

MÁV Tudógyógyintézet Szanatóriumi Osztály, Központi Fizikai  
Kutató Intézet Bioelektronikai Csoportja

Multifaktoriális vizsgálatok cluster analízissel a zsirvérőség  
kórismézésében

Ghyczy Kálmán dr, B. Nagy András, Wolf Tamás

A betegvizsgálatnak az a célja, hogy az adódó tünetek és jelek ismeretében felállithassuk a kórismét. Jó lenne, ha minden betegségnek volna külön-külön csak reá jellemző tünete vagy jele. Természetesen nem így van, legalább is nem, ami az ugynevezett általános betegvizsgálatot illeti.

A különleges vizsgálatok viszont éppen erre volnának hivatva. Sokszor már-már bizonyító erejüknek tűnnek. De tulajdonképpen olyan vizsgálat, ami egymagában egyértelműen bizonyít, talán nincs is.

Ebből az következik, hogy azt az általános vizsgálata-  
tot lehetőleg ki kell használni. Bár az így nyert adatok  
érzékenysége, fajlagossága általában szerény - tehát a  
vizsgálat során ugyanaz az adat több különböző betegség-  
ben fordulhat elő - a ráismerés mégis lehetséges. Egyrészt  
jellemző lehet az abban a betegségben előforduló tünetek,  
jelek sajátos összeválogatódása. Egyszóval a minta ismer-

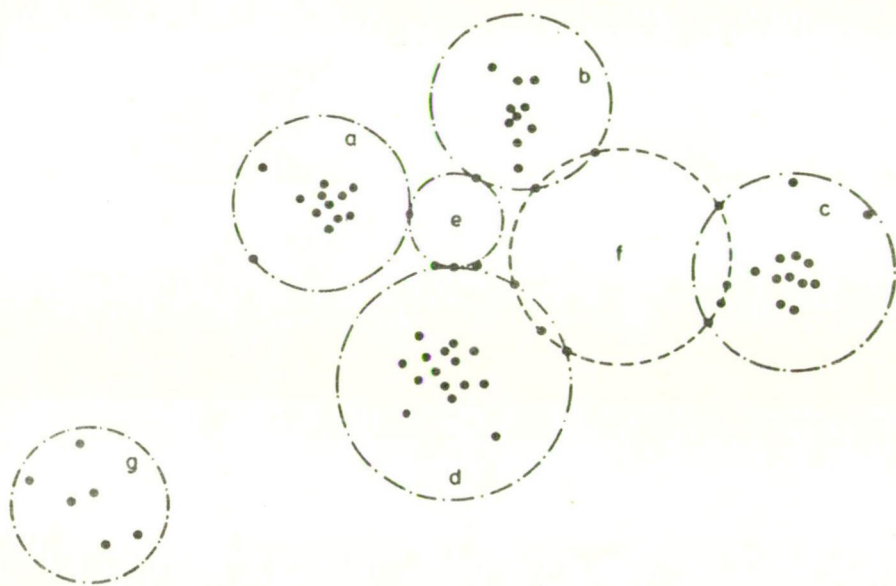
hető fel. Másrészt jellemző lehet az adatok különböző előfordulási gyakorisága az egyes betegségekben. Ez meg valószínűségi becslésre adhat alkalmat.

Az ilyen általános vizsgálat során mindig sok tényezőt leszünk kénytelenek mérlegelni. Ha csak önkényegesen nem hagyunk mindenfélét figyelmen kívül, akkor a megítélés mindig sokszempontu, azaz multifaktoriális lesz.

Nagyon hamar elveszíti az ember az áttekintést, ha egyszerre több változóval és azoknak egymáshoz való viszonyaival foglalkozik. Ezért terelődik a figyelem a klinikumban is mindinkább a sok adat feldolgozására való matematikai módszerekre.

Ilyen módszer pl. az ugynevezett cluster analysis is, ami egy adatsokaság értékeinek hasonlóságán ill. különbözőségén alapszik. A fogalomzavar elkerülésére néhány meghatározást ismertetünk abból az irodalomból, amihez magunkat tartjuk: taxonomia az osztályozás módszertana, numerikus taxonomia a matematikai módszerek bevezetése ugyanide, az empiria vagyis az egyszerű megfigyelés ill. az intuición - magyarul: ösztönös megérzés - helyett. Az orvosi taxonomia [4] az oki és a korbonctani összefüggéseket figyelmen kívül hagyva csak a betegről nyerhető vagy a rajta mutató tünetekből, jelekből, adatokból indul ki.

A hasonlóság - hasonló elrendeződés ill. közelálló érték - révén csoportok - cluster-ek, halmazok - állanak össze, és vagy ezek pontosabb körülhatárolása vagy esetleg új csoportok felfedezése válik lehetővé.



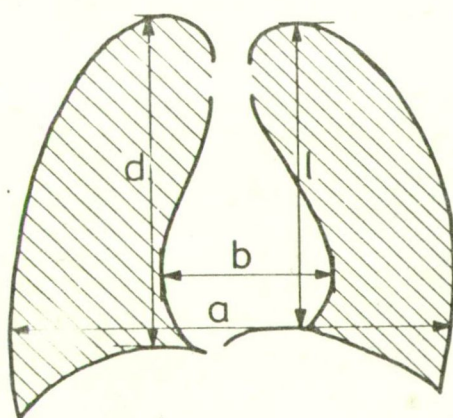
Az ábra Baron és Frazer cikkéből származik [1] és benne röviden összefoglalható a cluster analízis lényege. A pontok az egyedek. A köztük lévő távolság ill. közelség a különbség ill. hasonlóság mértéke. A csoporton belül az egyedek közti távolság kisebb mint a csoportok közti távolság.

A jelen munka során könnyen ellenőrizhető és átlátható témára volt szükségünk, hogy egy új - klinikai célra készült - cluster eljárást bemutathassunk.

Munkahipotézisünk: a különleges vizsgálat bizonyítóerejét lényegesen növeli a multifaktoriális vizsgálati módszer.

A MÁV Tüdőgyógyintézet szanatóriumi osztályán évek óta minden betegen egy, a szokásos orvosi vizsgálattól független és amennyire lehet szabványosított "elővizsgálatot" végeztünk.

Ez vagylagos fogalmazású önkitöltős kérdőivekből és különféle a betegről felvett számszerű adatokból áll. A mostani vizsgálat során figyelembe vett adatokat ábrába foglaltuk össze.



$$CTI = b/a$$

$$LTId = d/a$$

$$LTI_s = l/a$$

Testmagasság

Testsúly

Mellbőség

Hasbőség

Pulzusszám

Légzésszám

Összlipoid

Cholesterin

Triglicerid

Lipoproteid

Ezen felül még a Kiske /kisérő kérdőív/ jelű kérdőívünk 117 db vagylagos fogalmazásu elsősorban kardio-pulmonális vonatkozásu, kérdőívpontját értékeltük [2], [3].

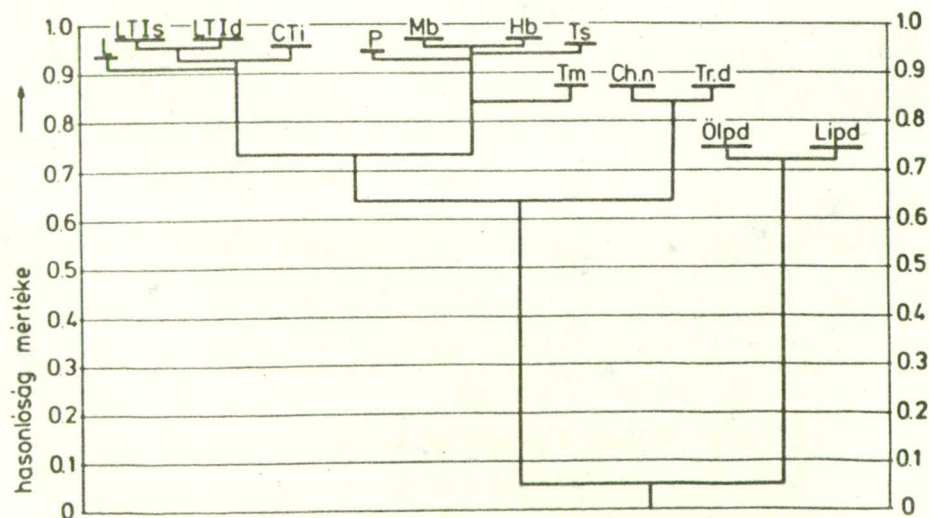
A kérdőív bináris, a többi mutató analógiás adatot szolgáltat. A feldolgozásra kiválasztottunk összesen 30 kórlapot. A kigyűjtés szempontjai: először kiemeltük 9 olyan betegünk adatait, akik ugynevezett zsirvérűségben /hiperlipaemia/ szenvedtek. Ez jellegzetes panaszt, vagy egyértelmű külső jelet nem idéz elő, viszont mint kockázati /risk/ tényező az életkilátásokat csökkenti. Felismerése tehát elemi érdek. Ezután kiválasztottunk minden - összesen 6 - olyan pacienst is, akik nem bizonyultak hiperlipaemiának /a szérumzsir értékeik ugynevezett határesetet képeztek/, pedig a benyomásunk alapján felmerült a gyanu. Ezért is végeztük el a célzott vizsgálatot. Végül véletlenszerűen emeltünk ki 15 olyan beteget - férfit és nőt vegyesen, kor szerint hasonló eloszlásban - akiknél a zsirvérűség gyanuja fel sem merülhet. A részben analógiás, részben bináris adatokat a Központi Fizikai Kutató Intézet egyik TPAi kieszámítógépén futtatott ABCL jelű programmal dolgoztuk fel.

Az anyag ismeretében a munkahipotézist ugy is meg lehetne fogalmazni, hogy a sok egyéb adat ne zavarja össze a hozzávetőleg már ugy is ismert csoportokat.

A program háromféle bináris hasonlósági számítás és kétféle analógiás távolságfüggvény alapján számított, együttes hasonlósági mátrix elemein, három típusu összekapcsolási módszer szerint képes osztályba sorolni.

Az adataink feldolgozása során szerzett tapasztalunk az, hogy a furthest neighbour /a legtávolabbi szomszéd/ nevű szeparálási módszer szolgáltatja az orvosi szempontból a legelfogadhatóbb osztályozást. Ez az eljárás a legelütőbb tulajdonság alapján teszi a vizsgáltat abba a csoportba, ahova - jobb híján - leginkább illik.

Osztályozni két szempontból lehet. Egyszer csoportosíthatók a betegek adatai alapján maguk a változók és másodszor csoportosíthatók ugyanazon adatok alapján maguk a betegek. Előbb 13 analógiás változó csoportosítása történt meg az említett furthest neighbour eljárás szerint. Ilyen módon az egyváltozók egymással való összefüggése tűnik elő.

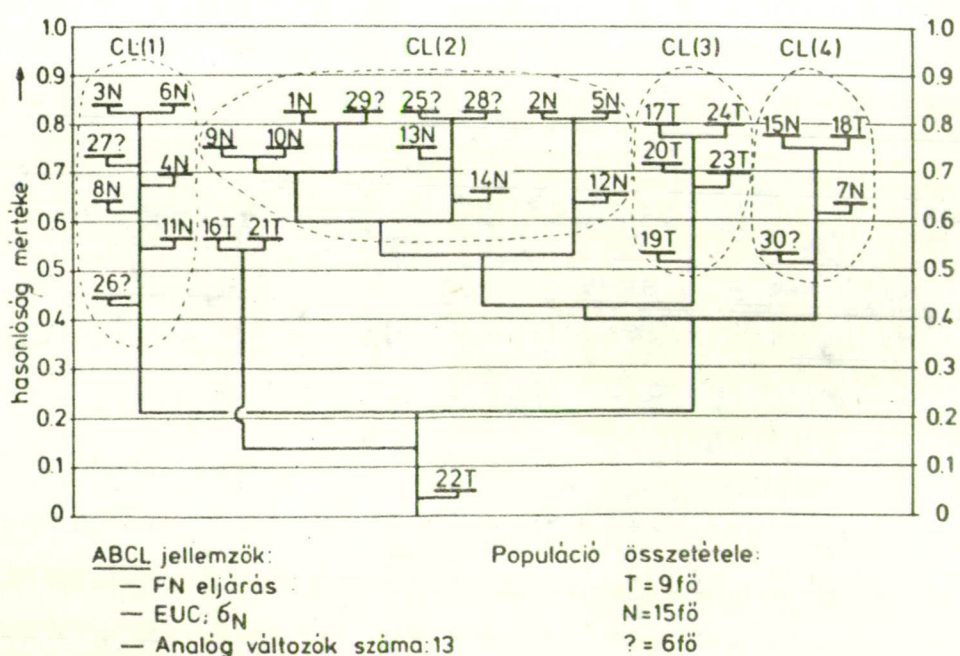


ABCL jellemzők:  
FN eljárás, változók szerint  
EUC;  $6_N$   
13analóg változó

Az rtg. ernyőképből nyert mellkas méretek ill. arányok nem csak egymással mutatnak szoros kapcsolatot, hanem a légzés szaporaságával is. A testméretek szoros összefüggésén túl pedig utóbbiakhoz társul az érverés szaporasága is. Ugyanakkor a különböző - a szérumszirra vonatkozó - laboratóriumi értékek az előbbiekkal és egymás között is sokkal lazább kapcsolatot mutatnak. /Megjegyzendő, hogy a szérumszirértékek eltérésén alapszik a különféle hiperlipaemia - típusok elkülönítése./

A csoportosítás éppen olyan mintha orvosi ismeretek alapján készült volna.

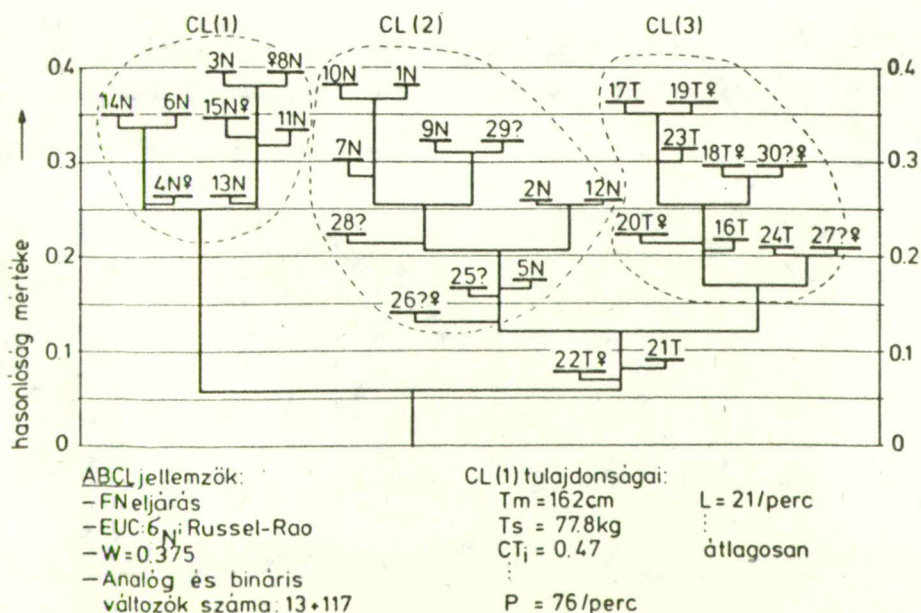
A betegek clusterezése ugyanezen módszer szerint két ízben történt meg. Először a 13 analógiás változó értékei alapján, majd másodsor ugyanezen adatok és a kérdőív bináris adatainak együttes értékelése révén került sor osztályozásra. Mindkettő eredményét ábrán mutatjuk be.



Az ábrán látható jelzések: a számjegy a paciens sor-száma, az utána látható irásjel értelme: T jelenti a hiperlipaemiás betegeket, ? a határeseteket és N az egyéb betegségben szenvedőket, akiknek szérumszir értéke normális. Ezen az ábrán tehát a 13 analógiás változó értékei szerint csoportosította az algoritmus a 30 beteget.

Négy csoport alakult ki, három eset pedig, szélsőséges értékei folytán csoporton kívül rekedt. Az első két csoport főleg normális és határeseteket foglal magába. A harmadik csak kóros esetekből áll, és azok nagyobb részét tartalmazza is. Az utolsó csoport vegyes és mindenféle van benne.

Ugyanezen pacienseket ezután az algoritmus újra csoportosította, de most már a kérdőív adatait - a bináris változókat is - figyelembe vette az osztályba sorolásakor. A feldolgozás eredménye a következő ábra.





Most három csoport alakult ki. Az első kettő normálisokat tartalmaz, a határesetek többsége a második csoportba társult be. Feltűnő, hogy míg az első csoport férfit-nőt vegyesen foglal magába, addig a második egy kivétellel férfiakból áll. Az összes hiperlipaemiás eset a harmadik csoportba került, kettő kivételével. Ezek a szélsőséges értékek folytán most is külön állanak. Az átlagértékek alapján három súlycsoportról /közép, félnehéz és nehézsúly/ lehetne szinte szó.

A kétféle csoportosítás tanulsága az, hogy az ilyen szempontból egyáltalán nem fajlagos kérdőív az osztályozást mégis javítani képes.

A munkahipotézis igazolódott: az általános vizsgálat adatai a különleges vizsgálat bizonyító erejét lényegesen növelik. Vagy más fogalmazásban: éppen a sok adat figyelembe vétele nem engedi összezavarni a csoportokat.

Ennek láttán egy analógia kínálkozik: az orvos a pácienssel beszélgetve tisztázza a kórelőzményeket és - látószólag intuitive - megsejt olyasmit is, amiről tulajdonképpen nem is volt szó. A cluster analízis bizonyos szempontból utánozza az orvos észjárását. Előbb felismeri a változók közti - az orvos szemében okinak tűnő - összefüggést, majd egy csomó általánosságból megsejti a különöset.

Az egészségügy egyik-másik területén fontos lehet ilyen módszer. Pl. a minket már régen érdeklő ugynevezett klinikumon belüli szűrővizsgálatban ígérkezik ez hatásos módszernek, amikor is - és ez a legfőbb erénye - bináris és analógiás adatokat vegyesen is képes kezelni ez az algoritmus.

#### I r o d a l o m

- [1] Baron D. N., Fraser P.: The digital computer in the classification and diagnosis of diseases, Lancet 1965/II 1066-1070.
- [2] Ghyczy K., B. Nagy A.: Kérdőívből operatív diagnózis, Neumann kollokvium Szeged 1975.
- [3] Ghyczy K., B. Nagy A.: Klinikai kérdőívek automatikus értékelése /eredmények/, Neumann kollokvium Szeged 1976.
- [4] Sokal R. R., Sneath P. A. H.: Principles of numerical taxonomy, Freeman London 1963.