A LOG-LINEÁRIS MODELL EGY ALKALMAZÁSA PULMONOLÓGIAI MEGBETEGEDÉSEK RIZIKÓTÉNYEZŐINEK VIZSGÁLATÁBAN

Nikodémusz Anna, Ábrahám Erzsébet

Semmelweis Orvostudományi Egyetem Számitástechnikai Csoport

A tömegesen elterjedt krónikus megbetegedések gyógyitásában és hatékony megelőzésében alapvetőek a rendszeresen ismétlődő szürővizsgálatok. A teljes lakosság minden betegségre kiterjedő - komplex - szürése megoldhatatlan, ezért kiemelkedő jelentőségü feladat az egyén veszélyeztetettségéhez igazitott szürési ritmus kialakitása, amelynek alapja - az adott betegség/ek/-re vonatkozó rizikótényezők ismeretében - az eltérő mértékben veszélyeztetett lakosságcsoportok elkülönitése.

Budapest X. kerületében a Tüdőgondozó Intézetben jelenleg is folyó epidemiológiai modellvizsgálat fő célja a legelterjedtebb pulmonológiai betegségek rizikócsoportjaira célzott szürési metodika kidolgozása. A fontossági sorrend figyelembe vételével a tbc-rizikócsoportok megbizható elkülönitése után, [1] a tüdődaganatos megbetegedések rizikócsoportjainak meghatározását tüzték ki célul.

Irodalmi adatok, valamint a megelőző modellkisérlet morbiditási mutatói szerint feltételzhető rizikótényezők voltak az életkoron kivül a nem, a dohányzás, az előző rtg-felvételeken látható elváltozás, a munkahelyi ártalmak, és a tartósan fennálló léguti panaszok.

A vizsgálat kezdetén az életkor alapján legveszélyeztetettebb rétegnél a 40-74 éves korosztálynál - a szokásos ernyőképszürő vizsgálattal egyidejüleg - a fenti rizikótényezőkre vonatkozó kérdőives felmérés történt, majd 3 illetve 6 év elteltével ujra megvizsgálták az érintett lakosságot.

Az adatokat, feltételezett rizikótényezők szerinti csoportositásban kontingencia-táblába rendezték, majd az első utánvizsgálat adatainak feldolgozásakor meghatározták az adott periódus alatt felfedezett uj betegek megoszlását - továbbiakban incidencia - a tábla cellái között.

Az incidencia-értékekre vonatkozó hipotéziseket χ^2-próbával ellenőrizve lényeges különbség mutatkozott a férfiak tüdőtumorinciden-ciája és a nőké között, /a férfiaké közel 6-szoros a nőkéhez képest/, lényeges rizikótényezőnek mutatkozott a dohányzás is; az erős dohányosok tüdőtumor-incidenciája szignifikánsan magasabb volt az egyáltalán nem vagy csak mérsékelten dohányzókénál. Veszélyeztetett rétegeket lehetett elkülöniteni az előző rtg-felvételeken látható elváltozás és a tartós léguti panaszok szerint is. Első közelitésben minden veszélyeztető faktort egyforma "sulyunak", halmozott előfordulásuk hatását pedig additivnak tételezve fel a következő rizikócsoportokat határoztuk meg:

3 rizikófaktorral rendelkezett a lakosok 7,1 %-a, 2-vel 26,2 %, 1-gyel 40 % és fennmaradt 26,7 % rizikómentesnek mutatkozott.

A több rizikófaktorral rendelkező erősen veszélyeztetett lakosságcsoport megbetegedési kockázata kiemelkedően magas volt az egyáltalán nem veszélyeztetettekéhez képest. Figyelemre méltó eredmény volt, hogy mig az egész népességben a nemek közti incidenciabeli eltérés szignifikáns volt, sem az erősen veszélyeztetett rétegben sem

a rizikómentes rétegben nem találtak lényeges különbséget. Ilyen és ehhez hasonló nem kellően tisztázott kérdések szükségessé tették, hogy a rizikótényezők egymáshoz való viszonyát figyelembe vevő elemzési keretet keressünk, s ezt a log-lineáris modellben találtuk meg.

Egy kontingencia-tábla log-lineáris reprezentációjánál azzal a feltevéssel élünk, hogy a várt cellagyakoriságok logaritmusát előállithatjuk bizonyos "természetes" paraméterek lineáris függvényeként, amelyek eleget tesznek néhány lineáris "kényszerfeltételnek". A "természetes" paraméterek a megfigyelt gyakoriságok logaritmusainak lineáris függvényei. Az irási kényelem és a könnyebb áttekinthetőség kedvéért példaként tekintsük egy háromszempontu kontingencia-tábla log-lineáris reprezentációját: Jelöljük a tábla általános elemét x_{ijk}-val, az osztályozási kategóriákat A-val, B-vel és C-vel, akkor feltevésünk szerint:

$$\ln x_{ijk}^{*} = \theta + \lambda_i^{A} + \lambda_j^{B} + \lambda_k^{C} + \lambda_{ij}^{AB} + \ldots + \lambda_{ijk}^{ABC} \qquad (1)$$

érvényes, és a kényszerfeltételek a marginális összegekre vonatkozó - természetes - feltételekből adódnak.

A λ-kat főhatásoknak, illetve interakcióknak nevezzük, és a felső index jelzi, hogy mely táblafaktor/ok/ hatására indulnak. Egy log-lineáris modellt röviden a benne szereplő hatások és interakciók szögletes zárójelek között való felsorolásával jelölünk.
Pl. az (1) modell röviditett jelölése a következő:

$$[A,B,C, \ldots, ABC]$$

Ezt teljes vagy szaturált modellnek nevezzük.

A log-lineáris modell kiválasztásánál az a cél, hogy a "jól" illeszkedő modellt találjunk, amely "takarékos" is egyben abban az értelemben, hogy a lehető legkevesebb paramétert használja. Takarékosság tekintetében a két extrém példa a telitett modell és a teljes homogenitás hipotézisének megfelelő egyenletes eloszlást feltételező. A két véglet között modellek tág varietása helyezkedik el, amelyeket ugy nyerünk, hogy a modellben bizonyos interakciókat 0-vá teszünk. Fontos osztályt képeznek az un. hierarchikus modellek, amelyekben nem fordulhat elő valamely magasabb rendü effektus anélkül, hogy összes olyan alacsonyabb rendü interakció ne szerepelne benne, amelynek faktorait ő részhalmazként tartalmazza. A hierarhikus modellek nagyon kellemes tulajdonságokkal rendelkező részosztályát képezik az un. felbontható vagy grafikus modellek [5].

Ezekre a modellekre az jellemző,
1/ hogy az MLE explicit alakban kifejezhető a marginálisok logaritmusainak lineáris függvényeként,
2/ a modell interpretálható a függetlenség, feltételes függetlenség, feltételes homogenitás és homogenitás terminusaiban,
3/ olyan gráfreprezentációjuk van, amely segit eldönteni mintegy "ránézésre" is a modell felbonthatóságát.

Ha a log-lineáris analizist a hierarchikus modellekre korlátozzuk, az analizis első fázisában a lényeges effektusok kiszürését végezzük el, a második fázisban a hipotetikus modelleket teszteljük a következő statisztikákkal.

$$G^2 = \sum_i \sum_j \sum_k x_{ijk} \ln (x_{ijk}/x_{ijk}^*) \qquad (2)$$

amely χ^2 eloszlásu aszimptotikusan n-p szabadságfokkal, ahol n az elemszám és p a becsült paraméterek száma.

Megjegyezzük, hogy G^2 additiv egymásbaágyazott hierarhikus modelleken, mig a Pearson-féle χ^2-statisztikánál ez nem mindig teljesül.

Végezetül visszatérve saját alkalmazásunkra - mely még korántsem átfogó igényü, és validálását majd az ujabb utánvizsgálati adatok teljes feldolgozása után lehet elvégezni - 2 példát mutatunk be; a log-lineáris analizist a BMDP programcsomag 3F programjának futtatásával végeztük. 3 éves utánvizsgálati incidenciaadatokra az [SD,MD] modell jól illeszkedett, /S a nem, D a dohányzás mértéke, M a morbiditás/, melyet ugy lehetett interpretálni, lévén felbontható modell, hogy a nem és a tüdőtumor morbiditás feltételesen függetlenek a dohányzási szokásokra vonatkozóan.

Itt az

$$x_{ijk}^* = x_{i.k} \cdot x_{.jk}/x_{..k} \qquad (3)$$

statisztika adja az "MLE" becslést.
Hasonlóan jó illeszkedést kaptunk - elemezve az erős dohányosok morbiditási tábláját - a következő szempontok szerint: nem, előző rtg-felvétel lelete, léguti panaszok. A szempontokat rendre, M,S,R,P betükkel jelölve, a modell

[MR,MP,RS,RP,SP]

alaku.

Eddigi eredményeink is igazolják, hogy a log-lineáris modell alkalmas matematikai-statisztikai eszköz lehet a krónikus megbetegedések incidencia-megoszlásának feltételezett rizikófaktorok segitségével való leirására, az egyes faktorok kölcsönhatásának tesztelésével mód nyilik az eltérő mértékben veszélyeztetett lakosságcsoportok megbizhatóbb elkülönitésére is.

Irodalom

[1] Ábrahám, E.: A programozott célzott szürés egészségpolitikai jelentősége, Bajcsy-Zsilinszky Kórház-Rendelőintézet Évkönyve 1981.

[2] Bishop, Fienberg, Holland: Discrete multivariate analysis, MIT Press 1975.

[3] Gochale, Kullback: The information in contingency tables, Marcel Decker Inc. 1979.

[4] BMDP Biomedical Computer Programs, P-series 1977. University of California Press.

[5] J.N. Darroch, S.L. Lauritzen, T.P. Speed: Markov fields and log-linear interaction models for contingency tables, 1980. The Annals of Stat. Vol. 8. No.3. 522-539.