RITKA ESEMÉNYEK ELŐFORDULÁSI GYAKORISÁGÁNAK TRENDELEMZÉSE R-KÖRNYEZETBEN

AGYÉRBETEGSÉGEK OKOZTA MAGYARORSZÁGI HALÁLOZÁSI ARÁNYSZÁMOK VIZSGÁLATA 1981 ÉS 2010 KÖZÖTT

Virág Katalin

Szegedi Tudományegyetem

Általános Orvostudományi Kar, Orvosi Fizikai és Orvosi Informatikai Intézet

A kutatás a TÁMOP 4.2.4.A/2-11-1-2012-0001 azonosító számú Nemzeti Kiválóság Program – Hazai hallgatói, illetve kutatói személyi támogatást biztosító rendszer kidolgozása és működtetése konvergencia program című kiemelt projekt keretében zajlott. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

2014. március 25.

BEVEZETÉS

Epidemiológiai kutatások során gyakran megbetegedések, halálozások előfordulási gyakoriságának időbeli változásait vizsgálják. Ezen cikk ritka események bekövetkezési gyakoriságának vagy a népességhez viszonyított arányának trendelemzését mutatja be konkrét példán keresztül. *A statisztikai elemzéseket a mindenki számára ingyenesen elérhető R-környezetben végezzük, és a felhasznált kódokat mellékeljük.*

ADATOK

A Központi Statisztikai Hivatal adatai alapján vizsgáljuk a központi idegrendszerre ható érsérülések okozta magyarországi halálozási rátát nemek szerint, 1981 és 2010 között (BNO-kód: 1981-1995: 430-438; 1996-2010: I60-I69).

ELEMZÉS

1. Kor szerint standardizált arányszámok számítása

Ha a halálozási ráta korcsoportonként eltérő, akkor a teljes populációra vonatkozó ráta a populáció korösszetételétől is függ. Eltérő korstruktúrájú populációk összehasonlíthatósága érdekében a halálozási vagy más arányszámokat standardizálni kell, ezzel biztosítva, hogy a halálozási ráták közötti különbség nem a populációk eltérő korösszetételének köszönhető. A standardizálás kétféleképpen történhet: direkt és indirekt módszerrel ([8]).

1.1. Direkt standardizálás

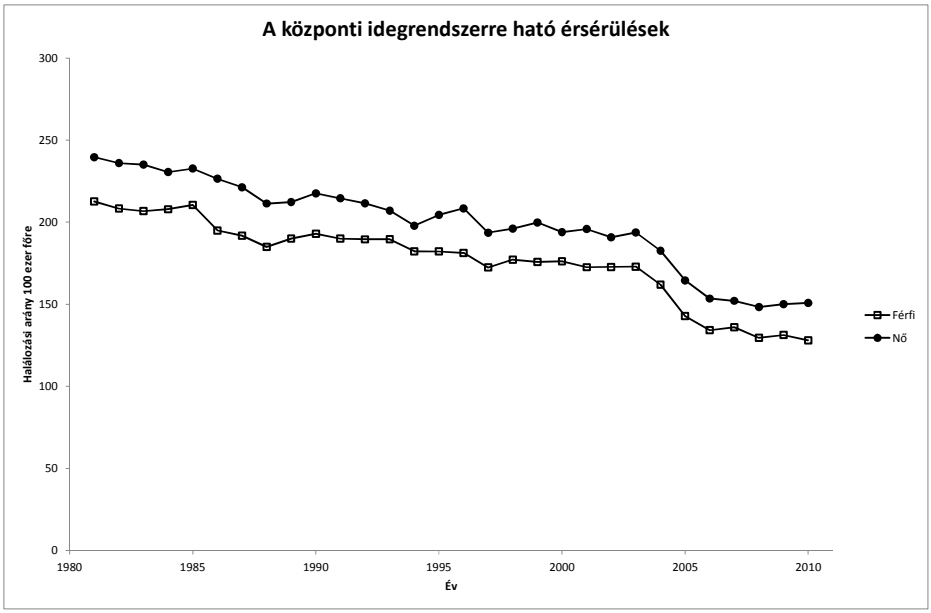
Ez az elterjedtebb módszer. A direkt standardizáláshoz egy standard populációra van szükség (referencia-népesség), amely lehet a svéd populáción alapuló Európai standard populáció, a Világ standard populáció, vagy akár a két összehasonlítandó populáció egyesítése is. Azt mutatja meg, hogy milyen lenne a teljes populációra vonatkozó halálozási ráta a referencia-népesség korstruktúrájának fennállása esetén.

1.2. Indirekt standardizálás

Az indirekt standardizáció standard rátákkal történik. Azt mutatja meg, hogy milyen lenne a populáció halálozási rátája a referenciaként használt kor-specifikus halálozási arányszámok fennállása esetén.

1.3. Agyérbetegségek okozta halálozási arányok vizsgálata

Az 1. ábra a központi idegrendszerre ható érsérülések okozta halálozási rátát mutatja be nemek szerint. Az évi halálozási arányok kiszámításához az évközepi lakosságszámot vettük figyelembe.



1. ábra. Nyers halálozási arányszámok

Az ábra alapján *nők esetén rendre magasabbak a halálozási arányok, mint férfiaknál*. Lehetséges, hogy ez a két csoport eltérő korstruktúrájának köszönhető, ezért az éves halálozási arányokat kor szerint standardizáljuk.

A direkt standardizálás módszerét alkalmazzuk, referencia népességnek az Európai standard populációt választva. Az európai standard populáció korstruktúráját az 1. táblázat tartalmazza.

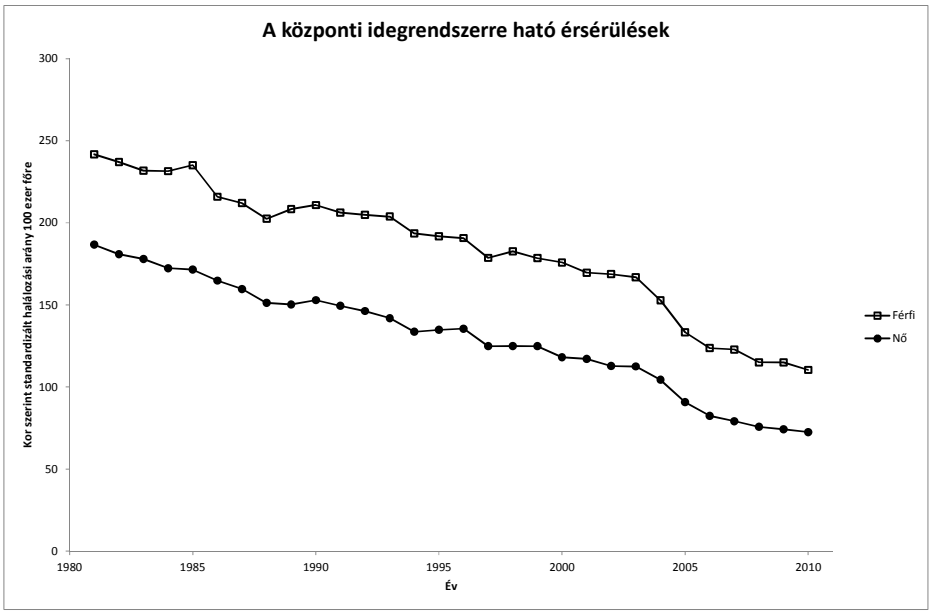
A standardizáció lépéseit az 1981-es év halálozási arányain mutatjuk be a 2. táblázatban. Direkt standardizációhoz először nemenként és korcsoportonként kiszámítjuk a halálozások várható számát (a halálozási arányokat megszorozzuk a standard populáció létszámával), majd a kapott eredményeket mindkét nem esetén összeadjuk.

Életkor	Európai standard populáció (%)	Európai standard populáció (100 000)
0-4	8	8000
5-9	7	7000
10-14	7	7000
15-19	7	7000
20-24	7	7000
25-29	7	7000
30-34	7	7000
35-39	7	7000
40-44	7	7000
45-49	7	7000
50-54	7	7000
55-59	6	6000
60-64	5	5000
65-69	4	4000
70-74	3	3000
75-79	2	2000
80-84	1	1000
85+	1	1000
Összesen	100	100000

1. táblázat. Európai standard populáció

Korcsoport	Standard populáció	Férfi		Nő	
		Halálozási arány	Halálozások várható száma	Halálozási arány	Halálozások várható száma
0-4	8000	0,0000073	0,0583	0,0000026	0,0205
5-9	7000	0,0000000	0,0000	0,0000025	0,0174
10-14	7000	0,0000053	0,0374	0,0000028	0,0199
15-19	7000	0,0000212	0,1482	0,0000097	0,0677
20-24	7000	0,0000293	0,2049	0,0000360	0,2522
25-29	7000	0,0000387	0,2709	0,0000444	0,3108
30-34	7000	0,0000983	0,6884	0,0000928	0,6497
35-39	7000	0,0001746	1,2220	0,0001348	0,9435
40-44	7000	0,0004333	3,0328	0,0002191	1,5336
45-49	7000	0,0006824	4,7771	0,0004478	3,1346
50-54	7000	0,0013328	9,3293	0,0008836	6,1854
55-59	6000	0,0022959	13,7756	0,0012806	7,6837
60-64	5000	0,0038618	19,3090	0,0022194	11,0971
65-69	4000	0,0072658	29,0630	0,0047994	19,1976
70-74	3000	0,0125759	37,7278	0,0093034	27,9103
75-79	2000	0,0218106	43,6213	0,0173390	34,6779
80-84	1000	0,0302788	30,2788	0,0276637	27,6637
85-X	1000	0,0482742	48,2742	0,0453801	45,3801
Összesen	100000		241,8190		186,7456
Kor szerint standardizált halálozási arány 100 ezer főre		**242**		**187**	

2. táblázat. Direkt standardizáció



2. ábra. Kor szerint standardizált halálozási arányszámok

A 2. ábra a kor szerint standardizált halálozási arányokat mutatja. Ezek a halálozási ráták *férfiak esetén rendre magasabbak, mint nőknél*. Standardizálás nélkül a nőknél megfigyelhető magasabb halálozási arány annak köszönhető, hogy *a nők körében magasabb az idősek aránya, mint férfiak esetén*.

2. Trendelemzés

Ha a függő változó valamilyen ritka esemény előfordulási gyakorisága vagy a népességhez viszonyított aránya, akkor az eloszlása általában jól közelíthető a Poisson- vagy a negatív binomiális eloszlások valamelyikével. A népességhez viszonyított arányok esetén gyakori a túlszóródás („overdispersion”), vagyis a variancia meghaladja az átlagot; sérül a hagyományos Poisson-regresszió feltétele, miszerint a várható érték és a variancia megegyezik ([1] – [7]).

2.1. Általánosított lineáris modellek

A Poisson- és a negatív binomiális regressziók az általánosított lineáris modellek családjába tartoznak. Az általánosított lineáris modellek az egyszerű lineáris modellek általánosításai:

- a függő változó eloszlása eltérhet a normális eloszlástól (akár diszkrét eloszlást is követhet);
- a függő változó várható értéke helyett annak valamilyen függvényét írják le a magyarázó változók lineáris függvényeként (egy kapcsolatfüggvény segítségével);

– megengedik a variancia átlagtól való függését.

2.2. Gyakorisági adatok regressziója: Poisson-regresszió

Ha a függő változó valamilyen ritka esemény előfordulási gyakorisága (pl. adott populációban adott időtartam alatt előforduló új megbetegedések száma), akkor a Poisson-modell használható. Ha *Y* Poisson eloszlású $\mu > 0$ várható értékkel, akkor:

$$\mathrm{P}(Y = k) = \frac{\mu^k e^{-\mu}}{k!} \quad (k = 0, 1, 2, \dots);$$

$$\mathrm{E}(Y) = \mathrm{Var}(Y) = \mu.$$

A lineáris modell (logaritmikus kapcsolatfüggvény használatával):

$$\ln(\mu) = \boldsymbol{X}^T \boldsymbol{\beta}$$

$$\mathrm{E}(Y) = \mu = e^{\boldsymbol{X}^T \boldsymbol{\beta}},$$

ahol

Y: függő változó, $Y \sim \mathrm{Poisson}(\mu)$;

$\boldsymbol{X}$: magyarázó változók vektora;

$\boldsymbol{\beta}$: regressziós együtthatók vektora.

2.3. Túlszóródás („overdispersion”)

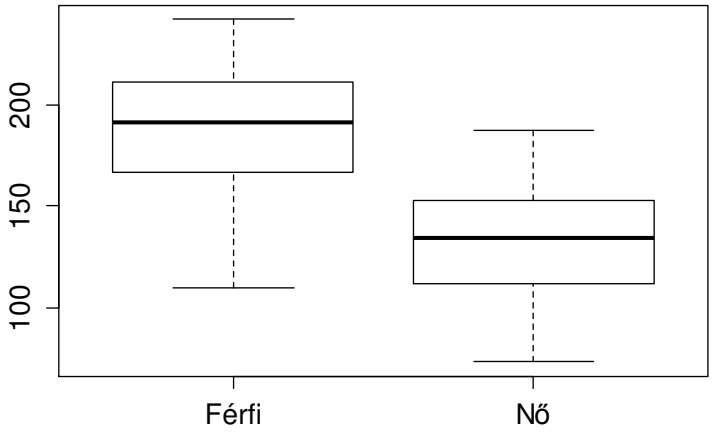
Poisson eloszlás esetén a várható érték és a variancia megegyezik, az ún. diszperziós paraméter 1-gyel egyenlő. Ha a megfigyelt gyakoriságokat a kockázatnak kitett népességhez viszonyítjuk, akkor gyakran a variancia meghaladja a várható értéket, vagyis túlszóródás figyelhető meg.

A Poisson-regresszió lehetséges általánosításai túlszóródás esetén:

- **Robusztus módszer:** a kovariancia mátrixot robusztus („szendvics”) módszerrel becsüljük, miközben a várható érték becslése változatlan.
- **Kvázi-Poisson modell:** a diszperziós paramétert a modellből becsüljük. A regressziós együtthatók változatlanok, a standard hibák nagyobbak.
- **Negatív binomiális modell:** keverék Poisson-eloszlás, ahol feltételezzük, hogy a függő változó várható értéke Gamma-eloszlást követ.
- **Zero-inflated modell:** ha az adatok között sok nulla szerepel.
- **Zero-truncated modell:** ha az adatok között egyáltalán nincs nulla.

- **Joinpoint-regresszió:** ha trendvizsgálat során az adatsorban töréspontok vannak.
- **Egzakt Poisson-regresszió:** nagyon kis gyakoriságok esetén.

2.4. Agyérbetegségek okozta halálozási arányok vizsgálata



3. ábra. Kor szerint standardizált halálozási arányok eloszlása

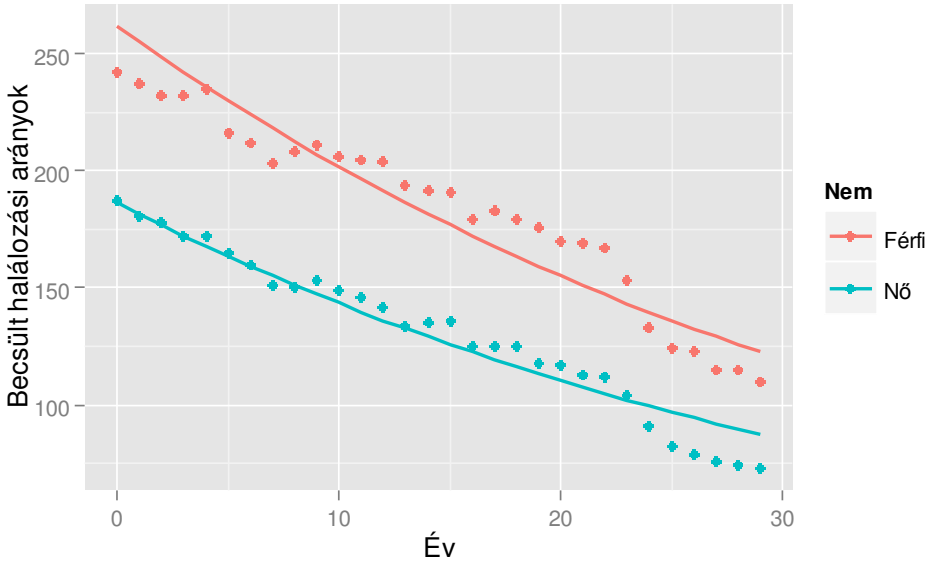
A kor szerint standardizált halálozási arányokra Poisson- és negatív binomiális regressziót illesztettünk, majd a két modellt likelihood-hányados próbával hasonlítottuk össze. Ez alapján az egyszerűbb Poisson-modellt választottuk, mert a negatív binomiális regresszió nem javította szignifikánsan a modelllilleszkedést. A regressziós együtthatók becsléseit és ezek konfidencia intervallumait (robusztus kovariancia mátrix használata esetén) a 3. táblázat tartalmazza.

Látható, hogy az agyérbetegségek következtében elhunytak aránya *szignifikánsan csökkenő (évenként körülbelül 2,6%-os) trendet* mutat, és a nemek között is szignifikáns különbség állapítható meg 5%-os szinten (*férfiak esetén majdnem 30%-kal magasabb a halálozási arány, mint nőknél*).

A 4. ábrán az eredeti adatpontokat és az illesztett (a Poisson-regresszió által becsült) görbéket láthatjuk.

	Becslés	**Konfidencia intervallum**		**p-érték**
		Alsó végpont	**Felső végpont**	
Intercept	5,569	5,532	5,605	< 0,0001
Év	-0,026	-0,029	-0,024	< 0,0001
Nem (nő)	-0,340	-0,373	-0,308	< 0,0001

3. táblázat. Regressziós együtthatók



4. ábra. Poisson-regresszió által illesztett modell

A STATISZTIKAI ANALÍZISHEZ ALKALMAZOTT R-KÓDOK

```
# a szükséges csomagok
library(MASS)
library(AER)
library(multcomp)
library(msm)
library(ggplot2)

# adatbázis beolvasása
adat <- read.csv2("Standardizalt_nemek szerint.csv")
adat <- within(adat, Nem <- factor(Nem, levels = 0:1, labels = c("Férfi", "Nő")))
attach(adat)

# átlag és variancia számítása
tapply(Agyerbetegseg, Nem, function(x){sprintf("Átlag (Variancia) = %1.2f (%1.2f)", mean(x), var(x))})

# hisztogramok készítése
hist(Agyerbetegseg[Nem == "Férfi"])
hist(Agyerbetegseg[Nem == "Nő"])

# doboz-ábra készítése
boxplot(Agyerbetegseg ~ Nem)

# Poisson-regresszió
summary(Poisson <- glm(Agyerbetegseg ~ Ev + Nem, family = "poisson"))
(Poisson_est <- cbind(Estimate = coef(Poisson), confint(Poisson), summary(Poisson)$coef[, "Pr(>|z|)"]))

# Robusztus kovariancia mátrix használatával
(Poisson_rest <- cbind(coeftest(Poisson, vcov = sandwich), confint.default(glht(Poisson, vcov = sandwich))))
exp(Poisson_rest[, 1])

# diszperziós paraméter becslése
deviance(Poisson)/df.residual(Poisson)
```

```
# Túlszóródás ellenőrzése
dispersiontest(Poisson)

# Modellilleszkedés ellenőrzése
with(Poisson, cbind(res.deviance = deviance, df = df.residual, p = pchisq(deviance, df.residual, lower.tail = FALSE)))

# Negatív binomiális regresszió
summary(NB <- glm.nb(Agyerbetegseg ~ Ev + Nem))

# Poisson-regresszió és negatív binomiális regresszió összehasonlítása
Poisson_NB <- list("Poisson" = Poisson, "Negatív binomiális" = NB)
cbind("2LogL" = 2*sapply(Poisson_NB, function(x) round(logLik(x), digits = 1)), "AIC" = sapply(Poisson_NB, function(x) AIC(x)))
cbind("2Log(L_NB)-2Log(L_Poisson)" = 2*(logLik(NB)-logLik(Poisson)), "Df" = 1, p = pchisq(2*(logLik(NB)-logLik(Poisson)), df = 1, lower.tail = FALSE))

# Illesztett modell ábrázolása
adat$becsles <- predict(Poisson, type = "response")
adat <- adat[with(adat, order(Nem, Ev)), ]
ggplot(adat, aes(x = Ev, y = becsles, colour = Nem)) + geom_point(aes(y = Agyerbetegseg)) + geom_line(size = 1) + labs(x = "Év", y = "Becsült halálozási arányok")
detach(adat)
```

HIVATKOZÁSOK

[1] A. Agresti. Categorical Data Analysis, Second Edition. Hoboken, New Jersey, John Wiley & Sons, Inc., 2002, pp. 115-164, 385-387, 559-563.

[2] C. Cameron. Advances in Count Data Regression Talk for the Applied Statistics Workshop, March 28, 2009. http://cameron.econ.ucdavis.edu/racd/count.html.

[3] Data Analysis Examples. UCLA: Statistical Consulting Group. from http://www.ats.ucla.edu/stat/dae/ (accessed October 12, 2013).

[4] J. Dobson. An Introduction to Generalized Linear Models, Second Edition. Boca Raton, Florida, Chapman & Hall/CRC, 2002, pp. 151-170.

[5] S. Everitt. Modern Medical Statistics: A Practical Guide. London, Arnold, 2003, pp. 1-20.

[6] A. Zeileis, C. Kleiber, and S. Jackman. „Regression Models for Count Data in R", Journal of Statistical Software, vol. 27(8), pp. 1-25, 2008.

[7] A. Pedan. Analysis of Count Data Using the SAS System. Proceedings of the Twenty-Sixth Annual SAS® Users Group International Conference, Cary, NC: SAS Institute Inc., 2001.

[8] O.B. Ahmad, C. Boschi-Pinto, A.D. Lopez, C.JL. Murray, R. Lozano, M. Inoue. Age standardization of rates: a new WHO standard. (GPE discussion paper series no. 31). Geneva, World Health Organization, 2001.

[9] R 2.15.3: A Language and Environment for Statistical Computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria)