NÉHÁNY BETEGSÉG STATISZTIKAI ADATAINAK IDŐSORI ELEMZÉSE

Doktori értekezés tézisei

TIME SERIES ANALYSIS OF STATISTICAL DATA FOR SOME DISEASES

PhD Theses

Fazekasné Kis Mária

Debreceni Egyetem
Debrecen, 2004

Tartalomjegyzék

1. BEVEZETÉS, CÉLKITŰZÉSEK ... 3

2. IDŐSOROK JELLEMZŐINEK ÁTTEKINTÉSE ... 3

2.1. Sztochasztikus folyamatok ... 3

2.2. ARMA(p,q) modellek ... 4

2.3. Sztochasztikus modellek megadása ... 5
- 2.3.1. Sztochasztikus modellek azonosítása ... 5
- 2.3.2. Sztochasztikus modell paramétereinek becslése ... 5
- 2.3.3. Sztochasztikus modellek ellenőrzése ... 5

2.4. AR(1) konfidencia intervallumai becsléseinek összehasonlításai ... 6

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK ... 8

3.1. Halálozási arányok feldolgozása ... 8
- 3.1.1. Vizsgálati módszerek ... 8
- 3.1.2. Eredmények ... 8
- 3.1.3. Következtetések a halálozási arányok elemzéseiből ... 12

3.2. Gyermekkori lymphoid leukaemiás betegek adatainak feldolgozása ... 14
- 3.2.1. A betegek születési dátumának elemzése ... 14
- 3.2.2. A betegség diagnosztizálási dátumának elemzése ... 14
- 3.2.3. Következtetések a lymphoid leukaemiás betegek adatainak vizsgálatából ... 15

4. ÚJ KUTATÁSI EREDMÉNYEK ... 15

5. PHD THESES – TIME SERIES ANALYSIS OF STATISTICAL DATA FOR SOME DISEASES ... 17

6. IDÉZETT IRODALOM ... 20

7. AZ ÉRTEKEZÉSHEZ KAPCSOLÓDÓ PUBLIKÁCIÓK ... 20

8. DISSZERTÁCIÓ TÉMAKÖRÉN KÍVÜLI PUBLIKÁCIÓK ... 21

1. BEVEZETÉS, CÉLKITŰZÉSEK

Az értekezés két egymással összefüggő témakörrel foglalkozik. Az értekezés első részében az idősorok jellemzőivel, sajátosságaival, törvényszerűségeivel foglalkozom. Idősornak nevezzük az időben diszkrét $\xi_{t_1}, \xi_{t_2}, \ldots$ $\left(t_1 < t_2 < t_3 < \ldots < t_n < \ldots\right)$ valószínűségi változók összességét, ahol a ξ_{t_i} valószínűségi változók nem függetlenek. Munkám második részében az idősorok számítógépes alkalmazásait mutatom be. A számítógépes feldolgozások során olyan humán megbetegedéseket vizsgáltam, melyek oktana komplex és tisztázatlan. Az idősori elemzések célja az volt, hogy vizsgáljam a környezeti hatások szerepét ezeknek a betegségeknek a kialakulásában, időbeli lefolyásában.

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 *(méhnyakrák, mellrák, emésztőszervi daganatok (gyomor- és vastagbélrák), légzőszervi daganatok (légcső-, hörgő-, tüdőrák))* okozta halálozási arányok idősor analízis módszereivel történő vizsgálata.
3. A környezeti hatások szerepének kimutathatóságának vizsgálata a fenti betegségek kialakulásában és időbeli lefolyásában.
4. A vizsgált halálozási arányokra kapott AR(1) együttható 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, a képletek és az irodalmi hivatkozások számozása megegyezik az értekezésbeli számozással. Az irodalomjegyzék csak a tézisek hivatkozásait tartalmazza. A tézisekben a képletek jelölései az idézett [14], [25], [34] irodalom szerint történtek.

2. IDŐSOROK JELLEMZŐINEK ÁTTEKINTÉSE

Az alábbiakban az idősorok olyan jellemzőit ismertetem, melyek a későbbiekben bemutatott számítógépes alkalmazásokhoz kapcsolódnak.

2.1. Sztochasztikus folyamatok

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 $t \in T$ paraméter halmazt. 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 részhalmazaiból alkotott σ-algebrát jelöli. A sztochasztikus folyamatot többféle módon szokás jelölni: $\xi_t(\omega)$ vagy ξ_t illetve $\xi(t,\omega)$ és $\xi(t)$. 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].

2.2. ARMA(p,q) modellek

A matematikai statisztikából ismert, hogy a legegyszerűbb idősor az u.n. fehérzaj sorozat, amikor az ε_t, ε_{t+1}, ε_{t+2}, ... független, azonos eloszlású véletlen változók sorozata $N(0,\sigma_\varepsilon^2)$ (normális) eloszlással. Egy véletlen folyamatot megfigyelünk a ... , t_{-2}, t_{-1}, t_0, t_1, t_2, ... időpontokban, a megfigyelt értékeket ..., z_{-2}, z_{-1}, z_0, z_1, z_2, … jelölje. Ha a z_t megfigyelési érték sztochasztikusan és lineárisan az előző z_{t-1}, ... , z_{t-p}, és ε_t, ... , ε_{t-q} értékektől függ, akkor p-edrendű autoregresszív és q-adrendű mozgóátlag modellnek, un. ARMA(p,q) modellnek nevezzük a következő kifejezést [14], ahol,

$$\tilde{z}_t = z_t - \mu$$

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

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: $B\tilde{z}_t = \tilde{z}_{t-1};\quad és \quad B^m \tilde{z}_t = \tilde{z}_{t-m}$, akkor a (2.35) a következő egyszerű alakban is felírható [14]:

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

$\phi(B)=1-\phi_1B-\phi_2B^2-\ldots-\phi_pB^p$ és $\theta(B)=1-\theta_1B-\theta_2B^2-\ldots-\theta_qB^q$.

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], [34], [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.

2.3. Sztochasztikus modellek megadása

2.3.1. 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 [14].

A stacionaritás vizsgálatával eldöntjük, hogy az adott idősorhoz illeszthető-e ARIMA modell, ha igen, milyen *d* dimenzióval. Továbbiakban a stacionárius vagy azzá transzformált adatokkal dolgozunk.

A modell azonosítás következő lépése során megkeressük a stacionárius vagy azzá transzformált idősorhoz milyen MA(q), AR(p), vagy ARMA(p,q) modelltípus(oka)t lehet illeszteni, amelynek jellemzőire leginkább hasonlítanak a tapasztalati idősorból számított empírikus jellemzők. Ehhez a modelltípusok elméleti jellegzetességeinek ismerete szükséges. A modell azonosításához az autokorrelációs függvények és a parciális autokorrelációs függvények adják a legfontosabb információkat.

2.3.2. Sztochasztikus modell paramétereinek becslése

A sztochasztikus modelltípus azonosítása és a paraméterek kezdőértékének megadása után a paraméterek végleges értékeit kell meghatározni. A paraméterek hatásos becslésének eléréséhez a rendelkezésre álló adatainkat hatékony módon kell felhasználni. Ehhez felhasználhatók a maximum likelihood és a legkisebb négyzetek elvén alapuló módszerek. Az értekezésben a maximum likehood becslést részletesen ismertetem.

2.3.3. Sztochasztikus modellek ellenőrzése

A sztochasztikus modell paramétereinek becslése után ellenőrizni kell a választott modell helyességét, jóságát. Az ARIMA modellek ellenőrzése a reziduumok alapján történik.

Helyes modell választása esetén a reziduumok autokorrelációt nem tartalmaznak, normális eloszlás szerint szóródnak nulla körül, konstans szórással. A reziduumok korrelációs függvénye alkalmazható a modell helyességének vizsgálatára [25]. A reziduumokat $\hat{a}$, a reziduumok autokorrelációinak becslését $r_k\left(\hat{a}\right)$ jelöli.

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

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

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

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

Ha a modell nem megfelelő, Q értéke erősen megnő [14], [25]. Ezt az ellenőrzési módot portmanteau-próbának nevezik. Ekkor 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 a becslést túlillesztésnek nevezzük. Ha a H_0 hipotézist a χ^2 próba nem igazolja, akkor korrigálni kell a választott modellt.

2.4. AR(1) konfidencia intervallumai becsléseinek összehasonlításai

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

Az elsőrendű autoregresszív folyamatot a $\Delta,\ 2\Delta,\ \ldots\ ,\ T\Delta\ (\Delta>0)$ 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]:

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

Ha $\phi \approx 1$ és $1-\phi \approx 1/T$, 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]:

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

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

Az AR(*1*) folyamatokra igaz, hogy a ϕ paraméter értéke az elsőrendű autokorrelációs együtthatóval becsülhető, azaz $r_1 \approx \hat{\phi}$. 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 $\bar{\kappa}$ 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]:

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

A fentiekben ismertetett módszerekkel történt az AR(*1*) együtthatók folytonos idejű modellel való becslésével a konfidencia intervallumok meghatározása.

3. SZÁMÍTÓGÉPES ALKALMAZÁSOK

3.1. Halálozási arányok feldolgozása

3.1.1. Vizsgálati módszerek

A számítógépes feldolgozások során nyolc betegség *(cerebrovasculáris betegség, isémiás szívbetegség, krónikus májbetegségek, légzőszervi betegségek, méhnyakrák, mellrák, emésztőszervi daganatok, légzőszervi daganatok)* okozta halálozási arányok elemzését végeztem el idősor analízis módszerével. A halálozási arányok elemzéseihez 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 komponenst 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 becslése.

A számítógépes alkalmazások során a leukaemiás 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 adtam meg.

Az értekezés 2.6. fejezetében foglaltak figyelembe vételével elkészültek az autokorrelációs- és a parciális autokorrelációs függvények. A lehetséges modelleket kiválasztottam. A férfiak és a nők halálozási arányaiból készült autokorrelációs függvényértékek exponenciálisan csökkentek és a parciális autokorrelációs függvények egy szignifikáns értéket mutattak, $k=1$ esetben. Ezekből arra lehet következtetni, hogy AR(*1*) megfelelő modell. A számítógépes programcsomag alkalmazásával a modellek illesztését elvégeztem, megkaptam a reziduumok varianciáit is.

A paraméterek becslésének elvégzése után a következő egyenleteket kaptam. A férfiak halálozási arányát leíró sztochasztikus egyenlet: $z_t=0,792z_{t-1}+\varepsilon_t;$ $\sigma^2_a=5542,023;$ a nők adataira: $z_t=0,809z_{t-1}+\varepsilon_t;$ $\sigma^2_a=3816,344$. A σ^2_a-t, a reziduumok varianciáit, a program megadta.

Az elsőrendű autoregresszív folyamatok paramétereinek megbízhatóságára az értekezés 2.8. fejezetében említett háromféle módon meghatározott konfidencia intervallumokat adtam meg. A normális eloszlás alkalmazásához és a White féle becsléshez a szükséges $x_{0,05}$ és $t_{0,05}$ értékeket a standard normális eloszlás illetve a Student eloszlás ismert táblázataiból kikerestem. A folytonos idejű modellhez a konfidencia intervallumok meghatározása az értekezés 2.8. fejezetében megadott (2.69); (2.70); (2.71) kifejezésekkel történt, ahol $T=28$, a ϕ paraméter közelítő értéke az elsőrendű autokorrelációs együttható, azaz r_1.

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. 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 is meghatároztam és a fenti módon összehasonlítottam. A számítások eredményeit az értekezés 5. és 6. táblázatai tartalmazzák.

A helyes modell választás igazolásához, a reziduumok tulajdonságait ellenőriztem. A (2.58) kifejezéssel megadott χ^2 próbát alkalmaztam.

A reziduumok autokorrelációiból számolt $\chi^2_{férfiak}=1,7478;$ $\chi^2_{nők}=3,886$ értékek és $\chi^2_{0,05}=11,07$ érték alapján a választott modellek helyesnek bizonyultak [25], [50], [51].

Helyes modell választásakor a reziduumok autokorreláltságot nem tartalmaznak. A 7. ábra a férfiak halálozási arányaira illesztett modell reziduumai autokorrelációit mutatja. Látható, hogy a függvény minden értéke a konfidencia intervallumon belül van. Hasonló ábrát kaptam a nők adataiból is.

```
Autocorrelations:
    Auto- Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75
1
  1   ,155   ,179              .       ⇔***   .
  2  -,049   ,176              .     *⇔       .
  3   ,070   ,173              .       ⇔*     .
  4  -,064   ,169              .     *⇔       .
  5   ,100   ,165              .       ⇔**    .
  6   ,134   ,162               .      ⇔***  .
Plot Symbols:      Autocorrelations *     Two Standard
Error Limits .
```

7. ábra: A férfiak halálozási arányaira illesztett modell reziduumainak autokorrelációi.

A férfiak és a nők halálozási arányait jellemző idősorok közötti kapcsolat elemzéséhez a keresztkorrelációs függvényt alkalmaztam. A reziduumok keresztkorrelációs függvényének egy szignifikáns értéke volt a k=0 időkésleltetésnél.

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 az értekezés 3.1.2.1. fejezetében leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatónak 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 az értekezés 2.8. fejezetében ismertetett háromféle módon végeztem el.

A reziduumok keresztkorrelációs függvényének egy szignifikáns értéke volt a k=0 időkésleltetésnél.

3.1.2.3. Krónikus májbetegség

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

Az adatok vizsgálatát az értekezés 3.1.2.1. fejezetében leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatónak 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 az értekezés 2.8. fejezetében ismertetett háromféle módon végeztem el.

A reziduumok keresztkorrelációs függvényének egy szignifikáns értéke volt a k=0 időkésleltetésnél.

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 az értekezés 3.1.2.1. fejezetében leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatóak. A férfiak adataiból a következő sztochasztikus egyenletet 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 az értekezés 2.8. fejezetében ismertetett háromféle módon végeztem el.

A reziduumok keresztkorrelációs függvényének egy szignifikáns értéke volt a k=0 időkésleltetésnél.

3.1.2.5. Méhnyakrák

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

Az adatok vizsgálatát az értekezés 3.1.2.1. fejezetében leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatóak. A 0-64 éves korcsoport halálozási arányát jellemző sztochasztikus kifejezés: $z_t=0,576z_{t-1}+\varepsilon_t$; $\sigma^2{}_a=0,260$; 65 év felettiekre: $z_t=0,703z_{t-1}+\varepsilon_t$; $\sigma^2{}_a=5,722$. A reziduumok varianciáit $\sigma^2{}_a$ jelöli.

A konfidencia intervallumok becslését az értekezés 2.8. fejezetében ismertetett háromféle módon végeztem el.

A reziduumok keresztkorrelációs függvényének nem volt szignifikáns értéke.

3.1.2.6. Mellrák

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

Az adatok vizsgálatát az értekezés 3.1.2.1. fejezetében leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatónak bizonyultak. A 0-64 éves korcsoportra a sztochasztikus egyenlet a következőnek 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 konfidencia intervallumok becslését az értekezés 2.8. fejezetében ismertetett háromféle módon végeztem el.

A reziduumok keresztkorrelációs függvényének nem volt szignifikáns értéke.

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 elemeztem.

Az adatok vizsgálatát az értekezés 3.1.2.1. fejezetében leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatónak bizonyultak. A férfiak adataiból a sztochasztikus egyenlet a következőnek 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 konfidencia intervallumok becslését az értekezés 2.8. fejezetében ismertetett háromféle módon végeztem el.

A reziduumok keresztkorrelációs függvényének nem volt szignifikáns értéke.

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 az értekezés 3.1.2.1. fejezetében leírtakhoz hasonlóan végeztem el. Az elemzés alapján az elsőrendű autoregresszív modellek elfogadhatónak bizonyultak. A férfiak adataiból a sztochasztikus egyenlet a következőnek 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 konfidencia intervallumok becslését az értekezés 2.8. fejezetében ismertetett háromféle módon végeztem el.

A reziduumok keresztkorrelációs függvényének nem volt szignifikáns értéke.

3.1.3. Következtetések a halálozási arányok elemzéseiből

A halálozási arányok elemzéseinek eredményei alapján a vizsgált betegségek két csoportba sorolhatóak. A betegségek egyik csoportjába tartozott a *cerebrovasculáris betegség, az isémiás szívbetegség, a krónikus májbetegség és a légzőszervi betegség.* Ezekre az a jellemző, hogy a két idősorra a modellek illesztése után kapott reziduumok hasonlóan viselkedtek. A reziduumok keresztkorrelációs függvényének egy szignifikáns értéke volt a $k=0$ időkésleltetésnél. Az idősorok nulla időkésleltetésnél hasonlóan viselkedtek, ezt úgy is mondhatjuk, hogy az idősorok „szinkronizáltak" voltak. A „szinkronizáció" alatt azt értjük, hogy a betegségenként képzett két idősor időben 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 betegségenként képzett mindkét idősor elemeire. Ez azt jelenti, 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 képezik *(méhnyakrák, mellrák, emésztőszervi-, légzőszervi daganatok)*. Ezekre az a jellemző, hogy a két idősorra a modellek illesztése után kapott reziduumok nem mutattak hasonlóságot. A reziduumok keresztkorrelációs függvényének nem volt szignifikáns értéke, az idősorok nem viselkedtek hasonlóan, 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 megbetegedések 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 betegségnél *(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 (légszőszervi daganatok, nemiszervek daganatai, mellrák) adatait vizsgálta idősori elemzéssel. A szerző az alkalmazott statisztikai módszerekkel 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.

3.1.3.1. A konfidencia intervallum számítások értékelése

Az AR(*1*) folyamat ϕ paraméterére vonatkozó 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 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 felhasználásával kapott intervallumokhoz viszonyítva.

Ha az AR(*1*) paraméter $\phi \approx 1$, akkor a normális eloszlás illetve a White féle közelítés nem használható, mert a konfidencia intervallum felső határa 1-nél nagyobb is lehet. Ekkor stacionárius esetben használhatatlan az eredmény. Az első két eljárással a konfidencia intervallumok felső határaként 1-nél nagyobb értékek is adódtak, melyek használhatatlanok [6]. A folytonos idejű modellből adódó becsléssel kapott konfidencia intervallumok felső határai minden esetben egy alatt maradtak. A fenti példák jól mutatják a folytonos idejű becslés előnyeit a másik két 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 határaként, melyek szintén nem elfogadhatóak [6].

Az elemzések arra is adatokat szolgáltattak, 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 alapján havonként csoportosítottam. Az elemzése során lényeges halmozódást nem figyeltem meg, így ebből semmilyen következtetés nem vonható le.

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 és a magzatot ért környezeti hatás (pl: vírusjárvány, légszennyezettség, stb) nem játszik lényeges szerepet a gyermekkori lymphoid leukaemia kialakulásában.

3.2.2. A betegség diagnosztizálási dátumának elemzése

A betegség diagnosztizálási dátumának elemzését a teljes korcsoportra és a medián alapján kettébontott adatsorokra is elvégeztem. Az összes beteg számának havi változásait tanulmányoztam. Az összes beteg számából a szezonális komponensek ábrázolásakor kilenc csúcsot kaptam, melyek közül hat a téli időszakra, egy az őszi, egy a tavaszi és egy a nyári hónapokra esett. A szezonális komponensek elemzése során azokat az emelkedéseket tekintettem csúcsnak, amikor a szezonális komponensek értéke a hatot elérte.

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 komponensek ábrázolása adekvált módszernek bizonyult 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 komponensek a téli hónapokban halmozódást mutattak. 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. A „szinkronizáció” alatt azt értjük, hogy betegségenként képzett két idősor időben azonos módon változott, növekedett vagy csökkent. 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 betegségeknél 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 nagy arányú halálozásokat okozó megbetegedések oktanához következtetéseket szolgáltattak.

5. A *gyermekkori lymphoid leukaemia* elemzése során arra a következtetésre jutottam, hogy a diagnosztizálás dátumának nagyobb a szerepe, mint a születés dátumának. A betegség diagnosztizálási dátumának elemzésekor azt 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 idősori 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 azért is szükségesek, mivel a határeloszlások, ha az AR(1) együttható egyhez közeli, nem minden esetben alkalmazhatók.

5. PHD THESES – TIME SERIES ANALYSIS OF STATISTICAL DATA FOR SOME DISEASES

Introduction, objectives

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.

Computers 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 tumouruous 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.

New research findings

1. My studies have proved that the analysis of time series models are suitable for the examination of the causes of human diseases (cerebrovascular disease, ischaemic heart disease, chronic hepatopathy, respiratory diseases, cervical cancer, breast cancer, tumours of the digestive organs and the respiratory organs). It is also shown that this statistical method can be efficiently used in case of diseases with a complex system of causes.
2. In the course of the analyses I have found that environmental effects have a decisive role in the development of a group of diseases (cerebrovascular disease, ischaemic heart disease, chronic hepatopathy, respiratory diseases).
3. From the analysis of the date on tumorous diseases with unknown causes I have come to the conclusion that the effects of the environment may, at the most, have an indirect role in the development and manifestation of these diseases (cervical cancer, breast cancer, tumours of the digestive organs and the respiratory organs).
4. In my studies I have endeavoured to systematically discuss all such medical data that could serve as the basis of conclusions on the causes of diseases with high mortality rates.
5. During the analysis of the number of childhood patients of lymphoid leukaemia, I have found that the date of diagnosis has a more important role than the date of birth. These show the important role of environmental effects in childhood in the manifestation of the disease.
6. The establishment of confidence intervals using the method of continuous estimation is also necessary, because the limit distributions cannot be used in all cases when the AR(1) coefficient is close to 1.

6. IDÉZETT IRODALOM

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 paraméterei becsléseinek eloszlására, *MTA SZTAKI Közlemények.* **8** Budapest, (1972), 3-34.
11. Benczúr, A.: Stacionárius Gauss Markov folyamat csillapodási paraméterének konfidencia határai, *MTA Számítástechnikai Központja Közlemények.* **6** Budapest, (1971), 1-14.
14. Box, G.E.P., Jenkins, G.M.: *Time series analysis forecasting and control.* Holden-Day, San Francisco, (1970).
25. Csáki, P.: ARMA folyamatok, in: (Idősorok analízise, Ed. Tusnády, G., Ziermann, M.) Műszaki Kiadó, Budapest, (1986), 50-84.
34. Éltető, Ö., Meszéna, Gy., Ziermann, M.: *Sztochasztikus módszerek és modellek,* Közgazdasági és Jogi Könyvkiadó, Budapest, (1982).
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.
55. Helfenstein, U.: Detecting hidden relations between time series of mortality rates, *Methods Inf Med.* **29** (1990), 57-60.
60. Hulyák, K.: *Idősorok sztochasztikus modelljei,* KSH Ökonometriai Laboratórium, Budapest, (1976).
110. SPSS – "Statistical Package for Social Sciences" version 9.0; http://www.spss.com.
126. White, J.S.: A t-test for the serial correlation coefficient, *Ann. Math. Stat.* **28** (1957), 1046-1048.

7. AZ ÉRTEKEZÉSHEZ KAPCSOLÓDÓ PUBLIKÁCIÓK

1. Fazekas, M.: Special Time series models for analysing 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 Patiens. 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, Czeh 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) (megjelenés alatt)

8. DISSZERTÁCIÓ TÉMAKÖRÉN KÍVÜLI PUBLIKÁCIÓK

8.1. 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).

8.2. Idegen nyelvű közlemények

18. Furka, I., Mikó, I., Tóth, K., Furka, A., Kappelmayer, J., Szikszai, Z., Góth, L., Kovács, J., Baráth, E., Imre, S., Fazekas, M.: Haematological, hemorheological and catalase level following spelenectomy 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 B_1 and other fusariotoxins in Hungary, *Acta Veterinaria Hungarica.* **44** (1996), 25-37.

8.3. 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.

8.4. 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, (1994), 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), 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), 390-400.
28. Fazekasné Kis M.: A ROC analízis oktatása az orvosdiagnosztikai laboratóriumi analitikus szakon. *Informatika a felsőoktatásban '99.* Debrecen, (1999), 228-233.

8.5. Poszter

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

8.6. Egyéb

30. Fazekasné Kis M., Kériné Fülöp I., Virágos Cs.: *Reference Manager 7.01 verzió, Ismertetés a program működéséről.* Szerkesztő: Pethő Attila. Kézikönyv, Debrecen, (1997).
31. Fazekas B., Kis M., Konczné Tar A., Tóthné Hajdú E.: Fumonisin B_1 tartalmú fusarium moniliforme gombatenyészet etetésének hatása napos korú kacsákra, *MTA Állatorvos-Tudományi Bizottsága, Akadémiai Beszámoló*. Budapest, (1999).