t - \lambda\varepsilon_{t-1})$$

Rendezés után a becslésre alkalmas függvény:

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + (\beta_1 - \lambda\beta_0) X_{t-1} + \lambda Y_{t-1} + (\varepsilon_t - \lambda\varepsilon_{t-1})$$

$$Y_t = \alpha' + \beta_0 X_t + \beta_1' X_{t-1} + \lambda Y_{t-1} + v_t$$

$$\alpha' = \alpha(1-\lambda)$$

$$\alpha = \frac{\alpha'}{(1-\lambda)}$$

$$\beta_1' = \beta_1 - \lambda\beta_0$$

$$\beta_1 = \beta_1' + \lambda\beta_0$$

Az egyenlet rendezése után tehát a becslésre alkalmas függvényt kapjuk, ahol az eredményváltozó Y_t , a magyarázó változók X_t , X_{t-1} és Y_{t-1}

Pascal eloszlás esetén a becslés,^{244 245} ha $r=2$

$$Y_t = 2\lambda Y_{t-1} - \lambda^2 Y_{t-2} + \beta(1-\lambda)^2 X_t + e_t$$

Pascal eloszlás esetén a becslés,^{246 247} ha $r=3$

$$Y_t = 3\lambda Y_{t-1} - 3\lambda^2 Y_{t-2} + \lambda^3 Y_{t-3} + \beta(1-\lambda)^3 X_t + e_t$$

4.6.4 Almon-féle polinom eloszlású osztott késleltetésű modellek

Tekintsük^{248 249} a következő – véges számú késleltetést figyelembe vevő – egyenlet becslését:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-k} + \varepsilon_t$$

Összevont formában:

$$Y_t = \alpha + \sum_{i=0}^k \beta_i X_{t-i} + \varepsilon_t$$

A nyilvánvaló multikollinearitás miatt nem kapunk megbízható becslést a paraméterekre, ezért ismételten szükségünk van feltevésekre. Almon azt feltételezte, hogy a késleltetett modell súlyként szereplő paraméterei előre adott fokszámú polinom szerint alakulnak. Az Almon-féle késleltetés jele $PDL(k,r)$ ²⁵⁰, ahol k ($i=1,2,3,\dots,k$) a késleltetések hossza, míg r ($r=2,3,\dots,m$) a polinom feltételezett fokszáma.

Általános formában:

$$\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \dots + \alpha_m i^m$$

Így $r = 2$ esetén:

$$\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2$$

Ekkor eredeti egyenletünk a következő alakra módosul:

$$Y_t = \alpha + \sum_{i=0}^k \beta_i X_{t-i} + \varepsilon_t = \alpha + \sum_{i=0}^k (\alpha_0 + \alpha_1 i + \alpha_2 i^2) X_{t-i} + \varepsilon_t$$

A zárójel felbontása után, valamint az X_t összegszerű tagjainak megfelelő helyettesítésével a következő egyenletet kapjuk:

²⁴⁴ Solow R. M. [1960]:397.

²⁴⁵ Kiss Tibor [1985]: 1004.

²⁴⁶ Solow R. M. [1960]:397.

²⁴⁷ Kiss Tibor [1985]: 1004.

²⁴⁸ G. S. Maddala [2004]: 465-472.

²⁴⁹ Gujarati Damodar N. [2003]: 687-693.

²⁵⁰ Polynomial Distributed Lag

$$\begin{aligned}
Y_t &= \alpha + \sum_{i=0}^k (\alpha_0 + \alpha_1 i + \alpha_2 i^2) X_{t-i} + \varepsilon_t = \\
&= \alpha + \alpha_0 \sum_{i=0}^k X_{t-i} + \alpha_1 \sum_{i=0}^k i X_{t-i} + \alpha_2 \sum_{i=0}^k i^2 X_{t-i} + \varepsilon_t = \\
&= \alpha + \alpha_0 Z_{0t} + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \varepsilon_t
\end{aligned}$$

Ugyanis:

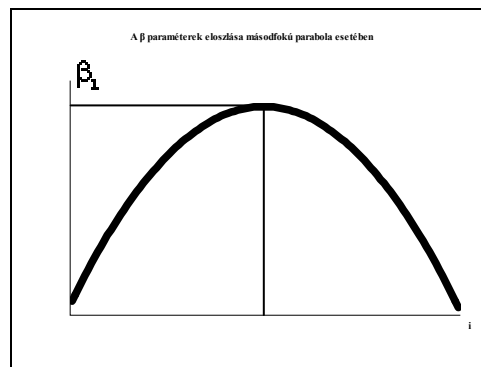
$$\begin{aligned}
Z_{0t} &= \sum_{i=0}^k X_{t-i} \\
Z_{1t} &= \sum_{i=0}^k i X_{t-i} \\
Z_{2t} &= \sum_{i=0}^k i^2 X_{t-i}
\end{aligned}$$

Legyen $r=2$, és $k=5$ akkor:

$$\begin{aligned}
Z_{0t} &= \sum_{i=0}^5 X_{t-i} = (X_t + X_{t-1} + X_{t-2} + X_{t-3} + X_{t-4} + X_{t-5}) \\
Z_{1t} &= \sum_{i=0}^5 i X_{t-i} = (X_{t-1} + 2X_{t-2} + 3X_{t-3} + 4X_{t-4} + 5X_{t-5}) \\
Z_{2t} &= \sum_{i=0}^5 i^2 X_{t-i} = (X_{t-1} + 4X_{t-2} + 9X_{t-3} + 16X_{t-4} + 25X_{t-5})
\end{aligned}$$

A másodfokú polinomális késleltetés sémáját az alábbi ábra mutatja, ahol látható, a késleltetés növekedésével ($i=1,2,3,\dots,k$) a regressziós paraméterek egy ideig nőnek, majd a csúcspont elérése után csökkennek.

Gyakran alkalmazott módszer az ún. végpontfeltételek alkalmazása, amelyek azt kötik ki, hogy a késleltetésnél figyelembe vett értékek előtti és utáni súlyok 0 értéket vegyenek fel, ekkor a becslés tovább egyszerűsödik.



4.11. ábra. A másodfokú polinomális késleltetés sémája

A **késleltetettmátrix.xls parancsfájl** kéri: a megfigyelt időpontokat, az eredményváltozó (Y_t) és magyarázóváltozó (X_t) adatsorokat melyek között késleltetett hatás feltételezhető. Ennek alapján elkészíti a grafikus ábrát, ahol már ellenőrizhető, hogy a magyarázóváltozó és az eredményváltozó között van-e késleltetett kapcsolat, ugyanis pl. ebben az esetben, az X növekedő szakaszát az Y növekedő szakasza követi és fordítva, az X csökkenő szakaszát az Y csökkenő követi. Ezt követően közli a program az adatsorok késleltetett értékeit, a késleltetett munkalapon, és elkészíti külön-külön munkalapokon a 12, az előzőekben bemutatott késleltetett modell adatmátrixát (ezek: Fisher1, Fisher2, Fisher3, Alt1, Alt2, Alt3, Alt4, Koyck1, Koyck2, Almon, Pascal2, Pascal3) készíti el, továbbá közli a regressziós együtthatókat és a vizsztatranszformált együtthatókat. Az adattranszformáció után a regresszio.xls parancsfájl felhasználásával lehet ellenőrizni a késleltetett regressziós modelleket.

4.7 A hatványkitevős, Cobb-Douglas termelési függvény (A termelési függvény átlag és határmuta-tói.xls parancsfájl működése)*

Az egy egyenletbe sűrített regressziós modelleknek az ágazati és a vállalati gyakorlatban egyaránt használt formái a termelési függvények, melyek nagy szerepet játszanak a termelési folyamatok leírásában.²⁵¹

²⁵²A termelési folyamat eredménye (volumene) nagyszámú műszaki, gazdasági, természeti és társadalmi tényezőtől függ. Közgazdasági és műszaki elemzés útján határozhatók meg azok a termelési tényezők, melyek az adott gazdasági egységnél a legnagyobb hatással vannak a kibocsátásra. A matematikai, statisztikai elemzés e hatások modellezésével és mértékének számszerűsítésével foglalkozik. A termelési ráfordításokat termelési tényezőknek hívjuk. A termelési tényezőkhez tartozik például a föld, a tőke, a nyersanyag, az energia és a munkaerő. A munkaerő személyi termelési tényező, míg a többi termelési tényező tárgyi tényező. A tőkejavakhoz tartoznak a termelő- és szállító berendezések, épületek, számítógépek és egyéb gépek. A tőke fogalmát gyakran egy üzleti vállalkozás beindításához vagy fenntartásához szükséges pénzösszeg meghatározására használják, amit mi pénztőkének hívunk, míg a termelésben előállított tényezőket fizikai tőkének nevezzük. A fizikai tőkét holt munkának is hívjuk megkülönböztetve a munkaerőtől, az élők munkától. A termelési függvények tehát, valamely gazdasági egység termelésének volumenét fejezik ki a vizsgált időszakban az élők munkára fordítások, a fizikai tőke ráfordítások és egyéb termelési tényezők függvényében. A termelés eredménye és a termelési tényezők közötti összefüggést sztochasztikus formában, technikai (technológiai) egyenletekkel írják le. Az alapvető két termelési tényező, az élők munka és holt munka ráfordítások mellett, a termelési függvények tartalmazhatják az anyag-, és energiaráfordításokat és egyéb tényezőket is. A termelési függvények segítségével a vállalat (általában a gazdaság) viselkedésének technológiai korlátait írhatjuk le. A technológiai korlátok azt jelentik, hogy egy adott mennyiségű kibocsátás csak bizonyos ráfordítás (input) kombinációk révén valósítható meg, a vállalat a technológiailag megvalósítható termelési tervekre kénytelen korlátozni a tevékenységét, vagyis már döntött az optimális termékösszetétel és technológia kérdésében.

A termelési függvények felírásánál feltételezzük, hogy:

- a vizsgált törvényszerűség időben állandó, vagy csak lassan változik, vagy ismert a változása;
- az elemzés tárgya, a termelési eredmény közvetlenül vagy közvetve mérhető;
- az elemzés tárgyára lényeges hatást gyakorló tényezők elhatárolhatók azoktól a tényezőktől, amelyeknek szerepe elhanyagolható;
- az adatok hozzáférhetők és összehasonlíthatók.

A termelés, a kibocsátás (output) mérése

A termelés eredménye esetében a legmegfelelőbb megoldás az lenne, ha azt természetes mértékegységben figyelnénk meg. A természetes mértékegységben történő számbavétel során olyan mennyiségi egységeket alkalmazunk, amelyek a termék fizikai tulajdonságaival, a használati értékkel vannak kapcsolatban. A természetes mértékegységben való számbavétel jelentőségét az húzza alá, hogy ez kifejezi az adott időszakban előállított használati értékek tömegét és ez képezi az összes többi számbavételi mód; az értéki számbavétel, a munkamértékegység, a terméksoros volumenindex stb. alapját. Például, a cementüzem esetében a termelés viszonylag homogén, és így az üzem kibocsátása, amit az adott időszak folyamán termeltek, tonnában mérhető. Általában a gazdasági egységeknél előállított termékek száma igen nagy, így a termelés eredményét legtöbbször valamilyen értéki mutatószámmal mérjük.

A termelés értéki mutatószámai többek között az alábbiak lehetnek:

- Bruttó termelési érték, egyenlő az értékesített termékek és szolgáltatások alapján keletkezett árbevételnek, a saját előállítású eszközök aktivált (állományba vett) értékének²⁵³ és a saját termelésű készletek állomány változásának az összegével.

²⁵¹ Cobb, C.W. – Douglas, P.H. [1928] volt az első publikáció a termelési függvények témakörében, a szerzők feltételezték, hogy a két magyarázóváltozós, hatványkitevős regressziós függvényben a kitevők összege 1.

²⁵² Kádas Kálmán [1944] publikációja volt Magyarországon az első, amelyik a Cobb-Douglas termelési függvény ágazati alkalmazásával foglalkozott.

²⁵³ A saját előállítású eszközök aktivált értékét mindig az előállítás tényleges költségeinek összegével kell számításba venni.

- Anyagmentes termelési érték, vagyis a hozzáadott érték, (GDP vállalkozás szintű megfelelője) a bruttó termelési értékből kivonjuk az anyagköltséget és az igénybe vett anyagjellegű szolgáltatás költségét.
- Nettó termelési érték, az anyagmentes termelési értékből kivonjuk az állóeszközök értékcsökkenésének a költségét.

A bruttó jellegű mutatószámok közös jellemzője, hogy ha az egymást követő termelő folyamatok értékeit összegezzük, akkor az összegzett termelési érték többszörös számbavételt tartalmaz, ezt halmozódásnak nevezzük. A nettó jellegű termelési mutatószámok a létrehozott termékeknek csak azon érték részét tartalmazzák, melyet a vizsgált időszakban a megfigyelt termelőegység hozott létre. Így a nettó termelési érték az átvitt munka egyetlen elemét sem tartalmazza. A nettó termelési érték megbízhatóan jellemzi a termelőegységnél előállított új értéket, nagysága nem függ sem az ipar szervezeti felépítésétől, sem az egyes iparvállalatok belső tagozódásától. A nettó termelési érték kiszámításánál sok esetben problémát okoz a vizsgált időszak termelését terhelő amortizáció megállapítása, melyhez a fizikai kopás mellett, az erkölcsi kopás figyelembevétele is szükséges lenne. Ezért kerül sor a gyakorlatban az anyagmentes termelési érték meghatározására, mely a teljes termelés, valamint az összes anyagköltség és egyéb anyagjellegű költség különbsége. Az anyagmentes termelési érték alkalmazása leegyszerűsíti a termelési függvényt, mivel a nyersanyag változókat nem szükséges explicit módon belefoglalni. Vállalati szinten végzett elemzéseknél általában megfelelő a bruttó termelési mutatószámmal történő mérés. Az ágazati szintű termelési függvény számításoknál viszont a jelentős halmozódás kiküszöbölése érdekében célszerűbb a nettó termelés mutatóját alkalmazni. A termelés volumenének értékbeli mérésekor, mivel a termelési érték idősorai képezik a vizsgálat adatbázisát nagyon fontos az összehasonlíthatóság biztosítása. Ez történhet árindexek felhasználásával. Az így nyert idősorok, a termelés volumenének változását mutatják. Ágazati szinten alkalmazzák a terméksorok alapján számított indexeket, amely a termelés mennyiségi változását jelzi. A terméksoros volumenindex valamely ágazat termelési volumenének változását, az adott ágazat termelő tevékenységét leginkább jellemző termékek mennyiségi adataiból számított egyedi indexek súlyozott átlagával méri. Súlyként általában a bér és amortizáció együttes összegét alkalmazzák. A terméksoros volumenindexszel közelíthetjük a bruttó termelés alakulását, ha a bruttó értékkel súlyozunk. Ha a munkások teljesített óráival súlyozunk, akkor a nettó termelést, ha a bér plusz amortizációval súlyozzuk, akkor az anyagmentes termelést közelítjük. A terméksorok alapján számított közelítő volumenindex eltérhet a termelési módszer alapján számított nettó termelési indextől. Ennek okaként a következők említhetők meg: a közelítő index nem tükrözi a fajlagos anyagfelhasználás változását; a számítás során alkalmazott súlyok aránya eltér a nettó értékek arányától; eltekint bizonyos minőségi jellemzőktől; reprezentatív jellege miatt nem vesz minden terméket figyelembe; és az új termékek késve kerülhetnek be a mintába; eltekint a szolgáltatásoktól, a félkész és befejezetlen termelés állományváltozásától, stb.

A ráfordítások (inputok) mérése

Az élőmunka ráfordítás az alábbi mutatószámokkal mérhető:

- Létszámadatokkal (az összes foglalkoztatottak, fizikai foglalkoztatottak, szakmunkások stb. létszámával).
- Teljesített munkanapok vagy teljesített munkaórák számával.
- Munkabér és közterhei adatokkal.

Az élőmunka-ráfordítást az összes foglalkoztatottak létszámával mérve az élőmunkát egyneműnek tekintjük, bár ez nyilvánvalóan egyszerűsítést jelent. Ezt a megközelítést az indokolja, hogy a termelés volumenét nemcsak a fizikai foglalkozásúak száma befolyásolja, hanem közvetetten a nem fizikai (alkalmazotti) apparátus is pl. a műszaki fejlesztés, az irányítás színvonala által. A fizikai foglalkozásúak létszámát tekintve, a termelés és a közvetlenül termelő munkát végző dolgozók közötti kapcsolat általában szorosabb, mint a termelés és a közvetve termelő munkát végző dolgozók között. A közvetve termelő munkát végzők tevékenysége a közvetlen termelő munkához kapcsolódik, azt irányítja, ellenőrzi, kiszolgálja, ellátja, pl. közvetlen termelésirányítók, anyagmozgatók, karbantartók. Célszerű ezért külön adatsorral közelíteni a

fizikaiak említett kategóriáit. A foglalkoztatottak adatsorai és a munkanapokra vonatkozó megfigyelések hátránya, hogy nem tükrözik az egy munkanapra jutó átlagos munkaóra számot illetve annak változását. Az élómunka-ráfordítás legjobb közelítésének a teljesített munkaórák száma tekinthető, mivel ez tükrözi a leginkább a munkaráfordítások nagyságát. Problémát az jelent, hogy a teljesített munkaórák száma a munkahelyen eltöltött időt jelenti, mely nem minden esetben azonos a munkában eltöltött tényleges idővel. Teljesített óráként jelentkezik ugyanis az állásidő, amikor a fizikai dolgozó pl. anyag, energiahiány, géphiba miatt nem végez termelő munkát. Tehát, amennyiben az adatok rendelkezésre állnak, célszerű a teljesített, a bélyegzett munkaórákból az állásidőt levonni.

A fizikai tőkeráfordítás, azaz az állóeszköz-felhasználás közelítésére a következő megoldások lehetségesek:

- Az összes termelő állóeszköz nyitó, záró vagy átlagos állománya.
- A termelő állóeszközökön belül a gépek, berendezések, felszerelések nyitó, záró vagy átlagos állománya.

Ha az év folyamán jelentős beruházást helyeztek üzembe, akkor célszerű a záró vagy átlagos állományt megfigyelni. Idősoros vizsgálat esetén az összehasonlíthatóság biztosítása céljából, lehetőleg összehasonlító árakon kell az állóeszközöket értékelni. Az állóeszközök számbavételekor problémát jelent, hogy nettó vagy bruttó értékben történjen a mérés. Elvileg a nettó érték tükrözi az állóeszköz-állomány műszaki-gazdasági állapotát, de figyelembe kell venni, hogy a lineáris leírási kulcsok általában nem tükrözik az állóeszköz-állomány tényleges elhasználódási fokát, azaz torzítanak. A magyar iparvállalatoknál magas a nullára leírt, de a termelésben továbbra is résztvevő gépek aránya, melyek csak "eszmei" értékkel szerepelnek a vállalati nyilvántartásban, hatékonyságuk azonban megkérdőjelezhető. Figyelembe kell venni a kapacitás kihasználását is, továbbá azt, hogy az állóeszközök különböző évjáratúak, és így teljesítőképességük, műszaki jellemzőik is mások.

Az anyag és energiaráfordítást a műszaki sajátosságoknak megfelelően értékben és naturális mutatókkal lehet közelíteni.

A termelési függvény értelmezése

A termelési függvény azt mutatja meg, hogy a különbözőfajta termelési tényezők kombinációinak felhasználásával, milyen maximális kibocsátási szint érhető el.²⁵⁴

Az egyetlen kibocsátást létrehozó termelőtevékenység az alábbi explicit függvénnyel írható le:

$$y=f(x_1, x_2, \dots, x_k)$$

ahol:

y = a termelés, a kibocsátás, az output, mint eredményváltozó,

x_1, x_2, \dots, x_k = a termelési tényezők (személyi és tárgyi)²⁵⁵, a ráfordítások, az inputok, mint magyarázó változók.

Tételezzük fel, hogy a termelés becsült volumene [\hat{y}] két tényezőtől; az élómunka (x_1) és a holtmunka (x_2) tényezőktől függ.

$$\hat{y}=f(x_1, x_2)$$

Tételezzük fel továbbá, hogy a függvény (f) hatványkitevős²⁵⁶, vagyis az eredeti változók logaritmusai között lineáris kapcsolat van:

²⁵⁴ Barancsuk János [2008]: 135.

²⁵⁵ FP=Termelési tényezők (Factors of Production). A termelési folyamatban résztvevő, illetve abban felhasznált személyi és tárgyi tényezők, környezeti adottságok, a termeléshez szükséges emberi, szellemi, fizikai, természeti, anyagi, technológiai és pénzügyi feltételek. Neoklasszikus értelemben a termelési folyamatban közreműködő természeti tényező (föld) valamint a munka és a tőke.

²⁵⁶ A hatványkitevős függvényeket, mivel az eredményváltozó és a tényezőváltozók logaritmusai között van lineáris kapcsolat log-log modelleknek is hívják az ökonometriai szakirodalomban. Ld. Ramu Ramanathan [2003]: 519-520.

$$\hat{y} = b_0 x_1^{b_1} x_2^{b_2}$$

$$\ln \hat{y} = \ln b_0 + b_1 \ln x_1 + b_2 \ln x_2$$

A modellspecifikáció így még hiányos, mert a regressziós modell sztochasztikus, vagyis véletlen hatások is jelentkeznek, (olyan hatások, amelyeket a modellbe bevont változókkal, x_1 és x_2 nem tudunk magyarázni). A véletlen hatást a reziduális változó $[u]$ testesíti meg.

A modell ekkor²⁵⁷:

$$y = b_0 x_1^{b_1} x_2^{b_2} u$$

$$\ln y = \ln b_0 + b_1 \ln x_1 + b_2 \ln x_2 + \ln u$$

A termelési függvény jellemzői, átlag- és határmutatók két tényezőváltozós és hatványkitevős függvény esetén^{258 259}

A termelési függvény értelmezési tartománya és értékkészlete

A termelési függvény értelmezési tartományának és értékkészletének meghatározásakor figyelembe kell venni, hogy mind a termelési tényezők, mind a kibocsátás csak pozitív értékeket vehetnek fel. Mivel a kibocsátás létrehozásához az x_1 és x_2 tényezőkre egyaránt szükség van, ezek nulla értéket nem vehetnek fel. Ha valamelyik tényező értéke nulla lenne, akkor az eredeti függvényben nem szerepelne, s az egy tényező termelési függvénné alakulna át, ami rendszerint csak részleges vizsgálatokat tesz lehetővé. A termelési tényezők pozitív intervallumának felső határát elméletileg a rendelkezésre álló tényezők mennyisége, gyakorlatilag természetesen a rendelkezésre álló termelési tényezőknek az a mennyisége szabja meg, ameddig a termelés még hatékony. Feltételezzük, hogy a termelési függvény folytonos. Ez azt jelenti, hogy az x_1 és x_2 termelési tényezők kismértékű megváltoztatására az y kibocsátás is csak kismértékben változik.

Határtermelékenység

Feltételezzük továbbá, hogy a termelési függvény x_1 és x_2 szerinti elsőrendű parciális deriváltjai léteznek és a határtermelékenységek²⁶⁰ pozitívak, azaz

$$MP_{x_1} = \frac{dy}{dx_1} > 0$$

$$MP_{x_2} = \frac{dy}{dx_2} > 0$$

Az elsőrendű parciális derivált pozitív volta azt fejezi ki, hogy a termelési tényezők növekedésével a kibocsátás is nő. Ez azt jelenti, hogy a munkatényező vagy az állóeszköz-tényező rögzítésével mindig találunk olyan befektetési lehetőséget vagy foglalkoztatási lehetőséget, amellyel az állóeszköz-felhasználás vagy a munkaráfordítás utolsó növekménye is elősegíti a kibocsátás növekedését. Valamely termelési tényező határtermelékenysége tehát megmutatja, hogy mennyi többletkibocsátást hoz létre a felhasznált termelési tényező többletköltsége, miközben a termelési függvényben szereplő másik termelési tényező mennyisége változatlan marad. Természetesen ezekkel a tulajdonságokkal több termelési tényező esetén is rendelkezik a termelési függvény, itt csak az egyszerűség és a könnyebb értelmezhetőség miatt írtuk fel a két tényezőváltozós függvényt. A termelési függvények közgazdasági-matematikai vizsgálata lehetővé teszi számos, a termelési függvény tartalmával és formájával összefüggő mutató felírását. Ezekből a mutatókból fontos következtetéseket lehet levonni a tényezők közötti összefüggés jellegéről.

A termelési függvény alapján meghatározható átlagmutatók²⁶¹:

²⁵⁷ A regresszio.xls Excel parancsfájllal először ellenőrizni kell, hogy a modell eleget tesz-e a matematikai-statisztikai elméleti feltételeknek.

²⁵⁸ Ld.: Kristó Zoltán [1979]: 13-23.

²⁵⁹ Ld.: Sipos Béla [1982]: 30-43.

²⁶⁰ MP=Határtermelékenység (Marginal Productivity). Használatos fogalom még: határtermék: MP: Marginal Product.

Átlagtermelőkenységek (y/x_1 és y/x_2)²⁶²:

Valamely termelési tényező átlagtermelőkenységén a megfelelő termelési tényező egységére vonatkozó kibocsátást értjük. Általában az átlaghatékonyságok amelyek a t időpontbeli vagy i térbeli átlaghatékonyságokat jelölik, a termelési tényezők függvényei. Az x_1 és x_2 termelési tényezők átlagtermelőkenységei a munka termelőkenysége és az eszközhatékonyság:

$$\bar{y}_1 = \frac{y}{x_1}$$

$$\bar{y}_2 = \frac{y}{x_2}$$

Technológiai koefficiens (x_1/y és x_2/y):

Az átlagtermelőkenység reciprokát technológiai koefficiensnek nevezzük. Valamely tényező technológiai koefficiense az egységnyi kibocsátáshoz szükséges termelési tényező átlagos mennyiségét adja meg:

$$\frac{1}{\bar{y}_1} = \frac{x_1}{y}$$

$$\frac{1}{\bar{y}_2} = \frac{x_2}{y}$$

Technikai felszereltség ($F=x_2/x_1$):

Az alábbi (F) hányadost technikai felszereltségnek nevezzük.

$$F = \frac{x_2}{x_1}$$

Az előzőekből következik ugyanis, hogy a technikai felszereltség és az eszközhatékonyság szorzata a munka termelőkenységével egyenlő.

$$\bar{y}_1 = \frac{y}{x_1} = \frac{y}{x_2} \cdot \frac{x_2}{x_1}$$

Az átlagmutatókon kívül a határmutatók is lényeges szerepet játszanak a termelési függvények elemzésében.

Határtermelőkenység (dy/dx_1 és dy/dx_2):

Hatványkitevős két tényezőváltozós termelési függvény esetében a határtermelőkenységek:

$$MP_{x_1} = \frac{d\hat{y}}{dx_1} = b_0 x_2^{b_2} b_1 x_1^{b_1-1} = b_1 \frac{\hat{y}}{x_1} = b_1 \bar{y}_1$$

$$MP_{x_2} = \frac{d\hat{y}}{dx_2} = b_0 x_1^{b_1} b_2 x_2^{b_2-1} = b_2 \frac{\hat{y}}{x_2} = b_2 \bar{y}_2$$

Ugyanis:

$$\hat{y} = b_0 x_1^{b_1} x_2^{b_2}$$
$$x_1^{b_1-1} = x_1^{b_1} x_1^{-1} = x_1^{b_1} \frac{1}{x_1}$$

²⁶¹ A zárójelben feltüntettük azt, hogy az adott mutatót hogyan jelöltük: a C-Dtermelésifüggvényátlag és határmutatói.xls Excel parancsfájlban.

²⁶² Átlagtermelőkenység (Average Productivity) vagy átlagtermék: Average Product, jele AP.

Az élőmunka határtermelékenysége az élőmunka változó (x_1) kitevője (b_1) és az átlagtermelékenység (\bar{y}_1) szorzatával egyenlő. Hasonló módon a holtmunka határtermelékenysége a holtmunka változó (x_2) kitevője (b_2) és az eszközhatékonyság (\bar{y}_2) szorzatával egyenlő.

Az x_1 és x_2 termelési tényező parciális rugalmassága (elaszticitása) ²⁶³:

$$e_1 = \frac{d\hat{y}}{\hat{y}} \cdot \frac{dx_1}{x_1} = \frac{d\hat{y}}{d\hat{x}_1} \cdot \frac{\hat{y}}{\hat{x}_1} = \frac{d\hat{y}}{d\hat{x}_1} \cdot \bar{y}_1 = MP_{x_1} \cdot \bar{y}_1$$

$$e_2 = \frac{d\hat{y}}{\hat{y}} \cdot \frac{dx_2}{x_2} = \frac{d\hat{y}}{d\hat{x}_2} \cdot \frac{\hat{y}}{\hat{x}_2} = \frac{d\hat{y}}{d\hat{x}_2} \cdot \bar{y}_2 = MP_{x_2} \cdot \bar{y}_2$$

Látható, hogy a parciális elaszticitás a termelési tényezők határ- és átlagtermelékenysége hányadosával is kifejezhető. Az e_1 az élőmunka parciális elaszticitása azt mutatja meg, hogy hány százalékkal nő a kibocsátás az élőmunka-ráfordítás egy százalékos növekedése mellett, miközben a holtmunka-ráfordítás változatlan marad. Hasonló módon értelmezzük az e_2 -t, amely a holtmunka parciális elaszticitását jelöli. A parciális elaszticitás reciprokát parciális flexibilitásnak (részleges érzékenységnek) hívjuk. Megmutatja, hogy hány százalékkal nő pl. a munkaerő-felhasználás, ha a termelést egy százalékkal növeljük, miközben a holtmunka-ráfordítás változatlan marad. A gazdálkodó egység számára hasznos lehet megtudni, hogyan nő a termelés, ha minden termelési tényező felhasználását növeli. A parciális elaszticitás csak egy termelési tényező változása esetén adja meg a termelés relatív változását. Hosszabb távon azonban az összes termelési tényező változik.

A hatványkitevős regressziós függvény esetében a parciális rugalmassági együttható a b_1 és b_2 paraméterekkel egyenlő:

$$e_1 = (b_1 \bar{y}) / (\bar{y}) = b_1$$

$$e_2 = (b_2 \bar{y}) / (\bar{y}) = b_2$$

A volumenhozadék (e_v) ²⁶⁴

A volumenhozadék (volumenelaszticitás) megmutatja, hogy hány százalékkal nő a termelés a termelési tényezők egy százalékos növekedése mellett. Ez a mutató tehát az összes termelési tényező egyidejű arányos relatív növekedésének hatását méri. A volumenhozadék az egyes termelési tényezők parciális elaszticitásainak összegével egyenlő:

$$e_v = e_1 + e_2$$

A volumen hozadék e_v azt fejezi ki, hogy a termelés relatív növekedése nagyobb ($e_v > 1$), vagy kisebb ($e_v < 1$), mint a ráfordítások relatív növekedése, illetve azzal egyenlő-e $e_v = 1$. A gazdasági fejlődés szempontjából az előnyös, ha $e_v > 1$.

Ha a termelési függvény k -ad fokú homogén, akkor

$$\hat{y}_a = f(ax_1, ax_2) = a^k f(x_1, x_2)$$

vagyis a termelési tényezők a -szoros növekedése a kibocsátás a^k -szoros növekedését idézi elő. A k -t a homogenitás fokának nevezik. A termelési függvény segítségével vizsgálhatjuk a termelési tényezők arányainak, helyettesítésének és kölcsönhatásának alakulását is. Ha $k=1$, akkor a termelési tényezők k -szoros növekedése a kibocsátás szintjén szintén k -szoros növekedéshez vezet. ($\hat{y}_a = a\hat{y}$). Ha $k>1$, akkor a termelési tényezők k -szoros növekedése a kibocsátás szintjén több mint k -szoros növekedéshez vezet. ($\hat{y}_a > a\hat{y}$). Ha $0 < k < 1$, akkor a termelési tényezők k -szoros növekedése a kibocsátás szintjén kevesebb, mint k -szoros növekedéshez vezet. ($\hat{y}_a < a\hat{y}$).

A k -ad fok, vagyis a homogenitás foka azonos a volumen hozadékával.

Euler-tétel alapján: a termelési tényezők hozadékát határtermelékenységük és mennyiségük szorzata határozza meg:

²⁶³ ME=Marginal Elasticity (marginális rugalmasság).

²⁶⁴ Használatos kifejezések: volumenhozadék, volumenrugalmasság (volumenelaszticitás), skálahozadék, nívóhozadék, mérethozadék, Return to Scale.

$$\hat{y} = MP_{x_1} * x_1 + MP_{x_2} * x_2 / \hat{y}$$

$$1 = MP_{x_1} * \frac{x_1}{\hat{y}} + MP_{x_2} * \frac{x_2}{\hat{y}}$$

Az Euler tételből következik, hogy elsőfokú homogén termelési függvény esetén a tényezők parciális rugalmasságainak összege 1, ugyanis az átlagtermelékenységek és a határtermelékenységek szorzata egyenlő eggyel.

A helyettesítési határány ²⁶⁵ (s_1 és s_2)

A termelési függvényben szereplő tényezők bizonyos határokon belül helyettesíthetik egymást. A helyettesítési határány azt fejezi ki, hogy az egyik termelési tényező egységnyi csökkentése esetén mennyivel kell megnövelni a másik tényezőt ahhoz, hogy a kibocsátás változatlan szinten maradjon.

$$s_1 = \frac{d\hat{y}}{dx_1} : \frac{d\hat{y}}{dx_2} = b_1 \frac{\hat{y}}{x_1} / b_2 \frac{\hat{y}}{x_2} = (b_1/b_2) \frac{x_2}{x_1}$$

$$s_2 = \frac{d\hat{y}}{dx_2} : \frac{d\hat{y}}{dx_1} = (b_2/b_1) \frac{x_1}{x_2}$$

A helyettesítési határány a termelési tényezők határtermelékenységeinek a hányadosával határozható meg. Az s_1 -gyel jelöltük annak a szükséges beruházásnak a nagyságát, amely egységnyi munkaerő állóeszközökkel történő helyettesítéséhez szükséges a kibocsátás változatlansága mellett. Az élőmunka helyettesítésének így kifejezett nagysága egyenesen arányos az élőmunka határtermelékenységeivel, és fordítottan arányos a holtmunka határtermelékenységeivel. Az s_1 változása a vizsgált függvény esetében a technikai felszereltség változásától függ. Az s_2 -vel jelöltük a kibocsátás változatlansága esetén az egységnyi állóeszköz munkaerővel való kiváltásának nagyságát. Ez egyenesen arányos a holtmunka határtermelékenységeivel, és fordítottan arányos az élőmunka határtermelékenységeivel.

Helyettesítési rugalmasság ²⁶⁶

A helyettesítési rugalmasság a technikai felszereltség relatív változásának és a helyettesítési határány relatív változásának hányadosával egyenlő. Közgazdasági szempontból fontos ez a mutató, mert értékétől függ például az, hogy a beruházások milyen mértékben terjeszthetők ki gazdaságosan valamely termelési egységnél.

A helyettesítési rugalmasság képlete:

$$\sigma = \frac{d(x_2/x_1)}{x_2/x_1} : \frac{ds_1}{s_1}$$

Kedvező az, ha $\sigma > 1$, ugyanis ebben az esetben a helyettesítési határány 1%-os növekedése esetén a technikai felszereltség 1%-nál nagyobb ütemben nő. A helyettesítési rugalmasságra érvényes, hogy:

$$0 \leq \sigma \leq \infty$$

Hatványkitevős regressziós függvény esetében a $\sigma=1$.

Akcelerátorok.

Közvetlen akcelerátor (akc_1 és akc_2)

Valamely termelési tényező határtermelékenységeinek változását, valamint a termelési tényezők változásának a kapcsolatát jellemző mutatókat akcelerátoroknak nevezzük. A termelés akcelerátora az egyik termelési tényező határtermelékenységeinek a változása, ha ennek a termelési tényezőnek felhasznált mennyiségét egységgel növelik (közvetlen akcelerátor), vagy a másik termelési tényező felhasznált mennyi-

²⁶⁵ A helyettesítési határány jele MRS=Marginal rate of Substitution.

²⁶⁶ A helyettesítési rugalmasság angolul: ES=Elasticity of substitution. A fogalmat John Hicks (1932) és Joan Robinson (1933) használta először.

ségét egy egységgel növelik (keresztakcelerátor). Valamely termelési tényező közvetlen akceleratorát a termelésnek a megfelelő termelési tényező szerinti másodrendű parciális deriváltja adja meg. A vizsgált termelési függvényhez a következő közvetlen akceleratorok tartoznak:

$$\frac{d^2 \hat{y}}{dx_1^2} = (b_0 x_2^{b_2} b_1 x_1^{b_1-1})' = b_0 x_2^{b_2} b_1 (b_1 - 1) x_1^{b_1-2} = \frac{b_1 (b_1 - 1)}{x_1^2} \hat{y}$$

$$\frac{d^2 \hat{y}}{dx_2^2} = (b_0 x_1^{b_1} b_2 x_2^{b_2-1})' = b_0 x_1^{b_1} b_2 (b_2 - 1) x_2^{b_2-2} = \frac{b_2 (b_2 - 1)}{x_2^2} \hat{y}$$

A közvetlen akceleratorok előjelétől függően a határtermelékenységek viselkedésének három esete különböztethető meg:

1. A termelési tényező határtermelékenysége nő a termelési tényező növekedésével, ha a közvetlen akcelerator pozitív. Feltétele: $b_1 > 1$ illetve $b_2 > 1$
2. A termelési tényező határtermelékenysége változatlan marad a termelési tényező növekedésével, ha a közvetlen akcelerator nulla. Feltétele: $b_1 = 1$ illetve $b_2 = 1$
3. A termelési tényező határtermelékenysége csökken a termelési tényező növekedésével, ha a közvetlen akcelerator negatív. Feltétele: $0 < b_1 < 1$ illetve $0 < b_2 < 1$

Keresztakcelerátor (keresztakc)

A közvetett akcelerator, a keresztakcelerator, amely megadja az egyik termelési tényező határtermelékenységének változását a másik termelési tényező egységnyi változása esetén. A differenciálás szabályai szerint, először az x_1 , majd az x_2 szerint végezzük el a deriválást:

$$\frac{d\hat{y}}{dx_1} \frac{d\hat{y}}{dx_2} = b_0 x_2^{b_2} b_1 x_1^{b_1-1} b_2 x_2^{b_2-1} = \hat{y} \frac{b_1 b_2}{x_1 x_2}$$

ami azt jelenti, hogy az egyik termelési tényező változása ugyanúgy befolyásolja a másik termelési tényező határtermelékenységét, mint ez utóbbi termelési tényező változása az előbbinek a határtermelékenységét. A fenti összefüggés azt mutatja, hogy a két keresztakcelerator mindig egyenlő egymással, mivel a deriválás szabályai szerint mindegy, hogy a parciális deriválásokat milyen sorrendben végezzük el. Mind-egy tehát, hogy a termelési függvényt először az élömunka és azután a holtmunka szerint, vagy először a holtmunka és utána az élömunka szerint deriváljuk.

A keresztakceleratorok felhasználásával eldönthetjük, hogy két termelési tényező helyettesíti, vagy kiegészíti egymást a termelésben. Ha két termelési tényező keresztakceleratora pozitív, vagyis az egyik termelési tényező növelése növeli a másik határtermelékenységét, akkor a két tényező kiegészíti egymást. Ha két termelési tényező keresztakceleratora negatív, vagyis az egyik mennyiségének növelése csökkenti a másik határtermelékenységét, akkor a két tényező egymást helyettesíti. Ha két termelési tényező keresztakceleratora nulla, akkor függetlenek egymástól. A helyettesítési viszonyban levő tényezők jellegzetessége, hogy ezekből többféle mennyiségi kombináció használható fel ugyanazon termékmennyiség előállításához, azaz bizonyos keretek között az egyik a másikat helyettesítheti a termelésben. Az egyik tényező növelése, a másik csökkentését vonja maga után. Feltételeztük, hogy a termelési függvény parciális deriváltjai pozitívak, így a b_1 és b_2 együtthatók is pozitívak, tehát a keresztakceleratorok is pozitívak.

Hozadékok, az átlagtermelékenységek deriváltjai (hozadékk₁ és hozadékk₂)

Képezzük az (\hat{y}/x_1) x_1 szerinti deriváltját, felhasználva a törzfüggvényekre vonatkozó deriválási szabályt:

$$\frac{d\bar{y}_1}{dx_1} = \frac{d\left(\frac{\hat{y}}{x_1}\right)}{dx_1} = \frac{x_1 \frac{d\hat{y}}{dx_1} - \hat{y}}{x_1^2} = \frac{1}{x_1} \left[\frac{d\hat{y}}{dx_1} - \frac{\hat{y}}{x_1} \right] = \frac{1}{x_1} \left[\frac{d\hat{y}}{dx_1} - \bar{y}_1 \right]$$

A holtmunka hozadéka hasonló módon vezethető le:

$$\frac{d\bar{y}_2}{dx_2} = \frac{d\left(\frac{\hat{y}}{x_2}\right)}{dx_2} = \frac{x_2 \frac{d\hat{y}}{dx_2} - \hat{y}}{x_2^2} = \frac{1}{x_2} \left[\frac{d\hat{y}}{dx_2} - \frac{\hat{y}}{x_2} \right] = \frac{1}{x_2} \left[\frac{d\hat{y}}{dx_2} - \bar{y}_2 \right]$$

Látható, hogy a hozadék nagysága a határtermelékenység és az átlagtermelékenység viszonyától függ:

A termelési tényező hozadéka nő, ha a határtermelékenység nagyobb, mint az átlagtermelékenység, nem változik, ha a határtermelékenység és az átlagtermelékenység egyenlő és csökken, ha a határtermelékenység kisebb, mint az átlagtermelékenység.

A vizsgált hatványkitevős regressziós függvény esetében az élómunka és a holtmunka hozadéka:

$$\frac{d\bar{y}_1}{dx_1} = \frac{d\left(\frac{\hat{y}}{x_1}\right)}{dx_1} = \frac{1}{x_1} \left[\frac{d\hat{y}}{dx_1} - \frac{\hat{y}}{x_1} \right] = \frac{1}{x_1} \left[b_1 \frac{\hat{y}}{x_1} - \frac{\hat{y}}{x_1} \right] = \frac{1}{x_1} \frac{\hat{y}}{x_1} [b_1 - 1]$$

$$\frac{d\bar{y}_2}{dx_2} = \frac{d\left(\frac{\hat{y}}{x_2}\right)}{dx_2} = \frac{1}{x_2} \left[\frac{d\hat{y}}{dx_2} - \frac{\hat{y}}{x_2} \right] = \frac{1}{x_2} \left[b_2 \frac{\hat{y}}{x_2} - \frac{\hat{y}}{x_2} \right] = \frac{1}{x_2} \frac{\hat{y}}{x_2} [b_2 - 1]$$

A termelési tényező hozadéka nő, ha $b_1 > 1$ illetve $b_2 > 1$, nem változik, $b_1 = 1$ illetve $b_2 = 1$ és csökken, ha $0 < b_1 < 1$ illetve $0 < b_2 < 1$.

A termelési függvények becsülhetők kettőnél több magyarázó változó esetében is, az előzőekben bemutatott parciális átlag és határmutatók hasonló módon értelmezhetők:

$$y = b_0 x_1^{b_1} x_2^{b_2} x_3^{b_3} \dots x_k^{b_k} u$$

$$\ln y = \ln b_0 + b_1 \ln x_1 + b_2 \ln x_2 + b_3 \ln x_3 \dots b_k \ln x_k + \ln u$$

A nem lineáris regressziós függvények jellemzői

A trendszámításnál bemutatott módon (ebben az esetben $t=x$) linearizálhatók az alábbi nem lineáris regressziós függvények is, a függvények jellemzőit a 4-3. táblában foglaljuk²⁶⁷ össze.

²⁶⁷ Ld. Ramu Ramanathan (2003): 258. és Pintér József – Rappai Gábor (2007): 464.

4-3. tábla: Lineárisra visszavezethető nem lineáris regresszió függvények jellemzői

Függvénytípus	A függvény	A függvény meredeksége	A függvény elaszticitása
Lineáris (lin-lin)	$\hat{y} = b_0 + b_1 x$	b_1	$b_1 \frac{x}{\hat{y}}$
Hatványkitevős (log-log)	$\hat{y} = b_0 x^{b_1}$ $\ln \hat{y} = \ln b_0 + b_1 \ln x$	$b_1 \frac{\hat{y}}{x}$	b_1
Exponenciális (log-lin)	$\hat{y} = b_0 b_1^x$ $\ln \hat{y} = \ln b_0 + x \ln b_1$	$\hat{y} \ln b_1$	$x \ln b_1$
Féllogaritmikus (lin-log)	$\hat{y} = b_0 + b_1 \ln x$	$b_1 \frac{1}{x}$	$b_1 \frac{1}{\hat{y}}$
Hiperbolikus (reciprok)	$\hat{y} = b_0 + b_1 \frac{1}{x}$	$-b_1 \frac{1}{x^2}$	$-b_1 \frac{1}{x\hat{y}}$
Log- reciprok	$\ln \hat{y} = b_0 + b_1 \frac{1}{x}$	$-b_1 \frac{\hat{y}}{x^2}$	$-b_1 \frac{1}{x}$
Másodfokú para- bolikus (kvadratikus)	$\hat{y} = b_0 + b_1 x + b_2 x^2$	$b_1 + 2b_2 x$	$\frac{(b_1 + 2b_2 x)x}{\hat{y}}$
Log- kvadratikus	$\ln \hat{y} = b_0 + b_1 x + b_2 x^2$	$\hat{y}(b_1 + 2b_2 x)$	$x(b_1 + 2b_2 x)$

C-Dtermelésifüggvényátlagéshatármutatói.xls parancsfájl működése²⁶⁸

A program kiszámítja először a hatványkitevős (Cobb-Douglas) termelési (regressziós) függvény paramétereit a logaritmizált függvény alapján, majd közli a transzformált b_0 és e_v volumenhozadék értékeket, továbbá a becsült y vektort logaritmizált ($\ln y_{\text{becsült}}$: $\ln \hat{y}$) és természetes formában ($y_{\text{becsült}}$: \hat{y}) Ezt követően kiszámítja az átlag- és a határmutatókat, a megfigyeléseknek és a magyarázóváltozóknak (x_1 és x_2) megfelelően a következő sorrendben: átlagtermelékenységek, technológiai koefficiensek, technikai felszereltség, határtermelékenységek, parciális rugalmasságok, helyettesítési határányok, közvetlen akceleratorok, keresztakcelerator, hozadékok. Az ábrák munkalapon pedig ábrázolja is a kiszámított átlag- és határmutatókat. mutatókat.

A jelölések:

y/x_1 y/x_2 x_1/y x_2/y $F=x_2/x_1$ dy/dx_1 dy/dx_2 s_1 s_2 akc_1 akc_2 $keresztake$ $hozadékx_1$ $hozadékx_2$

A mintapélda a következő adatállomány (cipőgyár adatai) alapján készült:

²⁶⁸ A regresszio.xls Excel parancsfájllal először ellenőrizni kell, hogy a modell eleget tesz-e az elméleti statisztikai feltételeknek. Regressziocipohatványkitevos.xls parancsfájl tartalmazza a tesztelés eredményeit. A modell eleget tesz az elméleti feltételeknek, csak a multikollinearitás zavaró, viszont a b_1 és b_2 paraméterek szignifikánsan különböznek a nullától. Az empirikus szignifikancia szint a p értéke: 0,0000231. A b_1 és b_2 paraméter pozitív, tehát a határtermelékenység is pozitív, így a modell a közgazdasági feltételnek is eleget tesz, elemzésre és előrejelzésre felhasználható.

Évek (t)	y	x ₁	x ₂
1	1009	1787	1008
2	1099	2178	1111
3	1476	2627	1253
4	1488	2611	1218
5	1522	2740	1219
6	1568	2723	1208
7	1570	2746	1300
8	1701	2882	1647
9	1699	2506	1499
10	1667	2792	1741
11	1681	2681	2126
12	1594	2559	2544
13	1557	2483	2766
14	1560	2403	2654
15	1562	2146	2695
16	1589	2098	3702
17	1601	2032	4042
18	1588	2010	3930
19	1592	1949	4074
20	1538	2032	5278
21	1330	1844	4976
22	1208	1651	5314
23	1000	1568	5346

Ahol:

y = a termelés volumene [ezer pár]

x₁ = a fizikai foglalkozásuk teljesített munkaórái [ezer óra]

x₂ = a gépek berendezések amortizációja [millió Ft]

A számítások eredményei:

$$\ln b_0 =$$

$$-2,3616$$

$$b_0 =$$

$$0,0943$$

$$b_1 =$$

$$1,0416$$

$$b_2 =$$

$$0,2071$$

$$e_v = b_1 + b_2$$

$$1,2487$$

y/x_1	y/x_2	x_1/y	x_2/y	$F=x_2/x_1$
0,564633	1,000992	1,77106	0,999009	0,564074
0,504591	0,989199	1,981802	1,010919	0,510101
0,561858	1,177973	1,77981	0,848916	0,47697
0,569897	1,221675	1,754704	0,818548	0,466488
0,555474	1,248564	1,800263	0,80092	0,444891
0,575835	1,298013	1,736607	0,770408	0,443628
0,571741	1,207692	1,749045	0,828025	0,473416
0,590215	1,032787	1,694297	0,968254	0,571478
0,677973	1,133422	1,474985	0,882284	0,598164
0,597063	0,957496	1,674865	1,044391	0,623567
0,627005	0,790687	1,594884	1,264723	0,792988
0,6229	0,626572	1,605395	1,595985	0,994138
0,627064	0,562907	1,594733	1,776493	1,113975
0,649189	0,587792	1,540385	1,701282	1,104453
0,727866	0,579592	1,37388	1,725352	1,255825
0,757388	0,429227	1,320327	2,329767	1,764538
0,787894	0,396091	1,269207	2,524672	1,989173
0,79005	0,404071	1,265743	2,474811	1,955224
0,816829	0,390771	1,224246	2,559045	2,090303
0,75689	0,291398	1,321196	3,43173	2,597441
0,721258	0,267283	1,386466	3,741353	2,698482
0,731678	0,227324	1,366722	4,399007	3,218655
0,637755	0,187056	1,568	5,346	3,409439

dy/dx_1	dy/dx_2	s_1	s_2	akc_1	akc_2	keresztakc	hozadékk ₁	hozadékk ₂
0,588115	0,207296	2,837077	0,352475	1,37E-05	-0,000163	0,000120826	1,314E-05	-0,000787397
0,525576	0,204854	2,565614	0,38977	1E-05	-0,000146	9,79674E-05	9,63465E-06	-0,000705981
0,585223	0,243947	2,398978	0,416844	9,26E-06	-0,000154	9,67232E-05	8,89448E-06	-0,000745432
0,593597	0,252997	2,346257	0,426211	9,45E-06	-0,000165	0,000100926	9,07702E-06	-0,000795302
0,578575	0,258566	2,23763	0,446901	8,78E-06	-0,000168	9,82915E-05	8,43078E-06	-0,00081214
0,599783	0,268806	2,231282	0,448173	9,16E-06	-0,000176	0,000102822	8,79437E-06	-0,000851993
0,595518	0,250102	2,381102	0,419974	9,02E-06	-0,000153	9,48662E-05	8,6587E-06	-0,000736608
0,61476	0,21388	2,874318	0,347909	8,87E-06	-0,000103	7,72987E-05	8,51668E-06	-0,000497211
0,706167	0,234721	3,00854	0,332387	1,17E-05	-0,000124	9,75588E-05	1,12509E-05	-0,000599534
0,621893	0,198288	3,136307	0,318846	9,26E-06	-9,03E-05	7,39737E-05	8,89322E-06	-0,000436075
0,65308	0,163744	3,988427	0,250725	1,01E-05	-6,11E-05	6,36155E-05	9,72586E-06	-0,000294893
0,648804	0,129757	5,000138	0,199994	1,05E-05	-4,04E-05	5,28149E-05	1,01228E-05	-0,000195289
0,653142	0,116573	5,602871	0,17848	1,09E-05	-3,34E-05	4,89007E-05	1,05024E-05	-0,000161364
0,676186	0,121726	5,554978	0,180019	1,17E-05	-3,64E-05	5,27625E-05	1,1235E-05	-0,000175609
0,758135	0,120028	6,316322	0,15832	1,47E-05	-3,53E-05	5,8257E-05	1,41051E-05	-0,000170525
0,788885	0,088889	8,874954	0,112677	1,56E-05	-1,9E-05	4,41304E-05	1,5013E-05	-9,19337E-05
0,82066	0,082027	10,00479	0,099952	1,68E-05	-1,61E-05	4,20462E-05	1,61249E-05	-7,77002E-05
0,822905	0,083679	9,834033	0,101688	1,7E-05	-1,69E-05	4,33628E-05	1,6346E-05	-8,15247E-05
0,850798	0,080925	10,51343	0,095116	1,82E-05	-1,58E-05	4,3248E-05	1,7429E-05	-7,60544E-05
0,788366	0,060346	13,06414	0,076545	1,61E-05	-9,07E-06	3,09328E-05	1,54904E-05	-4,37765E-05
0,751253	0,055352	13,57234	0,073679	1,69E-05	-8,82E-06	3,12655E-05	1,62661E-05	-4,25907E-05
0,762106	0,047077	16,18861	0,061772	1,92E-05	-7,02E-06	2,96998E-05	1,84301E-05	-3,39193E-05
0,664277	0,038737	17,14818	0,058315	1,76E-05	-5,75E-06	2,57324E-05	1,69146E-05	-2,77438E-05

Gyakorló feladatok: C-D-termelési függvény átlag és határmutatói Excel parancsfájl.xls alkalmazása. Nem lineáris, de linearizálható regressziós függvények becslése regresszio.xls Excel parancsfájllal*

A könyv adatai: termelési függvény alkönyvtárban található az adatok.

1. A cipőgyár adatainak felhasználásával (cipogyar.xls) végezzen Backward regresszió számítást, a regresszio.xls Excel parancsfájllal úgy, hogy egy élőmunka és egy holtmunka változó maradjon a modellben. Az idősor első 17 adatával dolgozzon. Értékelje matematikai-statisztikai szempontok alapján a modellt, majd határozza meg az átlag és határmutatókat a C-D termelési függvény átlag és határmutatói Excel parancsfájl felhasználásával.
2. Bővítse az 1. feladatban definiált termelési függvény magyarázó változóit az anyagfelhasználás (d=felsőbőr felhasználás) változóval. A számítások elvégzése után értelmezze az átlag és határmutatókat.
3. Végezze el a számításokat a Cobb és Douglas által elvégzett számítások eredeti adatállományával. (Cobb-Douglaseredetiadatok1899-1922.xls)
4. A könyv adatai: Egy főre jutó GDP és kapcsolódó mutatók országoként alkönyvtárban található: összes mutató feldolgozva Excel fájl. A feladat felhasználva a lineáris és lineárisra visszavezethető, de nem lineáris regresszió-analízis módszereit az, hogy az országokénti egy főre jutó GDP \$-ban mért nagysága és az egyes kiválasztott gazdasági-demográfiai és társadalmi mutatók között milyen kapcsolatot modellezünk. Vizsgálja meg, hogy hogyan változik a gazdaság növekedésével a születéskor várható élettartam, az életminőség, a jólét, a fiatalok kilátásait jellemző humán fejlettség mutatója (HDI), a munkanélküliség aránya, illetve bizonyos termékek esetében a fogyasztás, így az egy főre jutó gépkocsik száma, a fajlagos energia, illetve a villamos energia felhasználás. A regresszio.xls parancsfájl használata előtt az adatok transzformációit el kell végezni. A grafikus ábra alapján kell eldönteni, hogy a függvénytípus váltás körülbelül mekkora GDP/fő értéknél következik be.

4.8 A CES-függvény becslése. (CES1.xls, CES2.xls CES3.xls)*

A hatványkitevős regressziós (Cobb - Douglas – típusú) termelési függvények széleskörű elterjedése, népszerűsége a függvény egyszerű és könnyen kezelhető matematikai alakjának tulajdonítható. Probléma viszont az, hogy a helyettesítési rugalmasság (σ) értéke előre rögzített, eggyel egyenlő érték lehet csak. Ezt a korlátot oldották fel a CES (állandó helyettesítési rugalmasságú) termelési függvény kidolgozói. A CES függvénynél a helyettesítési rugalmasság értékét a konkrét termelési függvény eredményeként határozzák meg. Ebben az esetben a helyettesítési rugalmasság állandó, de nem feltétlenül egyenlő eggyel, viszont értéke nem lehet negatív. A CES függvény öt becsült paraméterével sokoldalúbban írja le a fejlődést, mint a három paraméteres Cobb-Douglas termelési függvény. Ugyanakkor a CES függvény becslése bizonyos problémákat vet fel. Szükség van a munkaerő és az állóeszköz "ára" becslésére és ez sok bizonytalanságot visz a modellbe. Iterációs eljárással viszont a munkaerő és az állóeszköz "ára" ismerete nélkül is meghatározható az a függvény, ami mellett a többszörös determinációs együttható (R^2) értéke a legnagyobb.

A CES²⁶⁹-függvény becslése.^{270 271 272}

A CES-függvény képlete eredeti és logaritmizált formában:

$$y = h \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right]^{-\frac{e_v}{p}} e^u$$

$$\ln y = \ln h + e_v \left\{ \left[-\frac{1}{p} \right] \ln \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right] \right\} + u$$

$$p = \frac{1}{\sigma} - 1$$

$$\sigma = \frac{1}{p + 1}$$

$$\sigma = \frac{d(x_2 / x_1)}{x_2 / x_1} \cdot \frac{ds_1}{s_1}$$

²⁷⁰ Arrow, K. J. - Chenery, H. B. - Minhas, B. S. – Solow, R. M. [1961]: 225-250. A szerzők vezetéknévük kezdetűi alapján a CES - függvényt ACMS vagy SMAC – függvénynek is hívják. Az állandó helyettesítési rugalmasságú függvény angol rövidítése CES. (Constant Elasticity of Substitution).

²⁷¹ Pintér József: [1987]: 187-206.

²⁷² Mátyás Antal [1973]: 305-310.

$$s_1 = \frac{dy}{dx_1} : \frac{dy}{dx_2}$$

Ahol:

- $h > 0$; $0 < A_2 < 1$; $p \geq -1$; $\sigma \geq 0$,
- y = a termelés (kibocsátás, output),
- x_1 = a munkatényező (az élőmunka-ráfordítás, input)²⁷³,
- x_2 = az állóeszköz-tényező (a holtmunka-ráfordítás, input),
- e_v = a volumenhozadék, azt fejezi ki, hogy hány százalékkal nő a termelés (a kibocsátás) a termelési tényezők (x_1 és x_2) egy százalékos növekedése mellett.
- A_2 = az elosztási paraméter, a két termelési tényező részesedését fejezi ki a termelés [\hat{y}] létrehozásában, ahol az A_2 a holtmunka, míg $(1 - A_2)$ az élőmunka részesedése.
- σ = a helyettesítés rugalmassága (a technikai felszereltség relatív változásának és a helyettesítési határány relatív változásának a hányadosa)
- s_1 = a helyettesítési határány (határráta) a vizsgált termelési függvény esetében az élőmunka parciális deriváltjának és a holtmunka parciális deriváltjának a hányadosa.
- p = a helyettesítési paraméter, amely a helyettesítési rugalmasság (σ) transzformációja, ha $p=0$, akkor $\sigma=1$, ez a Cobb-Douglas függvény esete, ha: $-1 < p < 0$, akkor $\sigma > 1$, végül ha $0 < p < \infty$, akkor $\sigma < 1$.
- h = a semleges hatékonysági paraméter, amelynek változása meghatározott ráfordítások esetén azonos arányban változtatja a termelést.

Ha ismernénk az A_2 és a p értékét, akkor egy háromváltozós lineáris modellre vezetnénk vissza a CES-függvényt és akkor a becslés megoldható lenne. A CES-függvény paraméterei ugyanis csak a p és A_2 ismeretében becsülhetők meg:

$$\ln y = \ln h + e_v \left\{ \left[-\frac{1}{p} \right] \ln \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right] \right\} + u$$

Legyen:

$$W = \left[-\frac{1}{p} \right] \ln \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right]$$

A_2 és p ismeretében a két hiányzó ($\ln h$ illetve h és e_v) paraméter becsülhető:

$$\ln y = \ln h + e_v W + u$$

A CES becslésének három módszere ismert.

1. Módszer CES1.xls

Keresés gombra kattintva az **Opt W** vektorban megkeresi azt a helyettesítés rugalmasság, illetve az ebből számított helyettesítési paraméter értéket és az elosztási paramétereket, A_2 illetve az ebből számított A_1 értéket, amelyek esetében az illesztés a legjobb, tehát a többszörös determinációs együttható (R^2) a legnagyobb. Ha A_2 és p ismert, – amit az Excel parancsfájl kiszámít az előzőekben leírt módon – a hiányzó $\ln h$ és e_v paraméterek az alábbi regressziós függvényvel becsülhetők:

$$\ln y = \ln h + e_v W + u$$

Az optimális regressziót, tehát ahol az adott A_2 és p értékek mellett az R^2 a legnagyobb, elfogadva a CES-függvény paramétereit a logaritmizált egyenletből ($\ln y_{\text{becs}}$) megkapjuk.

²⁷³ A munkatényezőt a nemzetközi irodalomban általában L-vel, (Labor, munkaerő) a holtmunka tényezőt K-val (Capital assets, állótöke) jelölik. A feladatot regressziós modellként értelmezzük, ezért a megfigyelt magyarázóváltozókat x_i -vel jelöljük, az élőmunka indexe 1, a holtmunkáé 2.

$$\ln y = \ln h + e_v \left\{ \left[-\frac{1}{p} \right] \ln \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right] \right\} + u$$

$h = e^{\ln h}$ transzformációval az eredeti függvény (y_{becs}) felírható, mert p , A_2 és e_v paraméterek ismertek.

Az ábra munkalapon az eredeti és a becsült (CES) adatok ábráját is meg lehet tekinteni.

A feldolgozható legnagyobb adatállomány esetében a megfigyelések száma 500.

2 Módszer CES2.xls

A CES függvény kidolgozói, abból indultak ki, hogy a munka átlagos termelékenysége és a munkabér közötti empirikus összefüggés magában foglal egy hatványkitevős regressziós függvényt. Az alábbi összefüggésből kiindulva, mindkét oldalt logaritmizálva, a munkabér (v) kitevőjét c -vel jelölve, a következő segéd függvényhez jutottak.

$$\frac{y}{x_1} = b_1 v^c e^u$$

$$\ln \frac{y}{x_1} = \ln b_1 + c \ln v + u$$

Ahol:

v = az egységnyi munkaerő-tényező ára (pl. az összes munkabér [vagy bérköltség, vagy reálbér stb.] osztva a figyelembe vett munkatényező [x_1] mennyiségével)

A fenti összefüggésekben a b_1 és c paraméterek, a legkisebb négyzetek módszerével, az $\ln[y/x_1]$ és $\ln v$ idősorából meghatározhatók. Bizonyították, hogy c állandó volumenhozadék (e_v) esetén a helyettesítési rugalmassággal σ egyenlő. A fenti összefüggésekben látható, hogy a munkabér kitevője (c) azt fejezi ki, hogy a munkabér (v) 1 százalékos növekedésével a termelékenység c %-kal változik. A $c = \sigma$ összefüggést felhasználva a helyettesítési paraméter (p) is meghatározható. Az egyenlet becslését úgy végezhetjük el, hogy az A_2 értékét változtatjuk, s azt a változatot fogadjuk el, ahol az R^2 a legnagyobb és a regressziós modell, az elméleti feltételeknek eleget tesz. A p tehát ismert, így az A_2 értékét kell változtatni a 0 és 1 intervallumban, és mindegyik A_2 érték esetében meg kell határozni a többszörös determinációs együttható (R^2) értékét, felhasználva az alábbi, korábban már megismert determinációs együttható (R^2) értékét, felhasználva az alábbi, korábban már megismert összefüggéseket. Amelyik A_2 értéknél a legnagyobb a többszörös determinációs együttható értéke (R^2), azt a függvényt fogadjuk el. A számítás tehát hasonlít az 1. módszerhez, a különbség az, hogy csak az A_2 értékét változtatjuk, a p viszont már ismert. Az adatállomány viszont bonyolultabb, szükség van a munkaerő árára is, hogy a p értékét megbecsüljük.

$$y = h \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right]^{-\frac{e_v}{p}} e^u$$

$$\ln y = \ln h + e_v \left\{ \left[-\frac{1}{p} \right] \ln \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right] \right\} + u$$

A p ismeretében a hiányzó az A_2 változtatásával paraméterek megbecsülhetők. A ces2.xls fájl közli az $A_2 = 0,1; A_2 = 0,2; A_2 = 0,3; A_2 = 0,4; A_2 = 0,5; A_2 = 0,6; A_2 = 0,7; A_2 = 0,8; A_2 = 0,9$: és az optimális A_2 (az A_2 bármilyen értéket felvehet 0 és 1 között, és ahol az R^2 a legnagyobb) esetében a következő mutatókat: R^2 , $\ln h$, h , e_v .

A W a megadott A_2 ($A_2 = 0,1; A_2 = 0,2; A_2 = 0,3; \dots = 0,9$: és az optimális) értékek szerint változik:

$$W = \left[-\frac{1}{p} \right] \ln \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right]$$

A megadott A_2 felhasználásával a hiányzó paraméterek a már ismert regressziós függvénnyel becsülhetők.

$$\ln y = \ln h + e_v W + u$$

A program közli az A_2 megadott értékei ($A_2 = 0,1; \dots; A_2 = 0,9$:) közül a legjobb becslést (ahol az R^2 a legnagyobb) adó CES függvény logaritmizált ($\ln y_{\text{becs}}$) és transzformált, eredeti (y_{becs}) értékeit, az eredeti adatok és a CES-függvény alapján számított reziduum négyzet értékeit. A K oszlopban először a CES paramétereket sorolja fel.

A program közli az A_2 optimális értékét is, (ahol az R^2 a legnagyobb) és ezt *-gal jelöli, CES függvény logaritmizált ($\ln y_becs^*$) és transzformált, eredeti (y_becs^*) értékeit, az eredeti adatok és a CES-függvény alapján számított REZIDUUM NÉGYZET* értékeket. A K oszlopban a CES paramétereket * megjelöléssel másodikként sorolja fel.

A feldolgozható legnagyobb adatállomány esetében a megfigyelések száma 500, ezért a számított paraméterek egy része az 501. sorban található.

Az ábra munkalapon az eredeti, és a becsült (CES) adatok ábráját is meg lehet tekinteni.

3 Módszer CES3.xls

A módszer alkalmazásához szükség van a munkaerő (v) és az állóeszköz (r) árára is. Feltételezzük, hogy a CES függvény folytonos és létezik az elsőrendű és másodrendű parciális deriváltja. Az elsőrendű parciális derivált [a határtermelékenység] pozitív, ami azt jelenti, hogy a megfelelő termelési tényező növelésével a termelés is nő. Ez a feltétel a termelési függvény növekvő jellegére utal. A parciális deriváltakat meghatározva, bizonyítható, hogy a CES termelési függvényénél, az élőmunka és a holtmunka határtermelékenységének hányadosa, a helyettesítési határárány s_1 :

$$\frac{v}{r} = s_1 = \frac{1 - A_2}{A_2} \left(\frac{x_2}{x_1} \right)^{p+1}$$

A helyettesítési határárány – nyereség maximum esetén – a tényezőváltozók árának arányával v/r egyenlő.

Ahol a még nem ismert jelölés:

r = az egységnyi állóeszköz-tényező ára, ami pl. az állóeszköz állomány megtérülési rátája, annak a kamatlábnak a meghatározására szolgál, amely egyszeri megtérülést biztosít az élettartamon belül, ez a belső kamatláb. A belső kamatláb megmutatja, hogy mekkora az a kalkulatív kamatláb, amely mellett a beruházás egyszeri és a működés folyamatos költségei a bevételekből éppen egyszer térülnek meg az élettartam alatt. Helyettesíthetjük a belső kamatlábat az eszközhatékonyság vagy a kibocsátás (termelés) növekedési ütemével, vagy az állóeszközök nettó értékére jutó amortizációval, vagy az állóeszközök nettó értékére jutó amortizációnak az egyéb tiszta jövedelmi elemekkel növelt tömegével.

E feltételek mellett a CES termelési függvényre a következő segéd függvényt írhatjuk fel. A fenti egyenletet logaritmizálva:

$$\ln \frac{v}{r} = \ln \frac{1 - A_2}{A_2} + (p + 1) \ln \left(\frac{x_2}{x_1} \right)$$

az A_2 és a p meghatározható.

Legyen:

$$e_0 = \ln \frac{1 - A_2}{A_2}$$

$$e_1 = p + 1$$

Ekkor:

$$A_2 = \frac{1}{1 + e^{e_0}}$$

$$p = e_1 - 1$$

Az A_2 és p ismeretében a CES függvény becsülhető.

Látható, hogy a becsléshez szükség van a munkaerő és az állóeszköz ár adatokra is.

$$\ln y = \ln h + e_v \left\{ \left[-\frac{1}{p} \right] \ln \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right] \right\} + u$$

$$W = \left[-\frac{1}{p} \right] \ln \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right]$$

$$\ln y = \ln h + e_v W + u$$

$h = e^{\ln h}$ transzformációval az eredeti függvény felírható, mert p és A_2 az első segédfüggvényből, míg e_v a fenti, logaritmizált CES regressziós függvény becsléséből meghatározhatók:

$$\hat{y} = h \left[A_2 x_2^{-p} + (1 - A_2) x_1^{-p} \right]^{-\frac{e_v}{p}}$$

A program automatikusan minden paramétert kiszámít a fenti egyenletek felhasználásával.

A feldolgozható legnagyobb adatállomány esetében a megfigyelések száma 500.

Az ábra munkalapon az eredeti és a becült (CES) adatok ábráját is meg lehet tekinteni.

Gyakorló feladatok CES1.xls, CES2.xls és CES3.xls*

A cipőgyár adatai felhasználásával végezze el a CES függvény becslését a 3 módszerrel. Az állóeszköz ára sort az eszközhatékonysággal becsülje. Végezze el a CES függvények, mint regressziós függvények tesztelését, felhasználva az Excel Adatelemzés-Regresszió programját. (elsőrendű autokorreláció, homoszkedaszticitás Gleiser próba)

4.9 Logisztikus regressziós függvények*

A logisztikus regressziós függvények kezdetben konvex, később konkáv függvénygörbét írnak le. A következőkben a logisztikus, és az általánosított Richards-féle regressziós-függvényeket mutatjuk be. A becslés hasonló a trendfüggvényeknél bemutatott eljárással, csak itt a magyarázó változó (x) tetszőleges értéket vehet fel, míg a trendbecslésnél az $x=t=1,2,\dots,n$. A függvények tulajdonságai is azonosak, csak a t változó helyett az x keresztmetszeti adatokból álló magyarázó változó növekvő sorrendbe rendezett értékeit használjuk a becsléseknél. Ezért a tulajdonságok ismertetéstől itt eltekintünk, azok azonosak a trendfüggvényeknél leírtakkal. A fájlban a cellák színezése jelentőséggel bír: a halvány sárga cellák szabadon változtathatók, a zöld cellák az egyes paraméterek javasolt kezdeti értékeit adják meg, míg a fehér cellák számítási (rész)eredményeket tartalmaznak. A színezés alapján látható, hogy a fájl maximálisan 1000 hosszúságú idősor feldolgozására képes. Az induló paraméterek természetesen nem minden esetben adnak tökéletes javaslatot, így lehetőség van a paraméterek kézi vezérlésére is. Valamennyi munkalap tartalmaz olyan parancsgombokat, melyek a paraméterek finomhangolását végzik el (Opt. mind). A parancsgombok az Excel beépített Solver funkcióját hívják meg, a célfüggvény pedig az R^2 maximalizálása az egyes paraméterek iteratív változtatásával. Lehetőség van arra is, hogy a Solver a (kézzel, szakértői becslés alapján beállított) telítődési paraméter értékén ne változtasson, ekkor csupán a többi paraméter nagyságát fogja a program meghatározni (Opt. K nélkül). Az Excel beépített Solver csomagja nem képes minden esetben globális optimumot találni, így érdemes az illesztést több különböző, kézzel beállított indulóértékkel elvégezni. A Solver ebben az esetben a túlságosan nagy paramétert nem mozdítja el kezdeti értékéről. A megoldás az eredeti adatsor dimenziójának változtatása (pl. 1000-rel való osztás). A fájl

$$R^2 = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y}_t)^2} = 1 - \frac{SSE}{SST}$$

módon számít, ahol SSE (Sum of Squared Errors) a reziduumok négyzetösszege; SST (Sum of Squares Total) a teljes eltérés négyzetösszeg.

A Pearl–Reed-féle logisztikus regressziós függvény

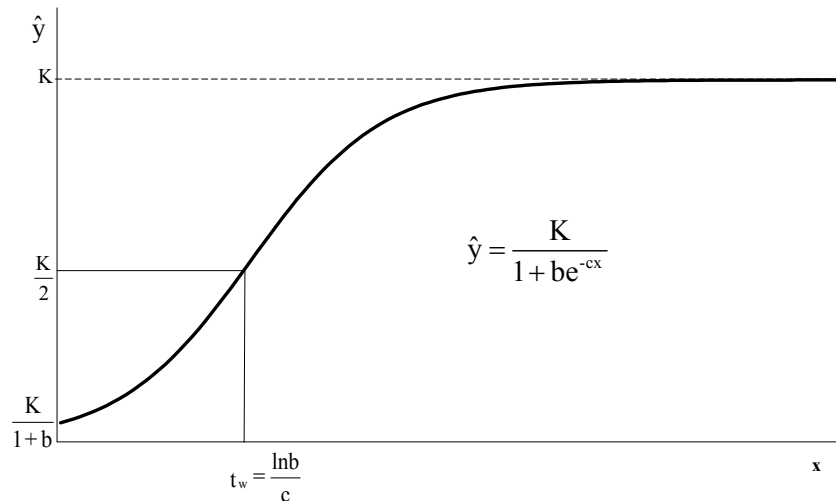
A függvényt leíró formula:

$$\hat{y} = \frac{K}{1 + b e^{-\epsilon x}}$$

ahol:

- \hat{y} – az eredményváltozó becslött értéke
- x – magyarázó változó
- ϵ – reziduális változó

- K – a telítettségi szint, ha $K > 0$;
- b – a helyzetparaméter, ha $b > 0$;
- c – növekedési sebességet jellemző paraméter, ha $c > 0$ logisztikus növekedésről, ha $c < 0$, akkor logisztikus csökkenésről van szó.



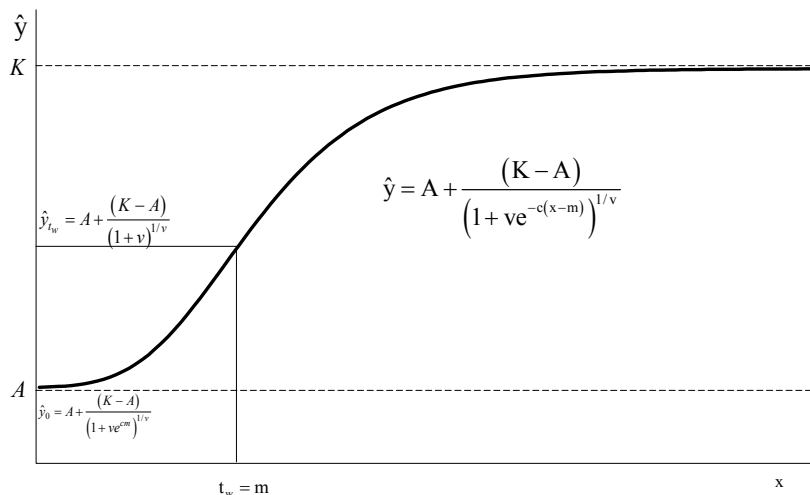
4-8. ábra: A Pearl-Reed-féle logisztikus regressziós függvény

Az általánosított Richards-féle logisztikus regressziós függvény

$$\hat{y} = A + \frac{(K - A)}{\left(1 + v e^{-c(x-m)}\right)^{1/v}}$$

Ahol

- y – az eredményváltozó becült értéke
- x – magyarázó változó
- ε – reziduális változó
- v – szabályozza az inflexiós pontban felvett függvényértéket, $v > 0$;
- c – növekedési sebességet jellemző paraméter, $c > 0$;
- m – az inflexiós pont, $m > 0$.



4-9. ábra: Az általánosított Richards-féle regressziós függvény

A többszörös determinációs együttható segítségével ellenőrizhetjük a modell egészét 5 %-os szignifikancia szinten a 4.1 pontban bemutatott módon, így ennek ismertetésétől itt eltekintünk. A p érték zöld értéke jelzi, hogy a modell létezik, a piros szám pedig a nullhipotézis elfogadását vagyis a modell elvetését jelenti.

A homoszkedaszticitás tesztelését a Glejser-próba alkalmazásával végzi el a program, a korábban bemutatott módon. A p-érték zöld értéke jelzi, hogy homoszkedasztikus a modell, a piros szám pedig a heteroszkedaszticitás jelenlétére utal.

4.10 A sztochasztikus kapcsolat elemzése, az asszociációs együtthatók Excel parancsfájl működése

A sztochasztikus kapcsolatok **típusai**²⁷⁴ (az ismérvek típusa szerint):

- **Asszociáció:** a minőségi (vagy területi) ismérvek közötti kapcsolat
- **Vegyes kapcsolat:** az ok szerepét minőségi (területi) az okozat szerepét mennyiségi ismérv tölti be
- **Korreláció:** a mennyiségi ismérvek közötti sztochasztikus kapcsolat

A sztochasztikus kapcsolatok **típusai** (a változók száma szerint):

- kétváltozós kapcsolat
- többváltozós (három-, négy-, stb.) kapcsolat

A minőségi (területi) ismérvek közötti sztochasztikus kapcsolatot **asszociáció**nak nevezzük. Ilyen például a nemhez való tartozás és a beosztás, az iskolai végzettség, vagy a szakképzettség és az alkalmazás minősége között lévő összefüggés.

Az asszociáció statisztikai elemzésének módszerei:

- a kombinációs (kontingencia) tábla elemzése viszonyszámokkal,
- kapcsolat-szorossági mérőszámok számítása.

A kétváltozós asszociáció vizsgálatakor a sokaság egységeit egyidejűleg, a két ismérv szerint kombinált módon csoportosítjuk. A kombinált csoportosítás eredménye egy **kétdimenziós kontingencia táblába** foglalható. Nevezzük a vizsgálatban szereplő két ismérvet „A”-nak (ok, X) és „B”-nek (okozat, Y). Rendelkezzon az „A” ismérv s, a „B” ismérv o ismérvváltozattal. Bármely egyed egyidejűleg mindkét ismérv egy-egy konkrét változatával jellemezhető, vagyis valamely megfigyelésünk – általánosságban – az (a, b) ismérvváltozat-párral írható le (a=1,2,...,s; b=1,2,...,o).

4-4. tábla: A kétdimenziós kontingencia tábla általános sémája

„A” ismérv változatai: f_{ab}	„B” ismérv változatai					Összesen
	B_1	...	B_b	...	B_o	
A_1	f_{11}	...	f_{1b}	...	f_{1o}	S_1
A_2	f_{21}	...	f_{2b}	...	f_{2o}	S_2
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
A_a	f_{a1}	...	f_{ab}	...	f_{ao}	S_a
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
A_s	f_{s1}	...	f_{sb}	...	f_{so}	S_s
Összesen	O_1	...	O_b	...	O_o	n

Ahol:

- n az összes elemszám,
- f_{ab} az „A” ismérv a-dik és a „B” ismérv b-edik változatához rendelt gyakorisága,
- S_a az a-dik sor (az „A” ismérv első változatához tartozó) gyakoriságának összege,
- O_b az b-dik oszlop (a „B” ismérv első változatához tartozó) gyakoriságának összege.

Belátható az alábbi összefüggés:

²⁷⁴ Pintér József – Rappai Gábor (szerkesztő) [2007]: Statisztika. 227-235.

$$\sum_{a=1}^s \sum_{b=1}^o f_{ab} = S_1 + S_2 + \dots + S_s = O_1 + O_2 + \dots + O_o = n$$

A sorok és oszlopok összegeit **peremgyakoriságok**nak is hívják. A tábla belsejében a kétszeres csoportosítás eredményeként keletkezett gyakoriságok, a tábla szélein külön-külön, az egyik, illetve a másik ismérv szerinti csoportosítással nyert gyakoriságok találhatóak, míg a tábla utolsó rovatában a gyakoriságok összege szerepel, mely a sokaság elemszámát mutatja.

Az adatokat az f_{ab} sárga mezőbe lehet beírni, a maximális méret $s=15$ és $o=15$, tehát 15×15 mátrix, amelyek lehet nem szimmetrikus is. Az összesen értékeket (S_a és O_b és n) a program kiszámítja.

A következő két tábla amit a program kiszámít, a függetlenség esetére feltételezett gyakoriságok (f_{ab}^*) táblája, majd a χ^2 -értékek számítási táblája. Ehhez a következő lépésekben jutunk el.

A fenti kontingencia tábla belsejében lévő gyakoriságok elhelyezkedése már szolgáltat bizonyos információkat a sztochasztikus kapcsolat meglétéről. Megállapítható ugyanis, hogy a két ismérv mely változatai járnak gyakrabban együtt („vonzzák egymást”), és melyek fordulnak elő ritkábban („taszítják egymást”). Ugyanez még jobban látszik, ha a 4.4 tábla gyakoriságaiból **megoszlási viszonyszámokat** számítunk.

4-5. tábla: Megoszlási viszonyszámokat tartalmazó kétdimenziós kontingencia tábla sémája

„A” ismérv vált- ozatai	„B” ismérv változatai					Ösz- zesen
	B_1	...	B_b	...	B_o	
A_1	p_{11}	...	p_{1b}	...	p_{1o}	S_1/n
A_2	p_{21}	...	p_{2b}	...	p_{2o}	S_2/n
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
A_a	p_{a1}	...	p_{ab}	...	p_{ao}	S_a/n
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
A_s	p_{s1}	...	p_{sb}	...	p_{so}	S_s/n
Ösz- zesen	O_1/n	...	O_b/n	...	O_o/n	1

$$p_{ab} = \frac{f_{ab}}{n}$$

$$\sum_{b=1}^o p_{ab} = O_b/n \quad \sum_{a=1}^s p_{ab} = S_a/n$$

$$\sum_{a=1}^s \sum_{b=1}^o p_{ab} = \sum_{a=1}^s S_a/n = \sum_{b=1}^o O_b/n = 1$$

Egy kétdimenziós kombinációs tábla (4.4 tábla) gyakoriságaiból más módon, ún. feltételes megoszlási viszonyszámokat is számolhatunk.

4-6. tábla: Feltételes megoszlási viszonyszámok (oszlop szerint)

„A” ismérv vált- ozatai	„B” ismérv változatai					Ösz- szesen
	B_1	...	B_b	...	B_o	
A_1	f_{11}/O_1	...	f_{1b}/O_b	...	f_{1o}/O_o	S_1/n
A_2	f_{21}/O_1	...	f_{2b}/O_b	...	f_{2o}/O_o	S_2/n
⋮	⋮	...	⋮	...	⋮	⋮
A_a	f_{a1}/O_1	...	f_{ab}/O_b	...	f_{ao}/O_o	S_a/n
⋮	⋮	...	⋮	...	⋮	⋮
A_s	f_{s1}/O_1	...	f_{sb}/O_b	...	f_{so}/O_o	S_s/n
Ösz- szesen	1	...	1	...	1	1

Feltételes megoszlási viszonyszámok (sor szerint)

„A” ismérv vált- ozatai	„B” ismérv változatai					Ösz- szesen
	B_1	...	B_b	...	B_o	
A_1	f_{11}/S_1	...	f_{1b}/S_1	...	f_{1o}/S_1	1
A_2	f_{21}/S_2	...	f_{2b}/S_2	...	f_{2o}/S_2	1
⋮	⋮	...	⋮	...	⋮	⋮
A_a	f_{a1}/S_a	...	f_{ab}/S_a	...	f_{ao}/S_a	1
⋮	⋮	...	⋮	...	⋮	⋮
A_s	f_{s1}/S_s	...	f_{sb}/S_s	...	f_{so}/S_s	1
Ösz- szesen	O_1/n	...	O_b/n	...	O_o/n	1

A fenti viszonyszámok segítségével már jellemezhetőek az ismérvek között kapcsolatok. Amennyiben a megoszlási viszonyszámok homogének, egyenletes eloszlásúak, vagy a feltételes megoszlási viszonyszámok soronként, illetve oszloponként azonosak, akkor nincs számottevő kapcsolat az ismérvek között. A sztochasztikus kapcsolat vizsgálatához a legcélravezetőbb az, ha a – feltételezésünk szerint – ok szerepét játszó ismérv szerinti csoportokban az okozat szerinti megoszlást vizsgáljuk. Ha az oknak tekinthető ismérv változatait a tábla oldalrovatában, az okozatnak tekinthetőét a fejtáblában helyezük el, akkor a tábla sorai fogják mutatni az ok szerinti csoportokban, az okozat szerinti megoszlást.

A sztochasztikus kapcsolatot a tábla utolsó, összesen sorában, illetve a felette lévő sorokban szereplő megoszlások összehasonlításával mutathatjuk ki. Ha az ok szerinti csoportokban (a tábla soraiban) számított megoszlások különböznek egymástól, és ilyenkor természetesen az egész sokaságra jellemző (az ösz-

szesen sorban szereplő) megoszlástól, akkor megállapítható a sztochasztikus kapcsolat megléte. Ha a tábla minden sorában (és így az összesen sorban is) ugyanolyan lenne a megoszlás, az a két ismerv **függetlenségét** jelezné. Ha a táblának csak egyik átlójában találnánk nullától különböző gyakoriságot, és így minden sorban csak egy, 1 értékű, illetve 100% -os megoszlási viszonyszámot, akkor a két ismerv *függhetlenségű kapcsolatban* lenne. Természetesen ez csak olyan táblával reprezentált összefüggések esetén lehetséges, amikor a két ismerv változatainak a száma megegyezik, azaz kvadrátikus a tábla (például 2×2-es, 3×3-as).

Természetesen az elsőként bemutatott 4.4 tábla logikájának megfelelően is elemezhetjük a sztochasztikus kapcsolat meglétét. A megoszlási viszonyszámok segítségével tulajdonképpen az alábbi azonosságokat vizsgáljuk:

$$\frac{f_{ab}}{O_b} = \frac{S_a}{n}$$

illetve

$$\frac{f_{ab}}{S_a} = \frac{O_b}{n}$$

Amennyiben a fenti összefüggések teljesülnek, a két ismerv *függetlennek* tekinthető. Ezzel eljutottunk tulajdonképpen az asszociációs kapcsolat mérésének alapgondolatához. A valószínűség-elméletből ismeretes az a megfogalmazás, amely szerint a **függetlenség feltétele**, hogy a feltételes valószínűség legyen egyenlő a feltétel nélküli valószínűséggel. Mindezt reprezentálják a fenti azonosságok. Az ismérvek függetlenségét megközelíthetjük abból a közismert valószínűségelméleti összefüggésből is, amely szerint két esemény független, ha együttes bekövetkezési valószínűségük megegyezik a két esemény valószínűségének szorzatával:

$$P(A \cap B) = P(A)P(B)$$

Mindez empirikus statisztikai jelölésekkel is leírható:

$$\frac{f_{ab}}{n} = \frac{O_b}{n} \times \frac{S_a}{n}$$

A továbbiakban azt a gyakoriságot keressük, amely a fenti feltételeknek megfelel. Jelöljük f_{ab}^* -gal a **függetlenség esetén feltételezett gyakoriságot**. A fenti összefüggésből kiindulva az ismérvek függetlenségének esetére feltételezett gyakoriságokat számíthatjuk ki az alábbi módon:

$$f_{ab}^* = n \cdot \frac{O_b}{n} \cdot \frac{S_a}{n} = n P_b^{(o)} P_a^{(s)} = \frac{O_b S_a}{n}$$

ahol:

$$\frac{O_b}{n} = P_b^{(o)} \quad \text{és} \quad \frac{S_a}{n} = P_a^{(s)}$$

a két ismerv szerint külön-külön számított (a tábla peremén szereplő) megoszlási viszonyszámok.

A feltételezett gyakoriságok számítása tehát azt jelenti, hogy a sokaságot a peremeloszlások alapján osztjuk szét. Ha a táblát a feltételezett gyakoriságokkal töltjük ki, minden sor megoszlása ugyanolyan lesz, ami megfelel a két ismerv függetlenségének. A sztochasztikus kapcsolat létezését jelzi tehát az, ha a ténylegesen megfigyelt és a függetlenség esetére feltételezett gyakoriságok nem egyeznek meg. Összehasonlításukat a **négyzetes kontingencia** mutatójával az ún. χ^2 -értékkel végezhetjük el.

Képlete:

$$\chi^2 = \sum_{a=1}^s \sum_{b=1}^o \frac{(f_{ab} - f_{ab}^*)^2}{f_{ab}^*}$$

Ha a tényleges és a feltételezett gyakoriságok megegyeznek, azaz az ismérvek *függetlensége* esetén, $\chi^2 = 0$.

A χ^2 -értéket felhasználva a kapcsolat szorossága a **Cramer-féle asszociációs együtthatóval** mérhető.

Képlete:

$$C = \sqrt{\frac{\chi^2}{n \times \min[s-1, o-1]}}$$

A min jelöli, hogy az s és o közül a kisebbiket kell választanunk, pl. s=3, o=2, akkor min=2.

A Cramer-együttható értéke 0, ha a két ismerv *független* egymástól, és 1 értéket vesz fel, ha *függvényszerű* a kapcsolat. Az együttható 0 és 1 közötti értékeit hagyományosan három részre osztjuk: *gyenge*, *közepes* és *szoros* kapcsolatot mutató tartományra. (Általában a 0,3 és 0,7 közötti értéket tekintjük közepes értéknek, ám meg kell jegyeznünk, hogy a felosztás inkább csak tájékoztató jellegű, a mérőszám megítélése függ pl. a sokaság nagyságától, vagy a tábla méretétől is.) Az együttható előjele (hiszen gyökvonással keletkezett) mindig pozitív, tárgyi értelme nincs.

A másik hasonló asszociációs mutatószám a **Csuprov-együttható**:

$$\text{Csuprov-együttható} = \sqrt{\frac{\chi^2}{n \sqrt{(s-1)(o-1)}}$$

Értelmezése azonos a Cramer-együtthatóéval.

A Csuprov-együttható a következő intervallumban helyezkedik el:

$$0 \leq \text{Csuprov-együttható} \leq \left(\frac{s-1}{o-1}\right)^{1/4} \quad \text{ahol : } s \leq o$$

A Csuprov-együttható értéke az ismérvek függetlensége esetén nulla lesz. Maximális értéke egy, amit csak akkor vehet fel, ha az ismérvek változatainak a száma azonos, tehát s=o. Ebben az esetben minden oszlopban és minden sorban csak egy helyen találunk gyakoriságot (f_{ab}). A mérőszám értéke minél közelebb van a maximális értékhez, annál erősebb, intenzívebb az asszociáció.

Ha az ismérvek változatainak a száma különböző, a Csuprov-együttható csak akkor veszi fel a maximális értéket ha vagy minden sorban, vagy minden oszlopban csak egy helyen találunk gyakoriságot. A Csuprov-együttható ebben az esetben:

$$\left(\frac{s-1}{o-1}\right)^{1/4} \quad \text{ahol : } s \leq o$$

vagy:

$$\left(\frac{o-1}{s-1}\right)^{1/4} \quad \text{ahol } o \leq s :$$

A Cramer-féle asszociációs együtthatót inkább akkor használjuk, amikor az asszociációs összefüggések térbeni vagy időbeni összehasonlításánál valamelyik ismerv változatainak számában eltérés mutatkozik. A Cramer-féle asszociációs együttható tulajdonképpen a Csuprov-együtthatót annak maximális értékéhez viszonyítja, vagyis a $\text{Csuprov} \leq \text{Cramer}$. Ha az ismérvek változatainak a száma azonos, tehát s=o, akkor a Cramer- és a Csuprov- együttható értéke azonos lesz.

$$\text{Cramer - együttható} = \frac{\text{Csuprov - együttható}}{\left(\frac{s-1}{o-1}\right)^{1/4}}$$

Vagy:

$$\text{Cramer - együttható} = \frac{\text{Csuprov - együttható}}{\left(\frac{o-1}{s-1}\right)^{1/4}}$$

Továbbá:

$$\text{Cramer - együttható} * \left(\frac{s-1}{o-1}\right)^{1/4} = \text{Csuprov - együttható}$$

Vagy:

$$\text{Cramer - együttható} * \left(\frac{o-1}{s-1}\right)^{1/4} = \text{Csuprov - együttható}$$

A függetlenség-vizsgálat módszere

A kapcsolat szignifikáns voltára vonatkozó feltevésünket a **függetlenség-vizsgálat** módszerével válaszolhatjuk meg. A teszteljárás a négyzetes kontingencia χ^2 mutatójára épül. A vizsgálat során alkalmazott hipotézis-rendszer:

$$H_0: \chi^2 = 0$$

$$H_1: \chi^2 > 0$$

Tehát a nullhipotézis szerint az ismérvek függetlenek, míg az alternatív hipotézis elfogadása a sztochasztikus kapcsolat szignifikáns voltát jelenti.

A nullhipotézis teljesülése esetén a négyzetes kontingencia χ^2 -eloszlást követ, $(s-1)(o-1)$ szabadságfokkal. Vagyis:

$$\chi_{empirikus}^2 = \sum_{a=1}^s \sum_{b=1}^o \frac{(f_{ab} - f_{ab}^*)^2}{f_{ab}^*}$$

Ennek ismeretében a hipotézisellenőrzés könnyen elvégezhető:

1) Meghatározzuk az empirikus χ^2 értéket ($\chi_{empirikus}^2$).

2) Kikeressük a χ^2 -eloszlás választott szignifikancia-színjéhez (alapeset 5%, 0,0500) tartozó kritikus értéket, ${}_{(s-1)(o-1)}\chi_{kritikus}^2$ ahol a szabadságfok: $(s-1)(o-1)$ vagy meghatározzuk a p-értéket, az empirikus szignifikancia értéket.

3) Amennyiben az általunk számított érték kisebb mint a kritikus érték,

$$\chi_{empirikus}^2 < {}_{(s-1)(o-1)}\chi_{kritikus}^2$$

úgy a nullhipotézist, ellenkező esetben:

$$\chi_{empirikus}^2 > {}_{(s-1)(o-1)}\chi_{kritikus}^2$$

az alternatív hipotézist fogadjuk el.

A p-érték, az empirikus szignifikancia érték azt a legkisebb valószínűséget mutatja, ami mellett még a H_0 nullhipotézist elutasíthatjuk. Ha a teszteléshez ha a p-értékeket használjuk, akkor: ha a p-érték magasabb, mint amit megengedtünk (pl. 0,05), akkor elfogadjuk azt a nullhipotézist, vagyis azt, hogy az ismérvek függetlenek, ha kisebb a p-érték mint pl. 0,05, akkor az alternatív hipotézist fogadjuk el, vagyis a sztochasztikus kapcsolat szignifikáns.

A Yule-mutató.

Az első kísérletek az asszociációs kapcsolat szorosságának meghatározására abból az egyszerűsítésből indultak ki, hogy a vizsgálandó mindkét ismerv alternatív, azaz csak két ismervváltozattal rendelkezik. A változók minden esetben alternatívává transzformálhatók az általunk legfontosabbnak tartott ismervváltozat kiválasztásával és a többi változat összevonásával; ám ez lényeges információk elvesztéséhez vezet. (Példa lehet erre a tantárgyak minősítése: az öt fokozatú skála (jeles, jó, közepes, elégséges, elégtelen - alternatívává alakítható - megfelelt, nem felelt meg.)

Alternatív ismérveket tartalmazó kombinációs tábla sémája

A/B	B ₁	B ₂	
A ₁	f ₁₁	f ₁₂	S _{1.}
A ₂	f ₂₁	f ₂₂	S _{2.}
	O _{.1}	O _{.2}	n

A kétállapotú (bináris) változók közötti kapcsolat elemzése során használható első mutatószám kidolgozása **G. U. Yule** nevéhez fűződik. Mivel függetlenség esetén:

Ha nincs kapcsolat, akkor

$$\frac{f_{11}}{f_{12}} = \frac{f_{21}}{f_{22}},$$

azaz

$$f_{11}f_{22} - f_{12}f_{21} = 0,$$

ha van kapcsolat, akkor

$$f_{11}f_{22} - f_{12}f_{21} \neq 0$$

A Yule-féle asszociációs együttható:

$$\text{Yule - mutató} = \frac{f_{11}f_{22} - f_{12}f_{21}}{f_{11}f_{22} + f_{12}f_{21}}$$

Y értéke minél közelebb van a 0-hoz, a kapcsolat annál lazább, gyengébb és minél közelebb van az 1-hez, annál szorosabb, erősebb. Ha a kapcsolat olyan, hogy az A_1 tulajdonsággal inkább B_1 tulajdonság és az A_2 -vel inkább B_2 jár együtt, az Y értéke "+" lesz, ellenkező esetben "-" lesz. A Yule-féle asszociációs együttható előjelére nem fordítunk külön figyelmet, ugyanis előjele attól függ, hogy a vizsgált ismérvek változatait milyen sorrendben írtuk fel, azaz melyik lesz az A- illetve B- változat, ezért abszolút értékének nagysága a kapcsolat erőségét jellemzi.

A bemutató példa:

A szakmunkásképzőt és szakiskolát, illetve középiskolát végzett magyarországi foglalkoztatottak megoszlása nemek szerint:

4-7. tábla: A foglalkoztatottak megoszlása iskolatípus és nemek szerint (ezer fő)

Nem f_{ab}	Szaktanács- képző és szakiskola	Középiskola	Összesen
Férfi	876	561	1 437
Nő	362	704	1 066
Összesen	1 238	1 265	2 503

Forrás: Magyar Statisztikai Évkönyv 2000

A tábla adatai alapján látható, hogy az iskolatípus és a nemhez való tartozás között van összefüggés, mivel a férfi nemhez való tartozás és a szakmunkásképző és szakiskola ismérvváltozat, illetve a női nemhez való tartozás és a középiskola ismérvváltozat járnak gyakrabban együtt. Ha kiszámítjuk a nemhez való tartozás (az ok) szerinti csoportokon belül az iskolatípus (az okozat) szerinti megoszlást, akkor a következő táblában szereplő megoszlási viszonyszámok megerősítik az előző megállapítást.

4-8. tábla: A foglalkoztatottak megoszlása iskolatípus és nemek szerint (%)

Nem	Szaktanács- képző és szakiskola	Középisko- la	Összesen
Férfi	61,0	39,0	100,0
Nő	34,0	66,0	100,0
Összesen	49,5	50,5	100,0

A férfiak és nők csoportjában az iskolatípus szerinti megoszlási viszonyszámok jelentősen különböznek egymástól és az összes foglalkoztatottra jellemző megoszlási viszonyszámától. Míg a férfiak 39%-a végzett középiskolát, a nőknél ez az arány 66%, az összes foglalkoztatottnál pedig, az előbbieket valamilyen köztes értéke, 50,5%. A kapcsolat szorosságának mérésére a Cramer-együtthatót használjuk, ezért a következő táblában az ismérvek függetlensége esetére feltételezett gyakoriságokat számítjuk ki.

4-9. tábla: A függetlenség esetére feltételezett gyakoriságok, f_{ab}^* (ezer fő)

Nem	Szaktanácsképző és szakiskola	Középiskola	Összesen
Férfi	711	726	1 437
Nő	527	539	1 066
Összesen	1 238	1 265	2 503

Példaként a szaktanácsképzőt és szakiskolát végzett férfiak feltételezett gyakorisága illetve a középiskolát végzett nők feltételezett gyakorisága:

$$f_{11}^* = n \frac{O_1 S_1}{n} = 2503 \times \frac{1238}{2503} \times \frac{1437}{2503} = 711$$

$$f_{12}^* = n \frac{O_2 S_1}{n} = 2503 \times \frac{1265}{2503} \times \frac{1437}{2503} = 726$$

$$f_{21}^* = n \frac{O_1 S_2}{n} = 2503 \times \frac{1238}{2503} \times \frac{1066}{2503} = 527$$

$$f_{22}^* = n \frac{O_2 S_2}{n} = 2503 \times \frac{1265}{2503} \times \frac{1066}{2503} = 539$$

Vegyük észre, hogy – korábbi megállapításainkból következő módon – a peremgyakoriságok értéke a tényleges, illetve függetlenség esetére feltételezett kontingencia táblában azonos.

A következő táblában a χ^2 -érték számítását közöljük:

4-10. tábla: A χ^2 érték számítása

Megnevezés	f_{ab}	f_{ab}^*	$(f_{ab} - f_{ab}^*)$	$\frac{(f_{ab} - f_{ab}^*)^2}{f_{ab}^*}$
Férfi-szaktanácsk.	876	711	165	38,42
Férfi-középisk.	561	726	-165	37,6
Nő-szaktanácsk.	362	527	-165	51,79
Nő-középisk.	704	539	165	50,69
Összesen	2 503	2503	-	178,5

A Cramer-együttható (C) értéke:

$$C = \sqrt{\frac{\chi^2}{n \times \min[s-1, o-1]}} = \sqrt{\frac{178,5}{2503 \times (2-1)}} = 0,267 \approx 0,3 \quad \text{ahol: } s = o = 2$$

A szaktanácsképzőt és szakiskolát, illetve középiskolát végzett magyarországi foglalkoztatottak nemhez való tartozása és a végzettségének megfelelő iskola típusa között közepesen gyengébb sztochasztikus kapcsolatot számszerűsítettünk.

A Csuprov-együttható értéke ebben az esetben megegyezik a Cramer-együttható értékével, mivel $s=o=2$:

$$\text{Csuprov-együttható} = \sqrt{\frac{\chi^2}{n\sqrt{(s-1)(o-1)}}} = \sqrt{\frac{178,5}{2503\sqrt{(2-1)(2-1)}}} = 0,267 \approx 0,3$$

A függetlenség vizsgálata:

$$\chi^2_{\text{empirikus}} = \sum_{a=1}^s \sum_{b=1}^o \frac{(f_{ab} - f_{ab}^*)^2}{f_{ab}^*} = 178,5$$

$${}_{(2-1)(2-1)}\chi_{0,05} = 3,841$$

A számított empirikus χ^2 érték (178,5) nagyobb mint a χ^2 eloszlás kritikus értéke (3,841) az 1 szabadságfoknál és 5 %-os szignifikancia szinten, így 5 %-os szignifikancia szinten a H_0 hipotézist elutasítjuk, az alternatív hipotézist fogadjuk el, tehát a kapcsolat szignifikáns.

A p-érték is ezt mutatja: 0.0000²⁷⁵

A Yule-mutató.

A bemutatott példában:

	B1	B2	Össz.
A1	876	561	1 437
A2	362	704	1 066
Össz.	1 238	1 265	2 503

$$Y = \frac{f_{11}f_{22} - f_{12}f_{21}}{f_{11}f_{22} + f_{12}f_{21}} = \frac{876 * 704 - 362 * 561}{876 * 704 + 362 * 561} = 0,505$$

Yule 0,505

4.11 Kendall-féle rangkonkordancia-mutató

Ha kettőnél több bíráló illetve döntéshozó ($m > 2$) állapít meg rangsort, akkor a konkordancia-mutató segít megválaszolni azt a kérdést, hogy milyen az összhang a bírálók véleménye között. Tegyük fel, hogy m számú bíráló rangsorol n darab objektumot, (egységet, értékelési tényezőt). Az egységek végső rangsorát kialakíthatjuk úgy, hogy összeadjuk az m bíráló adott egységre vonatkozó rangszámát, és ezen összegek alapján rangsoroljuk az egységeket.

Rangszámnak nevezzük azt a pozitív egész számot, amely megmutatja, hogy egy konkrét adat hányadik az adathalmaz emelkedő rangsorában, vagyis $R_i = k$ ha $x_i = x_{(k)}$. Könnyen belátható, hogy a minimális érték rangszáma 1; a maximális pedig n , a rangszámok pedig a természetes számokkal egyenlők 1-től n -ig. Ordinalis (sorrendi) skálán sorrendiségre vonatkozó relációk alapján rangsorba rendezzük a megfigyelt objektumokat, egyedeket. A sorrendi skálán az egyes egyedek egymástól nem feltétlenül egyenlő távolságban helyezkednek el. Ez még nem tekinthető tiszta kvantitatív skálának, habár használja a numerikus értékeket, és gyakran további műveletek is végezhetők a rangszámokkal.

Ha teljesen azonos a bírálók véleménye, úgy az első egységhez $1m$, a másodikhoz $2m$, ... az n -ikhez nm rangösszeg fog tartozni. Ha nem teljes az összhang, ugyanahhoz az egységhez az egyik kisebb, másik nagyobb sorszámot rendel, így az összegek egymás között kiegyenlítettebbek lesznek. Az egységek közti végső sorrend is kialakítható a rangszámösszegek segítségével, de ez akkor lesz megbízható, ha nagy az összhang a bírálók között (azaz m -hez közeli a rangösszegek különbsége).

Az egyes rangszámösszegek (C_j) átlaguktól (\bar{C}) vett eltérés négyzetösszege teljes összhang esetén lesz a legnagyobb:

$$C_{\max} = \frac{m^2(n^3 - n)}{12}$$

Egymással ellentétes sorszámozás, vagy sokféle sorszámozás esetén előfordulhat a sorösszegek kiegyenlítődése, így az átlagtól való eltérés 0 is lehet. Ebben az esetben a teljes egyet nem értés esete fordul elő²⁷⁶.

²⁷⁵ Az asszociációegyütthatók.xls parancsfájllal számítható ki.

Az egyetértési - mutató:

A döntést hozók véleményegyezését a W-mutatóval jellemezhetjük.

$$0 \leq W \leq 1$$
$$W = \frac{C}{C_{\max}} = \frac{12 \sum_{j=1}^m (C_j - \bar{C})^2}{m^2 (n^3 - n)}$$

Az egyetértési – mutató értéke teljes egyetértés esetén 1, teljes egyet nem értés esetén pedig nulla. A részleges egyetértés esetében a mutató nulla és egy között vesz fel értékeket és minél közelebb van az egyhez, annál nagyobb az egyetértés a bírálók között.

A Kendall-féle rangkonkordancia-mutató (W) szignifikanciavizsgálata²⁷⁷

A W szignifikancia vizsgálata során a $W=0$ nullhipotézist vizsgáljuk, vagyis azt, hogy nincs korreláció a vizsgált rangsorok között. Az alternatív hipotézis esetén nem a véletlennek tekintjük a W adott és nullánál nagyobb értékét, hanem az egyetértésnek.

A teszteljárás a W-mutató négyzetes khi-négyzet mutatójára (χ^2) épül. A vizsgálat során alkalmazott hipotézis-rendszer:

$$H_0: \chi^2 = 0$$

$$H_1: \chi^2 > 0$$

A nullhipotézis teljesülése esetén a W-mutató χ^2 -eloszlást követ, $(n-1)$ szabadságfokkal. Vagyis:

$$\chi_{\text{empirikus}}^2 = m(n-1)W$$

Ennek ismeretében a hipotézisellenőrzés könnyen elvégezhető:

1) Meghatározzuk az empirikus χ^2 értéket ($\chi_{\text{empirikus}}^2$).

2) Kikeressük a χ^2 -eloszlás választott szignifikancia-színjéhez (alapeset 5%, 0,0500) tartozó kritikus értéket, ${}_{(n-1)}\chi_{\text{kritikus}}^2$ ahol a szabadságfok: $(n-1)$ vagy meghatározzuk a p-értéket, az empirikus szignifikancia értéket.

3) Amennyiben az általunk számított érték kisebb mint a kritikus érték,

$$\chi_{\text{empirikus}}^2 < {}_{(n-1)}\chi_{\text{kritikus}}^2$$

úgy a nullhipotézist, ellenkező esetben:

$$\chi_{\text{empirikus}}^2 \geq {}_{(n-1)}\chi_{\text{kritikus}}^2$$

az alternatív hipotézist fogadjuk el.

²⁷⁶ Kerékgyártó Györgyné – Mundruczó György [1995]: Statisztikai módszerek a gazdasági elemzésben. 2. átdolgozott kiadás. Aula Kiadó. Budapest. 65-67.

²⁷⁷ Kindler József-Papp Ottó [1977]: Komplex rendszerek vizsgálata. Összemérési módszerek. Műszaki Könyvkiadó. 180-181.

Függelék

F.1 Internetes ingyenes szoftverek és adatbázisok

Az Excel parancsfájlok és a kézikönyv letölthető:

<http://www.gmi.ktk.pte.hu/index.php?mid=33#SiposB>

Ingyenes matematikai és statisztikai szoftverek.

<http://www.statisticalconsultants.co.nz/links.html>

<http://www.statisticalconsultants.co.nz/statssoftware.html>

Microsoft Mathematics 4.0

<http://www.microsoft.com/downloads/en/details.aspx?FamilyID=9caca722-5235-401c-8d3f-9e242b794c3a>

Gnu Regression, Econometrics and Time-series Library:

<http://gretl.sourceforge.net>

Ycas:

<http://yacas.sourceforge.net/homepage.html>

Maxima:

<http://maxima.sourceforge.net/>

<http://maxima.sourceforge.net/download.html>

JMulti ingyenes sztochasztikus idősorkutatási módszereket (ARCH, ARIMA, VAR, VECM) becsülő szoftver:

<http://www.jmulti.de/>

Wessa P., (2009), ARIMA Forecasting (v1.0.5) in Free Statistics Software (v1.1.23-r7), Office for Research Development and Education, URL http://www.wessa.net/rwasp_arimaforecasting.wasp/

The R code is based on :

Borghers, E, and P. Wessa, Statistics - Econometrics - Forecasting, Office for Research Development and Education, <http://www.xycoon.com/>

Leírás:

http://www.stat.pitt.edu/stoffer/tsa2/R_time_series_quick_fix.htm

(Partial) Autocorrelation Function - Free Statistics Software (Calculator):

http://www.wessa.net/rwasp_autocorrelation.wasp#output

ARIMA Backward Selection - Free Statistics Software (Calculator)

http://www.wessa.net/rwasp_arimabackwardselection.wasp#output

ARIMA előrejelzés, az interneten, R code:

http://www.wessa.net/rwasp_arimaforecasting.wasp

Magyar adatbázisok.

A Statisztikai Szemlében megjelent tanulmányok 1923-tól letölthetők:

http://www.ksh.hu/statszemle_archive/

Adatok keresése az interneten:

Név filetype:xls (vagy doc, ppt, pdf stb.)

Példa: gold production statistics filetype:xls

A ROPStat próbaverziója letölthető: <http://www.ropstat.com/>

A statisztikai tevékenység összefogását és irányítását a legtöbb modern államban egy külön e célra szervezett hivatal látja el. Ez Magyarországon az 1867 óta működő Statisztikai Hivatal.

A fontosabb KSH adatok letölthetők a Központi Statisztikai Hivatalnak honlapjáról:

<http://portal.ksh.hu/>

Startlap: <http://statisztika.lap.hu/>

ECOSTAT Kormányzati Gazdaság- és Társadalom-stratégiai Kutató Intézet adat szolgál -tatása:

<http://www.ecostat.hu/>

Magyar Nemzeti Bank (MNB):

<http://www.mnb.hu/>

- Statisztika:

BUX (A BUDAPESTI ÉRTÉKTŐZSDE HIVATALOS INDEXE) (1991. január 2.=1000)

Magyar tőzsdeindexek, Bux és más indexek:

<http://www.mnb.hu/Statisztika/statisztikai-adatok-informaciok/adatok-idosorok>

XII Értékpapírok-BUX

A jegybanki alapkamat és a monetáris politikai eszközökhöz kapcsolódó kamatlábak időSORA (2002. január 1-jétől, százalékpontban)

- Statisztika-statisztikai adatok-időSOROK

Fogyasztói árindex

Külkereskedelem

Háztartásoknak nyújtott fogyasztási hitelek

A lakáscélú hitelek állománya szektor, lejárat és deviza szerinti bontásban

<http://www.bet.hu/>

Magyar tőzsdeindexek, Bux és más indexek:

Kereskedési adatok-statisztikák

Historikus adatok letöltése – indexek historikus értékei (visszafelé lehet menni az időben, a napra rá kattintva.) stb.

MTA KTI (MTA Közgazdaságtudományi Intézet) adatbankja:

<http://adatbank.mtaki.hu>

Gazdasági versenyhivatal:

http://www.gvh.hu:80/gvh/alpha?do=2&st=1&pg=54&m5_doc=5635&m251_act=4

Hazai szakirodalom, keresési lehetőség, általában pdf fájlban letölthető

<http://www.matarka.hu/>

Nemzetközi adatbázisok.

Nominális és reál GDP, GDP/fő, népességi, fogyasztói árindex, \$árfolyama adatok 1969-2010 tényadatok, 2011-2030-ig prognózis. Folyamatosan frissítik. 228- 230 ország illetve régió.

<http://www.ers.usda.gov/Data/Macroeconomics/>

Egyesült Nemzetek Szervezete Statisztikai Hivatala:

<http://unstats.un.org/>

Nemzetközi Valutaalap (The International Monetary Fund; IMF):

<http://www.imf.org/>

Világbank (The World Bank):

<http://web.worldbank.org/>

Európai Unió statisztikai szervezete az EUROSTAT:

<http://epp.eurostat.ec.europa.eu/>

Demográfiai adatok, startlap:

<http://demografia.lap.hu/>

USA Cenzus Hivatal:

<http://www.census.gov/>

A világ országainak korfa adatai (nemek és kor szerinti, valamint évenkénti [általában 1990-2050] bontásban) valamint a népességi mutatók letölthetők:

<http://www.census.gov/ipc/www/idb/summaries.html>

Az adatok forrása és letöltése.

Az interneten számos fontos adat megtalálható a korfák elkészítésével kapcsolatban. A világ országainak (224 ország és a világ) korfáinak az adatai (korévek, nemek: férfi - nő és összesen bontásban és az évek általában 1990-2050 között, 2005 után extrapoláció) és adatok forrása:

A világ népességére vonatkozó adatok forrása:

<http://www.census.gov/cgi-bin/ipc/idbagg>

<http://www.census.gov/ipc/www/idb/summaries.html>

Country Summary-Válasszuk ki a vizsgálni kívánt országot és kattintunk a kérdőív elküldésére (go-ra).

Dinamikus ábrázolás (Dynamic) mutatja a változásokat grafikusán 2050-ig. Az eredmény típusa: (Type of output) mutatja a korfát, 2000-ben, 2025-ben és 2050-ben, válasszuk a több népesség korfán (More population pyramids) belül az összegzést (Summary). Az adatsorok letöltéséhez: válasszuk a legördülő menüben a 094 a népesség száma az év közepén kor és nem szerint adatsort (094 Midyear Population by

age and sex) és az összes évet (All years), utána következik, kérdőív elküldése (go) és megkapjuk szövegfájlban az adatokat általában 1990-től 2050-ig évenkénti bontásban korfa és nemek szerint, korfánként összesítve és évenként összesítve is. (Egyes esetekben 1990 előtti adatokat is letölthetünk, pl. az USA-nál 1950-től, Brazíliánál 1970-től, Pakisztánnál 1981-től, Afganisztánnál 1979-től, Nigériától 1953-tól található meg például az adatok.) A világ összesen adatai az alábbi módon érhetők el: World population - World population by age and sex-select year (a lehetőségek: 1996-2050) - submit query. Ahhoz, hogy az USA eredetű DOS-os adatfájlokat (a számoknál minden ezer után vessző van) az Excel kezelni tudja a következő lépéseket kell megtenni. Mentsük le a fájlt txt szövegfájlként, úgy, hogy jelöljük ki az összes adatot (CTRL A) és utána: Szerkesztés - Az összes kijelölése - Másol) és másoljuk be a jegyzetömbbe (notepad-ba) mentsük le text fájlként, a kiterjesztés txt. A Vezérlőpult-Dátum-idő-nyelvi és területi beállítások-on belül a számok, dátumok és az idő formátumának módosításánál a magyart változtassuk angolra (egyesült államokbelire). Ezt követően a lementett fájlt nyissuk meg az Excelben: az Excel felismeri a Fájl típust: Fix széles és a Fájl eredetét: Közép-európai (DOS), válasszuk a Tovább-t és ahol szükséges a Törésvonal áthelyezésével jelöljük ki a számoszlopokat, ellenőrizzük a legördülő sáv oldalirányú és lefelé való mozgatásával azt, hogy a számoszlopok kijelölése pontos volt-e, ha szükséges mosósítsunk, majd tovább. Az oszlop adattípusánál a korfa beosztásnál válasszuk a szöveget, a többinél maradjon az általános adattípus és válasszuk a befejezést. Ezt követően a regionális beállítást (angol) változtassuk vissza magyarra. Az adatállománnyal most már lehet Excelben dolgozni, le lehet menteni a Microsoft Office Excel munkafüzet formátumban. Angol windows esetén a lépések: A Control Panelben a nyelv és regionális (Regional and Language Options) beállításokat Hungary-ról változtassuk meg English (United States) - re, majd, az excel fájl megnyitása után állítsuk vissza Hungary-ra.

2009 10 24 után az adatszolgáltatás egyszerűbbé vált: Country ki kell választani az országot, Ctrl lenyomásával az összes év kiválasztható, utána Population Pyramids-t választva, mindegyik országra elkészíti a korfát és az utolsó korfa után az adatok letölthetők Excel/CSV formátumban.

A korfák elméleti kérdései:

<http://termtud.akg.hu/okt/10/2/103.htm>

A világ országainak különböző statisztikai adatai:

http://www.nationmaster.com/graph/peo_pop-people-population

Napi és az idei év statisztikáinak alakulása pl.: népességi adatok folyamatosan: születés, halálozás, népesség abszolút növekedése, megtermelt és elfogyasztott energia mennyiség változása folyamatosan, a jelenlegi év adatai: költség adatok: oktatási és fegyverkezési- kiadások millió dollárban, termelési adatok: autó- és kerékpár- termelése darabban, eladott számítógépek darabszáma, megszűnt erdőterület nagysága hektárban, egészségügyi statisztikák, stb. nyomon követése:

<http://www.worldometers.info/>

<http://www.peterrussell.com/Odds/WorldClock.php>

Regionális adatok (Regional and Country Links):

<http://www.internetworldstats.com/>

Az USA kormányának hivatalos energiastatisztika szolgáltató honlapja: Energy Information Administration (EIA) a világ országainak energiagazdálkodással kapcsolatos adatai, idősorok, stb.:

<http://www.eia.doe.gov/>

USA hosszú idősorok: (Excel fájlok letöltésekor a regionális beállítást meg kell változtatni: angol USA utána, ha megnyitottuk az Excel fájlt visszaállítani Magyarra):

<http://stats.bls.gov/data/home.htm>

<http://minerals.usgs.gov/minerals/pubs/>

<http://www.measuringworth.com/>

<http://www.economagic.com/>

USA konjunktúra ciklusok 1853-tól

<http://www.nber.org/cycles/cyclesmain.html>

Economic Report of the President.²⁷⁸ Az amerikai elnök részére készített gazdasági jelentés, évente készül el, 1959-től található meg az idősorok. A gazdasági jelentések (az Excel fájlok és szöveges jelentések pdf formátumban) megtalálhatók 1997-től 2008-ig évente, a gazdaság minden területéről:

<http://www.gpoaccess.gov/eop/download.html>

²⁷⁸ The Economic Report of the President is an annual report written by the Chairman of the Council of Economic Advisers.

Az amerikai elnök részére készített gazdasági jelentések az 1947-1996 években megtalálható:

<http://fraser.stlouisfed.org/publications/ERP/>

A Dow-Jones index elérhető az alábbi internet címen 1896-tól.

<http://djindexes.com/mdsidx/index.cfm?event=showavgIndexData>

http://globaledge.msu.edu/resourceDesk/_statisticalDataSources.asp

Az egyes statisztikai fogalmak felfedezőiről, az első publikációkról az alábbi internetcímről szerezhetünk adatokat:

<http://jeff560.tripod.com/mathword.html>

HDI-index. Az „emberi fejlődés” indexe (angolul: Human Development Index, rövidítése: HDI) egy mutatószám, amely a világ országainak összehasonlítását teszi lehetővé a várható élettartam, az írástudás, az oktatás és az életszínvonal alapján. Általánosan elfogadott eszköze a jólét mérésének, különösen a gyermekjólétének. A Human Development Report országonkénti adatai megtalálható az interneten:

<http://hdrstats.undp.org/indicators/>

<http://hdr.undp.org/en/statistics/data/>

Nemzetközi adatbázis, idősorok Excelben

<http://www.economicswebinstitute.org>

Nemzetközi szakirodalom, keresési lehetőség, általában pdf fájlban letölthető

<http://www.oxfordjournals.org/>

Gujarati Damodar N. [2003]: Basic econometrics. McGraw-Hill Higher Education.

<http://highered.mcgraw-hill.com/sites/0072335424/>

USA elnökének a blogja: Economic statistics briefing room

<http://www.whitehouse.gov/fsbr/esbr.htm>

Federal Reserve System

<http://www.federalreserve.gov/>

The Institute For International Economics. A private research institution devoted to the study of international economic issues .

<http://www.iie.com/>

Penn World Table 6.2 (188 countries, 1950-2004, 2000 as base year)

<http://datacentre.chass.utoronto.ca/pwt/>

World Development Indicators.

<http://www.helsinki.fi/WebEc/WebEc.html>

BEA. Bureau of Economic Analysis. International Economic Accounts. International Accounts Data, Comprehensive international accounts data from the Survey of Current Business.

<http://www.bea.gov/international/index.htm#bop>

Számos adatszolgáltató elérhető:

<http://personal.ashland.edu/jgarcia/links.htm>

F.2 A matrix.xls parancsfájl működése

A számítások céljára felhasznált mátrix ($n \times n$) minimum 2×2 maximum 15×15 kvadratikus (négyzetes) mátrix. A mátrix jelölése esetünkben: $\mathbf{X} = [x_{ij}]$. A kvadratikus (négyzetes) mátrix esetében a mátrix n sorból és n oszlopból áll, ahol az n a mátrix rendje. A matrix.xls fájl a következő számításokat végzi el a mátrix munkalapon:

Inverz mátrix (\mathbf{X}^{-1})

Transzponált mátrix (\mathbf{X}') és az Inverz mátrix (\mathbf{X}^{-1}) szorzata egyenlő az egység mátrixszal (\mathbf{E}), $\mathbf{X}' \mathbf{X}^{-1} = \mathbf{E}$, ami a számítások ellenőrzésére szolgál.

A mátrix megfelelő sorainak és oszlopainak felcserélésével keletkezett mátrixot nevezzük egy mátrix transzponáltjának, jelölése általában: \mathbf{X}' , \mathbf{X}^T , \mathbf{X}^* .

Két mátrix szorzata akkor értelmezhető, ha az első tényezőnek annyi oszlopa van ahány sora a másodikonak. Ekkor a szorzat mátrix i -dik sor j -edik elemét úgy kapjuk meg, hogy az első tényező i sorának és a második tényező j oszlopának megfelelő elemeit összeszorozzuk és az értékeket összeadjuk.

Az inverz mátrix (\mathbf{X}^{-1}): egy $\mathbf{X} = [x_{ij}]$. négyzetes mátrix inverze az a $\mathbf{X}^{-1} = [x^{-1}_{ij}]$ mátrix amelyre $\mathbf{X} \mathbf{X}^{-1} = \mathbf{X}^{-1} \mathbf{X} = \mathbf{E}$.

Egységmátrix (**E**): egy négyzetes mátrix egységmátrix ha a főátlójában csak 1, más helyeken a 0 szám található.

A kvadratikus mátrix rangját definiálhatjuk úgy, mint az **X** mátrix lineárisan független oszlopainak maximális számát. Igazolható, hogy ez egy jól definiált természetes szám és megegyezik a mátrix lineárisan független sorainak maximális számával (a sorrang tehát egyenlő az oszlopranggal). A lineáris algebraiban vektorok egy halmazát lineárisan függetlennek nevezzük, ha egyikük sem fejezhető ki a többi vektor lineáris kombinációjaként. Ellenkező esetben lineárisan összefüggő vektorokról beszélünk. Ha egy kvadratikus mátrix rendje és rangja azonos, a mátrix nem szinguláris, ebben az esetben van inverze. Ha egy kvadratikus mátrix rangja kisebb a rendjénél a mátrix szinguláris, és nincs inverze.

A kvadratikus mátrixokhoz egy olyan skalárt (számot) rendelünk, amelynek az értéke függ a mátrix elemeinek nagyságától. Ezt a skalárt az adott mátrix determinánsának nevezzük és az **X** mátrix esetében a **|X|** szimbólummal jelöljük. Ha egy kvadratikus mátrix oszlopai (sorai) lineárisan függetlenek, a mátrix nem szinguláris (rang=rend) és ebben az esetben a mátrix determinánsa különbözik nullától. Ha egy kvadratikus mátrix oszlopai (sorai) lineárisan nem függetlenek, a mátrix szinguláris (rang < rend), akkor a mátrix determinánsa nulla.

A program közli a mátrix sajátértékeit (a sajátértékek szorzata a determinánssal egyenlő) és a saját vektorokat.

Az általános sajátérték probléma:

Legyen az **A** és **B** két kvadratikus (n,n) típusú mátrix. A feladat, a λ számok és az $\underline{x} \neq \underline{0}$ hozzátartozó vektorok meghatározása.

$$A\underline{x} = \lambda B\underline{x}$$

Fejezi ki, az általános sajátérték problémát. A λ számot sajátértéknek, az \underline{x} vektort saját vektornak nevezzük.

A **B=E** speciális eset, ahol **E** egy n-ed rendű egységmátrix, azaz:

$$A\underline{x} = \lambda\underline{x}$$

$$(A - \lambda E)\underline{x} = \underline{0}$$

Legyen **A=X'X**

Az alábbi sajátérték feladat Excelben megoldható:

$$(X'X)\underline{x} = \lambda\underline{x}$$

$$[(X'X) - \lambda E]\underline{x} = \underline{0}$$

A magyarázó változók korrelációs mátrixának a determinánsa egyenlő a sajátértékek szorzatával.

$$R_k = \lambda_1 \lambda_2 \dots \lambda_k$$

Az **X** mátrix munkalapon a mátrix 2000x15 (n=2000, x=15) típusú lehet. Az adatok törlése után bemásolható az **X** mátrix. Az **X** mátrix standardizálás munkalapon a következő transzformációt végzi el a program:

A magyarázó változókat (i=1,2...2000; j=1,2...15) standardizáljuk az alábbi módon:

$$\tilde{x}_{ji} = \frac{x_{ji} - \bar{x}_j}{\sigma_j \sqrt{n}}$$

Az **X** mátrix műveletek munkalapon közöljük az **X'X** mátrixot és annak inverzét $(X'X)^{-1}$. Az **X'X** mátrix a tényezőváltozók korrelációs mátrixát adja. Az inverz mátrix diagonális elemei pedig a VIF_j mutatókat adják.

Ebben az esetben:

$$X'X = R$$

$$\text{A standardizált magyarázóváltozók esetén } (X'X)_{jj}^{-1} = R_{jj}^{-1} = \text{VIF}_j$$

Tehát az ismertett módon standardizált magyarázó változók **X'X** mátrixa egyenlő a magyarázó változók korrelációs mátrixával, ugyanis az x_i és x_j változók közötti korrelációs együttható:

$$r_{ij} = \frac{\sum (x_i - \bar{x})(x_j - \bar{x})}{n \sigma_i \sigma_j}$$

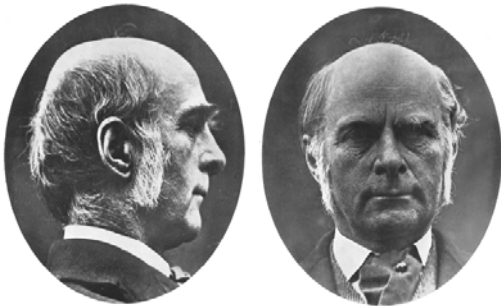
F3 Tudománytörténeti összefoglaló

A **matematikai statisztika** (mathematical statistics) fogalma először a következő publikációkban jelent^{279 280 281} meg:

Német nyelven 1867-ben található meg először a matematikai statisztika (mathematische Statistik) kifejezés: T. Wittstein [1867]: „Mathematische Statistik und deren Anwendung auf National-Oekonomie und Versicherungs-Wissenschaft” Forrás: (David, H. A. [1995]: "First (?) Occurrence of Common Terms in Mathematical Statistics," The American Statistician 49:2, May.) *Angol nyelven* a matematikai statisztika (mathematical statistics) fogalma először 1918-ban fordult elő: a következő könyvben: C. J. West [1918]: „Introduction to Mathematical Statistics” Forrás: (David, H. A. [1998]: "First (?) Occurrence of Common Terms in Probability and Statistics -- A Second List, with Corrections," The American Statistician 52:1, February).

Statisztikai sokaság (population) és **minta** (sample) először Francis Galton and W. F. R. Weldon munkájában 1877-ben jelent meg. Francis Galton and W. F. R. Weldon [1877]: "Typical laws of heredity," Nature, 15,, April 19th, p. 532.

Galton (1822-1911)²⁸²



Maximum és **minimum** kifejezések, mint értékek összehasonlítása először 1743-ban találhatók meg: W. Emerson, [1743]: Doctrine of Fluxions c. munkájában.

Módusz (mode) fogalmát Karl Pearson (1857-1936) angol kutató vezette be 1895-ben. Karl Pearson [1895]: "Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material," Philosophical Transactions of the Royal Society of London, Ser. A, 186, 343-414.

Görbe illesztése (curve fitting) Karl Pearson említi először [1902]: On the Mathematical Theory of Errors of Judgment, with Special Reference to the Personal Equation. Philosophical Transactions of the Royal Society of London. Series A, 198, 235-299.

Karl Pearson [1902]: On the Systematic Fitting of Curves to Observations and Measurements. Biometrika, 1, 265-303.

Karl Pearson fiatalkori képe²⁸³



²⁷⁹ Forrás: <http://members.aol.com/jeff570/mathword.html> (2008. február 7.) Earliest Known Uses of Some of the Words of Mathematics. Azokban az esetekben amikor más internetes forrásokat használtunk azt külön jelöljük.

²⁸⁰ A szoftverek (még a magyar nyelvű Excel is sok esetben) általában az angol nyelvű kifejezéseket használják.

²⁸¹ <http://www.maths.tcd.ie/pub/HistMath/People/RBallHist.html> áttekintést ad a XVII.-ik és XVIII.-ik század neves matematikusainak a munkásságáról. (2008. február 7.)

²⁸² Forrás: <http://galton.org/> (2008. február 7.)

²⁸³ Forrás: http://en.wikipedia.org/wiki/Karl_Pearson Közli életrajzát és publikációs jegyzékét. (2008. február 7.)

Karl Pearson időskori képe²⁸⁴



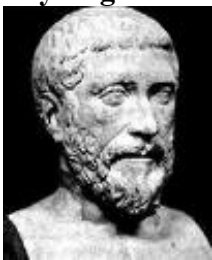
Medián (median) fogalmát először Cournot (Antoine Augustin Cournot (1801–1877) francia közgazdász, filozófus és matematikus) említi 1843-ban. Antoine A. Cournot [1843]: „Exposition de la théorie des chances et des probabilités” (pp. 119-20). David, H. A. [1998]: "First (?) Occurrence of Common Terms in Probability and Statistics -- A Second List, with Corrections," The American Statistician 52:1, February).

A. Augustin Cournot²⁸⁵



Átlag (mean) Thomas Little Heath (1861 -1940) angol matematikus 1921-ben „A History of Greek Mathematics” (Oxford: Clarendon Press, 1921, p. 85) c. művében azt írta, hogy Pythagoras görög matematikus (i. e. 580-572 – 500-490) fedezte fel számtani, a mértani és a harmonikus átlagokat. Az átlagok között a **geometriai átlag** (geometric mean) angol nyelven először megtalálható egy 1450 körül készült kéziratban, Mark Dunn [1450]: „The Art of Numbering”. A **számtani átlag** (arithmetical mean) megtalálható 1697-ben a következő publikációban: E. Halley "A Most Compendious and Facile Method for Constructing the Logarithms, Exemplified and Demonstrated from the Nature of Numbers, without any Regard to the Hyperbola, with a Speedy Method for Finding the Number from the Logarithm Given," Philosophical Transactions of the Royal Society of London, Vol. 19. (1695 - 1697), pp. 58-67.

Pythagoras²⁸⁶



Az átlag hibája. (mean error) a XIX.-ik században elfogadott kifejezés volt a hiba elméletében. Johann Carl Friedrich Gauss (1777 - 1855) vezette be az átlag hibája fogalmat 1821-ben „Theoria combinationis observationum erroribus minimis obnoxiae (Theory of the combination of observations least subject to error)” (1821, p. 7), c. munkájában.

Gauss fiatal korában.²⁸⁷

²⁸⁴ Forrás: <http://www-groups.dcs.st-and.ac.uk/~history/PictDisplay/Pearson.html> (2008. február 7.)

²⁸⁵ Forrás: <http://www.spock.com/Antoine-Augustin-Cournot> (2008. február 7.)

²⁸⁶ Forrás: <http://www.google.hu/search?q=Pythagoras++picture&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:hu:official&client=firefox-a> (2008. február 7.)

²⁸⁷ Forrás: <http://www-groups.dcs.st-and.ac.uk/~history/PictDisplay/Gauss.html> (2008. február 7.)



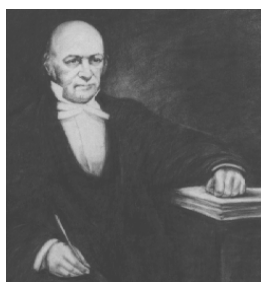
Gauss idős korában.



Négyzetes hiba (mean square) megtalálható 1838 – ban a Augustus De Morgan „An Essay on Probabilities, and Their Application to Life Contingencies and Insurance Offices” című könyvében.

Asszociációs kapcsolat vizsgálata. (association relationship) W. R. Hamilton²⁸⁸ [1848]: Researches respecting Quaternions: First Series. (Transactions of the Royal Irish Academy, 21. 199-296). Dolgozatát felolvasta a Felolvasta az Ír Királyi Akadémián 1843. nov. 13-án, de csak 1848-ban publikálta.

William Rowan Hamilton²⁸⁹ (1805-1865)



Variancia (variance, mean square deviation) Ronald Aylmer Fisher (1890 - 1962)²⁹⁰ 1922– ben megjelent munkájában fordul elő először. Ronald Aylmer Fisher [1922]: "On the mathematical foundations of theoretical statistics" Philosophical Transactions of the Royal Society, A, 222: 309-368. Fontos könyve, ami befolyásolta a XX.-ik századi statisztikai irodalmat: Ronald Aylmer Fisher [1925]: „Statistical Methods for Research Workers”²⁹¹ Kiadó: Edinburgh, Oliver és Boyd.

R. Aylmer Fisher²⁹²



Leontine Tintner festménye²⁹³

²⁸⁸ <http://www.maths.tcd.ie/pub/HistMath/People/Hamilton/> (2008. február 7.)

²⁸⁹ W. R. Hamilton publikációs jegyzéke: ld.: <http://www.maths.tcd.ie/pub/HistMath/People/Hamilton/Papers.html> (2008. február 7.)

²⁹⁰ Életrajzát ld.: J. C. Gower, Ronald Aylmer Fisher 1890-1962, Mathematical Spectrum 23 (1990-91), 76-86.

²⁹¹ R. Aylmer Fisher könyve az interneten elérhető: <http://psychclassics.yorku.ca/Fisher/Methods/> (2008. február 7.)

²⁹² <http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Fisher.html> (2008. február 7.)

²⁹³ <http://www-groups.dcs.st-and.ac.uk/~history/PictDisplay/Fisher.html> (2008. február 7.)

R. Aylmer Fisher²⁹⁴



Kvantilis. (quantile) mint általános fogalom, ami magában foglalja a kvartilist (negyedelőt), percentilist (századolót) stb. 1940-ben jelent meg először.

David, H. A. [2001]: "First (?) Occurrence of Common Terms in Statistics and Probability", Appendix B pp. 219-228. hivatkozott M. G. Kendall munkájára [1940 - 1941] "Note on the Distribution of Quantiles for Large Samples," Supplement to the Journal of the Royal Statistical Society, 7, 83-85.

Kvartilis (quartile). Alsó és felső kvartilis először 1879-ben fordul elő. Donald McAlister, The Law of the Geometric Mean, Proc. R. Soc. XXIX, p. 374.

Alsó és felső kvartilis megjelenik később 1881-ban F. Galton, [1881] "Report of the Anthropometric Committee," Report of the 51st Meeting of the British Association for the Advancement of Science, p. 245-260 .

Maurice George Kendall (1907 - 1983) angol statisztikus.

M. G. Kendall²⁹⁵



Momentum (moment) a mechanikában Newton használta először 1704-ben. Isaac Newton [1704]: De Quadratura Curvarum: "Momenta id est incrementa momentanea synchrona".

Momentum (moment) fogalmát a statisztikában a Karl Pearson vezette be, átvéve a mechanikából 1893-ban, Karl Pearson [1893]: "Asymmetrical Frequency Curves," Nature October 26th

Csúcsság (kurtosis) fogalmát először Karl Pearson használta 1905-ben: Karl Pearson [1905]: "Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson". A Rejoinder, Biometrika, 4, 169-212.

Ferdeség (skew curve) 1895-ben Karl Pearson munkájában jelenik meg először: Karl Pearson [1895]: Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material, Philosophical Transactions of the Royal Society A, 186, 343-414.

Kumulálás (cumulation) R. A. Fisher használta először 1929-ben és 1931-ben Wisharttal: R. A. Fisher [1929]: Moments and Product Moments of Sampling Distributions," Proceedings of the London Mathematical Society, Series 2, 30, 199-238. és R. A. Fisher és J. Wishart, [1931]: "The Derivation of the Pattern Formulae of Two-Way Partitions From Those of Simpler Patterns," Proceedings of the London Mathematical Society, Ser. 2, 33, p.195. Harold Hotelling (J. Amer. Stat Assoc., 28, 1933, 374).

Koncentráció (concentration) mérésére Lorenz dolgozta ki a Róla elnevezett görbét. Lorenz, M. O. [1905]: Methods of measuring the concentration of wealth Publications of the American Statistical Association. 9: 209-219.

Max Otto Lorenz (1880 – 1962) amerikai közgazdász. A Lorenz görbét először a következő könyvben publikálták: King, W. I. [1912]:. The Elements of Statistical Method. New York: Macmillan.

Lorenz, M. O.²⁹⁶

²⁹⁴ <http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Fisher.html> (2008. február 7.)

²⁹⁵ http://www-gap.dcs.st-and.ac.uk/~history/Mathematicians/Kendall_Maurice.html (2008. február 7.)

²⁹⁶ <http://www.spock.com/Max-O.-Lorenz> (2008. február 7.)



Durbin - Watson teszt (Durbin - Watson test) kidolgozói: Durbin, J., and Watson, G. S., [1950]: Testing for Serial Correlation in Least Squares Regression, I. *Biometrika* 37.

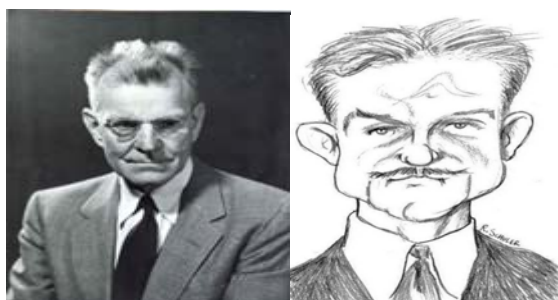
Benjamin Gompertz²⁹⁷ (1779-1865)



BENJAMIN GOMPERTZ

A Gompertz logisztikus trend felfedezője.²⁹⁸ Benjamin Gompertz²⁹⁹ biztosítási matematikus 1825-ben fölfedezte azt a halandósági törvényt, amit az állatokon végzett vizsgálatok is megerősítenek. Az emberi halandósági ráta a nemi érettség elérése idején a legkisebb, utána exponenciálisan emelkedik. Gompertz tétele szerint az ember elhalálzási esélye harminc és nyolcvanéves kora között hét évente megduplázódik, nyolcvanéves kora fölött viszont újra csökkenni kezd. Az újabb vizsgálatok alapján a halandóság erejének alakulását leíró görbe csak a születéstől a korai kamaszkori mélypontjáig és felnőttkori monoton jellegű emelkedésének a legfiatalabb életkortól kezdődően alakul úgy, ahogyan azt Gompertz leírta.³⁰⁰ Az egyre növekvő tapasztalati anyag elemzése kapcsán az is egyértelművé vált, hogy az orvostudomány fejlődése, az egészségügyi rendszabályok és a morbiditást és mortalitást befolyásoló egyéb tényezők a különböző életkorúakat eltérő arányban érintik, ezért a különböző nemű és korú népesség halandóságának egymáshoz viszonyított nagysága állandóan változik. A nyolcvanévesnél idősebb korosztályokban a nők már túlnyomó többségbe kerültek, így a Gompertz - egyenlet³⁰¹ ³⁰² lényegében csak a nőket támogatja.

Marion King Hubbert. (1903-1989)³⁰³ ³⁰⁴ ³⁰⁵



A texasi San Saba-ban született.³⁰⁶ Tanulmányait Chicagói Egyetemen folytatta, ahol geológiát, matematikát és fizikát tanult. B. S fokozatot 1926-ban, M. S.-t 1928-ban, PhD-t pedig 1937-ben szerzett. PhD tanulmányai alatt az Amerada olajtársaság geológus asszisztenseként dolgozott. 1943-tól 1964-ig a Shell

²⁹⁷ http://en.wikipedia.org/wiki/Benjamin_Gompertz

²⁹⁸ Gompertz, B., [1825].

²⁹⁹ http://en.wikipedia.org/wiki/Benjamin_gompertz

³⁰⁰ Ld. Valkovics Emil [2001] 121-141.

³⁰¹ Gompertz, B., [1825]: 513-585.

³⁰² Bartlett M. S. – Cox F. R. S. [1975] 16.

³⁰³ http://hu.wikipedia.org/wiki/M._King_Hubbert

³⁰⁴ <http://rsparlourtricks.blogspot.com/2005/10/m-king-hubbert.html>

³⁰⁵ <http://www.hubbertpeak.com/hubbert/>

³⁰⁶ http://hu.wikipedia.org/wiki/M._King_Hubbert

Oil Company alkalmazottja, ezután 1976-os nyugdíjba vonulásáig a United States Geological Survey kutató fizikusa volt. A Stanford Egyetemen geológiát és geofizikát tanított 1963-tól 1968-ig, a Berkeley Egyetemen pedig 1973-tól 1976-ig. A technokrata mozgalom aktív tagja, a 30-as években alakult Technocracy Incorporated szervezet egyik alapítója volt. Több fontos eredményt ért el a geológia és a geofizika területén. Nevéhez fűződik az olajtermelés időbeli alakulását egy adott területen modellező haranggörbbe, az ún. Hubbert-görbe, az olajhozam-csúcs elméletének egyik központi eleme.

F.4. Táblázatok

A Béta-eloszlás kritikus értékei³⁰⁷ a Szroeter teszthez, 5 %-os szignifikancia szinten, ha a lineáris regressziós egyenes konstanst is tartalmaz.

³⁰⁷ M. J. Harrison [1982]: Table 4. 165.

n	m=k+1=2	m=k+1=3	m=k+1=4	m=k+1=5	m=k+1=6
8	0,873	0,758	0,583	0,326	0,025
9	0,962	0,873	0,758	0,583	0,326
10	1,033	0,962	0,873	0,758	0,583
11	1,091	1,033	0,962	0,873	0,758
12	1,140	1,091	1,033	0,962	0,873
13	1,181	1,140	1,091	1,033	0,962
14	1,217	1,181	1,140	1,091	1,033
15	1,248	1,217	1,181	1,140	1,091
16	1,276	1,248	1,217	1,181	1,140
17	1,301	1,276	1,248	1,217	1,181
18	1,324	1,301	1,276	1,248	1,217
19	1,345	1,324	1,301	1,276	1,248
20	1,363	1,345	1,324	1,301	1,276
21	1,381	1,363	1,345	1,324	1,301
22	1,397	1,381	1,363	1,345	1,324
23	1,411	1,397	1,381	1,363	1,345
24	1,425	1,411	1,397	1,381	1,363
25	1,438	1,425	1,411	1,397	1,381
26	1,450	1,438	1,425	1,411	1,397
27	1,461	1,450	1,438	1,425	1,411
28	1,471	1,461	1,450	1,438	1,425
29	1,481	1,471	1,461	1,450	1,438
30	1,491	1,481	1,471	1,461	1,450
31	1,500	1,491	1,481	1,471	1,461
32	1,508	1,500	1,491	1,481	1,471
33	1,516	1,508	1,500	1,491	1,481
34	1,524	1,516	1,508	1,500	1,491
35	1,531	1,524	1,516	1,508	1,500
36	1,537	1,531	1,524	1,516	1,508
37	1,542	1,538	1,531	1,524	1,516
38	1,547	1,545	1,538	1,531	1,524
39	1,551	1,551	1,545	1,538	1,531
40	1,555	1,557	1,551	1,545	1,538
45	1,575	1,585	1,580	1,574	1,569
50	1,593	1,608	1,603	1,599	1,594
55	1,608	1,627	1,623	1,620	1,616
60	1,622	1,644	1,641	1,637	1,634
65	1,635	1,659	1,656	1,653	1,650
70	1,646	1,671	1,669	1,666	1,664
75	1,656	1,682	1,681	1,679	1,676
80	1,666	1,692	1,692	1,690	1,688
85	1,675	1,699	1,701	1,699	1,698
90	1,683	1,706	1,710	1,709	1,707
95	1,690	1,712	1,718	1,716	1,715
100	1,697	1,718	1,726	1,724	1,723

Standard normális eloszlás sűrűségfüggvény értékei ³⁰⁸

z	0	1	2	3	4	5	6	7	8	9
0,0	0,500	0,496	0,492	0,488	0,484	0,480	0,476	0,472	0,468	0,464
0,1	0,460	0,456	0,452	0,448	0,444	0,440	0,436	0,433	0,429	0,425
0,2	0,421	0,417	0,413	0,409	0,405	0,401	0,397	0,394	0,390	0,386
0,3	0,382	0,378	0,374	0,371	0,367	0,363	0,359	0,356	0,352	0,348
0,4	0,345	0,341	0,337	0,334	0,330	0,326	0,323	0,319	0,316	0,312
0,5	0,309	0,305	0,302	0,298	0,295	0,291	0,288	0,284	0,281	0,278
0,6	0,274	0,271	0,268	0,264	0,261	0,258	0,255	0,251	0,248	0,245
0,7	0,242	0,239	0,236	0,233	0,230	0,227	0,224	0,221	0,218	0,215
0,8	0,212	0,209	0,206	0,203	0,200	0,198	0,195	0,192	0,189	0,187
0,9	0,184	0,181	0,179	0,176	0,174	0,171	0,169	0,166	0,164	0,161
1,0	0,159	0,156	0,154	0,152	0,149	0,147	0,145	0,142	0,140	0,138
1,1	0,136	0,133	0,131	0,129	0,127	0,125	0,123	0,121	0,119	0,117
1,2	0,115	0,113	0,111	0,109	0,107	0,106	0,104	0,102	0,100	0,099
1,3	0,097	0,095	0,093	0,092	0,090	0,089	0,087	0,085	0,084	0,082
1,4	0,081	0,079	0,078	0,076	0,075	0,074	0,072	0,071	0,069	0,068
1,5	0,067	0,066	0,064	0,063	0,062	0,061	0,059	0,058	0,057	0,056
1,6	0,055	0,054	0,053	0,052	0,051	0,049	0,048	0,047	0,046	0,046
1,7	0,045	0,044	0,043	0,042	0,041	0,040	0,039	0,038	0,038	0,037
1,8	0,036	0,035	0,034	0,034	0,033	0,032	0,031	0,031	0,030	0,029
1,9	0,029	0,028	0,027	0,027	0,026	0,026	0,025	0,024	0,024	0,023
2,0	0,023	0,022	0,022	0,021	0,021	0,020	0,020	0,019	0,019	0,018
2,1	0,018	0,017	0,017	0,017	0,016	0,016	0,015	0,015	0,015	0,014
2,2	0,014	0,014	0,013	0,013	0,013	0,012	0,012	0,012	0,011	0,011
2,3	0,011	0,010	0,010	0,010	0,010	0,009	0,009	0,009	0,009	0,008
2,4	0,008	0,008	0,008	0,008	0,007	0,007	0,007	0,007	0,007	0,006
2,5	0,006	0,006	0,006	0,006	0,006	0,005	0,005	0,005	0,005	0,005
2,6	0,005	0,005	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004
2,7	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003
2,8	0,003	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002
2,9	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,001	0,001	0,001
3,0	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001

Kritikus értékek különböző szignifikancia-szintek esetén

Szignifikancia-szint (α)						
Egyoldalú	0,1000	0,0500	0,0250	0,0225	0,0100	0,0050
Kétoldalú	0,2000	0,1000	0,0500	0,0450	0,0200	0,0100
z	1,280	1,645	1,960	2,000	2,330	2,587

³⁰⁸ Pintér József – Rappai Gábor (szerkesztő) [2007]: 489.

Student-féle t-eloszlás kritikus értékei különféle szignifikancia-szint mellett³⁰⁹

Szabadság- fok	Szignifikancia-szint				
	0,1	0,05	0,025	0,01	0,005
1	3,078	6,314	12,706	31,821	63,656
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
50	1,299	1,676	2,009	2,403	2,678
60	1,296	1,671	2,000	2,390	2,660
70	1,294	1,667	1,994	2,381	2,648
80	1,292	1,664	1,990	2,374	2,639
90	1,291	1,662	1,987	2,368	2,632
100	1,290	1,660	1,984	2,364	2,626
150	1,287	1,655	1,976	2,351	2,609
200	1,286	1,653	1,972	2,345	2,601

³⁰⁹ Ramanathan Ramu [2003]: 609.

χ^2 -eloszlás kritikus értékei különféle szignifikancia-szintek mellett³¹⁰

Szabadságfok	Szignifikancia-szint					
	0,9900	0,9500	0,9000	0,1000	0,0500	0,0100
1	0.000	0.004	0.016	2.706	3.841	6.635
2	0.020	0.103	0.211	4.605	5.991	9.210
3	0.115	0.352	0.584	6.251	7.815	11.345
4	0.297	0.711	1.064	7.779	9.488	13.277
5	0.554	1.145	1.610	9.236	11.070	15.086
6	0.872	1.635	2.204	10.645	12.592	16.812
7	1.239	2.167	2.833	12.017	14.067	18.475
8	1.647	2.733	3.490	13.362	15.507	20.090
9	2.088	3.325	4.168	14.684	16.919	21.666
10	2.558	3.940	4.865	15.987	18.307	23.209
11	3.053	4.575	5.578	17.275	19.675	24.725
12	3.571	5.226	6.304	18.549	21.026	26.217
13	4.107	5.892	7.041	19.812	22.362	27.688
14	4.660	6.571	7.790	21.064	23.685	29.141
15	5.229	7.261	8.547	22.307	24.996	30.578
16	5.812	7.962	9.312	23.542	26.296	32.000
17	6.408	8.672	10.085	24.769	27.587	33.409
18	7.015	9.390	10.865	25.989	28.869	34.805
19	7.633	10.117	11.651	27.204	30.144	36.191
20	8.260	10.851	12.443	28.412	31.410	37.566
21	8.897	11.591	13.240	29.615	32.671	38.932
22	9.542	12.338	14.041	30.813	33.924	40.289
23	10.196	13.091	14.848	32.007	35.172	41.638
24	10.856	13.848	15.659	33.196	36.415	42.980
25	11.524	14.611	16.473	34.382	37.652	44.314
26	12.198	15.379	17.292	35.563	38.885	45.642
27	12.878	16.151	18.114	36.741	40.113	46.963
28	13.565	16.928	18.939	37.916	41.337	48.278
29	14.256	17.708	19.768	39.087	42.557	49.588
30	14.953	18.493	20.599	40.256	43.773	50.892
31	15.655	19.281	21.434	41.422	44.985	52.191
32	16.362	20.072	22.271	42.585	46.194	53.486
33	17.073	20.867	23.110	43.745	47.400	54.775
34	17.789	21.664	23.952	44.903	48.602	56.061
35	18.509	22.465	24.797	46.059	49.802	57.342
36	19.233	23.269	25.643	47.212	50.998	58.619
37	19.960	24.075	26.492	48.363	52.192	59.893
38	20.691	24.884	27.343	49.513	53.384	61.162
39	21.426	25.695	28.196	50.660	54.572	62.428
40	22.164	26.509	29.051	51.805	55.758	63.691
50	29.707	34.764	37.689	63.167	67.505	76.154
60	37.485	43.188	46.459	74.397	79.082	88.379
70	45.442	51.739	55.329	85.527	90.531	100.425
80	53.540	60.391	64.278	96.578	101.879	112.329
90	61.754	69.126	73.291	107.565	113.145	124.116
100	70.065	77.929	82.358	118.498	124.342	135.807
150	112.668	122.692	128.275	172.581	179.581	193.207
200	156.432	168.279	174.835	226.021	233.994	249.445
250	200.939	214.392	221.806	279.050	287.882	304.939

³¹⁰ Aczel A. D. [2002]: Table 4 alapján.

F-eloszlás kritikus értékei 5%-os egyoldalú (10%-os kétoldalú) szignifikancia-szint mellett³¹¹

Nevező szf	Számológó szabadságfoka																
	1	2	3	4	5	6	7	8	9	10	15	20	25	30	50	100	200
2	18,5	19,0	19,2	19,2	19,3	19,3	19,7	19,4	19,3	19,4	19,4	19,4	19,6	19,5	19,5	19,5	18,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,66	8,63	8,62	8,58	8,55	10,1
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,80	5,77	5,75	5,70	5,66	7,71
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,56	4,52	4,50	4,44	4,41	6,61
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87	3,83	3,81	3,75	3,71	5,99
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,44	3,40	3,38	3,32	3,27	5,59
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15	3,11	3,08	3,02	2,97	5,32
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94	2,89	2,86	2,80	2,76	5,12
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77	2,73	2,70	2,64	2,59	4,96
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,72	2,65	2,60	2,57	2,51	2,46	4,84
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,62	2,54	2,50	2,47	2,40	2,35	4,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,53	2,46	2,41	2,38	2,31	2,26	4,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,46	2,39	2,34	2,31	2,24	2,19	4,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33	2,28	2,25	2,18	2,12	4,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,35	2,28	2,23	2,19	2,12	2,07	4,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,31	2,23	2,18	2,15	2,08	2,02	4,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,27	2,19	2,14	2,11	2,04	1,98	4,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,23	2,16	2,11	2,07	2,00	1,94	4,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,12	2,07	2,04	1,97	1,91	4,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,18	2,10	2,05	2,01	1,94	1,88	4,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,15	2,07	2,02	1,98	1,91	1,85	4,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,13	2,05	2,00	1,96	1,88	1,82	4,28
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,11	2,03	1,97	1,94	1,86	1,80	4,26
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,09	2,01	1,96	1,92	1,84	1,78	4,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,07	1,99	1,94	1,90	1,82	1,76	4,23
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,06	1,97	1,92	1,88	1,81	1,74	4,21
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,04	1,96	1,91	1,87	1,79	1,73	4,20
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,03	1,94	1,89	1,85	1,77	1,71	4,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,93	1,88	1,84	1,76	1,70	4,17
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	1,96	1,88	1,82	1,79	1,70	1,63	4,12
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	1,92	1,84	1,78	1,74	1,66	1,59	4,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,87	1,78	1,73	1,69	1,60	1,52	4,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,84	1,75	1,69	1,65	1,56	1,48	4,00
75	3,97	3,12	2,73	2,49	2,34	2,22	2,13	2,06	2,01	1,96	1,80	1,71	1,65	1,61	1,52	1,44	3,97
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,77	1,68	1,62	1,57	1,48	1,39	3,94
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,72	1,62	1,56	1,52	1,41	1,32	3,89

³¹¹ Aczel A. D. [2002]: Table 5A alapján.

F-eloszlás kritikus értékei 2,5%-os egyoldalú (5%-os kétoldalú) szignifikancia-szint mellett

Nevező szf	Számológó szabadságfoka																
	1	2	3	4	5	6	7	8	9	10	15	20	25	30	50	100	200
2	38,5	39,0	39,2	39,3	39,3	39,3	39,4	39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5	39,5	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,4	14,4	14,2	14,1	14,1	14,0	14,0	13,9	13,9
4	12,2	10,6	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,66	8,56	8,50	8,46	8,38	8,32	8,29
5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,43	6,33	6,27	6,23	6,14	6,08	6,05
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,27	5,17	5,11	5,07	4,98	4,92	4,88
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,57	4,47	4,40	4,36	4,28	4,21	4,18
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,10	4,00	3,94	3,89	3,81	3,74	3,70
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,77	3,67	3,60	3,56	3,47	3,40	3,37
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,52	3,42	3,35	3,31	3,22	3,15	3,12
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,33	3,23	3,16	3,12	3,03	2,96	2,92
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,18	3,07	3,01	2,96	2,87	2,80	2,76
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,05	2,95	2,88	2,84	2,74	2,67	2,63
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	2,95	2,84	2,78	2,73	2,64	2,56	2,53
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,86	2,76	2,69	2,64	2,55	2,47	2,44
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,79	2,68	2,61	2,57	2,47	2,40	2,36
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,72	2,62	2,55	2,50	2,41	2,33	2,29
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,67	2,56	2,49	2,44	2,35	2,27	2,23
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,62	2,51	2,44	2,39	2,30	2,22	2,18
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,57	2,46	2,40	2,35	2,25	2,17	2,13
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,53	2,42	2,36	2,31	2,21	2,13	2,09
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,50	2,39	2,32	2,27	2,17	2,09	2,05
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,47	2,36	2,29	2,24	2,14	2,06	2,01
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,44	2,33	2,26	2,21	2,11	2,02	1,98
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,41	2,30	2,23	2,18	2,08	2,00	1,95
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,39	2,28	2,21	2,16	2,05	1,97	1,92
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,36	2,25	2,18	2,13	2,03	1,94	1,90
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,34	2,23	2,16	2,11	2,01	1,92	1,88
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,32	2,21	2,14	2,09	1,99	1,90	1,86
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,31	2,20	2,12	2,07	1,97	1,88	1,84
35	5,48	4,11	3,52	3,18	2,96	2,80	2,68	2,58	2,50	2,44	2,23	2,12	2,05	2,00	1,89	1,80	1,75
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,18	2,07	1,99	1,94	1,83	1,74	1,69
50	5,34	3,97	3,39	3,05	2,83	2,67	2,55	2,46	2,38	2,32	2,11	1,99	1,92	1,87	1,75	1,66	1,60
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,06	1,94	1,87	1,82	1,70	1,60	1,54
75	5,23	3,88	3,30	2,96	2,74	2,58	2,46	2,37	2,29	2,22	2,01	1,90	1,82	1,76	1,65	1,54	1,48
100	5,18	3,83	3,25	2,92	2,70	2,54	2,42	2,32	2,24	2,18	1,97	1,85	1,77	1,71	1,59	1,48	1,42
200	5,10	3,76	3,18	2,85	2,63	2,47	2,35	2,26	2,18	2,11	1,90	1,78	1,70	1,64	1,51	1,39	1,32

Durbin-Watson-próba, az 5 százalékos d_L és d_U értékek az egyoldali próbához. 312 313 314

5%	k=1		k=2		k=3		k=4		k=5		k=6		k=7		k=8		k=9		k=10	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,610	1,400	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	0,700	1,356	0,467	1,896	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	0,763	1,332	0,559	1,777	0,367	2,287	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	0,824	1,320	0,629	1,699	0,455	2,128	0,296	2,588	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,414	0,243	2,822	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
11	0,927	1,324	0,758	1,604	0,595	1,928	0,444	2,283	0,315	2,645	0,203	3,004	-----	-----	-----	-----	-----	-----	-----	-----
12	0,971	1,331	0,812	1,579	0,658	1,864	0,512	2,177	0,380	2,506	0,268	2,832	0,171	3,149	-----	-----	-----	-----	-----	-----
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094	0,444	2,390	0,328	2,692	0,230	2,985	0,147	3,266	-----	-----	-----	-----
14	1,045	1,350	0,905	1,551	0,767	1,779	0,632	2,030	0,505	2,296	0,389	2,572	0,286	2,848	0,200	3,111	0,127	3,360	-----	-----
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977	0,562	2,220	0,447	2,471	0,343	2,727	0,251	2,979	0,175	3,216	0,111	3,438
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935	0,615	2,157	0,502	2,388	0,398	2,624	0,304	2,860	0,222	3,090	0,155	3,304
17	1,133	1,381	1,015	1,536	0,897	1,710	0,779	1,900	0,664	2,104	0,554	2,318	0,451	2,537	0,356	2,757	0,272	2,975	0,198	3,184
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872	0,710	2,060	0,603	2,258	0,502	2,461	0,407	2,668	0,321	2,873	0,244	3,073
19	1,180	1,401	1,074	1,536	0,967	1,685	0,859	1,848	0,752	2,023	0,649	2,206	0,549	2,396	0,456	2,589	0,369	2,783	0,290	2,974
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828	0,792	1,991	0,691	2,162	0,595	2,339	0,502	2,521	0,416	2,704	0,336	2,885
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812	0,829	1,964	0,731	2,124	0,637	2,290	0,546	2,461	0,461	2,633	0,380	2,806
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797	0,863	1,940	0,769	2,090	0,677	2,246	0,588	2,407	0,504	2,571	0,424	2,735
23	1,257	1,437	1,168	1,543	1,078	1,660	0,986	1,785	0,895	1,920	0,804	2,061	0,715	2,208	0,628	2,360	0,545	2,514	0,465	2,670
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775	0,925	1,902	0,837	2,035	0,750	2,174	0,666	2,318	0,584	2,464	0,506	2,613
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767	0,953	1,886	0,868	2,013	0,784	2,144	0,702	2,280	0,621	2,419	0,544	2,560
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759	0,979	1,873	0,897	1,992	0,816	2,117	0,735	2,246	0,657	2,379	0,581	2,513
27	1,316	1,469	1,240	1,556	1,162	1,651	1,084	1,753	1,004	1,861	0,925	1,974	0,845	2,093	0,767	2,216	0,691	2,342	0,616	2,470
28	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747	1,028	1,850	0,951	1,959	0,874	2,071	0,798	2,188	0,723	2,309	0,649	2,431
29	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743	1,050	1,841	0,975	1,944	0,900	2,052	0,826	2,164	0,753	2,278	0,681	2,396
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,071	1,833	0,998	1,931	0,926	2,034	0,854	2,141	0,782	2,251	0,712	2,363
31	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735	1,090	1,825	1,020	1,920	0,950	2,018	0,879	2,120	0,810	2,226	0,741	2,333
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732	1,109	1,819	1,041	1,909	0,972	2,004	0,904	2,102	0,836	2,203	0,769	2,306
33	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730	1,127	1,813	1,061	1,900	0,994	1,991	0,927	2,085	0,861	2,181	0,796	2,281
34	1,393	1,514	1,333	1,580	1,271	1,652	1,208	1,728	1,144	1,808	1,079	1,891	1,015	1,978	0,950	2,069	0,885	2,162	0,821	2,257
35	1,402	1,519	1,343	1,584	1,283	1,653	1,222	1,726	1,160	1,803	1,097	1,884	1,034	1,967	0,971	2,054	0,908	2,144	0,845	2,236
36	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724	1,175	1,799	1,114	1,876	1,053	1,957	0,991	2,041	0,930	2,127	0,868	2,216
37	1,419	1,530	1,364	1,590	1,307	1,655	1,249	1,723	1,190	1,795	1,131	1,870	1,071	1,948	1,011	2,029	0,951	2,112	0,891	2,197
38	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722	1,204	1,792	1,146	1,864	1,088	1,939	1,029	2,017	0,970	2,098	0,912	2,180
39	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722	1,218	1,789	1,161	1,859	1,104	1,932	1,047	2,007	0,990	2,085	0,932	2,164
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,230	1,786	1,175	1,854	1,120	1,924	1,064	1,997	1,008	2,072	0,952	2,149
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720	1,287	1,776	1,238	1,835	1,189	1,895	1,139	1,958	1,089	2,022	1,038	2,088
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721	1,335	1,771	1,291	1,822	1,246	1,875	1,201	1,930	1,156	1,986	1,110	2,044
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724	1,374	1,768	1,334	1,814	1,294	1,861	1,253	1,909	1,212	1,959	1,170	2,010
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727	1,408	1,767	1,372	1,808	1,335	1,850	1,298	1,894	1,260	1,939	1,222	1,984
65	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731	1,438	1,767	1,404	1,805	1,370	1,843	1,336	1,882	1,301	1,923	1,266	1,964
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735	1,464	1,768	1,433	1,802	1,401	1,838	1,369	1,874	1,337	1,910	1,305	1,948
75	1,598	1,652	1,571	1,680	1,543	1,709	1,515	1,739	1,487	1,770	1,458	1,801	1,428	1,834	1,399	1,867	1,369	1,901	1,339	1,935
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743	1,507	1,772	1,480	1,801	1,453	1,831	1,425	1,861	1,397	1,893	1,369	1,925
85	1,624	1,671	1,600	1,696	1,575	1,721	1,550	1,747	1,525	1,774	1,500	1,801	1,474	1,829	1,448	1,857	1,422	1,886	1,396	1,916
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751	1,542	1,776	1,518	1,801	1,494	1,827	1,469	1,854	1,445	1,881	1,420	1,909
95	1,645	1,687	1,623	1,709	1,602	1,732	1,579	1,755	1,557	1,778	1,535	1,802	1,512	1,827	1,489	1,852	1,465	1,877	1,442	1,903
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758	1,571	1,780	1,550	1,803	1,528	1,826	1,506	1,850	1,484	1,874	1,462	1,898
150	1,720	1,747	1,706	1,760	1,693	1,774	1,679	1,788	1,665	1,802	1,651	1,817	1,637	1,832	1,622	1,846	1,608	1,862	1,593	1,877
200	1,758	1,779	1,748	1,789	1,738	1,799	1,728	1,809	1,718	1,820	1,707	1,831	1,697	1,841	1,686	1,852	1,675	1,863	1,665	1,874

³¹² Savin, N. E.-White K. J. [1977]: 1989-1996.

³¹³ Ramanathan Ramu [2003]: 614-615.

³¹⁴ Az n= a megfigyelések száma (n=1,2...200), k= a magyarázóváltozók száma (k=1,2...20).

5%	k=11		k=12		k=13		k=14		k=15		k=16		k=17		k=18		k=19		k=20	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
6	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
7	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
8	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
9	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
10	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
11	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
12	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
13	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
14	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
15	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
16	0,098	3,503	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
17	0,138	3,378	0,087	3,557	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
18	0,177	3,265	0,123	3,441	0,078	3,603	----	----	----	----	----	----	----	----	----	----	----	----	----	----
19	0,220	3,159	0,160	3,335	0,111	3,496	0,070	3,642	----	----	----	----	----	----	----	----	----	----	----	----
20	0,263	3,063	0,200	3,234	0,145	3,395	0,100	3,542	0,063	3,676	----	----	----	----	----	----	----	----	----	----
21	0,307	2,976	0,240	3,141	0,182	3,300	0,132	3,448	0,091	3,583	0,058	3,705	----	----	----	----	----	----	----	----
22	0,349	2,897	0,281	3,057	0,220	3,211	0,166	3,358	0,120	3,495	0,083	3,619	0,052	3,731	----	----	----	----	----	----
23	0,391	2,826	0,322	2,979	0,259	3,128	0,202	3,272	0,153	3,409	0,110	3,535	0,076	3,650	0,048	3,753	----	----	----	----
24	0,431	2,761	0,362	2,908	0,297	3,053	0,239	3,193	0,186	3,327	0,141	3,454	0,101	3,572	0,070	3,678	0,044	3,773	----	----
25	0,470	2,702	0,400	2,844	0,335	2,983	0,275	3,119	0,221	3,251	0,172	3,376	0,130	3,494	0,094	3,604	0,065	3,702	0,041	3,790
26	0,508	2,649	0,438	2,784	0,373	2,919	0,312	3,051	0,256	3,179	0,205	3,303	0,160	3,420	0,120	3,531	0,087	3,632	0,060	3,724
27	0,544	2,600	0,475	2,730	0,409	2,859	0,348	2,987	0,291	3,112	0,238	3,233	0,191	3,349	0,149	3,460	0,112	3,563	0,081	3,658
28	0,578	2,555	0,510	2,680	0,445	2,805	0,383	2,928	0,325	3,050	0,271	3,168	0,222	3,283	0,178	3,392	0,138	3,495	0,104	3,592
29	0,612	2,515	0,544	2,634	0,479	2,755	0,418	2,874	0,359	2,992	0,305	3,107	0,254	3,219	0,208	3,327	0,166	3,431	0,129	3,528
30	0,643	2,477	0,577	2,592	0,512	2,708	0,451	2,823	0,392	2,937	0,337	3,050	0,286	3,160	0,238	3,266	0,195	3,368	0,156	3,465
31	0,674	2,443	0,608	2,553	0,545	2,665	0,484	2,776	0,425	2,887	0,370	2,996	0,317	3,103	0,269	3,208	0,224	3,309	0,183	3,406
32	0,703	2,411	0,638	2,517	0,576	2,625	0,515	2,733	0,457	2,840	0,401	2,946	0,349	3,050	0,299	3,153	0,253	3,252	0,211	3,348
33	0,731	2,382	0,668	2,484	0,606	2,588	0,546	2,692	0,488	2,796	0,432	2,899	0,379	3,000	0,329	3,100	0,283	3,198	0,239	3,293
34	0,758	2,355	0,695	2,454	0,634	2,554	0,575	2,654	0,518	2,754	0,462	2,854	0,409	2,954	0,359	3,051	0,312	3,147	0,267	3,240
35	0,783	2,330	0,722	2,425	0,662	2,521	0,604	2,619	0,547	2,716	0,492	2,813	0,439	2,910	0,388	3,005	0,340	3,099	0,295	3,190
36	0,808	2,306	0,748	2,398	0,689	2,492	0,631	2,586	0,575	2,680	0,520	2,774	0,467	2,868	0,417	2,961	0,369	3,053	0,323	3,142
37	0,831	2,285	0,772	2,374	0,714	2,464	0,657	2,555	0,602	2,646	0,548	2,738	0,495	2,829	0,445	2,920	0,397	3,009	0,351	3,097
38	0,854	2,265	0,796	2,351	0,739	2,438	0,683	2,526	0,628	2,614	0,575	2,703	0,522	2,792	0,472	2,880	0,424	2,968	0,378	3,054
39	0,875	2,246	0,819	2,329	0,763	2,413	0,707	2,499	0,653	2,585	0,600	2,671	0,549	2,757	0,499	2,843	0,451	2,929	0,404	3,013
40	0,896	2,228	0,840	2,309	0,785	2,391	0,731	2,473	0,678	2,557	0,626	2,641	0,575	2,724	0,525	2,808	0,477	2,829	0,430	2,974
45	0,988	2,156	0,938	2,225	0,887	2,296	0,838	2,367	0,788	2,439	0,740	2,512	0,692	2,586	0,644	2,659	0,598	2,733	0,553	2,807
50	1,064	2,103	1,019	2,163	0,973	2,225	0,927	2,287	0,882	2,350	0,836	2,414	0,792	2,479	0,747	2,544	0,703	2,610	0,660	2,675
55	1,129	2,062	1,087	2,116	1,045	2,170	1,003	2,225	0,961	2,281	0,919	2,338	0,877	2,396	0,836	2,454	0,795	2,512	0,754	2,571
60	1,184	2,031	1,145	2,079	1,106	2,127	1,068	2,177	1,029	2,227	0,990	2,278	0,951	2,330	0,913	2,382	0,874	2,434	0,836	2,487
65	1,231	2,006	1,195	2,049	1,160	2,093	1,124	2,138	1,088	2,183	1,052	2,229	1,016	2,276	0,980	2,323	0,944	2,371	0,908	2,419
70	1,272	1,987	1,239	2,026	1,206	2,066	1,172	2,106	1,139	2,148	1,105	2,189	1,072	2,232	1,038	2,275	1,005	2,318	0,971	2,362
75	1,308	1,970	1,277	2,006	1,247	2,043	1,215	2,080	1,184	2,118	1,153	2,156	1,121	2,195	1,090	2,235	1,058	2,275	1,027	2,315
80	1,340	1,957	1,311	1,991	1,283	2,024	1,253	2,059	1,224	2,093	1,195	2,129	1,165	2,165	1,136	2,201	1,106	2,238	1,076	2,275
85	1,369	1,946	1,342	1,977	1,315	2,009	1,287	2,040	1,260	2,073	1,232	2,105	1,205	2,139	1,177	2,172	1,149	2,206	1,121	2,241
90	1,395	1,937	1,369	1,966	1,344	1,995	1,318	2,025	1,292	2,055	1,266	2,085	1,240	2,116	1,213	2,148	1,187	2,179	1,160	2,211
95	1,418	1,930	1,394	1,956	1,370	1,984	1,345	2,012	1,321	2,040	1,296	2,068	1,271	2,097	1,247	2,126	1,222	2,156	1,197	2,186
100	1,439	1,923	1,416	1,948	1,393	1,974	1,371	2,000	1,347	2,026	1,324	2,053	1,301	2,080	1,277	2,108	1,253	2,135	1,229	2,164
150	1,579	1,892	1,564	1,908	1,550	1,924	1,535	1,940	1,519	1,956	1,504	1,972	1,489	1,989	1,474	2,006	1,458	2,023	1,443	2,040
200	1,654	1,885	1,643	1,896	1,632	1,908	1,621	1,919	1,610	1,931	1,599	1,943	1,588	1,955	1,576	1,967	1,565	1,979	1,554	1,991

1%	k=1		k=2		k=3		k=4		k=5		k=6		k=7		k=8		k=9		k=10	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
6	0,390	1,142	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
7	0,435	1,036	0,294	1,676	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8	0,497	1,003	0,345	1,489	0,229	2,102	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9	0,554	0,998	0,408	1,389	0,279	1,875	0,183	2,433	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10	0,604	1,001	0,466	1,333	0,340	1,733	0,230	2,193	0,150	2,690	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
11	0,653	1,010	0,519	1,297	0,396	1,640	0,286	2,030	0,193	2,453	0,124	2,892	-----	-----	-----	-----	-----	-----	-----	-----
12	0,697	1,023	0,569	1,274	0,449	1,575	0,339	1,913	0,244	2,280	0,164	2,665	0,105	3,053	-----	-----	-----	-----	-----	-----
13	0,738	1,038	0,616	1,261	0,499	1,526	0,391	1,826	0,294	2,150	0,211	2,490	0,140	2,838	0,090	3,182	-----	-----	-----	-----
14	0,776	1,054	0,660	1,254	0,547	1,490	0,441	1,757	0,343	2,049	0,257	2,354	0,183	2,667	0,122	2,981	0,078	3,287	-----	-----
15	0,811	1,070	0,700	1,252	0,591	1,465	0,487	1,705	0,390	1,967	0,303	2,244	0,226	2,530	0,161	2,817	0,107	3,101	0,068	3,374
16	0,844	1,086	0,738	1,253	0,633	1,447	0,532	1,664	0,437	1,901	0,349	2,153	0,269	2,416	0,200	2,681	0,142	2,944	0,094	3,201
17	0,873	1,102	0,773	1,255	0,672	1,432	0,574	1,631	0,481	1,847	0,393	2,078	0,313	2,319	0,241	2,566	0,179	2,811	0,127	3,053
18	0,902	1,118	0,805	1,259	0,708	1,422	0,614	1,604	0,522	1,803	0,435	2,015	0,355	2,238	0,282	2,467	0,216	2,697	0,160	2,925
19	0,928	1,133	0,835	1,264	0,742	1,416	0,650	1,583	0,561	1,767	0,476	1,963	0,396	2,169	0,322	2,381	0,255	2,597	0,196	2,813
20	0,952	1,147	0,862	1,270	0,774	1,410	0,684	1,567	0,598	1,736	0,515	1,918	0,436	2,110	0,362	2,308	0,294	2,510	0,232	2,174
21	0,975	1,161	0,889	1,276	0,803	1,408	0,718	1,554	0,634	1,712	0,552	1,881	0,474	2,059	0,400	2,244	0,331	2,434	0,268	2,625
22	0,997	1,174	0,915	1,284	0,832	1,407	0,748	1,543	0,666	1,691	0,587	1,849	0,510	2,015	0,437	2,188	0,368	2,367	0,304	2,548
23	1,017	1,186	0,938	1,290	0,858	1,407	0,777	1,535	0,699	1,674	0,620	1,821	0,545	1,977	0,473	2,140	0,404	2,308	0,340	2,479
24	1,037	1,199	0,959	1,298	0,881	1,407	0,805	1,527	0,728	1,659	0,652	1,797	0,578	1,944	0,507	2,097	0,439	2,255	0,375	2,417
25	1,055	1,210	0,981	1,305	0,906	1,408	0,832	1,521	0,756	1,645	0,682	1,776	0,610	1,915	0,540	2,059	0,473	2,209	0,409	2,362
26	1,072	1,222	1,000	1,311	0,928	1,410	0,855	1,517	0,782	1,635	0,711	1,759	0,640	1,889	0,572	2,026	0,505	2,168	0,441	2,313
27	1,088	1,232	1,019	1,318	0,948	1,413	0,878	1,514	0,808	1,625	0,738	1,743	0,669	1,867	0,602	1,997	0,536	2,131	0,473	2,269
28	1,104	1,244	1,036	1,325	0,969	1,414	0,901	1,512	0,832	1,618	0,764	1,729	0,696	1,847	0,630	1,970	0,566	2,098	0,504	2,229
29	1,119	1,254	1,053	1,332	0,988	1,418	0,921	1,511	0,855	1,611	0,788	1,718	0,723	1,830	0,658	1,947	0,595	2,068	0,533	2,193
30	1,134	1,264	1,070	1,339	1,006	1,421	0,941	1,510	0,877	1,606	0,812	1,707	0,748	1,814	0,684	1,925	0,622	2,041	0,562	2,160
31	1,147	1,274	1,085	1,345	1,022	1,425	0,960	1,509	0,897	1,601	0,834	1,698	0,772	1,800	0,710	1,906	0,649	2,017	0,589	2,131
32	1,160	1,283	1,100	1,351	1,039	1,428	0,978	1,509	0,917	1,597	0,856	1,690	0,794	1,788	0,734	1,889	0,674	1,995	0,615	2,104
33	1,171	1,291	1,114	1,358	1,055	1,432	0,995	1,510	0,935	1,594	0,876	1,683	0,816	1,776	0,757	1,874	0,698	1,975	0,641	2,080
34	1,184	1,298	1,128	1,364	1,070	1,436	1,012	1,511	0,954	1,591	0,896	1,677	0,837	1,766	0,779	1,860	0,722	1,957	0,665	2,057
35	1,195	1,307	1,141	1,370	1,085	1,439	1,028	1,512	0,971	1,589	0,914	1,671	0,857	1,757	0,800	1,847	0,744	1,940	0,689	2,037
36	1,205	1,315	1,153	1,376	1,098	1,442	1,043	1,513	0,987	1,587	0,932	1,666	0,877	1,749	0,821	1,836	0,766	1,925	0,711	2,018
37	1,217	1,322	1,164	1,383	1,112	1,446	1,058	1,514	1,004	1,585	0,950	1,662	0,895	1,742	0,841	1,825	0,787	1,911	0,733	2,001
38	1,227	1,330	1,176	1,388	1,124	1,449	1,072	1,515	1,019	1,584	0,966	1,658	0,913	1,735	0,860	1,816	0,807	1,899	0,754	1,985
39	1,237	1,337	1,187	1,392	1,137	1,452	1,085	1,517	1,033	1,583	0,982	1,655	0,930	1,729	0,878	1,807	0,826	1,887	0,774	1,970
40	1,246	1,344	1,197	1,398	1,149	1,456	1,098	1,518	1,047	1,583	0,997	1,652	0,946	1,724	0,895	1,799	0,844	1,876	0,749	1,956
45	1,288	1,376	1,245	1,424	1,201	1,474	1,156	1,528	1,111	1,583	1,065	1,643	1,019	1,704	0,974	1,768	0,927	1,834	0,881	1,902
50	1,324	1,403	1,285	1,445	1,245	1,491	1,206	1,537	1,164	1,587	1,123	1,639	1,081	1,692	1,039	1,748	0,997	1,805	0,955	1,864
55	1,356	1,428	1,320	1,466	1,284	1,505	1,246	1,548	1,209	1,592	1,172	1,638	1,134	1,685	1,095	1,734	1,057	1,785	1,018	1,837
60	1,382	1,449	1,351	1,484	1,317	1,520	1,283	1,559	1,248	1,598	1,214	1,639	1,179	1,682	1,144	1,726	1,108	1,771	1,072	1,817
65	1,407	1,467	1,377	1,500	1,346	1,534	1,314	1,568	1,283	1,604	1,251	1,642	1,218	1,680	1,186	1,720	1,153	1,761	1,120	1,802
70	1,429	1,485	1,400	1,514	1,372	1,546	1,343	1,577	1,313	1,611	1,283	1,645	1,253	1,680	1,223	1,716	1,192	1,754	1,162	1,792
75	1,448	1,501	1,422	1,529	1,395	1,557	1,368	1,586	1,340	1,617	1,313	1,649	1,284	1,682	1,256	1,714	1,227	1,748	1,199	1,783
80	1,465	1,514	1,440	1,541	1,416	1,568	1,390	1,595	1,364	1,624	1,338	1,653	1,312	1,683	1,285	1,714	1,259	1,745	1,232	1,777
85	1,481	1,529	1,458	1,553	1,434	1,577	1,411	1,603	1,386	1,630	1,362	1,657	1,337	1,685	1,312	1,714	1,287	1,743	1,262	1,773
90	1,496	1,541	1,474	1,563	1,452	1,587	1,429	1,611	1,406	1,636	1,383	1,661	1,360	1,687	1,336	1,714	1,312	1,741	1,288	1,769
95	1,510	1,552	1,489	1,573	1,468	1,596	1,446	1,618	1,425	1,641	1,403	1,666	1,381	1,690	1,358	1,715	1,336	1,741	1,313	1,767
100	1,522	1,562	1,502	1,582	1,482	1,604	1,461	1,625	1,441	1,647	1,421	1,670	1,400	1,693	1,378	1,717	1,357	1,741	1,335	1,765
150	1,611	1,637	1,598	1,651	1,584	1,665	1,571	1,679	1,557	1,693	1,543	1,708	1,530	1,722	1,515	1,737	1,501	1,752	1,486	1,767
200	1,664	1,684	1,653	1,693	1,643	1,704	1,633	1,715	1,623	1,725	1,613	1,735	1,603	1,746	1,592	1,757	1,582	1,768	1,571	1,779

1%	k=11		k=12		k=13		k=14		k=15		k=16		k=17		k=18		k=19		k=20		
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	
6	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
7	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
8	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
9	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
10	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
11	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
12	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
13	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
14	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
15	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
16	0,060	3,446	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
17	0,084	3,286	0,053	3,506	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
18	0,113	3,146	0,075	3,358	0,047	3,557	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
19	0,145	3,023	0,102	3,227	0,067	3,420	0,043	3,601	----	----	----	----	----	----	----	----	----	----	----	----	----
20	0,178	2,914	0,131	3,109	0,092	3,297	0,061	3,474	0,038	3,639	----	----	----	----	----	----	----	----	----	----	----
21	0,212	2,817	0,162	3,004	0,119	3,185	0,084	3,358	0,055	3,521	0,035	3,671	----	----	----	----	----	----	----	----	----
22	0,246	2,729	0,194	2,909	0,148	3,084	0,109	3,252	0,077	3,412	0,050	3,562	0,032	3,700	----	----	----	----	----	----	----
23	0,281	2,651	0,227	2,822	0,178	2,991	0,136	3,155	0,100	3,311	0,070	3,459	0,046	3,597	0,029	3,725	----	----	----	----	----
24	0,315	2,580	0,260	2,744	0,209	2,906	0,165	3,065	0,125	3,218	0,092	3,363	0,065	3,501	0,043	3,629	0,027	3,747	----	----	----
25	0,348	2,517	0,292	2,674	0,240	2,829	0,194	2,982	0,152	3,131	0,116	3,274	0,085	3,410	0,060	3,538	0,039	3,657	0,025	3,766	----
26	0,381	2,460	0,324	2,610	0,272	2,758	0,224	2,906	0,180	3,050	0,141	3,191	0,107	3,325	0,079	3,452	0,055	3,572	0,036	3,682	----
27	0,413	2,409	0,356	2,552	0,303	2,694	0,253	2,836	0,208	2,976	0,167	3,113	0,131	3,245	0,100	3,371	0,073	3,490	0,051	3,602	----
28	0,444	2,363	0,387	2,499	0,333	2,635	0,283	2,772	0,237	2,907	0,194	3,040	0,156	3,169	0,122	3,294	0,093	3,412	0,068	3,524	----
29	0,474	2,321	0,417	2,451	0,363	2,582	0,313	2,713	0,266	2,843	0,222	2,972	0,182	3,098	0,146	3,220	0,114	3,338	0,087	3,450	----
30	0,503	2,283	0,447	2,407	0,393	2,533	0,342	2,659	0,294	2,785	0,249	2,909	0,208	3,032	0,171	3,152	0,137	3,267	0,107	3,379	----
31	0,531	2,248	0,475	2,367	0,422	2,487	0,371	2,609	0,322	2,730	0,277	2,851	0,234	2,970	0,193	3,087	0,160	3,201	0,128	3,311	----
32	0,558	2,216	0,503	2,330	0,450	2,446	0,399	2,563	0,350	2,680	0,304	2,797	0,261	2,912	0,221	3,026	0,184	3,137	0,151	3,246	----
33	0,585	2,187	0,530	2,296	0,477	2,408	0,426	2,520	0,377	2,633	0,331	2,746	0,287	2,858	0,246	2,969	0,209	3,078	0,174	3,184	----
34	0,610	2,160	0,556	2,266	0,503	2,373	0,452	2,481	0,404	2,590	0,357	2,699	0,313	2,808	0,272	2,915	0,233	3,022	0,197	3,126	----
35	0,634	2,136	0,581	2,237	0,529	2,340	0,478	2,444	0,430	2,550	0,383	2,655	0,339	2,761	0,297	2,865	0,257	2,969	0,221	3,071	----
36	0,658	2,113	0,605	2,210	0,554	2,310	0,504	2,410	0,455	2,512	0,409	2,614	0,364	2,717	0,322	2,818	0,282	2,919	0,244	3,019	----
37	0,680	2,092	0,628	2,186	0,578	2,282	0,528	2,379	0,480	2,477	0,434	2,576	0,389	2,675	0,347	2,774	0,306	2,872	0,268	2,969	----
38	0,702	2,073	0,651	2,164	0,601	2,256	0,552	2,350	0,504	2,445	0,458	2,540	0,414	2,637	0,371	2,733	0,330	2,828	0,291	2,923	----
39	0,723	2,055	0,673	2,143	0,623	2,232	0,575	2,323	0,528	2,414	0,482	2,507	0,438	2,600	0,395	2,694	0,354	2,787	0,315	2,879	----
40	0,744	2,039	0,694	2,123	0,645	2,210	0,597	2,297	0,551	2,386	0,505	2,476	0,461	2,566	0,418	2,657	0,377	2,748	0,338	2,838	----
45	0,835	1,972	0,790	2,044	0,744	2,118	0,700	2,193	0,655	2,269	0,612	2,346	0,570	2,424	0,528	2,503	0,488	2,582	0,448	2,661	----
50	0,913	1,925	0,871	1,987	0,829	2,051	0,787	2,116	0,746	2,182	0,705	2,250	0,665	2,318	0,625	2,387	0,586	2,456	0,548	2,526	----
55	0,979	1,891	0,940	1,945	0,902	2,002	0,863	2,059	0,825	2,117	0,786	2,176	0,748	2,237	0,711	2,298	0,674	2,359	0,637	2,421	----
60	1,037	1,865	1,001	1,914	0,965	1,964	0,929	2,015	0,893	2,067	0,857	2,120	0,822	2,173	0,786	2,227	0,751	2,283	0,716	2,338	----
65	1,087	1,845	1,053	1,889	1,020	1,934	0,986	1,980	0,953	2,027	0,919	2,075	0,886	2,123	0,852	2,172	0,819	2,221	0,789	2,272	----
70	1,131	1,831	1,099	1,870	1,068	1,911	1,037	1,953	1,005	1,995	0,974	2,038	0,943	2,082	0,911	2,127	0,880	2,172	0,849	2,217	----
75	1,170	1,819	1,141	1,856	1,111	1,893	1,082	1,931	1,052	1,970	1,023	2,009	0,993	2,049	0,964	2,090	0,934	2,131	0,905	2,172	----
80	1,205	1,810	1,177	1,844	1,150	1,878	1,122	1,913	1,094	1,949	1,066	1,984	1,039	2,022	1,011	2,059	0,983	2,097	0,955	2,135	----
85	1,236	1,803	1,210	1,834	1,184	1,866	1,158	1,898	1,132	1,931	1,106	1,965	1,080	1,999	1,053	2,033	1,027	2,068	1,000	2,104	----
90	1,264	1,798	1,240	1,827	1,215	1,856	1,191	1,886	1,166	1,917	1,141	1,948	1,116	1,979	1,091	2,012	1,066	2,044	1,041	2,077	----
95	1,290	1,793	1,267	1,821	1,244	1,848	1,221	1,876	1,197	1,905	1,174	1,943	1,150	1,963	1,126	1,993	1,102	2,023	1,079	2,054	----
100	1,314	1,790	1,292	1,816	1,270	1,841	1,248	1,868	1,225	1,895	1,203	1,922	1,181	1,949	1,158	1,977	1,136	2,006	1,113	2,034	----
150	1,473	1,783	1,458	1,799	1,444	1,814	1,429	1,830	1,414	1,847	1,400	1,863	1,385	1,880	1,370	1,897	1,355	1,913	1,340	1,931	----
200	1,561	1,791	1,550	1,801	1,539	1,813	1,528	1,824	1,518	1,836	1,507	1,847	1,495	1,860	1,484	1,871	1,474	1,883	1,462	1,896	----

F.4. A görög betűk

<i>A</i>	α	<i>Alfa</i>	<i>N</i>	ν	<i>Nű</i>
<i>B</i>	β	<i>Béta</i>	<i>E</i>	ξ	<i>Kszí</i>
<i>Γ</i>	γ	<i>Gamma</i>	<i>O</i>	o	<i>Omikron</i>
<i>Δ</i>	δ	<i>Delta</i>	<i>Π</i>	π	<i>Pí</i>
<i>E</i>	ε	<i>Epszilon</i>	<i>P</i>	ρ	<i>Ró</i>
<i>Z</i>	ζ	<i>Zéta</i>	<i>Σ</i>	σ	<i>Szigma</i>
<i>H</i>	η	<i>Éta</i>	<i>T</i>	τ	<i>Tau</i>
<i>Θ</i>	θ	<i>Théta</i>	<i>Υ</i>	υ	<i>Üpszilon</i>
<i>I</i>	ι	<i>Ióta</i>	<i>Φ</i>	φ	<i>Fi</i>
<i>K</i>	κ	<i>Kappa</i>	<i>X</i>	χ	<i>Khí</i>
<i>Λ</i>	λ	<i>Lambda</i>	<i>Ψ</i>	ψ	<i>Pszí</i>
<i>M</i>	μ	<i>Mű</i>	<i>Ω</i>	ω	<i>Ómega</i>

Felhasznált irodalom

- Abraham B.- Ledolter J. [1983]: Statistical methods for forecasting. John Wiley & Sons. New York.
- Aczel A. D. [2002]: Complete business statistics. (The Irwin/McGraw-Hill series: operations and decision sciences) McGraw - Hill Higher Education. 5. ed. McGraw -Hill/Irwin. Boston.
- Almon, S. [1965]: The distributed lag between capital appropriations and expenditures. *Econometrica*. 33.
- Andreich Jenő [1937]: A konjunktúrakutatás módszerei. MTA.
- Alt F. F. [1942]: Distributed lags. *Econometrica*. Vol. 10.
- Arrow K. J. - Chenery H. B. - Minhas B. S. – Solow R. M. [1961]: Capital - labor substitution and economics efficiency. *The Review of Economics and Statistics*.
- Baczoni Pál [2007]: Egyszerűen Microsoft Office Excel 2003. Panem. Budapest.
- Balázs M. Mócsai Andrea-Csetényi Arthur [2003]: Kvantitatív technikák. II. Tankönyv. Zsigmond Király Főiskola. Budapest.
- Balogh Márta [2000]: Statisztikai ismeretek. Perfekt. Budapest.
- Barancsik János [2008]: Mikrogazdaságtan. PTE KTK Pécs.
- Bártfai Barnabás [2002]: Office XP. World 2002. Excel 2002. Power Point 2002. Outlook Access 2002. BBS-Info Kft. Budapest.
- Bartlett M. S. – Cox F. R. S. [1975]: The analysis of time series: theory and practice. Chapman and Hall. London.
- Bedő Zsolt - Rappai Gábor [2006]: Is there causal relationship between the value of the news and stock returns (társszerző:). *Hungarian Statistical review*. Special number 10.
- Berenson Mark L. – Levine David M. – Krehbiel Timothy C. [2006]: Basic business statistics: Concepts and applications. 10th ed. Pearson/Prentice Hall.
- Bertalanffy, L. [1938]: A Quantitative Theory of Organic Growth. (Inquiries on Growth Laws II.) *Human Biology*, 10..
- Bertalanffy, L. [1960]: Principles and theory of growth, in: *Fundamental Aspects of Normal and Malignant Growth*, W. W. Nowinski (ed), Amsterdam.
- Besenyei Lajos - Gidai Erzsébet - Nováky Erzsébet [1977]: Jövőkutatás, előrejelzés a gyakorlatban. KJK. Budapest.
- Besenyei Lajos - Gidai Erzsébet - Nováky Erzsébet [1982]: Előrejelzés. Megbízhatóság. Valóság. KJK. Budapest.
- Black K. [2006]: Business statistics (4. rev. ed.) John Wiley & Sons. New York. 2006.
- Borli Károly - Sipos Béla [1977]: Iparvállalati prognóziskészítés matematikai, statisztikai módszerrel. Közgazdasági és Jogi Könyvkiadó. Budapest.
- Box G. E. P. - Cox D. R [1964]: An analysis of transformations. *Journal of the Royal Statistical Society. Series B* 26. 2.
- Box G.E.P. - Jenkins G. M. [1970]: Time series analysis. Forecasting and control. Holden-Day. San Francisco. CA.
- Box, G.E.P. - Cox, D.R [1964].: An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B* 26. 2. sz. pp. 211-252.
- Box G. E. P.-Pierce A. [1970]: Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Association*. Vol. 65.
- Braudel F. [1972]: A történelem és a társadalomtudományok. A hosszú időtartam. Századok. 4-5. sz.
- Braudel F. [2004]: Anyagi kultúra, gazdaság és kapitalizmus XV-XVIII. század. 1. köt. A mindennapi élet struktúrái: a lehetséges és a lehetetlen 2. kiad. Budapest. Gutta Könyvkiadó.
- Breusch, T. S.; A. R. Pagan [1979]: Simple test for heteroscedasticity and random coefficient variation. *Econometrica (Econometric Society)* 47 (5).
- Bródy András [1983]: A lassuló idő. Közgazdasági és Jogi Könyvkiadó. Budapest.
- Bronstejn I. N. – Szemengyajev K. A. – Musiol G. – Mühling H. [2004]: Matematikai kézikönyv. Typotex Kiadó. Budapest.
- Chan N. H. [2002]: Time series. John Wiley. Canada.
- Cobb C.W. – Douglas P.H. [1928]: A theory of production. *American Economic Review*. Vol. 18.
- Cochrane - Orcutt. [1949]: Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association* 44, 32–61.

- Colin P. D. Birch [1999]: A New Generalized Logistic Sigmoid Growth Equation Compared with the Richards Growth Equation. *Annals of Botany* 83.
- Dagum, C. (1985). Analyses of income distribution and inequality by education and sex in Canada. In *Advances in Econometrics*, Vol. IV, R. L. Basmann and G. F. Rhodes, Jr., eds. JAI Press, Greenwich, CN.
- Demográfiai évkönyv. KSH. Budapest. 1993. – 2006.
- Descartes [1961]: Válogatott filozófiai művek. Akadémiai Kiadó, Budapest.
- Divatszociológia [1982]: Válogatta, szerkesztette, lektorálta: Klaniczay Gábor - S. Nagy Katalin. Mémbrán könyvek. 1. köt. I.-III. - A Tömegkommunikációs Kutatóközpont Kiadása. Budapest.
- Duijn J. J. Van [1982]: The long wave in economic life. London George Allen és Unwin. Forrester J. W. [1982]: Nach jeder depression ein neuer aufschwung? *Bild der Wissenschaft*. évi. 2. sz.
- Durbin J. and Watson, G. S. [1950]: Testing for serial correlation in least squares regression. I. *Biometrika* 37.
- Éltető Ödön – Meszéna György – Ziermann Margit [1982]: Sztochasztikus módszerek és modellek. KJK. Budapest 1982.
- Encyclopedia of Statistical Sciences, 16 Volume Set, [2006] 2nd Edition Samuel Kotz, Campbell B. Read, N. Balakrishnan, Brani Vidakovic ISBN: 978-0-471-15044-2. Hardcover 9686 pages January
- Evans James R. [2007]: Statistics, data analysis, and decision modeling. Pearson-Prentice Hall. New Jersey.
- Ezekiel M. – Fox K. A. [1970]: Korreláció- és regresszió-analízis. KJK. Budapest.
- Faragó Tamás [2007]: Történeti mutatószám az „Emberi Fejlődés” ábrázolására Magyarországon (1910-2001). Elemzési kísérlet. *Demográfia*. 50. évf. 2-3. sz.
- Farrar D. E. és Glauber R. R. [1967]: Multicollinearity in regression analysis: the problem revisited. *Review of Economics and Statistics*.
- Farnum Nicholas R. - Stanton LaVerne W. [1989]: Quantitative forecasting methods. Boston. PWS-Kent Pub. Co.. 1989. I-II.
- Fisher I. [1937]: Note on a short-cut method for calculating distributed lags. *International Statistical Bulletin*. Vol. 29.
- Fokasz Nikosz [2006]: Növekedési függvények, társadalmi diffúzió, társadalmi változás. *Szociológiai Szemle*. 3. sz.
- Freschl György [1982]: Bevezetés az időszori módszerek gyakorlatába. Statisztikai módszertani füzetek. KSH. Budapest.
- Gál P. – Moldicz Cs. – Novák T. [2004]: Gazdasági ciklusok és gazdaságpolitika a 21. század elején. *Fejlesztés és finanszírozás*. 4. sz.
- Gary Koop [2008]: A közgazdasági adatok elemzése. Osiris. Budapest.
- Gazdasági képlet-gyűjtemény. [1978]: Összeállította Káldor Mihály. Közgazdasági és Jogi Könyvkiadó. Budapest.
- Glejser H. [1969]: A new test for heteroscedasticity. *Journal of the American Statistical Association*. 1. sz.
- Godfrey, L. [1978]: Testing for multiplicative heteroscedasticity. *Journal of the American Statistical Association*. 8. sz.
- Goldfeld S. M. – Quandt R. E. [1965]: Some Tests for Homoscedasticity. *Journal of the American Statistical Association*, Vol. 60, No. 310 (Jun)
- Gompertz B. [1825]: On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*. Vol. 115 (1825)
- Gazdag László [1990]: A hosszú hullámok problémája (Az évszázados gazdasági ciklusok). *Gazdasági Fórum*. 3. sz.
- Greene, William H. [2003]: *Econometric analysis*. 5th ed. Pearson Education International. Upper Saddle River, N. J. Prentice Hall.
- Gujarati Damodar N. [2003]: *Basic econometrics*. McGraw-Hill Higher Education.
- Hajdu Ottó - Kertész László - Sipos Béla [1984]: A munkabérek regressziós elemzése és a koncentráció vizsgálata I. rész. *Statisztikai Szemle*. A Központi Statisztikai Hivatal folyóirata. 62. évf. 4. sz.
- Hajdu Ottó - Kertész László - Sipos Béla [1984]: A munkabérek regressziós elemzése és a koncentráció vizsgálata. II. rész. *Statisztikai Szemle*. A Központi Statisztikai Hivatal folyóirata. 62. évf. 5. sz.

- Hajdu O. - Herman S. - Pintér J. – Rappai G. - Rédey K. [1994-95]: Statisztika I-II. JPTE Kiadó. Pécs.
- Hajdu Ottó - Herman Sándor - Pintér József - Rédey Katalin [1987]: Ökonometriai alapvetés. Tankönyvkiadó. Budapest 1987.
- Hajdu Ottó – Hunyadi László [1995]: Varianciafelbontás: előfeltevések és következtetések. Szigma. 1-2. sz.
- Hajdu Ottó [1997]: A szegénység mérőszámai. KSH Könyvtár és Dokumentációs Szolgálat. Budapest.
- Hajdu Ottó [2003]: Többváltozós statisztikai számítások. KSH. Budapest.
- M. J. Harrison [1982]: Tables of critical values for a Beta approximation to Szroeter's statistic for testing for heteroscedasticity. Oxford Bulletin of Economics and Statistics. 2 sz.
- Hartmann Nicolai [1972]: Lételméleti vizsgálódások. Budapest.
- Harvey G. [2000]: Excel 2000 for Windows for dummies. Kossuth Kiadó. Budapest.
- Harvey A. C. [1976]: Estimating regression models with multiplicative heteroscedasticity. 3. sz.
- Haustein H. D. [1972]: Prognózismódszerek a szocialista gazdaságban. Közgazdasági és Jogi Könyvkiadó.
- Heller Farkas [1943]: A közgazdasági elmélet története. Gergely R. Budapest.
- Herman Sándor [1985]: A szezonáltság - vizsgálat statisztikai módszerei. IGK. Időszerű gazdaságirányítási kérdések. Prodinform Műszaki tanácsadó vállalat. 4. szám.
- Herman Sándor – Varga József [1983]: A szezonális áttrendezősség vizsgálata. Statisztikai Szemle. 61. évf. 6. sz.
- Hill R. C., Griffiths W. E. Lim G. C. [2008]: Principles of Econometrics, 3rd Edition, Wiley
- Hoover jr. Edgar Malone [1936]: The measurement of industrial localization. Review of Economics and Statistics. Vol. 18. No.
- Hoós J. [2003]: Konjunktúra- és piackutatás. Aula. Budapest.
- Hubbert M. K. [1956]: Nuclear energy and the fossil fuels M. K. Hubbert, Presented before the Spring Meeting of the Southern District, American Petroleum Institute, Plaza Hotel, San Antonio, Texas, March 7-8-9. 1956.
- Hunyadi László [1980]: Megosztott késleltetésű ökonometriai modellek. Szigma 1-2. sz.
- Hunyadi László [2004]: A logisztikus függvény és a logisztikus eloszlás. Statisztikai szemle. 10 – 11. sz.
- Hunyadi László [2006]: A heteroszkedaszticitásról egyszerűbben. Statisztikai szemle. 1. sz.
- Hunyadi László [2000]: A mintavétel alapjai. BKÁE-Számalk. Budapest 2000.
- Hunyadi László [2001]: Statisztikai következtetésemélet közgazdászoknak. KSH. Budapest 2001.
- Hunyadi László [2002]: Grafikus ábrázolás a statisztikában. Statisztikai Szemle. 1. sz.
- Hunyadi László [2005]: A hányadosbecslés néhány tulajdonsága és egy új becslőfüggvénye. Statisztikai Szemle. 2. sz.
- Hunyadi László – Vita László [2002]: Statisztika közgazdászoknak. Budapest. KSH.
- Hunyadi László – Vita László [2008]: Statisztika I-II. Aula Kiadó.
- Iványi Attila Szilárd [1984]: Termékstratégia, gyártáspolitikai, műszaki fejlesztés. Műszaki könyvkiadó. Budapest.
- Jánosa András [2005]: Adatelemzés számítógéppel. Perfekt RT. Budapest.
- Jánossy Ferenc [1966]: A gazdasági fejlődés trendvonalai és a helyreállítási periódus. Közgazdasági és Jogi Könyvkiadó. Budapest.
- Jánossy Ferenc [1975]: A gazdasági fejlődés trendvonaláról. (második bővített kiadás) Magvető Könyvkiadó. Budapest.
- Johnson, Norman Lloyd [1949]: Systems of frequency curves generated by methods of translation, Biometrika, Vol. 36, No. 1-2.
- Juglar C. [1862]: Des crises commerciales et leur retour periodique en France, en Augleterre et aux Etats Unis. Franklin. Páris.
- Kádas Kálmán [1941]: Áralakulás irányítása és a piaci egyensúly. Közgazdasági könyvtár. XXV. kötet. Budapest.
- Kádas Kálmán [1944]: Az emberi munka termelékenységének statisztikai vizsgálata a magyar gyáripárban. (A Cobb-Douglas féle statisztikai törvény kiegészítése) Magyar Statisztikai Szemle. 7-8. sz.
- Kehl Dániel – Rappai Gábor [2006]: Mintaelemszám tervezése Likert-skálát alkalmazó lekérdezésekben. Statisztikai Szemle. 84. évfolyam. 9. sz.
- Kehl Dániel – Sipos Béla [2007a]: A bőr és a különböző bőrtermékek termelői árainak hosszú távú tendenciái az Amerikai Egyesült Államokban. I. rész. Bőr- és Cipőtechnika – Piac. LVII. évf. 1. sz.

- Kehl Dániel – Sipos Béla [2007b]: A bőr és a különböző bőrtermékek termelői árainak hosszú távú tendenciái az Amerikai Egyesült Államokban. II. rész. Bőr- és Cipőtechnika – Piac. LVII. évf. 2. sz.
- Kehl Dániel – Sipos Béla: [2007c]: Évszázados trendek és hosszú ciklusok az Amerikai Egyesült Államokban, Kínában és a világgazdaságban. Hitelintézeti Szemle. Hatodik Évfolyam. 3. szám.
- Kehl Dániel – Sipos Béla [2007d]:3-12. A gazdasági növekedés ciklikus változása az USA-ban. Fejlesztés és Finanszírozás. 4. sz.
- Kehl Dániel – Sipos Béla [2007e] 4. Secular Trends and Long Cycles in the US Economy. Development and Finance. 4. sz.
- Kehl Dániel – Dr. Sipos Béla [2009]: A telítődési, a logisztikus és életgörbe alakú trendfüggvények becslése Excel parancsfájl segítségével. Statisztikai Szemle. 87. évf. 4. sz.
- Kerékgyártó Györgyné – Mundruczó György [1995]: Statisztikai módszerek a gazdasági elemzésben. 2. átdolgozott kiadás. Aula Kiadó. Budapest.
- Kerékgyártó Györgyné – Mundruczó György – Sugár András [2001]: Statisztikai módszerek és alkalmazásuk a gazdasági, üzleti elemzésekben. Aula Kiadó. Budapest 2001.
- Kindler József-Papp Ottó [1977]: Komplex rendszerek vizsgálata. Összemérési módszerek. Műszaki Könyvkiadó.
- King Maxvel L. [1981]: A note on szroeter's bound test. Oxford Bulletin of Economics and Statistics. 3. sz.
- Kiss Tibor [1985]: Koyck és Solow modelljeinek felhasználása a döntéselőkészítésben. Statisztikai Szemle. 10. sz.
- Kiss Tibor - Sipos Béla - Szentmiklósi Miklós [1995]: Az üzleti ciklus modellezése és prognosztizálása EXPS- programmal. Statisztikai Szemle. A Központi Statisztikai Hivatal folyóirata. 73. évf. 8 - 9. sz.
- Kiss Tibor – Kruzslicz Ferenc - Sipos Béla - Szentmiklósi Miklós [1997]: A SABL- eljárás felhasználása elemzésre és prognosztizálásra. Statisztikai Szemle. A Központi Statisztikai Hivatal folyóirata. 75. évf. X. sz.
- Kiss Tibor – Sipos Béla [1998]: REGAL: Expert system for multiple linear regression analysis. Hungarian Statistical Review. 76. Volume. Special Number.
- Kiss Tibor – Sipos Béla [2000]: EXPS for Windows, a software application. Hungarian Statistical Review. 78. Volume. Special Number.
- Kitchin J. [1923]: Cycles and trends in economic factors. Review of Economic Statistics 5. évf. 1. sz.
- Kiss Tibor [1985]: Koyck és Solow modelljeinek felhasználása a döntéselőkészítésben. Statisztikai Szemle. 10. sz.
- Koyck, L. M. [1954]: Distributed lags and investment analysis. Amsterdam. North-Holland.
- Knüsel L. [1998]: On the accuracy of statistical distributions in Microsoft Excel 97. Computational Statistics and Data Analysis 26.
- Knüsel L. [2002]: On the reliability of Microsoft Excel XP for statistical purposes. Computational Statistics and Data Analysis 39.
- Knüsel L. [2005]: On the accuracy of statistical distributions in Microsoft Excel 2003. Computational Statistics and Data Analysis 48.
- Komjáti Zoltán [1975]: A termelési kapacitás és az átbocsátóképesség kihasználásának összefüggései az iparvállalatoknál. Ipari és Építőipari Statisztikai Értesítő.
- Кондратьев Н. Д. [1925]: Большие циклы конъюнктуры. Вопросы конъюнктуры.-М. 1925. - Т. 1. вып. 1.
- Kondratieff N. D. [1926]: Die langen wellen der konjunktur. Archiv für Sozialwissenschaft und Sozialpolitik. Berlin. 56. köt.
- Kondratieff N. D. [1979]: The long waves in economic life. Review. 2. évf. 4. sz.
- Kondratyev N. D. [1980]: A gazdasági fejlődés hosszú hullámai. Történelmi Szemle. 22. évf. 2. sz.
- Kovács Péter – Petres Tibor – Tóth László [2004]: Adatállományok redundanciájának mérése. Statisztikai Szemle. KSH. 6-7. sz.
- Kovács Péter [2008]: A multikollinearitás vizsgálata lineáris regressziós modellekben. Statisztikai szemle. 1. sz.
- Peter Kovacs – Tibor Petres – Laszlo Toth [2005]: A new measure of multicollinearity in linear regression models. International Statistical Review Volume 73 Number 3. International Statistical Institute. Voorburg. The Netherlands.

- Kovács Péter [2008a]: A multikollinearitás vizsgálata lineáris regressziós modellekben. Statisztikai szemle. 1. sz.
- Kovács Péter [2008b]: A statisztikaoktatás módszertanának modernizálása? Statisztikai Szemle. A Központi Statisztikai Hivatal folyóirata. 86. évf. 12. sz. 1143-1157.
- Knut Sydsaeter - Peter I. Hammond [2006]: Matematika közgazdászoknak. 2., jav. kiad. Budapest. Aula.
- Körösi Gábor-Mátyás László-Székely István [1990]: Gyakorlati ökonometria. KJK. Budapest.
- Kotler P. – Keller K. L. [2006]: Marketing – menedzsment. Akadémiai Kiadó. Budapest.
- Kovalcsik Géza [2000]: Excel 2000. ComputerBooks. Budapest.
- Kövér György [1988]: N. D. Kondratyev és a gazdasági konjunktúra nagy ciklusai. Magyar Filozófiai Szemle. 5-6. sz.
- Köves Pál [1981]: Indexelmélet és közgazdasági valóság Akadémiai Kiadó. Budapest 1981.
- Köves Pál - Párniczky Gábor [1981]: Általános statisztika I.-II. KJK. Budapest 1981.
- Krekó Béla [1966]: Lineáris programozás. KJK. Budapest.
- Krekó Béla [1966]: Mátrixszámítás. KJK. Budapest.
- Kristó Zoltán [1979]: Termelési függvények a gazdasági elemzésben. Ökonometriai füzetek 16. KSH. Budapest.
- Kruszlicz Ferenc-Kiss Tibor-Sipos Béla: SABL Decomposition of Time Series ©1997 JPTE Version 2.1.)
- Kuznets S. [1930]: Secular movements in production and prices. Houghton Mifflin Company. Boston és New York.
- Labrousse E. [1984]: Esquisse du mouvement des prix et revenus en France au XVIIIème siècle. 2 Vols. Edition des Archives Contemporaines. Paris.
- Laherrère J. H. [2000]: Learn Strengths, Weaknesses to Understand Hubbert curve. Oil and Gas Journal. April 17. <http://dieoff.org/page191.htm> (Elérés dátuma: 2009. január 28.)
- Lénárt Imre – Rappai Gábor [2001]: Néhány gondolat a varianciabecslés hibahatáráról. Statisztikai Szemle. 7. sz.
- Levenbach H. - Cleary J. P. [1982]: The beginning forecaster: The forecasting process through data analysis. Lifetime Learning Publications. Belmont. Kalifornia.
- Liao C. Y. - Podrázský V. V.- Liu G. B.[2003]: Diameter and height growth analysis for individual White Pine trees in the area of Kostelec nad Černými lesy. Journal of Forest Science. 49. (12). 544-551.
- Lilien, Gary L., Kotler, P. [1983]: Marketing Decision Making Harper & Row, Publishers.
- Maddala G. S. [2004]: Bevezetés az ökonometriába. Budapest. Nemzeti Tanankönyvkiadó Rt.
- Mátyás Antal [1973]: A modern polgári közgazdaságtan története. Közgazdasági és Jogi Könyvkiadó. Budapest.
- S. Makridakis-S. C. Wheelwright- V. Mcgee [1983]: Forecasting: methods and applications. 2. Edition. John Wiley and Sons. Inc. New York.
- S. Makridakis -S. C. Wheelwright- R. J. Hyndman. [1998]: Forecasting. John Wiley and Sons. Inc. New York.
- McCullough B.D. - Wilson B. [1999]: On the accuracy of statistical procedures in Microsoft EXCEL 97. Computational Statistics and Data Analysis 31. 27–37.
- McCullough B.D. - Wilson B. [2002]: On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP. Computational Statistics and Data Analysis 40. 713–721.
- Mellár Tamás - Rappai Gábor [2001]: Money Supply, GDP and Inflation: The Dynamic Econometric Analysis of Macro-Equilibrium (társszerző:). Hungarian Statistical Review. Special number 6.
- Meszéna György - Ziermann Margit [1981]: Valószínűségelmélet és matematikai statisztika. KJK. Budapest 1981.
- Mitscherlich E. A. [1919]: Das Gesetz des Pflanzenwachstums. Landwirtsch. Jahrb 53. 167-182.
- Molnár Gyöngyvér-Capó Benő [2003]: A képességek fejlődésének logisztikus modellje. Iskolakultúra. 2. sz.
- Mundruczó György [1981]: Alkalmazott regressziószámítás. Akadémiai Kiadó. Budapest. 1981.
- Mundruczó György [1982]: A minőségi ismérvek közötti kapcsolatok vizsgálata. I.-II. Statisztikai Szemle 60. évf. 6.sz. 635-648. és 7. sz. 730-737.
- Mundruczó György [1998]: Útmutató a statisztikai modellezéshez. Aula Kiadó. Budapest.
- Naisbitt John - Aburdene Patricia [1991]: Megatrendek 2000. Tíz új irányzat a kilencvenes években. Budapest. OMIKK.

- John C. Nash [2008] Teaching statistics with Excel 2007 and other spreadsheets. Computational Statistics and Data Analysis (article in press)
- Nerlove, Marc, [1972]: Lags in economic behavior. *Econometrica*. Econometric Society. vol. 40(2), March.
- Nováky Erzsébet [Szerk.] [1992]: Jövőkutatás. BKE. Budapest.
- Nyitrai Ferencné - Rédey Katalin [1974]: Statisztika III. (Korszerű statisztikai módszerek és alkalmazásuk a gyakorlati közgazdasági munkában). Tankönyvkiadó. Budapest.
- Park R. E. [1966]: Estimation with heteroscedastic error terms. *Econometrica*. vol. 34. no. 4. okt.
- Pawlowski Z. [1970]: Ökonometria KJK. Budapest.
- Pearson K. [1895]: Contributions to the mathematical theory of evolution. II: Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London*. 186.
- Pearson K. [1905]: Das fehlergesetz und seine verallgemeinerungen durch fechner und Pearson. A Rejoinder. *Biometrika*. 4.
- Pella JS and PK Tomlinson [1969]: A generalised stock-production model. *Bull. IATTC* 13.
- Pétery Kristóf [2003]: Táblázatkezelés Excel 2002. Kossuth Kiadó. Budapest.
- Petres Tibor-Tóth László [2008]: Statisztika. KSH.
- Pintér József [1991]: A heteroszkedaszticitás diagnosztizálása. *Statisztikai Szemle*. 69. évf. 1. szám.
- Pintér József [1987]: A termelési függvények vállalati alkalmazásai. *Statisztikai Szemle*. 2. sz.
- Pintér József [2000]: Bevezetés a statisztika módszereibe. Pécsi Tudományegyetem. Pécs. 2000.
- Pintér József - Rappai Gábor [2001]: A mintavételi tervek készítésének néhány gyakorlati megfontolása. *Marketing & Menedzsment*. 5. sz.
- Pintér József [2007]: A spektrálanalízisről. *Statisztikai Szemle*. 85. évf. 2. sz.
- Pintér József – Rappai Gábor (szerkesztő) [2007]: *Statisztika*. Pécsi Tudományegyetem. Közgazdaságtudományi Kar. Pécs.
- Pusztai L. [1987]: Gazdasági ciklus és bűnözés. *Belügyi Szemle*. 9. sz.
- Ramanathan Ramu [2003]: Bevezetés az ökonometriába alkalmazásokkal. Budapest. Panem.
- Raymond Pearl - Lowell J. Reed [1920]: On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Sciences*. Volume 6. June 15. Number 6.
- Raymond Pearl [1929]: *The Biology of Population Growth*. Howard Woolston. The American Journal of Sociology, Vol. 35, No. 3 (Nov.)
- Richards, F. J. [1959] A flexible growth function for empirical use. *Journal of Experimental Botany*. Volume 10. Number 2.
- Rappai Gábor [2001]: Üzleti statisztika Excellel. Központi Statisztikai Hivatal. Budapest.
- Rappai Gábor [2003]: Üzleti statisztika: új tudomány vagy marketing-fogás? *Statisztikai Szemle*. 5.
- Rappai Gábor szerk. [2007]: Egy életpálya három dimenziója. Tanulmánykötet Pintér József emlékére. PTE KTK.
- Rappai Gábor [2008]: Gondolatok a gazdaságtudományi képzési területen folyó statisztikaoktatásról. *Statisztikai Szemle*. A Központi Statisztikai Hivatal folyóirata. 86. évf. 9. sz. 829-849.
- Richards F. J. [1959] A flexible growth function for empirical use. *J. Exp. Bot.* 10.
- Savin, N. E.-White K. J. [1977]: The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. *Econometrica*, Vol. 45, No. 8. 1989-1996.
- Schumpeter I. A. [1939]: *Business cycles: A theoretical, historical and statistical analysis of the capitalist process*. New York and London: McGraw-Hill Book Co. Inc.. 1st ed. 2 vols.
- Simiand F. [1932]: *Le salaire, l'évolution sociale et la monnaie. Essai de théorie expérimentale du salaire*. 3 Vols. F. Alcan. Paris.
- Sipos Béla [1982]: Termelési függvények - vállalati prognózisok. Közgazdasági és Jogi Könyvkiadó. Budapest.
- Sipos Béla [1982]: Iparvállalati árprognózisok. PRODINFORM Műszaki Tanácsadó Vállalat. Időszerű Gazdaságirányítási Kérdések. 4. sz. (2. változatlan kiadás 1984.)
- Sipos Béla [1982]: Termelési függvények - vállalati prognózisok. Közgazdasági és Jogi Könyvkiadó. Budapest.
- Sipos Béla [1985]: Vállalati árelőrejelzések. Közgazdasági és Jogi Könyvkiadó. Budapest.
- Sipos B. 1986: A Kondratyev-ciklus empirikus vizsgálata és prognosztizálása. *Statisztikai Szemle*. A Központi Statisztikai Hivatal folyóirata. 64. évf. 12. sz.

- Sipos Béla [2003]: Vállalati prognosztika. (Elmélet – Módszertan - Szoftverek) PTE Kiadó. Pécs.
- Solow R. M.[1960]: On a family of lag distributions. *Econometrica*.
- Solow [1957]: Technical change and the aggregate production function. *Review of Economics and Statistics* 39 (August 1957).
- Spiegel Murray R. [1995]: Statisztika: Elmélet és gyakorlat: SI mértékegységekkel. Schaum – könyvek. Panem Kft. - McGraw-Hill.
- Statisztikai idősorok a Kínai Népköztársaságról. [1986]: Központi Statisztikai Hivatal. Budapest.
- Szilágyi György [2002]: Gondolatok a statisztika szakmai etikájáról. *Statisztikai Szemle*. 3. sz.
- Szroeter Jerzy [1978]: A class of parametric tests for heteroscedasticity in linear econometric models. *Econometrica*. 46. évf. 6. sz.
- Stuart A. és Ord J. K. [1994]: Kendall's advanced theory of statistics. Volume 1. Distribution Theory. Sixth Edition. Edward Arnold. London.
- Szentmiklósi Miklós [2000]: Pénzügyi előrejelzési modellek készítésének néhány elméleti és gyakorlati kérdése. Osiris Kiadó.
- Szentmiklósi Miklós – Rédey Katalin [2007]: Arima modellek alkalmazása idősorelemzésére és előrejelzésére. In.: Rappai Gábor szerk. [2007]: Egy életpálya három dimenziója. Tanulmánykötet Pintér József emlékére. PTE KTK.
- Theil H. [1961]: *Economic forecasts and Policy*. 2nd Edition. Amsterdam: North Holland.
- Theiss Ede [1943]: Konjunktúrakutatás. A Mérnöki Továbbképző Intézet kiadványai. 15. kötet. 11. füzet. Budapest.
- Theiss Ede szerk. [1958]: *Korreláció és trendszámítás*. Közgazdasági és Jogi Könyvkiadó. Budapest.
- Thomopoulos, N. T. [1980]: *Applied Forecasting Methods*. Prentice-Hall, Englewood Cliffs, NJ, p. 370.
- Töröcsik Mária [2003]: Fogyasztói magatartás. *Trendek. Új fogyasztói csoportok*. KJK-Kerszöv. Budapest.
- Vágási Mária [2001]: *Újtermék – marketing*. Nemzeti Tankönyvkiadó. Budapest.
- Valkovics Emil [2001]: A Gompertz - függvény felhasználási lehetőségei a demográfiai modellezésben. *Statisztikai szemle*. 79. évf.
- Varga József [2001]: *A valószínűségelmélet alapjai*. Pécsi Tudományegyetem. Pécs. 2001.
- Verhulst, P. F. [1838]: Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathematique et Physique*. 10. szám. évf.
- Vincze István [1975]: *Matematikai statisztika ipari alkalmazásokkal*. Műszaki Könyvkiadó. Budapest 1975.
- Világ gazdasági idősorok 1860–1960. szerk.: Kenessey Zoltán [1965]: Központi Statisztikai Hivatal. Közgazdasági és Jogi Könyvkiadó. Budapest.
- Yalta A.T. [2008]: The accuracy of statistical distributions in Microsoft® Excel 2007. *Computational Statistics and Data Analysis* (article in press)
- Yule G. U. – Kendall M. G. [1964]: *Bevezetés a statisztika elméletébe*. KJK. Budapest. 1964.
- Verhulst, P. F. [1838]: Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathematique et Physique*, 10.
- Vető Istvánné [1980]: A dinamika vizsgálata autoregresszív és osztott késleltetésű modellekkel. KSH Módszertani Főosztály. Bp.
- Xinyou Yin-Jan Goudriaan-Egbert A. Lantinga - Jan Vos-Huub J. Spiertz [2003]: A Flexible Sigmoid Function of Determinate Growth. *Annals of Botany*.
- http://wapedia.mobi/en/Generalised_logistic_function (Elérés dátuma: 2009. január 28.)
- <http://www.bioss.ac.uk/smart/unix/mgrow/slides/slide02.htm> (Elérés dátuma: 2009. január 28.)
- http://www.horticultureandlandscape.rdg.ac.uk/hlm_richards.htm (Elérés dátuma: 2009. január 28.)
- Wessa P., (2009), ARIMA Forecasting (v1.0.5) in Free Statistics Software (v1.1.23-r7), Office for Research Development and Education, URL http://www.wessa.net/rwasp_arimaforecasting.wasp/