# tanulmányok

160/1984





MTA Számitástechnikai és Automatizálási Kutató Intézet Budapest

A kiadásért felelős:
DR VÁMOS TIBOR

Főosztályvezető:
DEMETROVICS JÁNOS

ISBN 963 311 179 X
ISSN 0324 - 2951

SZÁMALK REPRO VGM 84/254

MAGYAR TUDOMÁNYOS AKADÉMIA
SZÁMITÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZETE

# VALOGATOTT SZÁMITASTECHNIKAI ES MATEMATIKAI MODSZEREK ORVOSI ALKALMAZASOKBAN

Kandidátusi disszertáció

Készitette:

Ratkó István

Tanulmányok 160/1984

# Tartalomjegyzék

# B E V E Z E T É S

A matematika orvosbiológiában történő alkalmazása nem ujkeletü. A valószinüségelmélet, valamint a matematikai statisztika elemei nélkül talán már nem is lehet uj kvantitativ eredményeket felmutatni. Természetesen napjainkban egyre bővül az alkalmazásra kerülő matematikai fejezetek száma. Igy pl. az elágazó sztochasztikus folyamatok, a Markov-féle sztochasztikus folyamatok, a sorbaállás elmélete, a statisztika nem-paraméteres módszerei, stb mind-mind ilyen területek.

Gnyegyenko [2] kitünő áttekintést ad ezekről az alkalmazási lehetőségekről. Következő megállapitásával csak egyet lehet érteni; "Semmi kétségem azonban afelől, hogy elkövetkezett az az idő, amikor matematikusok és biológusok kollektiváinak el kell kezdeniök a rendszeres együttes munkát, hogy megoldják a sarkalatos biológiai problémákat; ez olyan munka lesz, amelynek során a matematikusnak bele kell mélyednie a biológiai jelenségek lényegébe, a biológusnak pedig a matematikai módszerek alapgondolatai, s nem csak számolásokban való használhatósága által nyujtott lehetőségekbe."

Ez annál is fontosabb, mert a matematikai módszerek helytelen alkalmazása hamis eredményekre vezethet. Ilyen buktatókra mutatnak rá a statisztikai következtetésekkel kapcsolatban Vincze István [5] és A.J.Boreham [1]. Mark Kac [3] a matematika fontosságának hangsulyozása mellett már a számitógépek kiemelkedő szerepéről is szól.

Gnyegyenko emlitett megállapitását az együttműködés tekintetében a számitástechnika területére is kiterjeszthetjük.

Mindezek magyarországi felismeréseként hivta életre Kalmár László akadémikus a Szegeden rendszeres időközökben megrendezésre kerülő "Számitástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és biológiában" kollokviumot.

E dolgozat szerzője a hetvenes évek elején kapcsolódott be különböző orvosbiológiai kutatásokba. A kutatási feladatok megoldása során felvetődött matematikai és számitástechnikai problémák alátámasztják azt, amit D.E. Knuth [4] mond a matematika és számitástudomány kapcsolatáról; állandó kölcsönhatásban vannak egymással; összefüggések vannak közöttük a numerikus analizis, logika és számelmélet, stb. terén; a számitástudomány hatására a konstrukciók nagyobb hangsulyt kapnak a matematikában, stb.

A matematika és orvosbiológia is hat egymásra: matematikai eszközökkel orvosi (biológiai) következtetéseket vonhatunk le, ami aztán uj orvosi (biológiai) következtetés vagy kérdés feltevés alapja lehet; ugyanakkor orvosi (biológiai) kérdések megválaszolásához matematikai problémákat kell megoldani.

A dolgozatban a következő kérdésekről lesz szó: 1. logikai kifejezések használata programokban 2. Multifaktoriális kóreredetü betegség öröklődésének vizsgálata adott jellemző ismeretében. 3. Néhány statisztikai módszerről. 4. Betegség-regiszterek számitógép segitségével.

A tárgyalt kérdésekben elért eredmények hazai és külföldi konferenciákon ismertetésre kerültek, továbbá különböző folyóiratokban és egyéb kiadványokban megjelentek.

Ezeknek és a megfelelő hivatkozásoknak részletesebb ismertetése az egyes fejezetekben történik meg.

Hivatkozások

[1] Arthur John Boreham: How far should and could those who produce statistics engage in research and analysis?, 40th Session of the International Statistical Institute, Invited Paper, Warsaw, Sept, 1-9, 1975, pp 14/1-15.

[2] B.V. Gnyegyenko: A valószinüségelmélet bizonyos fejezeteiről, melyek közvetlen kapcsolatban vannak a biológiai és az orvostudomány problémáival, MTA III. Osztály közleményei, XV/2, pp. 165-173.

[3] Mark Kac: A mathematician looks at medicine, The american journal of medicine, Vol.66, May 1979, pp. 725-726.

[4] Donald E. Knuth: Computer Science and its relation to mathematics, The american mathematical monthly, 81/4. pp. 323-342.

[5] Vincze István: A statisztikai következtetés és korlátai, Magyar Tudomány, 11-12, 1981., pp. 902-912.

# I. LOGIKAI KIFEJEZÉSEK HASZNÁLATA PROGRAMOKBAN

Bármilyen tipusu adatfeldolgozásnál lépten-nyomon logikai kifejezéseket kell használnunk. "Lassu" gépnél vagy nagyméretü adatfeldolgozásnál már lehet annak szerepe, hogy milyen módon használjuk a logikai kifejezéseket. Ebben a fejezetben két ilyen tipusu feladat megoldását ismertetjük.

## 1. Logikai kifejezések kiértékelésének optimalizálása

A tárgyalt probléma egy nagyméretü adatfeldolgozás megoldása közben vetődött fel. [2o], [24].

Legyen adott egy adatfile, mely fix hosszuságu rekordokból áll. Az adatfile felhasználóját igen gyakran csak speciális logikai feltételeket kielégitő rekordok érdeklik. Más szóval azokról és csak azokról a rekordokról akar bizonyos táblázatokat, statisztikákat elkésziteni, amelyek valamilyen adott logikai feltételeket kielégitenek.

Nézzünk egy konkrét példát. A példa a kórházi morbiditási vizsgálatokban szereplő adatokon alapszik, de a könnyebb tárgyalhatóság kedvéért az ottani adatoknak csak egy részét szerepeltetjük. A részletes leirásra vonatkozóan ld [24].

### 1. Példa

Az adatfile rekordjainak a felépitése legyen a következő.

| karakter-pozició | a változó neve | tartalom |
|---|---|---|
| 1-2 | AMK | Az ápolást végző kórház megye kódja |
| 3-4 | AKK | Az ápolást végző kórház kódja az adott megyében |
| 5-6 | AOK | Az ápolást végző osztály kódja |
| 7-8 | AOF | Az osztály megjelölés további finomitása |
| 9-14 | TSZ | A beteg törzsszáma |
| 15-16 | SZE | A beteg születési évének utolsó két jegye (ha 19XY a feldolgozás éve, akkor az 18XY+1 előtt születetteknél XY+1 kódolandó) |
| 17-18 | SZH | A beteg születésének hónapja |
| 19-2o | SZN | A beteg születésének napja |
| 21-22 | ANK | A beteg anyja neve kezdőbetüjének sorszáma |
| 23 | NEM | A beteg neme (1=férfi, 2=nő) |
| 24-25 | FHO | A felvételi hónap sorszáma |
| 26-29 | NAP | Ápolási napok száma |
| 3o-31 | LMK | Állandó lakás megyekódja |
| 32-33 | LTJ | Állandó lakás település jellege |
| 34-37 | ILA | Az ideiglenes lakás adatai |
| 38-39 | FGA | Foglalkozási ágazatok |
| 4o | FGV | Foglalkozási viszony |
| 41-44 | BDK | Beutaló diagnózis kódszáma |
| 45-48 | AIK | Ápolást indokló fő kórisme kódszáma |

| karakter pozició | a változó neve | tartalom |
|---|---|---|
| 49-52<br>53-56<br>57-6o | KBK | Kisérő és következményes betegségek kódja |
| 61 | KSB | Kisérőbetegségek száma |
| 62-63 | FMN | A felvételtől a mütét időpontjáig eltelt napok száma |
| 64 | MSZ | Mütétek száma |
| 65 | BEI | A beutalás indoka (1=sürgős, 2=nem sürgős) |
| 66 | BEU | A beutaló (1=egyéb, 7=körzeti orvos 8=rendelő intézet, 9=kórházi áthelyezés) |
| 67-7o | HOK | A halál okának kódja |
| 71-74 | HAK | A halál okának alapbetegség kódja |

Az adatfile azon rekordjait akarjuk feldolgozni, amelyekre teljesülnek a következők:

a) A beutaló diagnózis, az ápolást indokló fő kórisme vagy a halál oka a szivvel kapcsolatos betegség volt (Reumás láz a sziv érintettségével: 391, Idült reumás szivbetegség: 393-398, Ischaemiás szivbetegség: 41o-414, A sziv veleszületett anomáliái: 745, A sziv-érrendszerre vonatkozó tünetek: 785). (A megadott kódszámok a BNO-Betegségek Nemzetközi Osztályozása-alapján vannak megadva.)

b) Település jellege: Bp (1), Megyei jogu város (2) v. egyéb város (3)

c) Foglalkozási ágazata: közszolgálat (8) v. nyugdijas-járadékos (9)

d) 3o és 5o év közötti szellemi foglalkozásu betegről van szó

e) Belgyógyászati (1), sebészeti (3), traumatológiai (4), intenziv (16), krónikus utókezelő (19) vagy szanatóriumi (2o) osztályon ápolt betegről van szó

A logikai feltétel valahogy igy irható:

```
IF ((BDK.EQ.391.OR.(BDK.LE.393.AND.BDK.GE.398).OR.
                    :
  BDK.EQ.785.OR.
          Ugyanaz, mint az eddigek, csak BDK helyett
          AIK-ra és HOK-ra
  HOK.EQ.785).AND.
  (LTJ.EQ.1.OR.LTJ.EQ.2.OR.LTJ.EQ.3).AND.
  (FGA.EQ.8.OR.FGA.EQ.9).AND.FGV.EQ.1.AND.
  (KOR.LE.50.AND.KOR.GE.30).AND.
  (AOK.EQ.1.OR.AOK.EQ.3.OR.AOK.EQ.4.OR.
   AOK.EQ.16.OR.AOK.EQ.19.OR.AOK.EQ.20))          ⊠
```

Megjegyzések:

1, A logikai kifejezés programba történő beépitésére adandó módszernek olyannak kell lennie, hogy egyrészt a felhasználó szemszögéből nézve minél egyszerübb legyen a használata, másrészt "független" legyen attól, hogy az egyes adatelemek kiválasztandó értékeinek a halmaza milyen strukturáju. Ez utóbbi példánkban pl. a következőt jelentheti: a beutaló diagnózis értékei között nincs két egymás utáni. Ekkor ez a feltételben igy jelentkezne:

$$BDK.EQ.d_1.OR.BDK.EQ.d_2......OR.BDK.EQ.d_n,$$

vagyis a felirt kifejezés konjunktiv normálformában lenne megadva.

2, Bár a logikai feltételt a FORTRAN programozási nyelven irtuk fel, a fejezetben mondottak függetlenek a nyelv megválasztásától.

3, Ennél jóval bonyolultabb feltétel is létezik, amikor is a felirt logikai kifejezés nem normálformában van megadva.

4, Az ebben a pontban tárgyalt problémának egy konkrét adatfeldolgozó rendszerbe történő beépitéséről a 2. pontban lesz szó.

5, Láthatjuk, hogy logikai kifejezésünknek nagyon sok tagja van, igy a hagyományos módon beépiteni a programba kényelmetlen, általános esetben reménytelen.

1.1. A modell

Álljon a rekord $N$ adatból. Jelölje $x_i$ az i-edik adat értékét és $H_i$ $x_i$ lehetséges értékeinek halmazát. $H_i$ lehet egy intervallum, de lehet bonyolultabb halmaz is.

Olyan rekordok kiválasztása a célunk, amelyek bizonyos logikai feltételeket kielégitenek. Ez alatt pontosabban a következőt értjük. Tegyük fel, hogy a logikai feltételekben $M$ adat szerepel, éspedig az $i_1$ -edik, $i_2$-edik,..., $i_M$-edik.

Könnyen látható, hogy bármilyen általános, az emlitett adatokra vonatkozó logikai kifejezés felirható a következő alakban:

$$\begin{aligned}&((x_{i_1} \in A_{1,1}) \wedge (x_{i_2} \in A_{1,2}) \wedge \ldots \wedge (x_{i_M} \in A_{1,M})) \vee \\ &\vee (x_{i_1} \in A_{2,1}) \wedge (x_{i_2} \in A_{2,2}) \wedge \ldots \wedge (x_{i_M} \in A_{2,M}) \vee \ldots \vee \\ &\vee (x_{i_1} \in A_{N,1}) \wedge (x_{i_2} \in A_{N,2}) \wedge \ldots \wedge (x_{i_M} \in A_{N,M})\end{aligned} \tag{1}$$

ahol $A_{k,j} \subseteq H_{i_j}$ $(j = 1,2,\ldots,M;\ k = 1,2,\ldots,N)$
de halmazegyenlőség semelyik rögzitett $k$ esetén sem lehet minden $j$ -re érvényes.

Az $(x_{i_j} \in A_{k,j})$ itélet is tulajdonképpen egy $|A_{k,j}|$ számu diszjunkciókból álló összetett itélet. A diszjunkciók

$$x_{i_j} .eq.\, z$$

alakuak, ahol $z$ végigfut $A_{k,j}$ elemein. ($|Y|$ az $Y$ halmaz elemeinek számát jelöli.) Nyilvánvaló az alábbi

<u>1. Lemma</u> Az (1) logikai kifejezésben szereplő elemi

itéletek száma

$$\sum_{j=1}^{M} \sum_{k=1}^{N} x_{k,j} \, |A_{k,j}|$$

ahol $x_{k,j} = \begin{cases} 1, \text{ ha} & A_{k,j} \subset H_{ij} \\ 0, \text{ ha} & A_{k,j} \equiv H_{ij} \end{cases}$ ⊠

2. példa

Határozzuk meg az 1. példa elemi itéleteinek számát (feltételezve, hogy az 1. megjegyzésben leirtaknak megfelelően a logikai feltételt konjunktiv normálformába átirtuk) ! ⊠

Az elemi itéletek száma:

(13+13+13)+3+2+21+1+5=76

## 1.2. Egy módszer

Redukáljuk az (1) logikai kifejezés elemi itéleteinek számát. Ezt a következő egyszerü ötlettel érhetjük el: definiáljuk a $Z_{k,j}(s)$ függvényt a következőképpen:

$$Z_{k,j}(s) = \begin{cases} 0, \text{ ha} & s \in A_{k,j} \\ 1, \text{ ha} & s \notin A_{k,j} \end{cases} \quad (s \in H_{ij}) \qquad (2)$$

Igy az $|A_{k,j}|$ számu $(x_{ij} \in A_{k,j})$ diszjunkció helyett egy egytagu logikai itélet használható:

$$Z_{k,j}(x_k).\text{eq}.\emptyset$$

Ugyanis $(x_k \in A_{k,j})$ akkor és csak akkor igaz, ha $Z_{k,j}(x_k).eq.\emptyset$ igaz.

3. példa

Határozzuk meg az 1. példához a $Z$ függvényeket ⊠

Elegendő az $A_{k,j}$ $(j=1,2,\dots,M;\ k=1,2,\dots,N)$ halmazokat megadni.

$A_{1,1}=\{391,393\text{-}398,410\text{-}414,745,785\}$

$A_{1,2}=\{391,393\text{-}398,410\text{-}414,745,785\}$

$A_{1,3}=\{391,393\text{-}398,410\text{-}414,745,785\}$

$A_{1,4}=\{1,2,3\}$ $A_{1,5}=\{8,9\}$ $A_{1,6}=\{1\}$

$A_{1,7}=\{30\text{-}50\}$ $A_{1,8}=\{1,3,4,16,19,20\}$

ahol ezek rendre a BDK, AIK, HOK, LTJ, FGA, FGV, KOR és AOK adatokra vonatkoznak. ⊠

Mint látjuk, ebben a példában (1) aránylag "egyszerübb" alaku, mivel N=1.

Nyilvánvaló igaz a

2. Lemma A $Z$ függvénnyel redukált logikai kifejezés elemi itéleteinek száma:

$$\sum_{j=1}^{M}\sum_{k=1}^{N} x_{k,j}$$ ⊠

Vegyük észre, hogy az (1) logikai kifejezést a (2) függvény segitségével tulajdonképpen diszjunktiv normálformára hoztuk.

Logikai kifejezésünket jelöljük $L$ -lel, az egyes diszjunkciókat $L_i$ -vel $(i=1,2,\dots,N)$ . Ekkor tehát

$$L = L_1 \vee L_2 \vee \dots \vee L_N$$ , ahol

$L_i$ konjunkciókból áll.

Nyilvánvaló, hogy az

```
IF(L.EQ.FALSE)GOTO2
```

utasitás "később" hajtódik végre, mint az

```
IF(L1.EQ.TRUE)GOTO 1
IF(L2.EQ.TRUE)GOTO 1
       :
IF(LN.EQ.TRUE)GOTO 1
GOTO2
1 ...
```

utasitáscsoport.

Ugyanis, mig az első esetben a számitógép minden diszjunkciót végig néz, addig a második esetben hamarabb befejeződik a kiértékelés.

Ha $L_i$ -ben az egyes konjunkciókat $L_{iq_i}$ -vel jelöljük, azaz ha

$$L_i = L_{i1} \wedge L_{i2} \wedge \dots \wedge L_{iq_i},$$

akkor az

```
      IF(L_i1.EQ.FALSE)GOTO 3
      IF(L_i2.EQ.FALSE)GOTO 3
              :
      IF(L_iq_i.EQ.FALSE)GOTO 3
      GOTO4
    3 ...
```

utasitáscsoport - hasonló indokok miatt - hamarabb hajtódik végre, mint az

$$\text{IF}(L_{i1}.\text{EQ}.\text{TRUE}.\text{AND}. \dots . L_{iq_i}.\text{EQ}.\text{TRUE})\text{GOTO4}$$

utasitás.

Természetes módon vetődik fel a következő probléma:

Legyen $m_1, m_2, \dots, m_N$ ill. $n_1, n_2, \dots, n_{q_i}$ az $1, 2, \dots, N$ számoknak egy permutációja $(i=1,2,\dots,N)$. Ha $L$ -ben a diszjunkciók sorrendje $L_{m_1}, L_{m_2}, \dots, L_{m_N}$ továbbá az $L_i$-ben a konjunkciók sorrendje $L_{in_1}, L_{in_2}, \dots, L_{in_{q_i}}$ akkor meghatározható a kifejezés kiértékelésének száma. A diszjunkciók és azokon belül az egyes konjunkciók milyen sorrendjénél lesz a kiértékelés száma a legkisebb?

A következő pontban precizen fogalmazzuk meg a problémát.

## 1.3. A probléma pontos megfogalmazása.

Vizsgáljuk az

$$L = L_1 \vee L_2 \vee \dots \vee L_N$$

logikai kifejezést, ahol

$$L_i = L_{i1} \wedge L_{i2} \wedge \dots \wedge L_{iq_i}$$

$$(i = 1, 2, \dots, N)$$

Definició A

$$\varrho(L) = \ell_1 + \ell_2 + \ldots + \ell_k + a_{k+1}$$

számot az $L$ logikai kifejezés kiértékelési számának nevezzük akkor, ha az $L_j$ diszjunkció $\ell_j$-edik konjunkciója hamis, az $1,2,\ldots,(\ell_j - 1)$ -edik konjunkciója igaz $(j=1,2,\ldots,k)$ , $L_{k+1}$ igaz és az $L_{k+1}$ konjunkcióinak száma $a_{k+1}$. Ha $L_1$ igaz, legyen $\varrho(L) = a_1$ ⊠

Nyilvánvaló, hogy $\varrho(L)$ függ a diszjunkciók sorrendjétől, illetve az egyes diszjunkción belüli konjunkciók sorrendjétől.

$\varrho(L)$ egy valószinüségi változó, hiszen (adott sorrendet feltételezve) értéke a véletlentől függ, rekordonként más és más.

Ebben a pontban $E\varrho(L)$ ,azaz $\varrho(L)$ várható értékének minimalizálásával foglalkozunk.

4. példa

Legyen $L = L_{11} \vee (L_{21} \wedge L_{22})$ és tegyük fel, hogy

$P(L_{21}.EQ.TRUE) \geq P(L_{22}.EQ.TRUE)$ és

$L_{11}.EQ.TRUE$, $L_{21}.EQ.TRUE$, $L_{22}.EQ.TRUE$ események függetlenek. Legyen továbbá $\tilde{L} = L_{11} \vee (L_{22} \wedge L_{21})$. $E\varrho(L)$ vagy $E\varrho(\tilde{L})$ a nagyobb? ⊠

Az $L_{11}.EQ.TRUE$, $L_{21}.EQ.TRUE$ és $L_{22}.EQ.TRUE$ eseményeket jelölje $B_i$ $(i=1,2,3)$. Ekkor nyilván

$$E\rho(L)= P(B_1B_2B_3)+3\,P(\bar B_1B_2B_3)+P(B_1\bar B_2B_3)+P(B_1B_2\bar B_3)+$$
$$+P(B_1\bar B_2\bar B_3)+3\,P(\bar B_1B_2\bar B_3)+2\,P(\bar B_1\bar B_2B_3)+2\,P(\bar B_1\bar B_2\bar B_3)$$

és
$$E\rho(\tilde L)= P(B_1B_2B_3)+3\,P(\bar B_1B_2B_3)+P(B_1\bar B_2B_3)+P(B_1B_2\bar B_3)+$$
$$+P(B_1\bar B_2\bar B_3)+2\,P(\bar B_1B_2\bar B_3)+3\,P(\bar B_1\bar B_2B_3)+2\,P(\bar B_1\bar B_2\bar B_3)$$

Ebből következik, hogy

$E\rho(L)\leqq E\rho(\tilde L)$ akkor és csak akkor, ha

$$P(\bar B_1B_2\bar B_3)\leqq P(\bar B_1\bar B_2B_3)$$

A függetlenségi feltétel miatt ez akkor és csak akkor teljesül, ha $(P(B_1)\neq 1)$

$P(B_2)\,P(\bar B_3)\leqq P(\bar B_2)\,P(B_3)$ azaz ha

$$(1-P(B_3))\,P(B_2)\leqq P(B_3)(1-P(B_2))$$

Ez pedig a $P(B_2)\leqq P(B_3)$ egyenlőtlenséggel ekvivalens, ami azt jelenti, hogy

$$E\rho(L)\leqq E\rho(\tilde L) \qquad \boxtimes$$

Szemléletesen ezt ugy mondhatjuk, hogy ha az $L_{21}$ és $L_{22}$ elemi itéletek sorrendjét megcseréljük, akkor $L$ kiértékelése hamarabb befejeződik, mint $\tilde L$ kiértékelése.

Ez általában nem igaz. Ezt mutatja az

5. példa

Az előbbi példa jelöléseit megtartva most a következőket tudjuk:

$$P(B_1 B_2 B_3) = 1/8 - p \qquad P(\bar{B}_1 B_2 B_3) = 1/8 - p$$
$$P(B_1 \bar{B}_2 B_3) = 1/8 + 2p \qquad P(B_1 B_2 \bar{B}_3) = 1/8$$
$$P(B_1 \bar{B}_2 \bar{B}_3) = 1/8 - p \qquad P(\bar{B}_1 B_2 \bar{B}_3) = 1/8 + p$$
$$P(\bar{B}_1 \bar{B}_2 B_3) = 1/8 \qquad P(\bar{B}_1 \bar{B}_2 \bar{B}_3) = 1/8$$

ahol $p < 1/8$

Mit tudunk mondani $E_\rho(L)$ és $E_\rho(\tilde{L})$ viszonyáról? ⊠

Most is igaz, hogy

$E_\rho(L) \leqq E_\rho(\tilde{L})$ akkor és csak akkor, ha

$$P(\bar{B}_1 B_2 \bar{B}_3) \leqq P(\bar{B}_1 \bar{B}_2 B_3)$$

Mivel a feltételek miatt

$$P(\bar{B}_1 B_2 \bar{B}_3) = 1/8 + p \quad \text{és} \quad P(\bar{B}_1 \bar{B}_2 B_3) = 1/8$$

ezért

$$E_\rho(L) > E_\rho(\tilde{L})$$ ⊠

A példák alapján láthatjuk, hogy két egymás melletti elemi itélet cseréje a kiértékelési szám várható értékét csökkentheti is, növelheti is.

Vizsgáljuk meg ezek után általánosan a kérdést. Erre vonatkozik az

<u>1. tétel</u> Legyen

$$L = L_1 \vee L_2 \vee \ldots \vee L_j \vee (\ldots \wedge L' \wedge L'' \wedge \ldots) \vee \ldots \vee L_N \quad \text{és}$$
$$\tilde{L} = L_1 \vee L_2 \vee \ldots \vee L_j \vee (\ldots \wedge L'' \wedge L' \wedge \ldots) \vee \ldots \vee L_N$$

ahol $L'$ és $L''$ elemi itéletek, $L_1, L_2, \ldots, L_N$ pedig elemi itéletek konjunkciója. Ekkor

$E_\rho(L) \leqq E_\rho(\tilde{L})$ akkor és csak akkor, ha

$P(L'=i \mid A) \leqq P(L''=i \mid A)$, ahol
$A=\{ L_k$ hamis $(k=1,2,\dots,j)$, $L_{j+1}$-ben az $L'$ előtti elemi
itéletek igazak$\}$ ⊠

<u>Bizonyitás</u>:

Vezessük be a következő jelöléseket:

$\xi$ : $L$ kiértékelésekor $L'$-ig ( azt nem beleértve) elvégzett kiértékelések száma

$\tilde{\xi}$ : $\tilde{L}$ kiértékelésekor $L''$-ig (azt nem beleértve) elvégzett kiértékelések száma

$\eta$ : $L$ kiértékelésekor $L''$ után ( azt nem beleértve ) elvégzett kiértékelések száma

$\tilde{\eta}$ : $\tilde{L}$ kiértékelésekor $L''$ után (azt nem beleértve) elvégzett kiértékelések száma

Természetesen $\eta=0$ illetve $\tilde{\eta}=0$ ha $L$ ill. $\tilde{L}$ kiértékelésekor az $L'$ ill. $L''$ utáni elemi itéletekre nem "kerül sor", mivel a kiértékelés már hamarabb befejeződött.

Nyilván

$$\rho(L)=\xi+\rho(L' \wedge L'')\cdot \chi_A+\eta \quad \text{és}$$
$$\rho(\tilde{L})=\tilde{\xi}+\rho(L'' \wedge L')\cdot \chi_A+\tilde{\eta} \quad \text{, ahol}$$

$\chi_A$ az $A$ esemény karakterisztikus változója.

$E\rho(L) \leqq E\rho(\tilde{L})$ akkor és csak akkor, ha
$$E[\rho(L' \wedge L'')\chi_A] \leqq E[\rho(L'' \wedge L')\chi_A]$$
ugyanis $E\xi=E\tilde{\xi}$ és $E\eta=E\tilde{\eta}$

Könnyen látható, hogy

$$E[\rho(L' \wedge L'')\chi_A] = 2P(L'=i, L''=i, A) + 2P(L'=i, L''=h, A) +$$

$$+ P(L'=h, L''=i, A) + P(L'=h, L''=h, A) \qquad \text{és}$$

$$E[\rho(L'' \wedge L')\chi_A] = 2P(L'=i, L''=i, A) + P(L'=i, L''=h, A) +$$

$$+ 2P(L'=h, L''=i, A) + P(L'=h, L''=h, A)$$

Ezért azt kapjuk, hogy

$E\rho(L) \leqq E\rho(\tilde{L})$ akkor és csak akkor, ha

$$P(L'=i, L''=h, A) \leqq P(L'=h, L''=i, A)$$

Elemi ekvivalens átalakitásokkal

$$P(L'=i, L''=h \mid A) \leqq P(L'=h, L''=i \mid A)$$

$$P(L'=i \mid A) - P(L'=i, L''=i \mid A) \leqq P(L''=i \mid A) - P(L'=i, L''=i \mid A)$$

azaz

$$P(L'=i \mid A) \leqq P(L''=i \mid A) \qquad \boxtimes$$

Azt mondjuk, hogy az $L'$ és $L''$ elemi itélet <u>független</u> egymástól, ha az $\{L'=i\}$ és $\{L''=i\}$ események függetlenek.

Két diszjunkció akkor <u>független</u> egymástól, ha bármely két olyan elemi itélet független egymástól, melyek egyike az egyik, másika a másik diszjunkcióhoz tartozik.

A következőkben megvizsgálunk néhány speciális esetet.

<u>1. Következmény</u>

Tegyük fel, hogy az L kifejezés bármelyik két itélete független egymástól. Ekkor

$$E_\rho(L) \leqq E_\rho(\tilde{L}) \quad \text{akkor és csak akkor, ha} \quad P(L' = i) \leqq P(L'' = i) \quad \boxtimes$$

Ez az 1. tétel közvetlen következménye.

2. Következmény

Tételezzük fel, hogy bármely két diszjunkció független egymástól. Legyen $B$ az az esemény, hogy $L_{j+1}$ -ben az $L'$ előtti elemi itéletek igazak. ( ld. 1. tétel jelöléseit).

Ebben az esetben

$$E_\rho(L) \leqq E_\rho(\tilde{L}) \quad \text{akkor és csak akkor, ha} \quad P(L' = i \mid B) \leqq P(L'' = i \mid B) \quad \boxtimes$$

Bizonyitás

Jelölje $A_k$ azt az eseményt, hogy $L_k$ hamis $(k = 1, 2, \dots, j)$

Ekkor $A = A_1 A_2 \dots A_j B$ miatt elemi ekvivalens átalakitásokkal azt kapjuk, hogy

$$P(L' = i \mid A) = \frac{P(L' = i, A_1 A_2 \dots A_j B)}{P(A_1 A_2 \dots A_j B)} = \frac{P(L' = i, B)\, P(A_1 A_2 \dots A_j)}{P(A_1 A_2 \dots A_j B)}$$

s igy

$$P(L' = i \mid A) \leqq P(L'' = i \mid A) \quad \text{akkor és}$$

csak akkor igaz, ha

$$P(L' = i \mid B) \leqq P(L'' = i \mid B) \quad \boxtimes$$

Vizsgáljuk meg ezek után, hogy két egymás melletti diszjunkció cseréje esetén mit mondhatunk a kiértékelési számokról? Nézzünk először egy példát.

6. Példa

Legyen

$$L = L_{11} \vee (L_{21} \wedge L_{22}), \qquad \tilde{L} = (L_{21} \wedge L_{22}) \vee L_{11}$$

és tegyük fel, hogy a

$B_1 = \{L_{11} = i\}$, $B_2 = \{L_{21} = i\}$, $B_3 = \{L_{22} = i\}$ események függetlenek.

Hasonlitsuk össze

$E\rho(L)$ és $E\rho(\tilde{L})$ értékét! ⊠

Mivel

$$E\rho(L) = P(B_1 B_2 B_3) + P(\bar{B}_1 B_2 B_3) + P(B_1 \bar{B}_2 B_3) + 2P(B_1 B_2 \bar{B}_3) + \\ + 2P(B_1 \bar{B}_2 \bar{B}_3) + 3P(\bar{B}_1 B_2 \bar{B}_3) + P(\bar{B}_1 \bar{B}_2 B_3) + 3P(\bar{B}_1 \bar{B}_2 \bar{B}_3)$$ és

$$E\rho(\tilde{L}) = 2P(B_1 B_2 B_3) + 2P(\bar{B}_1 B_2 B_3) + 3P(B_1 \bar{B}_2 B_3) + 2P(B_1 B_2 \bar{B}_3) + \\ + 3P(B_1 \bar{B}_2 \bar{B}_3) + 2P(\bar{B}_1 B_2 \bar{B}_3) + 2P(\bar{B}_1 \bar{B}_2 B_3) + 2P(\bar{B}_1 \bar{B}_2 \bar{B}_3)$$

Ezért $E\rho(L) \leqq E\rho(\tilde{L})$ akkor és csak akkor, ha

$$P(\bar{B}_1 B_2 \bar{B}_3) + P(\bar{B}_1 \bar{B}_2 \bar{B}_3) \leqq P(B_1 B_2 B_3) + P(\bar{B}_1 B_2 B_3) + 2P(B_1 \bar{B}_2 B_3) + \\ + P(B_1 \bar{B}_2 \bar{B}_3) + P(\bar{B}_1 \bar{B}_2 B_3)$$

A függetlenség miatt elemi átalakitásokkal következik, hogy ez ekvivalens a

$$\frac{1 - 2P(B_3)}{2 - P(B_2) - P(B_3)} \leqq P(B_1)$$

egyenlőtlenséggel.

Ha $P(B_3) \geqq 1/2$, akkor ez nyilván teljesül függetlenül attól, hogy a

$P(L_{11}=i)=P(B_1)$ és a $P(L_{21}\wedge L_{22}=i)=P(B_2B_3)$ valószinüségek milyen nagyságrendi viszonyban állnak egymással. ⊠

Anélkül, hogy a legutóbbi egyenlőtlenséget részletesen elemeznénk, látható, hogy két egymásmelletti diszjunkció cseréje ilyen egyszerü feltételezés mellett is "változatosabb képet mutat", mint két egymásmelletti konjunkció cseréje.

Nézzük meg általánosabban a kérdést.

Legyen $L=L_1\vee L_2\vee\dots\vee L_{j-1}\vee L_j\vee L_{j+1}\vee L_{j+2}\vee\dots\vee L_N,$

$$\tilde{L}=L_1\vee L_2\vee\dots\vee L_{j-1}\vee L_{j+1}\vee L_j\vee L_{j+2}\vee\dots\vee L_N \quad \text{és}$$

$$A=\{L_1=h,\ L_2=h,\dots,\ L_{j-1}=h\}$$

2. tétel

$E\rho(L)\leqq E\rho(\tilde{L})$ akkor és csak akkor, ha

$$a_j P(L_j=i,\ L_{j+1}=i,\ A)+E[\rho(L_j)\mid L_j=h,L_{j+1}=i,\ A]\cdot$$
$$\cdot P(L_j=h,\ L_{j+1}=i,\ A)\leqq$$
$$a_{j+1}\ P(L_j=i,\ L_{j+1}=i,\ A)+E[\rho(L_{j+1})\mid L_j=i,L_{j+1}=h,A]\cdot$$
$$\cdot P(L_j=i,\ L_{j+1}=h),$$

ahol

$a_k$ az $L_k$ konjunkcióinak a száma. ⊠

Bizonyitás

Vezessük be a következő jelöléseket:

$\xi$ : $L$ kiértékelésekor $L_j$-ig (azt nem beleértve) elvégzett kiértékelések száma

$\tilde{\xi}$ : $\tilde{L}$ kiértékelésekor $L_{j+1}$-ig (azt nem beleértve) elvégzett kiértékelések száma

$\eta$ : $L$ kiértékelésekor $L_{j+1}$ után (azt nem beleértve) elvégzett kiértékelések száma.

$\tilde{\eta}$ : $\tilde{L}$ kiértékelésekor $L_j$ után (azt nem beleértve) elvégzett kiértékelések száma

Természetesen $\eta=0$ illetve $\tilde{\eta}=0$, ha $L$ illetve $\tilde{L}$ kiértékelésekor az $L_{j+1}$ illetve $L_j$ utáni elemi itéletekre nem kerül sor, mivel a kiértékelés már hamarabb befejeződött.

Nyilván

$$\rho(L) = \xi + \rho(L_j \vee L_{j+1})\chi_A + \eta$$
$$\rho(\tilde{L}) = \tilde{\xi} + \rho(L_{j+1} \vee L_j)\chi_A + \tilde{\eta}$$

ahol $\chi_A$ az $A$ esemény karakterisztikus változója.

Mivel $E\xi = E\tilde{\xi}$ és $E\eta = E\tilde{\eta}$, ezért

$E\rho(L) \leq E\rho(\tilde{L})$ akkor és csak akkor,

ha

$$E[\rho(L_j \vee L_{j+1})\chi_A] \leq E[\rho(L_{j+1} \vee L_j)\chi_A] \qquad (3)$$

Az egyenlőtlenség bal oldala igy irható fel:

$$E[\varrho(L_j \vee L_{j+1}) \mid L_j = i,\, L_{j+1} = i,\, A]\, P(L_j = i,\, L_{j+1} = i,\, A) +$$

$$+ E[\varrho(L_j \vee L_{j+1}) \mid L_j = h,\, L_{j+1} = h,\, A]\, P(L_j = h,\, L_{j+1} = h,\, A) +$$

$$+ E[\varrho(L_j \vee L_{j+1}) \mid L_j = i,\, L_{j+1} = h,\, A]\, P(L_j = i,\, L_{j+1} = h,\, A) +$$

$$+ E[\varrho(L_j \vee L_{j+1}) \mid L_j = h,\, L_{j+1} = i,\, A]\, P(L_j = h,\, L_{j+1} = i,\, A) =$$

$$= a_j\, P(L_j = i,\, L_{j+1} = i,\, A) + E[\varrho(L_j \vee L_{j+1}) \mid L_j = h,\, L_{j+1} = h,\, A) \cdot$$

$$\cdot\, P(L_j = h,\, L_{j+1} = h,\, A) + a_j\, P(L_j = i,\, L_{j+1} = h,\, A) +$$

$$+ E[\varrho(L_j) \mid L_j = h,\, L_{j+1} = i,\, A]\, P(L_j = h,\, L_{j+1} = i,\, A) +$$

$$+ a_{j+1}\, P(L_j = h,\, L_{j+1} = i,\, A)$$

Hasonlóan felirható az egyenlőtlenség jobboldala. A kapott két kifejezést összevetve azt kapjuk, hogy (3) akkor és csak akkor igaz, ha

$$a_j P(L_j=i, L_{j+1}=i, A)+E[\varrho(L_j)|L_j=h, L_{j+1}=i, A]\cdot P(L_j=h, L_{j+1}=i, A) \leqq$$
$$\leqq a_{j+1} P(L_j=i, L_{j+1}=i, A)+E[\varrho(L_{j+1})|L_j=i, L_{j+1}=h, A] P(L_j=i, L_{j+1}=h, A),$$

s épp ezt kellett bizonyitani.

Vizsgáljunk meg egy speciális esetet.

3. Következmény

Tegyük fel, hogy az $L$ kifejezés bármelyik két diszjunkciója független egymástól. Ekkor

$$E\varrho(L) \leqq E\varrho(\tilde{L}) \quad \text{akkor és csak akkor, ha}$$

$$\frac{E\varrho(L_j)}{P(L_j=h)} \leqq \frac{E\varrho(L_{j+1})}{P(L_{j+1}=h)} \qquad \boxtimes$$

Bizonyitás

A függetlenségi feltétel miatt (3) -ból azonnal következik, hogy

$$E\varrho(L) \leqq E\varrho(\tilde{L}) \quad \text{akkor és csak akkor,}$$

ha $$E\varrho(L_j \vee L_{j+1}) \leqq E\varrho(L_{j+1} \vee L_j) \tag{4}$$

Mivel $\varrho(L_j \vee L_{j+1}) = \varrho(L_j)+\varrho(L_{j+1})\chi_{L_j=i}$ és

$$\varrho(L_{j+1} \vee L_j) = \varrho(L_{j+1})+\varrho(L_j)\chi_{L_{j+1}=i},$$

a függetlenséget kihasználva

következik, hogy (4) akkor és csak akkor igaz, ha

$$E\varrho(L_j)+E\varrho(L_{j+1})P(L_j=i) \geqq E\varrho(L_{j+1})+E\varrho(L_j)P(L_{j+1}=i)$$

Ennek átrendezésével adódik a bizonyitandó állitás. ⊠

A következőkben az 1.2. pont végén megfogalmazott kérdésre próbálunk választ adni a bizonyitott tételek és következmények alapján.

3. tétel

Legyen adott az

$$L = L_1 \vee L_2 \vee \dots \vee L_N$$

logikai kifejezés, ahol

$$L_j = L_{j1} \wedge L_{j2} \wedge \dots \wedge L_{jq_j} \quad (j = 1, 2, \dots, N)$$

Tegyük fel, hogy $L$ bármelyik két elemi itélete független egymástól.

Az egyszerüség kedvéért tételezzük fel a következőt: az indexeket ugy választottuk meg, hogy

$$\frac{E\rho(L_1)}{P(L_1 = h)} \leqq \frac{E\rho(L_2)}{P(L_2 = h)} \leqq \dots \leqq \frac{E\rho(L_N)}{P(L_N = h)} \quad \text{és} \qquad (5a)$$

$$P(L_{j1} = i) \leqq P(L_{j2} = i) \leqq \dots \leqq P(L_{jq_j}) \quad (j = 1, 2, \dots, N) \qquad (5b)$$

Legyen $\tilde{L}$ egy tetszőleges olyan logikai kifejezés, amely $L$ -ből konjunkciók ill. diszjunkciók már emlitett módon történő cseréjével keletkezett. Ekkor

$$E\rho(L) \leqq E\rho(\tilde{L}) \quad \text{, azaz}$$

$E\rho(L)$ a diszjunkciók ill. a konjunkciók olyan sorrendjére lesz minimális, amelyre (5a) és (5b) teljesül. ⊠

Bizonyitás

A tétel az 1. és 3. következmény alapján közvetlenül adódik, felhasználva azt a nyilvánvaló tényt, hogy az

$$\frac{E_\rho(L_j)}{P(L_j=h)}$$

kifejezés értéke ugyancsak az 1. következmény miatt az (5b)-beli sorrendnek megfelelően változik. Ez pontosabban a következőt jelenti: ha

$$L_j = \ldots \wedge L' \wedge L'' \wedge \ldots ,$$

$$\tilde{L}_j = \ldots \wedge L'' \wedge L' \wedge \ldots ,$$

$$P(L'=i) \leqq P(L''=i)$$ akkor

$$\frac{E_\rho(L_j)}{P(L_j=h)} \leqq \frac{E_\rho(\tilde{L}_j)}{P(\tilde{L}_j=h)}$$

hiszen $P(L_j=h) = P(\tilde{L}_j=h)$

Konkrét példákat a 2. pontban veszünk.

Az ismertetett módszert, a javasolt megoldást több helyen ismertettük, publikáltuk, igy [1o], [11], [12], [13], [16] és [17]-ben a részeredményekről, [18]-ban a végső stádiumról számoltunk be. [19] módszerünk szükségességét - érintőlegesen - hangsulyozza.

1.4. <u>Irodalmi áttekintés</u>

Ebben a fejezetben megpróbáljuk áttekinteni, hogy az általunk javasolt optimalizálási módszer, pontosabban a modellünk hogyan viszonyul az ismert modellekhez és módszerekhez.

Ebből a célból az alábbiakra szoritkozunk: a) döntési táblázatokhoz b) Boole-függvényrendszerek minimalizálásához

c) döntéshozatal analiziséhez való kapcsolatok vizsgálata

a) <u>Döntési táblázatokkal való kapcsolat</u>

A kapcsolat megvizsgálása előtt tekintsünk át néhány alapfogalmat.

A döntési táblázat információs rendszerek leirását elősegitő módszer. A döntési táblázat téglalap alaku táblázatban tartalmazza az információkat:

| | 1 | 2 | 3 | . . . | K |
|---|---|---|---|---|---|
| F1 | Y | N | Y | . . . | – |
| F2 | Y | – | N | . . . | Y |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ ⋮ ⋮ | ⋮ |
| FN | N | Y | Y | . . . | Y |
| T1 | X | – | X | . . . | X |
| T2 | – | X | – | . . . | – |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ ⋮ ⋮ | ⋮ |
| TM | X | – | X | . . . | – |

szabály számok

bejegyzés

szabály

Az F1,F2,...,FN különböző <u>feltételeket</u>, a T1,T2,..,TM különböző <u>tevékenységeket</u> jelölnek. A tevékenységek végrehajtása vagy végre nem hajtása a feltételek aktuális állapotától illetve a <u>bejegyzés</u>ektől függ.

Például a 3. szabály jelentése: ha F1=Y (azaz az F1 feltétel teljesül) <u>és</u> F2=N (F2 nem teljesül), <u>és</u> ... <u>és</u> FN=Y akkor hajtsd végre a T1 és TM tevékenységeket.

Meg kell jegyezni, hogy a feltételek és a szabályok sorrendje nem feltétlenül kötött, a tevékenységek felsorolásánál azonban az előbb végrehajtandó tevékenységnek "feljebb" kell szerepelnie, mint a később végrehajtandónak.

Nem térünk ki a döntési táblázatok legáltalánosabb definiciójára, azok különböző osztályozásaira. Csak annyit említünk meg, amennyi a döntési táblázatok és a modellünk összevetéséhez szükséges.

Az 1.1. pont jelöléseit megtartva, az 1. fejezetben tárgyalt probléma döntési táblázattal a következőképpen fogalmazható meg:
(a feltételek ÉS, a szabályok KIZÁRÓ VAGY kapcsolatban állnak)

| | SZ1 | SZ2 | | SZN | EGYÉB |
|---|---|---|---|---|---|
| $X_1$ | =... | =... | ... | =... | |
| $X_2$ | =... | =... | ... | =... | |
| : | | | | | |
| $X_N$ | =... | =... | ... | =... | |
| Szükség van a rekordra | X | X | ... | X | - |
| Nincs szükség a rekordra | - | - | ... | - | X |

Azonnal észrevehető, hogy ez egy speciális döntési táblázat, mivel az EGYÉB szabály kivételével azonos tevékenységet kell végezni.

A döntési táblázatok optimalizálásának három fázisa (szabályok összevonása, táblázat részekre bontása, a táblázat átrendezése) közül a harmadik mutat "rokonságot" az 1. pont problémájával, annak is az alábbi szempontjai: minimális számu döntéssel lehessen megtalálni egy keresett szabályt; a szabályok táblázatbeli sorrendje feleljen meg az egyes szabályok alkalmazási gyakoriságának. A különböző, döntési táblázatokkal kapcsolatos optimalizáló algoritmusok, pl. [6], [9], [21] többek között ennek elérését szolgálják.

A mondottakból nyilvánvaló, hogy mást optimalizálunk a döntési táblázatoknál és mást az 1. pont problémájánál.

b) <u>Boole-függvényrendszerek minimalizálásával való kapcsolat</u>

Boole-függvények megadásával és alapvető tulajdonságaival foglalkozik [2] és [3]. Boole-függvények kezelésére [7]-ben és [8] -ban található programrendszer.

Abból a célból, hogy rá tudjunk mutatni az ezekben található módszerek és az általunk használt módszer közötti különbségre, néhány fogalmat kell bevezetnünk.

Az $f(x_1, x_2, \ldots, x_n)$ $(x_i \in \{0,1\},\ i=1,2,\ldots,n)$ $\{0,1\}$ értékkészletü függvényt <u>n változós Boole-függvénynek</u> nevezzük.

$x_1$ és $x_2$ konjunkcióját $x_1 \wedge x_2$ (v. röviden $x_1 x_2$), diszjunkcióját $x_1 \vee x_2$, $x$ negáltját $\overline{x}$ jelöli; értelmezésük nyilvánvaló.

Az $x_{i_1}^{k_1} x_{i_2}^{k_2} \dots x_{i_r}^{k_r}$ formulát az $X^n = \{x_1, x_2, \dots, x_n\}$ halmaz feletti konjunkciónak nevezzük, ahol $k_j \in \{0,1\}$, $x_{i_j}^0 = \overline{x}_{i_j}$, $x_{i_j}^1 = x_{i_j}$, $i_j \in \{1,2,\dots,n\}$ minden $j = 1,2,\dots,r$ esetén.

Hasonlóan értelmezhető a diszjunkció is.

A konjunkció (diszjunkció) <u>elemi</u>, ha $j \neq l$ esetén $x_{i_j} \neq x_{i_l}$. Az $x_{i_j}$ alaku kifejezéseket <u>betüknek</u> fogjuk nevezni.

A
$$D = K_1 \vee K_2 \vee \dots \vee K_s$$
formuláról azt mondjuk, hogy <u>diszjunktiv normálforma</u>, ha $K_j\ (j=1,2,\dots,s)$ $X^n$ feletti elemi konjunkció. (A konjunktiv normálforma is értelemszerüen definiálható.) A <u>s</u> szám a diszjunktiv normálforma <u>hossza</u>.

A diszjunktiv normálforma <u>minimális</u>, ha a vele ekvivalens diszjunktiv normálformák közül legkevesebb betüt tartalmaz; <u>legrövidebb</u>, ha a hossza a vele ekvivalens diszjunktiv normálformák közül a legkisebb.

[2], [3], [6], [7] és [8] az előbb definiált minimalizálási eljárásokkal foglalkozik. Teljesen nyilvánvaló, hogy ezen minimalizálási feladatoknak <u>más</u> a céljuk, igy az ott kapott eredmények esetünkre nem alkalmazhatók.

c) Döntéselemzéssel való kapcsolat

A döntési modellekről [4] ad kitünő áttekintést. Saját eredményeinkkel való összevetés miatt röviden ismertetjük a legfontosabb alapfogalmakat.

Tegyük fel, hogy a döntéshozó az $S_1, S_2, \ldots, S_t$ stratégiák közül választhat. Bármelyiket is választja azonban, a $K_1, K_2, \ldots, K_r$ következmények valamelyikével kell számolnia. Az alábbi valószinüségeket

$$p_{i,j} = P(K_j \mid S_i)$$

ismertnek tételezzük fel, jelentésük nyilvánvaló. Természetesen

$$\sum_{i=1}^{r} p_{i,j} = 1 \qquad \text{minden } j \text{ esetén} .$$

Létezik egy a következmények halmazán értelmezett valós értékü $f$ függvény, amelyre: $f(K_j) \gtreqless f(K_\ell)$ akkor és csak akkor, ha a döntést végző a $K_j$ következményt nagyobb fontosságunak, azonos fontosságunak, kisebb fontosságunak tartja, mint $K_\ell$-t. Az $f$ függvény a lineáris transzformáció erejéig egyértelmüen meg van határozva. Ezen azt értjük, hogy ha $f$ és $g$ két a fenti tulajdonságnak eleget tevő függvény, akkor léteznek olyan $a$ és $b$ valós számok, melyekre

$$g(K_j) = af(K_j) + b \quad (j = 1, 2, \ldots, r)$$

A következő kérdést vizsgáljuk: a döntéshozó melyik stratégiát válassza, hogy döntése valamilyen értelemben optimális legyen. Nyilvánvalóan alapvető az, hogy mit értsünk optimális alatt.

i) Azt az $S_i$ stratégiát részesitsük előnyben, amelyikre

$$\sum_{j=1}^{r} p_{i,j} f(K_j) \quad \text{maximális.}$$

ii) Tegyük fel még a következőt is: léteznek olyan $E_1 \geq E_2 \geq \ldots E_n$ állapotok, (események) melyek nem befolyásolhatók vagy módosithatók a választott stratégiával, továbbá, hogy mindegyikük bekövetkezésének adott a $p_j = P(E_j)$ valószinüsége, és

$$\sum_{j=1}^{n} P(E_j) = 1$$

A következmények száma most $nt$, a stratégiák, állapotok és következmények viszonyát az alábbi ábra mutatja

| | $E_1$ | $E_2$ | $\cdots$ | $E_n$ |
|---|---|---|---|---|
| $S_1$ | $K_1$ | $K_2$ | $\cdots$ | $K_n$ |
| $S_2$ | $K_{n+1}$ | $K_{n+2}$ | $\cdots$ | $K_{2n}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ |
| $S_t$ | $K_{n(t-1)+1}$ | | $\cdots$ | $K_{nt}$ |

A $(p_{i,j})$ mátrix a következő:

| | $K_1$ | $K_2$ | $\cdots$ | $K_n$ | $K_{n+1}$ | $K_{n+2}$ | $\cdots$ | $K_{2n}$ | $\cdots$ | $K_{n(t-1)+1}$ | $\cdots$ | $K_{nt}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $S_1$ | $p_1$ | $p_2$ | $\cdots$ | $p_n$ | 0 | 0 | $\cdots$ | 0 | $\cdots$ | 0 | $\cdots$ | 0 |
| $S_2$ | 0 | 0 | $\cdots$ | 0 | $p_1$ | $p_2$ | $\cdots$ | $p_n$ | $\cdots$ | 0 | $\cdots$ | 0 |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | | $\vdots$ |
| $S_t$ | 0 | 0 | $\cdots$ | 0 | 0 | 0 | $\cdots$ | 0 | $\cdots$ | $p_1$ | $\cdots$ | $p_n$ |

Most is azt a stratégiát részesitjük előnyben, amelyikre

$$\sum_{j=1}^{nt} p_{i,j} f(K_j) \qquad \text{maximális.}$$

Nézzük meg ezek után, hogy saját vizsgálódásaink hogyan kapcsolhatók össze az (i) modellel.

Az $S_1, S_2, \ldots, S_t$ stratégiák most megfelelnek a diszjunkciók és azon belül a konjunkciók egy-egy rögzitett sorrendjének. Ha a logikai kifejezés kiértékelési száma $j$, akkor azt mondjuk, hogy a $K_j$ következmény realizálódott.

Legyen $f(K_j) = -j$

Kérdés: milyen i-re lesz maximális a

$$-\sum_{j=1}^{r} P(K_j | S_i) j$$

kifejezés?

Az (i) modell kérdése ilymódon azonos az 1.3. pontban felvetett kérdésünkkel.

A különbség a kiinduló adatok strukturájában van: a valószinüségek, amiket ismerünk, nem esnek egybe; egyiknek a kifejezése a másikkal és forditva komoly számolási nehézségekbe ütközik.

Ez természetesen nem zárja ki annak lehetőségét, hogy további vizsgálatok feltárják a két modellben elért eredmények kölcsönösen egymásra alkalmazhatóságát.

[1] -ben a nem ismételhető döntéshozatal elemzésének leirása található meg, az ottani eredmények sem alkalmazhatók modellünkben, ennek oka ugyanaz, mint az (i) modell esetén.

Természetesen a mondottak a (ii) modellre is érvényesek.

## 2. Egy válogatást-kihagyást végrehajtó program

A program a SIS77 statisztikai információs rendszer részét képezi. [20], [24].

A program célja adott file-ból kiválogatni bizonyos feltételeknek elegettevő rekordok adatainak egy részét. Azokról az adatokról, amelyek a szükebb file-t határozzák meg, azt mondjuk, hogy részt vesznek a kihagyásban. Azokat az adatokat, amelyeket beleveszünk a válogatott rekordok adatai közé, a válogatásban résztvevő adatoknak nevezzük.

7. példa
(Az 1. példa alapján.)
Az összes beteg közül csak azokra van szükségünk, akiket adott kórházakban ápoltak vagy 1960 után születtek; és a beteg adatai közül az alábbiak kellenek: születési év, nem, ápolási napok száma és az ápolást indokló fő kórisme. Ekkor a kihagyásban résztvevő adatok a kórházkód és a születési év, mig a válogatásban résztvevő adatok: születési év, nem ápolási napok száma és az ápolást indokló fő kórisme. ◻

Az, hogy egy rekord kihagyandó-e, bonyolultabb vagy egyszerübb logikai kifejezés értékétől függ. Ennek a logikai kifejezésnek a megadása kétféle módon történhet: 1. ha a kifejezés nem tul bonyolult, maga a felhasználó irja fel és adatkártyaként adja meg. 2. bonyolultabb kifejezés esetén a felhasználó paramétereket ad meg, s egy u.n. szerkesztő program ezek segitségével állitja össze a logikai vizsgálatokat végző programsorokat.

Az input file lehet karakterformáju, tömöritett bináris és direkt elérésü bináris. Mi most csak az első tipussal foglalkozunk.

A célprogramot amely tehát a kihagyást és a válogatást végzi egy szerkesztő program állitja össze paraméterkártyák segitségével.

2.1. Paraméterkartyák saját feltétel esetére

1. kártya Ezen a KARAKTERES szöveg áll ( más tipusu input-file esetén ez más).

2. kártya Hány kártyán fér el a beolvasást meghatározó FORMAT utasitás. (Értéke maximum 2). Nyilván csak azokat az adatokat érdemes beolvastatni, amelyek vagy a kihagyásban vagy a válogatásban résztvesznek.

3. kártya Ezen a kártyán van a beolvasást meghatározó FORMAT utasitás két zárójel közé zárva az első 6o pozición . Ha a 2. kártyán 2 áll, akkor a "3. kártyából" természetesen még egy van.

4. kártya Két adat szerepel ezen; a, a kihagyásban vagy válogatásban résztvevő adatok száma (mindkettő maximális értéke 2o ) I2, 1X, I2 formátumban.

5. kártya Kettős célt szolgál. Egyrészt megadja, hogy az input

rekord ( válogatásban v. kihagyásban résztvevő) adatelemei közül - az input rekordbeli sorrendet figyelembevéve hanyadikok azok, amelyek leválogatandók. Másrészt megmutatja, milyen sorrendben akarjuk kiirni ezeket az adatokat az output rekordba. Itt lényegében k db szám egy permutációja áll, ahol k a 4. adatkártya második száma. Az input rekord adatelemeire vonatkozó információk egy u.n. adminisztrációs file-on vannak elhelyezve. (Formátum: 20I2)

6. kártya A válogatandó adatelemek maximális értékeit tartalmazza az output rekordban elfoglalt helyük sorrendjében.( Nem az "abszolut" maximumról, hanem a leválogatandó értékek maximumáról van szó.) (Formátum: 15I4) . Ebből szükség esetén még egy kártya lehet.

7. kártya Azt mondja meg, hogy a szerkesztő program v. a felhasználó állitja össze a logikai feltételt. "Értéke" esetünkben: SAJÁT.

8. kártya Ennyi kártyán fér el a saját logikai feltétel.

9. kártya A saját logikai feltétel - amely akkor és csak akkor igaz, ha a rekord kihagyandó - megadása IF, kezdő és vég zárójel nélkül. A feltétel valahogy igy néz ki:

(K2.LT.12.OR.K5.GT.10).AND.K6.GE.200 ,

ahol a K betük után álló számok ( 2,5 és 6) azt mutatják meg, hogy a megfelelő részfeltétel az input rekordban szereplő ( válogatásban v. kihagyásban résztvevő) adat-

elemek közül hanyadikra vonatkozik. (Formátum A66)
Ilyen adatkártyából annyi db van, amilyen szám a 8. kártyán van.

10. kártya Az elkészitendő output file neve, amelyet a célprogram egyéb információkkal együtt elhelyez az adminisztrációs file-on.

11. kártya Jelölje v (ezt a jelölést a későbbiekben is használjuk) a 4. kártya második számát, azaz a válogatásban résztvevő adatelemek számát.
Az i-edik (i=1,2,..,v) leválogatandó adatelemhez a formátum által meghatározott sorrendet figyelembevéve tartozó adatnevet tartalmazza (formátum : A18). Ez az output file "adminisztrálásához kell".
Ebből a kártyából pontosan v db van.

12. kártya A válogatandó adatelemek minimális értékeit tartalmazza a rekordban elfoglalt helyük sorrendjében (nem az "abszolut" minimumot, hanem a leválogatandó értékek minimumát). Sorrend alatt a formátum által (3. kártya) meghatározott sorrend értendő. (Formátum: 15I4). A kártyából lehet még egy.

13. és 14. kártya Adminisztráláshoz szükséges kártyák (több van belőle), nem részletezzük

2.2. Paraméterkártyák megadása számitógéppel történő feltétel megadás esetére

1.-7. kártya Ugyanaz mint előző esetben az 1.-7. kártya. Röviditett jelöléssel: ua e/1.7.

8. kártya Egy adatelem több feltételben is szerepelhet más-más szempont szerint.

Pl. ha az egyik feltételben:

(nem.eq.férfi) $\wedge$ (életkor $\leq$ 3o) $\wedge$ (halál oka infarktus),

a másik feltételben:

(nem.eq.nő) $\wedge$ (életkor $\leq$ 4o) $\wedge$ (halál oka infarktus)

akkor a "nem" adatelem ( az "életkor" és a "halálok" adatelem is ) más-más értékeket vehet fel az egyes logikai feltételekben. Ezt ugy fejezzük ki, hogy egy adatelem több adattipust határozhat meg.

Ezen a kártyán a logikai feltételt megadó adattipusok száma van. Maximális értéke 2o.

9. kártya A logikai feltételben szereplő adattipusok "abszolut" minimális értékeit tartalmazza. ( ilyen kártya esetleg kettő van)

1o. kártya Mint az előző, csak az "abszolut" maximumot tartalmazza.

11. kártya Ezen három szám áll (I1,2I2 formátum szerint). Az első szám értéke 2. Ennek okaira most nem térünk ki . A második szám azt mutatja meg, hogy az input rekord hanyadik adatelemére vonatkozik a részfeltétel.

A harmadik szám azt mondja meg, hogy az előbbi adatelem hanyadik fajtáju részfeltételéről van szó. Ebből az adatkártyából annyi db van, ahány adattipus létezik, de ezek nem közvetlenül egymásután következnek. (Ld. a 13. kártya leirása utáni megjegyzést.)

A 2. és 3. számból képzett szám alkot egy "logikai változót", melyre majd a 16 kártyánál utalunk.

12. kártya Ezen több szám áll. (3X,I2,A71 szerint). Az első szám értéke

2, ha az előző kártyán definiált "logikai változóhoz a kihagyandókat soroljuk fel

1, mint az előbbi, csak az utolsó felsoroló kártyát jelenti

-2, ha az előző kártyán definiált "logikai változóhoz a nem kihagyandókat soroljuk fel

-1, mint az előbbi, csak az utolsó felsoroló kártyát jelenti.

A további számokkal az 1.2. pontban emlitett halmazokat adjuk meg. Részletesebben ld. [23] v. [24].

Az 1. szám éppen azért lehet negativ, mert előfordulhat, hogy az $A_{k,j}$ halmazt megadó számok nem férnek el egy kártyán.

Ebből a kártyából több van, ld. a 13. kártya leirasa utáni megjegyzést.

13. kártya Ezen egy szám áll, a három.

Megjegyzés a 11., 12. és 13. kártyákkal kapcsolatban

A következő sorrendben követik egymást:

11.,12.,12.,...,12.,11.,12.,12.,...,12.,...,11.,12.,12.,..., 12.,13.

14. kártya Azt mondja meg, milyen normálformát akarunk használni a rekordokat kiválasztó logikai kifejezésben. Értéke KONJUNKTIV v. DISZJUNKTIV.

15. kártya A normálforma tagjainak száma ( azaz hány konjunkcióból v. diszjunkcióból áll).
Értékét jelölje $t$ ($t \leqq 6o$).

16. kártya A normálforma i-edik ($i=1,2,...,t$) tagjában szereplő tényezők felsorolása. A tényezőket a 11. kártya leirásában bevezetett "logikai változók" alkotják. Ebből a kártyából $t$ db van .

17-21 kártya Ua e/1o.-14.

## 2.3. Példák

Az előző két pontban tárgyalt esetekre veszünk példákat. A példákban az 1. példa jelöléseit használjuk és tételezzük fel, hogy az input file a következő adatokat tartalmazza:

| karakter pozició | tartalom | jelölés |
|---|---|---|
| 1-2 | Az ápolást végző megye kódja | AMK |
| 3-4 | Az ápolást végző osztály kódja | AOK |
| 5-6 | A beteg életkora | KOR |

| karakter pozició | tartalom | jelölés |
|---|---|---|
| 7 | A beteg neme | NEM |
| 8-11 | Ápolási napok száma | NAP |
| 12-13 | Állandó lakás megyekódja | LMK |
| 14-17 | Beutaló diagnózis | BDK |
| 18-21 | Ápolást indokló fő korisme | AIK |
| 22-25 | A halál oka | HOK |

8. példa

Készitsünk egy olyan file-t, amely a következő adatokat tartalmazza: AMK,AOK,KOR,NEM és HOK.

A file-on csak azok az esetek legyenek, amelyekre:

a, 3o évnél nem idősebb, infarktusban meghalt férfi

vagy

b, 4o évnél nem idősebb, infarktusban meghalt nő

Az adatkártyákat mindkét esetre ( saját ill, számitógépes megadás ) adjuk meg! ⊠

Megoldás;

Most a kihagyásban résztvevő adatok a KOR, a NEM és a HOK, mig a leválogatandó adatok: az AMK,AOK,KOR,NEM és HOK (ennek csak az első három jegye kell).

saját megadás esetén az adatkártyák:

```
KARAKTERES
1
(3I2,I1,14X,I3)
 3  5
 1 2 3 4 5
  22  20  99   29999
SAJAT
1
(KOR.GT.30.OR.HOK.NE.410.OR.NEM.EQ.2). AND.
(KOR.GT.40.OR.HOK.NE.410.OR.NEM.EQ.1)
PROBA
AP.MEGY.KOD
:
HALAL OKA
   1   1   0   1 410
(2 számunkra lényegtelen kártya)
```

számitógépes megadás

```
KARAKTERES
1
(3I2,I1,14X,I4)
 3  5
 1 2 3 4 5
  22  20  99   29999
GEPI
 5
   0    0    0       1     1
  99   999999        2     2
2 301
     1  30-
2 302
     1  40-
2 501
     1  410
2 401
     1  2
2 402
     1  1
```

3

DISZJUNKTIV

2

301 501 401

302 501 402

PROBA

AP.MEGY.KOD

:

HALAL OKA

1 1 0 1 410

(2 számunkra lényegtelen kártya) ⊠

Bonyolultabb példák [24] -ben találhatók.

A program SIS77-beli alkalmazásánál az adatokkal kapcsolatos "gyakorisági, függetlenségi ismereteket" használtuk fel [5] .

## 3. Egy interaktiv programszerkesztési eljárás

Adatfeldolgozási munkák megoldása során gyakran merül fel a következő tipusu programozási feladat: adott input file-ból egy output file készitendő; az output file-on az input file olyan rekordjai legyenek, amelyek bizonyos feltételeknek eleget tesznek; ezek a feltételek esetenként változhatnak, s tartalmazhatnak számadatokat, amelyek szintén változtathatják értéküket.

Nézzünk erre egy példát.

### 9. példa

Az Országos Kardiológiai Intézet számára készitettük a Szivmütétre Várakozók Számitógépes rendszerét (részletesebben ld. I. fejezet 2. pontja). Ebben a rendszerben több lista előállitásához használjuk az alábbi input file-t a példa egyszerüsitése érdekében nem minden adatelemet részletezünk:

| karakterpozició | jelölés | tartalom |
|---|---|---|
| 1-24 | NEV | Beteg neve |
| 25 | NEV | Beteg nem |
| 26-49 | GON | A szülő ( gondozó) neve - gyermekeknél |
| 5o-52 | SEV | Születési év utolsó három jegye |
| 53-56 | IRS | Lakóhely irányitószáma |
| 57-75 | VAR | Lakóhely neve |
| 76-95 | UTC | Utca, házszám |
| 96-97 | MEG | Megyekód |
| 98-99 | EEV | A mütétre előjegyzés |
| 1oo-1o1 | EHO | éve, hónapja napja |
| 1o2-1o3 | ENP | (év utolsó két jegye) |
| 1o4-113 | ORV | Előjegyző orvos neve |
| 114-115 | MU1 | A betegen elvégzendő |
| 116-117 | MU2 | mütét tipusa, |
| 118-119 | MU3 | (maximum négy mütét |
| 12o-121 | MU4 | végezhető) |
| 122 | HAN | Hányadik mütüt ( primer v.nem) |
| 123-137 | BDG | A mütétet befolyásoló diagnózis |
| 138 | VER | Vércsoport |
| 139 | RHF | RH-faktora |
| 14o | EBE | Mütét sürgősségének jelzése (u.n. előjegyzési besorolás) |
| 141 | HEM | Venticulographia és/vagy aortographia jelzése |

| karakterpozició | jelölés | tartalom |
|---|---|---|
| 142 | COR | Coronarografia jelzése |
| 143-148 | IKC | Katheterezés v. coronarographia időpontja |
| 149 | OKE | Mütéti kivizsgálás teljessége |
| 15o | NYI | NYHA stádium |
| 151-152 | BEI | Mütéti behivás és archiválás visszajelzése |
| 153-154 | IBE | Ujabb mütéti behivás éve |
| 155-156 | IBH | hónapja |
| 157-158 | IBN | napja |
| 159-164 | IKI | Visszajelző lap kitöltési dátuma |
| 165-166 | IFE | Felülvizsgálat éve |
| 167-168 | IFH | hónapja |
| 169-17o | IFN | napja |
| 171-184 | EDG | Egyéb, mütétet befolyásoló diagnózis |
| 185-186 | BEM | Felülvizsgálat eredménye |
| 187-188 | UFE | Halasztást kérőkél a éve |
| 189-19o | UFH | mütét, a nem indikált hónapja |
| 191-192 | UFN | betegeknél az ujabb felülvizsgálat napja |
| 193-198 | IHE | Felülvizsgálat utáni heaemodinamikai vizsgálat ideje |
| 199-2o4 | ICO | Felülvizsgálat utáni coronarographia időpontja |
| 2o5 | ELL | Beteg ellenőrzésének státusza |
| 2o6-2o9 | KU | Beteg sorszáma |
| 21o-215 | BES | Beteg besorolási száma ( számitott érték ) |

Olyan lista előállitása a cél, amely segiti a mütütre való behivást. Ennek a listának az előállitásához többféle kivánságot (feltételt) kell figyelembe venni:

(1) Előfordulhat, hogy csak speciális mütétre várakozó beteget kell behivni. Ekkor természetesen meg kell adni a mütét tipusok számát és kódjait is

(2) Speciális vércsoportu beteg

(3) Csak adott besorolásu beteget hivhatunk

(4) Csak u.n. oke-s (akinek bizonyos vizsgálatai rendben vannak) beteg lehet a listán

(5) Csak u.n. Mütét előtti felülvizsgálaton átesett betegről lehet szó

(6) Hivhatunk olyan betegeket, akik mütétjük elhalasztását kérték

(7) U.n. coronarográfiás mütéteseket előnyben kell részesiteni

(8) Hivhatunk-e olyat, akik mütétre előkészités alatt vannak

(9) Csak férfiakat hivhatunk

(10) Csak nőket hivhatunk

(11) Csak adott korunál idősebb lehet

(12) Csak adott korunál fiatalabb lehet

(13) Csak speciális megyéből hivhatunk

(14) Csak olyan beteget hivhatunk, akinek mütét előtti felülvizsgálati lapja adott időnél nem régebbi

(15) Bizonyos kivizsgálás v. egyéb esemény megtörténte után hivható

(16) A különböző besorolásu betegek arányát meg akarjuk változtatni

(17) A behivandó betegek férfi-nő arányát meg akarjuk változtatni

A feltételekkel kapcsolatban a következőket jegyezzük meg:
a, nem feltétlenül kell mindegyik feltételre tekintettel lennünk b, vannak olyan feltételek, amelyeket ha figyelembe akarunk venni, egy v. több számadatot is meg kell hozzá adnunk. Ez a számadat futtatásról futtatásra változhat. Pl. ha az (1) feltételt használnánk, akkor meg kell adnunk a figyelembeveendő speciális mütétek számát és sorszámait. ⊠

Fogalmazzuk meg általánosan a kitüzendő programozási feladatot ugy, hogy közben a javasolt megoldási módszert is megadjuk.

## 3.1. A feladat megfogalmazása

Adott egy "nyers program" (röviden NyP), amelynek soraiból a felhasználó által megadott számadatok, karakteradatok alapján egy u.n. "szerkesztő program" ( SzP) állitja össze azt a célprogramot (CP), amelyikkel a kivánságnak megfelelő output file-t előállithatjuk.

A NyP a következő strukturáju:

```
  - - - - - - - - - -
i1.feltételhez tartozó programsorok
  - - - - - - - - - -
            :
  - - - - - - - - - -
in.feltételhez tartozó programsorok
  - - - - - - - - - -
            :
  - - - - - - - - - -
```

Ezzel kapcsolatban a következőket jegyezzük meg:

(i) A NyP nem minden sora kerül át a CP-ba

(ii) A ---- sor egy v. több olyan programsort jelöl, amely mindig átkerül a CP-ba

(iii) Egy adott feltételhez több helyen is tartozhatnak programsorok

(iv) A feltételekhez tartozó programsorok az alábbi tipusuak lehetnek:

a) egyszerü utasitás, mely a feltétel figyelembevétele esetén automatikusan átkerül a CP-ba.

b) Olyan utasitás, amely tartalmaz a felhasználó által változtatható számadatot.

Pl. ha a 11. példában a (12) feltételt akarjuk figyelembe venni, akkor az

IF(900+MAE-SEV.GT.14)GOTO1

utasitásban a 14 számadat a felhasználó kivánsága szerint esetenként más-más is lehet. (Itt MAE jelenti az aktuális programfuttatás évszámát, az (1) cimkénél történik az input file-ról az olvasás, ami esetünkben az előzőleg beolvasott rekord kihagyását jelenti)

A 14 számadat változtathatóságát olymódon érjük el, hogy a NyP-ban a fenti programsort

IF(900+MAE-SEV.GT.

1XXXXXX

2)GOTO1

alakban vesszük fel és a SzP fog gondoskodni arról, hogy a CP-ban XXXXXX helyett már az aktuális szám kerüljön (hogy milyen módon, azt nem sokára látni fogjuk).

c) U.n. "utasitáscsoport".

Ha a K változó lehetséges értékei: 1,2,...,N és a figyelembeveendő feltétellel azt akarjuk elérni, hogy az output file-ba csak azok a rekordok kerüljenek, amelynél a K változó értéke: I1,I2,...,IL , akkor ezt az

M ( 1 ) = 1
M ( 2 ) = 1
⋮
M ( N ) = 1

utasitáscsoport I1.,I2.,...,IL. sorának CP-ba való "beszerkesztésével" tudjuk elérni.

Természetesen a CP elején az M tömb lenullázása megtörténik.

(Az output file-ba kerülés feltétele ugyanis M(.) = 1)

Nézzük meg ezek után, hogy a NyP milyen módon ad információt a SzP számára arra vonatkozólag, hogy mi és hogyan kerüljön a CP-ba.

A szükséges információkat a NyP 73.-75. pozicióiban helyezzük el az alábbi módon:

- Ha azt akarjuk, hogy a programsor minden feltétel nélkül bekerüljön a CP-ba, akkor a 73.-75. poziciókat üresen hagyjuk.
- Egyéb esetben a 73.-74. pozicióban a CP kialakitásában résztvevő feltétel sorszámát kell megadni.

- A 75. pozición lévő szám értéke és jelentése a következő:

  a, 1: a programsor változtatás nélkül kell

  b, 2: utasitáscsoport <u>első soráról</u> van szó

  c, 3: utasitáscsoport <u>nem első</u> utasitását jelöli ki

  d, 4: ehhez a programsorhoz tartozik változtatható számadat

3.2. <u>A szerkesztő program müködése</u>

a) A SzP felteszi kérdéseit a felhasználónak. .......?

A felhasználó válasza: <u>i</u> vagy <u>n</u> vagy <u>v</u>, melyek jelentése:

<u>i</u>: a feltételt figyelembe kell venni a CP összeállitásánál

<u>n</u>: a feltételt nem kell figyelembevenni

<u>v</u>: sem a kérdezett feltételt, sem a soronkövetkező feltételeket nem kell figyelembevenni.

Bármilyen más válasz esetén a gép a következőket irja ki: ⊡ volt a válasza, s ez rossz ujra felteszem a kérdést .........?

b) Miután a felhasználó már minden kérdésre válaszolt vagy valamelyikre <u>v</u> választ adott, a gép megkérdezi

tehát ön a következőkre válaszolt igennel:

......... ?

......... ?

:

......... ?

Azaz felsorolja az összes olyan feltételre vonatkozó kérdést, amelyet a felhasználó figyelembe akar venni az output file összeállitásánál.

Erre a felhasználó válasza:

i: igen, ekkor a program futása folytatódik

n: bármi más a program leáll, a CP összeállitása nem történik meg.

c) Ezek után megkezdődik a CP összeállitása, összeszerkesztése. A SzP egyenként, egymás után megvizsgálja a NyP sorait. Ha a 73.-75. pozició üres, akkor ezt a sort átmásolja a CP-ba. Ha nem üres, megnézi, mi áll a 73.-74. pozición. Amennyiben ott olyan feltételnek a sorszáma áll, amit nem kell figyelembevennünk - azaz a megfelelő kérdésre n vagy v választ adtunk , akkor ezt a sort nem másolja át a CP-ba.

Ha a feltételt figyelembe kell vennünk, akkor a további teendők a 75. pozición álló számtól függően más-más módon alakulnak.

(i) A pozición 1 áll. A sort a SzP átmásolja a CP-ba.

(ii) A pozición 2 áll. A gép kérdése:

Kérem a ....... ? feltételhez tartozó tömbelemeket.

Erre 40I2 formátumnak megfelelően az alábbi számokat kell beirni:

tömbelemek száma, majd sorban a megfelelő tömbelem sorszámok

(iii) A pozición 3 áll. A sor a (ii) -re adott válasznak megfelelően bekerül vagy nem kerül be a CP-ba.

(iv) A pozición 4 áll. A gép kérdése:

Kérem a ........ ? feltételhez tartozó számot.

A megadott számmal a NyP

1XXXXXX

sorában az XXXXXX karaktersorozatot hehelyettesiti a gép, s az igy kapott uj sort szerkeszti be a CP-ba.

3.3. Alkalmazás, megjegyzések

Ebben a pontban először a SzP egy lehetséges változatát adjuk meg, majd a 11. példának megfelelő NyP-ot és végül néhány megjegyzést teszünk a módszerre vonatkozólag.

A szerkesztő program müködési algoritmusának megadásához néhány jelölést vezetünk be.

A feltételek számát jelölje kszam. Jelölje e(i) az i-edik kérdésre adott választ, mely "i", "n" v. "v" lehet (i=1,2,..., kszam). Az input file referencia kódja legyen 12-es.

A feltételek rendszere olyan, hogy a CP két darabba való megszerkesztését kivánja meg: a CP első része válogatást és a szortoláshoz a kulcs kijelölését végzi, a CP második része olyan, közvetlenül az előállitandó output file-hoz szükséges müveleteket csinál, amit az első részben még nem lehetett megcsinálni.

A CP első része a 14-es, második része a 15-ös file-ra lesz rászerkesztve.

Az input file első 72 karaktere (a FORTRAN utasitás "értékes" része) az r, a 73.-74. pozición álló szám a jfel és végül a 75. pozición álló szám a jel nevü változóba kerül beolvasásra.

Adjunk meg ezek után egy lehetséges algoritmust a SzP-ra:

1. kref ← 14
2. kszam beolvasása a 12-es file-ról
3. A feltételekre vonatkozó kérdések beolvasása a 12-es file-ról
4. e(i) feltöltése n-nel (i=1,2,...,kszam)
5. j ← 1
6. A j-edik kérdés megjelenitése a képernyőn
7. Az erre adott válasz beolvasása, értékét jelölje s.
8. Ha s ≠ "i" akkor menj 11-re
9. e(j) ← s
10. menj 14-re
11. Ha s= "v" akkor menj 17-re
12. Ha s= "n" akkor menj 14-re
13. A "rossz a válasza" szöveg kiirása, majd menj 6-ra
14. Ha j=kszam, akkor menj 17-re
15. j ← j+1
16. Menj 6-ra
17. Azon feltételekhez tartozó kérdések képernyőn való

megjelenitése, melyek "i" választ adott. És egy plusz kérdés: "tehát ön ezekre válaszolt igennel?"

18. Ez utóbbi kérdésre a válasz beolvasása
19. Ha ez a válasz "i" akkor stop
20. Egy rekord beolvasása a 13-as file-ról ez a NyP , azaz r, jfel és jel beolvasása. EOF esetén menj 50-re.
21. Ha jfel ≠ 99 akkor menj 25-re
22. EOF jel irása a 14-es file-ra, majd annak visszatekerése
23. kref ← 15
24. Menj 2o-ra
25. Ha jfel ≠ 0, akkor menj 28-ra
26. A rekord ( azaz r ) kiirása a kref referenciakódu file-ra
27. Menj 20-ra
28. Ha e(jfel) ≠ "i" akkor menj 20-ra (azaz a megfelelő sorszámu feltételre vonatkozó kérdésre nem i volt a válasz, ne szerkessze be a sort a CP-ba)
29. Ha jel ≠ 1 akkor menj 32-re
30. A rekord ( azaz r) kiirása a kref referenciakódu file-ra
31. Menj 20-ra
32. Ha jel ≠ 4, akkor menj 37-re
33. "Kérem a jfel-edik feltételhez tartozó számot" megjelenitése a képernyőn.
34. A szám beolvasása.
35. A szám kiirása a kref referenciakódu file-ra.
36. Menj 20-ra

37. Ha jel=3 akkor menj 49-re
38. A 3o elemü KU segédtömb nullával való feltöltése
39. "Kérem a jfel-edik feltételhez tartozó tömbelemeket" szöveg megjelenitése a képernyőn.
4o. A KUS ( a tömbelemek száma), majd a LU(1) , LU(2),..., LU(KUS) ( a szükséges tömbelemek sorszámai) értékek beolvasása
41. kuk ← 1
42. k ← 1
43. ku ( lu (k)) ← 1
44. ha k=kus, akkor menj 47-re
45. k ← k + 1
46. Menj 43-ra
47. Ha ku ( kuk)= 0, akkor menj 20-ra
48. A rekord ( azaz r ) kiirása a kref referenciakódu file-ra. Menj 2o-ra
49. kuk ← kuk+1. Menj 47-re.
5o. A 15-ös file-re EOF jel irás, majd a file visszatekerése
51. stop ⊠

Konkrét alkalmazásunkban a CP két Fortran programból áll. Természetesen a CP állhat akár csak egy, akár több programból is. A CP programnyelve sem feltétlenül kötött. Az SzP-ot szükség szerint át kell irni.

Az általunk javasolt interaktiv szerkesztési módszert [14] , [15] és [18] dolgozatokban ismertettük.

Hivatkozások az I. fejezethez

[1] Benedikt Szvetlána: Nem ismételhető döntéshozatal analizise kockázattal járó esetekben, MTA SZTAKI Tanulmányok, 111/198o.

[2] Gavrilov G.P., Szapozsenko: Diszkrét matematikai feladatgyüjtemény, Müszaki Könyvkiadó, 1981.

[3] Jablonszkij-Lupanov szerkesztők : Diszkrét matematika a számitástudományban, Müszaki Könyvkiadó, 198o.

[4] Fishburn.P.C.: Decision and value theory, John Wiley and sons, New York, 1964.

[5] Kramli András-Ratkó István-Ruda Mihály-Soltész János: A statisztikai adatfeldolgozás matematikai és számitástechnikai problémái, MTA SZTAKI Tanulmányok, 7o/1977.

[6] Z.M. Lotfi, A.J. Tosser: Fast arithmetic procedure for minimization of logical functions, Int. J. Electron, 1979, 47 No2, pp. 181-185

[7] Matavovszky Tibor, Pásztorné Varga Katalin: Programrendszer Boole-függvényrendszer együttes egyszerüsitésére vagy minimalizálására, MTA SZTAKI Tanulmányok, 43/1975.

[8] Pásztorné Varga Katalin: Rekurziv eljárás, MTA SZTAKI Tanulmányok, 1o2/198o.

[9] Solomon L. Pollack: Conversion of Limited-Entry Decision Tables to Computer Programs, Communications of the ACM, Vol 8/No 11/Nov. 1965, pp. 677-682

[1o] Ratkó István: Egy számitástechnikai eszköz bonyolult logikai kifejezések leirására orvosstatisztikai alkalmazásokban, "Számitástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biologiában" (SZKMAOB) 7. kollokvium, Szeged, 1977, pp. 279-286.

[11] István Ratkó: On optimization problems of logical expressions in programming languages, Colloquium on mathematical logic in programming, Salgótarján, 1978.

[12] Ratkó István: Döntések sztochasztikus optimalizációja adatfeldolgozásnál, MTA SZTAKI Tanulmányok, 11o/198o, pp. 19-26.

[13] István Ratkó: Kelkaj rimarkoj pri la statisztika prilaboro de datumoj, Apliko de esperanto en la scienco kaj tehniko, Žilina, 1981, pp. 11-15.

[14] Ratkó István: Egy interaktiv programszerkesztési módszer ismertetése, SZKMAOB 11. kollokvium, Szeged, 1982. pp. 469-471.

[15] István Ratkó: Dialoge funkciada metodo por konstrui programojn, Interkomputo'82, Teoriaj kaj praktikaj problemoj de la programado, Budapest, 1982. pp. 162-165.

[16] István Ratkó: Stokasta optimaligo de decidoj, Interkomputo' 82, Teoriaj kaj praktikaj problemoj de la programado, Budapest, 1982, pp. 125-129.

[17] István Ratkó: On evaluating of logical expressions in programming languages, Cybernetics and Systems Research 2., R. Trappl ed., Elsevier Science Publisher B.V. North-Holland, 1984. pp. 611-613.

[18] Ratkó István: Válogatott számitástechnikai és matematikai módszerek orvosi alkalmazásokban, Kandidátusi disszertáció, 1984.

[19] Ruda Mihály: Optimalizálási kérdések a statisztikai adatfeldolgozásban, MTA SZTAKI Tanulmányok, 110/1980, pp. 5-18.

[20] Ruda Mihály: A SIS77 statisztikai információs rendszer kialakitásának szempontjai, alkalmazásának és továbbfejlesztésének lehetőségei, MTA SZTAKI Tanulmányok, 86/1978.

[21] J.K. Sethi, B. Chatterjee: Conversion of Decision Tables to Efficient sequential testing procedures, Comm. of the ACM, Vol. 23/No. 5/May. 1980, pp. 279-293.

[22] Keith Shwayder: Conversion of Limited-Entry Decision Tables to Computer Programs - A proposed modification to Pollack's Algorithm, Comm. of the ACM, Vol. 14/No.2/Febr. 1971, pp. 69-73.

[23] Soltész János: Egy általánosan használható eljárás táblázatoknak a központi memóriában történő kitöltésére, MTA SZTAKI Tanulmányok 110/1980, pp. 27-57.

[24] A SIS77 statisztikai információs rendszer témadokumentációja, MTA SZTAKI, 1978.

## II. MULTIFAKTORIÁLIS KÓREREDETÜ BETEGSÉG ÖRÖKLŐDÉSÉNEK VIZSGÁLATA ADOTT JELLEMZŐ ISMERETÉBEN

A familiáris betegségeket három csoportra szokták osztani [3]: (1) az egyszerü Mendel szerint öröklődő egyetlen genetikai lókusznak köszönhető betegségek, (2) ismert kromoszóma rendellenességnek köszönhető betegségek, (3) egyéb familiáris betegségek.

A harmadik csoportba tartoznak - többek között - az u.n. multifaktoriális modellnek megfelelő betegségek v. más szóval a multifaktoriális kóreredetü betegségek . Ennek részletes elemzésével nem foglalkozunk, az megtalálható pl. a [3], [4], [24], [25], [31], [32] v. [33] helyeken.

A multifaktoriális modell szerint - adott rendellenességet vizsgálva - minden egyénhez hozzárendelhető egy $\acute{\xi}$ valós szám, amit hajlamnak is szoktak nevezni. Ezt a hajlamot a $\xi$ genetikai és az $\eta$ környezeti hatás együttesen alakitja ki:

$$\acute{\xi} = \xi + \eta$$

ahol $\xi$ és $\eta$ függetlenek, $\xi \sim N(0, h^2)$ , $\eta \sim N(0, 1-h^2)$. $h^2$ az u.n. örökölhetőségi együttható, az f-ed foku rokonok korrelációs együtthatója $h^2/2^f$ . ( Elsőfoku rokonok a szülők és a testvérek - egypetéjü ikreket nem számitva - ; másodfokuak a nagyszülők, nagybácsik; nagynénik; harmadfokuak a dédszülők, a nagyszülők testvérei, unokatestvérek; stb.)

Egy egyén akkor és csak akkor beteg, ha hajlama nagyobb egy - a vizsgált egyedet tartalmazó populációra jellemző küszöbnél.

Pl. valamely konkrét - nemtől függő - rendellenességet vizsgálva, ha az összes $N_i$ (i = 1 férfi, i = 2 nő) egyén közül $M_i$(i= 1,2) rendelkezik az adott rendellenességgel, akkor a küszöbök ( $k_1$ férfira,

$$k_i = \Phi^{-1}\left(1-\frac{M_i}{N_i}\right)$$

ahol $\Phi$ a Gauss függvény.

## 1. Számolási módszerek a multifaktoriális modellben

Ebben a pontban a többdimenziós normális eloszlásfüggvény számitási módszereivel foglalkozunk, röviden.

a/ [5] különböző Monte Carlo módszereket ismertet többdimenziós normális eloszlásfüggvény meghatározására. A módszereket egymással összehasonlitja. A programok Honeywell-Bull 66/60 és 66/20 számitógépre FORTRAN nyelven készültek a gépi kódban irt véletlen számgenerátorok kivételével.

b/ [30] által adott algoritmus alapja, egy a Jordan tételt felhasználó módszer. A program UNIVAC 1108-as gépre készült.

c/ [1] módszere kevésbé ismert, aminek talán az az oka, hogy eredményét nem publikálta. Mi az ő módszerét használtuk a kiinduló valószinüségek meghatározására, mivel programja elsősorban öröklődő betegségek kockázati értékének kiszámitására készült. Ezuton is köszönetet mondunk Neki segitségéért.

d/ [14], [15], [18] ötletét használva rekurziv algoritmust adhatunk a valószinüségek kiszámitására. Ennek alapja az alábbi

1. tétel:

Legyen $\mathbb{X}^{k\times 1}=(X_1,X_2,\dots,X_k)$ $N(\mu,\Sigma)$ eloszlásu,

ahol $\underline{\mu}=(\mu_1,\mu_2,\dots,\mu_k)'$

$$\Sigma=\begin{pmatrix}\sigma_1^2 & \dots & \sigma_{1k}\\ \vdots & & \\ \sigma_{k1} & \dots & \sigma_k^2\end{pmatrix},\qquad \sigma_{i,j}=\mathrm{cov}(X_i,X_j)$$

Legyen továbbá

$$\mathbb{X}_1=(X_1,X_2,\dots,X_\ell)',\quad \mathbb{X}_2=(X_{\ell+1},\dots,X_k)'$$

$$\underline{\mu_1}=(\mu_1,\dots,\mu_\ell)',\qquad \underline{\mu_2}=(\mu_{\ell+1},\dots,\mu_k)'$$

$$\Sigma_{11}=\begin{pmatrix}\sigma_1^2\dots & \sigma_{1\ell}\\ \vdots & \vdots\\ \sigma_{\ell 1}\dots & \sigma_\ell^2\end{pmatrix},\ \Sigma_{22}=\begin{pmatrix}\sigma_{\ell+1}^2\dots & \sigma_{\ell+1,k}\\ \vdots & \vdots\\ \sigma_{k,\ell+1}\dots & \sigma_k^2\end{pmatrix},\ \Sigma_{12}=\Sigma_{21}'=\begin{pmatrix}\sigma_{1,\ell+1},\dots & \sigma_{1k}\\ \vdots & \vdots\\ \sigma_{\ell,\ell+1},\dots & \sigma_{\ell k}\end{pmatrix}$$

Ekkor $\mathbb{X}_1$ $N(\underline{\mu_1},\Sigma_{11})$, $\mathbb{X}_2$ $N(\underline{\mu_2},\Sigma_{22})$ eloszlásu, továbbá $\mathbb{X}_1$ eloszlása az $\mathbb{X}_2=\underline{x_2}=(x_{\ell+1},\dots,x_k)'$ feltétel mellett

$$N\left(\underline{\mu_1}+\Sigma_{12}\Sigma_{22}^{-1}(\underline{x_2}-\underline{\mu_2}),\ \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)\qquad\boxtimes$$

e/ [27] és [28] genetikai osztályok ismeretében becsüli a valószinüségeket.

f/ [11] a többdimenziós Pareto eloszláson [17] alapuló számolási módszert javasol a kockázati értékek meghatározásához. Mivel ez a módszer kevésbé ismert, ezért erről egy kicsit részletesebben szólunk.

Legyen $P(x,y)$ annak a valószinüsége, hogy egy egyén hajlama nagyobb, mint x és egy másik egyén hajlama nagyobb, mint y. [11] felteszi, hogy $P(x,y)$ a kétváltozós Pareto eloszlásfüggvény, azaz

$$P(x,y) = (1+x+y)^{-c} \quad (x,y,c>0)$$

ahol a c paraméter a két egyén hajlamának korrelációjától függ. Ha a az a küszöb, aminél a betegség manifesztálódik, p a betegség populációs incidenciája, akkor

$$p = P(a,0) = (1+a)^{-c} \tag{1}$$

továbbá, ha q a beteg egyének elsőfoku rokonaiban a betegség incidencia, akkor

$$q = \frac{P(a,a)}{P(a,0)} = \frac{(1+2a)^{-c}}{(1+a)^{-c}} \tag{2}$$

p és q értékét ismertnek feltételezve (1) és (2) alapján

$$\frac{\log(1+a)}{\log(1+2a)} = \frac{\log p}{\log p + \log q}$$

egyenlőségből a, majd a

$$c = -\frac{\log p}{\log(1+a)}$$

egyenlőségből c becsülhető.

Nézzünk egy példát: egy szülő és egy meglévő gyermek beteg.

Mi a valószinüsége annak, hogy egy további gyermek beteg lesz?

Az együttes eloszlás függvény:

$$(1+x+y+z)^{-c}$$

Könnyen adódik, hogy a keresett valószinüség:

$$\frac{(1+3a)^{-c}}{(1+2a)^{-c}}$$

[11] -ben összehasonlitások is vannak a két modellre (Pareto és multifaktoriális) nézve. A szerző az összehasonlitások alapján, a többdimenziós normális eloszlás "nehéz kezelhetőségét" hangsulyozva arra a következtetésre jut, hogy a Pareto modell hasonló eredményeket ad, igy jól használható.

Ezzel kapcsolatban a következőket jegyeznénk meg:

- A cikkben nincs igazán jól megindokolva a Pareto eloszlás használatának jogossága. Ezzel szemben a multifaktoriális modellmagyarázat precizen kidolgozott matematikai "háttérrel" rendelkezik ld. pl. [3], [4].
- A Pareto modellnek vannak tisztázatlan pontjai, nem mindig alkalmazhatók "rendesen". Pl. apát, anyát és gyermeket tekintve a megfelelő eloszlás

  $$(1+x+y+xy+z)^{-c}$$

  lenne, ám ez nem eloszlásfüggvény.

  További problémákra nézve utalunk az eredeti cikkre.
- Mindezek ellenére ugy gondoljuk, hogy éppen a könnyebb kezelhetősége miatt érdemes a Pareto modellel foglalkozni, de inkább azt a kérdést kell vizsgálat tárgyává tenni,

hogy a két eloszlás (Pareto és a normális) milyen feltételek mellett lehet egymás "közelitése".

## 2. "Jellemző" figyelembevétele a modellben

Adott egy multifaktoriális öröklődésü betegség, amelyről még azt is tudjuk, hogy valamilyen "jellemző" pl. szembelnyomás mértéke, a szérum lipoprotein v. hugysav szintje, HLA-B27 antigén jelenléte, stb a betegség kiváltásával szoros kapcsolatban van. A következő kérdésre keressük a választ: mi a valószinüsége annak, hogy egy adott gyermek beteg lesz, ha tudjuk, hogy szülei közül kik betegek, s ismerjük mindhármójuk jellemzőjének értékét.

Olyan esetre, amikor a jellemző folytonos, vannak módszerek a keresett valószinüségek kiszámitására [24], [29], diszkrét esetre azonban nincs.

A továbbiakban a diszkrét jellemző esetével foglalkozunk.

A jellemzőről feltesszük, hogy csak a 0,1 vagy 2 értéket veheti fel és nem változik az ember élete során. Egy gyermek jellemzője a következőképpen "adódik össze" szülei jellemzőjéből: ha valamelyik szülő jellemzője 0 ill. 2, gyermekének jellemzőjéhez 0-val ill. 1-gyel járul hozzá. Ha a szülő jellemzője 1, akkor gyermekének jellemzőjéhez 1/2 valószinüséggel 0-val, 1/2 valószinüséggel 1-gyel járul hozzá.

Ilyen diszkrét jellemző pl. az SPA esetére a HLA-B27 antigén [2], [8], [12], [13].

2.1. A jellemző eloszlásának meghatározása

Jelölje $x_{esj}$ ill. $x_{bgj}$ a j nemü ( j=1 férfi, j=2 nő) egészséges szülő ill. a j nemü beteg gyermek jellemzőjét. Egyéb $x_{...}$ jelölések értelemszerüek.

Vezessük be még a következő jelöléseket:

$r_j$ = annak a valószinüsége, hogy az általános populáció egy véletlenszerüen választott egyedének jellemzője $j$ $(j=0,1,2)$.

$q_1 = P(x_{bs1} \neq 0)$ $\qquad$ $t_1 = P(x_{es1} \neq 0)$

$q_2 = P(x_{bs2} \neq 0)$ $\qquad$ $t_2 = P(x_{es2} \neq 0)$

$s_j$ = annak a valószinüsége, hogy egy j nemü egyed beteg $(j=1,2)$.

Az $r = r_1 + r_2$, $q_1$, $q_2$, $s_1$ és $s_2$ értékeket a megfelelő populációkból becsüljük.

A Hardy-Weinberg [6] szabályból elemi számolások után adódik az

1. lemma

$$r_0 = 1-r \qquad r_1 = 2r + 2\sqrt{1-r} - 2 \qquad r_2 = 2-r-2\sqrt{1-r} \qquad (1) \boxtimes$$

A teljes valószinüség tételének felhasználásával az 1. lemmából következik a

2. lemma

$$t_i = (r - q_i s_i)/(1 - s_i) \qquad (i = 1,2) \qquad \boxtimes$$

Olyan betegséggel foglalkozunk, amelyre:

$$P(\xi_{sj} > k_j \mid x_{sj}=1) = P(\xi_{sj} > k_j \mid x_{sj}=2) \quad (j=1,2) \qquad (2)$$

Azaz, amely betegségre igaz az, hogy az 1 jellemzőjü egyedek között a betegség ugyanolyan gyakori, mint a 2 jellemzőjü betegek között. Ilyen jellemző-betegség pár a HLA-B27 antigén-SPA.

1. Tétel

$$\frac{P(x_{b1}=1)}{P(x_{b1}=2)} = \frac{P(x_{e1}=1)}{P(x_{e1}=2)} = \frac{r_1}{r_2} \qquad \boxtimes$$

Bizonyitás

A (2) feltevésből következik, hogy

$$\frac{P(\xi_{s1} > k_1,\, x_{s1}=1)}{P(x_{s1}=1)} = \frac{P(\xi_{s1} > k_1,\, x_{p1}=2)}{P(x_{p1}=2)}$$

Ebből

$$\frac{P(\xi_{s1} > k_1,\, x_{s1}=1)}{P(\xi_{s1} > k_1,\, x_{s1}=2)} = \frac{P(x_{s1}=1)}{P(x_{s1}=2)}$$

A jobb oldal $(r_1/r_2)$ -vel egyenlő, igy átalakitás után kapjuk:

$$\frac{P(x_{s1}=1 \mid \xi_{s1} > k_1)\, P(\xi_{s1} > k_1)}{P(x_{s1}=2 \mid \xi_{s1} > k_1)\, P(\xi_{s1} > k_1)} = \frac{r_1}{r_2},$$

azaz a bizonyitandó állitás egyik részét megkaptuk. A másik rész majdnem triviális. ⊠

Az 1. tétel alapján a teljes valószinüség tételét használva kapjuk a

3. lemmát

$$P(x_{ej}=1)=r_1 t_j/(r_1+r_2)$$

$$P(x_{bj}=1)=r_1 q_j/(r_1+r_2) \qquad (j=1,2)$$ ⊠

## 2.2. Kockázat számolása a jellemző ismeretében

Célunk az alábbi valószinüségek meghatározása:

$$P\left(\xi_{gj}>k_j \,\middle|\, \xi_{s1}\overset{>}{\leq}k_1,\ \xi_{s2}\overset{>}{\leq}k_2,\ x_{gj}\overset{=}{\neq}0,\ x_{s1}\overset{=}{\neq}0,\ x_{s2}\overset{=}{\neq}0\right)$$
$(j=1,2)$.

A tömörebb irásmód kedvéért vezessük be a következő jelöléseket:

$\omega_{s1}=0$ , ha $\xi_{s1}\leq k_1$

$\omega_{s1}=1$ , ha $\xi_{s1}>k_1$

$\tau_{g1}=0$ , ha $x_{g1}=0$

$\tau_{g1}=1$ , ha $x_{g1}\neq 0$

Értelemszerüen használunk egyéb $\omega_{..}$, $\tau_{..}$ jelöléseket is.

A keresett valószinüségek ezekkel a jelölésekkel igy irhatók:

$$P(\omega_{gi}=1 \mid \omega_{s1}=j,\ \omega_{s2}=k,\ \tau_{s1}=l,\ \tau_{s2}=m,\ \tau_{gi}=n)$$
$(i=1,2;\ \ j,k,l,m,n=0,1)$.

Tekintsük át először vázlatosan ezen valószinüségek meghatározásának a módszerét.

A becsült adatok alapján kiszámitjuk egyrészt a

$$
\begin{array}{ll}
P(\tau_{g2}=i, \tau_{s2}=j, \omega_{s1}=k) & P(\tau_{g2}=i, \tau_{s1}=j, \tau_{s2}=k) \\
P(\tau_{g2}=i, \omega_{s1}=j, \omega_{s2}=k) & P(\tau_{g2}=i, \tau_{s1}=j, \omega_{s1}=k) \\
P(\tau_{g2}=i, \tau_{s2}=j, \omega_{s2}=k) & P(\tau_{s1}=i, \omega_{s1}=j, \omega_{s2}=k) \\
P(\tau_{g2}=i, \tau_{s1}=j, \omega_{s2}=k) & P(\tau_{s1}=i, \omega_{s1}=j, \omega_{s2}=k) \\
P(\tau_{s1}=i, \tau_{s2}=j, \omega_{s1}=k) & P(\tau_{s1}=i, \tau_{s2}=j, \omega_{s2}=k) \\
\multicolumn{2}{c}{P(\omega_{g2}=i, \omega_{s1}=j, \omega_{s2}=k)}
\end{array} \qquad (3a)
$$

(i,j,k 0 vagy 1, de közülük legfeljebb egy lehet 0),

másrészt a

$$
\begin{array}{lll}
P(\omega_{s1}=j, \omega_{s2}=k) & P(\omega_{s1}=j, \omega_{g2}=k) & P(\omega_{s2}=j, \omega_{g2}=k) \\
P(\tau_{s1}=j, \tau_{s2}=k) & P(\tau_{s1}=j, \tau_{g2}=k) & P(\tau_{s2}=j, \tau_{g2}=k) \\
P(\omega_{s1}=j, \tau_{s1}=k) & P(\omega_{s2}=j, \tau_{s2}=k) & P(\omega_{g2}=j, \tau_{g2}=k)
\end{array} \qquad (3b)
$$

$(j=0,\ k=1$ vagy $j=1,\ k=0)$

harmadrészt a

$$
P(\omega_{s1}=1) \qquad P(\omega_{s2}=1) \qquad (3c)
$$

valószinüségeket.

Ily módon (3a)-ból 44, (3b)-ből 18 és (3c)-ből 2, azaz összesen 64 valószinüséget határozunk meg.

Bevezetve az

$$
x_{ijkluv} = P(\omega_{g2}=i, \omega_{s1}=j, \omega_{s2}=k, \tau_{g2}=l, \tau_{s1}=u, \tau_{s2}=v)
$$

$(i,j,k,l,u,v = 0,1)$

64 ismeretlent, a

$$\sum \alpha_{ijkluv} \cdot x_{ijkluv} = \beta_{ijkluv}$$

egyenletrendszert irhatjuk fel, ahol

$\beta_{ijkluv}$ a (3a), (3b) és (3c) valószinüségekkel egyenlő, továbbá

$\alpha_{ijkluv}$ =1 vagy 0, aszerint, hogy

$$\sum \alpha_{ijkluv} \{\omega_{g2}=i, \omega_{s1}=j, \omega_{s2}=k, \tau_{g2}=l, \tau_{s1}=u, \tau_{s2}=v\} = A,$$

$$\beta_{ijkluv} = P(A)$$

Ezt az egyenletrendszert megoldva, az x ismeretlenek alapján a keresett valószinüségek (lánygyermekre) már meghatározhatók. (Fiugyermek esetén ugyanez a módszer alkalmazható.)

Nézzük meg ezek után a (3) valószinüségek meghatározását. Egyes esetekben csak speciális indexekre adjuk meg a módszert, de természetesen más indexre hasonlóan számolhatunk.

Használni fogjuk az alábbi természetes feltételt:

$$\begin{aligned} &P(\chi_{gu}=i \mid \chi_{s2}=j, \chi_{s1}=k, \omega_{sv}=l) = \\ &= P(\chi_{gu}=i \mid \chi_{s2}=j, \chi_{s1}=k) \qquad (4) \\ &(u,v=1,2;\ i,j,k=0,1,2;\ \ l=0,1) \end{aligned}$$

<u>2. tétel:</u>

$$P(\tau_{g2}=0, \tau_{s2}=1, \omega_{s1}=k) =$$

$$= \sum_{l=0}^{2} \sum_{j=1}^{2} P(\tau_{g2}=0 \mid \tau_{s2}=j, \tau_{s1}=l)\, r_j\, P(\tau_{s1}=l \mid \omega_{s1}=k)\, P(\omega_{s1}=k)$$

ahol a $P(\tau_{g2}=0 \mid \tau_{s2}=j, \tau_{s1}=\ell)$ valószinüség a Mendel és a Hardy-Weinberg szabályok alapján számolható, továbbá

$$P(\tau_{s1}=\ell \mid \omega_{s1}=k)=\begin{cases} P(\chi_{e1}=\ell) & \text{, ha } k=1 \\ P(\chi_{b1}=\ell) & \text{, ha } k=0 \end{cases}$$

és $P(\omega_{s1}=1)=s_1$ ⊠

Bizonyitás

$$P(\tau_{g2}=0, \tau_{s2}=0, \omega_{s1}=k)=\sum_{j=1}^{2} P(\chi_{g2}=0, \chi_{s2}=j, \omega_{s1}=k)=$$

$$=\sum_{\ell=0}^{2}\sum_{j=1}^{2} P(\chi_{g2}=0, \chi_{s2}=j, \omega_{s1}=k, \chi_{s1}=\ell)=$$

$$=\sum_{\ell=0}^{2}\sum_{j=1}^{2} \left[P(\chi_{g2}=0 \mid \chi_{s2}=j, \omega_{s1}=k, \chi_{s1}=\ell)\cdot P(\chi_{s2}=j, \omega_{s1}=k, \chi_{s1}=\ell)\right]$$

(4) -et felhasználva, s tovább alakitva kapjuk, hogy ez egyenlő az alábbival:

$$\sum_{\ell=0}^{2}\sum_{j=1}^{2}\left[P(\chi_{g2}=0 \mid \chi_{s2}=j, \chi_{s1}=\ell)\cdot r_j\cdot P(\chi_{s1}=\ell \mid \omega_{s1}=k)\, P(\omega_{s1}=k)\right]$$ ⊠

3. tétel

$$P(\tau_{g2}=0, \tau_{s1}=1, \tau_{s2}=1)=\sum_{j=1}^{2}\sum_{\ell=1}^{2} P(\chi_{g2}=0 \mid \chi_{s1}=j, \chi_{s2}=\ell)\cdot r_j\cdot r_\ell$$

ahol a $P(\chi_{g2}=0 \mid \chi_{s1}=j, \chi_{s2}=\ell)$ a Mendel és Hardy-Weinberg szabályokból számolható. ⊠

Bizonyitás

$$P(\tau_{g2}=0, \tau_{s1}=1, \tau_{s2}=1) = \sum_{j=1}^{2}\sum_{\ell=1}^{2} P(x_{g2}=0, x_{s1}=j, x_{s2}=\ell) =$$

$$= \sum_{j=1}^{2}\sum_{\ell=1}^{2} P(x_{g2}=0 \mid x_{s1}=j, x_{s2}=\ell)\cdot r_j \cdot r_\ell \quad \boxtimes$$

4. tétel

$$P(\tau_{g2}=1, \omega_{s1}=1, \omega_{s2}=1) = P(x_{bs1}=0)\cdot\left[\tfrac{1}{2}P(x_{bs2}=1)+P(x_{bs2}=2)\right]+$$

$$+P(x_{bs1}=1)\cdot\left[\tfrac{1}{2}P(x_{bs2}=0)+\tfrac{3}{4}P(x_{bs2}=1)+P(x_{bs2}=2)\right]+P(x_{bs1}=2) \quad \boxtimes$$

Bizonyitás

$$P(\tau_{g2}=1, \omega_{s1}=1, \omega_{s2}=1) = \sum_{i=1}^{2} P(x_{g2}=i, \omega_{s1}=1, \omega_{s2}=1) =$$

$$= \sum_{i=1}^{2}\sum_{j=0}^{2}\sum_{\ell=0}^{2} P(x_{g2}=i, \omega_{s1}=1, \omega_{s2}=1, x_{s1}=j, x_{s2}=\ell) =$$

$$= \sum_{i=1}^{2}\sum_{j=0}^{2}\sum_{\ell=0}^{2}\Big[P(x_{g2}=i \mid \omega_{s1}=1, \omega_{s2}=1, x_{s1}=j, x_{s2}=\ell)\cdot$$

$$\cdot P(\omega_{s1}=1, x_{s1}=j)\, P(\omega_{s2}=1, x_{s2}=\ell)\Big] =$$

$$= \sum_{i=1}^{2}\sum_{j=0}^{2}\sum_{\ell=0}^{2}\Big[P(x_{g2}=i \mid x_{s1}=j, x_{s2}=\ell)\, P(x_{s1}=j \mid \omega_{s1}=j)\cdot$$

$$\cdot P(\omega_{s1}=1)\, P(x_{s2}=\ell \mid \omega_{s2}=1)\, P(\omega_{s2}=1)\Big]$$

s ebből már a bizonyitandó egyenlőség jobb oldala egyszerüen adódik. ⊠

5. tétel

$$P(\tau_{g2}=0,\tau_{s1}=1,\omega_{s1}=0)=$$

$$=\sum_{\ell=0}^{2}\sum_{j=1}^{2}[P(x_{g2}=0\mid x_{s1}=j,x_{s2}=\ell)P(x_{s2}=\ell)P(x_{s1}=j\mid P(\omega_{s1}=0)P(\omega_{s1}=0) \boxtimes$$

Bizonyitás:

$$P(\tau_{g2}=0,\tau_{s1}=1,\omega_{s1}=0)=\sum_{\ell=0}^{2}\sum_{j=1}^{2}P(x_{g2}=0,x_{s1}=j,\omega_{s1}=0,x_{s2}=\ell)=$$

$$=\sum_{\ell=0}^{2}\sum_{j=1}^{2}[P(x_{g2}=0\mid x_{s1}=j,\omega_{s1}=0,x_{s2}=\ell)P(x_{s1}=j,\omega_{s1}=0,x_{s2}=\ell)]$$

(4) -et felhasználva, s tovább alakitva adódik az állitás. ⊠

A további valószinüségekre vonatkozó összefüggéseket egy tételbe gyüjtjük össze. Bizonyitásuk az eddigiekhez hasonló módon történhet.

6. tétel

(i) $$P(\tau_{g2}=0,\tau_{s2}=1,\omega_{s2}=0)=$$

$$=\sum_{\ell=0}^{2}\sum_{j=1}^{2}[P(x_{g2}=0\mid x_{s2}=j,x_{s1}=\ell)P(x_{s1}=\ell)P(x_{s2}=j\mid\omega_{s2}=0)P(\omega_{s2}=0)$$

(ii)

$$P(\tau_{s1}=i,\omega_{s1}=j,\omega_{s2}=k)=P(\tau_{s1}=i\mid\omega_{s1}=j)P(\omega_{s1}=j)P(\omega_{s2}=k)$$

(iii) $$P(\tau_{g2}=0,\tau_{s1}=1,\omega_{s2}=k)=$$

$$=\sum_{\ell=0}^{2}\sum_{j=1}^{2}P(\tau_{g2}=0\mid\tau_{s1}=j,\tau_{s2}=\ell)r_j P(\tau_{s2}=\ell\mid\omega_{s2}=k)P(\omega_{s2}=k)$$

(iv) $$P(\tau_{s2}=i,\omega_{s1}=j,\omega_{s2}=k)=P(\tau_{s2}=i\mid\omega_{s2}=k)P(\omega_{s2}=k)P(\omega_{s1}=j)$$

(v) $$P(\tau_{s1}=i,\tau_{s2}=j,\omega_{s1}=k)=P(\tau_{s1}=i\mid\omega_{s1}=k)P(\omega_{s1}=k)P(\tau_{s2}=i)$$

(vi) $$P(\tau_{s1}=i,\tau_{s2}=j,\omega_{s2}=k)=P(\tau_{s2}=i\mid\omega_{s2}=k)P(\omega_{s2}=k)P(\tau_{s1}=i)$$

(VII) $$P(\omega_{g2}=i,\ \omega_{s1}=j,\ \omega_{s2}=k) = P(\omega_{g2}=i\,|\,\omega_{s1}=j,\omega_{s2}=k)P(\omega_{s1}=j)P(\omega_{s2}=k)$$

(VIII) $$P(\omega_{s1}=j,\ \omega_{s2}=k) = P(\omega_{s1}=j)\,P(\omega_{s2}=k)$$

(ix) $$P(\omega_{s1}=j,\ \omega_{g2}=k) = \sum_{i=0}^{1} P(\omega_{g2}=k\,|\,\omega_{s2}=i,\omega_{s1}=j)P(\omega_{s1}=j)P(\omega_{s2}=i)$$

(x) $$P(\omega_{s2}=j,\omega_{g2}=k) = \sum_{i=0}^{1} P(\omega_{g2}=k\,|\,\omega_{s1}=i,\omega_{s2}=j)P(\omega_{s1}=i)P(\omega_{s2}=j)$$

(xi) $$P(\tau_{s1}=j,\ \tau_{s2}=k) = P(\tau_{s1}=j)\,P(\tau_{s2}=k)$$

(xii) $$P(\tau_{s1}=1,\ \tau_{g2}=1) =$$
$$= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{\ell=0}^{2}\left[P(\chi_{g2}=j\,|\,\chi_{s1}=i,\chi_{s2}=\ell)P(\chi_{s1}=i)P(\chi_{s2}=\ell)\right]$$

(XIII) $$P(\tau_{s2}=1,\tau_{g2}=1) =$$
$$= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{\ell=0}^{2}\left[P(\chi_{g2}=j\,|\,\chi_{s2}=i,\chi_{s1}=\ell)P(\chi_{s2}=i)P(\chi_{s1}=\ell)\right]$$
(xiv)

(XIV) $$P(\omega_{s1}=j,\ \tau_{s1}=k) = P(\tau_{s1}=k\,|\,\omega_{s1}=k)\,P(\omega_{s1}=k)$$

(xv) $$P(\omega_{s2}=j,\tau_{s2}=k) = P(\tau_{s2}=k\,|\,\omega_{s2}=j)\,P(\omega_{s2}=j)$$

(xvi) $$P(\omega_{g2}=j,\tau_{g2}=k) = P(\tau_{g2}=k\,|\,\omega_{g2}=j)\,P(\omega_{g2}=j)$$

## 2.3. Egy alkalmazás

Már a század elején észrevették, hogy a spondylitis ankylopoetica v.más néven Bechterew-féle betegség egyes családokban halmozottan fordul elő. 1967-ben a vizsgálatok alapján megállapitották [7] , hogy a SA multifaktoriális kóreredetü. 1973-ban vált ismertté, hogy a SA és a HLA-B27 antigén között szoros kapcsolat van: az SA betegek között jóval gyakoribb az antigénnel rendelkezők száma, mint a kontroll populációban [2], [26].

Magyarországon 1976-ban fejeződtek be az ez irányu vizsgálatok [8]. Ennek során a következők adódtak:

a) Megerősitést nyert az SA multifaktoriális jellege: [33] módszerével megtörtént a GAMT modell tesztelése. (A módszer leirása [4] -ben is megtalálható.) Számszerü eredmények:

$h^2 = 0,7925$

$s_1 = 0,004$ $\qquad s_2 = 0,0008$

b) A HLA-B27 antigénnel kapcsolatos becslések:

$r = 0,1278$ $\qquad q_1 = 0,8298$ $\qquad q_2 = 0,75$

Az 1. lemma alapján:

$r_0 = 0,8722$ $\qquad r_1 = 0,1235$ $\qquad r_2 = 0,0043$

mig a 2. lemmából

$t_1 = 0,1250$ és $t_2 = 0,1273$ adódik.

A $P(\chi_{ej}=k)$ és a $P(\chi_{bj}=k)$ $(j=1,2;\ k=0,1,2)$ valószinüségek az 1. tétel és a 3. lemma segitségével számolhatók. A számszerü eredményekre vonatkozóan ld. [19], [20].

Bene módszerével [1] számoltuk a

$$P(\omega_{gk}=1 \mid \omega_{s1}=i,\ \omega_{s2}=j) \qquad (k=1,2;\ i^2+j^2=1)$$

értékeket (a táblázat vastagon szedett számai).

A 64 ismeretlenes egyenletrendszereket TI-59 programozható zsebszámológéppel oldottuk meg. Az egyenletrendszerből kapott ismeretlenek alapján a

$$P(\omega_{gi}=1 \mid \xi_{s1}=j,\ \xi_{s2}=k,\ \tau_{s1}=l,\ \tau_{s2}=m,\ \tau_{gi}=n)$$

$$(i=1,2;\quad j,k,l,m,n=0,1)$$

valószinüségek meghatározhatók. Ezek közül mi csak tizenhatot adtunk meg, a genetikai tanácsadás igényeinek megfelelően, az értékek az alábbi táblázatban láthatók.

| | | Apa beteg | | Anya beteg | |
|---|---|---|---|---|---|
| | | Az apának antigénje | | Az anyának antigénje | |
| | | van | nincs | van | nincs |
| A fiunak antigénje | van | 0,0976 | 0,0202 | 0,1416 | 0,0293 |
| | | **0,0559** | | **0,0785** | |
| | nincs | 0,0047 | 0,0010 | 0,0079 | 0,0012 |
| A lánynak antigénje | van | 0,0273 | 0,0067 | 0,0424 | 0,0104 |
| | | **0,0166** | | **0,0251** | |
| | nincs | 0,0022 | 0,0005 | 0,0039 | 0,0010 |

1. táblázat

A "vastagon irt számok" közül pl. a **0,0559** annak a valószinüsége, hogy egy SA beteg apának a fia is SA beteg lesz.

A kapott kockázati értékek orvosi értelmezése nem feladatunk, az megtalálható pl. [9] -ben és [10]-ben. Eredményeinket több helyen ismertettük [9], [10], [19], [20], [21], [22], [23] és [34].

## 2.4. Megjegyezések

a) Abban az esetben, ha a (2) feltétel nem teljesül, a

$$P(\chi_{ej}=k)\,, \quad P(\chi_{bj}=k) \qquad (k,j=1,2)$$

valószinüségeket is a megfelelő populációkból kell becsülni. A 2.2.-ben mondottakat mindez nem érinti, ugyanis azokban a (2) feltételt nem használtuk.

b) Ha a betegség nem multifaktoriális kóreredetü, a 2.2.-ben leirt módszert akkor is alkalmazhatjuk, csupán a

$$P(\omega_{gi}=j \mid \omega_{s1}=k,\, \omega_{s2}=\ell)$$
$$(i=1,2\,; \quad j,k,\ell=0,1)$$

valószinüségeket kell a megfelelő populációkból becsülni vagy a multifaktoriális modell helyett érvényes másik modell alapján kiszámolni.

c) A (3) valószinüségeket ugy választottuk ki a lehetséges valószinüségek közül, hogy az egyenletrendszer determinánsát nullától különbözővé tegyük.

d) Feltétlenül hasznos lenne olyan, mikrogépre kidolgozott programcsomag, amely egy vizsgált betegségre vonatkozóan elvégzett teszt-sorozat után tisztázná a betegség genetikai hátterét matematikai szempontból . Azért javasolnánk a mikrogépes kidolgozást, mert véleményünk szerint a mikrógép a kórházak számára is elérhető közelségbe került. Ebben az esetben a 1. pontban felsorolt módszerek közül a d) rekurziv eljárást fogjuk használni.

e) A 2.3. pontban megadott "vastagon irt" valószinüségeket a Pareto modell alapján is kiszámoltuk. Az eredmények:

| | Szülők egészségesek | | Csak anya beteg | | Csak apa beteg | | Mindkét szülő beteg | |
|---|---|---|---|---|---|---|---|---|
| Modell | Lány | Fiu | Lány | Fiu | Lány | Fiu | Lány | Fiu |
| Multifaktoriális | 0,0007 | 0,0037 | 0,0251 | 0,0785 | 0,0166 | 0,0559 | 0,2238 | 0,4288 |
| Pareto | 0,0007 | 0,0037 | 0,0230 | 0,0655 | 0,0473 | 0,1148 | 0,2650 | 0,4059 |

2. táblázat

Látható*, hogy komoly különbségek vannak a két modell kockázati értékei között.

f) A disszertációban ismertetett módszer végleges kidolgozása előtt egyéb matematikai eredmények felhasználásával is próbálkoztunk, (pl. [16]) de azok nem adtak megfelelő eredményt.

* statisztikai próbával is bizonyitható

Hivatkozások a II. fejezethez

[1] Bene Béla: Reichentransformationen zur gemeinsamen Normalverteilung von familiaren Erkrankungsneigungen, Proc. of the Symp. on Comp. Stat, Wien, 1974.

[2] D.A. Brewerton et al: Ankylosing spondylitis and HLA-B27, Lancet i, 1973, pp. 904-907.

[3] R.N. Curnow-C.Smith: Multifactorial models for familial diseases in man, J.R. Statist. Soc. A., 138, part 2, 1975, pp. 131-169.

[4] A. Czeizel-G. Tusnády: Aetiological Studies of isolated common congenital abnormalities in Hungary, Akadémiai Kiadó, Budapest, 1984.

[5] Deák István: Monte Carlo módszerek a többdimenziós térben elhelyezkedő halmazok valószinüségének meghatározására normális eloszlás esetén, Alkalmazott Matematikai Lapok, 4, 1978, 35-94.

[6] R.C. Elandt-Johnson: Probability models and statistical methods in genetics, John Wiley & sons, New York, 1971.

[7] A. E. H. Emery - J. S. Lawrence: Genetics of ankylosing spondylitis, J. Med. Genet. 4, 1967, pp. 239-243.

[8] Gömör Béla: A spondylitis ankylopoetica epidemiológiája és genetikája, Kandidátusi disszertáció, Budapest, 1976.

[9] B. Gömör - I. Ratkó: Determining the risk of ankylosing spondylitis recurrence risk by combining data of polygenic model and B27 characteristics, Rheumatológiai kongresszus, Párizs, 1981, pp. A6/0235.

[1o] B. Gömör - I. Ratkó: Calculation of recurrence risk in the case of ankylosing spondylitis taking into consideration the antigen HLA-B-27, Ann. Hum. Genet., 48, 1984, pp. 79-86.

[11] T.P. Hutchinson: An easy method of calculating approximate recurrence risk using multifactorial model of disease transmission, Ann, Hum. Genet. London, 43, 198o, pp. 285-293.

[12] K.K. Kidd et al: Genetic analysis of HLA-associated diseases: the "illness-susceptible", gene frequency and sex ratio in ankylosing spondilytis, HLA and Disease, eds. J. Dausset, A. Svejgaard, Munksgaard, Copenhagen, 1977, pp. 72-79.

[13] K.K. Kidd - R. Ceppelini: A Bayesian method for estimation of HLA-associated illness-susceptibility Is , Allele frequencies, I. Dominant Susceptibility, HLA and Disease, eds. J. Dausset A. Svejgaard, Munksgaard, Copenhagen, 1977, pp. 81-83.

[14] K. Lange - R.C. Elston: Extensions to pedigree analysis, I. Likelihood calculations for simple and complex pedigrees, Hum. Hered. 25, 1975, pp. 95-1o5.

[15] K. Lange et al: Extensions to pedigree analysis, II. Recurrence risk calculation under the polygenic threshold model, Hum. Hered., 26, 1976, pp. 337-348.

[16] Medgyesi Pál: Sürüségfüggvények és diszkrét eloszlások szuperpoziciójának felbontása, MTA III. osztály közleményei, 21, 1972, pp. 129-2oo.

[17] K.V. Mardia: Multivariate Pareto distributions, Annals of Math. Stat., 33, 1962, pp. 1oo8-1o15.

[18] J. Ott: Maximum likelihood estimation by counting methods under polygenic and mixed models in human pedigrees, Am. J. Hum. Genet., 31, 1979, pp. 161-175.

[19] Ratkó István: Spondylitis ankylopoetica a genetikai tanácsadásban, Neumann J. Sz.T. SzKMAOB 9. kollokviuma, Szeged, 1978, pp. 31-39.

[2o] Ratkó István - Gömör Béla: Öröklődés kockázatának számitása multifaktoriális modell esetén adott jellemző ismeretében, Neumann J. Sz.T. SzKMAOB lo. kollokviuma, Szeged, 1980, pp. 153-162.

[21] I. Ratkó - B. Gömör: Calculation of recurrence risk in the case of spondylitis ankylopoetica in the knowledge of the antigen HLA-B27,3-rd Hungarian Biometric Conference, Budapest, 1981, pp. 217-221.

[22] I. Ratkó - B. Gömör: Calculation of recurrence risk for multifactorial disease for given trait, 3-rd. International Conference on System Science in Health Care, München, 1984, pp. 487-490.

[23] I. Ratkó - B. Gömör: Kalkulo de risko de heredigo surbaze de multifaktora modelo, Matematiko Translimen

[24] T. Reich et al: The use of multiple thresholds in determining the mode of transmission of semi-continuous traits, Ann. Hum. Genet. London, 36, 1972, pp. 163-184.

[25] T. Reich et al: The use of multiple thresholds and segregation analysis in analyzing the phenotypic heterogenity of multifactorial traits, Ann. Hum. Genet. London, 42, 1972, pp. 371-390.

[26] L. Schlosstein et al: High association of an HLA antigen, W27, with ankylosing spondylitis, New Engl. J. Med. 288, 1973, pp. 704-706.

[27] C. Smith: Computer programme to estimate recurrence risks for multifactorial familial disease, British Medical Journal, 19, febr. 1972, pp. 495-497.

[28] C. Smith: Recurrence risks for multifactorial inheritance, Amer. J. Hum. Genet., 23, 1971, pp. 578-588.

[29] C. Smith - N.R. Mendell: Recurrence risks from family history and metric traits, Ann. Hum. Genet. London, 37, 1974, pp. 275-286.

[30] Szántai Tamás: Egy eljárás a többdimenziós normális eloszlásfüggvény és gradiense értékeinek meghatározására, Alkalmazott Matematikai Lapok, 2, 1976, pp. 27-39.

[31] Tusnádi Gábor: A multifaktoriális öröklődés, Matematikai Lapok, 20, 1969, pp. 389-396.

[32] Tusnády Gábor - Telegdi László - Czeizel Endre: Gyakori veleszületett rendellenességek öröklődésmenetének vizsgálata, Alkalmazott Matematikai Lapok, 4, 1978, pp. 1-25.

[33] G. Tusnády - A. Czeizel - L. Telegdy: ML-fitting of multifactorial threshold models, Periodica mathematica hungarica, Vol. 12., 3, 1981, pp. 205-216.

[34] И. Ратко-Б.Гомор: Вычисление риска наследственности по мультифакторной модели со знанием данной характерности, монография РГ-I6 КНВВТ, в печати, I984, Москва.

# III. NÉHÁNY STATISZTIKAI MÓDSZERRŐL

Az, hogy egy adott adathalmaz statisztikai feldolgozásához milyen statisztikai adatfeldolgozó rendszert használjunk, függ az adathalmaz méretétől.

Közepes méretü adathalmaz feldolgozására alkalmas a BMDP (Biomedical Computer Programs) [10], a KBAЗAP [41] és a GUHA (General Unary Hypotheses Automaton) [26].

Nagyméretü adathalmaz feldolgozására nálunk is készült adatfeldolgozó rendszer [36].

Hazánkban talán a BMDP használata terjedt el leginkább.

A disszertáció szerzője a 70-es évek közepétől vesz részt olyan gyakorlati munkákban, amelyek különböző statisztikai módszerek használatát igényelték [2], [15], [17], [18], [19], [20], [21], [22], [28], [29], [33], [34].

Ezek a munkák a más szakemberekkel (orvosok, szervezők, stb.) való szoros eggyüttmüködést követelték meg. Ezen együttmüködés alapján alakult ki a statisztikai módszereknek egy olyan halmaza, amelyek elemei lépten-nyomon alkalmazásra kerültek.

Ennek a fejezetnek a célja egyrészt egy a gyakorlati munkák során kapott speciális valószinüségszámitási eredmény ismertetése, másrészt egy olyan "statisztikai módszer-gyüjtemény" ajánlása, amely kisebb méretü mikrogépen megoldható adathalmaz feldolgozására is alkalmas.

1. Egy mintavételi feladat

Statisztikai vizsgálatoknál általában nem készitünk a teljes populációról adatfelvételt. Ennek egyik alapvető oka a költségek csökkentése. Mintavételnél mégis két fontos szempontot figyelembe kell venni; a) a minta lehetőleg pontosan a teljes populáció egy meghatározott hányada ( mondjuk h %-a) legyen b) a minta reprezentativ legyen, azaz a vizsgált populáció egyes részei arányosan kerüljenek a mintába.

Mi most csak az a) kérdéssel foglalkozunk.

Az alábbi modellel dolgozunk: (i) A populáció minden egyede N osztály valamelyikébe tartozik; (ii) Minden egyednek van két u.n. részazonositója: d és m; d lehetséges értékei 1,2,..., D; m lehetséges értékei: 1,2,...,M. Bármely egyedhez egyetlen d és egyetlen m érték tartozik. (iii) minden egyednek vannak további adatai (iv) egy egyedről azt mondjuk, hogy többszörösen nyilvántartott, ha a populációban található két v. több olyan egyed, melyeknek bizonyos adatai (az u.n. azonositók) megegyeznek. (v) A populáció két részazonositó szerinti eloszlása egyenletes. (vi) az adatok egymástól függetlenek.

A cél olyan mintavételi eljárás megadása, amely minden osztályra h %-os mintát biztosit. Vizsgáljuk meg ebből a szempontból az alábbi minta kiválasztó algoritmust mely minden egyes osztályra külön-külön hajtandó végre .

1. Válasszuk ki a populációból az összes olyan egyedet, amelyeknek d-részazonositójának értéke: 1,2,..., $\left[\frac{Dh}{1oo}\right]$
2. Legyen H=M
3. B < S esetén menj 4-re, egyébként menj 9-re, ahol B= a mintába belevett, utolsó utasitásnak elegettevő egyedek száma,

   S = az adott osztály összes egyedei számának h %-a.
4. Vegyük be a mintába a H m-részazonositóju és $\left[\frac{Dh}{1oo}\right]+ 1$ d-részazonositóju egyedeket.
5. Ha B < S akkor menj 7-re.
6. Hagyjuk el a mintából a H m-részazonositóju, $\left[\frac{Dh}{1oo}\right]+ 1$ d-részazonositóju egyedekből véletlenszerüen annyit, hogy S egyed maradjon.

   STOP
7. Ha H=1, akkor STOP.
8. Legyen H=H-1. Menj 4-re.
9. Hagyjuk el a mintából a H m-részazonositóju, 1 d-részazositóju egyedeket.
10. Ha B < S menj a 11-re egyébként a 12-re.
11. Vegyük bele a mintába a H m-részazonositóju, 1 d-részazositóju egyedekből véletlenszerüen annyit, hogy S egyedet kapjunk. STOP.
12. Ha H=1, akkor STOP.
13. Legyen H=H-1. Menj 9-re.

Az alábbi tétel biztositja, hogy a leirt algoritmus nagy valószinüséggel h %-os mintát eredményez.

<u>1. tétel:</u> Jelölje $P_i$ annak valószinüségét, hogy egy egy adott L létszámu osztály-populáció 1 v. 2 v...., v. $\frac{Dh}{100}$* d-részazonosítóju egyedeinek a száma legalább $\frac{Lh}{100}$* (azaz h %-os). Ekkor

$$P_1 \leq 0{,}00001, \text{ ha } L \geq \frac{\left[2\left(\frac{Dh}{100}-1\right)\left(1-\frac{\frac{Dh}{100}-1}{D}\right)+1\right]^2 5\ln 20}{2\left(\frac{Dh}{100}-1\right)\left(1-\frac{\frac{Dh}{100}-1}{D}\right)\frac{1}{D}} \tag{1}$$

$$\text{és} \quad \frac{100}{h}+\frac{100}{h\sqrt{1-h/100}} \leq D \tag{2}$$

$$P_2 \approx 0{,}5$$

$$P_3 \geq 0{,}9999, \text{ ha } L \geq \frac{\left[2\left(\frac{Dh}{100}+1\right)\left(1-\frac{\frac{Dh}{100}+1}{D}\right)+1\right]^2 5\ln 20}{2\left(\frac{Dh}{100}+1\right)\left(1+\frac{\frac{Dh}{100}+1}{D}\right)\frac{1}{D}} \tag{3}$$

és

$$\left(1+\frac{h}{100}+\sqrt{1+\frac{3h}{100}}\right)\Big/\left[\frac{h}{100}\left(1-\frac{h}{100}\right)\right] \leq D \tag{4}$$

∎

<u>Bizonyitás</u>

Egyrészt nyilvánvaló, hogy

$$P_2 = P\left(\ell \geq \frac{Lh}{100}\right) = \sum_{k=\frac{Lh}{100}}^{L} \binom{L}{k}\left(\frac{\frac{Dh}{100}}{D}\right)^k\left(1-\frac{\frac{Dh}{100}}{D}\right)^{L-k}$$

* Az [ ] jelet itt és a későbbiekben elhagyjuk, ez nem változtat a bizonyitás lényegén

ahol $\ell$ az L létszámu osztálypopuláció 1 v 2 v. ..., $\frac{Dh}{100}$ d-részazonositóju egyedeinek a száma.

A Moivre-Laplace tétel miatt

$$P_2 \approx 1 - \phi\left(\frac{\frac{Lh}{100} - L\frac{\frac{Dh}{100}}{D}}{L\frac{\frac{Dh}{100}}{D}\left(1 - \frac{\frac{Dh}{100}}{D}\right)}\right) \approx 0,5$$

Másrészt

$$P_1 = P\left(\ell \geqq \frac{Lh}{100}\right) = P\left(\frac{\ell}{L} \geqq \frac{h}{100}\right)$$

Legyen $\quad q_1 = \dfrac{\frac{Dh}{100} - 1}{D}$

Mivel

$$\frac{\ell}{L} \geqq \frac{h}{100} \Rightarrow \frac{\ell}{L} - q_1 \geqq \frac{1}{D}$$

ezért

$$P_1 = P\left(\frac{\ell}{L} \geqq \frac{h}{100}\right) \leqq P\left(\left|\frac{\ell}{L} - q_1\right| \geqq \frac{1}{D}\right)$$

(2) -ből következik, hogy

$$0 < \frac{1}{D} \leqq q_1(1 - q_1)$$

ezért alkalmazhatjuk a Bernstein-egyenlőtlenséget

$$P_1 \leqq P\left(\left|\frac{\ell}{L} - q_1\right| \geqq \frac{1}{D}\right) \leqq$$

$$\leqq 2\exp - \left[\frac{L/D^2}{2q_1(1-q_1)\left(1 + \frac{1/D}{2q_1(1-q_1)}\right)^2}\right] =$$

$$= 2\exp - \left[\frac{2Lq_1(1-q_1)}{(2Dq_1(1-q_1) + 1)^2}\right]$$

(1) felhasználásával azonnal következik,

$$P_1 \leqq 0,00001 \qquad \text{(1) és (2) mellett.}$$

Végül

$$1-P_3 = P\left(\ell < \frac{Lh}{100}\right) = P\left(\frac{\ell}{L} < \frac{h}{100}\right)$$

Legyen $$q_3 = \frac{Dh/100+1}{D}$$

Mivel $\ell/L < h/100 \Rightarrow q_3 - \ell/L > 1/D$ ezért

$$1-P_3 = P\left(\frac{\ell}{L} < \frac{h}{100}\right) \leqq P\left(\left|q_3 - \frac{\ell}{L}\right| \geqq \frac{1}{D}\right)$$

(4) -ből következik, hogy

$$0 < 1/D < q_3(1-q_3)$$

ezért most is alkalmazhatjuk a Bernstein-egyenlőtlenséget (ugyanugy átalakitva):

$$1-P_3 \leqq P\left(\left|q_3 - \frac{\ell}{L}\right| \geqq \frac{1}{D}\right) \leqq$$

$$\leqq 2\exp -\left[\frac{2Lq_3(1-q_3)}{(2Dq_3(1-q_3)+1)^2}\right]$$

(3) felhasználásával kapjuk, hogy

$$1-P_3 \leqq 0{,}00001, \text{ azaz}$$

$$P_3 \geqq 0{,}99999 \quad (3) \text{ és } (4) \text{ mellett.} \quad \boxtimes$$

1. példa (részletesen ld. [29])

Az évenkénti kórházi morbiditási adatokból osztályonként lo %-os mintát kell kiválasztani. A mintába kerülés feltétele az, hogy (3o napos hónapot feltételezve) két, három v. négy nap (pl. 4.-e, 14.-e, 24.-e és 6.-a) valamelyikén legyen a beteg születésnapja. A tétel jelöléseit megtartva - most a születésnap napja a d-részazonositó - a következők adódnak:

$$P_1 \leqq 2\exp\left(-\frac{28L}{5041}\right), \qquad P_2 \approx 0{,}5, \quad P_3 \geqq 1-\exp\left(-\frac{52L}{14161}\right)$$

Az 1972-73 évi adatokra, a legkisebb létszámu intenziv osztályra ez azt jelenti, hogy

$$P_1 \leqq 0{,}00008, \qquad P_3 \geqq 0{,}9974 \qquad \boxtimes$$

Megjegyzés

Az általunk adott algoritmus azt eredményezheti, hogy lehetnek olyan többszörösen nyilvántartott egyedek, akiknek nem mindegyik "példánya" kerül bele a mintába, s ezért bizonyos adatszolgáltatások pontossága megkérdőjelezhető.

A többszörösen nyilvántartott egyedek matematikai vizsgálatát Soltész János végezte el [29], [33], ennek eredményeivel most nem foglalkozunk.

## 2. Statisztikai eljárások használatáról

Orvosok, biológusok és egyéb - nem matematikus - szakemberek lépten-nyomon használnak különböző statisztikai eljárásokat. Ugy tünik azonban, nem minden alkalommal a vizsgált problémának legjobban megfelelő eljárást alkalmazzák, sőt néha teljesen rossz, a problémára nem is alkalmazható módszert használnak.

Mit tegyünk ekkor? A válasz megadásánál vegyük azt is figyelembe, hogy a kérdéseire választ kereső szakember mikrogépet is igénybe vehet. A javasolt teendőnél [5] és [39] észrevételeit vettük figyelembe.

Első lehetőség: a szakember "adattömegét" odaadja a matematikusnak, elmondja kérdéseit, a matematikus az adattömeget mikrogépre viszi, s a szükséges programokat lefuttatja.

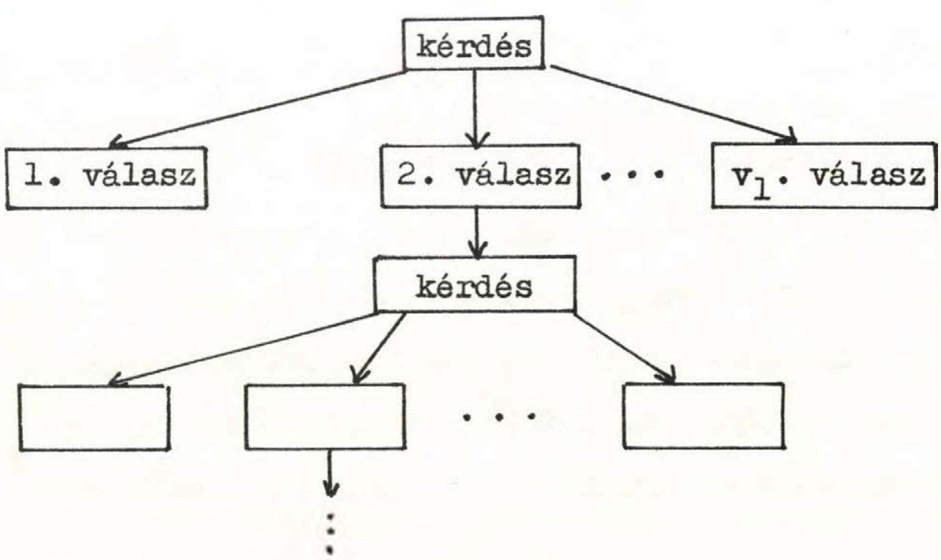
Második lehetőség: a szakember (v. másképpen felhasználó) odaül a mikrogép elé, saját maga v. megfelelő adminisztrátor mikrogépre viszi az adattömeget, s a mikrogép kérdéseire válaszolgatva (itt már a szakember és nem az adminisztrátor) a mikrogép választja ki a kapott válaszok alapján a végrehajtandó statisztikai eljárás oka t.

Mi most itt csak a - véleményünk szerint előnyösebb - második lehetőséggel foglalkozunk.

Először is megjegyezzük, hogy az adatbevitellel egy időben a következőket is elvégzi az adatbevitelt ellenörző program:

- adatok ellenőrzése: a) adott határok közé esnek-e a megadott adatok b) nincs-e valamelyik adatpár v. adathármas értékei között összeférhetetlenség (pl. ha az egyik adat a nem, a másik adat az ápolási kórforma, a harmadik adat az életkor, akkor pl. férfi nem szülhet v. 80 évnél idősebb nő nem szülhet)
- A bevitt adatokat "figyeli" a mikrogép és az adatbevitel befejezésekor megjegyzi, hogy az adatok közül melyik tekinthető diszkrétnek, melyik folytonosnak. Ez azért lényeges, mert ha csak a felhasználóra bizzuk a döntést, előfordul, hogy diszkrét változót folytonosnak kezel v. forditva. Ugyancsak az adatbevitel befejezése után az is "kiderül", hogy az egyes változók közül van-e speciális eloszlásu (Normális, Poisson, exponenciális, binomiális, stb).

A statisztikai módszert kiválasztó algoritmus strukturája egy irányitott G gráffal szemléltethető:



A kérdések célja az alábbiak eldöntése:

- Mely adatok vesznek részt a statisztikai feldolgozásban? Az adatok értékei között vannak-e átkódolandók?
- Az adatok nominális, ordinális v. intervallum skálájuak. Ez ugyanis fontos a legjobb módszer kiválasztása miatt.
- Kapcsolat keresés; átlag, szórás számitás, összehasonlitás; eloszlások összehasonlitása; egyéb feladatelvégzés - a kitüzött cél.

Megjegyzés az "egyébről": természetesen arra is fel kell készülni, hogy esetleg a felhasználó kérdésére a programcsomag programjai közül egyik sem adja meg a választ; ki kell jelölni az ebben az esetben szükséges teendőt is, ami lehet az is: forduljon a matematikushoz ujabb tanácsért.

Az irányitott gráf terminális elemei mutatják meg a használandó statisztikai módszert.

Az összes egyéb információt (a minták függetlensége; egymintás, kétmintás, k-mintás; speciális eloszlásról van-e szó, minták elemszáma stb) az adattömeg és a kérdésekre adott válaszok alapján megkapjuk.

A G gráf majdnem rögzitett. Egyrészt ugyanis a terminális elemek "tartalma" a vélt legfrisebb eredmények alapján változhat, azaz egy adott esetben eddigi tudásunk alapján a legjobbnak itélt módszert használjuk, azonban pl. egy megjelent uj cikk esetleg egy jobb módszert közöl, ezért a programunknak ezt a részét ujra irjuk, kicseréljük. Másrészt az egyéb válaszhoz tartozó terminális elem megszünhet terminális elem lenni, uj "ágak" nönek, uj programrészek keletkezhetnek.

Egy ilyen tervezett G irányitott gráf, a hozzátartozó statisztikai módszereknek a gyüjteménye található [35] -ben. Annak kialakitásában felhasználtuk [1], [3], [4], [6], [7], [8], [9], [11], [12], [13], [14], [16], [23], [24], [25], [27], [30], [31], [32], [37], [38] és [40] eredményeit.

A szóbakerült programok elkészitése COMMODORE 64 mikrógépre folyamatban van, VARYTER mikrógépre pedig tervbe van véve.

Ez utóbbi gépen a PASCAL nyelv használata a programok elkészitését elegánsabbá, a rendszer müködését gyorsabbá teszi.

Hivatkozások a III. részhez

[1] Arató Mátyás: Fejezetek a matematikai statisztikából számitógépes alkalmazásokkal I-II, SZÁMKI Közlemények, 22/1979.

[2] M. Bély, I. Ratkó et al: Clinical and histological evaluation of synovial needle - biopsies in patients suffering from rheumatoid arthritis. I. Relationship between clinical activity and histological pattern, Acta Morphologica Hungarica, 32 ( 1), 1984, pp. 57-65.

[3] B.O. Bengtsson, - G. Thomson: Measuring the strength of associations between HLA antigens and diseases, Tissue antigens, 18, 1981, pp. 356-363.

[4] C. Bonaiti: Genetic counselling of consanguineous families, Journal of Medical genetics, 15, 1978, pp. 109-112.

[5] A.J. Boreham: How far should and could those who produce statistics engage in research and analysis?, 40 th Session of the International Statistical Institute, Invited Paper, Warsaw, Sept. 1-9, 1975, pp 14/1-15.

[6] Börzsönyi László: Egymással kölcsönhatásban levő betegségek erősségének uj mérési lehetőségei, SzKMAOB 9. kollokviuma, Szeged, 1980, pp. 93-100.

[7] H.J. Chen: On selecting a subset which contains all populations better than a control, Commun. Statist., Theor. meth., A9(8), 1980, pp. 851-864.

[8] M.A.A. Cox - R.L. Plackett: Small samples in contingency tables, Biometrika, 67, 1980, pp. 1-13.

[9] L. Devroye: The computer generation of Poisson random variables, Computing, 26, 1981, pp. 197-2o7.

[1o] W.J. Dixon, chief editor: BMDP stat. softw., 1981, University of California press.

[11] T.G. Donelly: Bivariate normal distribution, Comm. of the ACM, 16, 1973, pp. 462.

[12] R.C. Elandt-Johnson: Probability models and statistical methods in genetics, John Wiley & sons, New York, 1971.

[13] Bo Eriksson: Confidence levels using normal approximation to modifed t-statistics for dependent variables, J. Statist. Comput. Simul., Vol 13, 1981, pp. 131-148.

[14] L. S. Freedman: Watson's $U_N^2$ statistic for a discrete distribution, Biometrika, 68,3, 1981, pp. 708-711.

[15] Garádi János - Krámli András - Ratkó István - Ruda Mihály: Statisztikai és számitástechnikai módszerek alkalmazása kórházi morbiditás vizsgálatokban, MTA SZTAKI Tanulmányok, 35/1975.

[16] J.J. Gart: Statistical analyses of the relative risk, Environmental health perspectives, Vol 32., 1979, pp. 157-167.

[17] Gáspárdi Géza - et.al. - Ratkó István: A láb megbetegedése rheumatoid arthritisben II., Magyar Reumatológia, XXI, 1981, pp. 39-46.

[18] Gáspárdy Géza - Ratkó István et al: A metatarsalgia gyakorisága reheumatoid arthritises betegeken, Magyar Reumatológia, XXIII., 1982, pp. 21-26.

[19] Gáspárdy Géza - et al. - Ratkó István: A láb fájdalma és nyomásérzékenysége rheumatoid arthritisben, Magyar Reumatológia, XXIII., 1982, pp. 147-151.

[20] Gáspárdy Géza - Ratkó István - et al.: A láb megbetegedése rheumatoid arthritisben III., Magyar Reumatológia, XXIII., 1982, pp. 241-247.

[21] Genti György - Ratkó István - et al.: Fájdalom mérése artrosisos betegeken különböző tipusu skálákkal, Magyar Reumatológia, XXI, 1980, pp. 156-159.

[22] Gömör Béla - Ratkó István - et. al.: 31 SPA-ra veszélyeztetett HLA-B27-es fiatal egyén megfigyelésének eredményei, Szocialista Országok Rheumatológusainak Kongresszusa, Magyarország, 1981, pp. 183-184.

[23] Anders Green: The epidemiologic approach to studies of association between HLA and disease, Tissue antigens, 19, 1982, pp. 245-268.

[24] J.E. Grizzle - G.G. Koch: Some application of categorical data analysis to epidemiological studies, Environmental health perspectives, Vol 32, 1979, pp. 169-179.

[25] J.D.F. Habbema - et al.: The measurement of performance in probabilistic diagnosis, Methods of Information in Medidine, 17/4, Okt. 1978, pp. 217-246.

[26] P. Hajek - T. Havránek: The GUHA method - its aims and techniques twenty - four questions and answers , Int. J. Man- Machine Studies, 1o, 1978, pp. 3-22.

[27] Hajtman Béla: A Mc Nemar - próba néhány általánositása, MTA III. osztály közleményei, 1974, pp. 127-137.

[28] K. Konrád - et.al. - I. Ratkó: Histological evaluation of synovial needle bippsy taken prior to chemical synovectomy, on the basis of the therapeutic results, Hungarian Rheumatology, 1983 Supplementum, pp. 27-32.

[29] Krámli András - Ratkó István - Ruda Mihály - Soltesz János: A statisztikai adatfeldolgozás matematikai és számitástechnikai problémái, MTA SZTAKI Tanulmányok, 7o/1977.

[30] D. Krewski - B. Junkins: Sample size determination for the interval estimation of the mean or median of distribution, J. Statist. Comput. Simul., 13, 1981, pp. 169-17[illegible].

[31] E. Peritz: Exact tests for matched pairs: studies with covariates, Commun. Statist.-Theor. Meth., 11 ( 19), 1982, pp. 2157-2167.

[32] M.J. Phillips: The statistical analysis of a disparity test, J. Opl. Res. Soc., 31, 1979, pp. 159-167.

[33] Ratkó István - Soltész János: Az 1974 óta végzett kórházi morbiditásvizsgálat által felvetett statisztikai kérdések, NJSZT SzKMAOB 9. kollokvium, Szeged, 1978, pp. 52o-532.

[34] I. Ratkó: Kelkaj rimarkoj pri la statistika prilaboro de datumoj, Apliko de experanto en la scienco kaj tehniko, Zsolna, 1981, pp. 11-15.

[35] Ratkó István: Statisztikai próbák interaktiv használata, MTA SZTAKI, Kézirat, 1984.

[36] Ruda Mihály: A SIS77 statisztikai információs rendszer kialakitásának szempontjai, alkalmazásának és továbbfejlesztésének lehetőségei, MTA SZTAKI Tanulmányok, 86/1978.

[37] L. Sachs: Applied Statistics, a handbook of techniques, Springer-Verlag, New York, Heidelberg, Berlin, 1982

[38] G. Thomson - W. Bodmer: The genetic analysis of HLA and disease association, HLA and disease, Copenhagen, 1977, pp. 84-93.

[39] Vincze István: A statisztikai következtetés és korlátai, Magyar Tudomány, 11-12, 1981, pp. 902-912.

[40] А.Афифи-С.Эйзен: Статистический анализ-подход с использованием ЭВМ,Москва,"МИР",I982.

[41] В.Д.Мазуров и.т.д.:Пакет КВАЗАР прикладных программ распознавания образов,УНЦ АН СССР,Свердловск,I97[illegible].

# IV. BETEGSÉGREGISZTEREK

Számitógépek alkalmazásának jelentősége betegek, betegségek nyilvántartásával kapcsolatban közismert. Ld. pl. [1], [3], [5], [6], [7], [8], [9], [21] és [26].

Ebben a részben két regisztert ismertetünk, melyek megvalósitásra kerültek.

Az egyik az Infarktusregiszter, amely a CDC-3oo-as gépen müködik, a másik a Szivmütétre várakozók regisztere, amely az IBM 3031 gépre lett kidolgozva (de kezdetben a HwB 66/2o gépen müködött).

A Regiszterek célja többféle lehet:

- a betegségek incidenciájának jobb ismerete
- egészségügyi szervezési programok kiértékelésében segitségadás
- a még meglévő manuális nyilvántartás helyettesitése, a nyilvántartás pontosságának fokozása
- az összegyült adatok birtokában a tudományos kutatás elősegitése (erre példa [2]); stb.

## 1. Az infarktus regiszter

A Délpesti Infarctus Regiszter 1970 óta müködik, biztositja a terület infarctus incidencia változás megismerését és alapul szolgál a gondozási rendszerek és a bevezetett egészségügyi programok hatékonyságának felmérésére.

1976-ban az Egészségügyi Minisztérium támogatásával kezdhettük meg az igen nagy adminisztrációt igénylő regiszterre egy adatkezelő rendszer kidolgozását. 1978 óta a rendszer müködik, ezzel kapcsolatban megjelent közleményeinket az irodalomjegyzékben ismertetjük. Az adatkezelő rendszer lehetővé tette a megbizhatóbb és gyorsabb értékelést.

Az infarctus incidenciájának ismerete az egyetlen lehetőség arra, hogy meghatározhassuk a betegség halálozási arányaiban bekövetkezett változások okát. Néhány országban az ischaemiás szivbetegség halálozási aránya jelentősen csökkent. A változások okát részint a betegség kezelésében és ellátásában történő jelentős változás eredményeként könyvelik el, másrészt a preventiv intézkedések hatásának tartják. Az ok eldöntése nem akadémikus kérdés, mivel az egészségügy anyagi eszközei még a leggazdagabb országokban is korlátozottak, a leghatékonyabb felhasználása gyakorlati probléma. Amennyiben a halálozás csökkenése változatlan incidencia mellett következik be, a megjavitott betegellátásnak /intenziv osztályok, koronaria bypass mütétek, uj gyógyszerek, stb./ tulajdonitható inkább a halálozási arányban elért eredmény, mig az incidencia csökkenés elsődlegessége esetén a preventiv eszközök bevetésének köszönhető a fentemlitett jótékony hatás.

Az Egészségügyi Világszervezet /EVSz/ az 198o-as évekre várható változások miatt speciális incidencia monitorozási rendszerek létrehozását javasolja a tagországoknak.

Az általunk kidolgozott rendszer szerepel az EVSz dokumentumai között, mint az egyik megvitatásra javasolt program.

A rendszerrel kapcsolatban megjelent és megjelenés alatt lévő közlemények: [4], [1o], [11], [12], [13], [16], [17], [24] és [25].

1.1. Az infarctus regiszter célja, szervezete és müködése

Mind a hazai, mind a fejlett országok mortalitási statisztikái egyértelmüen mutatják a keringési betegségek, ezen belül az acut myocardialis infarctus /AMI/ jelentőségét. Hazánkban a keringési betegségek az összhalálozás több mint feléért felelősek, s az AMI következtében elhaltak az összmortalitás több mint 1o %-át teszik ki. Az AMI gyakorisága és rokkantságot okozó hatása az aktiv kereső életkorban a betegség elleni küzdelmet még sürgetőbbé teszi.

Ennek első lépéseként az EVSz javaslatára 197o-ben a délpesti egészségügyi ellátás területén megszerveztük az infarctus regisztert /IR/. 1972 óta egyszerüsitett formájában tovább folytattuk.

Az infarctus regiszter célja

1/ Meghatározni az AMI incidenciáját, mortalitását és letalitását, egy területileg pontosan meghatározott populációban.

2/ Megismerni a betegség természetrajzát, különös tekintettel a betegség korai, kórházon kivüli szakaszára, a rohamot megelőző tünetek jelentőségére, a feltételezett rizikó-faktorok prevalenciájára, a betegség klinikai formária és kórbonctanára vonatkozóan.

3/ Feltárni az AMI ellátási problémáit; az acut szakaszban a kórházon kivül és belül, majd a kórházi ellátást követően.

4/ Megfigyelni az évek során a betegség incidenciájának mortalitásának, letalitásának, természetrajzának változását az életkörülmények az egészségügyi ellátás megváltoztatásának hatására.

5/ Felállitani egy olyan ellátási modellt, mely információival alapja a további kutatásoknak, egészségügyi szervezési programok kiértékelésének és mint operativ egység a kardiológiai gondozás részeként szerepelhet a későbbiekben.

Az infarctus regiszter müködése és szervezete

Az infarctus regiszter a volt délpesti egészségügyi ellátás területén müködik, mely a IX.,X.,XVII.,XVIII., XIX., és XX. kerület egészét foglalja magába. A terület lakossága félmillió fő.

A vizsgált területen az IR felkutat és nyilvántartásba vesz mindenkit AMI-val és annak gyanujával, és minden hirtelen és váratlanul elhunytat.

Az eredmények értéke a nyilvántartás teljességétől függ, csakugy mint minden más epidemiológiai vizsgálaté. A teljesség biztositására - hogy minden infarctus gyanus eset bekerüljön - nyilvántartásunknál a következő módszereket alkalmazzuk;

1/ Megfelelő együttmüködést épitettünk ki a területen dolgozó körzeti, kórházi orvosokkal, személyes és csoportos megbeszélések révén még az előkészitő szakaszban. Hasonlóképpen jó kapcsolatot alakitottunk ki a más minisztériumokhoz tartozók kórházakkal is, ahová a területi betegek egyéb jogosultságuk révén kerülhettek. Nagy segitségünkre volt a Fővárosi Tanács VB Egészségügyi Osztálya a megfelelő kerületi tanácsok egészségügyi osztályai, a délpesti területi igazgatói tanács, az Országos Mentőszolgálat, a MÁOTE és az üzemorvosok vezetősége, akik ezzel a munkával kapcsolatba kerülhettek.

2/ Egyidejüleg több bejelentési forrást használunk, igy egy-egy infarctus-gyanus személyről több bejelentés is érkezhet.

A következő információs forrásokat használjuk fel: az Országos Mentőszolgálat, kórházak ill. klinikák felvételi irodájának nyilvántartása, illetve az Igazságügyi Orvostani Intézet, kórbonctani intézetek és osztályok sectiói leletei. Felhasználjuk ezeken kivül a körzeti orvosi bejelentéseket és a halotti anyakönyv adatait is. A nyilvántartandó betegekről elsősorban aktiv felkutatás utján szerzett tudomást az IR. Minden bejelentésről az IR adminisztrátorai "Bejelentő lapot" töltenek ki, mely a beteg személyi adatain kivül, a bejelentési forrást, a beteg állapotát, a kórházba szállitás helyét, idejét, halálozás esetén a halál időpontját tartalmazza.

Minden bejelentettről a regiszter orvosai egységes "Alapvizsgálati lapot" töltenek ki, legkésőbb 28 nappal a roham kezdete után. Az Alapvizsgálati lap a beteg személyi adatain kivül a beteg állapotát, a roham és az ellátás időpontjait és helyét, és a diagnózis, valamint az epidemiológiai diagnózis kritériumait tartalmazza.

Az epidemiológiai diagnózis meghatározása négy alapvető kritérium alapján történik: 1. a fájdalom jellege az anamnézisben, 2. EKG, 3. enzimváltozások, 4. kórbonctani lelet. Ennek alapján a diagnosztikus kategóriák a következők:

1. Kétségtelen acut myocardialis infarctus /KAMI/
2. Lehetséges acut myocardialis infarctus /LAMI/

A regiszter teljességének biztositására félévenként a kórházi osztályok nyilvántartásait, a tanácsok halotti anyakönyvi kivonatait ismételten átnézik a regiszterek adminisztrátorai és egyeztetik a regiszterbe felvett esetekkel. A revizió megközelitőleg lo % uj eset felvételeit jelenti.

Az évente regisztrálásra kerülő szivrohamok száma 2ooo-25oo között mozog.

Az Infarctus Regiszter, müködésének megindulásakor, szervezetileg az Országos Kardiológiai Intézethez tartozott. Az első évek tapasztalatai azt bizonyitották, hogy tevékenysége kapcsolódik a területi munkához, regionalis jellegü, ezért 1973 óta a Főv. István Kórház-Rendelőintézet részlegeként müködik, kapcsolódva az I. Belosztályhoz, ahol a Regiszter orvosai osztályos munkájukat is végzik.

Az Országos Kardiológiai Intézet a számitógépes adatfeldolgozás terén segiti az Infarctus Regiszter tevékenységét.

A nyilvántartás gépesitésének szükségessége

Az 197o-ben elkezdett felmérés 1971 december végén zárult, az egységes utánvizsgálat pedig 1972-ben. Az adatfeldolgozást csak ezt követően tudtuk megkezdeni, 1973-ban, igy az eredmények az adatgyüjtés megkezdése után 4-5 évvel álltak csak rendelkezésünkre. Mivel az infarctus regiszter operativ nyilvántartás, és a döntés előkészitésében van jelentősége, a visszajelentés lassusága az egészségügyi szolgálat hatékonyabb müködését gátolja.

Az infarctus regiszter adatainak kézi tárolása nehézkes. Igy a rendszeres adatszolgáltatás a terület egészségügyi szakemberei számára csak igen nagy adminisztrativ apparátussal lenne csak megoldható. Ez tette szükségessé az infarctus regiszter számitógépes rendszerének kidolgozását. Ennek célja az, hogy rendszeres adatszolgáltatással, az egészségügy vezetőienk tájékoztatása mellett, a körzeti orvosokat segitse az infarctusos betegek gondozásában.

Az infarctus regiszter az egészségügyi ellátás és az egészségügyi igazgatás különböző szintjeinek szolgáltat információt:

1/ A körzeti orvos számára saját betegeiről,

2/ A délpesti terület kórházai számára a kürházigazgatónak és a kórházi osztályvezető főorvosoknak,

3/ A délpesti kerületek vezető főorvosainak,

4/ Tudományos kutatási célokra.

Tapasztalatok

Az immár 14 éve müködő infarctus regiszter program gépi rendszerének kidolgozása után a következő előnyöket tapasztaltuk:

1/ a nyilvántartás adminisztrációját nagymértékben csökkenti.

2/ a nyilvántartás pontosságát fokozza az állandó ellenőrzés biztositásával, 3/ jobban áttekinthetővé és könnyebben kezelhetővé teszi a nyilvántartást, mely a jobb gondozást elősegiti,

4/ rendszeres visszajelenetéssel nemcsak a nyilvántartást javitja, hanem módot nyujt a gyorsabb beavatkozásra,

5/ szükség esetén megfelelő információk a megszabott időpontokon kivül is gyorsan elkészülhetnek.

## 1.2. A rendszer általános leirása

Ebben a fejezetben leirjuk azt, milyen adatok kezelésére alkalmas az általunk kidolgozott rendszer.

### A rendszer szervezeti felépitése

A rendszer feltételez bizonyos forrásokat, ahonnan a rendszer nyilvántartását végző központ munkatársai információkat beszerzik vagy megkapják. Az információk objektumokra vonatkoznak és kétféle adatlapon jelentkezhetnek, egy u.n. bejelentő lapon /röviden BL/ és egy u.n. alapnyilvántartó lapon /röviden AL/. A BL-on az objektumok azonositásához szükséges adatok állnak, továbbá annak a forrásnak a megjelölése, ahonnan a BL-hez szükséges adatok beszerzése megtörtént. Az AL-on az azonositási adatok /ugyanaz mint BL-nél, esetleg ujabbakkal kiegészitve/, továbbá az objektumokra vonatkozó részletes speciális adatok szerepelnek, mely azok "történetére", "következményeire" utalnak. Az adatok között van néhány olyan adat, amelynek értékei az <u>összes objektum közül meghatározzák azokat, amelyeket be kell vennünk a rendszerbe.</u>

Vezessük be a következő jelöléseket:

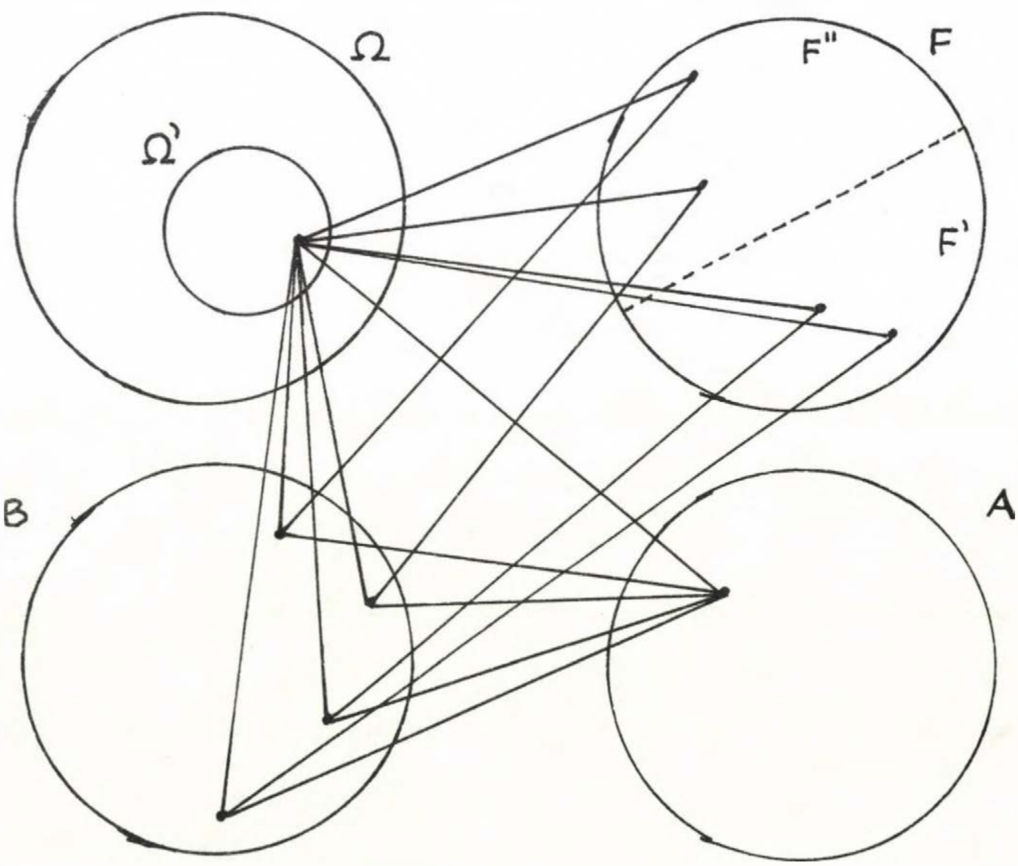
$\Omega$ : összes objektumok halmaza

$\Omega'$ : a rendszerbe beveendő objektumok halmaza

F : források halmaza  B : bejelentő lapok halmaza

A : alapnyilvántartó lapok halmaza

AL-ot nem minden forrás szolgáltat, csak F egy részhalmaza: F'. Legyen F"=F-F'. A halmazok elemeinek egymás közötti kapcsolatát az alábbi 1. ábra szemlélteti.




1. ábra

Ez a következőket jelenti:

/i/ Minden $\omega \in \Omega'$ objektum legalább egy F' beli f forrásnál nyilván van tartva.

/ii/ Minden $\omega \in \Omega'$ objektumhoz legalább egy b $\in$ B BL tartozik.

/iii/ Minden $\omega \in \Omega'$ objektumhoz pontosan egy a $\in$ A AL tartozik.

/iv/ Ha $\omega \in \Omega'$ objektumhoz az a $\in$ A tartozik, akkor B-nek ugyanazon elemei tartoznak $\omega$ -hoz, mint a-hoz.

/v/ Ha $\omega \in \Omega'$ objektumhoz az a $\in$ A tartozik, akkor F'-nek ugyanazon elemei tartoznak $\omega$ -hoz, mint a-hoz.

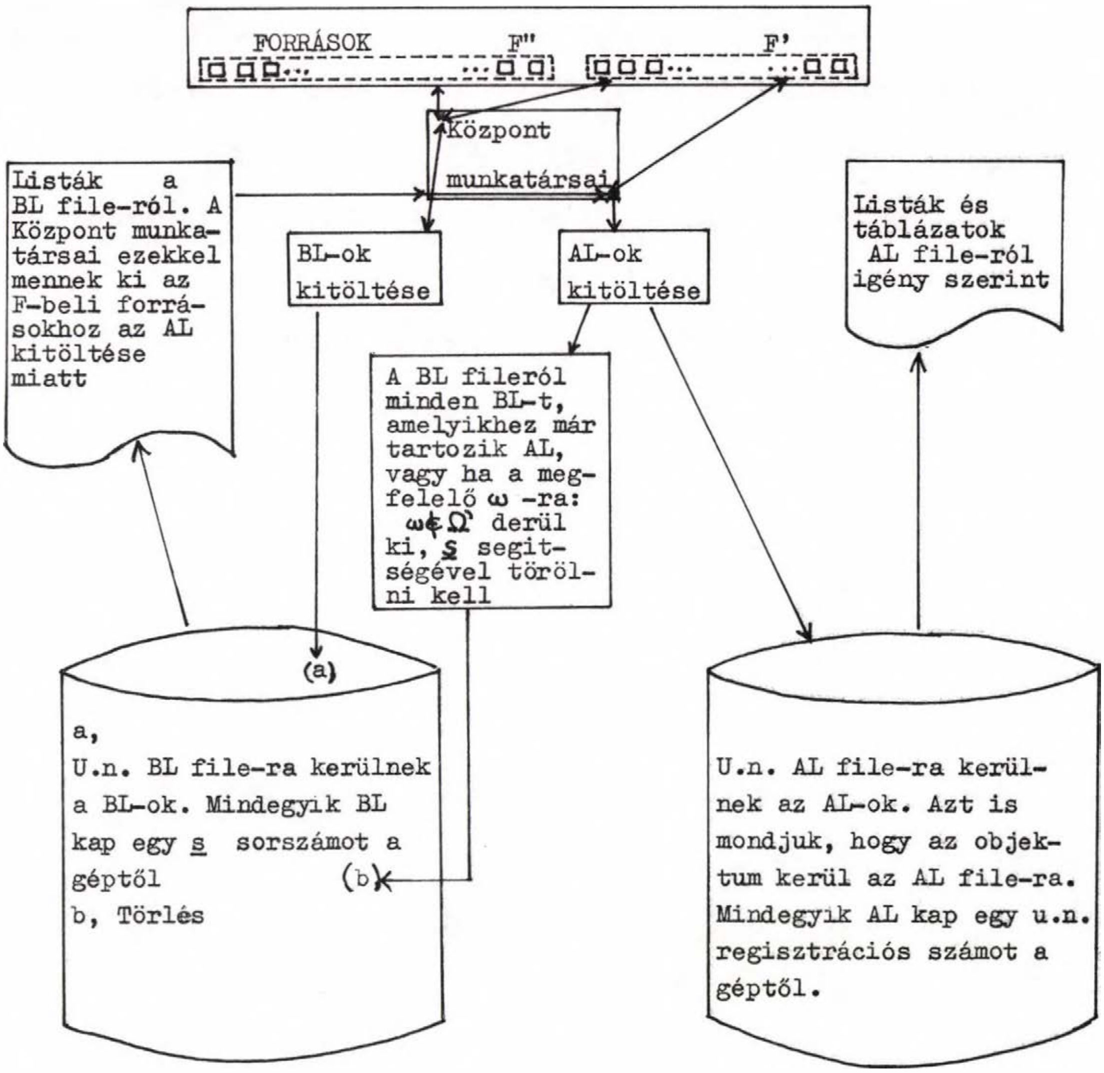
F,F',F" időben változatlan, $\Omega, \Omega'$,A,B változik, általában bővül. Azt, hogy mely objektumok tartoznak bele $\Omega'$-be, a Központ munkatársai is megmondhatják. Mivel tévedhetnek szükség van arra, hogy a B,A és $\Omega'$ halmazokat szükitsük.

A rendszer müködésének főbb mozzanatai az alábbi 2. ábrán láthatók.

Az adatok ellenőrzése

A BL-okról hetenként készül lista. Mivel a cél az, hogy minden $\Omega'$-beli objektumhoz elkészüljön az AL, a BL-nek csak közvetitő szerepe van. Mindezek miatt a BL hibaellenőrzése nem történik meg, csak az AL-oké.

A már emlitett SIS77-ben ([19], [20]) az adatok ellenőrzése speciális irányitott gráffal történik. Ennek alkalmazása akkor célszerü, ha nagytömegü adatot kevés alkalommal akarunk ellenőrizni. Esetünkben azonban az a helyzet, hogy sokszor kell kistömegü adatot ellenőrizni.


FORRÁSOK
F"
F'
Központ munkatársai
Listák a BL file-ról. A Központ munkatársai ezekkel mennek ki az F-beli forrásokhoz az AL kitöltése miatt
BL-ok kitöltése
AL-ok kitöltése
Listák és táblázatok AL file-ról igény szerint
A BL fileról minden BL-t, amelyikhez már tartozik AL, vagy ha a megfelelő ω -ra: ω∉Ω' derül ki, s segítségével törölni kell
(a)
a, U.n. BL file-ra kerülnek a BL-ok. Mindegyik BL kap egy s sorszámot a géptől
(b)
b, Törlés
U.n. AL file-ra kerülnek az AL-ok. Azt is mondjuk, hogy az objektum kerül az AL file-ra. Mindegyik AL kap egy u.n. regisztrációs számot a géptől.

2. ábra

Az adatok ellenőrzését végző program megvizsgálja, hogy az egyes adatok a megengedett értékek közé esnek-e, másrészt összeférhetetlenségi vizsgálatokat végez. Annak érdekében, hogy az egyes adatelemek lehetséges értékei egymásutáni egész számok legyenek - ekkor könnyü az ellenőrzés - bizonyos átkódolást kell végezni. Pl. ha egy adatelem lehetséges értékei 1,2,3,4 és 9, akkor a 9-et 5-re "átkódoljuk". Az adatelemek minimális és maximális értékei megfelelő adatfile-ról olvassuk be. Ily módon a korlátok közé esés vizsgálata egyetlen ciklusban elvégezhető.

Valamivel bonyolultabb a helyzet az összeférhetetlenségi vizsgálatokkal.

Tegyük fel, hogy $n$ darab összeférhetetlenségi vizsgálatot akarunk végeztetni. Legyen

$$\ell(i,j,k) = 1 \qquad (i = 1,2,\dots, n)$$

akkor és csak akkor, ha az i-edik összeférhetetlenség azt jelenti, hogy a

/1. adatelem értéke=j/ $\wedge$ /2. adatelem értéke=k/

konjunkció igaz /itt pl. az 1. adatelem az i-edik összeférhetetlenségi vizsgálatban résztvevő adatelemek közül az első, stb./

Az $\ell$ tömb kezdeti értékének beállitása a program elején történik meg.

Ezek után az összeférhetetlenségi vizsgálat - beolvasott vagy átkódolt j, k esetén -

IF( $\ell(i,j,k)$.EQ.1) i-tipusu összeférhetetlenség alaku utasitás megadását jelenti.
Hibás AL esetén az azonositásra használt adatelemek, valamint a hibakódok kerülnek kilistázásra.

2. Szivmütétre várakozók regisztere

1978 óta az Országos Kardiológiai Intézet korszerüen felszerelt Sebészeti Osztályán történik az ország extracorporalis szivmütéteinek több mint egyharmada. A mütéti várakozó listán szereplő betegek száma meghaladja az 1000 főt. A "nagyüzemi" méretek, a már-már áttekinthetetlen és állandóan változó adathalmaz miatt elkerülhetetlenül szükségessé vált a változásokat is rugalmasan követni tudó számitógépes nyilvántartás. A szivmütétre várakozó betegek adatait előzetes ellenőrzés, javitások és kiegészitések után 1980-ban géprevitelre kész állapotra hoztuk. Kidolgozott és további fejlesztésre alkalmas rendszerünk az 1980. év eleje óta müködik.
A rendszer müködéséről több helyen beszámoltunk. [4], [14], [15], [16], [17], [22] és [23], a részletes leirást [18] adja.

2.1. A rendszer célja

1. Szivmütétre várakozó betegek nyilvántartása, a mütét előttiffelülvizsgálatra való besorolásuk.

2. A mütétre váró betegekről adott feltételeknek eleget tevő, u.n. mütéti behivásra szolgáló listák előállitása.
3. A már mütött vagy valamilyen oknál fogva mütétre nem kerülő /pl. meghalt/ betegek archiválása.
4. Szivmütéten átesett betegek megyei gondozásának elősegitése, irányitása és ellenőrzése.
5. Mütétre várakozó betegek megyei gondozásának segitése, irányitása és ellenőrzése.
6. Különböző táblázatok, statisztikák készitése melyek az orvosok további munkáját /tudományos és szervezői/ elősegitik.

2.2. Az adatlapok tartalma és célja

Eddig háromféle adatlapot használtunk:

a/ Mütétre várakozók lapja /röviden: VÁL/. Ezen történik a várakozókról azoknak a fontos adatoknak a felvétele, melyek a behivás alapjául szolgálnak és információt adnak a kivizsgálás teljességéről.

b/ Mütét előtti felülvizsgálati lap /röviden: MEFL /. Célja, hogy a mütéti behivásnál figyelembe vehessük a felülvizsgálat során esetleg módosult illetve eddig hiányzó uj adatokat. /Pl. uj mütéti besorolás, halasztás stb./. További célja az előzetes szürővizsgálat eredményével csökkenteni a téves mütéti behivások számát. A MEFL tartalmazza a mütét előtti kivizsgálással kapcsolatos adatokat, valamint azt, hogy azok a betegek, akiket még mütétre nem javasoltak, hol és milyen szintü ellenőrzés alatt állnak. Módot nyujt ezek ujabb felülvizsgálati behivására is.

c/ Mütéti visszajelzési lap /röviden: VIL/. Több célt szolgál:

- Változásjelentés. A mütéti behiváskor fellépő változások jelzése, melyek miatt a mütét esetleg halasztásra kerül vagy végleg elmarad. Az uj mütéti besorolás jelzése, ha változott.
- Információt ad a számitógépnek arról, hogy kiket hivtak be mütétre, nehogy azokat a gép ujra kiirja, mint behivandókat, és ezzel megzavarja a behivás rendjét.
- Elrendeli azon egyéneknek archiválását, akiket megmütöttek, vagy akiknek a mütétjére nem kerül sor.
- Jelzést ad a számitógépnek azokról akiket valamilyen okból később kell hivni, - meghatározott vagy határozatlan időpontban.
- Azok archiválását is szolgálja, akiket törölni kell a várakozó listáról.

## 2.3. Az adatlapok használata

a/ VÁL

Az uj várakozókról az előjegyző orvos kitölti a VÁL-t /kódolni nem kell/ és hetente egyszer átadja a számitástechnikai csoportnak. Az adatok felvitele után a VÁL visszakerül a sebészeti osztályra, ott tárolják.

A számitógép minden betegnek ad egy sorszámot mely a azonositója lesz. Ez a szám kerül a VIL-re és a MEFL-re is.

b/ MEFL

Mütét előtt kb. két hónappal mütét előtti felülvizsgálat történik. A felülvizsgálatra behivandók listáját adott számu beteggel /ez a szám a szivsebészet kivánsága szerint változtatható/ meghatározott időnként adjuk meg, külön a felnőttekről, külön a gyermekekről. A lista átadása után a felülvizsgáló orvos dönthet ugy, hogy a felülvizsgálat szükségtelen, - visszajelzést akkor is kell adni "behivható" jelzéssel. Amennyiben a felülvizsgálat eredménye alapján a beteg archiválására kerül sor, ennek oka is kódolható.

Archivált beteg az eredeti előjegyzési időponttal visszahozható a várakozó listára, de ezt MEFL-pal kérni kell.

c/ VIL

VIL-pal kell jelezni a mütéti listáról behivottakat. /A már fentebb emlitett ok miatt./

Jelezni kell a behivás utáni történéseket /mütét megtörtént, - kontraindikált, - halasztott, nem jelent meg stb./.

Amennyiben halasztás esetén az ujabb behivás időpontja nincs megadva, a beteg 1 évig archiválásra kerül.

Acut betegek adataival kapcsolatos teendők

Azok az uj /előjegyzés nélküli/ acut betegek, akiknek a mütéte azonnal megtörténik, nem kerülnek a várakozók listájára. Ezért, hogy adataikat archiválhassuk a VÁL adatait a VIL "mütét megtörtént" jelzésével együtt visszük a gépbe.

Régi várakozó beteg soronkivüli, acut mütétének jelzése szintén VIL-on történik.

## 2.4. A betegberendelési rendszer szervezése és szolgáltatásai

Számitógépes futtatás hetente /egyszer/ történik. A behivandók száma mindig több /pl. kétszeres/, mint a mütétre kerülők száma, - egyrészt a kiválasztási lehetőség miatt, másrészt a már behivottak, de meg nem jelentek pótlásának érdekében. A szivsebészet előre közli: van-e valamilyen változtatás a régi feltételekben /ami az általánosan megadottaktól eltérő/ amit a következő behivásnál figyelembe kell venni. Pl. más férfi-nő, vércsoport, kor, mütéti besorolási arány stb. Bármikor lehet a szokásostól eltérő, speciális kivánság a behivásra, melyet egyedi esetként a számitógép teljesit. /pl. "A" Rh poz. vércsoportu és 1 ao.-coronaria saphena bypass graftra váró budapesti betegre lenne szükség/ A listán a mindhárom adatlapon felvett információk redundancia mentesen kerülnek kinyomtatásra. A szivsebészeti osztály hetente átadja az adatlapokat a számitástechnikai csoportnak, - a számitógépbe az adatok bevitele és hibaellenőrzés hetente történik. A behivási lista előjegyzési időpont és besorolási kategória szerint rendezett. Meghatározott időszakonként listát adunk azokról, akiket mütétre már behivtak és több mint 3o nap eltelt, de további sorsukról nem történt visszajelzés. A számitógép havonta megvizsgálja, hogy m nnyi a mütéti feltételeknek eleget tevő várakozók száma. Ha 6o alá esik ezek száma, a számitógép jelzést ad.

A várakozókról bármikor listákat szolgáltathatunk különböző szempontok szerint:

(1) Az ujonnan felvitt VÁL-okról, VIL-okról és MEFL-okról listák ha szükséges, a hibák feltüntetésével
(2) Mütéti behivást elősegitő lista
(3) Felülvizsgálatra behivó lista
(4) Speciális feltételeknek elegettevő névsor szerint
(5) Speciális feltételeknek elegettevő-valamilyen képzett szám v. adott érték szerint rendezett listák
(6) Megyei listák
(7) Budapesti lista, kerületenként
(8) Statisztikák

Mi ezen listák célja?

(1) Az ujonnan felvitt betegek ill. az adatváltozások könnyebb nyomonkövetésének elősegitése.
(2) - (3) A mütétre behivás alapvetően fontos feladatát ez a két lista kivánja megoldani.
(4) - (5) Az adott feltételknek megfelelő betegek esetleges nyomonkövetése, a listák távlati tervekben történő hasznositása.

(6) - (7) Az illetékes irányitó szervek, főorvosok, orvosok megfelelő szintü tájékoztatása.

(8) Információk gyüjtése a tudományos kutatás céljaira, továbbá a behivás rendszer javitására.

Statisztikák is készülnek, pl.

a./ mütéti tipus, vércsoport, kor

b./ mütéti besorolás, mütéti tipus, kor

c./ mütéti tipus, lakhely /pl. megyék/, kor

A felvitt betegekről egyéb statisztikák is adhatók. pl. 1 év alatt milyen mütéti besorolásu és mütéti tipusu betegek kerültek archiválásra, stb.

2.5. A mütéti behivást meghatározó logikai feltételek

A mütéti besorolás a következő feltételek figyelembevételével történik: besorolási kategória /soron kivül, nagyon sürgős, egyéb/, előjegyzési időpont, nem kor, vércsoport és a mütét nehézségi foka - nehéz mütét pl. két müDillentyüre váró beteg/.

1. A behivás sorrendjét az előjegyzés időpontjának évéből és hónapjából valamint a besorolási kategóriából állapitjuk meg. Az aorto-coronaria saphens bypass graftra váró betegeknél a coronarographia időpontját is figyelembe vesszük /ez a vizsgálat 6 hónap alatt elévül/.

2. A férfiak és a nők aránya ugyanaz az adott érték. Ez az arány igény esetén változtatható.

3. A 6 évesnél nem idősebb gyermeknél a nemre nem vagyunk tekintettel, /gyermek = a 14 évnél nem idősebb beteg/.

4. A különböző vércsoportokból a kivánt arányban szerepelhetnek a listán.

5. A mütét nehézségi fokánál figyelembeveendő feltétel: a mütétek legfeljebb kétharmad része lehet nehéz.

6. A listán a besorolási kategóriák adott arányban szerepelhetnek.

7. A felsorolt feltételeken kivül természetesen egyéb feltételek is megadhatók, de adott feltételek figyelmen kivül is hagyhatók.

A feltétel figyelembevételével kapcsolatban az I. fejezetre utalunk.

A már több mint egy éve müködő rendszerünkben eddig semmi lényeges korrekcióra nem volt szükség. A szivsebészet korábbi kartotékrendszere - ha szabad igy kifejezni - archivumba került. A gépre vitt adatok bármikor tetszés szerinti csoportositásban, bontásban lehozhatók. Igy kap például az ország megyénként információkat tőlünk az országos kardiológiai hálózat munkájának segitése céljából.

Dolgozunk a rendszer továbbfejlesztésén. Kibővitjük az adatbázist a praeoperativ vizsgálatokról /haemodynamika, rtg, stb./, a mütéti eseményekről, a közvetlen és késői postoperativ időszakról. Az adatlapok ehhez már zömmel elkészültek.

Célunk a centrum és az országos kardiológiai hálózat közötti folyamatos információcsere elérése. Igy pontos képet nyerhetünk a késői mütéti eredményekről is és a mütétre nem kerültek sorsáról is.

Óriási, csak számitógép segitségével kezelhető adathalmazokat teszünk könnyen hozzáférhetővé a tudományos feldolgozáshoz, remélve, hogy segitséget nyujtunk a még jobb sebészi, gondozási és szervezési eredmények eléréséhez. Az eddigi tapasztalataink biztatóak és lelkesitően ösztökélőek a már müködő rendszer továbbfejlesztéséhez.

Hivatkozások a IV. részhez

[1] Report of Finland, Scientific group on monitoring of health effects of environmental agents, Geneve, 7-13, Decembre, 1976, pp. 1-12.

[2] Gyárfás Iván et al.: Az acut myocardialis infarctus napi periodicitásának vizsgálata,Cardiológia Hungarica, 4/3-4, 1975, pp 1-8.

[3] P. Kerékfy - I. Ratkó - M. Ruda: Microcomputer-based medical information systems, MEDINFO'83, Amszterdam, pp. 733.

[4] P. Kerékfy - I. Ratkó - M. Ruda: Patient registers on microcomputers, Cybernetics and Systems Research 2, R. Trappl ed. , Elsevier Science Publisher B.V. North-Holland , 1984, pp. 519-522.

[5] P. Kerékfy - M.Ruda : A System model for microcomputers in Health Care Applications, 3-rd International Conference on System Science in Health Care, München, 1984, pp. 1419-1422.

[6] H.M. Lewis et al.: Desktop database management in vascular surgery

[7] S.C. Lloyd: Programming systems for medical data base management,

[8] K. Neumann: Role and function of the analysis and interpretation system of the automated statistical information system, 4oth Session of the International Statistical Institute, Invited Paper, Warsaw, Sept. 1-9, 1975, pp. 7o / 1-1o

[9] A.L. Pai: A computerized cardiac hemodynamic data base system, Computers in cardiology, 198o, pp. 145-152.

[1o] Ratkó István - et al.: Acut miocardialis infarktussal vagy annak gyanujával ápolt betegek számitógépes nyilvántartása, vizsgálata, NJSzT SzKMAOB 7. kollokvium, Szeged, 1977, pp. 251-258.

[11] Ratkó István et al.: Adatfeldolgozó rendszer a cardiovascularis megbetegedések felkutatására és gondozására, MTA SzTAKI, Working Paper, 198o.

[12] I. Ratkó - M. Csukás: A data base management system for patients suffering from acute myocardial infarction, in: R. Trapple ed. Progress in Cybernetics and Systems Research, IX, Hemisphere, Washigton , 198o, pp. 497-5o1.

[13] I. Ratkó - M. Csukás: A statistical valuation of data of an infarction registers by computer, 3rd Hungarian Biometric Conference, Budapest, 1981, pp. 213-214.

[14] I. Ratkó - M. Csukás - P. Vaszary: Computer registration of patients waiting for cardiac operation, Cybernetics and Systems Research, North-Holland Publishing Company, 1982, pp. 651-653.

[15] Ratkó István et al.: Szivmütétre várakozók számitógépes nyilvántartása, NJSzT SzKMAOB 11. kollokcium, Szeged, 1982, pp. 464-468.

[16] I. Ratkó - M. Ruda: Cardiolόgical system for computer, in: Bemmel J.H. ed. Proceedings of MEDINFO 83 North-Holland, Amsterdam , pp. 1286.

[17] I. Ratkó - M. Ruda: Laboroj de la komputıl-statistika divizio de la Esplorinstituto pri la Komputtekniko kaj Automatizado de Hungara Science Akademio, krociĝantaj al kurac-biologiaj demandoj, Unua Simpozio pri Komputiko, Rennes, Julio, 1981, pp. 261-265.

[18] Ratkó István - Ruda Mihály: Szivmütétre várakozók számitógépes rendszere, Rendszerleirás, 1982.

[19] Ruda Mihály: A SIS77 statisztikai információs rendszer kialakitásának szempontjai, alkalmazásának és továbbfejlesztésének lehetőségei, MTA SZTAKI Tanulmányok, 86/1978.

[20] Ruda Mihály: Optimalizálási kérdések a statisztikai adatfeldolgozásban, MTA SZTAKI Tanulmányok, 110/1980, pp. 5-18.

[21] J. Ruszkowski - J. Milewski: Relational model for the evaluation of medical diagnostic tests based on expert opinions: differentiation of Jaundice etiology

[22] Vaszary Péter - Csukás Andrásné - Ratkó István: Szivsebészeti nyilvántartási és gondozást segitő rendszer, A Magyar Egészségügyi Szervezők Tudományos Egyesületének VI. kongresszusa, Budapest, 1983, pp. 267-272.

[23] И.Ратко-М.Руда: Несколько замечаний о работе отдела статистики Исследовательского Института Вычислительной Техники и Автоматизации ВАН, MTA SZTAKI Közlemények, 28/1982, pp. 59-62.

[24] И.Ратко: Учёт инфарктов, MTA SZTAKI Közlemények, 28/1982, pp. 45-58.

[25] И.Ратко: Учёт инфарктов с помощью вычислительной машины, монография РГ-I6 КНВВТ, в печати,I984,Москва

[26] М.Руда-Я.Деметрович: Микрокомпютерная система обработки данных с медицинскими применениями, монография РГ-I6 КНВВТ, в печати,I984,Москва.
(bővitett angol nyelvü változata:
P. Bakonyi, A. Békéssy, J. Demetrovics, P. Kerékfy, M. Ruda: A microcomputer-network based decision support system for health-care organizations, IFAC 9th World Congress, Vol XI,pp. 85-92.)

# BEFEJEZÉS

A disszertációban több különálló rendszert, s az azokkal kapcsolatos eredményeket ismertettem. A rendszereket határidős munkákként kellett elkésziteni. Igy azok megszületése és a megfelelő matematikai, számitástechnikai eredmények nem mindig párhuzamosan alakultak.

Véleményem szerint az ilyen határidős munkákra is gyakran elmondható ugyanaz, amit Knuth Előd mond az operációs rendszerek által felvetett problémákról. [1] : " A gyakorlat nem várakozhat a felvetett kérdések megoldására, a rendszereknek «tüzön-vizen» át el kell készülniök."

Ugyanakkor azt is meg kell jegyezni, hogy olyan rendszer ( pl. betegségregiszter), amely minden esetben alkalmazható, nem létezik. Arató Mátyás frappáns fogalmazásában: [2] "Az a követelmény, hogy mindenre alkalmas és hatékony operációs rendszert hozzunk létre, durván fogalmazva, ekvivalens azzal, hogy statisztikai kisérletnél az első és másodfaju hibát egyszerre tegyük zérussá, ennek lehetetlenségét minden statisztikus jól ismeri." (Itt is az operációs rendszer helyett betegségregisztert is mondhatunk, akkor is -véleményem szerint- igaz marad a kijelentés.)

Végezetül szeretnék köszönetet mondani

- az MTA SZTAKI Számitógéptudományi Főosztály több dolgozójának, többek között Dr Demetrovics János főosztályvezetőnek és Dr Békéssy András osztályvezetőnek azért a hat-

hatos bátoritásért, amellyel a disszertáció megszületését elősegitették; Dr Ruda Mihály tudományos munkatársnak és Dr Krámli András tudományos főmunkatársnak az évek óta tartó közös munkákért.

- Az ORFI több orvosának, többek között Dr Gömör Béla egyetemi tanárnak, osztályvezető főorvosnak a genetikai téma felvetéséért és hasznos tanácsaiért.
- Dr Csukás Andrásnénak, az Országos Kardiológiai Intézet tudományos munkatársának a betegségregiszterek kialakitásában adott segitségéért.
- És nem utolsó sorban Dr Arató Mátyásnak, a SZÁMALK igazgatói tanácsadójának, akinek köszönhetem, hogy bekapcsolódhattam ezekbe a kutatási feladatokba, s akitől kezdetben igen sok ösztönzést és segitséget kaptam.

## Hivatkozások

[1] Knuth Előd: Bevezető, Operációs Rendszerek Elmélete, Visegrádi Téli Iskola, MTA SZTAKI Közlemények, 15/1975, pp. 7-8.

[2] Arató Mátyás: Kerekasztal megbeszélés, Operációs Rendszerek Elmélete, Visegrádi Téli Iskola, MTA SZTAKI Közlemények, 15/1975, pp.147.

A TANULMÁNYOK SOROZATBAN 1983-BAN MEGJELENTEK:

140/1983 Operations Research Software Descriptions (Vol. 1.) Edited by A. Prekopa and G. Keri

141/1983 Ngo The Khanh: Prefix-mentes nyelvek és egyszerü determinisztikus gépek

142/1983 Pikler Gyula: Dialógussal vezérelt interaktiv gépészeti CAD rendszerek elméleti és gyakorlati megfogalmazása

143/1983 Márkusz Zsuzsanna: Modellelméleti és univerzális algebrai eszközök a természetes és formális nyelvek szemantikaelméletében

144/1983 Publikációk '81 Szerkesztette: Petróczy Judit

145/1983 Telcs András: Belső állapotu bolyongások

146/1983 Varga Gyula: Numerical Methods for Computation of the Generalized Inverse of Rectangular Matrices

147/1983 Proceedings of the joint Bulgarian-Hungarian workshop on "Mathematical Cybernetics and Data Processing" /Szerkesztette: Uhrin Béla/

148/1983 Sebestyén Béla: Fejezetek a részecskefizikai elektronikus kisérleteinek adatgyüjrő, - feldolgozó rendszerei köréből

149/1983 L. Keviczky - J. Hethéssy: A general approach for deterministic adaptive regulators based on explicit identification

150/1983 IFIP TC.2 WORKING CONFERENCE "System Description Methodologies" May 22-27. 1983. Kecskemét, /Szerkesztette: Knuth Előd/

151/1983 Márkusz Zsuzsanna: On First Order Many-Sorted LOGIC

152/1983 Operations Research Software Descriptions (Vol. 2.)
Edited by A. Prekopa and G. Keri

153/1983 T.M.R. Ellis: The automatic generation of user-adaptable application-oriented language processors based on quasi-parallel modules

154/1983 Publikációk '82 /Szerkesztette: Petróczy Judit/

A TANULMÁNYOK SOROZATBAN 1984-BEN MEGJELENTEK: