



Néhány betegség statisztikai adatainak időszori elemzése

Doktori (PhD) értekezés

Fazekasné Kis Mária

Debreceni Egyetem
Debrecen, 2004

Ezen értekezést a Debreceni Egyetem TTK Matematika- és Számítástudomány Doktori Iskola Informatika programja keretében készítettem a Debreceni Egyetem TTK doktori (PhD) fokozatának elnyerése céljából.

Debrecen, 2004.

.....
dr. Fazekasné Kis Mária

Tanusítom, hogy Dr. Fazekasné Kis Mária doktorjelölt 1997-2004 között a fent megnevezett Doktori Iskola Informatika programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javaslom.

Debrecen, 2004.

.....
Prof. Dr. Arató Mátyás

TARTALOMJEGYZÉK

<u>1. BEVEZETÉS, CÉLKITŰZÉSEK</u>	1
<u>2. IDŐSOROK JELLEMZŐINEK ÁTTEKINTÉSE</u>	4
<u>2.1. Sztochasztikus folyamatok jellemzői</u>	4
<u>2.2. Empirikus kifejezések</u>	9
<u>2.3. Kétdimenziós stacionárius folyamatok</u>	12
<u>2.4. Speciális idősorok</u>	14
<u>2.5. Idősorok komponensei</u>	15
<u>2.6. Sztochasztikus modellek megadása</u>	21
<u>2.6.1. Sztochasztikus modellek azonosítása</u>	21
<u>2.6.2. Sztochasztikus modell paramétereinek becslése</u>	27
<u>2.6.3. Sztochasztikus modellek ellenőrzése</u>	30
<u>2.7. AR(1) együttható maximum likelihood becslése időben diszkrét esetben</u>	32
<u>2.8. AR(1) konfidencia intervallumai becsléseinek összehasonlításai</u>	34
<u>2.9. Számítógépes alkalmazások irodalmának összefoglalása</u>	36
<u>2.9.1. ARIMA modellek alkalmazásai</u>	36
<u>2.9.2. Szezonális idősorok alkalmazásai</u>	38
<u>3. SZÁMÍTÓGÉPES ALKALMAZÁSOK</u>	40
<u>3.1. Halálozási arányok feldolgozása</u>	40
<u>3.1.1. Vizsgálati módszerek</u>	40
<u>3.1.2. Eredmények</u>	41
<u>3.1.2.1. Cerebrovasculáris betegség</u>	41
<u>3.1.2.2. Isémiás szívbetegség</u>	49
<u>3.1.2.3. Krónikus májbetegség</u>	51
<u>3.1.2.4. Légzőszervi betegségek</u>	54
<u>3.1.2.5. Méhnyakrák</u>	56
<u>3.1.2.6. Mellrák</u>	61
<u>3.1.2.7. Emésztőszervi daganatok</u>	64
<u>3.1.2.8. Légzőszervi daganatok</u>	66

TARTALOMJEGYZÉK

3.1.3. Következtetések a halálozási arányok elemzéseiből	68
3.2. Gyermekkori lymphoid leukaemiás betegek adatainak feldolgozása	72
3.2.1. A betegek születési dátumának elemzése	72
3.2.2. A betegség diagnosztizálási dátumának elemzése	76
3.2.3. Következtetések a lymphoid leukaemiás betegek adatainak vizsgálatából	81
4. ÚJ KUTATÁSI EREDMÉNYEK	82
5. ÖSSZEFOGLALÁS	84
5.1. Idősorok	84
5.2. Számítógépes alkalmazások	84
5.2.1. Halálozási arányok elemzése	84
5.2.1.1. Eredmények	85
5.2.2. Lymphoid leukaemiás betegek adatainak feldolgozása	88
6. SUMMARY	90
7. KÖSZÖNETNYILVÁNÍTÁS	95
8. IRODALOM	96
9. DISSZERTÁCIÓHOZ KAPCSOLÓDÓ PUBLIKÁCIÓK	111
10. DISSZERTÁCIÓ TÉMAKÖRÉN KÍVÜLI PUBLIKÁCIÓK	113
10.1. Jegyzetek	113
10.2. Idegen nyelvű közlemények	113
10.3. Magyar nyelvű közlemények	114
10.4. Magyar nyelvű előadások	114
10.5. Poszter	115
10.6. Egyéb	115

1. BEVEZETÉS, CÉLKITŰZÉSEK

Az informatika és a matematika fejlődése alapvetően meghatározza más tudományágak fejlődését. Különösen vonatkozik ez az élettelen természettudományokra. Az ókortól napjainkig a fizika, kémia és más természettudományok fejlődésére jelentős hatással voltak a matematikai módszerek kiteljesedése. Az utóbbi évtizedekben tanúi vagyunk annak, hogy a matematikai módszerek az élő természettudományok fejlődését is jelentősen befolyásolják. Sőt az utóbbi években a matematikai módszereket a társadalomtudomány területén is alkalmazzák (közgazdaságtan). E tudományterületek fejlődése mögött gyakran valamilyen új matematikai módszer bevezetése áll. A számítógépek minőségi fejlődésével lehetővé váltak a számítás igényes matematikai módszerek gyors feldolgozásai. Tudományos területeken, a napi munkában, sőt egyre több háztartásban is az informatikai alkalmazások ma már nélkülözhetetlenek.

Az orvostudományi adatok feldolgozásához már régóta alkalmaznak különböző matematikai, illetve statisztikai módszereket. A nagy tömegű adatok gyors, hatékony elemzését támogatta a számítógépek teljesítményének fokozódása és a megfelelő statisztikai programcsomagok elérhetővé válása. Az elmúlt évtizedben a világ fejlett országaiban az időszori modelleket is elkezdték alkalmazni olyan problémák vizsgálatára, melyeket a korábban alkalmazott módszerekkel nem tudtak vizsgálni. Angliában, Spanyolországban, USA-ban és más

országokban is valóságos tudományos iskolák alakultak, melyek ezeket a módszereket alkalmazzák.

Magyarországon a statisztikai módszerek alkalmazása az orvostudomány területein is évtizedekre tekint vissza, az időszori modellek alkalmazásáról viszont eddig keveset tudunk.

Az értekezésem az időszori modellek orvostudományi alkalmazásait mutatja be. Az értekezés két összefüggő témakörrel foglalkozik. Az értekezés első részében az idősorok jellemzőivel, sajátosságaival, törvényszerűségeivel foglalkozom. Idősornak nevezzük az időben diszkrét $\xi_{t_1}, \xi_{t_2}, \dots$ ($t_1 < t_2 < t_3 < \dots < t_n < \dots$) valószínűségi változók összességét, ahol a ξ_{t_i} valószínűségi változók nem függetlenek.

A munkám második részében ezekkel a módszerekkel olyan humán megbetegedéseket vizsgáltam, melyek oktana komplex és részben tisztázatlan. Az időszori elemzések célja az volt, hogy vizsgáljam a környezeti hatások szerepét ezeknek a betegségeknek a kialakulásában.

Célkitűzések:

1. A magyarországi lakosság körében előforduló nem fertőző megbetegedések (*cerebrovasculáris betegség, isémiás szívbetegség, krónikus májbetegség (májzsugor), légzőszervi betegségek (bronchitis, tüdő emphysema, asztma)*) okozta halálozási arányok idősor analízis módszereivel történő vizsgálata.
2. A magyarországi lakosság körében előforduló tumoros megbetegedések a (*méhnyakrák, mellrák, emésztőszervi daganatok (gyomor- és vastagbélrák), légzőszervi daganatok (légcső-, hörgő-, tüdőrák)*) okozta halálozási arányok idősor analízis módszereivel történő vizsgálata.

3. A környezeti tényezők szerepének kimutathatóságának vizsgálata a fenti betegségek kialakulásában és időbeli lefolyásában.
4. A vizsgált betegségekre kapott sztochasztikus kifejezések AR(1) együtthatói konfidencia intervallumainak elemzése.
5. Az *akut lymphoid leukaemiás (ALL)* gyermekek születési dátumának és a betegség diagnosztizálási dátumának elemzése szezonális idősorokkal.

Technikai jellegű megjegyzés: az ábrák, a táblázatok és a képletek számozása folyamatos az egész dolgozaton belül.

2. IDŐSOROK JELLEMZŐINEK ÁTTEKINTÉSE

A véletlen szerepét is figyelembe vevő idősorokat, a sztochasztikus idősorokat, az 1970-es években kezdték intenzívebben vizsgálni. Box és Jenkins ebben az időszakban megjelent könyvei [14], [62] áttekintették az idősor analízis valamennyi területét, és ezzel megalapozták a tudományterület további fejlődését. A tudományterület jelentőségét jelzik a megjelent külföldi és hazai szerzők kézikönyvei is [4], [5], [6], [19], [48], [80], [90]. A kézikönyvek mellett nagyszámú közlemény is megjelent az autoregresszív integrált mozgóátlag (ARIMA) modellek alkalmazásairól [7], [8], [20], [26], [27], [39], [42], [58], [59], [67], [68], [70], [71], [72], [73], [92], [96], [97], [98], [104], [112]. E módszereket alkalmazó számítógépes feldolgozások számának gyarapodását nagyban elősegítette a számítógépek teljesítményének növekedése és a megfelelő programcsomagok kifejlesztése.

Az alábbiakban az idősorok olyan jellemzői kerülnek ismertetésre, melyek a későbbiekben bemutatott számítógépes alkalmazásokhoz kapcsolódnak.

2.1. Sztochasztikus folyamatok jellemzői

A $\xi(t)$ sztochasztikus (véletlen) folyamat a valószínűségi változóknak egy t paramétertől függő sokaságát jelenti. A paraméter véges vagy végtelen halmaz értékeit veheti fel és gyakran az időt jelenti, jelöljük T -vel a $t \in T$

paraméterhalmazt. Matematikai értelemben akkor beszélünk sztochasztikus folyamatról, ha adott az $\{\Omega, A, P\}$ valószínűségi téren (mezőn) értelmezett a t paramétertől függő $\{\xi_t(\omega), t \in T, \omega \in \Omega\}$ valószínűségi változók összessége. Ω az eseményteret, P a valószínűségi mértéket, A az eseménytér részhalmazaiából alkotott σ -algebrát jelöli. A sztochasztikus folyamatot többféle módon szokás jelölni: $\xi_t(\omega)$ vagy ξ_t illetve $\xi(t, \omega)$ és $\xi(t)$. A jelöléseket ekvivalensnek tekintjük a következőkben. Ha a $\xi_t(\omega)$ függvény ω változóját rögzítjük, és t befutja a T paraméterhalmazt, akkor egy valós függvényt kapunk, amelyet a sztochasztikus folyamat realizációjának nevezünk. Egy realizáció tehát a folyamat egy konkrét kimenetelét jellemzi. Ha t értékét rögzítjük, akkor bármely rögzített $t_0 \in T$ paraméterhez egy $\xi_{t_0}(\omega)$ valószínűségi változó tartozik. Azt a halmazt, amelyből ξ_t értékeit felveszi, állapottérnek nevezzük. Ha az állapottér megszámlálható elemet tartalmaz, diszkrét állapotterű sztochasztikus folyamatról, a nem megszámlálható esetben folytonos állapotterű sztochasztikus folyamatról beszélünk. Ha a t paraméter folytonosan változik, akkor időben folytonos paraméterű, ha t diszkrét értékeket vesz fel, akkor időben diszkrét paraméterű sztochasztikus folyamatról beszélünk [5].

A ξ_t sztochasztikus folyamatot gyengébb értelemben vett stacionáriusnak, röviden stacionáriusnak nevezzük, ha $E\xi_t = \mu$; $Cov(\xi_t, \xi_{t+h}) = c(h)$; $\sigma_t^2 = \sigma^2$. A kovariancia függvény (ezután jelöljük C -vel) csak h -tól függ, amelyet késleltetésnek szokás nevezni. Stacionárius folyamatra az autokorrelációs függvény [34]:

$$(2.1) \quad \rho(h) = \frac{C(h)}{C(0)}.$$

A fenti kifejezésből látható, hogy $\rho(0)=1$ és $\rho(h)=\rho(-h)$.

A valószínűségi vektorváltozó meghatározása [5]. Legyenek $\xi_1, \xi_2, \dots, \xi_k$ valószínűségi változók. A $\underline{\xi}=(\xi_1, \dots, \xi_k)^T$ vektort valószínűségi vektorváltozónak nevezzük, ahol $(\xi_1, \dots, \xi_k)^T$ a $\xi_1, \dots, \xi_k$ valószínűségi változókból képzett vektor transzponáltját jelöli. A valószínűségi vektorváltozó koordinátáit oszlopvektorba írjuk fel.

Jelölje $\underline{\xi}(t)$ a k -dimenziós valószínűségi vektorváltozót [5]:

$$(2.2) \quad \underline{\xi}(t) = \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \\ \dots \\ \xi_k(t) \end{bmatrix}.$$

A ξ_i (skaláris) valószínűségi változót a $\underline{\xi}(t)$ valószínűségi vektorváltozó i -edik koordinátájának nevezzük. A k -dimenziós folyamat várható érték vektorfüggvényének meghatározása [5]:

$$(2.3) \quad E\underline{\xi}(t) = \begin{bmatrix} E\xi_1(t) \\ E\xi_2(t) \\ \dots \\ E\xi_k(t) \end{bmatrix}.$$

Az autokorrelációs függvény (autokorrelációs mátrix) a következő alakú [5]:

$$(2.4) \quad D(s,t)=E(\underline{\xi}(s)\underline{\xi}^T(t)) = \begin{bmatrix} E\xi_1(s)\xi_1(t) & E\xi_1(s)\xi_2(t) & \dots & E\xi_1(s)\xi_k(t) \\ E\xi_2(s)\xi_1(t) & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ E\xi_k(s)\xi_1(t) & \dots & \dots & E\xi_k(s)\xi_k(t) \end{bmatrix}.$$

A kovariancia függvény (kovariancia mátrix) a $(\underline{\xi}(t)-E\underline{\xi}(t))$ vektorfolyamat várható értékéből határozható meg [5].

A parciális korrelációs együttható meghatározásához a feltételes várható érték fogalmát használjuk fel [94], [125]. Ha ξ és η diszkrét valószínűségi változók, akkor a ξ valószínűségi változó $\eta=y$ feltétel melletti várható értékén az alábbi összeget értjük [125]:

$$(2.5) \quad E(\xi|\eta=y) = E(\xi|\eta) = \sum_i x_i P(\xi = x_i | \eta = y).$$

Ha a ξ és az η folytonos valószínűségi változók együttes sűrűségfüggvénye $f(x,y)$ és $f_2(y)$ az η valószínűségi változó peremsűrűségfüggvénye, akkor a ξ folytonos valószínűségi változó $\eta=y$ feltételre vonatkozó feltételes sűrűségfüggvényén az alábbi függvényt értjük [5]:

$$(2.6) \quad f(x|\eta=y) = \frac{f(x,y)}{f_2(y)}, \quad f_2(y) \neq 0, \quad x \in R.$$

A ξ folytonos valószínűségi változónak az η folytonos valószínűségi változóra az $\eta=y$ feltétel melletti feltételes várható értéke a következő

kifejezés [125], feltéve $\int_{-\infty}^{+\infty} |x| f(x|y) dx < \infty$:

$$(2.7) \quad E(\xi|\eta=y) = \int_{-\infty}^{\infty} x f(x|y) dx.$$

Jelölje $\xi_1, \xi_2, \dots, \xi_k$ a valószínűségi változókat, és tegyük fel, hogy ξ_1 -et becsüljük a $\xi_2, \dots, \xi_k$ -val. A becsléshez alkalmazzuk az $E(\xi_1|\xi_2=x_2, \dots, \xi_k=x_k) = h(x_2, \dots, x_k)$ $(k-1)$ változós függvényt. A $\xi_2, \dots, \xi_k$ konkrét értéke esetén ξ_1 becslését kapjuk.

Tekintsük a $\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_j, \dots, \xi_k$ valószínűségi változókat. Az i -k és a j -k közötti közvetlen hatás mérésére alkalmazható $\rho_{ij} = \rho_{ij.12,\dots,i-1,i+1,\dots,j-1,j+1,\dots,k}$ a parciális korrelációs együttható, miközben a többi változó hatását kizárjuk [94], [120]. A ξ_i -t becsüljük a többi

változóra vonatkozó feltételes várható értékkel. A maradékot jelölje $\eta_{i,1,2, \dots, j-1,j+1, \dots, k}$. A ξ_j változót is hasonlóan becsüljük, a maradékot jelöljük $\eta_{j,1,2, \dots, i-1,i+1, \dots, k}$ -val. Az így kapott maradékok közönséges értelemben vett korrelációs együtthatóját képezzük, ezt nevezzük a ξ_i és ξ_j parciális korrelációs együtthatójának. A ξ_i és ξ_j kapcsolatából a többi változó hatását kiszűrtük.

$$(2.8) \quad \rho_{ij,1,2, \dots, i-1,i+1, \dots, j-1,j+1, \dots, k} = \frac{E(\eta_{i,1,2, \dots, j-1,j+1, \dots, k} \eta_{j,1,2, \dots, i-1,i+1, \dots, k})}{D(\eta_{i,1,2, \dots, j-1,j+1, \dots, k}) D(\eta_{j,1,2, \dots, i-1,i+1, \dots, k})}.$$

Az indexben a pont előtt azon változók indexei szerepelnek, amelyek kapcsolatát vizsgáljuk, a pont után, pedig olyan változók indexei, amelyek hatásától eltekintünk.

Ha C_k -val jelöljük a $\xi_1, \xi_2, \dots, \xi_k$ valószínűségi változók kovariancia mátrixát, akkor a (2.8) kifejezés felírható a kovariancia mátrixból képzett megfelelő aldeterminánsokkal, ahol C_{ij} a mátrix (i,j) indexű tagjához tartozó aldetermináns. Hasonlóan értelmezzük C_{ii} -t és C_{jj} -t is [94], [120].

$$(2.9) \quad \rho_{ij} = -\frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}.$$

Speciálisan három változó esetében a ξ_1 és a ξ_2 kapcsolatának vizsgálata során egyetlen ξ_3 változó hatását küszöböljük ki. Ekkor a ξ_1 és a ξ_2 változók ξ_3 -ra vonatkoztatott parciális korrelációs együtthatója [120]:

$$(2.10) \quad \rho_{12,3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1 - \rho_{13}^2)(1 - \rho_{23}^2)}}.$$

A fenti képlet szerint a $\rho_{12,3}$ parciális korrelációs együttható kifejezhető a három változó páronkénti totális korrelációs együtthatóival.

Ha a ξ_1 és a ξ_2 kapcsolatát a ξ_3 és a ξ_4 hatásának kiküszöbölésével vizsgáljuk, akkor a parciális korrelációs együttható [120]:

$$(2.11) \quad \rho_{12.34} = \frac{\rho_{12.4} - \rho_{13.4}\rho_{23.4}}{\sqrt{(1 - \rho_{13.4}^2)(1 - \rho_{23.4}^2)}}.$$

2.2. Empirikus kifejezések

Matematikai statisztikai szempontból egy megfigyelendő mennyiség egy valószínűségi változó, jelöljük ezt X -lel. Megfigyeljük a jelenséget N -szer, egymástól függetlenül. A megfigyelési eredményeket mintának nevezzük, jelöljük $X_1, X_2, X_3, \dots, X_N$ -lel, ezek a mennyiségek is valószínűségi változók. Rögzített $\omega \in \Omega$ esetén az $x_1 = X_1(\omega), x_2 = X_2(\omega), \dots, x_N = X_N(\omega)$ szám N -est a minta realizációjának nevezzük, Ω az eseményteret jelöli.

Legyen $X_1, X_2, \dots, X_N$ N elemű minta X -re. Empirikus középnek vagy minta átlagnak nevezzük az alábbi kifejezést [35]:

$$(2.12) \quad \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i.$$

Az empirikus szórásnégyzetet a következő kifejezés határozza meg [35]:

$$(2.13) \quad s_N^2 = \frac{1}{N} \sum_{i=1}^N \left(X_i - \bar{X} \right)^2.$$

Ha az X és az Y valószínűségi változókat egyszerre megfigyeljük N -szer, akkor az $(X, Y)^T$ kétdimenziós vektorváltozóra kapunk értékeket. A valószínűségi vektorváltozó $1/N$ valószínűséggel vesz fel minden $(x_i, y_i)^T$ értéket, mely a minta realizáció i -edik eleme. A két koordináta közötti kovariancia [35]:

$$(2.14) \quad c = \frac{1}{N} \sum_{i=1}^N \left(x_i - \bar{x} \right) \left(y_i - \bar{y} \right) .$$

Empirikus kovarianciának nevezzük az alábbi mennyiséget [35]:

$$(2.15) \quad C = \frac{1}{N} \sum_{i=1}^N \left(X_i - \bar{X} \right) \left(Y_i - \bar{Y} \right) .$$

Az empirikus korrelációs együtthatót az alábbi kifejezés adja meg [35]:

$$(2.16) \quad R = \frac{\sum_{i=1}^N \left(X_i - \bar{X} \right) \left(Y_i - \bar{Y} \right)}{\sqrt{\sum_{i=1}^N \left(X_i - \bar{X} \right)^2} \sqrt{\sum_{i=1}^N \left(Y_i - \bar{Y} \right)^2}} .$$

Legyenek $z_{t_1}, z_{t_2}, \dots, z_{t_N}$ valamely idősor megfigyelt értékei a $t_1, t_2, \dots, t_N$ időpontokban. Ezután egyszerűbb jelöléssel a következőképpen hivatkozunk rá: $z_1, z_2, \dots, z_N$ idősor a $t=1, 2, \dots, N$ időpontokban. Ha z_t stacionárius, és teljesül, hogy $Ez_t=0$, akkor

$$(2.17) \quad c_k = Ez_t z_{t+k}, \quad t, k=0; \pm 1; \pm 2; \dots .$$

c_k jelenti a k -adik autokovarianciát, azaz az idősorban az egymástól k távolságra lévő tagok közötti kovarianciát [25].

A c_k autokovariancia becslése a következő, ahol $k=1, 2, \dots, L$ és $L \ll N$ [34]:

$$(2.18) \quad \hat{c}_k = \frac{1}{N-k} \sum_{t=1}^{N-k} z_t z_{t+k} .$$

Speciálisan $c_0 = \sigma^2 = D^2(z_t) = E(z_t^2)$ becslése [34]:

$$(2.19) \quad c_0 = \sigma^2 = \frac{1}{N} \sum_{t=1}^N z_t^2 .$$

A c_k autokovarianciák becsléseit empirikus autokovarianciáknak nevezzük. Az autokorrelációs függvény, $k=0; \pm 1; \pm 2; \dots$ esetén [25]:

$$(2.20) \quad r_k = \frac{c_k}{c_0} \quad .$$

Az r_k jelenti az idősorban az egymástól k távolságra lévő tagok közötti korrelációt. Az autokorrelációs együtthatók becslése [34]:

$$(2.21) \quad \hat{r}_k = \frac{\sum_{t=1}^{N-k} z_t z_{t+k}}{\left(\sum_{t=1}^{N-k} z_t^2 \right)^{1/2} \left(\sum_{t=1}^{N-k} z_{t+k}^2 \right)^{1/2}}, \quad k = 1, 2, \dots, L.$$

Ha a z_t stacionárius idősorban $Ez_t = 0$, akkor r_k becslése [34]:

$$(2.22) \quad \hat{r}_k = \frac{\frac{1}{N-k} \sum_{t=1}^{N-k} z_t z_{t+k} - \frac{1}{(N-k)^2} \left(\sum_{t=1}^{N-k} z_t \right) \left(\sum_{t=1}^{N-k} z_{t+k} \right)}{\left(\frac{1}{N-k} \sum_{t=1}^{N-k} z_t^2 - \frac{1}{(N-k)^2} \left(\sum_{t=1}^{N-k} z_t \right)^2 \right)^{1/2} \left(\frac{1}{N-k} \sum_{t=1}^{N-k} z_{t+k}^2 - \frac{1}{(N-k)^2} \left(\sum_{t=1}^{N-k} z_{t+k} \right)^2 \right)^{1/2}}.$$

Az r_k becsléseit empirikus autokorrelációs együtthatóknak illetve szeriális korrelációs együtthatóknak nevezzük. Az r_k becsléseit k függvényében ábrázolva az idősor korrelogramját kapjuk. Korrelálatlan változókból álló stacionárius idősor esetén $r_0 = 1$, és $r_k = 0$, ha $k \neq 0$. Ha egy idősor teljesen véletlen, akkor N nagy értékeire valamennyi empirikus autokorreláció nullához tart. Stacionárius idősorokra jellemző, hogy gyakran tartalmaznak rövid távú autokorrelációt, pl: r_1 viszonylag magas, r_2 és r_3 is szignifikánsan különbözhet nullától, de k növelésével r_k zérushoz tart [34].

Autokorrelációs mátrix [25]:

$$(2.23) \quad R_k = \begin{pmatrix} 1 & r_1 & \dots & r_j & \dots & r_{k-1} \\ r_1 & 1 & \dots & r_{j-1} & \dots & r_{k-2} \\ \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ r_j & r_{j-1} & \dots & 1 & \dots & r_{k-j+1} \\ \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ r_{k-1} & r_{k-2} & \dots & r_{k-j+1} & \dots & 1 \end{pmatrix}.$$

A parciális autokorrelációs függvény az R_k mátrix determinánsából áll elő [25]:

$$(2.24) \quad \rho_k = \frac{\det R_k^*}{\det R_k}, \quad k = 1, 2, \dots$$

Az R_k^* mátrixot úgy kapjuk meg az R_k mátrixból, hogy az R_k utolsó sorát az $(r_1, r_2, \dots, r_k)$ sorral helyettesítjük. Speciálisan:

$$(2.25) \quad \rho_1 = r_1 \quad \text{és}$$

$$(2.26) \quad \rho_2 = \frac{r_2 - r_1^2}{1 - r_1^2}.$$

A ρ_k a z_t és z_{t+k} valószínűségi változók parciális korrelációs együtthatója a $z_{t+1}, z_{t+2}, \dots, z_{t+k-1}$ változókra, mint feltételekre vonatkozóan.

2.3. Kétdimenziós stacionárius folyamatok

Megfigyelhetünk olyan sztochasztikus jelenségeket is, melyeknek két jellemző tulajdonságuk van. Ekkor a kétdimenziós sztochasztikus folyamat realizációját egy idősort párral adhatjuk meg. Az egyik idősort jelöljük: $\dots, Z(t_{-1}), Z(t_0), Z(t_1), Z(t_2), \dots$ a másikat: $\dots, Y(t_{-1}), Y(t_0), Y(t_1), Y(t_2), \dots$ sorozattal. Feltételezve, hogy egyenlő időközönként keletkeztek a megfigyelési adataink, akkor az egyik idősor jelölése legyen: $\dots, Z(-1), Z(0), Z(1), Z(2), \dots$ a másiké $\dots, Y(-1), Y(0), Y(1), Y(2), \dots$. Egyszerűbben jelölve:

... , Z_{-1} , Z_0 , Z_1 , Z_2 , ... és ... , Y_{-1} , Y_0 , Y_1 , Y_2 , A kétdimenziós folyamat (Z_t, Y_t) nem szükségképpen stacionárius, de stacionáriussá tehető, megfelelő számú differenciaképzéssel [14]. Jelöljük z_t -vel és y_t -vel a differenciaképzés után kapott stacionárius idősorokat, azaz $z_t = \nabla^d Z_t$; $y_t = \nabla^d Y_t$, rendszerint $d=0$; 1 ; 2 . Ha $d=1$, akkor az elsőrendű differenciák képzése: $\nabla Z_t = Z_t - Z_{t-1}$; ha $d=2$, akkor a másodrendű differenciák képzése:

$$\nabla^2 Z_t = \nabla(Z_t - Z_{t-1}) = \nabla Z_t - \nabla Z_{t-1} = Z_t - Z_{t-1} - (Z_{t-1} - Z_{t-2}) = Z_t - 2Z_{t-1} + Z_{t-2}.$$

A kétdimenziós folyamatot alkotó stacionárius idősorok várható értékei (jelöljük: μ_z -vel és μ_y -nal) és varianciáik (jelöljük σ_z^2 -tel és σ_y^2 -tel) állandóak. A folyamat mindkét idősorára az autokovariancia együtthatókat a szokásos módon értelmezzük. A $C_{zz}(k)$ és a $C_{yy}(k)$ jelöli a z_t és az y_t idősorok autokovarianciáit [14].

$$(2.27) \quad C_{zz}(k) = E[(z_t - \mu_z)(z_{t+k} - \mu_z)] = E[(z_t - \mu_z)(z_{t-k} - \mu_z)],$$

$$C_{yy}(k) = E[(y_t - \mu_y)(y_{t+k} - \mu_y)] = E[(y_t - \mu_y)(y_{t-k} - \mu_y)].$$

Speciálisan vizsgálhatjuk a kétdimenziós folyamatot alkotó két idősor kapcsolatát. A két idősor keresztkovariancia függvénye a $k=0$; 1 ; 2 ; ... időkésléptetéseknél [14]:

$$(2.28) \quad C_{zy}(k) = E[(z_t - \mu_z)(y_{t+k} - \mu_y)], \quad k = 0, 1, 2, \dots,$$

$$C_{yz}(k) = E[(y_t - \mu_y)(z_{t+k} - \mu_z)], \quad k = 0, 1, 2, \dots.$$

Az idősorok közötti keresztkorrelációs függvény a következő [14]:

$$(2.29) \quad \gamma_{zy}(k) = \frac{C_{zy}(k)}{\sigma_z \sigma_y} \quad k = 0; \pm 1; \pm 2; \pm \dots.$$

A keresztkorrelációs függvény általában nem szimmetrikus a $k=0$ értékre. A keresztkovariancia és a keresztkorrelációs függvények

becsléséhez az empirikus idősorok átlagait ($\bar{z}$ és $\bar{y}$) használjuk fel. A keresztkovariancia függvény becslt értékeit $c_{zy}(k)$ -l, a keresztkorreláció becsléseit $r_{zy}(k)$ -l jelöljük, az s_z és az s_y pedig a szórásnégyzetek (σ_x és a σ_y) becslései [14]:

$$(2.30) \quad c_{zy}(k) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k} \left(z_t - \bar{z} \right) \left(y_{t+k} - \bar{y} \right) & \text{ha } k \geq 0 \\ \frac{1}{n} \sum_{t=1}^{n-k} \left(y_t - \bar{y} \right) \left(z_{t-k} - \bar{z} \right) & \text{ha } k \leq 0 \end{cases},$$

$$(2.31) \quad r_{zy}(k) = \frac{c_{zy}(k)}{s_z s_y},$$

$$(2.32) \quad s_z = \sqrt{c_{zz}(0)},$$

$$(2.33) \quad s_y = \sqrt{c_{yy}(0)}, \quad k = 0, \pm 1, \pm 2, \dots$$

A keresztkorrelációk standard hibáinak becslésére alkalmazható az alábbi kifejezés [14]:

$$(2.34) \quad s_{zy} \cong \frac{1}{\sqrt{n-k}}.$$

2.4. Speciális idősorok

Matematikai statisztikából ismert, hogy a legegyszerűbb idősor az u.n. fehérzaj sorozat, amikor az $\varepsilon_t, \varepsilon_{t+1}, \varepsilon_{t+2}, \dots$ független, azonos eloszlású véletlen változók sorozata $N(0, \sigma_\varepsilon^2)$ (normális) eloszlással. Egy véletlen folyamatot megfigyelünk a $\dots, t_{-2}, t_{-1}, t_0, t_1, t_2, \dots$ időpontokban, a megfigyelt értékek jelölésére használjuk a $\dots, z_{-2}, z_{-1}, z_0, z_1, z_2, \dots$. Ha a z_t megfigyelési érték lineárisan és sztochasztikusan az előző $z_{t-1}, \dots, z_{t-p}$, és $\varepsilon_t, \dots, \varepsilon_{t-q}$ értékektől függ, akkor p -edrendű autoregresszív és q -adrendű

mozgóátlag modellnek, un. ARMA(p, q) modellnek nevezzük a következő kifejezést, ahol $\tilde{z}_t = z_t - \mu$; [14]:

$$(2.35) \quad \tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \Theta_2 \varepsilon_{t-2} - \dots - \Theta_q \varepsilon_{t-q}.$$

Ha $p=0$ és $q=0$, akkor a fenti kifejezés a fehérzaj folyamatot jellemezi.

Bevezetve a visszaléptetés operátort, melyet jelöljünk B -vel, és az értelmezése: $B \tilde{z}_t = \tilde{z}_{t-1}$; és $B^m \tilde{z}_t = \tilde{z}_{t-m}$, akkor a (2.35) a következő egyszerű alakban felírható [14]:

$$(2.36) \quad \phi(B) \tilde{z}_t = \theta(B) \varepsilon_t, \quad \text{ahol}$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad \text{és} \quad \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q.$$

A (2.36) kifejezés baloldala p -edrendű autoregresszív- (AR(p)), a jobb oldala q -adrendű mozgóátlag (MA(q)) folyamatot jelent. Az AR, MA, és az ARMA modellek az un. stacionárius idősorok elemzésére alkalmazhatók [14], [25], [60]. Vannak olyan idősorok, melyek stacionárius idősorrá transzformálhatóak d -szeres differenciaképzéssel, ahol rendszerint $d=0$; 1; 2. Vagyis $w_t = \nabla^d z_t$, ahol w_t már stacionárius idősor, ∇ jelöli a differenciák képzését. A differenciák képzésével stacionáriussá transzformált idősorok esetében integrált autoregresszív mozgóátlag (ARIMA) modellekről beszélünk. Az első- és a másodrendű differenciák képzését az előzőekben megadtam.

2.5. Idősorok komponensei

Valamely időben változó jelenség egymást követő egyenlő időpontokban megfigyelt értékei a tapasztalati (empirikus) idősort alkotják.

Feltételezzük, hogy az empirikus idősor mindazokat az információkat tartalmazza, mint a sztochasztikus folyamat, amiből a megfigyelési értékeink származnak. Így az empirikus idősor valamely diszkrét vagy folytonos paraméterű sztochasztikus folyamat realizációja.

Időbeli alakulásuk sajátosságai alapján az idősorokat három összetevő vagyis három komponens eredőjeként tekintjük: trend vagy alapirányzat, periodikus ingadozás, véletlen ingadozás [34].

A trend a vizsgált jelenségben a tartósan érvényesülő tendenciát, az idősor alakulásának fő irányát jelenti.

A periodikus ingadozás alatt a szezonális ingadozásokat és a ciklikus változásokat értjük. A szezonális vagy idényszerű ingadozás (röviden szezonális) állandó periódushosszúságú hullámmozgás, ritmikus ingadozás, amely szabályosan ismétlődő időközönként mindig azonos irányba téríti el az idősor értékeit az alapirányzattól. A szezonális az évszakok változásaival van összefüggésben. Megfigyelhetünk olyan periodikus ingadozásokat is, melyek nem az évszakok változásaihoz kapcsolódnak, ezeket ciklikus ingadozásoknak nevezzük.

A véletlen ingadozás az idősorban található szabálytalan mozgás. Ez a komponens igen sok, az idősor alakulása szempontjából nem jelentős tényező együttes hatását képviseli. Általában minél rövidebb időszakra vonatkozik az idősor, annál jelentősebb a véletlen ingadozás hatása.

Az idősor elemzés hagyományos módszere az egyes komponensek elkülönítése. Az idősor összetevőinek kapcsolódása szerint beszélhetünk additív és multiplikatív kapcsolatokról. Jelölje z_{ij} az idősor elemeit, $i=1, 2, \dots, n$ a periódusok számát (pl: évek); $j=1, 2, \dots, m$ a perióduson belüli időszakok számát (pl: hónapok), s szezonális tényezőt, v véletlen ingadozást, $\hat{z}$ a trendet. Ekkor az additív modell megadása [34]:

$$(2.37) \quad \hat{z}_{i,j} = z_{i,j} + s_j + v_{i,j}.$$

Multiplikatív modell [34]:

$$(2.38) \quad \hat{z}_{i,j} = z_{i,j} * s_j * v_{i,j}.$$

A trend megadása történhet analitikus trendszámítással vagy mozgóátlagolású trendszámítással. Az analitikus trendszámítás az idősorban lévő tartós tendenciát megfelelően választott függvénnyel (egyenes, parabola, exponenciális görbe) adja meg és ennek paramétereit becsüli.

A mozgóátlagolás a trendet az idősor dinamikus átlagaként állítja elő, a véletlen hatását tompítja. A mozgóátlagolás tagszámaként a periódus részidőszakainak számát vagy annak egész számú többszörösét vesszük, így biztosítjuk, hogy valamennyi trendadat (mozgóátlag) előállításában minden egyes szezon megjelenjen. A mozgóátlagok képzése: meghatározzuk az átlagolandó értékek számát (k), és vesszük az első k adat átlagát. Ez az érték az első trendadat, melyet az időszak közepéhez írunk. Folytatjuk az eljárást, elhagyjuk az első adatot, helyette vesszük az előbbi adatokat követő értéket. Ezek átlaga képezi a második trendadatot. Az idősor utolsó k adatának felhasználásáig ismételjük az eljárást. Jelölje $z_1, z_2, \dots, z_n$ a megfigyelt idősor elemeit. Ekkor a k tagú mozgóátlagok [34]:

$$(2.39) \quad \begin{aligned} \hat{z}_1 &= \frac{z_1 + z_2 + \dots + z_k}{k}, \\ \hat{z}_2 &= \frac{z_2 + z_3 + \dots + z_{k+1}}{k}, \\ &\dots \\ \hat{z}_{n-k+1} &= \frac{z_{n-k+1} + z_{n-k+2} + \dots + z_n}{k}. \end{aligned}$$

A trendadatok száma kevesebb lesz, mint az idősor elemeinek száma, az idősor elejéhez és végéhez nem képződnek átlagok. Páratlan k tagszám esetén $(k-1)$ taggal, páros tagszám esetén k taggal rövidül a sor. Ha k szám páros, akkor a mozgó átlagolással kapott értékek az eredeti idősor két időszaka közé kerülnek. Ilyenkor a mozgóátlagolással kapott értékeket még középre kell igazítani (un. centírozás), ami a szomszédos mozgó átlagok további, páronkénti átlagolásával történik.

Hosszabb idősoroknál vagy nagyobb véletlen ingadozást tartalmazó idősoroknál célszerű nagyobb tagszámot venni. A sor elejéről és végéről „elvesző” tagok nem okoznak problémát, a véletlen hatását pedig jobban ki lehet szűrni.

A felírt kifejezések a súlyozatlan mozgóátlagok képzésére vonatkoznak. A súlyozott mozgóátlagok bemutatásával most nem foglalkozom.

Ha az idősorban a ciklus hossza elég nagy, akkor a mozgóátlagolás jól követi azt, és a mozgóátlag nem a trendet, hanem valójában a trend és a ciklus összegét adja meg.

Az idősorok második fő összetevője a periodikus ingadozás. A periodikus ingadozások többfélék lehetnek, most részletesebben csak a szezonálitásra térek ki. A szezonálitás állandó periódus hosszúságú ingadozás. A periódus hossza egy év, vagy ennél rövidebb. Szezonálitás vizsgálatakor azt jellemezzük, hogy a szezonális hatás a periódus egyes szakaszaiban milyen mértékben, vagy arányban téríti el az idősor értékeit a trendtől, vagyis az alapirányzattól. Beszélhetünk állandó- és változó szezonálitásról. Állandó szezonálitás esetén a szezonális összetevő viselkedése lehet:

- additív: ekkor a szezonális hatás abszolút nagysága, a hullámváz amplitúdója állandó, nem függ az idősor értékétől,
- multiplikatív: a szezonális hatás relatív nagysága állandó.

Mindkét esetben az egyes szezonokra (a perióduson belüli szakaszokra) valamennyi periódusra állandó értéket kapunk. Változó szezonálitásnál a szezonális kilengések nem állandóak, periódusonként változik az egyes szezonokhoz tartozó ingadozás.

Additív modellben a szezonális hatást *szezonális eltéréssel* jellemezzük. A szezonális hatás kimutatásakor az idősből kiszűrjük a trend és a véletlen tényező hatását (a szezonálitáson kívüli további ciklikus hatásoktól eltekintünk). A trend kiszűrése:

$$(2.40) \quad \hat{z}_{i,j} - z_{i,j} = s_j + v_{i,j} = z_{i,j}^*.$$

A $z_{i,j}^*$ a trendhatástól mentes idősort jelöli, melyben a szezonhatás és a véletlen szerepel. A véletlen hatás gyengítésére, kiszűrésére a $z_{i,j}^*$ idősor elemeit a periódusok i száma szerint átlagoljuk:

$$(2.41) \quad \hat{s}_j = \frac{\sum_{i=1}^n z_{i,j}^*}{n}.$$

Ahhoz, hogy a szezonális eltérések összege, illetve átlaga nulla legyen, korrekció szükséges. A korrigált szezonális eltérések, s_j^* , meghatározása:

$$(2.42) \quad s_j^* = \hat{s}_j - \frac{\sum_{j=1}^m \hat{s}_j}{m}.$$

A szezonális eltérés azt mutatja meg, hogy a szezonális hatás miatt az adott időszakban (pl: az első negyedévben) mennyivel tér el az idősor adata az alapirányzatnak megfelelő értéktől.

Multiplikatív modellben a szezonális hatást *szezonindexekkel* jellemezzük. A trendhatástól mentes idősor:

$$(2.43) \quad \frac{z_{i,j}}{\hat{z}_{i,j}} = s_j * v_{i,j} = z_{i,j}^*.$$

A véletlen hatás gyengítésére, illetve kiszűrésére szintén átlagolást alkalmazunk:

$$(2.44) \quad \hat{s}_j = \frac{\sum_{i=1}^n z_{i,j}^*}{n}.$$

A szezonindexek korrekciójával érhetjük el, hogy az $\hat{s}_j$ értékek átlaga 1 legyen. A korrigált szezonindexek meghatározása:

$$(2.45) \quad s_j^* = \frac{\hat{s}_j}{\frac{\sum_{j=1}^m \hat{s}_j}{m}}.$$

A szezonindex azt fejezi ki, hogy a szezonhatás miatt az adott szezonban relatíve mennyivel (hány százalékkal) nagyobb vagy kisebb az idősor adata, mint a megfelelő trendadat.

A trend és a szezonális tényező kiszűrése után a véletlen komponens maradéktagként, reziduumként, kapjuk. A reziduumok számítása additív modellben:

$$(2.46) \quad v_{i,j} = z_{i,j} - \hat{z}_{i,j} - \hat{s}_j.$$

A reziduumok multiplikatív modellben:

$$(2.47) \quad v_{i,j} = \frac{z_{i,j}}{\hat{z}_{i,j} * \hat{s}_j}.$$

Ha az idősor komponenseit helyesen határozzuk meg, a reziduumok véletlenszerűen viselkednek.

Az idősorok statisztikai analízisének legfontosabb vizsgálati szempontjai:

- a determinisztikus (nem véletlentől függő) komponensek függvényalakjának meghatározása,
- az idősorok sztochasztikus modelljeinek megalkotása, azaz az idősorok véletlen komponenseinek modellezése.

Továbbiakban csak a véletlen komponenst tartalmazó idősorok azonosításával, modellezésével foglalkozom.

2.6. Sztochasztikus modellek megadása

2.6.1. Sztochasztikus modellek azonosítása

A sztochasztikus folyamat modellezésének első szakaszát a modell azonosításának vagy identifikációjának nevezzük. Először megállapítjuk, hogy a vizsgált idősor stacionárius-e, illetve ha nem, akkor megpróbáljuk azt alkalmas transzformációval stacionáriussá tenni. Előzetes számítások alapján kezdőértékeket is adunk a modell p , d , q paramétereinek [14].

A modellezés kiindulását jelentő tapasztalati idősor az elméleti idősort alkotó valószínűségi változók realizációja. A tapasztalati idősor akkor csak tekinthető olyan mintának, melyből az elméleti idősort jelentő sztochasztikus folyamat jellemzői becsülhetők, ha az elméleti idősor jellemzői (várható érték, szórásnégyzet, autokorrelációs együtthatói) időben állandóak, azaz függetlenek a t változótól. A stacionárius idősorok nem tartalmaznak trendhatást, az idősor értékei egy állandó konstans szint körül ingadoznak állandó szórással. A stacionárius idősorban a belső összefüggéseket kifejező autokorrelációs együtthatók időben állandóak, nem függenek a t időtől, csak a változók egymás közötti távolságától. Az autokorrelációs függvényben az elsőrendű autokorrelációs együttható az

egymásután következő időszaki értékek, a másodrendű autokorrelációs együttható az egymástól két időegységnyi távolságra lévő időszaki értékek korrelációját mutatja. Az autokorreláció vizsgálakor az időszaki elemek távolságát késleltetésnek nevezzük. A stacionárius idősorok gyakran tartalmazhatnak rövid távú autokorrelációt, vagyis az elsőrendű autokorreláció viszonylag magas, a másodrendű- és a harmadrendű autokorrelációk is szignifikánsan különböznek nullától, de a késleltetés növekedésével az autokorreláció értékei nullához tartanak. A stacionaritás vizsgálatával eldöntjük, hogy az adott időszaki sorhoz illeszthető-e ARIMA modell, ha igen, milyen d dimenzióval. Továbbiakban a stacionárius vagy azzá transzformált adatokkal dolgozunk.

A modellazonosítás során megkeressük a stacionárius vagy azzá transzformált időszaki sorhoz milyen MA(q), AR(p), vagy ARMA(p, q) modelltípus(oka)t lehet illeszteni, amelynek jellemzőire leginkább hasonlítanak a tapasztalati időszaki sorból számított empirikus jellemzők. Ehhez a modelltípusok elméleti jellegzetességeinek ismerete szükséges. A modell azonosításához az autokorrelációs függvények és a parciális autokorrelációs függvények adják a legfontosabb információkat. A tapasztalati időszaki sorból becsült autokorrelációs együtthatók ábrája, a korrelogram, alapján lehet tanulmányozni az autokorrelációs együtthatók viselkedését, a késleltetés (k) függvényében. Becsüljük az autokorrelációs és parciális autokorrelációs együtthatókat. Ezek az elméleti autokorrelációs és parciális korrelációs függvények közelítései, így viselkedésükből az elméleti függvények tulajdonságaira lehet következtetni [14], [60]. A becsült autokorrelációs együtthatók konfidencia intervalluma alapján megállapíthatók a nullától szignifikánsan különböző értékek, ezek a konfidencia sávon kívül vannak.

Az $MA(1)$ folyamat autokorrelációs függvényére az a jellemző, hogy csak első autokorreláció értéke nem nulla, a parciális autokorrelációs függvénye közel exponenciálisan vagy csillapodó szinusz görbe szerint csökken.

A $MA(q)$ folyamatoknál az autokorrelációk válnak nullává q darab nem nulla érték után, míg a parciális autokorrelációk csökkennek exponenciálisan vagy csillapodó szinusz görbe szerint.

Az $AR(1)$ folyamat akkor stacionárius, ha a $z_t = \phi_1 z_{t-1} + \varepsilon_t$ egyenletben a paraméterre teljesül, hogy $|\phi_1| < 1$. Az autokorrelációs függvény közel exponenciálisan csökken, ha ϕ_1 pozitív érték, a függvény csillapodó szinusz görbe szerint csökken, ha ϕ_1 negatív érték. Az $AR(1)$ autokorrelációs együtthatói a ϕ_1 paraméterből származtathatók: $r_k = \phi_1^k$. A parciális autokorrelációs függvény első értéke nem nulla, a többi érték szignifikánsan nem különbözik nullától. Az $AR(1)$ folyamatban nincs rejtett periódus, az $AR(2)$ tartalmaz rejtett periódust.

A $AR(p)$ folyamat autokorreláció értékei exponenciálisan vagy exponenciálisan és csillapodó szinusz görbéhez hasonlóan csökkennek, a parciális autokorrelációi p darab nem nulla érték után nullává válnak.

Az $ARMA(p,q)$ vegyes folyamat autokorrelációs- és parciális autokorrelációs függvényei néhány jellemző érték után exponenciálisan vagy csillapodó szinusz görbéhez hasonlóan csökkennek [14], [60]. A gyakrabban előforduló modellekre az említett függvények jellemzőit az 1. táblázatban összefoglaltam.

2. IDŐSOROK JELLEMZŐINEK ÁTTEKINTÉSE

1. táblázat: Az autokorrelációs- és parciális autokorrelációs együtthatók értékei ARIMA modellekben.

Modelltípus ARIMA(p,d,q)	Autokorrelációs együtthatók (ρ_k)	Parciális autokorrelációs együtthatók (ϕ_{kk})
AR(1) (1,d,0)	Exponenciálisan csökkennek, ha $\rho_1 > 0$, csillapodó szinusz görbe szerint csökkennek, ha $\rho_1 < 0$	ϕ_{11} ha $k=1$ 0 ha $k>1$
AR(2) (2,d,0)	Exponenciálisan és/vagy csillapodó szinusz görbe szerint csökkennek	ϕ_{11} ha $k=1$ ϕ_{22} ha $k=2$ 0 ha $k>2$
MA(1), ha $d=0$ vagy IMA(d,1) (0,d,1)	ρ_1 ha $k=1$ 0 ha $k>1$	Exponenciálisan vagy csillapodó szinusz görbe szerint csökkennek
MA(2), ha $d=0$, vagy IMA(d,2) (0,d,2)	ρ_1 ha $k=1$ ρ_2 ha $k=2$ 0 ha $k>2$	Exponenciálisan és/vagy csillapodó szinusz görbe szerint csökkennek
ARMA(1,1) ha $d=0$, vagy ARIMA (1,d,1) (1,d,1)	Exponenciálisan vagy csillapodó szinusz görbe szerint csökkennek a második értéktől kezdődően	Exponenciálisan vagy csillapodó szinusz görbe szerint csökkennek a második értéktől kezdődően
ARMA(1,2) ha $d=0$ vagy ARIMA(1,d,2) (1,d,2)	$\rho_0=1$ és ρ_1 után exponenciálisan csökkennek	Exponenciálisan és/vagy csillapodó szinusz görbe szerint csökkennek
ARMA(2,1) ha $d=0$ vagy ARIMA(2,d,1) (2,d,1)	Exponenciálisan és/vagy csillapodó szinusz görbe szerint csökkennek	$\phi_{11}=\rho_1$ és ϕ_{22} után exponenciálisan csökkennek
ARMA(2,2) ha $d=0$ vagy ARIMA(2,d,2) (2,d,2)	$\rho_0=1$ után exponenciálisan csökkennek	$\phi_{11}=\rho_1$ után Exponenciálisan és/vagy csillapodó szinusz görbe szerint csökkennek

2. IDŐSOROK JELLEMZŐINEK ÁTTEKINTÉSE

A modelltípus azonosítása után a modell paramétereire az induló értékek meghatározásához szintén a becsült autokorrelációkat és parciális autokorrelációkat használjuk fel. Autoregresszív folyamatoknál a Yule-Walker egyenletekből származtatjuk a ϕ autoregresszív paraméterek induló értékeit. Mozgóátlagolású folyamatokban a q darab nem nulla értékű becsült autokorreláció felírásából keletkezett egyenletekből fejezzük ki a θ paraméterek induló értékét. Vegyes modellekben a paraméterértékek és az autokorrelációk közötti összefüggésekből számolunk induló értékeket. A 2. táblázat tartalmazza a leggyakrabban használt modellekben az induló értékek származtatását.

2. táblázat: A paraméterek induló értékeinek meghatározása ARIMA modellekben.

Modelltípus ARIMA(p,d,q)	A paraméterek induló értékének származtatása	A paraméterek lehetséges értékei
(1,d,0)	$\phi_1 = \rho_1$	$-1 < \phi_1 < 1$
(2,d,0)	$\phi_1 = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2}$ $\phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$	$-1 < \phi_2 < 1$ $\phi_2 + \phi_1 < 1$ $\phi_2 - \phi_1 < 1$
(0,d,1)	$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2}$	$-1 < \theta_1 < 1$
(0,d,2)	$\rho_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}$ $\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$	$-1 < \theta_2 < 1$ $\theta_2 + \theta_1 < 1$ $\theta_2 - \theta_1 < 1$
(1,d,1)	$\rho_1 = \frac{(1 - \theta_1\phi_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}$ $\rho_2 = \rho_1\phi_1$	$-1 < \phi_1 < 1$ $-1 < \theta_1 < 1$

A becsült autokorrelációkat és parciális autokorrelációkat ellenőrizni kell, mennyire szignifikánsak, milyenek a standard hibák. Ehhez a Bartlett formula alkalmazható. A becsült autokorrelációkat r_k -val jelöljük, ekkor az r_k autokorrelációk standard hibája [14]:

$$(2.48) \quad \hat{\sigma}(r_k) \cong \frac{1}{\sqrt{n}} \left\{ 1 + 2(r_1^2 + r_2^2 + \dots + r_q^2) \right\}^{1/2} \quad k > q.$$

A parciális autokorrelációk standard hibáira teljesül a következő [14]:

$$(2.49) \quad \hat{\sigma}(\hat{\phi}_{kk}) \cong \frac{1}{\sqrt{n}} \quad k > p.$$

A becsült autokorrelációs- és parciális autokorrelációs együtthatókról Anderson kimutatta, hogy közepes n érték esetén olyan együtthatókra, amelyek elméleti értéke nulla, a becsült együtthatók eloszlása közelítőleg normális eloszlású. Ez segít, annak eldöntésében, hogy az együtthatók lényegesen különböznek-e nullától. Ha az együtthatók abszolút értéke nagyobb, mint a kétszeres standard hiba, akkor az együtthatókat nullától lényegesen különbözőknek tekintjük [1], [14].

Az identifikáció jóságáról a reziduumok varianciájának előzetes becslése ad tájékoztatást. Az $AR(p)$ modellekben a reziduumok varianciáját az alábbi kifejezéssel becsülhetjük, ahol c_0 a γ_0 variancia becsült értéke [14]:

$$(2.50) \quad \hat{\sigma}_a^2 = c_0 \left(1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2 - \dots - \hat{\phi}_p r_p \right).$$

$MA(q)$ modellekben a reziduumok varianciájának becslése [14]:

$$(2.51) \quad \hat{\sigma}_a^2 = \frac{c_0}{1 + \theta_1 + \dots + \theta_q}.$$

$ARMA(1,1)$ modellekben a reziduumok varianciáinak becslése [14]:

$$(2.52) \quad \hat{\sigma}_a^2 = \frac{1 - \hat{\phi}_1^2}{1 + \hat{\theta}_1^2 - 2\hat{\phi}_1\hat{\theta}_1} c_0.$$

2.6.2. Sztochasztikus modell paramétereinek becslése

A modelltípus azonosítása és a paraméterek kezdőértékének megadása után a paraméterek végleges értékeit kell meghatározni. A paraméterek hatáson becslésének eléréséhez a rendelkezésre álló adatainkat hatékony módon kell felhasználni. Ehhez felhasználhatók a maximum likelihood és a legkisebb négyzetek elvén alapuló becslések.

Likelihood függvény

Ha az idősor N elemű, akkor az idősor elemeit tekinthetjük egy N dimenziós véletlen változó egy mintájának. A véletlen változó eloszlása az ismeretlen α paraméterektől függ, azaz $p(z/\alpha)$. Az α paraméter a p számú ϕ paraméter, a q számú θ paraméter és a σ paraméterek együttesét jelenti, tehát $p+q+1$ számú paramétert. A $p(z/\alpha)$ rögzített α paraméterek mellett a z különböző kimeneteleinek valószínűségét adja meg. Becslés esetén fordított a helyzet. Ismerjük z egyetlen kimenetelét, az N elemű z idősort, és arra vagyunk kíváncsiak, hogy az idősor milyen α paraméterekhez tartozhat. A likelihood függvény, $L(\alpha/z)$, rögzített z mellett vizsgálja az α paramétereket. Jelölje l a $L(\alpha/z)$ logaritmusát, azaz $l = \log L(\alpha/z)$. A L illetve a l függvény maximumához tartozó paraméter érték olyan, amely mellett a z megfigyelésünk a legesélyesebb, a z -ből következő legvalószínűbb α érték, az L maximumához tartozik [14], [60].

Gyakorlati szempontból fontos a reziduumok négyzetösszeg függvényének vizsgálata, jelöljük ezt $S(\alpha)$ -vel. Az idősor elemeit jelöljük $z_0, z_1, z_2, \dots, z_N$ -nel, és feltételezzük, hogy $N=n+d$ számú megfigyelésünk van az idősorból. Továbbá feltételezzük, hogy a z_t idősor a (p,d,q) paraméterű ARIMA modell generálja. A z_t értékeiből d -edfokú differenciaképzéssel a $w_1, w_2, \dots, w_n$ stacionárius idősorot kapjuk. Ehhez az ARMA(p,q) modellt illesztjük, és becsüljük az autoregresszív ϕ , valamint a mozgóátlag θ paramétereket. A modellből a reziduumokat kifejezve a következőt kapjuk [14], [60]:

$$(2.53) \quad a_t = \tilde{w}_t - \phi_1 \tilde{w}_{t-1} - \phi_2 \tilde{w}_{t-2} - \dots - \phi_p \tilde{w}_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}, \quad \text{ahol} \quad w_t = \nabla^d z_t; \quad \tilde{w}_t = w_t - \mu; \quad Ew_t = \mu.$$

A reziduumok kiszámítására szükség van p darab w^* és q darab a^* kezdeti értékre. Ezek miatt az L függvényt és az S függvényt is csak feltételesen, a w^* és az a^* kezdeti értéktől függően számíthatjuk ki, jelöljük L^* -gal és S^* -gal. Ha a reziduumok normális eloszlását feltételezzük, akkor

$$(2.54) \quad l^*(\phi, \theta, \sigma_a) = -n \ln \sigma_a - \frac{S^*(\phi, \theta)}{2\sigma_a^2}, \quad \text{ahol}$$

$$S^*(\phi, \theta) = \sum_{t=1}^n a_t^2(\phi, \theta, w^*, a^*, w).$$

Konstans σ_a esetén l^* lineáris függvénye S^* -nak, így l^* vizsgálata helyett S^* vizsgálata elegendő. Ahol l^* -nak maximuma van, ott S^* -nak minimuma van. Normális eloszlás feltételezése mellett, a legkisebb négyzetek felhasználásával nyert becslések maximum likelihood becslések is [14], [60].

A reziduumok négyzetösszegének feltételes és feltétel nélküli számítása

Az ARIMA modellek paraméter becslésére alkalmas programcsomagokban lehetőség van a reziduumok négyzetösszegének feltételes és feltétel nélküli választására is. A (2.53) egyenletből látszik, hogy az a_t reziduumok kiszámításához szükséges d -edfokú differenciaképzés után nyert idősor és a reziduumok t időt megelőző $(t-p)$ illetve $(t-q)$ értékei. Tehát a $\underline{w}^*=(w_{t-1}, w_{t-2}, \dots, w_{t-p})$ és $\underline{a}^*=(a_{t-1}, a_{t-2}, \dots, a_{t-q})$ kezdőértékek szükségesek. A kezdőértékek befolyásolják a reziduumok négyzetösszegét, így csak feltételes S^* függvényt lehet meghatározni. Sok esetben a feltételes S^* függvénnyel jól közelíthető a feltétel nélküli S^* függvény. A feltételes S^* függvény kiszámításához szükséges w^* és a^* kezdőértékének a feltétel nélküli várható értéküket vesszük, vagyis Ew^* -t és Ea^* -t. Hosszú idősorok esetén a reziduumok kiszámítását csak a_{p+1} -től kezdjük, kezdőértékként $w_1, w_2, \dots, w_p$ tényleges értékeit, $a_1, a_2, \dots, a_p=0$ -t alkalmazzuk [14], [60].

A reziduumok négyzetösszegének feltétel nélküli kiszámításakor a szükséges kezdőértékeket is becsléssel állapítjuk meg. Ez a módszer bonyolultabb, de előnyösebb, mivel a kezdőértékek megadása torzíthatja az eredményt. A rendelkezésre álló számítógépes programokkal az iteratív módszer is gyorsan elvégezhető. Általában tíztől kisebb iterációs lépésben megkaphatjuk az $S(a)$ függvény minimum értékének jó közelítését.

2.6.3. Sztochasztikus modellek ellenőrzése

A modell becslése után ellenőrizni kell a választott modell helyességét, jóságát. Az ARIMA modellek ellenőrzése a reziduumok alapján történik. Előzőek szerint következőképpen írható fel egy ARIMA folyamat [14]:

$$(2.55) \quad \phi(B)\tilde{w}_t = \theta(B)a_t, \quad \tilde{w}_t = w_t - \mu, \quad w_t = \nabla^d z_t.$$

Ebből kifejezzük a reziduumokat:

$$(2.56) \quad \hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)\tilde{w}_t.$$

Ha a model adekvált, akkor $\hat{a}_t = a_t + o\left(\frac{1}{\sqrt{n}}\right)$.

A reziduumokról, helyes modell választása esetén, Anderson [1], [14] kimutatta, hogy korrelálatlanok és közelítőleg normális eloszlásúak, nulla várható értékkel, n^{-1} szórásnégyzettel, a standard hibájuk $n^{-1/2}$. A ϕ paraméterű AR(I) folyamat esetén az $r_1\left(\hat{a}\right)$ szórásnégyzetéről Durbin igazolta, hogy a $\phi^2 n^{-1}$ -el becsülhető, mely lényegesen kisebb, mint n^{-1} . Tetszőleges ARMA folyamatra Box és Pierce igazolták [15], hogy a becsült reziduumok korrelációs együtthatóinak standard hibáira az $n^{-1/2}$ kifejezés csak mint felső határ alkalmazható, nem mint pontos standard hiba. Ha alacsony késleltetéseknél ezt alkalmazzuk, akkor az eltérések jelentőségét alábecsüljük [32].

A reziduumokról helyes modell választása esetén kimutatható, hogy fehér zaj folyamatot képeznek. Ekkor a reziduumok autokorrelációt nem tartalmaznak, normális eloszlás szerint szóródnak nulla körül, konstans szórással. A reziduumok korrelációs függvénye alkalmazható a

modell helyességének vizsgálatára [14], [25]. A reziduumokat $\hat{a}$, a reziduumok autokorrelációinak becslését $r_k(\hat{a})$ jelöli.

A reziduumok autokorrelációi eloszlásának közelítéseként használható az alábbi kifejezés [14], [25]:

$$(2.57) \quad r_k(\hat{a}) = \frac{\sum_{t=1}^{n-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=1}^n \hat{a}_t^2}.$$

Az autokorrelációk függetlenségük esetén közelítőleg $(K-p-q)$ szabadságfokú χ^2 eloszlást követnek.

$$(2.58) \quad Q = n \sum_{k=1}^K r_k^2(\hat{a}).$$

Ha a modell nem megfelelő, Q értéke erősen megnő [14], [25]. Ezt az ellenőrzési módot portmanteau-próbának nevezzük. A H_0 hipotézis: a megfigyelt folyamat $ARMA(p, q)$. A H_0 hipotézissel szemben a H_1 hipotézis: a megfigyelt folyamat $ARMA(p_1, q_1)$; $p_1 \geq p$; $q_1 \geq q$. Ezt becslést túlillesztésnek nevezzük. Ha a H_0 hipotézist a χ^2 próba nem igazolja, akkor korrigálni kell a választott modellt.

Ha a megfigyelt véletlen folyamat jellemzésére több kiinduló modell választása is helyesnek bizonyul, akkor azt a modellt célszerű a folyamat becslésére továbbiakban felhasználni, amelyikben a paraméterek száma kevesebb. Ha a modell ellenőrzése során bebizonyosodik, hogy nem megfelelő, akkor újból kezdjük a modellkészítés folyamatát egy más típusú modellel [25].

2.7. AR(1) együttható maximum likelihood becslése időben diszkrét esetben

A stacionárius Gauss-Markov-folyamat a $E\xi_n=0$ feltétel mellett a $\xi_n=\phi\xi_{n-1}+\sigma\varepsilon_n$, kifejezéssel állítható elő, ahol ε_n fehérzaj; $n=0, \pm 1, \pm 2, \dots$. Ez elsőrendű autoregresszív folyamatot jelent. Ha r_n a korrelációs-, c_n az autokovariancia függvényt jelöli, akkor érvényesek az alábbi összefüggések [4]:

$$(2.59) \quad \begin{aligned} r_n = \phi^n &= \frac{c_n}{c_0} = \frac{E\xi_{k+n}\xi_k}{D^2\xi_k}, \\ \phi &= \frac{E\xi_k\xi_{k-1}}{D^2\xi_k}, \\ \sigma^2 &= (1-\phi^2)c_0. \end{aligned}$$

A $\xi_1, \xi_2, \dots, \xi_n$ valószínűségi változók együttes sűrűségfüggvénye, feltéve $E\xi_k=m$, a következő alakú [4]:

$$(2.60) \quad \begin{aligned} p_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n) &= \\ &= (2\pi c_0)^{-\frac{n}{2}} (1-\phi^2)^{-\frac{n-1}{2}} * \\ &\exp\left\{-\frac{1}{2c_0(1-\phi^2)}\left[(x_1-m)^2(1-\phi^2) + \sum_{i=2}^n (x_i - \phi x_{i-1} - m(1-\phi))^2\right]\right\} = \\ &= (2\pi)^{-\frac{n}{2}} \sigma^{-n} (1-\phi^2)^{-\frac{1}{2}} * \\ &\exp\left\{-\frac{1}{2\sigma^2}\left[(x_1-m)^2(1-\phi^2) + \sum_{i=2}^n (x_i - \phi x_{i-1} - m(1-\phi))^2\right]\right\}. \end{aligned}$$

A fenti kifejezés mátrix alakban is felírható [4]:

$$(2.61) \quad \begin{aligned} p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) &= \\ &= (2\pi c_0)^{-\frac{n}{2}} (1-\phi^2)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2c_0(1-\phi^2)}(X-m)^T B_n^{-1}(X-m)\right\}. \end{aligned}$$

$$(2.62) \quad B_n^{-1} = \begin{pmatrix} 1 & -\phi & 0 & 0 & . & . & . & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 & . & . & . & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi & . & . & . & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & & & & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & 0 & & & & -\phi & 1 \end{pmatrix}.$$

$(X-m)^T = (x_1-m, \dots, x_n-m)$, T a transzponálást jelenti.

A $\xi_2, \dots, \xi_n$ valószínűségi változók feltételes sűrűségfüggvénye a $\xi_1=x_1$ feltétel mellett [4]:

$$(2.63) \quad p(x_2, \dots, x_n | \xi_1 = x_1) = \\ = (2\pi)^{-\frac{n-1}{2}} \sigma^{-(n-1)} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=2}^n (x_i - \phi x_{i-1} - m(1-\phi))^2\right).$$

A fenti (2.63) kifejezésben szereplő ismeretlenek (m, σ^2, ϕ) megoldására a maximum likelihood becsléssel a kifejezés logaritmusát képezzük. A kifejezést az ismeretlenek szerint parciálisan deriváljuk, és megoldjuk az így kapott egyenleteket [4]. Az elméleti értékek becslésére kapott eredményeket jelölje $\hat{m}, \hat{\phi}, \hat{\sigma}^2$.

$$(2.64) \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=2}^n \left(x_i - \hat{m} - \hat{\phi} (x_{i-1} - \hat{m}) \right)^2,$$

$$(2.65) \quad \hat{\phi} = \frac{\sum_{i=2}^n (x_i - \hat{m})(x_{i-1} - \hat{m})}{\sum_{i=2}^n (x_{i-1} - \hat{m})^2},$$

$$(2.66) \quad \hat{m} = \frac{x_n - \hat{\phi} x_1 + (1 - \hat{\phi}) \sum_{i=2}^n x_i}{(n-1)(1 - \hat{\phi})}.$$

A (2.67) kifejezést a (2.65)-ből kapjuk, $m=0$ esetén.

$$(2.67) \quad \hat{\phi} = \frac{\sum_{i=2}^n x_i x_{i-1}}{\sum_{i=1}^{n-1} x_i^2} \quad .$$

A (2.68) kifejezés szórásnégyzetére igaz, hogy aszimptotikusan $(1-\phi^2)$ -tel egyenlő [4]:

$$(2.68) \quad \sqrt{n-1} \left(\hat{\phi} - \phi \right) \quad .$$

A $\hat{\phi}$ becslés aszimptotikusan normális eloszlású minden fix ϕ értékre, a normális eloszláshoz való konvergencia csak a $-1+\varepsilon < \phi < 1-\varepsilon$ intervallumban egyenletes, ahol $\varepsilon > 0$. Így konfidencia intervallum ϕ -re csak a $(-1, 1)$ intervallum belsejében szerkeszthető [4].

2.8. AR(1) konfidencia intervallumai becsléseinek összehasonlításai

Az AR(1) folyamatot jellemző ϕ paraméter becslésére három eljárás összehasonlítását adtam meg.

Az elsőrendű autoregresszív folyamatot a $\Delta, 2\Delta, \dots, T\Delta$ ($\Delta > 0$) pontokban megfigyelve olyan $\xi(n\Delta) = \xi_n$ diszkrét sorozatot kapunk, amely eleget tesz a $\xi_n = \phi \xi_{n-1} + \sigma \varepsilon_n$ egyenletnek, ahol $\phi = e^{-\lambda \Delta}$; $\sigma^2 = (1-\phi^2)c_0$. Ha Δ fix (legyen $\Delta=1$), ekkor $\phi = e^{-\lambda}$, akkor ϕ becslése aszimptotikusan normális eloszlású, $\left(\phi, \frac{1-\phi^2}{T-1} \right)$ paraméterekkel. A ϕ paraméter becslésére konfidencia intervallum a normális eloszlás – mint közelítés –

alkalmazásával az alábbi összefüggés alapján nyerhető, ahol x_p a normális eloszlás p -kvantilisét, T a megfigyelések számát jelöli [4]:

$$(2.69) \quad \hat{\phi} - x_p \sqrt{\frac{1 - \hat{\phi}^2}{T - 1}} < \phi < \hat{\phi} + x_p \sqrt{\frac{1 - \hat{\phi}^2}{T - 1}}.$$

Ha $\hat{\phi} \approx 1$ és $1 - \hat{\phi} \approx 1/T$, akkor a normális eloszlással való közelítés nem használható. Ekkor a konfidencia intervallum felső határa 1-nél nagyobb is lehet, ami stacionárius esetben használhatatlan eredményt ad [4].

A (2.69) közelítés helyett White a (2.70) közelítést javasolta, ahol t_p az $(T+1)$ szabadságfokú Student eloszlás p -kvantilise [4], [126]:

$$(2.70) \quad \hat{\phi} - t_p \sqrt{\frac{1 - \hat{\phi}^2}{T + 1}} < \phi < \hat{\phi} + t_p \sqrt{\frac{1 - \hat{\phi}^2}{T + 1}}.$$

Időben diszkrét folyamat ϕ paraméterére vonatkozó konfidencia intervallum szerkesztésére felhasználhatók a folytonos folyamatra vonatkozó eredmények [4], [11]. Ha az $AR(1)$ folyamatot a $\Delta, 2\Delta, 3\Delta, \dots, T\Delta$ pontokban megfigyeljük, diszkrét sorozatot kapunk. A folytonos idejű modellből adódó becsléssel az $AR(1)$ folyamatra $\hat{\phi} = \exp(-\Delta \hat{\lambda})$ összefüggés alapján becsülhető a ϕ paraméter. Ha az idősor adatai egyenlő időközönként adóttak a $[0, T]$ intervallumban, akkor

$$\Delta = \frac{1}{T}. \quad \text{Behelyettesítve az előző kifejezésbe: } \hat{\phi} = \exp\left(-\frac{\hat{\lambda}}{T}\right). \quad A$$

$\hat{\lambda}$ paraméter szokásos elnevezése "*csillapodási tényező*", mely utal arra, hogy az $AR(1)$ autokorrelációs együtthatói exponenciálisan vagy exponenciálisan és csillapodó szinusz görbéhez hasonlóan csökkennek. A

kifejezés átalakítása után adódik: $\hat{\lambda} = -T \ln \hat{\phi}$. Vezessük be a $\hat{\kappa} = \hat{\lambda} T$ jelölést. A [4] irodalomban megadott táblázat - „*A csillapodási tényező becslésének az eloszlása*” - tartalmazza a $\hat{\kappa}$ paraméter értékeit, melyek felhasználhatók az $AR(1)$ folyamatok paramétereinek megbízhatóságát jellemző konfidencia intervallumok becslésére [4].

Az $AR(1)$ folyamatokra igaz, hogy a ϕ paraméter értéke az elsőrendű autokorrelációs együtthatóval becsülhető, azaz $r_1 \approx \phi$. Ezt figyelembe véve a fentiek alapján $\hat{\lambda}$ meghatározható, azaz $\hat{\lambda} = -T \ln r_1$. Az említett táblázatban a választott p érték oszlopában találhatók $\hat{\kappa}$ értékek, így visszakereshető a csillapodási tényező értéke. A választott p értékhez tartozó csillapodási tényező jelölése legyen λ_1 . Hasonlóan meghatározható λ_2 is az említett táblázat $(1-p)$ oszlopának felhasználásával. A konfidencia intervallum ϕ_1 és ϕ_2 végpontjai az alábbi kifejezésekkel kaphatók meg [4], [11]:

$$(2.71) \quad \phi_1 = e^{\frac{-\lambda_1}{T}} \quad \text{és} \quad \phi_2 = e^{\frac{-\lambda_2}{T}}.$$

A fentiekben ismertetett módszerrel történt az $AR(1)$ együtthatók konfidencia intervallumainak meghatározása folytonos idejű modellel.

2.9. Számítógépes alkalmazások irodalmának összefoglalása

2.9.1. ARIMA modellek alkalmazásai

Az ARIMA modelleket közgazdasági folyamatok elemzésére gyakran felhasználták, melyekről számos közlemény jelent meg [9], [13], [76],

[86], [93], [109]. Az egészségügyi alkalmazások is széles körben elterjedtek. Ezek közül különös jelentőséggel bír azoknak a WHO adatoknak az elemzése, melyek a nem fertőző illetve a tumoros betegségek okozta halálozási arányok idősori modelljeit ismerteti. Az elemzéssel az 1980-as években Svájcban arra kerestek választ, vajon a környezeti tényezőknek van-e hatása e két betegségcsoport okozta halálozási arányokra [55]. A levegőszennyeződés mértékét és a halálozások számának alakulását sztochasztikus idősorokkal modellezték [89]. Az influenza vírus terjedése és a szövődményes esetek halmozott előfordulása között idősori módszerekkel összefüggést kerestek [117]. A napi és a heti dohányzási szokások alakulását ARIMA modellekkel vizsgálták [103]. Illinois államban idősori módszerekkel elemezték a 65mph sebességhatár bevezetésének a balesetekre, a halálozásokra és a sebesülésekre gyakorolt hatását [102]. Spanyolországban a levegő hőmérséklete és a halálozás közötti összefüggést vizsgálták [105]. Japánban [106], Argentínában [74], Svájcban [33] és a spanyolországi Catalónia tartományban [31] az influenza által előidézett halálozások számának alakulását elemezték. Stockholmban idősori modellekkel vizsgálták az alkoholfogyasztás hatását a halálozásokra és ezzel összefüggő kórházi kezelésekre [75]. ARIMA modellekkel elemezték a homeopátiás kezelésnek a krónikus fejfájásra gyakorolt hatását [121]. A meningococcusok okozta agyhártyagyulladás és az influenza kapcsolatát idősori modellel jellemezték [85]. Madridi adatok alapján a nagy nyári melegnek az időskori halálozásra gyakorolt hatását elemezték [28], [29]. A környezeti zajszint hatására bekövetkezett kórházi felvételek követésére [114], a csecsemőhalandóság vizsgálatára [12], [23], az akut légzőszervi fertőzések jellemzésére [45], a véletlen sérülések

előfordulásának figyelésére [69], a motoros balesetek okozta halálesetek számának alakulására és az öngyilkosságok, illetve gyilkosságok számának vizsgálatára a Box-Jenkins modelleket alkalmazták [100].

A fentieken túl számos közleményben számoltak be más egészségügyi alkalmazásokról: [2], [10], [21], [24], [56], [57], [65], [79], [82], [87], [95], [99], [101], [115], [116]. A halálozási adatok valamint a hazai megbetegedések elemzéséről magyar publikációk is megjelentek, de ezek nem időszori elemzésekkel készültek [40], [61], [63], [64], [71].

2.9.2. Szezonális idősorok alkalmazásai

Az idősorok komponenseire bontásával és ennek alkalmazásaival gyakran találkozhatunk [53], [54], [111], [123]. A szezonális idősorokat számos területen alkalmazták. Ezzel a módszerrel elemezték az Atlanti Óceán plankton és baktérium tartalmának változását [78]. Monte Carlóban a pénzügyi folyamatok [3], az Egyesült Királyságban a negyedéves infláció változásának elemzésére szezonális idősorokat alkalmaztak [38]. A közgazdasági alkalmazásokat kézikönyvben foglalták össze [43]. A mélytengeri üledék változásának jellemzésére [41], a gyorsforgalmi utak forgalmának elemzésére [122] és az ausztráliai [77] és a Hong-Kong-i turisták számának szezonális ingadozásának kimutatására szezonális idősorokat alkalmaztak [44]. USA-ban a légkörben jelenlévő penészgomba spórák mennyiségének elemzését [66], Hollandiában a hőmérsékleti adatok jellemzését [37], Olaszországban az évi csapadék változását [17], Kanadában a nagy esőzések előfordulásának elemzését [124] és USA-ban az időjárási klíma változásának vizsgálatát ezzel a módszerrel végezték [16].

Orvosi alkalmazásokkal szép számban találkozhatunk. Mexikóban a gyermekkori rotavírus fertőzöttség járványát elemezték szezonális idősorokkal [81]. Az asztma előfordulásának szezonálisitását, e betegséggel összefüggő kórházi kezelések, halálozások számának ciklikusságát [36] vizsgálták. Finnországban a tüdőgyulladás következtében szükségessé váló kórházi kezelések szezonális ismétlődését elemezték [107], [119]. Dániában a humán és az állati szalmonella fertőzések trendjét és szezonálisitását vizsgálták [113]. A meningococcusok okozta fertőzések szezonális előfordulásait Spanyolországban elemezték [18]. A stroke betegség ciklikusságát is vizsgálták [88]. Egészségügyi területhez kapcsolódóan további szezonális vizsgálatokról is jelentek meg publikációk [30], [46], [52], [83], [91]. A gyermekkori akut lymphoid leukaemia előfordulásának elemzésével több közleményben is foglalkoztak [22], [49], [84], [108], [118].

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

3.1. Halálozási arányok feldolgozása

3.1.1. Vizsgálati módszerek

A számítógépes feldolgozások során nyolc betegség (cerebrovasculáris betegség, isémiás szívbetegség, krónikus májbetegség, légzőszervi betegségek, méhnyakrák, mellrák, emésztőszervi daganatok, légzőszervi daganatok) okozta halálozási arányok elemzését végeztem el idősor analízis módszerével. A halálozási arányok elemzéséhez az adatokat a WHO adatbázisából szereztem be. Az adatbázis 1970-től 1997-ig tartalmazta az adatokat. Az elemzések során betegségenként (életkor vagy nem alapján) két-két idősort vizsgáltam. A számítások során az idősor elemeiből az átlagot, a trendet és a szezonális komponenszt levontam. Továbbiakban az idősor véletlen komponensének azonosítását adtam meg. A vizsgálatok során a fontosabb lépések a következők voltak:

- az autokorrelációs és parciális autokorrelációs függvények vizsgálata,
- a sztochasztikus egyenletek megadása,
- a modell választás helyességének igazolása,
- a reziduumok ellenőrzése,
- a keresztkorrelációs függvények vizsgálata,
- konfidencia intervallumok becslései.

Idősor analízissel azt vizsgáltam, hogy a betegségenként képzett két-két idősor vajon hasonlóképpen viselkedik-e, azaz a függvények „szinkronizáltak”-e.

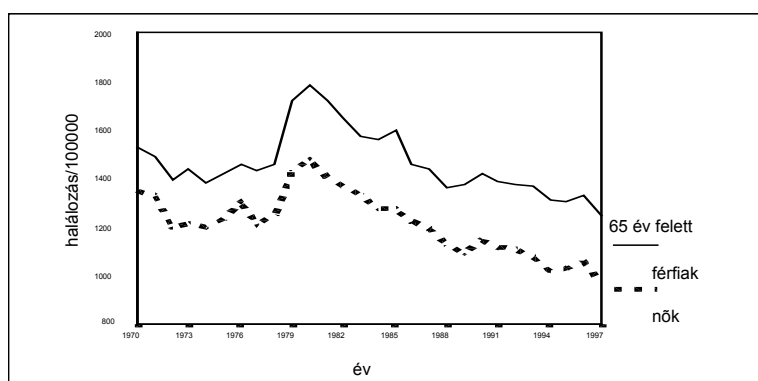
A számítógépes alkalmazások során a leukémiás megbetegedések adatait is vizsgáltam. Az adatokat a Magyarországi Gyermekonkológiai Munkacsoport adatbázisából szereztem be.

Az elemzéseket mind a kétféle alkalmazásnál az SPSS9.0 programcsomaggal végeztem [110].

3.1.2. Eredmények

3.1.2.1. Cerebrovasculáris betegség

A cerebrovasculáris betegség okozta halálozási arányok elemzését a 65 év feletti férfiak (1. idősor) és nők (2. idősor) adataira végeztem el (1. ábra).

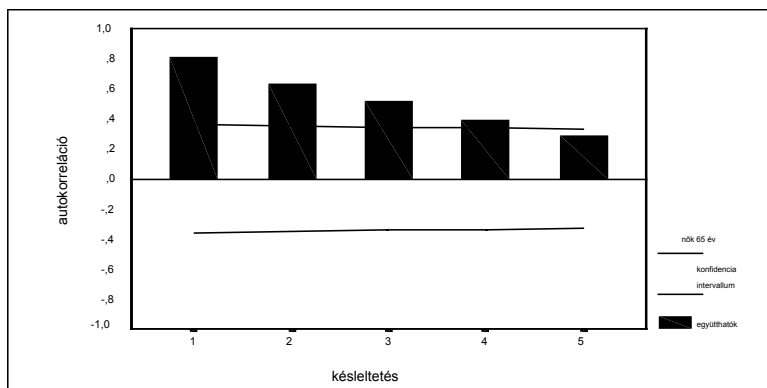


1. ábra: Cerebrovasculáris betegség okozta halálozási arányok a 65 év feletti férfiakra és nőkre.

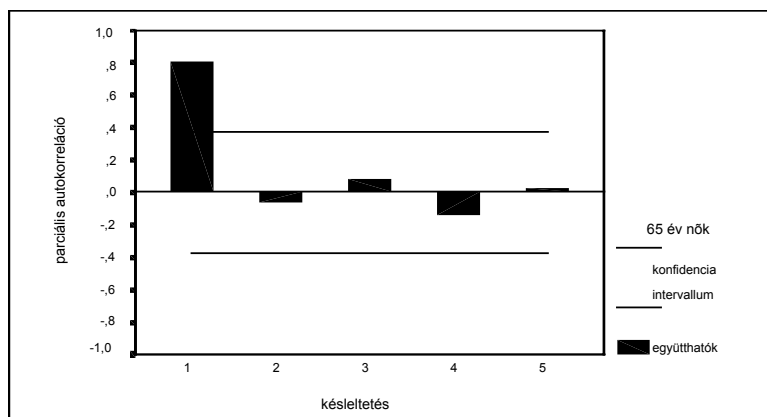
A 2.6. fejezetben foglaltak figyelembe vételével elkészítettem az autokorrelációs- és a parciális autokorrelációs függvényeket (2/a. és 2/b. ábra). A lehetséges modelleket kiválasztottam. A férfiak és a nők

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

halálozási arányaiból készült autokorrelációs függvényértékek exponenciálisan csökkentek (2/a. ábra) és a parciális autokorrelációs függvények egy szignifikáns értéket mutattak $k=1$ esetben (2/b. ábra). Ezekből arra lehet következtetni, hogy $AR(1)$ megfelelő modell. A számítógépes programcsomag alkalmazásával a modellek illesztését elvégeztem, megkaptam a reziduumok varianciáit.



2/a. ábra: Cerebrovasculáris betegség okozta halálozási arányok autokorrelációs függvénye 65 év feletti nők adataiból.



2/b. ábra: Cerebrovasculáris betegség okozta halálozási arányok parciális autokorrelációs függvénye a 65 év feletti nők adataiból.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

Az 3. ábra a férfiak, a 4. ábra a nők halálozási arányaiból számolt autokorrelációs függvények értékeit és standard hibáit mutatja. Az 5. ábra a férfiak, a 6. ábra a nők parciális autokorrelációs függvényeinek értékeit és standard hibáit ábrázolja.

```
Autocorrelations:
      Auto- Stand.
Lag  Corr.   Err. -1  -.75  -.5  -.25   0   .25  .5   .75
1
□↓↓↓↓↓↑↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓□
1  ,792   ,179                .      ⇔*****.*****
2  ,588   ,176                .      ⇔*****.*****
3  ,411   ,173                .      ⇔*****.*
4  ,233   ,169                .      ⇔*****.
5  ,115   ,165                .      ⇔**
6  -,024  ,162                .      *
Plot Symbols:      Autocorrelations *      Two Standard
Error Limits .
```

3. ábra: A férfiak halálozási arányaiból számolt autokorrelációk és standard hibái.

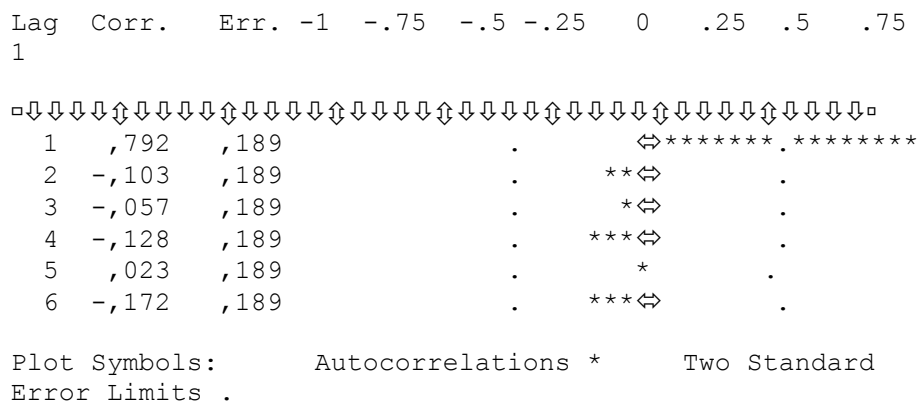
```
Autocorrelations:
      Auto- Stand.
Lag  Corr.   Err. -1  -.75  -.5  -.25   0   .25  .5   .75
1
□↓↓↓↓↓↑↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓↑↑↑↑↓□
1  ,809   ,179                .      ⇔*****.*****
2  ,632   ,176                .      ⇔*****.*****
3  ,522   ,173                .      ⇔*****.***
4  ,388   ,169                .      ⇔*****.*
5  ,286   ,165                .      ⇔*****.
6  ,161   ,162                .      ⇔***
Plot Symbols:      Autocorrelations *      Two Standard
Error Limits .
```

4. ábra: A nők halálozási arányaiból számolt autokorrelációk és standard hibái.

Partial Autocorrelations:

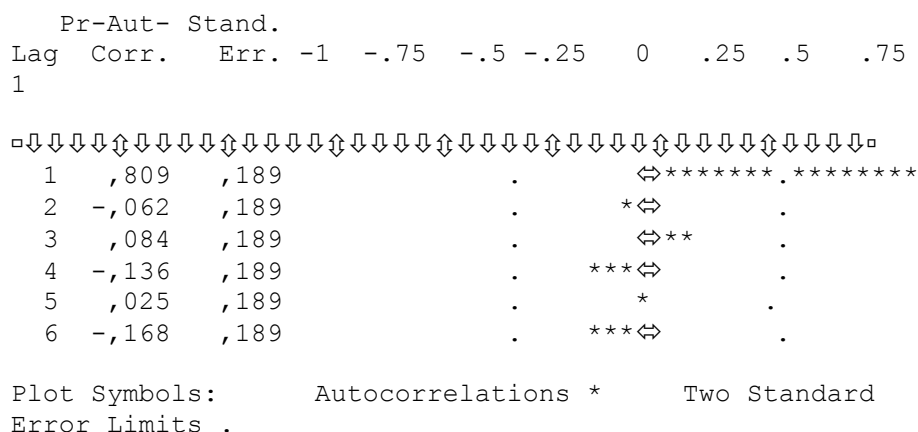
Pr-Aut- Stand.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK



5. ábra: A férfiak halálozási arányaiból számolt parciális autokorrelációk és standard hibái.

Partial Autocorrelations:



6. ábra: A nők halálozási arányaiból számolt parciális autokorrelációk és standard hibái.

A programcsomag alkalmazásával a paraméterek becslésre kerültek. Ezekután a következő egyenleteket kaptam. A férfiak halálozási arányát leíró sztochasztikus egyenlet: $z_t = 0,792z_{t-1} + \varepsilon_t$, $\sigma_a^2 = 5542,023$; a nők adataira: $z_t = 0,809z_{t-1} + \varepsilon_t$, $\sigma_a^2 = 3816,344$. A σ_a^2 -t, a reziduumok varianciáit, a program megadta.

Az elsőrendű autoregresszív folyamatok paramétereinek megbízhatóságára a 2.8. fejezetben ismertetett háromféle módon meghatározott konfidencia intervallumokat adtam meg. Az $AR(1)$ folyamatok konfidencia intervallumai becslési módszerei az alábbiakban $p=0,05$ -ra részletesen ismertetésre kerülnek. A normális eloszlás alkalmazásához és a White féle közelítéshez a szükséges $x_{0,05}$ és $t_{0,05}$ értékeket a standard normális eloszlás illetve a Student eloszlás ismert táblázataiból kikerestem. A folytonos idejű modellel a konfidencia intervallumok meghatározása a 2.8. fejezetben megadott (2.69); (2.70); (2.71) kifejezésekkel történt, ahol $T=28$, a ϕ paraméter közelítő értéke az elsőrendű autokorrelációs együttható, azaz r_1 .

Következőkben részletesen ismertetem a konfidencia intervallum meghatározását folytonos idejű modellből adódó becslés alkalmazásával. A cerebrovasculáris betegség okozta halálozási arányok esetében a nők adataiból becsült autokorrelációs együtthatóra ismertetem a számítás menetét. Továbbiakban a konfidencia intervallumok meghatározásai folytonos idejű becsléssel hasonló számítási elvek alapján történt. $\hat{\phi} \approx r_1$;
 $r_1=0,809$; $T=28$; $p=0,05$; $1-p=0,95$; $\hat{\kappa} = -T \ln r_1 = 5,9347$; $\lambda_1=0,59468$;
 $\lambda_1/T=0,02123$; $\phi_1=e^{-0,012138}=0,9789$; $\lambda_2=10,602578$; $\lambda_2/T=0,3786$;
 $\phi_2=e^{-0,3786}=0,6848$. Így a folytonos idejű becsléssel kapott konfidencia intervallum: $(0,6848; 0,9789)$.

A férfiak halálozási arányaiból meghatározott $AR(1)$ együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 3. táblázat.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

3. táblázat: $AR(1)$ együttható konfidencia intervallumai p különböző értékeinél a férfiak adataiból.

$\phi \approx 0,792$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,5956;0,9884)	(0,5523;1,0317)	(0,4736;1,1104)
White féle	(0,5993;0,9847)	(0,5596;1,0244)	(0,4791;1,1105)
Folytonos becs.	(0,6949;0,9424)	(0,6649;0,9723)	(0,6116;0,9978)

A konfidencia intervallumok összehasonlítását a 4. táblázatban foglaltam össze.

4. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,792$	$p=0,1$		$p=0,05$		$p=0,01$	
Normális elosz.	0,3928	100%	0,4794	100%	0,6368	100%
White féle	0,3854	98,1%	0,4648	96,9%	0,6314	99,5%
Folytonos becs.	0,2475	63%	0,3074	64,1%	0,3862	60,5%

A nők halálozási arányaiból meghatározott $AR(1)$ együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza az 5. táblázat.

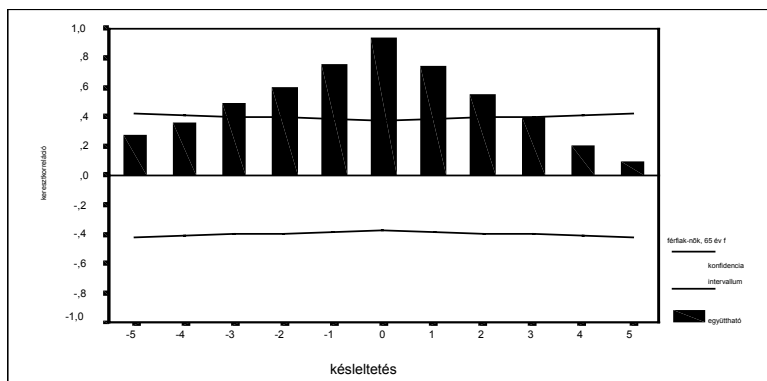
5. táblázat: $AR(1)$ együttható konfidencia intervallumai p különböző értékeinél a nők adataiból.

$\phi \approx 0,809$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,6201;0,9979)	(0,5783;1,0397)	(0,5025;1,1155)
White féle	(0,6235;0,9945)	(0,5852;1,0327)	(0,5078;1,1103)
Folytonos becs.	(0,7149;0,9527)	(0,6848;0,9788)	(0,6329;0,9985)

A konfidencia intervallumok összehasonlítását tartalmazza a 6. táblázat.

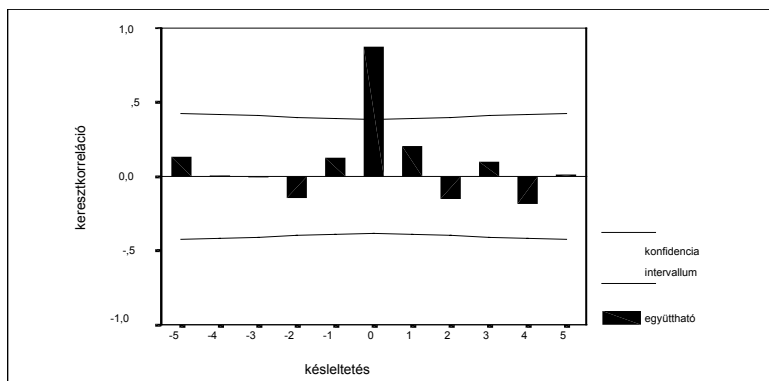
6. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,809$	$p=0,1$	$p=0,05$	$p=0,01$
----------------------	---------	----------	----------



8. ábra: Cerebrovasculáris betegség okozta halálozási arányokból a keresztkorrelációs függvény a férfiak és a nők adataiból.

Az ábrán folytonos vonal jelöli a konfidencia intervallumot, a nulla körüli kétszeres standard hibahatárt. A nők és a férfiak halálozási arányai idősorokból képzett keresztkorrelációs függvény együtthatói nulla körüli kétszeres standard hiba intervallumon kívül és belül is előfordultak (8. ábra). A reziduumok idősorokból képzett keresztkorrelációs függvénynek $k=0$ esetben szignifikáns értéke volt (0,865). A többi együttható nullától szignifikánsan nem különbözött (9. ábra). Az $AR(1)$ modellek illesztése után a reziduumok keresztkorrelációs függvénye olyan időkésleltetésnél adott szignifikáns értéket, mikor az idősorok között korreláció van [14], [50], [51], [55]. Az idősorok hasonlóan viselkedtek, azaz az idősorok „szinkronizáltak” voltak. Azokban az években, mikor növekedett a férfiak halálozási aránya, akkor a nőké is növekedett, ugyanez igaz a csökkenésekre is.



9. ábra: Cerebrovasculáris betegség okozta halálozási arányok reziduuumainak keresztkorrelációs függvénye.

3.1.2.2. Isémiás szívbetegség

Az isémiás szívbetegség okozta halálozási arányokat a 0-64 éves korcsoportban a férfiak (1. idősor) és a nők (2. idősor) adataira elemeztem.

Az adatok vizsgálatát a 3.1.2.1. fejezetben leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatóaknak bizonyultak. A férfiak adataiból a sztochasztikus egyenlet a következőnek adódott: $z_t = 0,884z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 39,291$; a nők adataiból az egyenlet: $z_t = 0,720z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 3,684$.

A konfidencia intervallumok becslését a 2.8. fejezetben ismertetett háromféle módon végeztem el. A férfiak halálozási arányaiból meghatározott AR(1) együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 7. táblázat.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

7. táblázat: $AR(1)$ együttható konfidencia intervallumai p különböző értékeinél a férfiak csoportjában.

$\phi \approx 0,884$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,7338;1,0342)	(0,7005;1,0675)	(0,6402;1,1278)
White féle	(0,7364;1,0316)	(0,706;1,0619)	(0,6444;1,1236)
Folytonos becs.	(0,8095;0,9864)	(0,7828;0,9579)	(0,7332;0,9725)

A konfidencia intervallumok összehasonlítását tartalmazza a 8. táblázat.

8. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,884$	$p=0,1$		$p=0,05$		$p=0,01$	
Normális elosz.	0,3004	100%	0,367	100%	0,4876	100%
White féle	0,2952	98,3%	0,3559	96,9%	0,4792	98,2%
Folytonos becs.	0,1769	58,8%	0,1751	47,7%	0,2393	49%

A nők halálozási arányaiból meghatározott $AR(1)$ együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 9. táblázat.

9. táblázat: $AR(1)$ együttható konfidencia intervallumai p különböző értékeinél a nők csoportjában.

$\phi \approx 0,72$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,497;0,943)	(0,4476;0,9924)	(0,3581;1,0819)
White féle	(0,501;0,939)	(0,456;0,984)	(0,3644;1,0756)
Folytonos becs.	(0,615;0,8906)	(0,5826;0,9323)	(0,5242;0,9894)

A konfidencia intervallumok összehasonlítását tartalmazza a 10. táblázat.

10. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,72$	p=0,1		p=0,05		p=0,01	
Normális elosz.	0,446	100%	0,5448	100%	0,7238	100%
White féle	0,438	98,2%	0,528	96,9%	0,7112	98,2%
Folytonos becs.	0,2756	61,7%	0,3497	64,1%	0,4652	64,3%

A helyes modell választás igazolásához a (2.58) kifejezéssel megadott χ^2 próbát alkalmaztam. A reziduumok autokorrelációiból az alábbi értékek adódtak: $\chi^2_{\text{férfiak}} = 10,795$; $\chi^2_{\text{nők}} = 6,56$. Mivel $\chi^2_{0,05} = 11,07$, így a választott modellek helyesnek bizonyultak [25], [50], [51].

A reziduumok autokorrelációi nem tartalmaztak korreláltságot, a 7. ábrához hasonlókat kaptam rájuk.

A férfiak és a nők halálozási arányait jellemző idősorok közötti kapcsolat elemzéséhez a keresztkorrelációs függvényt alkalmaztam. A keresztkorrelációs függvény a modell illesztése előtt hasonlóan viselkedett, mint a cerebrovasculáris betegségnél. A modell illesztése után a reziduumok keresztkorrelációs függvényének $k=0$ értéknél szignifikáns értéke volt (0,733) 95%-os szignifikancia szinten. Ebből arra következtethetünk, hogy szinkronizáció van az idősorok között. A szinkronizáció alatt azt értjük, hogy azokban az években, mikor a férfiak halálozási aránya növekedett, akkor a nők is növekedett. Hasonló igaz a csökkenésekre is.

3.1.2.3. Krónikus májbetegség

A krónikus májbetegség (májzsugor) okozta halálozási arányokat a férfiak (1. idősor) és a nők (2. idősor) adataira elemeztem.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

Az adatok vizsgálatát a 3.1.2.1. fejezetben leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatóaknak bizonyultak. A férfiak adataiból a sztochasztikus egyenlet a következőnek adódott: $z_t = 0,919z_{t-1} + \varepsilon_t$ $\sigma_a^2 = 81,117$; nőkre az egyenlet: $z_t = 0,917z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 8,631$.

A konfidencia intervallumok becslését a 2.8. fejezetben ismertetett háromféle módon végeztem el. A férfiak halálozási arányaiból meghatározott $AR(1)$ együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 11. táblázat.

11. táblázat: $AR(1)$ együttható konfidencia intervallumai p különböző értékeinél a férfiak csoportjában.

$\phi \approx 0,919$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,7922;1,04558)	(0,7642;1,0737)	(0,7134;1,1246)
White féle	(0,7945;1,0435)	(0,7689;1,069)	(0,717;1,121)
Folytonos becs.	(0,858;0,9942)	(0,8342;0,9984)	(0,7898;0,9998)

A konfidencia intervallumok összehasonlítását tartalmazza a 12. táblázat.

12. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,919$	$p=0,1$		$p=0,05$		$p=0,01$	
Normális elosz.	0,2536	100%	0,3095	100%	0,4112	100%
White féle	0,249	98,2%	0,3001	96,9%	0,404	98,2%
Folytonos becs.	0,1362	53,7%	0,1642	53,1%	0,21	51%

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

A nők halálozási arányaiból meghatározott AR(1) együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 13. táblázat.

13. táblázat: AR(1) együttható konfidencia intervallumai p különböző értékeinél a nők csoportjában.

$\phi \approx 0,917$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,7887;1,0452)	(0,7604;1,0736)	(0,709;1,125)
White féle	(0,7911;1,0429)	(0,7652;1,0688)	(0,7126;1,1214)
Folytonos becs.	(0,8551;0,9939)	(0,8314;0,9983)	(0,7863;0,9998)

A konfidencia intervallumok összehasonlítását tartalmazza a 14. táblázat.

14. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,917$	$p=0,1$		$p=0,05$		$p=0,01$	
Normális elosz.	0,2565	100%	0,3132	100%	0,416	100%
White féle	0,2518	98,2%	0,3036	96,9%	0,4088	98,3%
Folytonos becs.	0,1388	54,1%	0,1669	53,3%	0,2135	51,3%

A helyes modell választás igazolásához a (2.58) kifejezéssel megadott χ^2 próbát alkalmaztam. A reziduumok autokorrelációiból a következő értékek adódtak: $\chi^2_{\text{férfiak}}=8,971$; $\chi^2_{\text{nők}}=2,990$. Mivel $\chi^2_{0,05}=11,07$, így a választott modellek helyesnek bizonyultak [25], [47], [50], [51].

A reziduumok autokorrelációi nem tartalmaztak korreláltságot, a 7. ábrához hasonlókat kaptam rájuk.

A férfiak és a nők halálozási arányait jellemző idősorok közötti kapcsolat elemzéséhez a keresztkorrelációs függvényt alkalmaztam. A

keresztkorrelációs függvény a modell illesztése előtt hasonlóan viselkedett, mint a cerebrovasculáris betegségnél. A modell illesztése után a reziduumok keresztkorrelációs függvényének $k=0$ értéknél szignifikáns értéke volt 95%-os szignifikancia szinten (0,953). Ebből arra következtethetünk, hogy szinkronizáció van az idősorok között. A szinkronizáció alatt azt értjük, hogy azokban az években, mikor a férfiak halálozási aránya növekedett, akkor a nők is növekedett. Hasonló igaz a csökkenésekre is.

3.1.2.4. Légzőszervi betegségek

A légzőszervi betegségek (bronchitis, tüdő emphysema, asztma) okozta halálozási arányokat a férfiak (1. idősor) és a nők (2. idősor) adataira hasonlítottam össze.

Az adatok vizsgálatát a 3.1.2.1. fejezetben leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatóak. A férfiak adataiból a következő sztochasztikus egyenlet kaptam: $z_t = 0,78z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 62,825$; a nők adataiból az egyenlet: $z_t = 0,702z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 14,138$.

A konfidencia intervallumok becslését a 2.8. fejezetben ismertetett háromféle módon végeztem el. A számítások eredményeit az alábbi táblázatokban foglaltam össze.

A férfiak halálozási arányaiból meghatározott $AR(1)$ együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 15. táblázat.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

15. táblázat: $AR(1)$ együttható konfidencia intervallumai p különböző értékeinél a férfiak csoportjában.

$\phi \approx 0,78$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,5258;1,0342)	0,4695;1,0905)	(0,3654;1,0925)
White féle	(0,5303;1,0297)	(0,4781;1,0819)	(0,3746;1,1854)
Folytonos becs.	(0,6812;0,9345)	0,6567;0,9667)	(0,5963;0,9971)

A konfidencia intervallumok összehasonlítását tartalmazza a 16. táblázat.

16. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,78$	$p=0,1$		$p=0,05$		$p=0,01$	
Normális elosz.	0,5084	100%	0,6255	100%	0,8271	100%
White féle	0,4994	98,2%	0,6038	96,5%	0,8108	98%
Folytonos becs.	0,2533	49,8%	0,31	49,6%	0,4008	48,4%

A nők halálozási arányaiból meghatározott $AR(1)$ együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 17. táblázat.

17. táblázat: $AR(1)$ együttható konfidencia intervallumai p különböző értékeinél.

$\phi \approx 0,702$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,4731;0,9309)	(0,4224;0,9816)	(0,3306;1,0734)
White féle	(0,4773;0,9267)	(0,431;0,973)	(0,3371;1,0669)
Folytonos becs.	(0,5953;0,876)	(0,5625;0,9198)	(0,5031;0,9847)

A konfidencia intervallumok összehasonlítását tartalmazza a 18. táblázat.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

18. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,702$	p=0,1		p=0,05		p=0,01	
Normális elosz.	0,4578	100%	0,5592	100%	0,7428	100%
White féle	0,4494	98,2	0,542	96,9%	0,7298	98,2%
Folytonos becs.	0,2807	61,3	0,3573	63,9%	0,4816	64,8%

A helyes modell választás igazolásához a (2.58) kifejezéssel megadott χ^2 próbát alkalmaztam. A reziduumok autokorrelációiból számolt $\chi^2_{\text{férfiak}} = 0,9983$; $\chi^2_{\text{nők}} = 1,3399$ értékek adódtak. Mivel $\chi^2_{0,05} = 11,07$, így a választott modellek helyesnek bizonyultak [25], [50], [51].

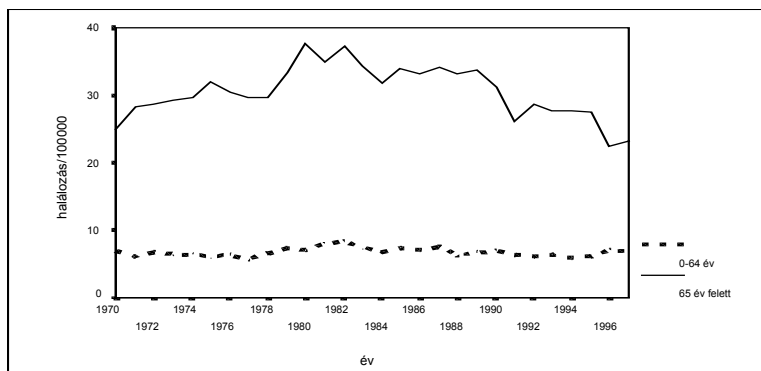
A reziduumok autokorrelációi nem tartalmaztak korreláltságot, a 7. ábrához hasonló ábrát kaptam rájuk.

A férfiak és a nők halálozási arányait jellemző idősorok közötti kapcsolat elemzéséhez a keresztkorrelációs függvényt alkalmaztam. A keresztkorrelációs függvény a modell illesztése előtt hasonlóan viselkedett, mint a cerebrovasculáris betegségnél. A reziduumok keresztkorrelációs függvényének a modell illesztése után a $k=0$ értéknél szignifikáns értéke (0,910) volt 95%-os szignifikancia szinten. Ebből arra következtethetünk, hogy szinkronizáció van az idősorok között. A szinkronizáció alatt azt értjük, hogy azokban az években, mikor a férfiak halálozási aránya növekedett, akkor a nők is növekedett. Hasonló igaz a csökkenésekre is.

3.1.2.5. Méhnyakrák

A méhnyakrák okozta halálozási arányokat a 0-64 éves (1. idősor) és 65 év feletti (2. idősor) korcsoportok adataira elemeztem (10. ábra).

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK



10. ábra: Méhnyakrák okozta halálozási arányok a 0-64 és a 65 év feletti korcsoportokban.

Az elemzés hasonlóan történt, mint az előző halálozási arányok esetében. A 0-64 éves korcsoport halálozási arányát jellemző sztochasztikus kifejezés: $z_t = 0,576z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 0,260$; a 65 év felettiekre: $z_t = 0,703z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 5,722$. A reziduumok varianciáit σ_a^2 jelöli.

Az 11/a. ábra a 0-64 éves, a 11/b. ábra a 65 év feletti korcsoportok halálozási arányai időszoraiból számolt autokorrelációkat és standard hibáit mutatja. A 12. ábra a 0-64 éves, a 13. ábra a 65 év felettiek halálozási arányaiból számolt parciális autokorrelációs függvényeket és standard hibáit adja meg.

```
Autocorrelations:
      Auto- Stand.
Lag  Corr.   Err. -1  -.75  -.5  -.25   0   .25  .5   .75
1
  1   ,576   ,179                .   ⇔*****.*
  2   ,387   ,176                .   ⇔*****.*
  3   ,268   ,173                .   ⇔*****.
  4   ,076   ,169                .   ⇔**
  5  -,131   ,165                .   ***⇔
  6  -,144   ,162                .   ***⇔
Plot Symbols:      Autocorrelations *      Two Standard
Error Limits .
```

11/a. ábra: Méhnyakrák okozta halálozási arányok autokorrelációi és standard hibái a 0-64 éves korcsoportban.

Autocorrelations:

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

```

Auto- Stand.
Lag Corr. Err. -1 -.75 -.5 -.25 0 .25 .5 .75
1
□↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓□
1 ,703 ,179 . ⇔*****.*****
2 ,484 ,176 . ⇔*****.***
3 ,368 ,173 . ⇔*****
4 ,248 ,169 . ⇔*****
5 ,214 ,165 . ⇔*****
6 ,120 ,162 . ⇔**

Plot Symbols: Autocorrelations * Two Standard
Error Limits .

```

11/b. ábra: Méhnyakrák okozta halálozási arányok autokorrelációi és standard hibái a 65 év feletti korcsoportban.

```

Partial Autocorrelations:
Pr-Aut- Stand.
Lag Corr. Err. -1 -.75 -.5 -.25 0 .25 .5 .75
1
□↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓↑↓↓↓↓↓□
1 ,576 ,189 . ⇔*****.*****
2 ,083 ,189 . ⇔**
3 ,024 ,189 . *
4 -,160 ,189 . ***⇔
5 -,221 ,189 . *****⇔
6 ,037 ,189 . ⇔*

Plot Symbols: Autocorrelations * Two Standard
Error Limits .

```

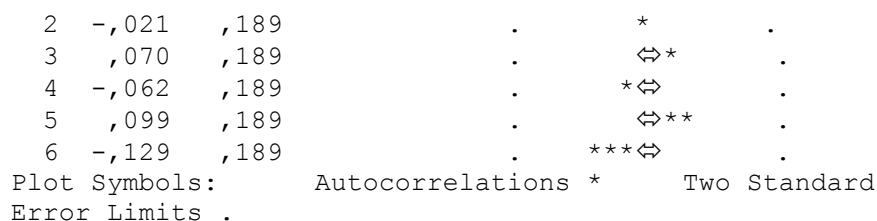
12. ábra: Méhnyakrák okozta halálozási arányok parciális autokorrelációi és standard hibái a 0-64 éves korcsoportban.

```

Partial Autocorrelations:
Pr-Aut- Stand.
Lag Corr. Err. -1 -.75 -.5 -.25 0 .25 .5 .75
1
1 ,703 ,189 . ⇔*****.*****

```

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK



13. ábra: Méhnyakrák okozta halálozási arányok parciális autokorrelációi és standard hibái a 65 év feletti korcsoportban.

Az $AR(1)$ együtthatók becslésére a konfidencia intervallumokat $p=0,1$; $p=0,05$ és $p=0,01$ -ra az előző fejezetben leírtak szerint elkészítettem, a kapott számítási eredményeket az alábbi táblázatokban foglaltam össze.

19. táblázat: $AR(1)$ együttható konfidencia intervallumai p különböző értékeinél a 0-64 éves korcsoportban.

$\phi \approx 0,576$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,3133;0,8387)	(0,2551;0,8969)	(0,1497;1,0002)
White féle	(0,3183;0,8337)	(0,3112;0,8872)	(0,1576;0,9944)
Folytonos becs.	(0,4634;0,7609)	(0,4319;0,8139)	(0,4442;0,9157)

A konfidencia intervallumok összehasonlítását tartalmazza a 20. táblázat.

20. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,576$	$p=0,1$		$p=0,05$		$p=0,01$	
Normális elosz.	0,5254	100%	0,6418	100%	0,8505	100%
White féle	0,5154	98,1%	0,576	89,7%	0,8368	98,4%
Folytonos becs.	0,2934	55,8%	0,382	59,5%	0,4715	55,4%

21. táblázat: $AR(1)$ együttható konfidencia intervallumai p különböző értékeinél a 65 év feletti korcsoportban.

$\phi \approx 0,703$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,4743;0,9317)	(0,4238;0,9822)	(0,3321;1,0739)

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

White féle	(0,4785;0,9275)	(0,4323;0,9737)	(0,3386;1,0674)
Folytonos becs.	(0,5964;0,8769)	(0,5636;0,9205)	(0,5043;0,985)

A konfidencia intervallumok összehasonlítását tartalmazza a 22. táblázat.

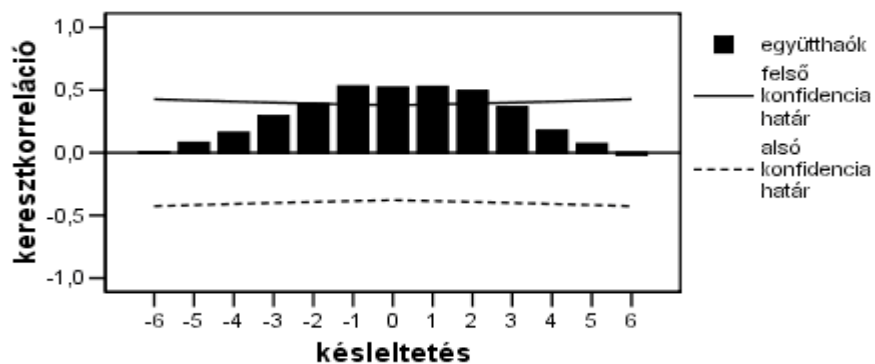
22. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,703$	p=0,1		p=0,05		p=0,01	
Normális elosz.	0,4574	100%	0,5584	100%	0,7418	100%
White féle	0,449	98,2%	0,5414	96,9%	0,7288	98,2%
Folytonos becs.	0,2805	61,3%	0,3569	63,9%	0,4807	64,8%

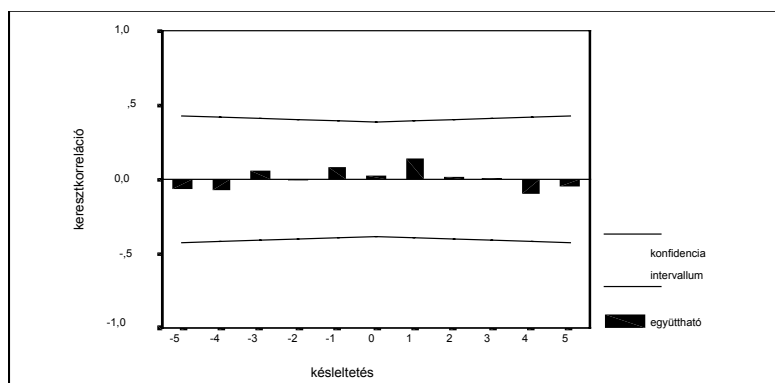
A reziduumok autokorrelációiból számolt $\chi^2_{0-64}=1,956$; $\chi^2_{65felett}=2,304$; és $\chi^2_{0,05}=11,07$ alapján a modellek választása helyesnek bizonyult [25], [50], [55].

A reziduumok autokorrelációi nem tartalmaztak korreláltságot, a 7. ábrához hasonlóan kaptam rájuk.

A két korcsoport halálozási arányainak kapcsolatát a keresztkorrelációs függvényekkel vizsgáltam (14. ábra). A modell illesztése előtt a keresztkorrelációs függvény értékei a kétszeres standard hiba intervallumon kívül és belül is előfordultak. A reziduumok keresztkorrelációs függvényének értékei a nulla körüli kétszeres standard hiba intervallumon belül maradtak, szignifikáns érték nem adódott (15. ábra). A keresztkorrelációs függvény legnagyobb értéke 0,073. Ilyen esetben a két idősor között kapcsolat nem mutatható ki [14], [50], [55], az idősorok nem viselkedtek hasonlóan, azaz az idősorok között "szinkronizáció" nem mutatkozott.



14. ábra: Méhnyakrák okozta halálozási arányokból a 0-64 éves és a 65 év feletti nők korcsoportjainak keresztkorrelációs függvénye.



15. ábra: Méhnyakrák okozta halálozási arányokból a 0-64 éves és a 65 év feletti nők korcsoportja reziduumaiknak keresztkorrelációs függvénye.

3.1.2.6. Mellrák

A mellrák okozta halálozási arányokat 0-64 éves (1. idősor) és a 65 év feletti (2. idősor) korcsoportok adataira elemeztem.

Az adatok vizsgálatát a 3.1.2.1. fejezetben leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatóaknak bizonyultak. A 0-64 éves korcsoportra a sztochasztikus

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

egyenlet az alábbiak adódott: $z_t = 0,762z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 0,797$; a 65 év felettiekre az egyenlet: $z_t = 0,818z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 50,197$.

A 0-64 éves korcsoport halálozási arányaiból meghatározott AR(1) együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 23. táblázat.

23. táblázat: AR(1) együttható konfidencia intervallumai p különböző értékeinél a 0-64 éves korcsoportban.

$\phi \approx 0,762$	p=0,1	p=0,05	p=0,01
Normális elosz.	(0,554;0,97)	(0,508;1,016)	(0,4244;1,0996)
White féle	(0,5576;0,9664)	(0,5155;1,0085)	(0,4301;1,0939)
Folytonos becs.	(0,6613;0,9223)	(0,6303;0,9574)	(0,5745;0,9956)

A konfidencia intervallumok összehasonlítását mutatja a 24. táblázat.

24. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,762$	p=0,1		p=0,05		p=0,01	
Normális elosz.	0,416	100%	0,508	100%	0,6752	100%
White féle	0,4088	98,3%	0,493	97%	0,6638	98,3%
Folytonos becs.	0,261	62,7%	0,3271	64,4%	0,4211	62,5%

A 65 év felettiek halálozási arányaiból meghatározott AR(1) együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 25. táblázat.

25. táblázat: AR(1) együttható konfidencia intervallumai p különböző értékeinél a 65 év feletti korcsoportban.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

$\phi \approx 0,818$	p=0,1	p=0,05	p=0,01
Normális elosz.	(0,6331;1,0029)	(0,5922;1,044)	(0,5181;1,1179)
White féle	(0,6364;0,9996)	(0,5991;1,0369)	(0,5233;1,1127)
Folytonos becs.	(0,7257;0,9577)	(0,6954;0,9822)	(0,6444;0,9987)

A konfidencia intervallumok összehasonlítását mutatja a 26. táblázat.

26. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,818$	p=0,1		p=0,05		p=0,01	
Normális elosz.	0,3698	100%	0,4518	100%	0,5998	100%
White féle	0,3632	98,2%	0,4378	96,9%	0,5894	98,3%
Folytonos becs.	0,232	62,7%	0,2868	63,5%	0,3543	59,1%

A helyes modell választás igazolásához a (2.58) kifejezéssel megadott χ^2 próbát alkalmaztam. A reziduumok autokorrelációiból a következő értékek adódtak: $\chi^2_{0-64\text{év}} = 6,56$; $\chi^2_{65\text{felett}} = 9,96$. Mivel $\chi^2_{0,05} = 11,07$, így a választott modellek helyesnek bizonyultak [25], [50], [51].

A reziduumok autokorrelációi nem tartalmaztak korreláltságot, a 7. ábrához hasonlókat kaptam rájuk.

A korcsoportokat jellemző idősorok közötti kapcsolat elemzéséhez keresztkorrelációs függvényeket alkalmaztam. A keresztkorrelációs függvény a modell illesztése előtt hasonlóan viselkedett, mint a méhnyakrákos halálozási arányoknál. A modell illesztése után a reziduumok keresztkorrelációs függvényének szignifikáns értéke nem volt 95%-os szignifikancia szinten. A keresztkorrelációs függvény legnagyobb értéke 0,358. Ebből arra következtethetünk, hogy az idősorok nem viselkedtek hasonlóan, azaz nincs szinkronizáció az idősorok között.

3.1.2.7. Emésztőszervi daganatok

Az emésztőszervi daganatok (gyomor- és vastagbélrák) okozta halálozási arányokat a 65 év feletti férfiak (1. idősor) és a nők (2. idősor) adataira hasonlítottam össze.

Az adatok vizsgálatát a 3.1.2.1. fejezetben leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatóaknak bizonyultak. A férfiak adataiból a sztochasztikus egyenlet az alábbiak adódott: $z_t = 0,742z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 521,097$; nőkre az egyenlet: $z_t = 0,756z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 67,902$.

A férfiak halálozási arányaiból meghatározott AR(1) együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 27. táblázat.

27. táblázat: AR(1) együttható konfidencia intervallumai p különböző értékeinél a 65 év feletti férfiak csoportjában.

$\phi \approx 0,742$	p=0,1	p=0,05	p=0,01
Normális elosz.	(0,5266;0,9574)	(0,4788;1,0052)	(0,3924;1,0916)
White féle	(0,5304;0,9537)	(0,4868;0,9972)	(0,3984;1,0856)
Folytonos becs.	(0,6392;0,9076)	(0,6074;0,9462)	(0,5503;0,9932)

A konfidencia intervallumok összehasonlítását tartalmazza a 28. táblázat.

28. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,742$	p=0,1		p=0,05		p=0,01	
Normális elosz.	0,4308	100%	0,5264	100%	0,6992	100%
White féle	0,4233	98,3%	0,5104	96,9%	0,6872	98,3%

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

Folytonos becs.	0,2684	62,3%	0,3388	64,4%	0,4429	63,3%
-----------------	--------	-------	--------	-------	--------	-------

A nők halálozási arányaiból meghatározott AR(1) együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 29. táblázat.

29. táblázat: AR(1) együttható konfidencia intervallumai p különböző értékeinél a 65 év feletti nők csoportjában.

$\phi \approx 0,756$	$p=0,1$	$p=0,05$	$p=0,01$
Normális elosz.	(0,5456;0,9664)	(0,499;1,013)	(0,4146;1,0974)
White féle	(0,5494;0,9626)	(0,507;1,005)	(0,4205;1,0915)
Folytonos becs.	(0,6547;0,9180)	(0,6234;0,9541)	(0,5672;0,9949)

A konfidencia intervallumok összehasonlítását tartalmazza a 30. táblázat.

30. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,756$	$p=0,1$		$p=0,05$		$p=0,01$	
Normális elosz.	0,4208	100%	0,514	100%	0,6828	100%
White féle	0,4132	98,2%	0,498	96,9%	0,671	98,3%
Folytonos becs.	0,2633	62,6%	0,3307	64,3%	0,4277	62,6%

A helyes modell választás igazolásához a (2.58) kifejezéssel megadott χ^2 próbát alkalmaztam. A reziduumok autokorrelációiból a következő értékek adódtak: $\chi^2_{\text{férfiak}}=8,475$; $\chi^2_{\text{nők}}=5,794$. Mivel $\chi^2_{0,05}=11,07$, így a választott modellek helyesnek bizonyultak [25], [50], [51].

A reziduumok autokorrelációi nem tartalmaztak korreláltságot, a 7. ábrához hasonlót kaptam rájuk.

A férfiak és a nők halálozási arányait jellemző idősorok közötti kapcsolat elemzéséhez keresztkorrelációs függvényeket alkalmaztam. A keresztkorrelációs függvény a modell illesztése előtt hasonlóan viselkedett, mint a méhnyakrák okozta halálozási arányoknál. A reziduumok keresztkorrelációs függvényének a modell illesztése után szignifikáns értéke nem volt 95%-os szignifikancia szinten. A keresztkorrelációs függvény legnagyobb értéke 0,118. Ebből arra következtethetünk, hogy az idősorok nem viselkedtek hasonlóan, nincs szinkronizáció az idősorok között.

3.1.2.8. Légzőszervi daganatok

A légzőszervi daganatok (légcső-, hörgő-, tüdőrák) okozta halálozási arányokat a 0-64 éves korcsoportban a férfiak (1. idősor) és a nők (2. idősor) adataira hasonlítottam össze.

Az adatok vizsgálatát a 3.1.2.1. fejezetben leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatóaknak bizonyultak. A férfiak adataiból a sztochasztikus egyenlet az alábbiak adódott: $z_t = 0,913z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 9,729$; a nők adataira az egyenlet: $z_t = 0,902z_{t-1} + \varepsilon_t$; $\sigma_a^2 = 0,616$.

A férfiak halálozási arányaiból meghatározott AR(1) együtthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 31. táblázat.

31. táblázat: AR(1) együttható konfidencia intervallumai p különböző értékeinél a férfiak csoportjában.

$\phi \approx 0,913$	$p=0,1$	$p=0,05$	$p=0,01$
----------------------	---------	----------	----------

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

Normális elosz.	(0,7819;1,0441)	(0,7529;1,0731)	(0,7003;1,1257)
White féle	(0,7843;1,0417)	(0,7578;1,0682)	(0,7041;1,1219)
Folytonos becs.	(0,8494;0,9933)	(0,8254;0,9981)	(0,7792;0,9998)

A konfidencia intervallumok összehasonlítását mutatja a 32. táblázat.

32. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,913$	p=0,1		p=0,05		p=0,01	
Normális elosz.	0,2622	100%	0,3202	100%	0,4254	100%
White féle	0,2574	98,2%	0,3104	96,9%	0,4178	98,2%
Folytonos becs.	0,1439	54,9%	0,1727	53,9%	0,22,6	51,8%

A nők halálozási arányaiból meghatározott AR(1) együttthatóra a különböző szignifikancia szintekhez ($p=0,1$; $p=0,05$; $p=0,01$) tartozó konfidencia intervallumokat tartalmazza a 33. táblázat.

33. táblázat: AR(1) együtttható konfidencia intervallumai p különböző értékeinél a nők csoportjában.

$\phi \approx 0,902$	p=0,1		p=0,05		p=0,01	
Normális elosz.	(0,7634;1,040)		(0,7327;1,0713)		(0,6771;1,1269)	
White féle	(0,7657;1,0383)		(0,7376;1,0664)		(0,6806;1,1234)	
Folytonos becs.	(0,8341;0,991)		(0,8088;0,9973)		(0,7616;0,9998)	

A konfidencia intervallumok összehasonlítását tartalmazza a 34. táblázat.

34. táblázat: Konfidencia intervallumok összehasonlítása.

$\phi \approx 0,902$	p=0,1		p=0,05		p=0,01	
Normális elosz.	0,2772	100%	0,3386	100%	0,4498	100%
White féle	0,2726	98,3%	0,3288	97,1%	0,4428	98,4%

Folytonos becs.	0,1569	56,6%	0,1885	55,7%	0,2382	53,8%
-----------------	--------	-------	--------	-------	--------	-------

A helyes modell választás igazolásához a (2.58) kifejezéssel megadott χ^2 próbát alkalmaztam. A reziduumok autokorrelációiból a következő értékek adódtak: $\chi^2_{\text{férfiak}}=6,147$; $\chi^2_{\text{nők}}=8,045$. Mivel $\chi^2_{0,05}=11,07$, így a választott modellek helyesnek bizonyultak [25], [50], [51].

A reziduumok autokorrelációi nem tartalmaztak korreláltságot, a 7. ábrához hasonlóan kaptam rájuk.

A férfiak és a nők halálozási arányait jellemző idősorok közötti kapcsolat elemzéséhez keresztkorrelációs függvényeket alkalmaztam. A keresztkorrelációs függvény a modell illesztése előtt hasonlóan viselkedett, mint a méhnyakrák okozta halálozási arányoknál. A reziduumokból készült keresztkorrelációs függvénynek szignifikáns értéke nem volt 95%-os szignifikancia szinten. A keresztkorrelációs függvény legnagyobb értéke 0,376. Ebből arra következtethetünk, hogy az idősorok nem viselkedtek hasonlóan, azaz nincs szinkronizáció az idősorok között.

3.1.3. Következtetések a halálozási arányok elemzéseiből

A halálozási arányok elemzéseinek eredményei alapján a vizsgált betegségek két csoportba sorolhatóak.

Az egyik csoportba tartozott a *cerebrovasculáris betegség, az isémiás szívbetegség, a krónikus májbetegség és a légzőszervi betegség*. Ezekre az a jellemző, hogy a modellek illesztése után kapott reziduumok hasonlóan viselkedtek. A reziduumok keresztkorrelációs függvényének egy szignifikáns értéke volt a $k=0$ időképletelésnél. Ez azt jelenti, hogy nulla időeltolódással, azaz azonos években a függvények azonos módon

viselkedtek, úgyis mondhatjuk, hogy „szinkronizáltak” voltak. A „szinkronizáció” alatt azt értjük, hogy a betegségként képzett két idősor azonos módon változott, növekedett vagy csökkent. A két idősor hasonló viselkedéséből arra lehetett következtetni, hogy a környezeti tényezők a betegség kialakulása szempontjából azonos módon hatottak a betegségként képzett mindkét idősor elemeire. Ez arra utal, hogy az adott betegség kialakulásában a környezet hatása lényeges szerepet játszott [50], [51], [55]. A környezeti tényezők alatt a külső hatások összességét értjük, az egyes betegségek kialakulásában szerepet játszó specifikus hatások vizsgálata nem képezte az értekezés tárgyát. Ezek az eredmények összhangban vannak az orvostudományi kutatások eredményeivel. A cerebrovasculáris betegség és az isémiás szívbetegség kialakulásában a zsírdús táplálkozásnak, a mozgásszegény életmódnak, a stressznek, a krónikus májbetegség kialakulásában a nagymértékű alkoholfogyasztásnak lényeges szerepet tulajdonítanak.

A halálozási arányok másik csoportját a daganatos megbetegedések képezték (*méhnyakrák, mellrák, emésztőszervi daganatok, légzőszervi daganatok*). Ezekre az a jellemző, hogy a modellek illesztése után kapott reziduumok nem mutattak hasonlóságot. A reziduumok keresztkorrelációs függvényének nem volt szignifikáns értéke. Ebből az idősorok azonos módon történő viselkedésére nem lehetett következtetni, azaz nem volt „szinkronizált” a két csoportot jellemző idősor. Ebből arra lehet következtetni, hogy a környezeti tényezők hatásai a daganatos betegségek kialakulása szempontjából eltérőek. Ez arra utal, hogy ezeknek a betegségeknek a kialakulásában a környezet hatása ezzel a módszerrel nem igazolható, bár egyértelműen ki sem zárható [50], [51], [55]. Ez összhangban van az orvostudományi

kutatások eredményeivel is, ugyanis a daganatos betegségek oktana csak részben tisztázott. Egyes daganatos betegségek kialakulásában az örökletes tényezőknek (hibás gének átörökítése) lényeges szerepet tulajdonítanak, ugyanakkor a környezeti tényezők szerepét elhanyagolhatónak vélik (pl. mellrák kialakulása). Ezzel szemben más daganatos betegségek esetében (pl. tüdőrák) a környezeti tényezőknek lényeges szerepet tulajdonítanak. Az elemzés eredménye és az orvostudományi eredmények közötti esetleges ellentmondás látszólagos, ugyanis e betegségek oktana csak részben tisztázott. Így az sem bizonyított, hogy a környezeti tényezőknek milyen szerepe lehet az egyes daganatos betegségek kialakulásában. Bár az általam alkalmazott módszerrel a környezet közvetlen hatása nem volt igazolható, de ennek szerepét egyértelműen kizárni nem lehetett.

Összesen nyolc betegség okozta halálozási arányt vizsgáltam, melyek az összes halálozások kb. 80%-t adják. Az elemzésekből arra következtettem, hogy négy betegség (*cerebrovasculáris betegség, isémiás szívbetegség, krónikus májbetegség, légzőszervi betegségek*) kialakulásában a környezeti tényezőknek lényeges szerepük van. A másik négy betegség esetében (*méhnyakrák, mellrák, emésztőszervi daganatok, légzőszervi daganatok*) ilyen következtetés nem vonható le. Mindkét csoportba 4-4 betegség sorolható, így a levonható következtetések még jobban megalapozottak.

A magyarországi adatokból levont fenti következtetések megegyeznek más országokban korábban hasonló módszerekkel végzett elemzésekkel [55]. Az idézett közleményben a szerző az USA-ból származó 1948-1978 közötti halálozási arányokat elemzett. A cerebrovasculáris betegség, az isémiás szívbetegség és tumoros

betegségek (légszervi daganatok, nemiszervek daganatai, mellrák) adatait vizsgálta. A szerző egy módszert ismertetett, amely segített annak eldöntésében, vajon az időben változó környezeti faktor befolyásolja-e az adott betegség kialakulását. A népességet két természetes csoportban vizsgálta (férfiak és nők, vagy két korcsoportban). Az ARIMA modellt meghatározta. Ha a reziduumok keresztkorrelációs függvénye a két csoport között szignifikáns csúcsot adott a nulla időkésleltetésnél, ebből arra következtetett, hogy környezeti faktor hatása érvényesül. A szerző megállapította, hogy a cerebrovasculáris betegség és az isémiás szívbetegség kialakulásában a környezeti tényezők jelentős szerepet játszanak, míg a tumoros megbetegedéseknél ilyen hatást nem lehetett igazolni.

3.1.3.1. A konfidencia intervallum számítások értékelése

Az $AR(I)$ folyamat ϕ paraméterének konfidencia intervallum becsléseiből látható, hogy a normális eloszlás felhasználásával és a White féle becsléssel körülbelül azonos nagyságú konfidencia intervallumokat kaptam, bár a White féle becslés kismértékben jobbnak bizonyult, mint a normális eloszlás alkalmazása. 2-4%-kal csökkentek a konfidencia intervallumok a White féle becsléssel a normális eloszlás felhasználásával nyert intervallumokhoz képest. 30-40%-kal csökkentek a konfidencia intervallumok a folytonos idejű becsléssel a normális eloszlás alkalmazásával kapott intervallumokhoz viszonyítva (4. 6. 8. 10. 12. 14. 16. 18. 20. 22. 24. 26. 28. 30. 32. 34. táblázat).

Ha az $AR(I)$ paraméter $\phi \approx 1$, akkor a normális eloszlás illetve a White féle becslés nem használható, mert a konfidencia intervallum felső határa 1-nél nagyobb is lehet (3. 5. 7. 9. 11. 13. 15. 17. 19. 21. 23. 25. 27.

29. 31. 33. táblázat). Ekkor stacionárius esetben használhatatlan az eredmény. Az első két eljárással a konfidencia intervallumok felső hataraként 1-nél nagyobb értékek is adódtak, melyek használhatatlanok [6]. A folytonos idejű modellből adódó becsléssel kapott konfidencia intervallumok felső határai minden esetben 1 alatt maradtak (3. 5. 7. 9. 11. 13. 15. 17. 19. 21. 23. 25. 27. 29. 31. 33. táblázat). A példák jól mutatják a folytonos idejű becslés előnyeit a többi módszerhez képest.

A teljesség miatt érdemes megemlíteni, bár a bemutatott példákban nem adódtak $\phi \approx 0$ értékek, hogy a normális eloszlás illetve a White féle becsléssel is kaphatunk negatív eredményeket a konfidencia intervallum alsó hataraként, melyek szintén nem elfogadhatóak [6].

Az elemzések arra is adatokat szolgáltatottak, hogy az $AR(1)$ együtthatók konfidencia intervallumokkal történő becslésére alkalmazott módszerek közül a hazai kutatók által kidolgozott folytonos becslés előnyösebben alkalmazható az irodalomban korábban leírtakhoz képest [4], [6], [7], [8], [11].

3.2. Gyermekkori lymphoid leukaemiás betegek adatainak feldolgozása

Szezonális idősor analízissel elemeztem az 1988. január 1. és 2000. december 31. között diagnosztizált 0-18 éves akut lymphoid leukaemiás (ALL) gyermekek adatait. A gyermekkori lymphoid leukaemiás betegek számának elemzéséhez az SPSS9.0 programcsomagot alkalmaztam [110].

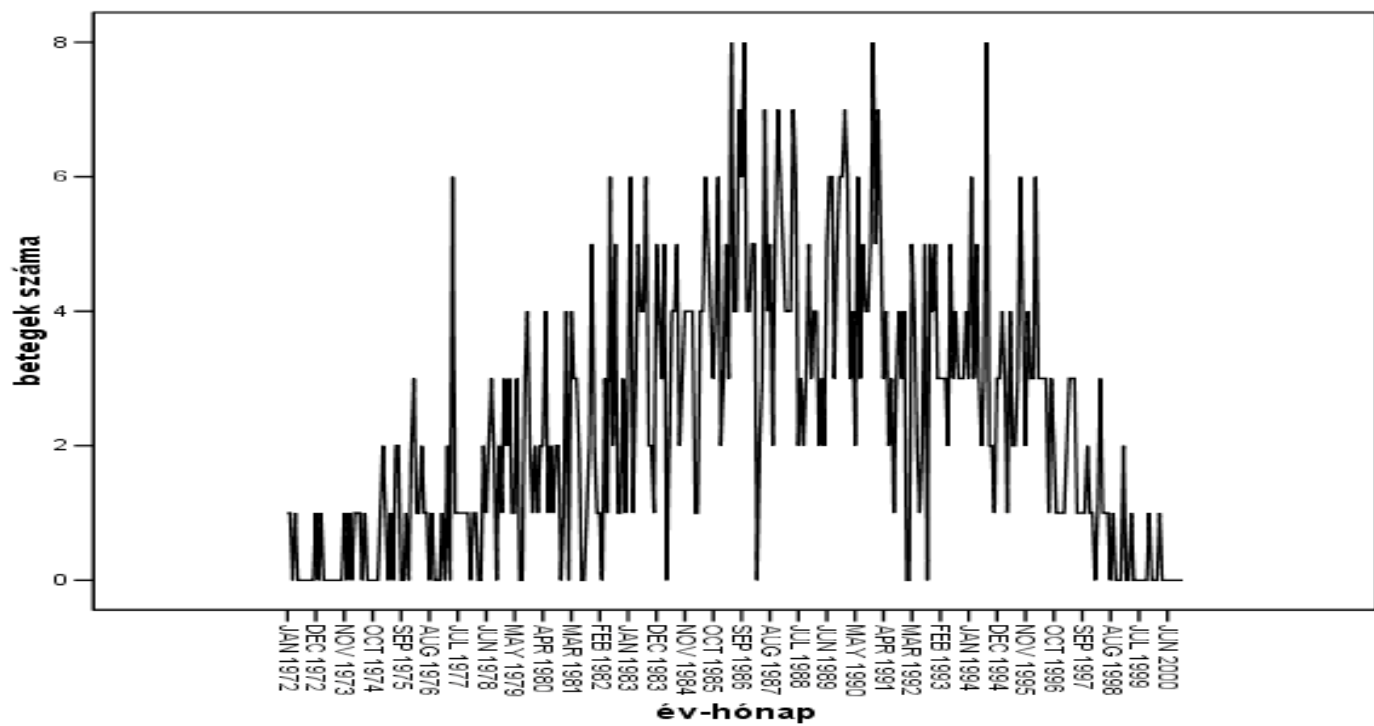
3.2.1. A betegek születési dátumának elemzése

A gyermekkori lymphoid leukaemiában megbetegedett gyermekek születési dátumát elemeztem. A betegek számát születési dátumuk szerint havonként csoportosítottam, melyek alapján elkészítettem a 16. ábrát. Az

ábra elemzése során lényeges halmozódást nem figyeltem meg, így ebből semmilyen következtetés nem vonható le. A szezonális komponens alakulását a 17. ábra szemlélteti.

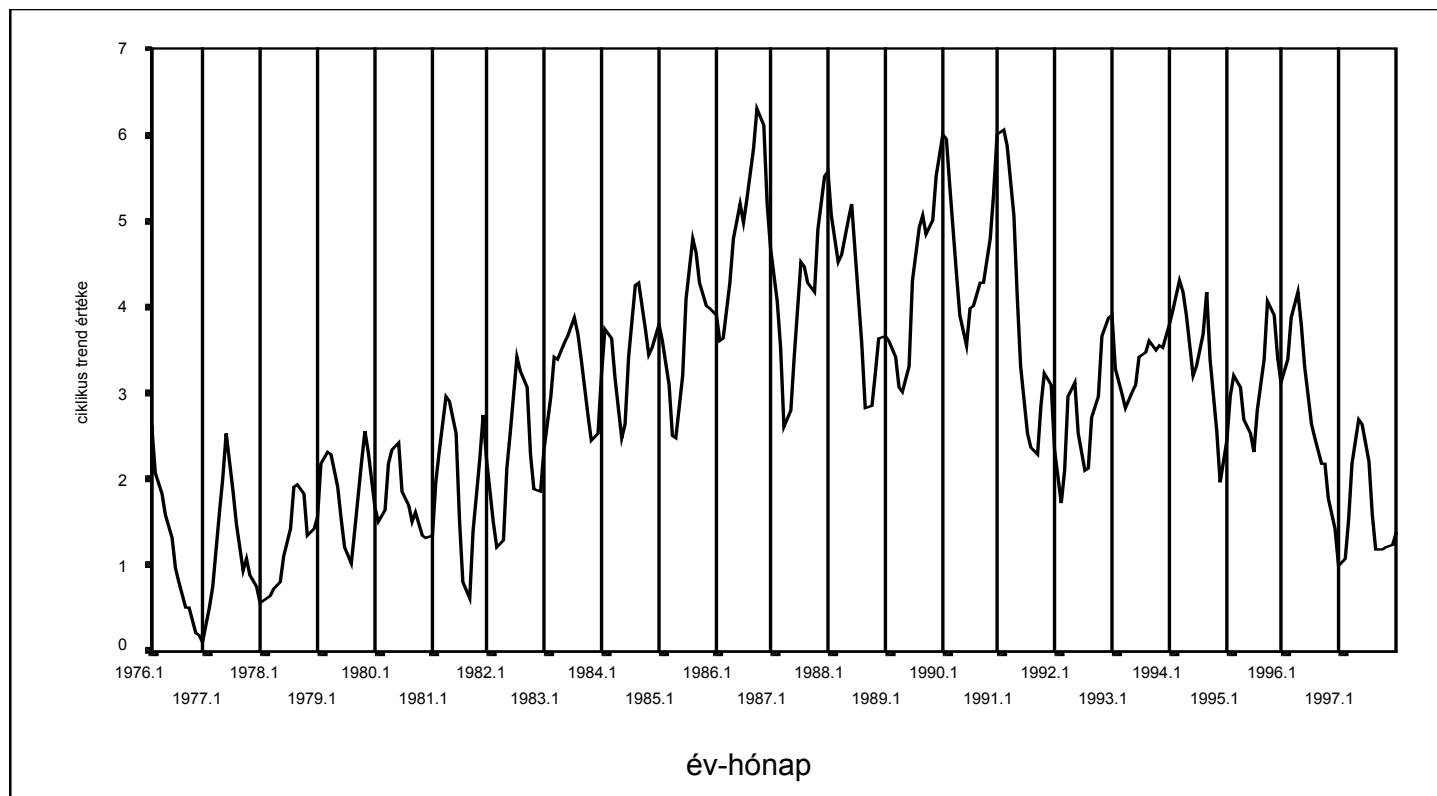
A szezonális komponens vizsgálata során megállapítottam, hogy egyes években egy, máskor kettő, esetenként három csúcs is adódott (17. ábra). Vizsgáltam, hogy a csúcsok az év mely szakára (az év eleje, közepe, vége) estek. Mindhárom időszakra közel azonos számú kimagasló érték adódott. Ezek alapján megállapítottam, hogy jellegzetes halmozódás nem figyelhető meg. A szezonális komponens ábrázolása során a csúcsok szabálytalanul követték egymást, a csúcsok közötti időtartam változó volt. A szezonális komponens csúcsainak egy-egy harmada az év elejére, közepére és végére esett, időbeli eloszlásuk szabálytalan volt. A csúcsok szabálytalan előfordulása a véletlenszerű hatást valószínűsíti. Ebből arra lehet következtetni, hogy a fogamzás vagy a terhesség időszakában az anyát és a magzatot ért környezeti hatás (pl: vírusjárvány, légszennyezettség, stb) nem játszik lényeges szerepet a gyermekkori lymphoid leukaemia kialakulásában.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK



16. ábra: A gyermekkori lymphoid leukaemiás betegek száma a születési dátumaik alapján csoportosítva.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

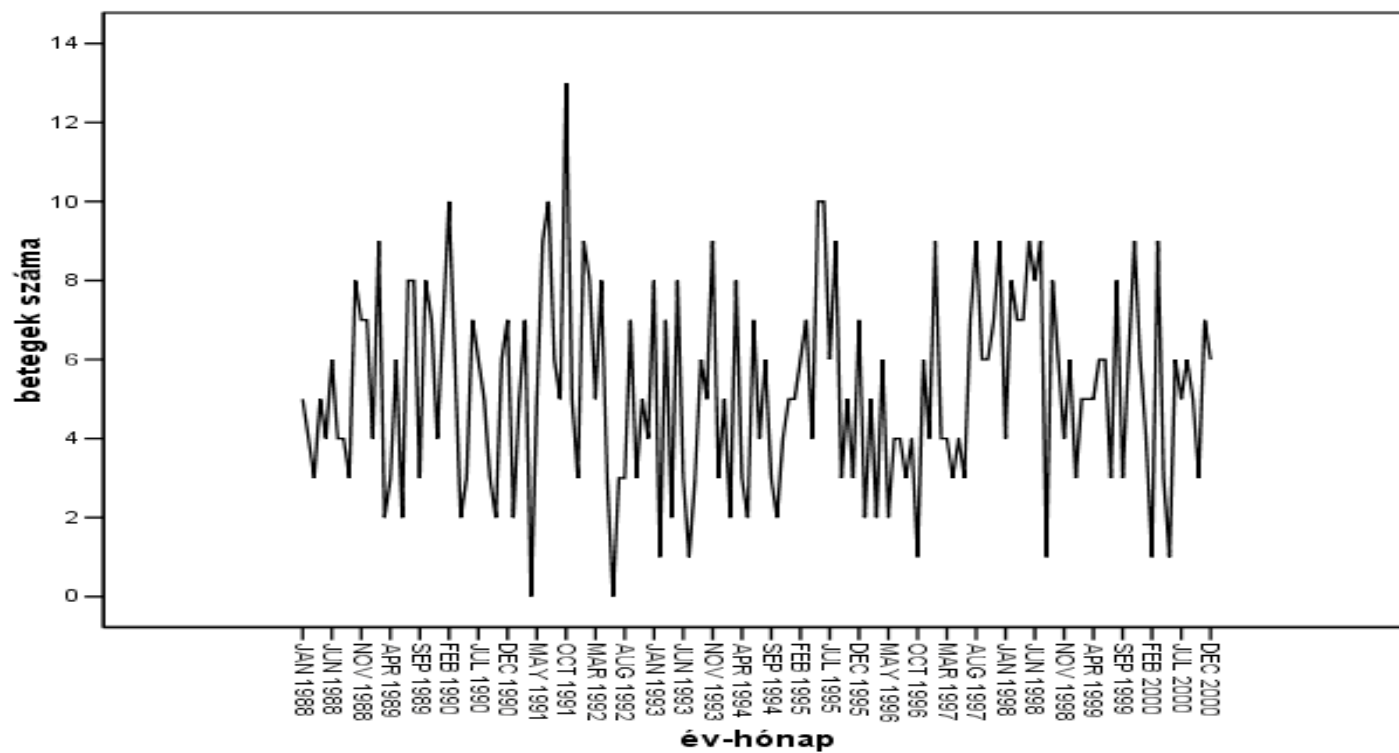


17. ábra: A gyermekkori lymphoid leukaemiás betegek születési dátuma alapján készült szezonális komponens.

3.2.2. *A betegség diagnosztizálási dátumának elemzése*

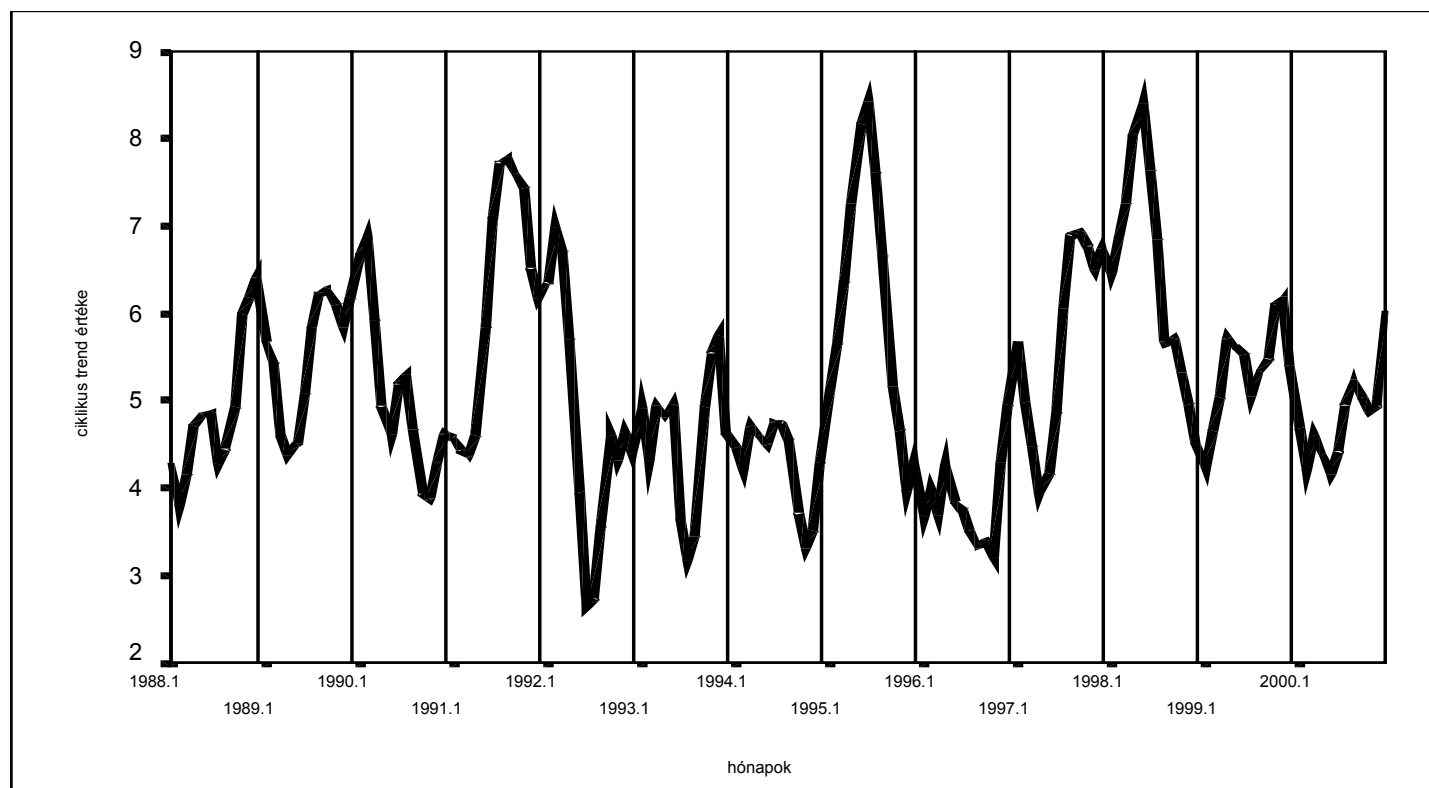
A betegség diagnosztizálási dátumának elemzését a teljes korcsoportra és a medián alapján kettébontott adatsorokra is elvégeztem. A 18. ábra mutatja az összes beteg számának havi változásait. Az összes beteg számából készült szezonális komponenseket a 19. ábra szemlélteti. A szezonális komponensek ábrázolása során kilenc csúcsot kaptam, melyek közül hat a téli időszakra, egy az őszi, egy a tavaszi és egy a nyári hónapokra esett. A szezonális komponens elemzésekor azokat az emelkedéseket tekintettem csúcsnak, amikor a szezonális komponens értéke a hatot elérte (19. ábra). A fiatalabb és az idősebb betegcsoportok adatai alapján készült szezonális komponensek alakulását a 20. ábra és a 21. ábra mutatja. A fiatalabb korcsoport szezonális komponenseinek elemzésekor (20. ábra) hét téli, két őszi, egy tavaszi és egy nyári csúcs adódott. Az idősebbek szezonális komponenseinek vizsgálatakor (21. ábra) a téli hónapokra hét, az ősziekre egy, a tavasziakra egy és a nyári hónapokra négy csúcsot kaptam. A fiatalabb és az idősebb korcsoport vizsgálata során csúcsnak tekintettem azokat az emelkedéseket, amikor a szezonális komponens értéke a hármat elérte. Mindhárom esetben a szezonális komponensekben a csúcsok többsége a téli hónapokra esett. Ez arra utal, hogy a téli időszakban valamilyen külső hatás (gyaníthatóan vírusfertőzés) elősegíthette a betegség manifesztálódását.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK



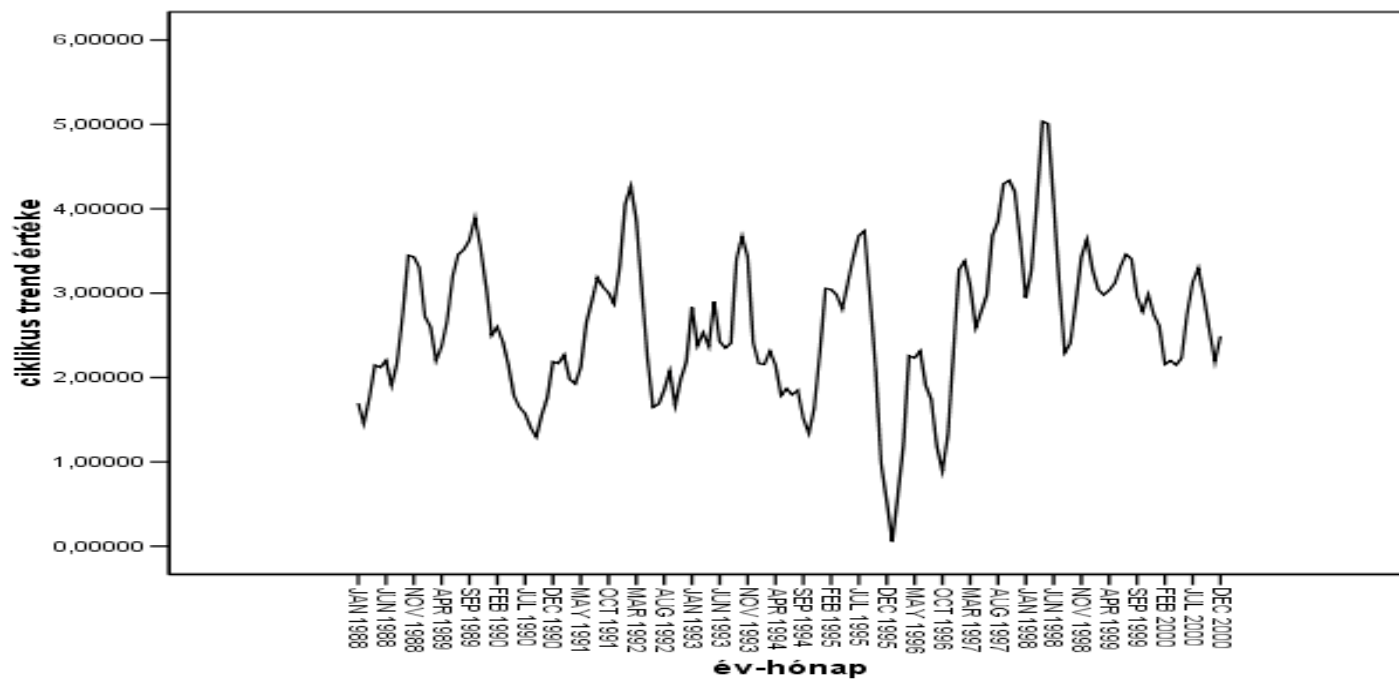
18. ábra: A gyermekkori lymphoid leukaemiás betegek száma a betegség diagnosztizálási dátuma alapján.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK



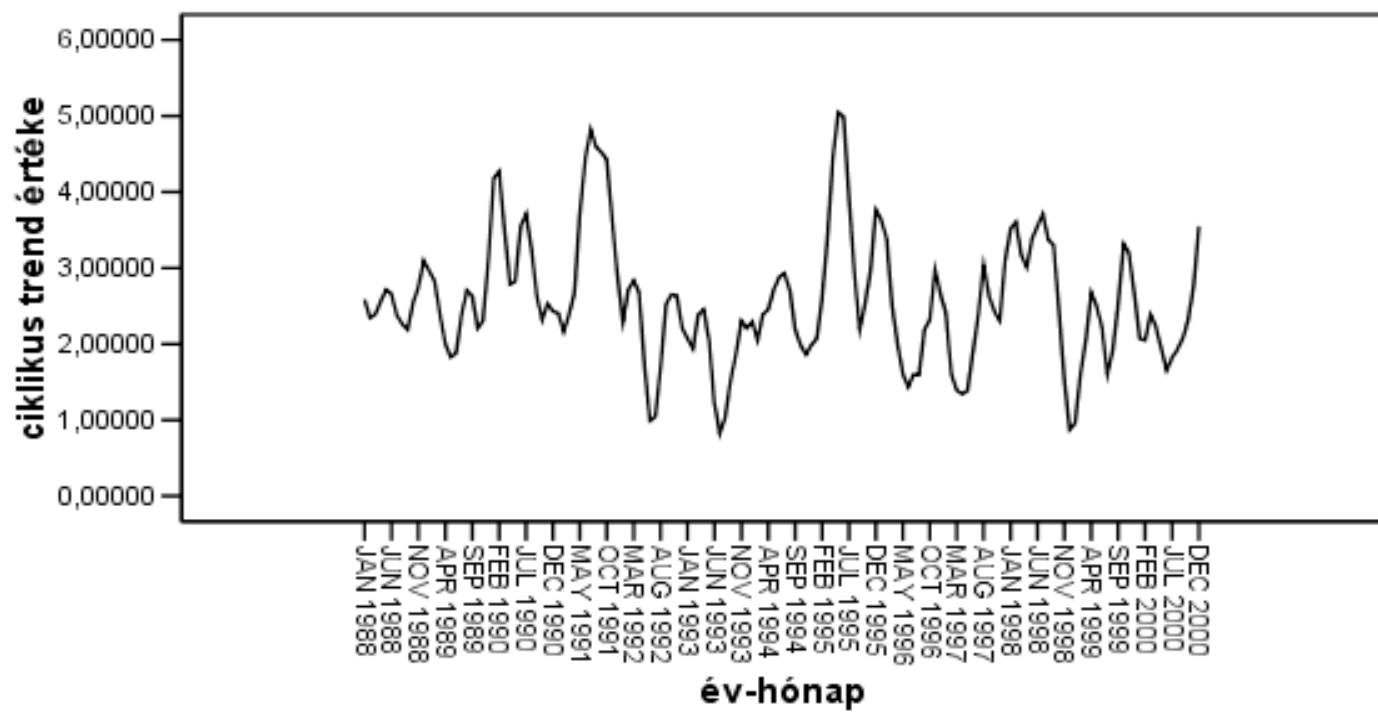
19. ábra: A gyermekkori lymphoid leukaemiás betegség diagnosztizálási dátuma alapján készült szezonális komponens az összes betegszámra.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK



20. ábra: A gyermekkori lymphoid leukaemiás betegség diagnosztizálási dátuma alapján készült szezonális komponens a fiatalabb korcsoportban.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK



21. ábra: A gyermekkori lymphoid leukaemiás betegség diagnosztizálási dátuma alapján készült szezonális komponens az idősebb korcsoportban.

3.2.3. Következtetések a lymphoid leukaemiás betegek adatainak vizsgálatából

Az összes betegszám elemzéséből lényeges következtetések nem vonhatók le, mert a betegszám időbeli ingadozásai nem értékelhetők. A szezonális komponens ábrázolása adekvált módszernek bizonyolult a betegszám időbeli ingadozásainak elemzésére és következtetések levonására.

Az elemzések során először a betegek születési dátumát vizsgáltam. A vizsgálat során jellegzetes halmozódásokat nem figyeltem meg. Ez arra utal, hogy a terhesség alatt az anyát és a magzatot ért külső környezeti hatások nem játszottak lényeges szerepet a betegség kialakulásában.

A betegség diagnosztizálásának dátuma alapján végzett elemzés során a szezonális komponensben a téli hónapokban halmozódást figyeltem meg. Ez arra enged következtetni, hogy a betegség manifesztálódásában a külső környezetnek lényeges hatása van. Gyaníthatóan a téli hónapokban előforduló vírusjárványoknak lényeges szerepük van a betegség manifesztálódásában.

4. ÚJ KUTATÁSI EREDMÉNYEK

1. Vizsgálataimmal igazoltam, hogy az idősori modellek elemzése alkalmas módszer humán megbetegedések (*cerebrovasculáris betegség, isémiás szívbetegség, krónikus májbetegség, légzőszervi betegség, méhnyakrák, mellrák, emésztőszervi daganatok, légzőszervi daganatok*) oktanának vizsgálatára. Az is megállapítható, hogy a komplex oktanú megbetegedéseknél ez a statisztikai módszer hatékonyan alkalmazható.
2. Az elemzések során megállapítottam, hogy a betegségek egy részében (*cerebrovasculáris betegség, isémiás szívbetegség, krónikus májbetegség, légzőszervi betegség*) a reziduumok keresztkorrelációs függvényének $k=0$ időkésleltetésnél szignifikáns értéke volt, ami azt jelenti, hogy az idősorok „szinkronizáltak” voltak. Ebből arra következtettem, hogy ezeknek a betegségeknek a kialakulásában a környezeti tényezőknek döntő szerepe van.
3. Az ismeretlen oktanú daganatos megbetegedések (*méhnyakrák, mellrák, emésztőszervi daganatok, légzőszervi daganatok*) elemzésekor a reziduumok keresztkorrelációs függvényének nem volt szignifikáns értéke, azaz az idősorok ezeknél a betegségeknek nem voltak „szinkronizáltak”. Ebből arra következtettem, hogy a környezet közvetlen szerepe a betegségek kialakulásában, manifesztálódásában, nem igazolható, de a közvetett hatások nem zárhatók ki.

4. Az értekezésben olyan orvostudományi adatokat dolgoztam fel, melyek nagyarányú halálozásokat okozó megbetegedések oktanához következtetéseket szolgáltatnak.
5. A *gyermekkori lymphoid leukaemiás* adatok elemzése során arra a következtetésre jutottam, hogy a diagnosztizálás dátumának nagyobb a szerepe, mint a születés dátumának. A betegség diagnosztizálási dátumának elemzésekor úgy tapasztaltam, hogy a megbetegedések a téli időszakban halmozódtak. Ezek a gyermekkori környezeti hatások lényeges szerepét mutatják a betegség manifesztálódásában (pl: vírusjárvány előfordulása).
A gyermekkori lymphoid leukaemiás betegek adatainak időszaki elemzéssel történő vizsgálatára első alkalommal került sor hazánkban.
6. Az $AR(1)$ együttható konfidencia intervallumait háromféle módon határoztam meg: normális eloszlás felhasználásával, White féle becsléssel és a folytonos idejű modell alkalmazásával. A konfidencia intervallumok folytonos idejű becsléssel történő megadásai szükségesek, mivel a határeloszlások, ha az $AR(1)$ együttható egyhez közeli, nem minden esetben alkalmazhatók.

5. ÖSSZEFOGLALÁS

A dolgozat két részre tagolódik, az első részben az idősorok jellemzőit ismertettem, a második részben orvostudományi adatok számítógépes feldolgozásait írtam le, melyekhez idősori elemzéseket alkalmaztam.

5.1. Idősorok

Áttekintettem a sztochasztikus folyamatok jellemzőit, a kétdimenziós stacionárius folyamatokat, az $ARIMA(p,d,q)$ modelleket, az idősorok autokorrelációs- és parciális autokorrelációs együtthatóinak sajátosságait. Becsültem az autokorrelációs- és parciális autokorrelációs együtthatók megbízhatóságát és standard hibáit. Az $AR(1)$ folyamat paraméterére háromféle módon (normális eloszlás, mint becslés alkalmazásával, White féle becsléssel, folytonos idejű modellel való becsléssel) konstruált konfidencia intervallumok összehasonlítását bemutattam. Megállapítottam, hogy a három becslés közül a folytonos idejű modell alkalmazása adta a legmegbízhatóbb eredményt.

5.2. Számítógépes alkalmazások

5.2.1. Halálozási arányok elemzése

A számítógépes feldolgozások során nyolc betegség (*cerebrovasculáris betegség, isémiás szívbetegség, krónikus májbetegség, légzőszervi*

betegségek, valamint a méhnyakrák, mellrák, emésztőszervi daganatok, légzőszervi daganatok) okozta halálozási arányok elemzését végeztem el idősor analízis módszerével. A halálozási arányok elemzéséhez az adatokat a WHO adatbázisából szereztem be. Az elemzések során betegségenként (életkor vagy nem alapján) két-két idősort vizsgáltam. A számítások során az idősorok elemeiből az átlagokat, a trendeket és a szezonális komponenseket levontam. Továbbiakban az idősorok véletlen komponenseinek azonosítását adtam meg. A vizsgálatok során a fontosabb lépések a következők voltak:

- az autokorrelációs- és parciális autokorrelációs függvények vizsgálata,
- a sztochasztikus egyenletek megadása,
- a modell választás helyességének igazolása,
- a reziduumok ellenőrzése,
- a keresztkorrelációs függvények vizsgálata,
- a konfidencia intervallumok becslései.

Idősor analízissel azt vizsgáltam, hogy a betegségenként képzett két-két idősor vajon hasonlóképpen viselkedik-e, azaz az idősorok „szinkronizáltak”-e.

5.2.1.1. Eredmények

A halálozási arányok elemzéseinek eredményei alapján a vizsgált betegségek két csoportba sorolhatóak.

Az egyik csoportba a *cerebrovasculáris betegség, az isémiás szívbetegség, a krónikus májbetegség és a légzőszervi betegség* tartozott. Ezekre az a jellemző, hogy a reziduumok keresztkorrelációs függvényének egy szignifikáns értéke volt a $k=0$ időkésltetésnél. Az idősorok a nulla időlésltetésnél hasonlóan viselkedtek, ezt úgy is

mondhatjuk, hogy az idősorok „szinkronizáltak” voltak. A két idősor hasonló viselkedéséből arra lehetett következtetni, hogy a környezeti tényezők a betegség kialakulása szempontjából azonos módon hatottak a betegségenként képzett mindkét idősor elemeire. Ez arra utal, hogy az adott betegség kialakulásában a környezet hatása lényeges szerepet játszott [50], [51], [55]. A környezeti tényezők alatt a külső hatások összességét értjük, az egyes betegségek kialakulásában szerepet játszó specifikus hatások vizsgálata nem képezte az értekezés tárgyát. Ezek az eredmények összhangban vannak az orvostudományi kutatások eredményeivel. A cerebrovasculáris betegség és az isémiás szívbetegség kialakulásában a zsírdús táplálkozásnak, a mozgásszegény életmódnak, a stressznek, a krónikus májbetegség kialakulásában a nagymértékű alkoholfogyasztásnak lényeges szerepet tulajdonítanak.

A halálozási arányok másik csoportját a daganatos megbetegedések képezték (*méhnyakrák, mellrák, emésztőszervi daganatok, légzőszervi daganatok*). Ezekre az a jellemző, hogy a reziduumok keresztkorrelációs függvényének nem volt szignifikáns értéke, azaz az idősorok nem voltak „szinkronizáltak”. Ebből arra lehetett következtetni, hogy a környezeti tényezők hatásai a daganatos betegségek kialakulása szempontjából eltérőek. Ez arra utal, hogy ezeknek a betegségeknek a kialakulásában a környezet hatása ezzel a módszerrel nem igazolható, bár egyértelműen ki sem zárható [50], [51], [55]. Ez összhangban van az orvostudományi kutatások eredményeivel is, ugyanis a daganatos betegségek oktana csak részben tisztázott. Egyes daganatos betegségek kialakulásában az örökletes tényezőknek (hibás gének átörökítése) lényeges szerepet tulajdonítanak, ugyanakkor a környezeti tényezők szerepét elhanyagolhatónak vélik (pl. mellrák kialakulása).

Ezzel szemben más daganatos betegségek esetében (pl. tüdőrák) a környezeti tényezőknek lényeges szerepet tulajdonítanak. Az elemzés eredménye és az orvostudományi eredmények között esetleges ellentmondás látszólagos, ugyanis e betegségek oktana csak részben tisztázott. Így az sem bizonyított, hogy a környezeti tényezőknek milyen szerepe lehet az egyes daganatos betegségek kialakulásában. Bár az általam alkalmazott módszerrel a környezet közvetlen hatása nem volt igazolható, de ennek szerepét egyértelműen kizárni sem lehetett.

A magyarországi adatokból levont fenti következtetések megegyeznek más országokban korábban hasonló módszerekkel végzett elemzésekkel [55]. Az idézett közleményben a szerző az USA-ból származó 1948-1978 közötti halálozási arányokat elemzett. A cerebrovasculáris betegség, az isémiás szívbetegség és tumoros betegségek (légszöszervi daganatok, nemiszervek daganatai, mellrák) adatait vizsgálta időszori elemzéssel. A szerző az alkalmazott statisztikai módszerrel azt vizsgálta, hogy az időben változó környezeti faktor befolyásolja-e az adott betegség kialakulását. A népességet két természetes csoportban vizsgálta (férfiak és nők, vagy két korcsoportban). Az ARIMA modellt meghatározta. Ha a reziduumok keresztkorrelációs függvénye a két csoport között szignifikáns csúcsot adott a nulla időkésleltetésnél, ebből arra következtetett, hogy környezeti faktor hatása érvényesül. A szerző megállapította, hogy a cerebrovasculáris betegség és az isémiás szívbetegség kialakulásában a környezeti tényezők jelentős szerepet játszanak, míg a tumoros megbetegedéseknél ilyen hatást nem lehetett igazolni.

Az elemzések arra is adatokat szolgáltatottak, hogy az AR(1) együtthatók konfidencia intervallumokkal történő becslésére alkalmazott

módszerek közül a hazai kutatók által kidolgozott folytonos becslés előnyösebben alkalmazható az irodalomban korábban leírtakhoz képest [4], [6], [7], [8].

5.2.2. *Lymphoid leukaemiás betegek adatainak feldolgozása*

A számítógépes alkalmazások során a leukaemiás megbetegedések adatait is vizsgáltam, melyhez az adatokat a Magyarországi Gyermekonkológiai Munkacsoport adatbázisából szereztem be.

A vizsgálatok során a betegek születési dátumának és a betegség diagnosztizálási dátumának elemzését végeztem el. A betegek születési dátuma alapján a szezonális komponenseket ábrázoltam. A szezonális komponensek vizsgálata során megállapítottam, hogy egyes években egy, máskor kettő, esetenként három csúcs is adódott. Vizsgáltam, hogy a csúcsok az év mely szakára (az év eleje, közepe, vége) estek. Mindhárom időszakra közel azonos számú kimagasló érték adódott. Ezek alapján megállapítottam, hogy jellegzetes halmozódás nem figyelhető meg. A szezonális komponensek csúcsai szabálytalanul követték egymást, a csúcsok közötti időtartam változó volt. A szezonális komponensek csúcsainak egy-egy harmada az év elejére, közepére és végére esett, időbeli eloszlásuk szabálytalan volt. A csúcsok szabálytalan előfordulása a véletlenszerű hatást valószínűsíti. Ebből arra lehet következtetni, hogy a fogamzás vagy a terhesség időszakában az anyát ért környezeti hatás (pl: vírusjárvány, légszennyezettség, stb) nem játszik lényeges szerepet a gyermekkori lymphoid leukaemia kialakulásában.

A betegség diagnosztizálási dátumának elemzését a teljes korcsoportra és a medián alapján kettébontott adatsorokra is elvégeztem.

5. ÖSSZEFOGLALÁS

A teljes korcsoportra vonatkozóan kilenc csúcsot kaptam, melyek közül hat a téli időszakra, egy az őszi, egy a tavaszi és egy a nyári hónapokra esett. A fiatalabb korcsoport szezonális komponenseiben hét téli, két őszi, egy tavaszi és egy nyári csúcs adódott. Az idősebbek szezonális komponenseiben a téli hónapokra hét, az ősziekre egy, a tavasziakra egy és a nyári hónapokra négy csúcsot kaptam. A három idősből kapott szezonális komponensek csúcsai bár nem teljesen szabályos időközönként követték egymást, de a csúcsok többsége mindhárom görbén a téli hónapokban jelentkezett. Ez arra utal, hogy a téli időszakban valamilyen külső hatás (gyaníthatóan vírusfertőzés) elősegíthette a betegség manifesztálódását.

6. SUMMARY

The dissertation basically deals with two related subject matters. In the first part of the dissertation, I discuss the characteristic features and rules applying to time series. In the second part of my work, I use these methods for the examination of human diseases the causes of which are complex and unclear. The objective of the time series analyses was to examine the roles of environmental effects in the development of these diseases.

The dissertation examines mortality rates caused by non-infectious and tumorous diseases in the Hungarian population with the use of time series analysis methods. Relying on WHO data, the time series models and their reliability are analysed in detail with respect to mortality rates caused by cerebrovascular disease, ischaemic heart disease, chronic hepatopathy, respiratory diseases, as well as cervical cancer, breast cancer, tumours of the digestive organs and the respiratory organs. With the help of autoregressive integrated moving average (ARIMA) models and autocorrelation and partial autocorrelation functions, as well as cross-correlation functions I examined whether external factors (environmental effects) had a significant role in the causes of death. I examined the reliability of the autoregressive (AR) coefficient with the help of confidence intervals gained using the methods of normal distribution, White's estimation and continuous time distribution, known from the literature. The objective of the analyses was to apply methods of mathematical statistics for the determination of the causes of diseases in

order to draw conclusions as to what roles environmental effects have in the development of diseases.

Using seasonal time series analysis, I examined the birth and diagnosis dates of children aged between 0 and 18 years and diagnosed with acute lymphoid leukaemia (ALL) between January 1, 1988 and December 31, 2000. The aim of the analysis was to determine whether characteristic aggregation (seasonality) may be observed on the basis of the birth or diagnosis dates of ALL patients, since the role of environmental effects can be concluded in case of seasonality.

Computer applications

Using time series analysis computer applications the mortality rates caused by eight diseases (cerebrovascular disease, ischaemic heart disease, chronic hepatopathy, respiratory diseases, as well as cervical cancer, breast cancer, tumours of the digestive organs and the respiratory organs) were analysed. The data for the analysis was retrieved from the database of WHO. The database included the mortality rates for the period between 1970 and 1997. I examined two time series for each disease (on the basis of age or sex).

By the help of the time series analysis I examined whether the cross-correlation function between the residual had a significant value, whether the two time series resulting for each of the diseases behaved similarly, in other words, whether the functions were “synchronised.”

In the course of the computer analysis it was also examined the data on leukaemia from the database of the Hungarian Paediatric Oncology Workgroup. For the statistical analysis I used the SPSS software package [110].

Results, conclusions from the analysis of the mortality rates

On the basis of the analysis of mortality rates, the results may be divided into two groups. In the first group (including cerebrovascular disease, ischaemic heart disease, chronic hepatopathy, respiratory diseases) it was found that the cross-correlation function between the residual had a significant value (with $k=0$ time lag), the two time series behaved similarly, in other words, the functions were “synchronised.” From the similar behaviour of the two time series it could be concluded that the elements of both time series had similar environmental effects. This means that environmental effects play an important role in the development of the given disease [50], [51], [55]. These results are in accordance with the findings of medical research. In the development of cerebrovascular disease and ischaemic heart disease, it is diets rich in fat, insufficient physical exercise and stress, while in the case of chronic liver disease, excessive consumption of alcohol are considered to play significant roles.

In the other group of diseases (cervical cancer, breast cancer, tumours of the digestive organs, tumours of the respiratory organs) it was found that the cross-correlation function between the residual did not have a significant value, the behaviour of the two time series did not show a similarity, in other words, the functions were not “synchronised.” No similar environmental effects on the elements of the two time series could be deduced here. This means that the role of environmental effects in the development of these diseases cannot be proved, although it cannot be clearly excluded either [50], [51], [55]. This is also in accordance with the results of medical research, since the causes of tumorous diseases are not

yet fully investigated. Hereditary factors (the inheritance of faulty genes) are considered to play a significant role in case of certain tumorous diseases, while the role of environmental effects is not regarded as important (e.g. the development of breast cancer). By contrast, in case of other tumorous diseases (e.g. lung cancer), environmental effects are considered to have a significant role. This is only an apparent contradiction between the results of the analysis and the findings of medical research, since the causes of these diseases are only partially known. Thus, it is also not proved what roles environmental effects may play in the formation of tumorous diseases. Although the direct effect of the environment could not be proved by the method used in the present study, its role could not be clearly excluded either. The conclusions drawn from the Hungarian data also correspond to the findings of analyses earlier performed in other countries with similar methods [55].

The evaluation of the confidence interval calculations

It can be seen from the estimates of the confidence intervals for the ϕ parameter of the AR(1) process that approximately the same confidence intervals were found with the use of normal distribution and with White's method, although the latter method of estimation proved to be slightly more accurate than the former. Confidence intervals were 2 to 4% lower in case of White's estimation in comparison with the method of normal distribution. Confidence intervals were 30 to 40% lower when the method of continuous distribution was used in comparison with the intervals gained using normal distribution.

If the AR(1) parameter $\phi \approx 1$, then the methods of normal distribution and White's estimation cannot always be used, because the upper limit of the confidence interval may be higher than 1. In a stationary case, the result will be useless. Using the first two methods, the upper limits of the confidence intervals may be above 1, in which case the results cannot be interpreted. The upper limits gained from estimation based on a continuous time distribution model were always below 1. The above examples well illustrate the advantages of the continuous time method over the other two [6], [7], [8], [11].

Conclusions from the analysis of the data of lymphoid leukaemia patients

No significant conclusions may be drawn from the analysis of the total number of patients, since the fluctuation of the number of patients in time cannot be evaluated. The description of the seasonal components proved to be an appropriate method for the analysis of the fluctuation in the number of patients and the drawing of the conclusions.

In the course of the analyses I first examined the birth dates of the patients, where I observed no characteristic aggregations. This indicates that external environmental effects on the mother and the embryo played no significant role in the development of the disease.

From the analysis on the basis of the diagnosis dates of the disease I observed an aggregation during the winter months. This seems to suggest that the external environment plays a significant role in the manifestation of the disease. It is suspected that the viral epidemics occurring during the winter months have a significant role in the manifestation of the disease.

7. KÖSZÖNETNYILVÁNÍTÁS

Hálásan köszönöm Dr. Arató Mátyás Professzor Úrnak iránymutató tanítását, támogatását és segítségét, mellyel lehetővé tette munkám rendszerebe foglalását, megvalósítását.

Köszönetem fejezem ki Dr. Herdon Miklós tanszékvezető egyetemi docens úrnak a szakmailag jelentős nemzetközi és hazai konferenciákon való részvételémhez nyújtott segítségéért, támogatásáért.

Tisztelettel köszönöm a Pécsi Orvostudományi Egyetemen dolgozó Dr. Méhes Károly akadémikus úrnak és Dr. Kajtár Pál professzor úrnak a témában nyújtott orvosszakmai segítséget és a közös publikációban való együttműködési lehetőséget.

8. IRODALOM

1. Anderson, R.L.: Distribution of the serial correlation coefficient, *Ann. Math. Stat.* **13** (1942).
2. Allard, R.: Use of time series analysis in infectious disease surveillance, *Bull World Health Organ.* **76** (1998), 327-33.
3. Andrade, I.C., Clare, A.D., O'Brien, R.J., Thomas, S.H.: Tests for stochastic seasonality applied to daily financial time series, *Manchester School.* **67** (1999), 39-59.
4. Arató, M., Benczúr, A.: Gauss-Markov folyamatok maximum-likelihood becslésének egzakt eloszlása, in: Idősorok analízise, Ed. Tusnády, G., Ziermann, M.) Műszaki Kiadó, Budapest, (1986), 85-117.
5. Arató, M., Knuth, E.: *Sztochasztikus folyamatok elemei.* Tankönyvkiadó, Budapest, (1970).
6. Arató, M.: *Linear stochastic systems with constant coefficients: A statistical approach.* Springer, Berlin, (1982).
7. Arató, M.: Komplex stacionárius Gauss Markov folyamat „csillapodási” paraméterének becslése és konfidencia intervallumainak megszerkesztése, *MTA SZTAKI Közlemények.* (1972), 122-161.
8. Arató, M., Benczúr, A.: Szimulációs eredmények az elemi Gauss folyamat paramétereinek becsléseinek eloszlására, *MTA SZTAKI Közlemények.* **8** Budapest, (1972), 3-34.

9. Armstrong, J.S.: Forecasting with econometric methods: folklore versus fact with discussion, *Journal of Business*. **51** (1978), 549-600.
10. Austin, P.C., Mamdani, M.M., Chan, B.T., Lin, E.: Anxiety-related visits to Ontario physicians following September 11, 2001, *Can. J. Psychiatry*. **48** (2003), 416-419.
11. Benczúr, A.: Stacionárius Gauss Markov folyamat csillapodási paraméterének konfidencia határai, *MTA Számítástechnikai Központja Közlemények*. **6** Budapest, (1971), 1-14.
12. Bishai, D.M.: Infant mortality time series are random walks with drift: are they cointegrated with socioeconomic variables? *Health Econ*. **4** (1995), 157-167.
13. Box, G.E.P., Tiao, G.C.: Intervention analysis with applications to economic and environmental problems, *Journal of the American Statistical Association*. **70** (1975), 70-79.
14. Box, G.E.P., Jenkins, G.M.: *Time series analysis forecasting and control*. Holden-Day, San Francisco, (1970).
15. Box, G.E.P., Pierce, D.A.: Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American Statistical Association*. **64** (1970).
16. Boyles, R.P., Raman, S.: Analysis of climate trends in North Carolina (1949-1998), *Environment International*. **29** (2003), 263-275.
17. Buffoni, L., Maugeri, M., Nanni, T.: Precipitation in Italy from 1833 to 1996, *Theoretical and Applied Climatology*, **63** (1999), 33-40.

18. Cani, L., Rios, M., Sanchez, J.: Meningococcal disease in Spain: seasonal nature and recent changes, *Gac. Sanit.* **15** (2001), 336-340.
19. Chatfield, C.: *The analysis of time series*. Chapman&Hall, London, (1992).
20. Chan, B.Y.N.H., Wei, C.Z.: Asymptotic inference for nearly nonstationary AR(1) processes, *The Annals of Statistics*, **15** (1987), 1050-1063.
21. Choi, K., Thacker, S.B.: An evaluation of influenza mortality surveillance, 1962-1979. II Percentage of pneumonia and influenza deaths as an indicator of influenza activity, *A. J. Epidemiol.* **113** (1981), 227-35.
22. Cohen, P.: The influence on survival of onset of childhood acute leukaemia (ALL), *Chronobiol. Int.* **4** (1987), 291-297.
23. Costa, M.da C., Monta, E.L., Paim, J.S., Silva, L.M., Teixeira, Mda G., Mendes, C.M.: Infant mortality in Brazil during recent periods of economic crisis, *Rev. Saude Publica.* **37** (2003), 699-706.
24. Crabtree, B.F., Ray, S.C., Schmidt, P.M., O'Connor, P.J., Schmidt, D.D.: The individual over time: time series applications in health care research, *J. Clin. Epidemiol.* **43** (1990), 241-60.
25. Csáki, P.: ARMA folyamatok, in: (Idősorok analízise, Ed. Tusnády, G., Ziermann, M.) Műszaki Kiadó, Budapest, (1986), 50-84.
26. Daniels, H.E.: The approximate distribution of serial correlation coefficients, *Biometrika.* **43** (1956), 169- .

27. David, A., Dickey, Wayne, A., Fuller: Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica*. **49** (1949), 1057-1072.
28. Diaz, J., Jordan, A., Garcia, R., Lopez, C., Alberdi, J.C., Hernandez, E., Otero, A.: Heat waves in Madrid 1986-1997: effects on the health of the elderly, *Int. Arch. Occup. Environ. Health*. **75** (2002), 163-70.
29. Diaz, J., Garcia, R., Velazquez de Castro, F., Hernandez, E., Lopez, C., Otero, A.: Effects of extremely hot days on people older than 65 years in Seville (Spain) from 1986 to 1997, *Int. J. Biometeorol.* **46** (2002), 145-149.
30. Dimitrov, B.D., Shangova-Grigoriadi, S., Grigoriadis, E.D.: Cyclicity in variations of incidence rates for breast cancer in different countries, *Folia Med.* **40** (1998), 66-71.
31. Dominguez, A., Munoz, P., Martinez, A., Orcau, A.: Monitoring mortality as an indicator of influenza in Catalonia, Spain, *J. Epidemiol. Community Health*. **50** (1996), 293-298.
32. Durbin, J.: Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables, *Econometrica*. **37** (1969).
33. Egger, M., Jennings, S., Spuhler, T., Zimmermann, H.P., Paccaud, F., Somaini, B.: Mortality in influenza epidemics in Switzerland 1969-1985, *Schweiz Med. Wochenschr.* **119** (1989), 434-439.
34. Éltető, Ö., Meszéna, Gy., Ziermann, M.: *Sztochasztikus módszerek és modellek*, Közgazdasági és Jogi Könyvkiadó, Budapest, (1982).
35. Fazekas, I.: *Bevezetés a matematikai statisztikába*, Debrecen, (1997).

36. Fleming, D.M., Cross, K.W., Sunderland, R., Ross, A.M.: Comparison of the seasonal pattern of asthma identified in general practitioner episodes, hospital admissions and deaths, *Thorax*. **8** (2000), 662-665.
37. Franses, P.H., Neele, J., van Dijk, D.: Modeling asymmetric volatility in weekly Dutch temperature data, *Environmental modelling&software*. **16** (2001), 131-137.
38. Franses, P.H., Ooms, M.: A periodic long-memory model for quarterly UK inflation, *International Journal of Forecasting*. **13** (1997), 117-126.
39. Gál, Z., Iglói, E., Terdik, Gy.: Nagysebességű informatikai hálózat adatforgalmának matematikai statisztikai jellemzése, *Alkalmazott Matematikai Lapok*. **19** (1999), 29-38.
40. Garádi, J., Krámli, A., Ratkó, I., Ruda, M.: Statisztikai és számítástechnikai módszerek alkalmazása kórházi morbiditás vizsgálatokban, *MTA SZTAKI Tanulmányok*. **35** (1975), 3-31.
41. Gehlen, M., Rabouille, C., Ezat, U., GuidiGuilvard, L. D.: Drastic changes in deep-sea sediment porewater composition induced by episodic input of organic matter, *Limnology and Oceanography*. **42** (1997), 980-986.
42. Gerencsér, L., Michaletzky, Gy., Vágó, Zs.: Risk-sensitive identification of ARMA processes, *Proceedings of the 37th IEEE Conference on Decision and Control*. **1** (1998), 215-216.
43. Ghysels, E., Osborn, D.: *Themes in modern economics*, Cambridge, (2001).

44. Goh, C., Law, R.: Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention, *Tourism Management*. **23** (2002), 499-510.
45. Gonzalez Ochoe, E., Armas Perez, L., Perez Rodriguez, A., Goyenechea Hernandez, A., Aguirre Jaime, A.: Progress in surveillance of acute respiratory infections in the Republic of Cuba, *Rev. Sanid. Hig. Publica*. **67** (1993), 305-14.
46. Grech, V., Savona-Ventura, C., Agius-Muscat, H., Janulova, L.: Seasonality of births in associated with seasonality of marriages in Malta, *J. Biosoc. Sci.* **35** (2003), 95-105.
47. Hajtman, B.: *Bevezetés a matematikai statisztikába*, Akadémiai Kiadó, Budapest, (1971).
48. Hannan, E. J.: *Time series analysis*. Butler&Tanner Ltd, (1962).
49. Harris, R.E., Harrel, F.E., Patil, K.D., Al-Rashid, R.: The seasonal risk of paediatric/childhood acute lymphocyte leukaemia in the United States, *J Chronic Dis*. **40** (1987), 915-923.
50. Haugh, L.D., Box, G.E.P.: Identification of dynamic regression (distributed lag) models connecting two time series, *J Am Stat Assoc*. **72** (1977), 121-130.
51. Haugh, L.D.: Checking the independence of two covariance stationary time series: a univariate residual cross-correlation approach, *J Am Stat Assoc*. **71** (1976), 378-385.
52. Haywood, J., Wilson, G.T.: Selection and estimation of component models for seasonal time series, *Journal of Forecasting*. **19** (2000), 393-417.

53. Haywood, J., Wilson, G.T.: An improved state space representation for cyclical time series, *Biometrika*. **87** (2000), 724-726.
54. Heiler, S., Feng, Y.: Data-driven decomposition of seasonal time series, *Journal of Statistical Planning and Inference*. **91** (2000), 351-363.
55. Helfenstein, U.: Detecting hidden relations between time series of mortality rates, *Methods Inf Med*. **29** (1990), 57-60.
56. Helfenstein, U., Wanner, H.U.: The enviromental accident at 'Schweizerhalle' and respiratory diseases in children: a time series analysis, *Statistics in Medicine*. **10** (1991), 1481-1492.
57. Helfenstein, U.: The use of transfer function models, intervention analysis and related time series methods in epidemiology, *Internatioinal Journal of Epidemiology*. **20** (1991), 808-815.
58. Huhn, E.: ARMA folyamatok egzakt sűrűségfüggvénye, *Alkalmazott Matematikai Lapok*. **10** (1984).
59. Huhn, E.: Rekurzív algoritmus ARMA folyamatok likelihood függvényének számolására, *Alkalmazott Matematikai Lapok*. **15** (1990-91), 303-307.
60. Hulyák, K.: *Idősorok sztochasztikus modelljei*, KSH Ökonometriai Laboratórium, Budapest, (1976).
61. Izsák, J.: A halálozási struktúra változásainak vizsgálata disszimilitási indexekkel, *Statisztikai Szemle*. **75** (1997), 130-140.
62. Jenkins, D.M., Watts, D.G.: *Spectral analysis and its applications*. Holden-Day, San Francisco, (1968).

63. Józan, P.: Epidemiológiai válság Magyarországon a kilencvenes években (I.), *Statisztikai Szemle*. **72** (1994), 5-20.
64. Józan, P.: Epidemiológiai válság Magyarországon a kilencvenes években (II.), *Statisztikai Szemle*. **72** (1994), 101-113.
65. Juvancz, I., Lipták, T.: A matematikai statisztika orvosi-biológiai alkalmazásának néhány problémájáról, *MTA Alkalmazott Matematikai Intézet Közleményei*. **1** (1952), 175-205.
66. Katial, R.K., Zhang, Y., Jones, R.H., Dyer, P.D.: Atmospheric mold spore counts in relation to meteorological parameters, *International Journal of Biometeorolgy*. **41** (1997), 17-22.
67. Knuth, E.: Központi egység kihasználtságának szimulációs vizsgálata másodrendű autoregresszív folyamatok alkalmazásával, *MTA SZTAKI Közlemények*. **15** (1975), 141-145.
68. Kormos, J.: Hipotézis vizsgálat közel nem stacionárius AR(1) esetén, *Alkalmazott Matematikai Lapok*. **14** (1989), 27-35.
69. Kopjar, B., Guldvog, B.: Time variations in injury incidence, *NIPH Ann.* **15** (1993), 3-11.
70. Kovács, K.: Idősorok alapján történő előrejelzés főbb módszertani kérdései: különös tekintettel a Box-Jenkins modellekre, disszertáció. MKKE, Budapest, 1984.
71. Krámlí, A., Ratkó, I., Ruda, M., Soltész, J.: A statisztikai adatfeldolgozás matematikai és számítástechnikai problémái, Hospitalizált morbiditási statisztikákkal kapcsolatos megfontolások (Esettanulmány), *MTA SZTAKI Tanulmányok*. **70** (1977), 5-67.

72. Krámlí, A., Pergel, J.: The connection between Gaussian Markov processes and autoregressive-moving average processes, *MTA SZTAKI Közlemények*. **13** (1974), 53-57.
73. Krámlí, A., Pergel, J.: Auttoregressziós típusú folyamatok által generált mértékek Radom-Nikodym deriváltjáról, *Alkalmazott Matematikai Lapok*. **1** (1975), 73-79.
74. Kusznierz, G.F., Imaz, M.S., Zerbini, E.V., Savy, V., Knez, V., Sequeira, M.D.: Effect of influenza epidemics on mortality in Santa Fe, Argentina, during 1992-1999, *Rev. Panam. Salud. Publica*. **12** (2002), 26-36.
75. Leifman, H., Romelsjö, A.: The effect of changes in alcohol consumption on mortality and admissions with alcohol-related diagnoses in Stockholm County- a time series analysis, *Addiction*. **92** (1997), 1523-1536.
76. Liéli, R.: Egyváltozós időszori modelleken alapuló inflációs előrejelzések, *Statisztikai Szemle*. **77** (1999), 522-549.
77. Lim, C., McAleer, M.: Monthly seasonal variations Asian tourism to Australia, *Annals of Tourism Research*. **28** (2001), 68-82.
78. Liu, H., Imai, K., Suzuki, K., Nojiri, Y., Tsurushima, N., Saino, T.: Seasonal variability of picophytoplankton and bacteria in the western subarctic Pacific Ocean at station KNOT, *Deep-Sea Research II*. **49** (2002), 5409-5420.
79. Lopez del Val J.A., Calvete Fernandez, H.I., Carreter Oronez, C.A., Abaurrea Leon, J., Muniesa Cuenca, M.P., Garcia Mata, J.R., Hernandez Navarrete, M.J., Arribas Llorente, J.L.: Time – series analysis applied to nosocomial infection, *Med. Clin.* **99** (1992), 52-526.

80. Makridakis, S., Wheelwright, S.C.: *Forecasting Methods for Management*. Boston, (1989).
81. Marco, V.J., Ruth, F.B.: Scaling properties and symmetrical patterns in the epidemiology of rotavirus infection, *Philosophical Transactions of The Royal Society of London. Series B: Biological Sciences*. **358** (2003), 1625-1641.
82. Martinez-Schnell, B., Zaidi, A.: Time series analysis of injuries, *Statistics in Medicine*. **8** (1989), 1497-1508.
83. Matsuda, S., Kahyo, H.: Geographic differences in seasonality of preterm births in Japan, *Hum. Biol.* **70** (1998), 919-935.
84. Meltzer, A.A., Annegers, J.F., Spitz, M.R., Month-of-birth and incidence of acute lymphoblastic leukaemia in children, *Leuk. Lymph.* **23** (1996), 85-92.
85. Moreno-Civantos, A., Diaz-Jimenez, J., Dominguez-Berjon, M.F.: Shared surveillance: meningococcal disease vs influenza, *Gac. Sanit.* **14** (2000), 422-428.
86. Nelson, C.R., Plosser, C.I.: Trends and random walks in macroeconomic time series: some evidence and implications, *Journal of Monetary Economics*. **10** (1982), 139-162.
87. Norstrom, T., Skog, O.J.: Saturday opening of alcohol retail shops in Sweden: an impact analysis, *J. Stud. Alcohol*. **64** (2003), 393-401.
88. Oberg, A.L., Ferguson, J.A., McIntire, L.M., Horner, R.D.: Incidence of stroke and season of the year: evidence of an association, *Am. J. Epidemiol.* **152** (2000), 558-564.
89. Odriozola, A.J.C., Jimenez, D.J., Rubio, M.J.C., Perez, M.I.J., Ortiz, P.M.S., Rodrigues, R.P.: Air pollution and mortality in

- Madrid, Spain: a time series analysis, *Arch. Occup. Environ. Health*. **71** (1998), 543-549.
90. Pandit, S.M., Wu, S.: *Time series and system analysis with applications*. John Wiley and Sons, New York (1983).
 91. Paronen, T., Haukka, J., Viilo, K., Hakko, H., Pirkola, S., Isometsa, E., Lonnqvist, J., Sarkioja, T., Vaisanen, E., Rasanen, P.: Cyclic time patterns of death from suicide in northern Finland, *J. Affect Disord.* **78** (2004), 11-19.
 92. Phillips, P.C.B.: Approximations to some finite sample distributions associated with a first-order stochastic difference equation, *Econometrica*. **45** (1977), 463-485.
 93. Pierce, D.A.: Relationship - and the lack there of - between economic time series, with special reference to money and interest rates, *Journal of the American Statistical Association*. **72** (1977), 11-26.
 94. Prékopa, A.: *Valószínűségelmélet műszaki alkalmazásokkal*. Műszaki Könyvkiadó, Budapest (1972).
 95. Prochazka, B., Benes, C.: Evaluation of time series trends in the weekly count of diseases, *Epidemiol. Mikrobiol. Immunol.* **48** (1999), 52-9.
 96. Quenouille, M.H.: Approximate tests of correlation in time series, *Jour. Royal Stat. Soc.* **B11** (1949).
 97. Rao, M.M.: Asymptotic distribution of an estimator of the boundary parameter of an unstable process, *The Annals of statistics*. **6** (1978), 185-190.

98. Ratkó, I., Ruda, M.: On an estimate for the parameter of a multidimensional stationary Gaussian Markov process, and an application, *MTA SZTAKI Közlemények*. **13** (1974), 21-29.
99. Ratkó, I.: Microcomputer in epidemiology, in: (Lecture Notes in Medical Informatics, Ed. Lindberg, D.A.B., Reichertz, P.L.) Springer-Verlag Berlin (1984), 477.
100. Reinfurt, D.W., Stewart, J.R., Weaver, N.L.: The economy as a factor in motor vehicle fatalities, suicides, and homicides, *Accid. Anal. Prev.* **23** (1991), 453-462.
101. Rios, M., Garcia, J.M., Cubedo, M., Perez, D.: Analysis of temporary series in the epidemiology of typhoid fever in Spain, *Medicina Clinica*. **106** (1996), 686-689.
102. Rock, S.M.: Impact of the 65 mph speed limits on accidents, deaths, and injuries in Illinois, *Accid. Anal. Prev.* **27** (1995), 207-14.
103. Rosel, J., Elosegui, E.: Daily and weekly smoking habits: a Box-Jenkins analysis, *Psychol. Rep.* **75** (1994), 1639-1648.
104. Ruda, M.: Parameter estimation in the first order autoregressive process, *Biometrika*. **61** (1974), 632-633.
105. Saez, M. Sunyer, J., Castellsague, J., Murillo, C., Anto, J.M.: Relationship between weather temperature and mortality: a time series analysis approach in Barcelona. *Int. J. Epidemiol.* **24** (1995), 576-582.
106. Shindo, N.: Morbidity and mortality of influenza in Japan, *Nippon Rinsho*. **58** (2000), 2187-2191.

107. Saynajakangas, P., Keistinen, T., Tuuponen, T.: Seasonal fluctuations in hospitalisation for pneumonia in Finland, *Int. J. Circumpolar Health*. **60** (2001), 34-40.
108. Sorenson, H.T., Pedersen, L., Olse, J.H., et al.: Seasonal variation in month of birth and diagnosis of early childhood acute lymphoblastic leukaemia, *J. A. M. A.* **285** 168-169.
109. Sugár, A.: Determinisztikus vagy sztochasztikus modell? Az időszori modellezés problémái, disszertáció. BKTE, Budapest, 1991.
110. SPSS – “Statistical Package for Social Sciences” version 9.0
<http://www.spss.com>.
111. Tam, W. K., Reinsel, G.C.: Tests for seasonal moving average unit root in ARIMA models, *Journal of the American Statistical Association*. **92** (1997), 725-738.
112. Terdik, Gy.: Az autoregressziós mező mozgó átlag előállítás és együtthatóinak maximum likelihood becslése, *Alkalmazott Matematikai Lapok*. **3** (1977), 379-388.
113. Tine, H., Strodl, A.: Trends and seasonal variations in the occurrence of salmonella in pigs, pork and humans in Denmark, 1995-2000, *BMTW Heft*. **114** (2001), 346-349.
114. Tobias, A., Diaz, J., Saez, M., Alberdi, J.C.: Use of poisson regression and Box-Jenkins models to evaluate the short-term effects of environmental noise levels on daily emergency admissions in Madrid, Spain, *Eur. J. Epidemiol.* **17** (2001) 765-71.
115. Tong, S. Bi.P., Donald, K., McMichael, A.J.: Climate variability and Ross River virus transmission, *J. Epidemiol. Community Health*. **56** (2002), 617-621.

116. Tong, S., Hu, W.: Climate variation and incidence of Ross river virus in Cairns, Australia: a time series analysis, *Environ. Health Perspect.* **109** (2001), 1271-1273.
117. Upshur, R.E., Knight, K., Goel, V.: Time-series analysis of the relation between influenza virus and hospital admissions of the elderly in Ontario, Canada, for pneumonia, chronic lung disease, and congestive heart failure, *Am. J. Epidemiol.* **149** (1999), 85-92.
118. Vienna, N.J., Polan, A.K.: Childhood lymphatic leukaemia prenatal seasonality and possible association with congenital varicella, *Am. J. Epidemiol.* **103** (1976), 321-332.
119. Vilkinen, S., Keistinen, T., Tuuponen, T., Kivela, S.L.: Seasonal variation in hospital admissions for chronic obstructive pulmonary disease in Finland, *Artic. Med. Res.* **55** (1996), 182-186.
120. Vincze, I.: *Matematikai statisztika*, Tankönyv Kiadó, Budapest, (1980).
121. Walach, H., Lowes, T., Mussbach, D., Schamell, U., Springer, E., Stritzl, G., Haag, G.: The long-term effects of homeopathic treatment of chronic headaches: one year follow-up and single case time series analysis, *Br. Homeopath. J.* **90** (2001), 63-72.
122. Williams, B.M., Durvasula, P.K., Brown, D.E.: Urban freeway traffic flow prediction – Application of seasonal autoregressive integrated moving average and exponential smoothing models, *Taffic flow theory transportation research record.* **1644** (1998), 132-141.
123. Xu Yu, J., Hg, K.M., Huang, J.Z.: Patterns discovery based on time-series decomposition, in: (*PAKDD 2001, LNAI 2035, Eds.*

- Cheung, D., Williams, G.J., Li, Q.*) Springer-Verlag, Berlin Heidelberg (2001), 336-347.
124. Zhang, X.B., Hogg, W.D., Mekis. E.: Spatial and temporal characteristics of heavy precipitation events over Canada, *Journal of Climate*. **14** (2001), 1923-1936.
125. Ziermann, M., Meszéna, Gy.: *Valószínűségszámítás*, Budapest, (1980).
126. White, J.S.: A t-test for the serial correlation coefficient, *Ann. Math. Stat.* **28** (1957), 1046-1048.

9. DISSZERTÁCIÓHOZ KAPCSOLÓDÓ PUBLIKÁCIÓK

1. Fazekas, M.: Special time series models for analysing of mortality data, *Lecture Notes in Computer Science*. **2199** (2001), 81-87.
2. Kajtár P., Fazekasné Kis M., Méhes K.: A születés időpontja gyermekkori lymphoid leukaemiában, *Orvosi Hetilap*. **144** (2003) 1869-1871.
3. Fazekasné Kis M.: Idősori analízis gyakorlati alkalmazása egészségügyi területen, *Debreceni Egyetem Agrártudományi Közlemények*. **2** (2001), 14-22.
4. Fazekas, M.: Applications of seasonal time series for analysing the occurrence of childhood leukaemia in Hungary, *Controlled Clinical Trials*; **24** (2003), 101.
5. Kis, M.: Analysis of the time series for some causes of death, Surján, G. et al. Eds. in *Health Data in the Information Society. Studies in Health Technology and Informatics*. **90** (2002), 439-443.
6. Fazekas, M.: Analysing the occurrence of childhood leukaemia using seasonal time series, Damiani, E. et al. Eds. in *Knowledge-Based Intelligent Information Engineering Systems&Allied Technologies. Frontiers in Artificial Intelligence and Applications*. **82** (2002), 950-954.
7. Fazekas, M.: Applications of seasonal time series for analysing the occurrence of childhood leukaemia, In *The New Navigators: from Professionals to Patients*. IOS Press. *Proceedings published in hard back by IOS Press. MIE 2003 France, Saint Malo*. (2003).

8. Fazekas, M.: Application of ARIMA models, *3rd International Conference on Telecommunications for Training. Prague, Czech Republic*, (2001), 50-54.
9. Fazekasné Kis M.: Idősor analízis alkalmazása halálozási okok adatainak elemzésére, *XXII. Neumann Kollokvium, A számítástechnika orvosi és biológiai alkalmazásai. Veszprém*, (2000), 16-19.
10. Fazekasné Kis M.: Autoregresszív integrált mozgóátlag modellek egészségügyi alkalmazása, *Média-Informatika-Kommunikáció. Veszprém*, (2001), 58-62.
11. Fazekasné Kis M.: ARIMA modellek alkalmazása halálozási adatok elemzésére, *Informatika a felsőoktatásban. Debrecen*, (2002).
12. Fazekas, M.: Application of special time series model, *EFITA 2001. Montpellier, France*, (2001), 93-94.
13. Fazekas, M.: Application time series models on medical research, *6th International Conference on Applied Informatics. Eger*, (2004).

10. DISSZERTÁCIÓ TÉMAKÖRÉN KÍVÜLI PUBLIKÁCIÓK

10.1. Jegyzetek

14. Fazekasné Kis M.: *Informatika előadások*, Főiskolai jegyzet, (1999).
15. Fazekasné Kis M.: *Informatika, Gyakorlati feladatok*, Főiskolai jegyzet, (1999).
16. Fazekasné Kis M.: *Matematika és statisztika elemei*, Főiskolai jegyzet, (1999).
17. Fazekasné Kis M.: *Matematika és statisztika, gyakorló feladatok*, Főiskolai jegyzet, (1999).

10.2. Idegen nyelvű közlemények

18. Furka, I., Mikó, I., Tóth, K., Furka, A., Kappelmyer, J., Szikszai, Z., Góth, L., Kovács, J., Baráth, E., Imre, S., Fazekas, M.: Hematological, hemorheological and catalase level following splenectomy and spleen autotrasplantations in experimental animals – Early results, *Acta Chirurgica Austriaca*. Jahrgang **29** Supplement Nr. 137. 31-32.
19. Fazekas B., Kis M., Hajdúné Tóth E.: Data on the contamination of maize with fumanisin B₁ and other fusariotoxins in Hungary, *Acta Veterinaria Hungarica*. **44** (1996), 25-37.

10.3. Magyar nyelvű közlemények

20. Fazekasné Kis M., Pethő A.: Integrált egészségügyi és gazdasági információs rendszerek értékelésének fontosabb szempontjai a Debreceni Orvostudományi Egyetemen, *Egészségügyi Gazdasági Szemle*. **32** (1994), 560-568.
21. Pethő A., Fazekasné Kis M., Fehértói J.-né., Hadházi A.: Hogyan fejlődik a Debreceni Orvostudományi Egyetem számítógépes információs rendszere? *Lage Artis Medicine – Új Magyar Orvosi Hírmondó*. **5** (1995), 814-819.
22. Fazekasné Kis M.: A szakértői rendszerek alkalmazási lehetőségeiről az egészségügyben – szakirodalmi adatok alapján, *Egészségügyi Gazdasági Szemle*. **34** (1997), 231-235.
23. Fazekasné Kis M., Góth L.: A ROC analízis alkalmazása az enzim diagnosztikában, *Klinikai és Kísérletes Laboratóriumi Medicina*. **26** (1999), 184-186.
24. Fazekasné Kis Mária: ROC analízis alkalmazása, *Debreceni Egyetem Agrártudományi Közlemények*. **1** (2002), 4-7.

10.4. Magyar nyelvű előadások

25. Fazekasné Kis M.: Integrált egészségügyi és gazdasági információs rendszerek értékelésének fontosabb szempontjai a Debreceni Orvostudományi Egyetemen, *Medicomp '94 Számítástechnikai és kibernetikai módszerek az orvostudományban és a biológiában*. Szeged, 144-146.
26. Pethő A., Fazekasné Kis M., Fehértói J.-né., Hadházi A.: Hogyan fejlődik a Debreceni Orvostudományi Egyetem számítógépes

- információs rendszere? Neumann János Számítógéptudományi Társaság VI. Országos Kongresszus kiadványa. Siófok, 1995. május 28-31. 536-546.
27. Fazekasné Kis M.: Az egészségügyi szakértői rendszerek a DOTE-n folyó informatikai oktatás szemszögéből, *Informatika a felsőoktatásban '96 & Networkshop '96*. Debrecen, 1996. Augusztus 27-30. 390-400.
28. Fazekasné Kis M.: A ROC analízis oktatása az orvosdiagnosztikai laboratóriumi analitikus szakon. *Informatika a felsőoktatásban '99*. Debrecen, 1999. Augusztus 27-29. 228-233.

10.5. Poszter

29. Kis, M.: Receiver operating characteristic plots for clinical laboratory tests, *23rd Annual Conference of International Society for Clinical Biostatistics*. Dijon, France, 9-13 September 2002.

10.6. Egyéb

30. Fazekasné Kis M., Kériné Fülöp I., Virágos Cs.: *Reference Manager 7.01 verzió, Ismertetés a program működéséről*. Szerkesztő: Pethő, A. Kézikönyv, Debrecen, (1997).
31. Fazekas B., Kis M., Konczné Tar A., Tóthné Hajdú E.: Fumonisin B₁ tartalmú fusarium moniliforme gombatenyészet etetésének hatása napos korú kacsákra, *MTA Állatorvos-Tudományi Bizottsága, Akadémiai Beszámoló*. Budapest, (1999).

NÉHÁNY BETEGSÉG STATISZTIKAI ADATAINAK IDŐSORI ELEMZÉSE

Értekezés a doktori (PhD) fokozat megszerzése érdekében
az Informatika tudományágban

Írta: Fazekasné Kis Mária okleveles matematikus

Készült a Debreceni Egyetem Matematika- és Számítástudomány
Doktori Iskolája Informatika programja keretében

Témavezető: Prof. Dr. Arató Mátyás

A doktori szigorlati bizottság:

elnök:

Dr.

tagok:

Dr.

Dr.

A doktori szigorlat időpontja: 200... ..

Az értekezés bírálói:

Dr.

Dr.

Dr.

A bírálóbizottság:

elnök: Dr.

tagok: Dr.

Dr.

Dr.

Dr.

Az értekezés védésének időpontja: 200... ..