

Szeged megyei városi Tanács VB Egészségügyi Osztálya, Csongrád megyei
Kórház Gyermekosztálya, Szentes

Ujzülöttek születési súlyát befolyásoló tényezők
számítógépes vizsgálata

Török Rozália, Kovács Zoltán és Kovalovszki Lajos

Az elmúlt évtizedekben jelentősen növekedett a koraszülések százalékos aránya. Míg 1936-ban hazánkban 100 szülés közül 3-4 esetben volt koraszülés, addig ma minden 10. szülés esetén 2500 gr alatti az ujzülött súlya. Az orvosi szakirodalomban számtalan, ezzel a kérdéssel foglalkozó közlemény látott napvilágot. Az eddigi eredmények azt mutatják, hogy egy vagy akár néhány (2-3) tényező vizsgálata nem ad elegendő felvilágosítást a kérdés megoldásához. Feltehetően több, közel azonos fontosságú tényező befolyásoló hatásával kell számolnunk. A vázolt kérdés vizsgálatánál igen nagy mennyiségű adat tárolására, analizálására van szükség, ezért ez csak elektronikus számítógép segítségével lehetséges.

A Szentesi Megyei Kórház Gyermekosztályán egy, ezzel a témával kapcsolatos statisztikai felmérést végeztek. Ebbe olyan kérdéseket vettek fel, melyek az orvosi tapasztalat alapján feltehetően befolyásolják az ujzülött születési súlyát.

Először ki kellett válogatni a kérdések közül azokat, melyek a koraszüléssel szoros kapcsolatban vannak. Erre azért volt szükség, mert a nem lényeges tényezők vizsgálata jelentősen megnövelné a program gépi idő szükségletét, memória igénye is nagyobb lenne, ugyanakkor az eredmény lényeges javulásához nem vezetne.

A feladat további része az volt, hogy a megmaradt tényezők figyelembevételével a mintában rejlő információ alapján olyan függvényt konstruáljunk, mely nagy valószínűséggel választja szét a koraszülöttek csoportját a nem koraszülöttekétől.

A programot INZSENYER nyelven írtuk és a JATE Kibernetikai Laboratórium MINSZK-22 típusú számológépén próbáltuk ki. Az adatok előkészítését, mágnesszalagra történő felírását a FELIRÁS KONVERTÁLÁSSAL program végezte. Az ujzülött születési súlyának és az egyes befolyásoló tényezők közötti kapcsolat erősségét X^2 -próbbal határoztuk meg.

Tekintsük az i -dik tényezőt és a születési sulyt.

A_1, A_2, \dots, A_r az i -dik kérdésre adható válaszok a
 B_1, B_2, \dots, B_s születési suly lehetséges értékeinek
($r, s \geq 2$) megfelelő kódok.

Mindkét rendszer teljes eseményrendszer.

χ^2 - próbával történő függetlenségvizsgálat esetén a H_0 hipotézis:

$$P(A_i, B_j) = P(A_i)P(B_j) \quad i = 1, 2, \dots, r, \quad j = 1, 2, \dots, s.$$

Megalkotva a

$$\chi_n^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{\left(\sqrt{|j-n} \frac{v_i}{n} \cdot \frac{v_j}{n} \right)^2}{n \cdot \frac{v_i}{n} \cdot \frac{v_j}{n}}$$

kifejezést, nagyszámu minta esetén a H_0 hipotézis teljesülésekor ez $(r-1)(s-1)$ szabadságfoku χ^2 eloszlást követ. A számítást és értékelést a TÉNYEZŐK FÜGGETLENSÉGÉNEK VIZSGÁLATA program végezte.

A próba akkor alkalmazható, ha $np_i \geq 10 \forall i$ esetén. A program ezt a feltételt ellenőrizte és azokban az esetekben, amikor ez nem teljesül, az ASR-en adott jelzést, nem végezve el a számítást és értékelést.

A koraszülés és az egyes tényezők kapcsolatának vizsgálatakor ilyen hiba nem adódott, de éppen amiatt a feltétel miatt már pl. a koraszülés és két vagy több tényező kapcsolatának vizsgálata ebből az adathalmazból nem végezhető el. A χ^2 -próba feltétele ezekben az esetekben ugyanis sokszor nem teljesülne, ilyen vizsgálati eredményből nem vonhatnánk le következtetéseket, ezért a program csak két-két tényező függőségi viszonyát vizsgálja. Megkíséreltük a lényeges tényezők kiválogatását az információelmélet felhasználásával is.

Ebben az esetben adott egy függő változó és n darab független változó, kérdés, hogyan választhatók ki a független változók közül a függő változó szempontjából leglényegesebbek. Y legyen a függő változó, X_1, X_2, \dots, X_n legyenek a független változók. $M(Y, X)$ lesz X fontosságának mértéke Y vonatkozásában, ha teljesíti a következő feltételeket:

- 1) M figyelembe veszi X saját fontosságát és X -nek Y -ra vonatkozó feltételes fontosságát.
- 2) M minimális értékű, ha Y és X statisztikailag függetlenek.

3) $M(Y_1, X_2) > M(Y_1, X_1)$, ha X_2 lényegesebb Y -ra nézve mint X_1 .

Az információelméletben használt kölcsönös információ használható ezen M érték definiálásához.

Esetünkben az X_1, X_2, \dots, X_s független tényezőkről a **TÉNYEZŐK FÜGGETLENSÉGÉNEK VIZSGÁLATA** program eredménye alapján feltehetjük, hogy függetlenek.

$$\text{Ekkor } I(Y_1, X_1, X_2, \dots, X_s) = I(Y_1, X_1) + I(Y_1, X_2) + \dots + I(Y_1, X_s)$$

A χ^2 - próba és az információtartalom vizsgálatának összehasonlításakor a két program eredményében nagyfokú egyezőséget találtunk. Azok a tényezők, melyeket a **TÉNYEZŐK FÜGGETLENSÉGÉNEK VIZSGÁLATA** program a koraszülést befolyásoló tényezők közé sorolt, általában informatívabbak voltak a kimaradtaknál.

A további vizsgálatra szánt kérdések kiválogatására ez a program szolgáltatott alapot, azonban a végleges döntés kialakítása orvosi vélemény figyelembevételével történt.

További vizsgálatra 22 tényezőt tartottunk meg.

Munkánk célja ezután az volt, hogy a születési súlynak ezen tényezőktől való függését vizsgáljuk és a mintákból nyert információ alapján közelítő függvénykapcsolatot állapítsunk meg.

Az Y értékét X_1, X_2, \dots, X_s lineáris függvényével közelítjük.

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_s X_s$$

Az együtthatókat a legkisebb négyzetek módszere alapján becsüljük.

A lineáris közelítéshez készített tesztprogram során egy P pontot elválasztópontnak nevezünk, ha a lineáris közelítés együtthatóival egy adott mintapontban a függvény értékét meghatározva, abban az esetben döntünk úgy, hogy koraszülés várható, ha a P -nél kisebb érték adódott. A **TESZTPROGRAM** megállapítja, hogy a P különböző helyeken történő rögzítésekor a lineáris függvény hány százalékban dönt helyesen, hány százalékos a "negatív hiba" (koraszülöttnek nem koraszülötté történt nyilvántartása), és hány százalékos a "pozitív hiba". A P rögzítésének csak a "nem-koraszülött" - "koraszülött" intervallumban van értelme, ahol a "nem-koraszülött" a lineáris függvény minimuma a nem koraszülött mintaelemeken, a "koraszülött" pedig a függvény maximuma a koraszülött mintaelemeken.

Az eredmény értékelése során az elválasztópontnak $P = 6,3417$ értéknél való rögzítése esetén a **LINEÁRIS KÖZELÍTÉS**-sel kapott

függvény 61 %-ban helyesen diagnosztizált, a pozitív hibaszázalék 29, a negatív hibaszázalék 10 volt. Ezt az eredményt azért tekinthetjük pozitívnak, mert az a legfontosabb, hogy felhívjuk a figyelmet azokra az esetekre, amikor várhatólag alacsony lesz a születési súly. A "nem-koraszülött" és "koraszülött" értékek megadják a koraszülöttek, illetve a nem-koraszülöttek csoportjában a lineáris függvény legkisebb és legnagyobb értékét. Ezáltal egy olyan intervallum adódik, amelyen kívül eső mintaelemekre, pontosabban ahol a függvény értéke ezen intervallumon kívül esik, biztos előrejelzést adhatunk.

A potenciálfüggvényes alakfelismerési módszer tanításon alapszik. A tanítás folyamán az M tanulóprogram sorozathoz hozzárendeljük az \hat{M} sorozatot:

Az \hat{M} sorozatot tanított sorozatnak nevezzük.

A hozzárendelés módjától függően megkülönböztethetünk

a) egyszerű tanítást, ahol

$$\hat{x}_i = \begin{cases} 1 & \text{ha } \underline{x}^i \in A \\ -1 & \text{ha } \underline{x}^i \in B \end{cases}$$

b) összetett tanítást, ahol

$$\hat{x}_i = y_i \quad - \text{itt } y \text{ az } i\text{-dik tanulóponthoz az } f(x) \text{ függvény értéke, esetünkben a születési súly kódja.}$$

Az összetett tanítás több osztályra való bontást is lehetővé tesz.

A tanítás célja olyan $f_n(\underline{x}) = f_n(\underline{x}; \{\hat{x}_i\}_{i=1}^n; \{\underline{x}^i\}_{i=1}^n)$

függvénysorozat meghatározása, amely valamilyen értelemben a szétválasztó függvényt $f(\underline{x})$ -et megközelíti.

Az algoritmus általános alakja:

$$f_n(\underline{x}) = \sum_{k=1}^n r_{k-1}(\underline{x}^k) K(\underline{x}, \underline{x}^k)$$

A módszer előnye, hogy lehetséges a szétválasztó függvényt közelítő $f_n(\underline{x})$ módosítása a folyamatosan érkező mintaelemek segítségével.

A POTENCIÁL-FÜGGVÉNYES MÓDSZER programja a következő feladatokat oldja meg:

- 1) Tanítóprogramként működik.
- 2) Tesztprogramként működik.

A probléma megoldási módjainak összehasonlítása során a LINEÁRIS KÖZELÍTÉS -sel kapott eredmények azt mutatják, hogy néhány négyzetes és vegyes tag figyelembevételével orvosi szempontból elfogadható eredményt érünk el. A módszer alkalmazása ellen szól azonban, hogy minden változtatás esetén újra kell számítani az együtt-hatókat. A korábbi számítások eredményéből nem végezhető el a módosítás. Így a gépi idő egyre növekedne, azonkívül minden mintaelemet meg kell őriznünk.

A POTENCIÁLFÜGGVÉNYES MÓDSZER - bár eredményei nem olyan jók - éppen ezeket a hátrányokat küszöböli ki. Itt a közelítő-függvény módosítása egyszerűen úgy történik, hogy a korábbi tanuló-pontok által szolgáltatott együtt-hatókat az új tanuló-pontok esetén megváltoztathatjuk.

Anyagunknál a POTENCIÁLFÜGGVÉNYES MÓDSZER alkalmazásakor feltételezhető, hogy azért van nagy eltolódás a "pozitív hiba" irányába, mert a tanuló-pontok összetétele nem volt megfelelő. Anyagunkban az esetek fele koraszülött volt. A természetes arány a koraszülöttek és nem koraszülöttek között 10 % és 90 % körül ingadozik.