

Eltiraĵo el
Acta Sanmarinensia 2.5/1992
ISBN 83-85033-07-1

Prezentata en TTT sub la URL:
<http://www.forst.uni-muenchen.de/publ/quednau/softo.html>

De tiu-ĉi publikaĵo estas haveblaj tradukaĵoj en la lingvoj :
germana

Evoluigo de portebla programar-sistemo por statistikaj analizoj

de H. D. Quednau, München (D)

(Prelego prezentita dum SUS 8 en Białystok (PL), septembro 1990)

Resumo

Estas prezentita libere akirebla program-sistemo por efektiviĝi statistikajn analizojn. La ĝis nun verkitaj modjuloj* konsistas el uzanto-afabla daten-interfaco kaj el programoj por aŭtomata deriveado*, por la analizo de Ĝeneralaj Linearaj Modeloj kaj por la optimumigo de maksimum-verŝajneco-funkcioj*. Ĉiuj programoj estas portebloj de MS/DOS-PK ĝis komputilego; se necese la uzanto povas kombini ilin kun propraj rutinoj kaj/aŭ laŭplaĉe modifi.

Zusammenfassung

Es wird ein frei zugängliches Programmsystem zur Durchführung statistischer Analysen vorgestellt. Die bisher entwickelten Module bestehen aus einer benutzerfreundlichen Datenschnittstelle und aus Programmen zum automatischen Differenzieren, zur Auswertung von Allgemeinen Linearen Modellen und zur Optimierung von Maximum-Likelihood-Funktionen. Alle Programme sind portabel vom MS/DOS-PC bis zum Großrechner; sie können bei Bedarf vom Benutzer mit eigenen Routinen kombiniert und/oder nach Belieben verändert werden.

La plej granda parto de la sciencistoj, almenaŭ en la okcidento, uzas por statistikaj analizoj unu el la konataj kaj ŝatataj komercaj program-pakaĵoj, kiel ekzemple SPSS, BMDP, SAS, GLIM kaj multajn aliajn. Oni povas diri, ke iu nove elpensita algoritmo estas fakte uzata maljam tiam, kiam ĝi enirintas en unu el tiuj programar-sistemoj. La ekskluziva uzado de tiuj pretaj program-pakaĵoj evidente havas avantaĝojn: Oni akiras rezultojn multe pli rapide ol se oni devus unue serĉi taŭgajn algoritmojn kaj poste skribi kaj testi la koncernan komputil-programon. Kutime la uzanto ricevas krom la dezirataj rezultoj multajn pliajn, inter kiuj li povas libere elekti tiujn, kiuj plej plaĉas al li. Kompreneble tia proced-maniero tute kontraŭas la plej bazajn principojn de la statistiko, tamen estas oftege farata.

Tiun afablecon de la program-pakaĵo la uzanto dankas per tia fideleco, kia en nia tempo aliloke apenaŭ estas trovebla: Sciencisto eble kelkfoje estas malfidela al sia edz(in)o, sed **neniam** al sia statistika programar-sistemo. Tiu kortuŝa fideleco bedaŭrinde kunportas kelkajn gravajn malavantaĝojn por la scienca esplorado, el kiuj mi menciuj jenajn:

I.: La uzanto limigas sin al tiuj metodoj, kiuj pli malpli hazarde estas enigitaj en lian sistemon. Laŭ mia sperto, ordinara uzanto neniel estas persvadebla apliki metodon, kiun lia pakaĵo ne enhavas, eĉ se ĝi estas multe pli adekvata al lia problemo ol la tradicie uzata.

II.: Dua malavantaĝo estas, ke la aĉeteblaj sistemoj ne ebligas eĉ la plej etajn modifojn. Al mi ekzemple okazis, ke funkcion*-optimumiga programo kaptiĝis en sel-punkto, aŭ ke la nombro de variabloj estis limigita al valoro tro malgranda por mia tasko. Se mi disponintus pri la fonto-teksto de la koncerna programo, mi ja facile povintus solvi la problemon. Sed la fonto-tekston la programarfirmaoj ja ne disdonas, pro la timo de ŝtel-kopiado.

III.: Pro la sama kaŭzo provizoraj rezultoj, kiujn la programo ne eldonas, perdiĝas sen ke la uzanto havas la eblecon akiri ilin por propraj kalkuloj.

IV.: Ofte la programar-pakaĵo efektivigas nur en difinitaj komputilsistemoj, ekz. nur en MS/DOS-komputiloj (la tiel-nomataj IBM-kongruaj) aŭ nur en kelkaj grand-komputiloj. Do, se oni uzontas alian komputilon aŭ eble nur alian mastruman sistemon, oni samtempe devas ŝanĝi la statistikan metodaron.

V.: La prezoj de la programarsistemoj kelkfoje estas konsiderindaj. En lando, kiu havas valutajn problemojn, ofte estas tre malfacile aĉeti la programojn, kiuj uzatas en la okcidentaj industri-landoj. Tio ege malhelpas la liberan internacian interŝanĝon de sciencaj informoj.

Malgraŭ tiuj malavantaĝoj mi tute ne neas la meritojn de la komercaj statistikaj program-pakaĵoj. Ili estas multe helpintaj enkonduki statistikajn analizmetodojn en multajn scienco-branĉojn. Tamen mi kredas, ke oni facilanime fordonas multajn eblecojn, se oni kvazaŭ katenas sin al ili, kiel fakte plej ofte okazas. Estas dezirinde, ke krom la konataj komercaj sistemoj ekzistu ankaŭ aliaj, kiuj laŭeble evitu la priparolitajn malavantaĝojn. Tia sistemo

1 - devas disponi pri interfacio al uzanto, kiu estu same komforta kiel ĉe la aĉeteblaj sistemoj

2 - estu plene manipulebla. Per tiu mi celas, ke ĝi ne nur permesu sed eĉ apogu ŝanĝojn en ĉiu nivelo. Tio inkluzivas, ke oni havas alireblon al ĉiuj provizoraj rezultoj.

3 - estu portebla. Tio signifas, ke post kompilado de la sama fonto-teksto ĝi estu efektivigebla sur ĉiu-specaj komputiloj, de simpla MS/DOS-aparato ĝis la plej alt-nivelaj grand- aŭ super-komputiloj.

4- estu senpage disdonata almenaŭ por sciencaj aplikoj.

Pluajn postulojn al malfermaj, "re-uzebaj" programoj resumas FÖSSMEIER (1990) (ĉapitro 6 : Programado kiel re-uzo).

Tian sistemon mi estas evoluiganta kune kun studentoj de la TU München. La unuajn plano-ŝtupojn mi estas prezentinta en QUEDNAU (1990). Nun estas pretaj kelkaj modjuloj*, kiujn mi ŝatus prezenti ĉi-tie. Por la programado ni uzis la programlingvon "C", kiu montriĝis la plej taŭga por nia celo. Por certigi la porteblecon, ni strikte limigis nin al la ĝenerala normo, sen iam uzi specialaĵojn de la mastruma sistemo. Pro la sama kaŭzo ni ne uzis grafikan interfacon, sed la komunikado inter la uzanto kaj niaj programoj efektiviĝas per entajpitaĵoj komandoj aŭ per komando-dosieroj en tute kutima ASCII-formato.

La modjuloj ĝis nun ellaboritaj estas montrataj ĉi-tie :

XDM : Ĝenerala daten-interfaco (Kick 1989)
 LINMOD : Ĝenerala Lineara Modelo (uzas modjulon XDM)
 AUTDIFF : Aŭtomata deriveado* (Bareuther 1990)
 OPTIM : Optimumigo (Staudinger 1990)
 MVSJ : Maksimum-verŝajneco (uzas modjulojn XDM,
 AUTDIFF kaj OPTIM)

La unua estas la ĝenerala interfaco inter la dosiero, kiu enhavas la observitajn datenojn, alivorte la daten-dosiero, kaj la analiz-programoj. Kutime la datenaro estas matrico, kies horizontaloj reprezentas esplorunuojn, ekz. arbojn, kaj la vertikaloj enhavas ties atributojn (komparu QUEDNAU 1989).

	specio	alto	diametro	sanstato	...
arbo 1					
arbo 2					
...					

La unua paŝo por statistika analizo de tia datenaro estas elekti kelkajn esplorunuojn (ekz. la fagojn), eltiri kelkajn atributojn de la elektitaj esplorunuoj, eble efektivi transformojn (kelkfoje en sufiĉe komplika maniero), kaj determini, kiuj analizoj fakte estas farendaj. Tiu tasko estas plenumata de la modjulo XDM (DM = datenmanipulado). La uzanto skribas siajn dezirojn en komando-dosieron, kiu konsistas el pluraj paragrafoj. Ĉiu paragrafo entenas unu aŭ plurajn asigno-preskribojn. Jen eta ekzemplo:

fonto

```
fontdosiero = "arbaro.dat" ;
atributoj = 4 ;
```

variabloj

```
nomoj = Specio, Alto, Diametro, Sanstato, LDiametro, LAlto ;
nombrokategorioj de Sanstato = 0,1,2,3,4 ; \ generigas fantoman vektoron
```

transformoj

```
se Specio != 3 tiam forjhetu ; finose ;
se Diametro == 0 tiam forjhetu; alie LDiametro = log(Diametro) ; finose ;
LAlto = log(Alto) ;
```

problemo

```
dependaj = LAlto ;
modelo = 1,LDiámetro,Sanstato,LDiámetro*Sanstato ;
Testo = "LDiámetro*Sanstato" ;           \ testo pri interago
Testo = Sanstato,"LDiámetro*Sanstato" ;   \ testo pri sanstato
```

fino

La unua paragrafo, nomita “fonto”, deskriptas* la daten-dosieron, nome ties nomon kaj la nombron de atributoj. La sekventa paragrafo listigas la variablojn de la analizo. Tiuj estas unue la atributoj en la datenaro, kaj due pluaj grandoj, kiujn oni akiras per transformado de ili (vidu la sekvan paragrafon). La aserto “nombrokategorioj de ... = ...” igas la programon transformi la koncernan variablon en tiel-nomatan **fantoman vektoron** (angle dummy vector), kies dimensio estas $d - 1$, kie d estas la nombro de la kategorioj. Ĝiaj komponantoj estas formataj jene:

$$\begin{aligned} \text{Se } var < d &\Rightarrow fan.vec(var) = 1 \text{ kaj } fan.vec(j \neq var) = 0 \\ \text{Se } var = d &\Rightarrow fan.vec(j = 1, \dots, d - 1) = -1. \end{aligned}$$

Ĉe nia ekzemplo d egalas al 5, sekve la nombro-kategorioj estas transformataj jene:

$$\begin{array}{l} 0 \rightarrow 1 \quad 0 \quad 0 \quad 0 \\ 1 \rightarrow 0 \quad 1 \quad 0 \quad 0 \\ 2 \rightarrow 0 \quad 0 \quad 1 \quad 0 \\ 3 \rightarrow 0 \quad 0 \quad 0 \quad 1 \\ 4 \rightarrow -1 \quad -1 \quad -1 \quad -1 \end{array}$$

Fantomaj vektoroj estas bezonataj en multaj statistikaj analizoj, ekz. kun linearaj kaj ĝeneraligitaj linearaj modeloj. Bedaŭrinde mi ne povas pli funde detaliĝi tion ĉi-tie.

En la paragrafo “transformoj” oni ne nur deskriptas simplajn transformojn, kiel $LAlto = \log(Alto)$, sed oni povas ankaŭ ekskludi difinitajn esplorunuojn de la analizo (tion signifas la vorto “forĵetu”), kaj oni povas enigi la preskribojn pri forĵetado aŭ transformado en kondiĉo-strukturon, kiu havas la formon

“se..tiam..finose” aŭ “se..tiam..alie..finose”.

Nun sekvas la plej grava paragrafo, nomita “problemo”. Ĝi preskribas, kion la modjulo fakte eligu. Ĉe tiu-ĉi ekzemplo la modjulo transformas ĉiun ne-forĵetitan atributaron en du vektorojn, el kiu unu, la dependo-vektoro, entenas nur la logaritmon de la alto, dum la alia, la modelo-vektoro, estas multe pli komplika. Jen eta ekzemplo: Se iu esploruono havas la atributojn :

specio = 3 ; alto = 27 ; diametro = 16 ; sanstato = 2 ;

tiam ni ricevas :

$$\begin{array}{l} 3 \ 27 \ 16 \ 2 \quad \rightarrow \text{dep.vec : } 3.30 \\ \quad \quad \quad \rightarrow \text{mod.vec : } 1 \quad 2.77 \quad 0 \ 0 \ 1 \ 0 \quad 0 \ 0 \ 2.77 \ 0 \end{array}$$

La koncerna dependo-vektoro enhavas la logaritmon de $27 = 3.30$, kaj la modelo-vektoro enhavas unue la konstanton 1, tiam la logaritmon de la diametro kaj tiam la

fantoman vektoron, kiu respondas al 2. Post la sanstato estas indikita en la modelo-deklaro la **interago** inter $\log(\text{Diametro})$ kaj sanstato, kies fantoman vektoron oni akiras simple per multipliko de la koncernaj ĉef-efikoj (do $\log(\text{Diametro})$ kaj sanstato). La dependo- kaj la modelo-vektoroj estas generataj por ĉiuj ne-forĵetitaj esplor-unuoj. Male, la du lastaj informoj (Testo = "LDiametro*Sanstato" ; Testo = Sanstato,"LDiametro*Sanstato") estas pludonataj sensanĝe. La malsupren-streko "\ " indikas la komencon de komento.

La modjulo XDM, kiun mi preparolis ĝis nun, estas nur la interfaco inter la uzanto kaj la analiza programo. Ĝi mem faras neniujn statistikajn analizojn, tamen ĝia tasko estas tre grava kaj multe pli malfacile programebla ol estas la statistikaj analizalgoritmoj.

La plej simpla kaj plej bone konata modelo por analizi tian datenaron estas la **Ĝenerala Lineara Modelo**, kiun pritraktas la modjulo LINMOD. Ĝi supozas, ke la celvariablo estas gaŭse distribuita kun konstanta varianco kaj kun ekspekto, kiu kalkuliĝas kiel la skalarproduko el tia vico de struktur-variabloj, kian konsistigas la modelo-vektoro, kaj samlonga vektoro de nekonataj parametroj:

$$\begin{aligned}
 E(y_i) &= b_0 * (x_0 = 1) + b_1 * x_1 + \dots + b_r * x_r \\
 \hat{y}_i &= b_0 * 1 + b_1 * 2.77 + b_2 * 0 + b_3 * 0 + b_4 * 1 + b_5 * 0 \\
 &\quad + b_6 * 0 + b_7 * 0 + b_8 * 2.77 + b_9 * 0
 \end{aligned}$$

La parametro-vektoron \underline{b} oni povas tute simple kalkuli per la metodo de kvadratminimumigo.

La unua el la du testhipotezoj asertas, ke ne ekzistas interago inter la influvariabloj LDiametro kaj Sanstato, do

$$H_0^1 : b_6 = b_7 = b_8 = b_9 = 0,$$

kaj la dua, ke la sanstato ne efikas, nek kiel ĉefefiko, nek kiel interago:

$$H_0^2 : b_2 = b_3 = b_4 = b_5 = b_6 = b_7 = b_8 = b_9 = 0.$$

Ankaŭ tiaj testoj estas fareblaj per kelkaj relative simplaj matric-kalkuladoj. Jen ebla testrezulto, donita de la modjulo LINMOD :

TESTREZULTOJ

lg	F	m.a.a.	efiko
4	1.941	0.10537	LDiametro*Sanstato
8	4.134	0.00021	Sanstato, LDiametro*Sanstato

187 3.67

RESTVARIANCO

lg = liberogrado

F = testvariablo, distribuita laŭ F-distribuado,

se la koncerna testhipotezo estas vera

m.a.a. : donas la voston probablon de la koncerna F-distribuado, kiun mi nomas "malakceptiga adedo*".

Se la malakceptiga adedo malplias ol 0.05, tiam la testrezulto estas signifika, kaj ni devas malakcepti la testhipotezon. Ĉe la supre montrita ekzemplo ni vidas, ke ne estas statistike certigebla la ekzisto de interagoj, sed ja estas certigita kun grandega signifiko, ke la sanstato havas korelativon al la alto.

Mi jam diris, ke en la praktiko la malfacilaĵoj ne konsistas en la programado de la analiza algoritmo; tio estas farebla dum mallonga tempo ankaŭ de ne spertega programisto. Kondiĉo por tio estas nur, ke la analizprogramo ĝuste ricevas de ĉiu esplorunu la dependo- kaj la modelo-vektorajn. Aliflanke, tre malfacile estas komuniki al la komputilo en homkonforma lingvaĵo, kiamaniere konstrui tiujn vektorajn el la datenaro. La modjulo XDM ĝuste tiun taskon plenumas.

Komprenoble ĉiu komerca program-pakaĵo disponas pri simila interfac-modjulo, alie ĝi ja preskaŭ ne povus labori. Tamen ekzistas gravaj diferencoj inter la priparolita XDM kaj la komercaj interfac-modjuloj:

Unue: Preskaŭ ĉiam la rezultantaj vektoroj estas direktitaj nur al la analizprogramoj, sen ke la uzanto havas la eblecon uzi ilin por enigo ekz. al propra programo. La konsekvenco de tio estas la jam menciita fakto, ke la uzanto ne povas aldoni pliajn metodojn al la sistemo. Kun XDM la situacio estas tute alia: Ĝiaj komponantoj povas libere esti ligitaj al propraj programoj, kaj oni povas uzi ilin ankaŭ sendepende, ekz. por sendi la rezultantajn vektorojn en dosieron.

Dua avantaĝo de XDM estas, ke la fonto-teksto estas senkoste havebla, tiel ke oni povas laŭplaĉe ŝanĝi ĝin. Tia ŝanĝo estas aparte simpla, se oni disponas pri la mastruma sistemo UNIX. Mi donu etan, tamen sufiĉe gravan ekzemplon. En la dosieraro, kiu konsistigas la modjulon, ekzistas dosiero nomita "language.tab". Jen la komenco de la origina versio de ĉi-tiu dosiero:

P_ SOURCE	"quelle"
ATTRIBUTES	"attribute"
INFILE	"eingabe"
P_ VARIABLE	"variable"
NAMES	"namen"
OF	"von"
...	...

Ĝi enhavas la leksikonon de la komando-lingvo, kiu origine konsistis el germanaj vortoj. Alia dosiero, la tiel-nomata "makefile", enhavas la UNIX-komandojn por generi el tiu language.tab kaj aliaj pra-programeroj la pretajn "C"-programojn. Por havi internacilingvan version de XDM, mi tutsimple verkis alian "language.tab", anstataŭigante ties dekstran parton per vortoj en llo:

P_ SOURCE	"fonto"
ATTRIBUTES	"atributoj"
INFILE	"fontdosiero"
P_ VARIABLE	"variabloj"
NAMES	"nomoj"
OF	"de"
...	...

Same simple oni povus generi komando-lingvon kun polaj aŭ litovaj aŭ latinlitere skribitaj rusaj vortoj. Sed ne nur la vortaron, sed ankaŭ la gramatikon de la lingvo oni povas ŝanĝi, kvankam ne tiom simple. Bedaŭrinde mi ne havas la spacon por detaligi tion. Por tiuj, kiuj iom konas UNIX 'on, mi nur mallonge menciui, ke la skana kaj la persa programoj estas verkitaj helpe de la UNIX-iloj "lex" respektive "yacc" (bonan enkondukon en ilian uzadon donas Schreiner kaj Friedman 1985). Por modifi la strukturon de la komando-lingvo de XDM, oni povas ŝanĝi la "lex"- kaj "yacc"-specifajn enigajn dosierojn.

Post kiam mi estas priparolinta la bazajn ideojn pri la verkado de portebla statistika programo iom pli detale per la ekzemplo de XDM, mi nun povas pritrakti la aliajn modjulojn pli mallonge. La modjulo LINMOD mi jam menciis kiel la plej simplan analizprogramon uzantan la rezultojn de XDM. Sekvas nun la modjulo AUTDIFF. Ĝi efektivas aŭtomatan deriveadon. La deriveendan funkcionon oni deskriptas denove en komando-dosiero, kiu ĉi-foje povas aspekti tiel (komentoj troviĝas inter "{..}"-formaj krampoj):

```

FIKSA      h , n;
VAR        a,b,c;
FUNKCIONOJ z,f,g ;
PROBLEMO   z = (h - a) / b ;
           f = log(c) - log(b) + (c-1)*log(z) - z ^ c ;
           g = - n * f ; { Temas ĉi-tie pri la Weibull-funkciono }

FINO

```

La deriveenda funkcio nomiĝas "g". Ĝi dependas de 3 variabloj "a,b,c", laŭ kiuj ĝi estas deriveenda, kaj du konstantoj "h,n". Ĝi estas difinita jene: $g = - n * f$, kie f estas mem funkcio, dependanta krom de variabloj b kaj c ankaŭ de la funkcio z , kiu siaparte dependas de la variabloj a kaj b kaj la konstanto h . La tuta estas la tiel-nomata Weibull-funkciono, kiu rolas ekz. en la analizo de la distribuaĵo de arbo-altecoj. Oni vidas, ke la logaritmo estas senpere uzebila; la samo validas kompreneble por aliaj ofte uzataj funkcioj kiel sinuso, kosinuso ktp. Krome estas tre simple enkonduki novajn funkciojn en la sistemon; oni bezonas nur verki etan programon por kalkuli la valoron kaj la deriveaĵojn de la novenkondukenda funkcio. Por demonstri, kiel funkcias la modjulo, mi verkis etan ĉefprogramon, kiu faras nenion alian ol demandi la uzanton pri la valoroj de konstantoj kaj variabloj kaj poste elskribi la rezultojn. Ĉe nia ekzemplo de la Weibull-funkciono la interago inter programo kaj uzanto povus efektiviĝi jene:

```

Komputilo: Bv. enigi valorojn por 2 "FIKSA"-variabloj
Uzanto:    20 5
Komputilo: Bv. enigi valorojn por 3 "VAR"-variabloj
Uzanto:    -0.5 1.5 2
Komputilo: Rezulto :
           f = 919.375680
           Gradiento =
                 -90.867209 -1238.518519 2426.507083
           Hesse-matrico =
                 4.456342
                 121.481481 2485.925926
                 -283.563544 -3875.368435 6387.195895

```

Kutime la modjulo AUTDIFF estas uzata ne aparte, sed kiel help-modjulo de alia sistemero. Ekzemplo por tio estas la lasta preparolenda modjulo, nome MVSI. Ĝia tasko estas optimumigo de, mi diru angle "maximum likelihood"- funkcioj. Por la internacie vaste konata termino "maximum likelihood" ni povus diri maksimuma verŝajneco aŭ versimilo. Tiu maksimum-verŝajneco-metodo estas principe tre fleksebla kaj vaste aplikebla, tamen la fakta apliko limiĝas per tio, ke ne ekzistis afabla interfacio al uzanto por deskripti la specifajn modelojn. Mi esperas, ke la modjulo MVSI en kunlaboro kun la modjuloj XDM, AUTDIFF kaj OPTIM tian interfacon nun disponigas. Ni rigardu tre tipan ekzemplon: Ni havas denove por ĉiu observaĵo dependo-vektoron \underline{y} kaj model-vektoron \underline{x} . Tiuj 2 vektoroj estas eltiritaj el la observo-datenaro per la modjulo XDM. Denove ni kalkulas la skalarprodukon de la model-vektoro \underline{x} kaj la parametro-vektoro \underline{b} kaj nomas ĝin z .

$$\sum_{j=0}^r b_j * x_j = z$$

La negativa logaritmo de la verŝajneco de \underline{y} kalkuliĝas pere de la negativa logaritmita verŝajneco-funkcio nlv , kiu dependas, krom de \underline{y} , de la skalarproduko z kaj eble aliaj nelinearaj parametroj $a_1, a_2, ..$

$$-\log[V(\underline{y})] = nlv(\underline{y}, z, a_1, a_2, ..)$$

La stimaĵojn por la vektoro \underline{b} kaj la nelinearaj parametroj $a_1, a_2, ..$ oni akiras per la minimumigo de la funkcio nlv . La minimumigajn algoritmojn enhavas la modjulo OPTIM, kiun helpas unue la modjulo XDM, eltirante la modelo-vektoron el la observo-datenaro, kaj due la modjulo AUTDIFF, difinante la minimumigendan funkcionon nlv kaj ties derivaĵojn, kiujn OPTIM bezonas por trovi la minimumon. Sekve la komandodosiero por MVSI konsistas el du partoj, unu por XDM kaj unu por AUTDIFF, kiel estas montrite ĉi-tie :

fonto

atributoj=3; fontdosiero="unkel2.dat"; observoj=20;

variabloj

nomoj = DOZO , \ [mg/l]
n , \ Nombro de observitaj skaraboj
r ; \ Nombro de mortigitaj skaraboj

transformoj

se DOZO > 0 tiam DOZO = log(DOZO) ;
alie DOZO = -1000000 ; finose ;

problemo

modelo = 1, DOZO ; dependaj = r , n ;
INIT = 0.346, 2.76, 0.163 ; \ Komencaj stimajhoj akiritaj kiel deskriptite sube

fino

FIKSA r,n ;
VAR z , { z = b0 + b1 * log(dozo) }
s ; { s = spontana kvoto }
FUNKCIONOJ p0, p, f;
PROBLEMO
p0 = 1 / (exp(-z)+1) ;
p = s + (1-s)*p0 ;
nlv = -r*log(p) - (n-r)*log(1-p);

FINO

Temas ĉi-tie pri eksperimento per insekticidoj, deskriptita de UNKELBACH KAJ WOLF (1985). La modelo-vektoro konsistas, krom el la konstanto, nur el la logaritmo de la dozo de la veneno. La dependo-vektoro ĉi-kaze konsistas el la nombro r de la mortigitaj kaj la nombro n de la entute traktitaj skaraboj. El la model-vektoro kalkuliĝas la skalarproduko z, kiu estas pludonata kiel la unua VAR-variablo. Krome ekzistas plua, nelineara parametro, kiu indikas, kiom da skaraboj mortintus eĉ sen la traktado per insekticido. La celfunkcio nlv estas la negativa logaritmo de la dunomiala distribuado, kies parametro p estas la sumo de la spontana kvoto s kaj ono p0 de la ne spontane mortintaj. La parametro p0 siaparte dependas de z pere de la logistika funkcio. Komencaj stimajhoj por b estas akireblaj per tio, ke oni unue laboras per komando-dosiero, en kiu mankas la parametro s; komenca stimajho por s estas la empiria kvoto de mortaj skaraboj je dozo 0. Post kelkaj iteracioj ni akiras la stimajhojn de la parametroj b_0, b_1 kaj s, kiel estas montrite jene:

REZULTO : (rc=0,it=7)

FUNKCIONA VALORO:

220.499552

PARAMETROJ:

0.471042 5.087078 0.181942

GRADIENTO:

0.000000 0.000000 0.000000

INVERSO DE HESSE-MATRICO:

0.205369	-0.399767	-0.014423
-0.399767	1.037412	0.030751
-0.014423	0.030751	0.001719

Helpe de la inverso de la Hesse-matrico ĉe la celpunkto oni povas testi hipotezojn pri la parametro-vektoro.

Mi pretas sendi la fonto-programojn aŭ/kaj ties MS/DOS-kompilaĵojn al interesitoj. Mi tre ĝojus, se la programaro estus utila al plejeble multaj gekolegoj el la internacia kaj precipe internacilingva sciencistaro.

Glosaro

adedo : stokasta variablo aŭ nombro kalkulita el samplo aŭ populacio, ekz. averaĝo, varianco, F-test-adedo (el la arabezo*, propono fare de mi)

deskripti : prezenti iun fakton tiamaniere, ke la adresito ĝin komprenu (Propono de Wells). Laŭ PIV tio estus “priskribi”, sed tiu vorto estas erariga.

derivei : kalkuli $d f(x)/d x$ (PIV Supl. 1987). Kontraŭe al PIV Supl., mi prenis kiel eliran vorton la verbon, por esti konforma al la sistemo de Kiselman (1988)

detektiva : kapabla kun granda probablo malakceptigi malverajn testhipotezojn (propono fare de mi)

-ez- : ne-oficiala sufikso por indiki lingvon

funkciono : 3-a signifo de “funkcio” en PIV (bildigo en aron da nombroj) (propono fare de mi)

krita nombro : nombro, je kies transpaŝo ŝanĝiĝas grava eco (PIV Supl. 1987)

(mal)pliedi : a (mal)pliedas ol b signifas, ke $a \leq b$ (resp. $a \geq b$) (propono de C.O.Kiselman)

malakceptiga adedo : Probablo por tio, ke oni akiras test-adedon*, kiu samgrade aŭ eĉ pli kontraŭas H_0 'on, samtempe favorante H_1 'on, kiel (ol) la akirita test-adedo (kondiĉe ke H_0 veras). Tiu adedo en la etnolingva literaturo kutime, sed malprecize nomatas “p-valor” (germane: “Überschreitungswahrscheinlichkeit”) (propono fare de mi).

modjulo : Sendependa parto de programara sistemo, kiu plenumas difinitan taskon (PIV Supl. 1987)

Literaturo

- Bareuther, A.** , 1990 : AUTDIFF - ein C-Modul zum Automatischen Differenzieren für nichtlineare Optimierungsprobleme.
Fortgeschrittenenpraktikum für Informatiker; TU München.
- Fößmeier, R.** , 1990 : Enkonduko al la problemoj de programaj-produktado.
Acta Sanmarinensia 1,L1.
- Kick, U.** , 1989 : XDM - eine Datenmanipulationssprache zur Vorbereitung statistischer Analysen. Fortgeschrittenenpraktikum für Informatiker; TU München.
- Kiselman, C.O.** , 1991 : Nomoj de matematikaĵoj en Esperanto.
Serta Gratulatoria in Honorem Juan Régulo Vol. IV, pp. 683 – 697.
Universidad de La Laguna
- Quednau, H.D.** , 1990 : Baza statistika metodaro. Acta Sanmarinensia 1,L3.
- Quednau, H.D.** , 1990 : Software-Werkzeuge zur Entwicklung und Auswertung unkonventioneller Modelle für Dosis-Wirkungsbeziehungen.
Proc. Conf. on Forest Statistics, Freiburg Sep. 13-15, 1989.
- Schreiner, A.T. kaj Friedman, G.** , 1985 : Compiler bauen mit UNIX.
Hanser-Verlag, Wien.
- Staudinger, M.E.J.** , 1990 : OPTIM - ein C-Modul zur flexiblen Optimierung nichtlinearer Funktionen. Fortgeschrittenenpraktikum für Informatiker; TU München.
- Unkelbach, H.D. kaj Wolf, T.** , 1985 : Qualitative Dosis-Wirkungs-Analysen.
Gustav Fischer-Verlag, Stuttgart.

Adreso de la aŭtoro :

OProf. H.D. Quednau dr.,
Forstwiss. Fakultät der LMU,
Hohenbachernstr.22, D-85354 Freising
email : quednau@lrz.uni-muenchen.de

Postrimarko Post la presado de tiu-ĉi publikaĵo estas verkitaj pluaj modjuloj, kiuj ekzemple faciligas la analizon, se la celvariablo mezuratas en ŝtupoj, kaj ebligas la analizon de mallongaj tempo-serioj (sinsekvaj mezuraĵoj ĉe la sama objekto).

Manlibretoj por la uzado de la opaj modjuloj (en la lingvoj internacia (Esperanto) kaj germana) estas parte jam finpreparitaj (kvankam ankoraŭ ne en latex-formato), parte en preparado.