

Tukey's Median Polish voor de rxk-tabel

Johan P. Velema *

samenvatting

De "median polish" is het via een iteratieve procedure splitsen van een matrix van waarnemingen in een additieve matrix van geschatte verwachtingen en een matrix van residuën waarvan de medianen per rij en per kolom nul zijn.

Twee alternatieve procedures worden voorgesteld. De drie methoden worden toegepast op een voorbeeld en de resultaten vergeleken. Extreme waarnemingen blijken op de drie procedures een verschillende invloed te hebben en daarom tot discrepantie tussen de uitkomsten te leiden. Wanneer het erom gaat afwijkingen van het aangepaste model te ontdekken verdient de "median polish" de voorkeur boven de twee andere procedures.

inleiding

In 1977 publiceerde Tukey zijn boek "Exploratory Data Analysis". Het bevat een verzameling technieken, waarmee zowel statistici als niet-statistici op eenvoudige wijze inzicht kunnen krijgen in voor hen relevante getallen van allerlei aard. De meest bekende technieken zijn de "stem and leaf plot" en de "box and whisker plot", later uitgebreid tot de "notched box plot" (McGill et al., 1978).

Hoewel Tukey alle methodes als hand-methodes presenteert, zijn intussen computer-programma's gepubliceerd in basic, fortran en APL (McNeil, 1977; Velleman & Hoaglin, 1981). Verschillende technieken zijn beschikbaar als subroutines in IMSL (Gentle & Gregory, 1980). Bij het instituut voor epidemiologie van de Erasmus Universiteit wordt op dit moment een statistisch programma pakket ontwikkeld waarin een aantal van de door Tukey voorgestelde technieken worden opgenomen. Aan het biomedisch centrum van de rijksuniversiteit Limburg zijn reeds diverse programma's gerealiseerd.

In dit artikel wordt nader ingegaan op een methode voor de analyse van een rxk-tabel die door Tukey de "median polish" wordt genoemd. Het resultaat van de "median polish" is enerzijds een beschrijving van het waarnemingsmateriaal in termen van additieve rij- en kolom-effecten en

*Thoraxcentrum en Instituut voor Biostatistica, Erasmus Universiteit, Postbus 1738, 3000 DR ROTTERDAM

anderzijds een residuen-matrix, die de niet verklaarde informatie bevat.

De methode wordt hier beschreven en toegepast op een voorbeeld. Enkele andere mogelijke benaderingen worden vermeld en op dezelfde gegevens toegepast. Vervolgens worden de verschillende oplossingen vergeleken.

de methode

Toepassen van de "median polish" op een $r \times k$ matrix van waarnemingen (O) is in feite het splitsen van die matrix in een $r \times k$ matrix van geschatte verwachte (ofwel aangepaste) celfrequenties (E) en een $r \times k$ residuen-matrix (R), zodat geldt:

- $O = E - R$
- de rijen resp. kolommen van E zijn uit elkaar te berekenen door optellen van een constante (additiviteit)
- van elke rij en elke kolom van R is de mediaan nul

stel O is de matrix:

1	2	3
8	6	1
4	20	7

dan wordt R als volgt gevonden. Eerst wordt per rij van O de mediaan bepaald en afgetrokken:

1	2	3		2	2	2		-1	0	1
8	6	1	-	6	6	6	=	2	0	-5
4	20	7		7	7	7		-3	13	0
	0								0'	

Vervolgens wordt per kolom van O' de mediaan bepaald en afgetrokken:

-1	0	1		-1	0	0		0	0	1
2	0	-5	-	-1	0	0	=	3	0	-5
-3	13	0		-1	0	0		-2	13	0
	0'								R	

De medianen per rij en per kolom van R zijn nu alle nul en daarmee is in dit voorbeeld de procedure reeds na 1 iteratie ten einde. Meestal zijn 3 of 4 iteraties nodig. Door aftrekken ($O-R$) vinden we E:

1	2	3		1	2	2		0	0	1
8	6	1	=	5	6	6	+	3	0	-5
4	20	7		6	7	7		-2	13	0
	0				E				R	

Optellen van de constante 1 bij de eerste kolom van E levert de tweede kolom van E; kolom 2 en kolom 3 verschillen een constante nul; de rijen 1 en 2 een constante 4; de rijen 2 en 3 een constante 1.

Door deze additiviteit (die een logisch gevolg is van de gebruikte procedure) kunnen we de elementen van E beschouwen als de som van drie termen nl. een kolom-effect, een rij-effect en een algemene waarde. Als algemene waarde kunnen we een willekeurig element van E kiezen. Deze aftrekken van de rij waarin zij staat levert de kolom-effecten. De algemene waarde aftrekken van de kolom waarin zij staat levert de rij-effecten. Een gebruikelijke keus is de "common median", d.w.z. de mediaan van alle elementen van E. In dit geval is de common median 6. De kolom-effecten zijn dan: 1,0,0 en de rij-effecten: -4,0,1. We hadden echter elk willekeurig element van E als algemene waarde kunnen kiezen en de bijbehorende rij- en kolom-effecten berekenen. Dit komt omdat de essentiële informatie van de rij- en kolom-effecten niet is gelegen in hun niveau maar in hun onderlinge afstand.

Opgemerkt moet worden, dat E en R niet uniek bepaald zijn. Immers, hadden we de procedure in omgekeerde richting doorlopen (eerst medianen per kolom en daarna per rij) dan was het resultaat geweest:

1	2	3		1	3	0		0	-1	3
8	6	1	=	4	6	3	+	4	0	-2
4	20	7		8	10	7		-4	10	0
	0				E				R	

De hieruit volgende rij- en kolom-effecten verschillen van de eerder gevonden waarden. Hoewel in praktijkgevallen deze verschillen niet groot zijn, is het mogelijk de procedure in beide richtingen te doorlopen en de uitkomsten te vergelijken.

De "median polish" veronderstelt geen ordening van rijen of kolommen. Immers, van de response-variabele wordt de mediaan bepaald en niet van de rij- of kolom-variabele. De response-variabele kan continu zijn (lengte, gewicht, temperatuur etc.) of discreet (frequenties, percentages).

Om na te gaan of een additief model de waarnemingen adequaat beschrijft introduceert Tukey de "comparison value". Deze waarde, hier verder aangeduid als index-waarde, wordt per cel berekend als:

rij-effect x kolom-effect/algemene waarde

De index-waarde is positief bij gelijk en negatief bij ongelijk teken van de rij- en kolom-effecten. Uitgaande van de "common median" als algemene waarde worden de residuen uitgezet tegen de index-waarden. De helling van de bij deze puntenwolk passende regressielijn wordt gebruikt als indicatie

voor een transformatie van de data:

Als de helling ongeveer 1 is, kan de fit van het additieve model verbeterd worden door aan het oorspronkelijke model:

$$e_{ij} = a + r_i + k_j \quad (1)$$

(met a = algemene waarde; r_i = rij-effect; k_j = kolom-effect) de indexwaarde toe te voegen:

$$e_{ij} = a + r_i + k_j + \frac{r_i k_j}{a}$$

hetgeen identiek is met:

$$e_{ij} = a \left(1 + \frac{r_i}{a}\right) \left(1 + \frac{k_j}{a}\right) \quad (2)$$

een multiplicatief model. Het is in te zien, dat wanneer grote niveauverschillen bestaan tussen rijen en kolommen, een multiplicatief model de voorkeur verdient boven een additief model. Als alle elementen van 0 positief zijn, kan bij een helling van ongeveer 1 een log-transformatie worden toegepast op de oorspronkelijke waarnemingen. Dit is niet identiek met het aanpassen van model (2) en leidt vaak tot een beter resultaat (Tukey, 1977 p.398).

alternatieven

Een voor de hand liggende variatie op de "median polish" is niet de medianen af te trekken maar het gemiddelde. Dit levert:

1	2	3		.55	5.55	-.11	0.45	-3.55	3.11
8	6	1	=	3.55	8.55	2.89	+ 4.45	-2.55	-1.89
4	20	7		8.88	13.88	8.22	-4.88	6.11	-1.22
O				E				R	

Niet de medianen maar de gemiddelden van rijen en kolommen van R zijn nu nul. E en R zijn nu uniek bepaald. Er geldt:

$$e_{ij} = \text{rijgemiddelde} + \text{kolommengemiddelde} - \text{algemeen gemiddelde}$$

In feite komen we hiermee op het terrein van de variantie-analyse.

Een andere mogelijkheid is de verwachte celfrequenties te schatten door:

$$\frac{n_{i.} \times n_{.j}}{N} \quad (3)$$

waarin $n_{i.} = \sum_j n_{ij}$, $n_{.j} = \sum_i n_{ij}$ en $N = \sum_{ij} n_{ij}$

In feite resulteert dit model als we in plaats van het gemiddelde of de mediaan de som per rij en daarna per kolom aftrekken, zodat geldt:

$$e_{ij} = n_{ij} + n_{i.} - N$$

en met $a = N$ en (1) volgt:

$$r_i = n_{i.} - N \text{ en } k_j = n_{.j} - N$$

Invullen in het multiplicatieve model (2) levert:

$$e_{ij} = N \times \frac{n_{i.}}{N} \times \frac{n_{.j}}{N}$$

wat weer gelijk is aan (3)

In ons voorbeeld:

1	2	3		1.50	3.23	1.27	-0.50	-1.23	1.73
8	6	1	=	3.75	8.08	3.17	4.25	+ -2.08	-2.17
4	20	7		7.75	16.69	6.56	-3.75	3.31	0.44
	0				E			R	

Opnieuw zijn de gemiddelden van rijen en kolommen van R nul. E is multiplicatief in plaats van additief d.w.z. de rijen en kolommen van E volgen uit elkaar door vermenigvuldiging in plaats van vermeerdering. De op deze wijze geschatte verwachtingen kunnen worden gebruikt voor het uitvoeren van de chi-kwadraat toets voor onafhankelijkheid. We zullen daarom deze methode aanduiden als het onafhankelijkheidsmodel.

voorbeeld

Gegevens worden beschouwd van 5045 patienten tussen 10 en 89 jaar van wie een ambulant 24-uurs ECG werd geregistreerd en geanalyseerd. Een ambulant 24-uurs ECG is een vereenvoudigd ECG (2 in plaats van 12 afleidingen), dat gedurende een etmaal wordt opgenomen op een cassetteband tijdens de normale dagelijkse activiteiten van de (poliklinische) patient (Pool et al., 1979). Dit diagnostisch onderzoek wordt gedaan op initiatief van een car-

dioloog of internist, die vermoedt dat bepaalde klachten van de patient (hartkloppingen, duizeligheid, wegrakingen) het gevolg zijn van hart-ritmestoornissen (aritmieën). Ook geeft bij patienten die coronaire hart-ziekte (CHD) of een myocardiinfarct hebben (gehad) een 24-uurs ECG inzicht in de stabiliteit van het hart. Tevens kan het effect van anti-aritmische medicatie of het functioneren van een pacemaker worden beoordeeld.

In Rotterdam worden door de stichting Cardiolab 24-uurs ECG's geanalyseerd ten dienste van specialisten die over heel Nederland zijn verspreid. De hier gepresenteerde gegevens zijn afkomstig van Cardiolab en betreffen alle patienten, van wie tussen 1 januari 1976 en 1 juli 1978 een ambulante 24-uurs ECG werd geregistreerd.

Tabel 1 bevat aantallen patienten per leeftijd en indicatie (d.w.z. de reden voor het maken van een 24-uurs ECG). 81 patienten van wie de indicatie ontbrak werden uitgesloten. Van patienten met meer dan 1 indicatie werd alleen de belangrijkste indicatie beschouwd. Ter nadere aanduiding van "belangrijkste" staan in tabel 1 de indicaties van boven naar beneden in volgorde van belangrijkheid. Deze volgorde werd bepaald op grond van medische overwegingen. De vraag bij tabel 1 is: Is er associatie tussen leeftijd en indicatie en zo ja, in welke richting.

Tabel 1

indeling naar leeftijd en indicatie van 5045 patienten van wie een ambulante 24-uurs ECG werd gemaakt tussen 1 januari 1976 en 1 juli 1978.

	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	totaal
overige	7	22	37	69	73	56	34	8	306
pacemaker	1	1	2	7	10	10	19	3	53
infarct	0	4	24	109	238	203	84	6	668
CHD	1	1	13	51	110	73	31	1	281
palpataties	40	184	274	350	399	280	128	18	1673
wegrakingen	17	30	42	76	144	173	154	51	687
duizelingen	3	23	31	41	71	80	70	27	346
therapie	3	4	2	20	33	32	13	0	107
aritmie	28	83	115	165	230	175	105	23	924
totaal	100	352	540	888	1308	1082	638	137	5045

berekeningen

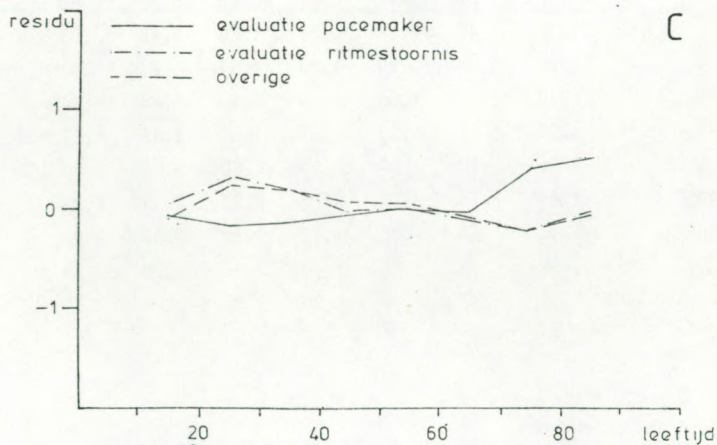
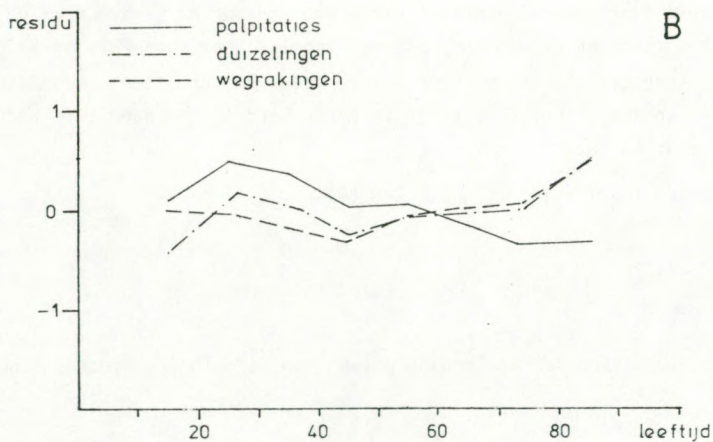
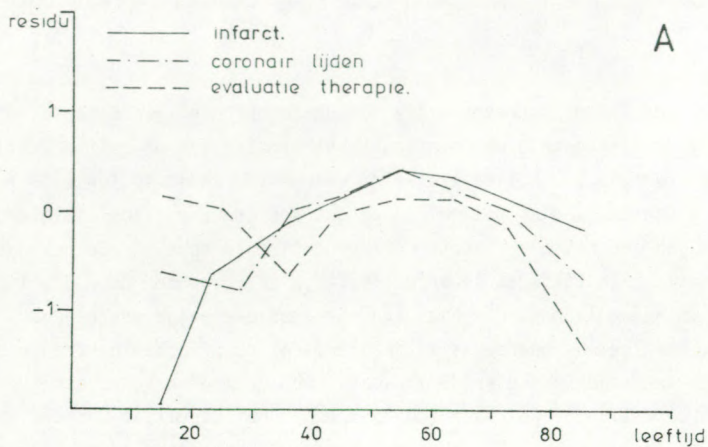
Na eerst op de bovenomschreven wijze met de oorspronkelijke data te hebben gewerkt, werd vastgesteld dat een log-transformatie van de celfrequenties de voorkeur verdiende. Nullen kregen de waarde .1, zodat $10 \log (.1) = -1$. Toepassing van de "median polish" op de getransformeerde tabel 1 leverde de matrix van geschatte verwachte celfrequenties in tabel 2. De residuen per indicatie zijn uitgezet tegen de leeftijd en verbonden door een lijn. Zo ontstaan negen lijnen, die hier in drie grafieken zijn weergegeven (fig. 1). De algemene waarde is zo gekozen, dat de effecten behorende bij de laagste leeftijdsklasse en de indicatie met de mediane geschatte verwachting (duizelingen) nul zijn (dus $a = e_{71}$). Het resultaat van de "median polish" was in dit voorbeeld identiek wanneer de procedure in omgekeerde richting werd doorlopen. De variantie-analyse benadering werd uiteraard eveneens toegepast op de getransformeerde tabel 1. Het onafhankelijkheidsmodel werd toegepast op de oorspronkelijke data. Om vervolgens vergelijking met de twee andere methoden mogelijk te maken werd de gevonden E-matrix ge-logarithmeerd. Er geldt dan:

$$e_{ij} = \log (n_{i.}) + \log (n_{.j}) - \log (N)$$

Tabel 2

E-matrix na toepassen van de "median polish" op de gelogarithmeerde tabel 1.

	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-80
overige	.88	1.13	1.40	1.82	1.88	1.88	1.78	.88
pacemaker	0	.25	.52	.94	1.00	1.00	.90	0
infarct	.94	1.19	1.46	1.88	1.94	1.94	1.84	.94
CHD	.59	.84	1.11	1.53	1.59	1.59	1.49	.59
palpataties	1.60	1.85	2.12	2.54	2.60	2.60	2.50	1.60
wegrakingen	1.23	1.48	1.75	2.17	2.23	2.23	2.13	1.23
duizelingen	.92	1.17	1.44	1.86	1.92	1.92	1.82	.92
therapie	.35	.60	.88	1.29	1.35	1.35	1.25	.35
aritmie	1.36	1.61	1.88	2.30	2.36	2.36	2.26	1.36



In tabel 3 en 4 zijn de rij- en kolom-effecten en algemene waarden vermeld, die met behulp van de drie genoemde technieken werden berekend. Hierbij werd ook voor de twee alternatieve methoden het element e_{71} als algemene waarde gekozen. De indicaties in tabel 3 zijn nu gerangschikt naar de grootte van de rij-effecten.

Tabel 3

Rij-effecten na toepassen op de gelogarithmeerde tabel 1 van de drie in het artikel beschreven methoden.

	mediaan	gemiddelde	onafh. model
palpataties	.68	.66	.69
aritmie	.44	.46	.43
wegrakingen	.31	.32	.30
infarct	.02	-.20	.29
duizelingen	0	0	0
overige	-.04	-.04	-.05
CHD	-.33	-.47	-.09
therapie	-.57	-.77	-.51
pacemaker	-.92	-.88	-.82

Tabel 4

Kolom-effecten en algemene waarden na toepassen op de gelogarithmeerde tabel 1 van de drie in dit artikel beschreven methoden.

leeftijd	mediaan	gemiddelde	onafh. model	gemiddelde*	onafh. model*
10-19	0	0	0	-.26	.09
20-29	.25	.50	.55	.24	.64
30-39	.52	.80	.73	.54	.82
40-49	.94	1.21	.95	.95	1.04
50-59	1.00	1.41	1.12	1.15	1.21
60-69	1.00	1.35	1.03	1.09	1.12
70-79	.90	1.16	.81	.90	.90
80-89	0	.20	.14	-.06	.23
alg. waarde	.92	.67	.84	.93	.75

* kolom-effecten na gelijktrekken algemene waarden

discussie

Door de waarnemingen te logarithmeren alvorens de "median polish" toe te passen ontstaat een model dat additief is op log-schaal. De zo gevonden rij- en kolom-effecten kunnen desgewenst worden teruggetransformeerd naar de oorspronkelijke schaal. Uiteraard zijn zij dan multiplicatief. De teruggetransformeerde residuen zijn dan niet (O-E) maar (O/E).

De additieve rij- en kolom-effecten geven een samenvatting van de waarnemingen. Daarbij wordt verondersteld dat rij-effecten constant zijn over alle kolommen en omgekeerd, dat kolom-effecten constant zijn over alle rijen. De algemene waarde is constant over zowel rijen als kolommen. De rij- en kolom-effecten geven een "gestileerde" werkelijkheid weer. Wanneer bepaalde combinaties van (in dit voorbeeld) leeftijd en indicatie veel meer of veel minder voorkomen dan door het additieve model wordt gespecificeerd, dan zal dit resulteren in zeer grote (positieve of negatieve) residuen. In zo'n geval spreken we van associatie. Afhankelijk van de vraagstelling zal de onderzoeker vooral in beschrijving geïnteresseerd zijn, i.e. in de rij- en kolom-effecten, dan wel primair belangstelling hebben voor de informatie, die niet in het model is vervat i.e. de residuen.

In figuur 1A suggereert de lijn behorende bij de indicatie infarct een parabolisch verloop met een maximum tussen 50 en 70 jaar. Dit wijst op een positieve associatie met leeftijden tussen 50 en 70 jaar en een negatieve associatie met lage leeftijden. De indicatie palpitations blijkt geassocieerd met lagere leeftijden (figuur 1B). "Evaluatie pacemaker" kwam op hoge leeftijden vaker voor dan het model aangeeft (figuur 2C).

De in tabel 3 vermelde rij-effecten blijken voor de meeste indicaties goed overeen te stemmen. Wanneer we het gebruik van mediaan en gemiddelde vergelijken, valt alleen voor "infarct", "CHD" en "evaluatie therapie" een duidelijke discrepantie te constateren. Dit zijn de indicaties uit figuur 1A en deze worden gekenmerkt door zeer groot negatieve residuen. Deze komen voor in cellen met zeer lage frequenties.

Nu hebben extreme waarnemingen op de voor de drie methoden kenmerkende maten een verschillende invloed:

- De mediaan is ongevoelig voor zeer hoge of zeer lage waarnemingen ongeacht eventuele transformaties.
- Het gemiddelde wordt door lage waarnemingen onlaaggetrokken en door hoge waarnemingen omhooggetrokken. Door logarithmeren hebben lage waarnemingen een sterkere invloed dan hoge waarnemingen.

- Het randtotaal wordt door lage waarnemingen omlaaggetrokken en door hoge waarnemingen omhooggetrokken. Omdat eerst het randtotaal wordt bepaald en daarna gelogarithmeerd hebben hoge en lage waarnemingen even grote invloed.

Omdat de rij- en kolom-effecten direct bepaald worden door hetzij de mediaan, het gemiddelde of het randtotaal per rij resp. kolom, is het begrijpelijk dat de rij-effecten voor de drie genoemde indicaties bij gebruik van het gemiddelde lager zijn dan bij gebruik van de mediaan. Het onafhankelijkheidsmodel resulteert in rij-effecten die hoger zijn dan die met de "median polish" worden gevonden. De hoge celfrequenties in de leeftijden 50-70 trekken de schatting van het rij-effect kennelijk sterker omhoog dan deze door de lage celfrequenties naar beneden wordt getrokken. Voor "evaluatie therapie" is het verschil tussen "median polish" en onafhankelijkheidsmodel gering omdat daar de "hoge" celfrequenties een orde van grootte lager zijn dan die voor de andere twee indicaties.

Bekijken we de kolom-effecten, dan zien we dat er een niveau-verschil is tussen de effecten die geschat worden m.b.v. het gemiddelde en de overige twee schattingen. De oorzaak hiervan is, dat de leeftijds-klasse 10-19 als nulpunt werd gekozen. Deze keuze is ongelukkig omdat op grond van bovenstaande overwegingen juist in deze kolom discrepantie tussen de drie schattingen te verwachten valt (veel lage celfrequenties). Een betere keus is de leeftijds-klasse 70-79 waarin geen extreme waarnemingen voorkomen (ratio hoogste en laagste waarneming = 12). Liever dan op nul fixeren we de schattingen in deze kolom op de waarde .90 welke voor de "median polish" werd gevonden.

Vergelijking van de gecorrigeerde kolom-effecten (tabel 4) laat zien dat tussen "median polish" en variantie-analyse model goede overeenstemming bestaat. Uitzonderingen zijn de eerste kolom (zoals verwacht) en de leeftijdsklassen 50-59 en 60-69, waar de zeer hoge celfrequenties het gemiddelde omhoogtrekken. Het onafhankelijkheidsmodel leidt in alle kolommen tot hogere schattingen dan de "median polish" en het variantie-analyse model. Oorzaak zijn vooral de in vrijwel elke kolom voorkomende zeer hoge frequenties voor de indicatie "palpitaties".

evaluatie

De "median polish" is een exploratieve techniek. Zij biedt de mogelijkheid een beter inzicht te krijgen, zonder dat direct tot toetsen wordt overgegaan. Exploratie wordt door velen die statistiek toepassen onvoldoende gewaardeerd. Het is in dit verband opmerkelijk, dat velen variantie-analyse gebruiken voor het uitvoeren van toetsen, maar dat de mogelijkheid m.b.v. variantie-analyse de data weer te geven als in dit artikel getoond, veel minder bekend is. Evenzo is de chi-kwadraat toets wijd verbreid, terwijl de mogelijkheid de verwachtingen-matrix op de bovenomschreven wijze te gebruiken zelden wordt genoemd. Daarom is het van belang de gebruiker van statistiek te wijzen op concrete mogelijkheden om zijn gegevens te exploreren. De "median polish" is zo'n mogelijkheid en heeft daarom praktische betekenis.

Als we de "median polish" vergelijken met de twee in dit artikel genoemde alternatieven zijn een aantal technische punten te noemen.

- Bij de "median polish" zijn de E- en R-matrix niet uniek bepaald. Dit introduceert een stuk onzekerheid. In de praktijk zullen de twee mogelijke uitkomsten echter niet veel verschillen. In het hier gepresenteerde voorbeeld was er zelfs geen enkel verschil.
- De "median polish" kan worden toegepast zonder gebruik te maken van rekenapparatuur. Het is duidelijk dat dit voordeel vervalst wanneer voor de "median polish" een computerprogramma wordt gebruikt.
- Bij de "median polish" zijn een aantal iteraties nodig; meestal niet meer dan drie of vier. In het hier gegeven voorbeeld werden (m.b.v. een computer) 10 iteraties uitgevoerd. Later bleek dat na 4 iteraties geen noemenswaardige veranderingen waren opgetreden.

Vergelijken we de schattingen voor rij- en kolom-effecten die de drie methoden leveren, dan zijn een aantal conclusies te trekken:

- De "median polish" is het minst gevoelig voor extreme waarnemingen.
- De schattingen van het variantie-analyse model zijn niet zeer verschillend van die welke de "median polish" oplevert. Verschillen ontstaan vooral daar waar zeer lage waarnemingen voorkomen, althans op log-schaal. Wanneer niet wordt gelogarithmeerd zijn grotere verschillen te verwachten omdat dan zeer hoge waarnemingen een grotere invloed doen gelden.
- Het onafhankelijkheidsmodel is meer dan de andere twee methoden gevoelig voor extreme waarnemingen.

De consequenties van deze conclusies voor de praktijk zijn afhankelijk van het doel van de analyse. Indien ons doel is, de waarnemingen zo goed mogelijk te schatten m.b.v. een statistisch model, d.w.z. de residuen te mini-

maliseren, dan zal het variantie-analyse model of het onafhankelijkheids-model de beste keus zijn. Immers, naarmate de schattingen meer in de richting van de extreme waarnemingen worden getrokken, zullen de corresponderende residuen kleiner zijn. In het hier gepresenteerde voorbeeld zijn we echter vooral geïnteresseerd in afwijkingen van het model, d.w.z. de residuen. In dat geval willen we extreme waarnemingen niet verdoezelen door schattingen die naar die extremen toetrekken en zal de "median polish" de voorkeur verdienen.

naschrift

Ik dank Prof. R. van Strik en drs. J.G.P. Tijssen voor hun waardevolle commentaren, de referee voor een nuttige aanvulling en H.E. Busselmann voor het vervaardigen van de figuur.

literatuur

- Gentle JE, Gregory WC: Recent enhancements to the IMSL subroutine library. The Am Statistician 34 (1980) 246.
- McGill R, Tukey JW, Larsen WA: Variations of box plots. The Am Statistician 32 (1978) 12-16.
- McNeil DR: Interactive data analysis; a practical primer. Wiley, 1977.
- Pool J, Tiggelaar H, Roodenburg A, Glaser B: Dynamische electrocardiografie; aritmiedetectie door ambulante 24-uurs ECG-registratie. Hart Bulletin, juni 1979.
- Tukey JW: Exploratory data analysis. Addison Wesley, 1977.
- Velleman PF, Hoaglin CD: Application, basics and computing of exploratory data analysis. Boston, 1981.