Eötvös Loránd Tudományegyetem

Természettudományi Kar

Fodor Péter

Halandósági táblák előrejelzése

BSc Szakdolgozat

Témavezető:

Zempléni András

Valószínűségelméleti és Statisztika Tanszék

Budapest, 2016

Köszönetnyilvánítás

Szeretném megköszönni Zempléni Andrásnak, hogy elvállalta a témavezetői feladatkört, valamint, hogy rendszeres konzultációkkal segítette szakdolgozatom elkészítését. Továbbá köszönettel tartozom feleségemnek és szüleimnek a támogatásukért.

Tartalomjegyzék

Bevezetés

Szakdolgozatom témájának kiválasztásakor arra törekedtem, hogy olyan témát válasszak, ami a biztosításmatematikának gyakorlati alkalmazása. Így esett a választásom a halandósági táblák előrejelzésére, ami biztosítótársaságok gyakori feladata.

Szakdolgozatom vezérvonala [1] cikk volt, és a fő cél a cikkben ismertetett módszerek gyakorlati tesztelése volt, amire az azóta eltelt idő adott lehetőséget.

Az első fejezetben bemutatjuk, hogy a halandósági táblák milyen adatokat tartalmaznak és megmutatjuk hogy a megfigyelt adatokból hogyan konstruálhatjuk meg. Majd ismertetünk simítási módszereket amiket nyers adatok kiegyenlítésére használunk.

A második fejezetben részletezünk néhány statisztikát amelyekkel összehasonlíthatunk halandósági táblákat. Továbbá összehasonlítjuk a magyar halandósági táblákat múltbeli USA táblákkal és leellenőrizzük a 10 évvel ezelőtti eredmény használhatóságát.

A harmadik fejezetben bemutatjuk a Lee-Carter modellt, mortalitás előrejelzésére. Majd előrejelzést készítünk magyar férfi és női adatokból. Az előrejelzések viszonyáról kapunk képet a két módszer összehasonlítása révén.

Az eredményeim megbízhatóságának növelése érdekében a számításokat és az ábrákat R programkörnyezetben készítettem.

1. fejezet

Halandósági táblák

1.1. Halandósági táblák konstruálása

A biztosításmatematika egyik legfontosabb és legnehezebb feladata az, hogy meghatározza egy konkrét biztosítás esetén a biztosítási díjat. Az életbiztosításoknál a díjkalkuláció alapjául a halandósági táblák szolgálnak. A halandósági táblák megadják,hogy 100000 db élve született csecsemőből előreláthatólag hány éli meg az 1, 2, 3... életkort. A táblázat egy ω életévvel ér véget, ami után feltételezzük, hogy mindenki meghal.

A táblázat tartalmazza az életkort, ezt jelöljük x-szel, az x életkort elérő emberek számát ezt jelöljük l_x-el, továbbá azt, hogy egy évben az x-edik és az $x+1$-edik életéveik között hány ember halálozott el, ez jelöli $d_x = l_x - l_{x+1}$.

Megmutatjuk, hogy egy férfi halandósági táblát hogyan konstruálhatunk meg nyers adatokból (Ezek az adatok fellelhetők a KSH kiadványaiban minden évben). Tudjuk, hogy egy adott évben hány x éves férfi volt, ezt jelöljük P_x-el. Emellett azt is, hogy hány olyan x éves férfi halt meg, aki x-edik születésnapját megérte, jelöljük $\overline{D}'_x$-el, hasonlóan $\overline{D}''_x$-el jelöljük azoknak az x éveseknek a számát akik elhaláloztak az x. születésnapjuk előtt. A szakirodalomban elterjedt Böck-féle módszer a következő értékeken alapul:

$$p_x^{(1)} = \frac{P_x - \overline{D}''_x}{P_x}, \qquad p_x^{(2)} = \frac{P_{x-1} - \overline{D}''_{x-1} - \overline{D}'_{x-1}}{P_{x-1} - \overline{D}'_{x-1}}.$$

Ahol $p_x^{(1)}$ annak a valószínűsége, hogy az x évesek betöltik az x. születésnapjukat, $p_x^{(2)}$ annak a valószínűsége, hogy azok az $x-1$ évesek, akik betöltötték az $x-1$. születésnapjukat megérik az x. évüket. Legyen $p_x = p_x^{(1)} p_x^{(2)}$ túlélési valószínűség, azaz p_x annak a valószínűsége, hogy egy $x-1$ éves férfi legalább egy évet él még. Ezekből a halandósági tábla x-edik évre vonatkoztatott nyers halálozási valószínűségei a következő módon származtathatóak:

$$q_x = 1 - p_x^{(1)} p_x^{(2)}.$$

A q_x annak a valószínűsége, hogy egy $x-1$ éves férfi meghal egy éven belül. Előfordul, hogy nagyok a különbségek a közel azonos korúak valószínűségei között és néhol olyan is, hogy a fiatalabbak nagyobb valószínűséggel halnak meg. Ez lehet véletlen vagy az adatok hibás felvétele miatt. Ekkor a nyers valószínűségeket célszerű simítani. A hibák kiegyenlítésére interpolációs és extrapolációs módszereket használunk az alábbiak szerint(vázlatosan):

1. 0-3 éves életkorokban nincs kiegyenlítés
2. 4-15 éves életkorokban analitikus kiegyenlítés: negyedfokú polinommal, Jordan féle ortogonális polinomok alkalmazásával.
3. 15-75 éves életkorokban mechanikus módszerrel, ami egy Karup által módosított Newton-interpoláció.
4. 75 éves kor fölött pedig Gompertz-Makeham függvény illesztéssel, King-Hardy féle csoportképzési eljárással.

Ezen simítási módszerek segítségével a nyers valószínűségekből megkapjuk a $\overline{q}_x$ kiegyenlített halálozási valószínűségeket. Egy halandósági tábla a nyers valószínűségeket csak 84 éves korig tartalmazza, 84 év fölött extrapolációval történik a Gompertz eloszlás illesztése. Ezek az életkorok annyira nem fontosak, mert ritkán fordulnak elő a biztosítási gyakorlatban.

Ezután kiindulásként vegyünk $l_0 = 100000$ darab fiktív élveszületett csecsemőt. Az l_x halandósági függvény értékeit a következőképp számítjuk ki: $l_x = (1 - \overline{q}_x) l_{x-1}$ $x = 1, 2, \ldots, \omega$ (ezeket egészre kerekítjük). A tábla d_x oszlopa a következőképpen határozható meg : $d_x = l_x - l_{x-1}$. Ezekből az adatokból több, a biztosítóknak fontos valószínűséget számíthatunk ki:

- Egy x éves férfi még minimum k évig fog élni:

$$ {}_kp_x = l_{x+k}/l_x $$

- Egy x éves férfi az $x+k$-edik és az $x+k+1$-edik életévei között halálozik el:

$$ {}_{k|1}q_x = d_{x+k}/l_x $$

- Egy x éves férfi meghal k éven belül:

$$ {}_kq_x = 1 - {}_kp_x = (l_x - l_{x+k})/l_x $$

A tábla tartalmaz még további két adatot: az x éves férfiak várható hátralévő élettartamát és az élettartamainak szórását. A várható hátralevő élettartamot a következőképp számítjuk ki:
Egy $y_1, \dots, y_m$ adatsor statisztikai várható értéke az adatok átlaga, amit E-vel jelölünk:

$$ E = \frac{y_1 + y_2 + \cdots + y_m}{m}. $$

Például újszülött csecsemők várható élettartamát úgy számoljuk ki, hogy 100000 élve született csecsemő életét végigkísérjük és feljegyezzük, hogy hány évet éltek(ez legyen $y_1 + y_2 + \cdots + y_{10000}$) és ezeknek vesszük az átlagát:

$$ \frac{y_1 + y_2 + \cdots + y_{100000}}{100000}. $$

Hasonlóan számítjuk ki egy x éves férfi várható élettartamát. x éves férfiből l_x van összesen, akik $y_1 + y_2 + \cdots + y_{l_x}$ évesen halnak meg, vagyis várható életkoruk az $y_1 + y_2 + \cdots + y_{l_x}$ évek átlaga. Az átlagban $x+1$, $x+2$, ..., $\omega - 1$, ω, $\omega + 1$ évek szerepelnek(ha egy férfi i és $i+1$ éves kora között hal meg, akkor $i+1$-et veszünk). Ekkor $x+1$ d_x-szer,$x+2$ d_{x+1}-szer,..., ω $d_{\omega-1}$-szer és az $\omega+1$ d_ω-szer szerepel, vagyis a várható élettartam becslése:

$$ E_x = \frac{y_1 + y_2 + \cdots + y_{100000}}{100000} = \frac{1}{l_x}\Big((x+1)d_x + (x+2)d_{x+1} + \cdots + \omega d_{\omega-1} + (\omega+1)d_\omega\Big). $$

Mivel $d_x = l_x - l_{x+1}$(ha $x < \omega$) és $d_\omega = l_\omega$, ezért

$$ E_x = x + \frac{l_x + l_{x+1} + \cdots + l_{\omega-1} + l_\omega}{l_x}. $$

Így egy x éves férfi várható hátralevő élettartam becslése $\frac{l_x + l_{x+1} + \cdots + l_{\omega-1} + l_\omega}{l_x}$. Mivel a számításoknál az elhalálozási életkort mindig felfelé kerekítettük, pontosabb képletet kapunk, ha csökkentjük 1/2 évvel. Vagyis a hátralevő várható élettartam:

$$e_x^0 = \frac{l_x + l_{x+1} + \cdots + l_{\omega-1} + l_\omega}{l_x} - \tfrac{1}{2}.$$

A statisztikai szórás megmutatja, hogy $y_1, \ldots, y_m$ számok átlagosan mennyivel térnek el az $E = \dfrac{y_1 + y_2 + \cdots + y_m}{m}$ várható értéktől, ezt jelöljük σ-val:

$$\sigma = \sqrt{\tfrac{1}{m}\sum_{i=1}^{m}(y_i - E)^2}$$

Bár $m-1$-gyel osztva kapunk torzítatlan becslést a szórásnégyzetre a nagy m érték és a kényelmesebb számolási mód miatt a szakirodalom így definiálja. Ebben a képletben az $(y_i - E)^2$ az adatok négyzetes eltérése az E átlagtól, vagyis $\tfrac{1}{m}\sum_{i=1}^{m}(y_i - E)^2$ az adatok átlagos négyzetes eltérése E-től, ez a szórásnégyzet, aminek a négyzetgyöke a szórás. A szórásnégyzetet így egyszerűsíthetjük:

$$\begin{aligned}
\sigma^2 = \tfrac{1}{m}\sum_{i=1}^{m}(y_i - E)^2 &= \frac{1}{m}\sum_{i=1}^{m}(y_i^2 - 2Ey_i + E^2) \\
&= \tfrac{1}{m}\sum_{i=1}^{m} y_i^2 - 2E\frac{1}{m}\sum_{i=1}^{m} y_i + \frac{1}{m}\sum_{i=1}^{m} E^2 \\
&= \tfrac{1}{m}\sum_{i=1}^{m} y_i^2 - 2E^2 + E^2 = \frac{1}{m}\sum_{i=1}^{m} y_i^2 - E^2.
\end{aligned}$$

Legyen $\mu^2 = \tfrac{1}{m}\sum_{i=1}^{m} y_i^2$, az adatok négyzetes átlaga. Vagyis a szórás rövidebb alakja: $\sigma = \sqrt{\mu - E^2}$.

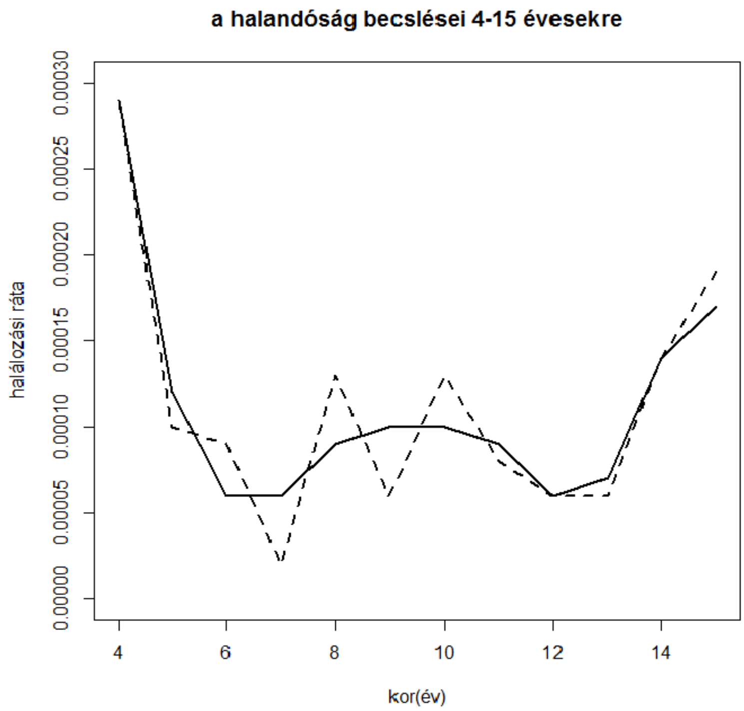
Ezek alapján kiszámoljuk l_x darab x éves férfi életkorának négyzetes átlagát a következőképp:

$$\begin{aligned}
\mu_x &= \frac{y_1^2 + \cdots + y_{l_x}^2}{l_x} \\
&= \frac{1}{l_x}\Big((x+1)^2 d_x + (x+2)^2 d_{x+1} + \cdots + \omega^2 d_{\omega-1} + (\omega+1)^2 d_\omega\Big)
\end{aligned}$$

Mivel $d_x = l_x - l_{x+1}$(ha $x < \omega$) és $d_\omega = l_\omega$, ezért

$$\mu_x = x^2 + \frac{1}{l_x}\Big((2x+1)l_x + (2x+3)l_{x+1} + \cdots + (2\omega-1)l_{\omega-1} + (2\omega+1)l_\omega\Big).$$

Vagyis $\sigma_x = \sqrt{\mu_x - E_x^2}$ érték megadja az x éves férfiak élettartamainak szórását.



1.1. ábra. 2011-es magyar férfi nyers(szaggatott) és kiegyenlített(folytonos) halálozás

1.2. Simítási módszerek

Az előző alfejezetben említettük, hogy előfordul a nyers halálozási valószínűségeknél, hogy nagy különbségek vannak a közel azonos korúak között, sőt olyan is előfordulhat, hogy az öregebbek kisebb valószínűséggel halnak meg. Ennek több oka lehet: véletlen, hibás adatfelvétel vagy bizonyos korcsoportokra ez jellemző (például a csecsemőknél nagyobb az érték, mivel nekik még gyengébb az immunrendszerük). Ezért célszerű a nyers valószínűségeket simítani.

A 0-3 év közötti korcsoportra nem alkalmazunk kiegyenlítést. A nagy és gyors változás miatt minden módszer torzítana és a valóságot teljesen elferdítené.

A 4-15 éves korcsoportot analitikus kiegyenlítéssel, negyedfokú polinommal, Jordan féle ortogonális polinomok alkalmazásával simítjuk az alábbiak szerint. Feltételezzük, hogy az észlelt függvény menetét a következő alakú

$$\overline{q}_x = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$$

negyed fokú parabola ívvel tudjuk leírni, akkor a probléma az a_0, a_1, a_2, a_3, a_4 paraméterek meghatározására redukálódik.

A parabola együtthatóit a legkisebb négyzetek módszerével határozhatjuk meg, ami megköveteli, hogy a nyers és a kiegyenlített értékek eltéréseinek négyzetösszege minimális legyen, vagyis:

$$S = \sum_{x=4}^{15} (\overline{q}_x - q_x)^2 = \text{minimum}.$$

A minimumhoz szükséges feltételek a következők:

$$\frac{\partial S}{\partial a_i} = 0 \qquad i = 0, 1, 2, 3, 4.$$

Az így nyert 5 egyenlet meghatározza a_0, a_1, a_2, a_3, a_4 ismeretleneket. Ennek a korosztálynak a kiegyenlítését az (1.1)-es ábrán láthatjuk.

A 15-75 évek értékeit két ütemben végzett mechanikus eljárással egyenlítjük ki. Első lépésben 5 éves egyenlő távolságú korokra úgynevezett főpontokat határozunk meg a nyers adatokból. Második lépésben a főpontok közötti 4 értéket a Karup-féle eljárással állapítjuk meg.

A főpontokat Newton-interpolációs képletével számítjuk ki az alábbiak szerint:

$$\overline{q}_{x+5} = Z_x + \Delta Z_x - 0,04\Delta^2 Z_x \tag{1.1}$$

$$\text{ahol } Z_x = \frac{q_{x-2} + q_{x-1} + q_x + q_{x+1} + q_{x+2}}{5}$$

$$\Delta Z_x = Z_{x+5} - Z_x$$

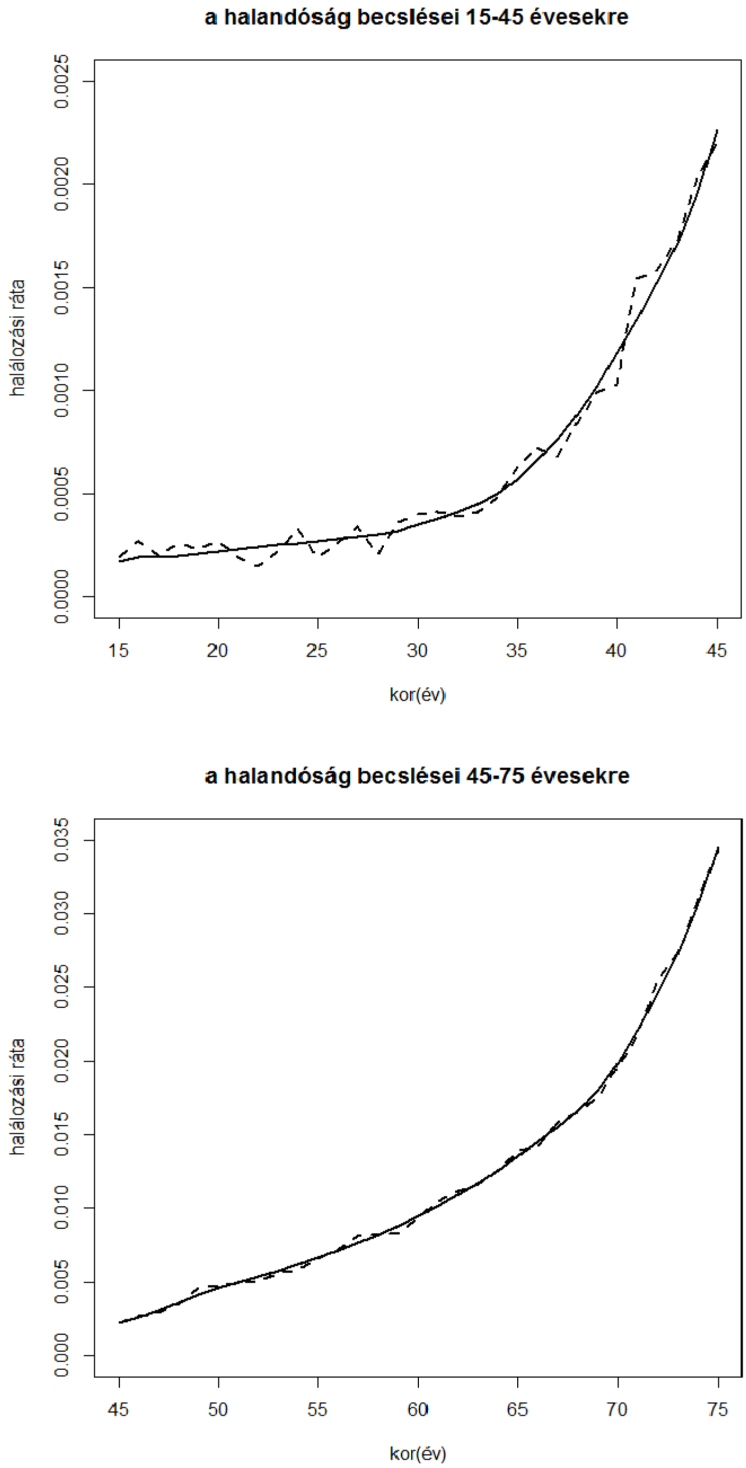
$$\Delta^2 Z_x = \Delta Z_{x+5} - \Delta Z_x$$

Ha az 1.1-es képletben a különbségeket feloldjuk és $x + 5$ életkorról áttérünk x-re, akkor megkapjuk a gyakorlatban használt képletet:

$$\overline{q}_x = 1,08Z_x - 0,04(Z_{x-5} + Z_{x+5}) \text{ ahol } x = 15, 20, 25, \dots, 75.$$

A közbeeső értékek kiszámítására a következő Karup-féle formulák szolgálnak:

$$\overline{q}_{x+1} =$$
$$0,00256Z_{x-10} - 0,1056Z_{x-5} + 0,9808Z_x + 0,1456Z_{x+5} - 0,024Z_{x+10} + 0,00064Z_{x+15}$$
$$\overline{q}_{x+2} =$$
$$0,00288Z_{x-10} - 0,1056Z_{x-5} + 0,7376Z_x + 0,432Z_{x+5} - 0,0688Z_{x+10} + 0,00192Z_{x+15}$$
$$\overline{q}_{x+3} =$$
$$0,00192Z_{x-10} - 0,0688Z_{x-5} + 0,432Z_x + 0,7376Z_{x+5} - 0,1056Z_{x+10} + 0,00288Z_{x+15}$$
$$\overline{q}_{x+4} =$$
$$0,00064Z_{x-10} - 0,024Z_{x-5} + 0,1456Z_x + 0,9808Z_{x+5} - 0,1056Z_{x+10} + 0,00256Z_{x+15}$$



1.2. ábra. 2011-es magyar férfi nyers(szaggatott) és kiegyenlített(folytonos) halálozás

Ennek a korosztálynak a kiegyenlítését az (1.2)-es ábrán láthatjuk.

Exponenciális függvénnyel a halandósági táblát csak olyan intervallumban lehet kiegyenlíteni, ahol a feltételezés szerint a halálozási intenzitás monoton növekvően változik a korral. Ez a hipotézis 75 év felett elfogadható. Így 75 év felett Gompertz-Makeham függvény illesztésével simítunk.
Feltételezésünk szerint l_x kifejezhető az x életkor függvényében a következőképpen:

$$l_x = kS^x e^{c^x}.$$

A paraméterek meghatározása céljából az előbbi képlettel kifejezzük $\overline{p}_x$ továbbélési valószínűséget:

$$l_x = kS^x e^{c^x} \; ; \; l_{x+1} = kS^{x+1} e^{c^{x+1}}$$

$$\overline{p}_x = \frac{l_{x+1}}{l_x} = Se^{(c-1)c^x} \tag{1.2}$$

A gyakorlati számításokat megkönnyítve logaritmáljuk az 1.2 egyenletet:

$$\log \overline{p}_x = \log S + (c-1)c^x$$

Legyen $a = \log S$, $b = c - 1$ vagyis:

$$\log \overline{p}_x = a + bc^x \tag{1.3}$$

Az a, b és c paraméterek meghatározására az úgynevezett King-Hardy csoportképzési eljárást és a legkisebb négyzetek elvén alapuló eljárást alkalmazzuk.
A számításoknál használt képletek a következők :

$$H_1 = \sum_{x=75}^{79} \log p_x \; ; \; H_2 = \sum_{x=80}^{84} \log p_x \; ; \; H_3 = \sum_{x=85}^{89} \log p_x$$

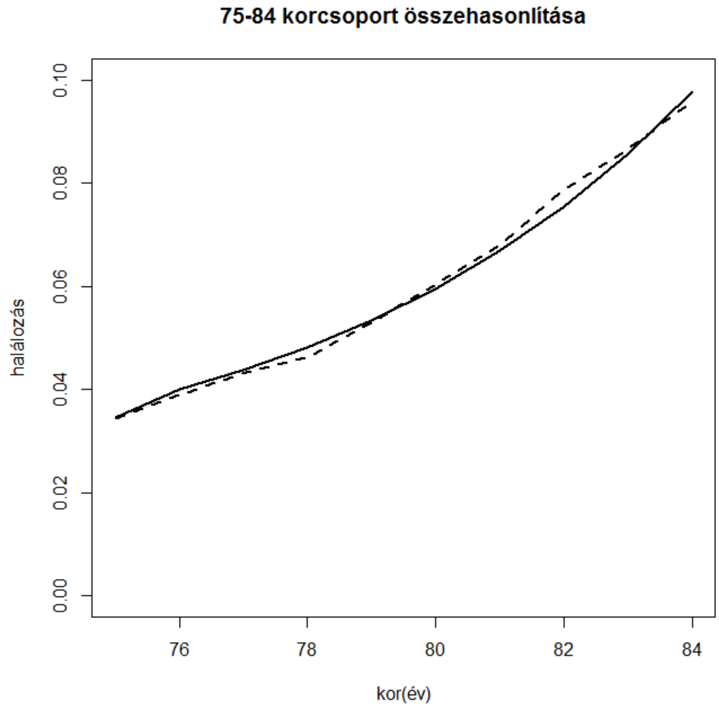
$$c^5 = \frac{H_3 - H_2}{H_2 - H_1}.$$

Ebből már c kiszámítható.

$$A = H_1 + H_2 + H_3$$

$$B = \sum_{x=75}^{89} c^x \; ; \; C = \sum_{x=75}^{89} c^{2x} \; ; \; D = \sum_{x=75}^{89} c^x \log p_x$$

$$a = \frac{AC - BD}{15C - B^2} \; ; \; b = \frac{15D - AB}{15C - B^2}$$



1.3. ábra. 2011-es magyar férfi nyers(szaggatott) és kiegyenlített(folytonos) halálozás

Ezekből megkapjuk a-t és b-t. Ismerve az összes paramétert az 1.3 képlet alapján megkapjuk a kiegyenlített továbbélési valószínűségek logaritmusait ($\log \overline{p}_x$), amiből már könnyen meghatározhatóak $\overline{p}_x$-ek, amiből meghatározható a kiegyenlített halálozási valószínűségek $\overline{q}_x = 1 - \overline{p}_x$.
A 1.3 képlet a 75-89 évek nyers adatainak kiegyenlítése mellett még arra is lehetőséget ad, hogy a további évek elhalálozási valószínűségeit extrapoláljuk és így a halandósági táblákat 100 éves korig bővíthetjük. Ennek a korosztálynak a kiegyenlítését az (1.3) ábrán láthatjuk.

2. fejezet

Halandósági táblák összehasonlítása

2.1. Hasonlósági mérőszámok

A fő kérdés az az, hogy hogyan mérjük a hasonlóságot a halandósági táblák között. Ebben a fejezetben megvizsgáljuk a hasonlósági mérőszámokat, amelyekkel jellemezhetjük a halandósági táblákat. A Society of Actuaries főoldalán található számos halandósági tábla, ezeken kívül a Központi Statisztikai Hivatal halálozási tábláit használjuk forrásként.

Jelöljük a közelíteni kívánt táblát q_1-el és elemeit q_{i1}-el. Az első statisztika, melyet részletezünk a súlyozott négyzetes szórás $(QDEV)$:

$$QDEV = \sum_{i=K}^{N} \frac{P_i(q_{i1} - q_{i0})^2}{q_{i0}}, \tag{2.1}$$

ahol P_i az i évesek számát jelöli, q_0 pedig az a választott tábla, amit a közelítésre jelöltünk ki (ennek elemeit q_{i0}-val jelöljük). K és N pedig az első és utolsó év amiket a rendelkezésre álló adatokhoz és a probléma megoldásához alkalmasan tudunk választani.

Ennek a statisztikának a határeloszlása khi-négyzet, vagyis egy statisztikai teszt a két tábla egyenlőségére az $N - K + 1$ szabadságfokú khi-négyzet eloszlás kritikus értékeire alapozhat. Habár, ezek az értékek a gyakorlatban a halandósági tábláknak mindig statisztikailag szignifikánsak. De ez nem egészen ugyanaz, mint a gyakorlatban is lényeges eltérés, ami azt jelentené, hogy a két tábla nem felcserélhető gyakorlati számításokhoz. A következő statisztikák alkalmasabbak a gyakorlatban is lényeges eltérés kimutatására.

A következő két statisztika megtalálható Mitchell és McCarthy cikkében [6]. Az első alternatíva az úgynevezett A/E statisztika, a következőképp megadva:

$$A/E = 100 \cdot \frac{\sum_{i=K}^{N} l_{i0} q_{i1}}{\sum_{i=K}^{N} l_{i0} q_{i0}}, \tag{2.2}$$

ahol l_{i0} annak a valószínűsége, hogy az i. évben még él a közelítő tábla szerint, azaz $l_{i+1,0} = l_{i0}(1 - q_{i0})$ és $l_{K0} = 1$. Azaz a tagok a nevezőben arányosak a népességben az adott életkorra becsült halálozások számával, míg a számláló hasonló mennyiséget ad, ami a közelítő tábla népesség eloszlásán és a vizsgált tábla által meghatározott kockázatokon alapszik. Úgy is értelmezhető mint egy Laspeyres index, ahol a két adatsor (q_{i1}, q_{i0}), a súlyokat(l_{i0}) az alap táblából véve.
A várható hátralevő élettartamot a következőképpen definiálhatjuk:

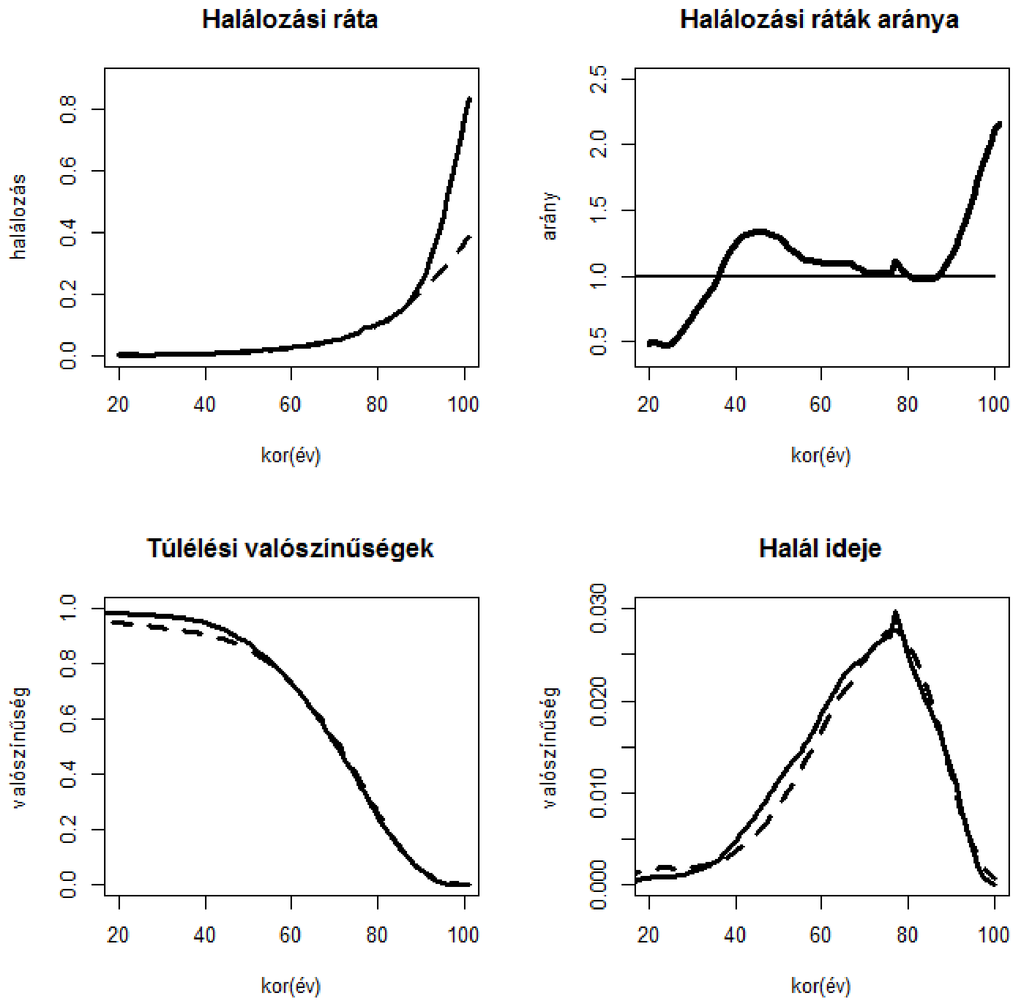
$$ERL = 100 \cdot \frac{\sum_{i=K}^{N} l_{i1} - 0.5}{\sum_{i=K}^{N} l_{i0} - 0.5}, \tag{2.3}$$

ahol $l_{K0} = l_{K1} = 1$. (2.3)-as formula megadja a várható hátralevő élettartamok arányát(K. évvel kezdve az N.-évig) a két tábla között.A 100-as szorzó miatt százalékban kapjuk meg a hányadost. Annak ellenére, hogy a fő érdekeltség (különösen a járulékok) az idősebb korosztály halandóságában rejlik, az együtthatókat $K = 20$,$K = 30$,$N = 60$ és $N = 70$ választással használjuk. Egy tipikus biztosító társaságnak rendelkezésére állnak az adatok ebben az intervallumban.

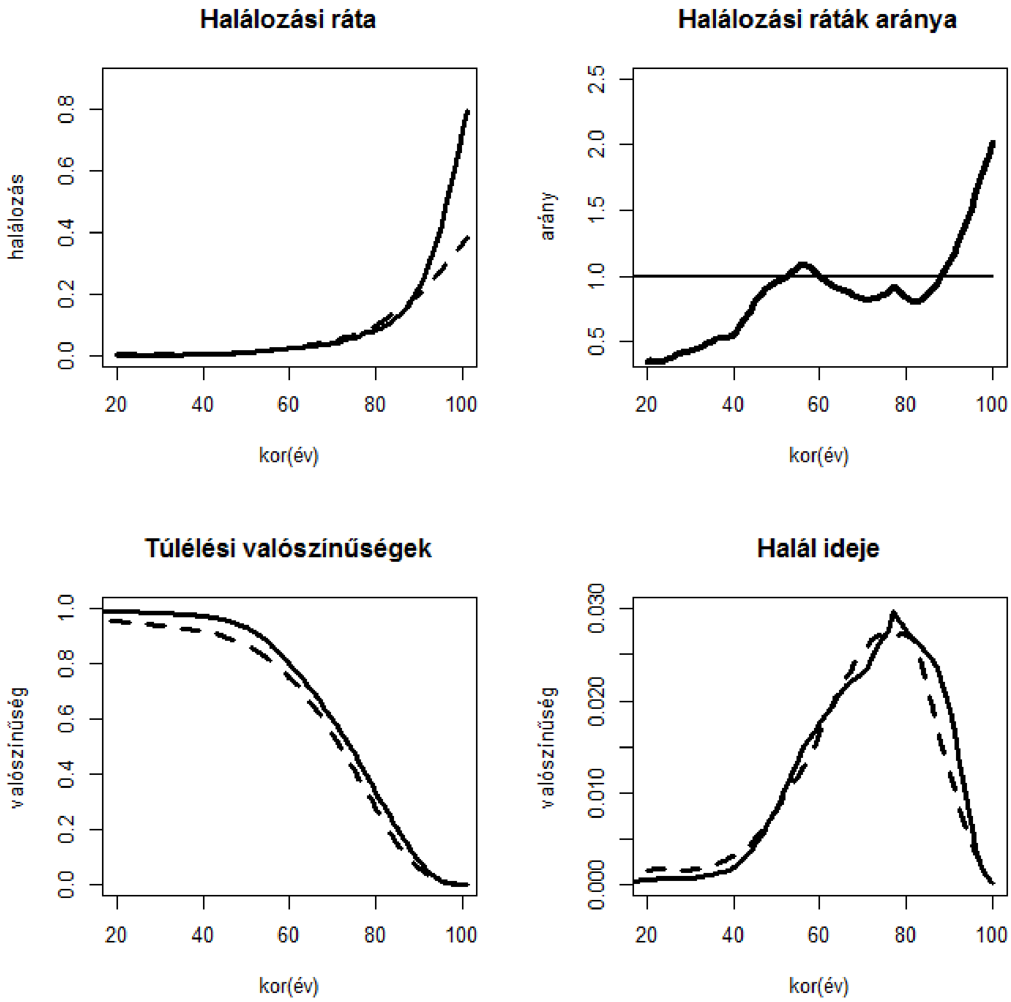
2.2. Alkalmazások

Ebben a fejezetben a fenti módszerekre mutatunk alkalmazást. Aktuális magyar halandósági táblákat hasonlítunk össze múltbeli USA halandósági táblákkal. Itt az illeszkedés általában nem olyan jó a biztosítótársaságok járandósági tábláihoz, de mi még megfelelő illeszkedést kapunk.

Számos táblát kipróbáltunk mint lehetséges közelítés, de az eredmények alapján(és a [1] cikk szerint is) ezeket találtuk a legjobbnak.



2.1. ábra. Magyar férfi halandóság (2000.év,folytonos) és az USA férfi halandóság (1950.év,szaggatott) összehasonlítása



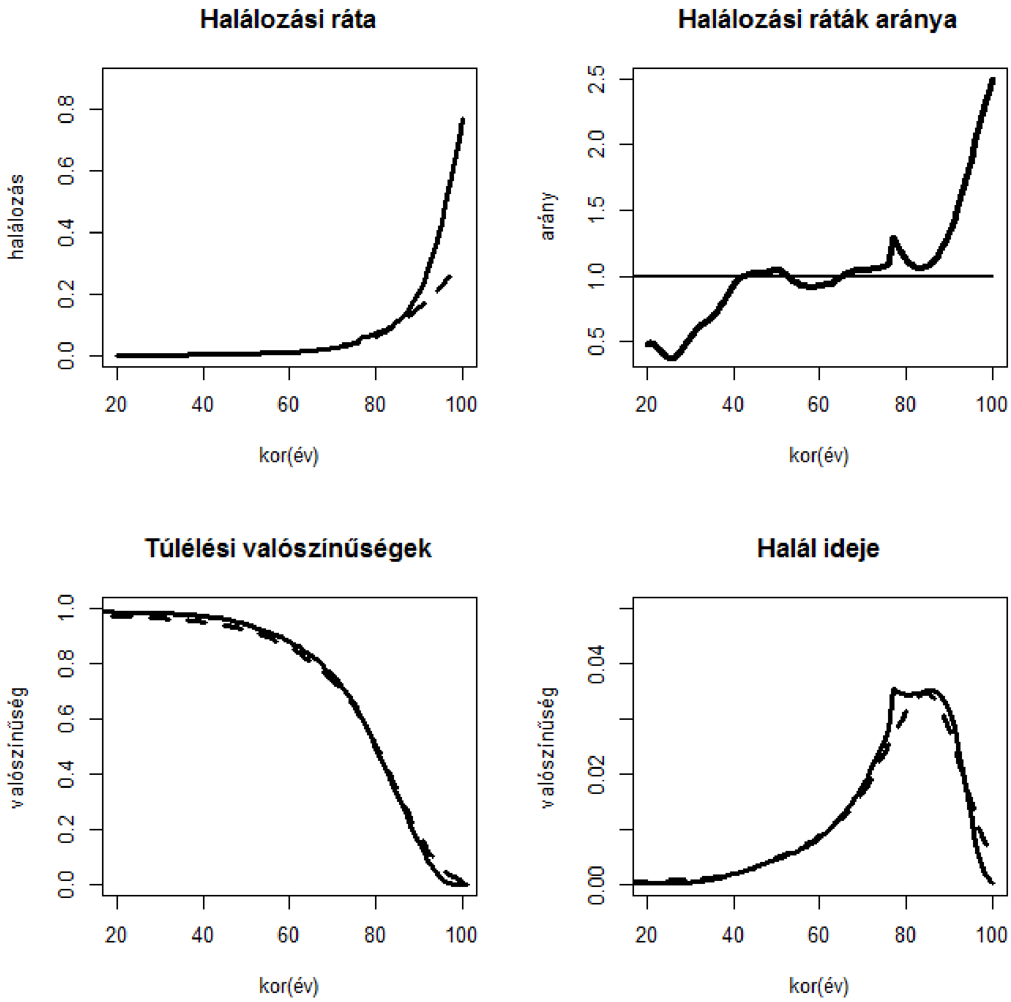
2.2. ábra. Magyar férfi halandóság (2010.év,folytonos) és az USA férfi halandóság (1960.év,szaggatott) összehasonlítása

Az 2.1-es ábrán láthatjuk a 2000. évi magyar férfi halandósági ráta és az 1950. évi USA férfi halandósági ráta összehasonlítását. Különösen a jobb felső ábrából látszik(ahol a referencia az USA-beli halandóság), hogy a magyar adatok kisebb halálozást mutatnak 20 és 40 év között, míg 40 és 60 között és 90 fölött nagyobb halálozást mutatnak. Viszont, 60 és 90 év között az illeszkedés nagyon jó (és például a nyugdíjszámítás esetén ezek a legfontosabb korok). A 2.2-es ábrán hasonló összehasonlítást láthatunk mint a 2.1-esen csak a magyar adatokat 2010. évből az USA adatokat pedig 1960. évből vettük. Látszik, hogy itt már kevésbé jó az illeszkedés, mivel a magyar halandóság nagyobb mértékben csökkent. Itt a magyar halandóság kisebb 45 év alatt, 90 év fölött nagyobb, míg 45-90 év közötti korosztálynál kapunk jó illeszkedést.

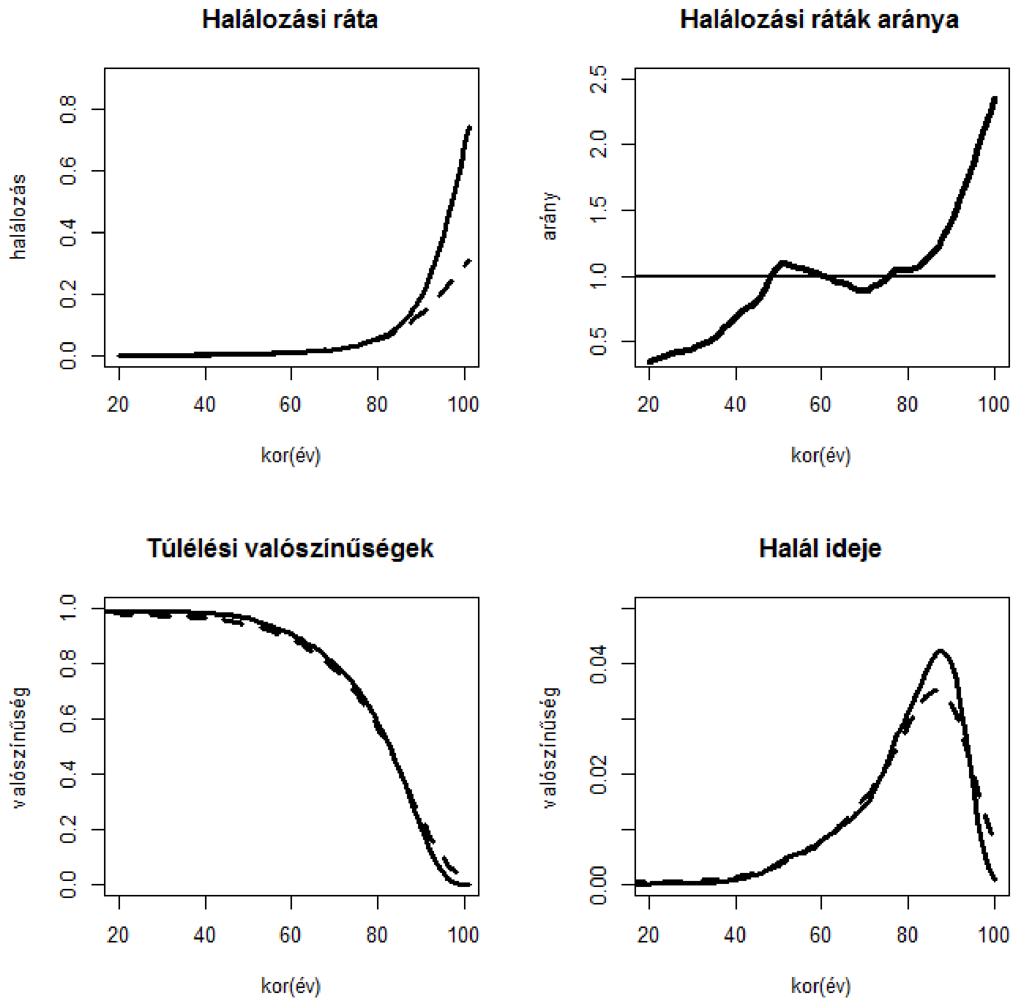
A 2.3-es ábrán összehasonlítottuk 2000. évi magyar női halandósági rátát és az 1970. évi USA női halandósági rátával. Ezen az ábrán is a halálozási ráták arányából láthatjuk, hogy 20 és 40 év között a magyar halálozási ráta kisebb, míg 90 év fölött nagyobb halálozást mutat. Emellett 40 és 90 év között jó illeszkedést kapunk, kivéve a 76-78 éveseknél, ott a magyar halandósági ráta kicsit megugrik, amit szintén láthatunk a jobb alsó ábrából. A 2.4-es ábrán a 10 évvel későbbi összehasonlítást láthatjuk. Mindkét halandósági ráta csökkent közel azonos mértékben, kivéve 40-50 év között, ott a magyar adatok jobban csökkentek. Vagyis itt csak az 50-90 közötti korosztálynál kapunk jó illeszkedést.

Az alábbi táblázatban láthatjuk a statisztikák értékeit, magyar táblákat közelítve USA táblákkal

Magyar tábla	USA tábla	QDEV(30,70)	A/E(30,70)	ERL(30,70)	ERL(20,60)
férfi,2000	férfi,1950	964,7867	112,9041	102,0939	100,0717
férfi,2010	férfi,1960	947,9737	90,46882	98,29569	97,98171
női,2000	női,1970	127,5261	97,59053	99,6666	99,37032
női,2010	női,1980	282,8437	94,06687	99,51969	99,30953



2.3. ábra. Magyar női halandóság (2000.év,folytonos) és az USA női halandóság (1970.év,szaggatott) összehasonlítása



2.4. ábra. Magyar női halandóság (2010.év,folytonos) és az USA női halandóság (1980.év,szaggatott) összehasonlítása

3. fejezet

Modellezés és szimuláció

Ebben a fejezetben az a célunk, hogy úgy modellezzük a halandóságot, hogy a modell alapján előrejelzést is tudjunk készíteni. Ezt az előrejelzést össze tudjuk hasonlítani az előző fejezetben látott, analógián alapuló módszer eredményeivel. Ilyen módszer a Lee-Carter modell, melyet a következő alfejezetben bemutatunk.

3.1. Lee-Carter modell

A módszert Ronald D.Lee és Lawrence R.Carter publikálta 1992-ben. Amerikai mortalitási rátákat modelleztek, amely során egy extrapolatív módszert találtak ki és alkalmaztak. A modellt korévenként kettő (α_x, β_x) és a megfigyelt időpontonként (κ_t) egy paraméter írja le. Lee és Carter modelljét a következő formula adja meg:

$$q_{x,t} = e^{\alpha_x + \beta_x \kappa_t + \epsilon_{x,t}}.$$

Amelyben $q_{x,t}$ egy x éves ember t. időszakra vonatkozó halálozási rátáját adja meg, α_x a mortalitási ráta alapértéke, β_x és κ_t az időtől való függés erősségét, $\epsilon_{x,t}$ pedig a hibatag.
Az egyenletet logaritmáljuk és kapjuk a következő egyenletet:

$$\log q_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t} \tag{3.1}$$

A Lee-Carter modell a következő 3 lépésből áll:

1. A modell paramétereit becsüljük

2. A modellezett halálozások számát a megfigyelt halálozások számához igazítjuk

3. Előrejelzést készítünk a mortalitási rátákra vonatkozóan.

A Lee-Carter modell túlparaméterezett abban az értelemben, hogy invariáns az alábbi transzformációkra:

$$\{\alpha_x, \beta_x, \kappa_t\} \to \{\alpha_x, \beta_x/c, \kappa_t \cdot c\}$$

$$\{\alpha_x, \beta_x, \kappa_t\} \to \{\alpha_x - c \cdot \beta_x, \beta_x, \kappa_t + c\}$$

Ezért Lee és Carter a következő feltételeket vezette be:

$$\sum_{x=0}^{N} \beta_x^2 = 1, \tag{3.2}$$

$$\sum_{t=1}^{T} \kappa_t = 0, \tag{3.3}$$

ahol T jelöli az évek számát, N pedig a legmagasabb életkor, ami a megfigyelésben szerepel. A 3.3 és 3.2 azért szükségesek, hogy a paraméterek egyértelműen meghatározhatóak legyenek.

3.1.1. Paraméterek becslése

Elsőnek becsüljük α_x paramétert. Az alábbi függvényt minimalizáljuk 3.2 és 3.3 feltételek mellett:

$$L(\alpha, \beta, \kappa) = \sum_{t,x} (\log(q_{x,t}) - \alpha_x - \beta_x \kappa_t)^2 \tag{3.4}$$

Ekkor $\dfrac{\partial L}{\partial \alpha_x} = \dfrac{\partial L}{\partial \beta_x} = \dfrac{\partial L}{\partial \kappa_t} = 0$ feltételek miatt a következő egyenleteket kapjuk:

$$\frac{\partial L}{\partial \alpha_x} = -2 \cdot \sum_t (\log(q_{x,t}) - \alpha_x - \beta_x \kappa_t) = 0 \tag{3.5}$$

$$\frac{\partial L}{\partial \beta_x} = -2 \cdot \sum_t (\log(q_{x,t}) - \alpha_x - \beta_x \kappa_t) \cdot \kappa_t = 0 \tag{3.6}$$

$$\frac{\partial L}{\partial \kappa_t} = -2 \cdot \sum_x (\log(q_{x,t}) - \alpha_x - \beta_x \kappa_t) \cdot \beta_x = 0 \tag{3.7}$$

A 3.5 egyenlet és 3.3 feltétel segítségével kapjuk α paraméter becslését:

$$\sum_t^T \alpha_x = \sum_t^T \log q_{x,t} - \sum_t^T \beta_x \cdot \kappa_t \Rightarrow \sum_t^T \alpha_x = \sum_t^T \log q_{x,t}$$

Ebből megkapjuk α_x paraméter becslését:

$$\widehat{\alpha}_x = \frac{1}{T}\sum_{t}^{T} \log q_{x,t}$$

Vezessük be az M mátrixot melyet a következő egyenlettel definiálunk:

$$M_{x,t} = \log q_{x,t} - \widehat{\alpha}_x = \beta_x \cdot \kappa_t$$

Az M mátrix segítségével a következő egyenleteket írhatjuk föl β_x és κ_t paraméterekre:

$$\sum_t M_{x,t} \cdot \kappa_t = \sum_t (\log q_{x,t} - \widehat{\alpha}_x) \cdot \kappa_t = \sum_t (\beta_x \cdot \kappa_t) \cdot \kappa_t = \beta_x \sum_t \kappa_t^2 \tag{3.8}$$

$$\sum_x M_{x,t} \cdot \beta_x = \sum_x (\log q_{x,t} - \widehat{\alpha}_x) \cdot \beta_x = \sum_x (\beta_x \cdot \kappa_t) \cdot \beta_x = \kappa_t \sum_x \beta_x^2 = \kappa_t \tag{3.9}$$

Vezessük be a következő jelölést:

$$\sum_t \kappa_t^2 = b$$

Ezt felhasználva kapjuk a következő egyenleteket:

$$M \cdot \kappa = b \cdot \beta \qquad M^T \cdot \beta = \kappa$$

A második egyenletet M-mel szorozva kapjuk, hogy:

$$(M \cdot M^T) \cdot \beta = M \cdot \kappa = b \cdot \beta$$

Vagyis β az $M \cdot M^T$ mátrix b sajátértékéhez tartozó sajátvektor. Az előbbi összefüggéseket felhasználva $L(\alpha, \beta, \kappa)$-ra a következőket kapjuk:

$$L(\alpha, \beta, \kappa) = \sum_{t,x} (log(q_{x,t}) - \widehat{\alpha}_x - \beta_x \kappa_t)^2 = \tag{3.10}$$

$$= \sum_{t,x} (M_{x,t} - \beta_x \kappa_t)^2 = \tag{3.11}$$

$$= \sum_{t,x} M_{x,t}^2 + \sum_{t,x} \beta_x^2 \cdot \kappa_t^2 - 2 \cdot \sum_t \kappa_t (\sum_x \beta_x \cdot M_{x,t}) = \tag{3.12}$$

$$= \sum_{t,x} M_{x,t}^2 + 1 \times b - 2 \cdot \sum_t \kappa_t^2 = \tag{3.13}$$

$$= \sum_{t,x} M_{x,t}^2 + b - 2 \cdot b = \tag{3.14}$$

$$= \sum_{t,x} M_{x,t}^2 - b \tag{3.15}$$

Tehát az $L(\alpha, \beta, \kappa)$ akkor lesz minimális, ha b maximális. Vagyis $\widehat{\beta}_x$ az $M \cdot M^T$ mátrix maximális sajátértékéhez tartozó sajátvektor lesz, míg $\widehat{\kappa}_t = M^T \cdot \widehat{\beta}_x$.

Vagyis a Lee-Carter módszer szerint az így nyert $\widehat{\alpha_x}$ paramétereket időfüggetlen korspecifikus paraméterként; a $\widehat{\kappa}_t$ paramétereket pedig korfüggetlen, időfüggő, látens folyamatként értelmezzük. A $\widehat{\beta}_x$ paraméterek azt fejezik ki, hogy melyik korspecifikus ráta változik gyorsan vagy lassan a κ_t paraméter egy egységnyi változásának hatására. Az $\epsilon_{x,t}$ a mortalitási ráták körüli véletlen szerepet jelöli.

3.1.2. $\widehat{\kappa}_t$ kiigazítása

Második lépés során a $\widehat{\kappa}_t$ paramétereket igazítjuk ki, hogy a megfigyelt és a modellezett halálozások száma minden évben megegyezzen egymással. Erre azért van szükség, mert a modell paramétereinek becslésekor a fiatal korok halálozási rátái ugyanakkora súlyt kapnak, mint az idős koroké, pedig a fiatal korok jóval kisebb mértékben járulnak hozzá az összes halálozás számához.
A $\widehat{\kappa}_t$-t helyettesítjük $\widetilde{\kappa}_t$-vel, amelyet a következő egyenletből egyértelműen meghatározhatunk:

$$\sum_x D_{x,t} = \sum_x N_{x,t} \cdot \exp(\widehat{\alpha}_x + \widehat{\beta}_x \widetilde{\kappa}_t)$$

ahol

$D_{x,t}$ a t. évben x évesen meghaltak száma,
$N_{x,t}$ a t. évben az x évesek száma,
$\widehat{\beta}_x$ és $\widehat{\alpha}_x$ az első lépésben becsült paraméterek.

3.1.3. $\widetilde{\kappa}_t$ előrejelzése

A Lee-Carter modell nagy előnye, hogy egyetlen időfüggő paraméter van csak a rendszerben, vagyis a $\widetilde{\kappa}_t$, így a halálozási ráták előrejelzéshez csak ezt a paramétert kell becsülnünk. Lee és Carter arra a következtetésre jutottak, hogy a halálozási ráták előrejelzésénél az alábbi formula tudja a legjobban leírni a $\widetilde{\kappa}_t$ jövőbeli értékeit:

$$\widetilde{\kappa}_t = \widetilde{\kappa}_{t-1} + \theta + \eta_t \tag{3.16}$$

ahol
θ az úgynevezett drift-(vagy trend-) paraméter, ami a mortalitás csökkenésének várható tendenciáját, mértékét ragadja meg,

η_t pedig $N(0, \delta^2)$ eloszlású hiba.
A drift paraméter csak a $\widetilde{\kappa}_t$ első és utolsó elemétől függ és a következőképp becsülhető:

$$\widehat{\theta} = \frac{\widetilde{\kappa}_T - \widetilde{\kappa}_1}{T - 1}$$

Továbbá a maximum likelihood módszer a következő becslést adja a hiba varianciájára és a trend sztenderd hibájára:

$$\widehat{\delta}^2 = \frac{1}{T-1} \sum_{t=1}^{T-1} (\widetilde{\kappa}_{t-1} - \widetilde{\kappa}_t - \widehat{\theta})^2$$

$$\mathrm{SE}(\widehat{\theta}) = \frac{\widehat{\delta}}{T - 1}$$

Ha a $(T + s)$-edik időpontra szeretnénk előrejelezni (ahol T időpontig vannak megfigyelt adataink), akkor iteratívan behelyettesítve 3.16 egyenletbe a megelőző összefüggéseket kapjuk a következő összefüggést:

$$\widetilde{\kappa}_{T+s} = \widetilde{\kappa}_T + s \cdot \theta + \sum_{n=1}^{s} \eta_{T+n}$$

A becsült paramétereket felhasználva becsüljük az új $\widehat{M}_{x,t}$ mátrixot, és kiszámoljuk a becsült halálozási rátákat:

$$\log \widehat{q}_{x,t} = \widehat{\alpha}_x + \widehat{\beta}_x \cdot \widetilde{\kappa}_t$$

3.2. Előrejelzés

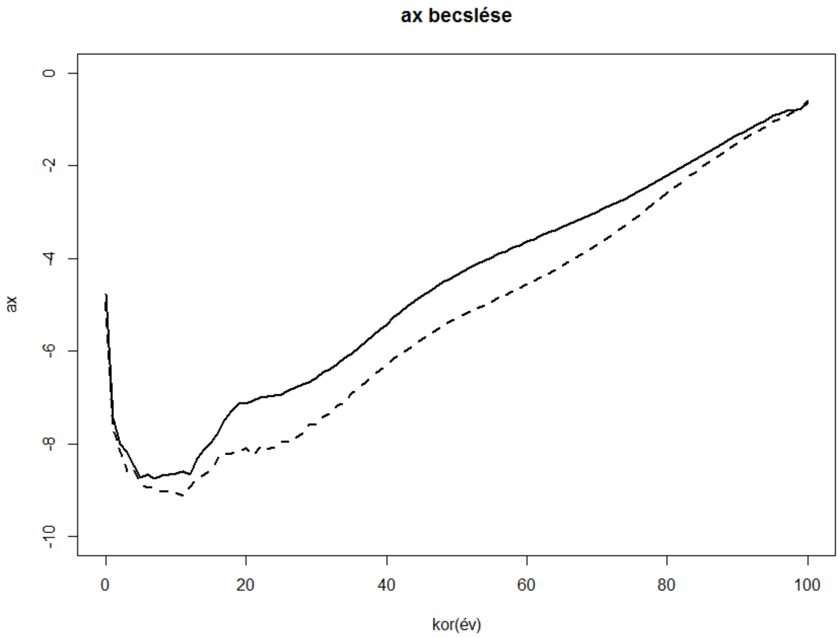
Ebben az alfejezetben a Lee-Carter modellt fogjuk használni előrejelzésre. Magyar férfi és női adatokat használunk 1990-2014 évekre és 0-100 éves korosztályokra. A számításokat R programkörnyezetben végeztük és a demography kiegészítést használva készítettük az előrejelzést.

A 3.1 ábrán láthatjuk a férfi és női α_x paraméterek becslését. Láthatjuk, hogy a vizsgált 25 naptári év azt mutatja, hogy csecsemőkorban jelentős halálozás van, ami 12 éves korig csökken, utána 20 éves korig egy intenzív növekedés és végül fokozott emelkedés látható.
A 3.2 ábrán láthatjuk a férfi és női β_x paraméterek becslését. Az első 40 évet erős

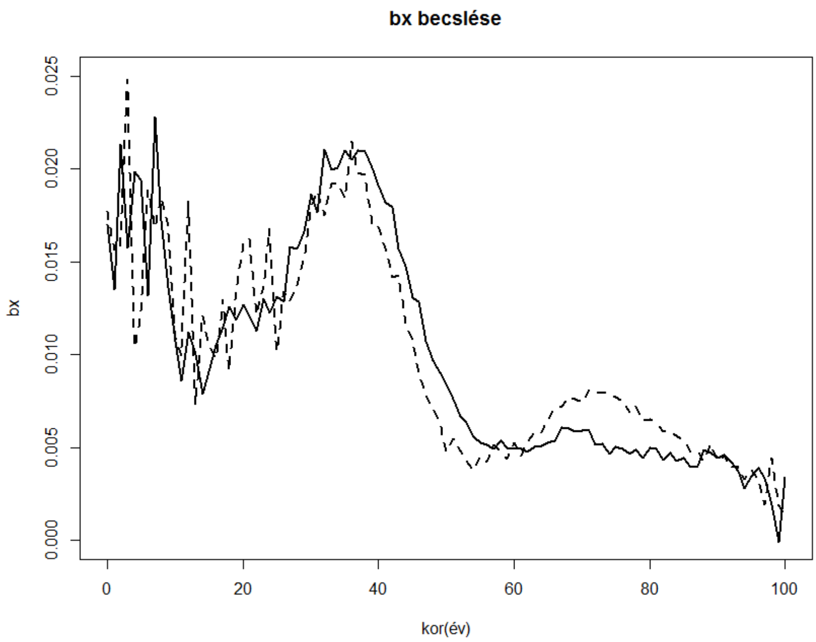
ingadozás jellemzi, majd a magas érzékenység 60 éves korig csökken. 65 és 80 között láthatunk egy újabb magas érzékenységet. A férfiaknál kisebb az ingadozás 40 éves korig és 60 fölött csökkenő trendet mutat minimális hullámzással.
A 3.3 ábrán láthatjuk a férfi és női $\widetilde{\kappa}_t$ paramétereket, előrejelzést 2020-ig és az előrejelzés 0,8-as konfidencia tartományának határait. A $\widetilde{\kappa}_t$ 1990 és 2014 között kis ingadozással csökkenő trendet mutat és az előrejelzés alapján tovább fog csökkenni.
A 3.4-es ábra mutatja az $q_{x,2010}$-es halálozási rátákat és a 2020-ra előrejelzett $q_{x,2020}$ adatokat férfiak és nők esetében. Láthatjuk, hogy mind a férfiak mind a nők esetében csökkenni fog a mortalitás az előrejelzésünk alapján.

3.3. Összehasonlítás

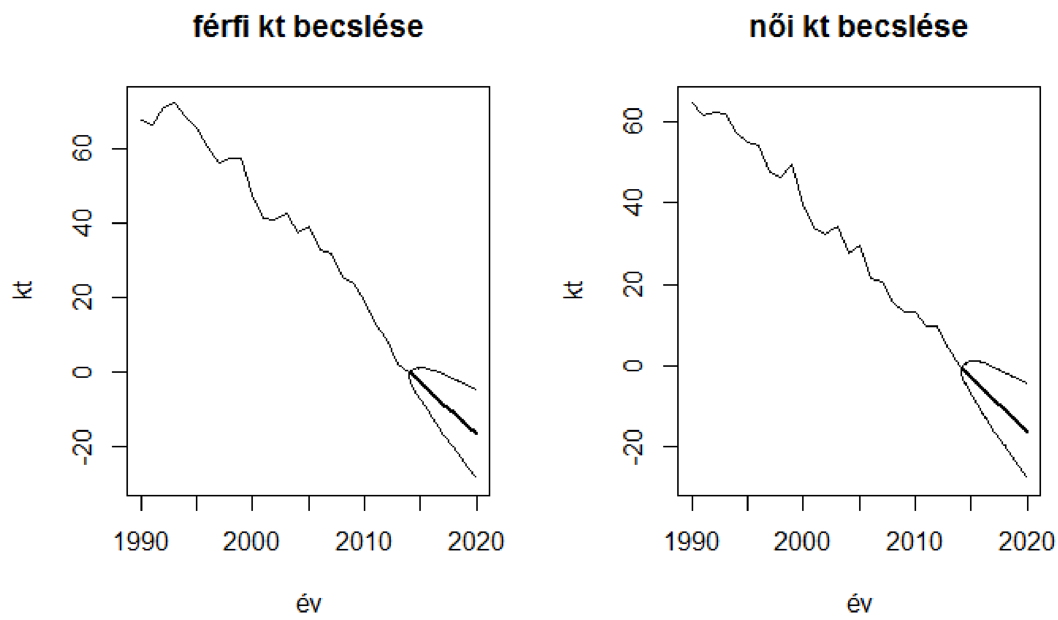
Ebben a fejezetben az előrejelzett magyar mortalitás segítségével készítünk összehasonlítást az amerikai múltbeli adatokkal a 2.2-es fejezethez hasonlóan. A 3.5-ös és 3.6-os ábrákon láthatjuk a férfi és női adatokra az összehasonlítást. A férfiaknál és a nőknél is csökkenő tendenciát mutat az arány az idő múlásával . A férfiaknál a magyar adatok minden korban alacsonyabbak, míg a nőknél is az 50-65 éves korosztály kivételével ugyanez a megfigyelés igaz.



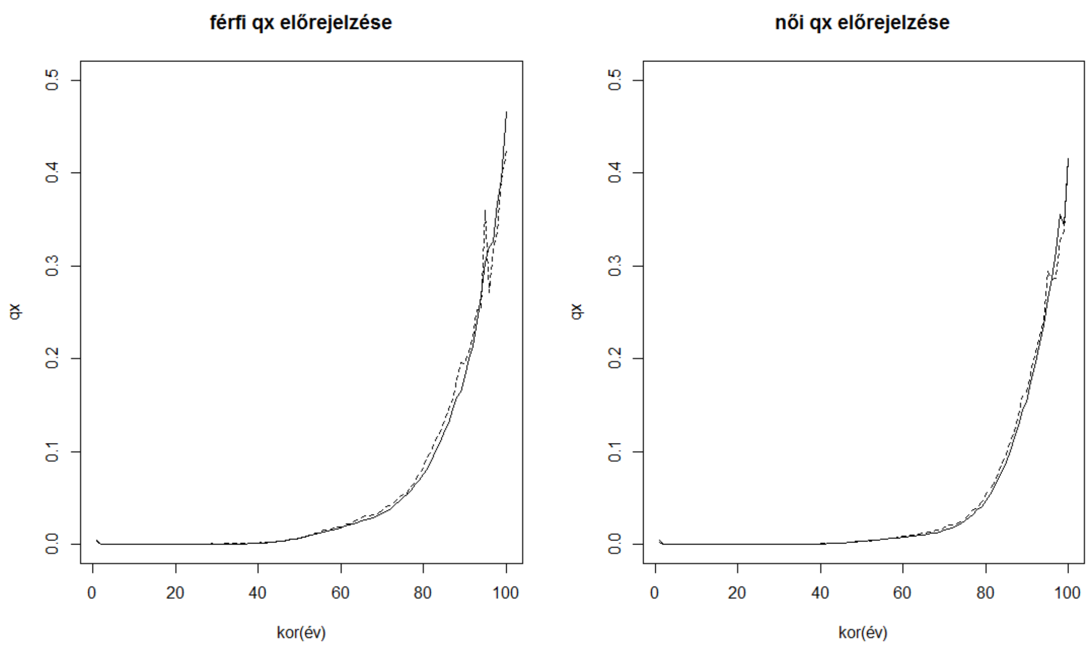
3.1. ábra. férfi(folytonos) és női(szaggatott) $\widehat{\alpha}_x$ paraméterek



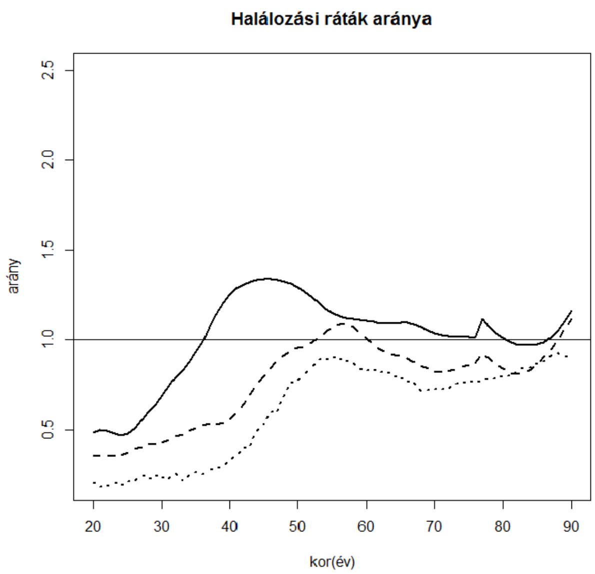
3.2. ábra. férfi(folytonos) és női(szaggatott) $\widehat{\beta}_x$ paraméterek



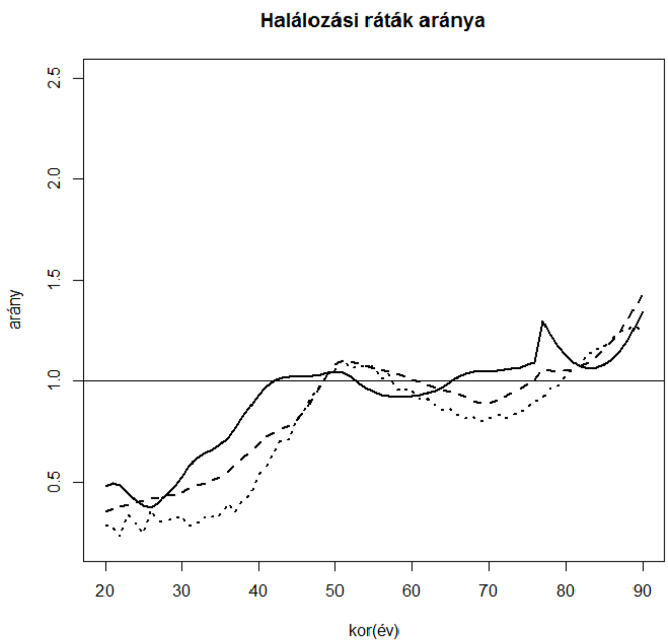
3.3. ábra. férfi és női $\widetilde{\kappa}_t$ paraméterek



3.4. ábra. férfi és női $q_{x,2010}$(szaggatott) halálozási ráták és $q_{x,2020}$(folytonos) Lee-Carter előrejezés



3.5. ábra. Magyar és USA férfi mortalitások aránya 2020/1970(pontozott);2010/1960(szaggatott);2000/1950(folytonos)



3.6. ábra. Magyar és USA női mortalitások aránya 2020/1990(pontozott);2010/1980(szaggatott);2000/1970(folytonos)

4. fejezet

Összefoglalás

A szakdolgozat első fejezetében megmutattuk, hogy nyers adatokból hogyan konstruálhatunk meg egy halandósági táblát, majd bemutattunk simítási módszereket nyers adatok kiegyenlítésére.

A második fejezetben összehasonlítottunk aktuális magyar és múltbeli USA halandósági táblákat (az [1]-es cikk mintájára) és megnéztük, a cikkhez képest 10 évvel későbbi adatok összehasonlítását és azt kaptuk eredményül, hogy a halálozási ráták aránya csökkent és még jobb közelítést kaptunk.

Végül a harmadik fejezetben ismertettük a Lee-Carter modellt, amely segítségével előrejelzést készítettünk. Az előrejelzés felhasználásával csináltunk egy újabb összehasonlítást, ami azt mutatta, hogy a halálozási ráták aránya csökkent az idő múlásával, mivel a magyar mortalitás nagyobb csökkenést mutatott, viszont az illeszkedés rosszabb lett.

Irodalomjegyzék

[1] Arató Miklós, Bozsó Dávid, Elek Péter, Zempléni András: *Forecasting and simulating mortality tables-Mathematical and computer modelling.* 49(2009) 805-813.

[2] Pallós Emil, *Magyarország halandósági táblái.* KSH Népességtudományi Kutató Intézet Közleményei, 34. sz. Budapest, 1974/2.

[3] Szabó László Imre, Viharos László: *Az életbiztosítás alapjai.* Szeged, 2001.

[4] Ronald D. Lee, Lawrence R. Carter *Modelling and Forecasting U.S mortality.*Journal of the American Statistical Association, 87(419), 1992 szeptember, 659-671

[5] Májer István, Dr.Kovács Erzsébet*Élettartam-kockázat-a nyugdíjrendszerre nehezedő egyik teher*, Statisztikai szemle, 89.évfolyam, 7-8. szám, 2011

[6] O.S Mitchell, D.G. Mcarthy,*Estimating international adverse selection in annuities*, North American Actuarial Journal 6(4), 2002, 38-54

NYILATKOZAT

Név: Fodor Péter

ELTE Természettudományi Kar, szak: Matematika BSc

NEPTUN azonosító: YOG9M2

Szakdolgozat címe: Halandósági táblák előrejelzése

A szakdolgozat szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló munkám eredménye, saját szellemi termékem, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2016. május 31.

a hallgató aláírása