

De univariate voorspellingen lijken weinig realistisch. Van een economische stagnatie is hier geen sprake, gezien de voorspellingen van de consumptie, de investeringen, de werkgelegenheid en de afzet van bedrijven. Bovendien liggen in alle gevallen de geschatte standaarddeviaties van de univariate voorspellingen boven die van de GRECON-voorspellingen.

3. Conclusies.

Over het algemeen lijkt het GRECON-model een betere voorspelkracht te hebben dan het door ons geschatte univariate model, zowel waar het ex-post voorspellingen betreft als in geval van de onderzochte ex-ante voorspellingen. Dit blijkt uit de waarden van de ongelijkheidscoëfficiënten en uit de grafische confrontaties. Voor relatief veel variabelen hebben bovendien de ongelijkheidscoëfficiënten van de univariate voorspellingen een waarde groter dan 1.

De betere voorspelkwaliteit van het GRECON-model is vooral duidelijk als we te maken hebben met een structuurbreuk. Dit is zowel het geval bij de oliecrisis (1974/1975) als bij de huidige recessie (1980).

De GRECON-voorspellingen voor 1982 lijken vooralsnog zowel realistisch als betrouwbaarder dan de univariate voorspellingen.

Literatuur.

- G.E.P. Box en G.M. Jenkins: "Time Series Analysis, Forecasting and Control", revised edition. Holden-Day Inc., San Francisco, 1976.
- C.W.J. Granger en P. Newbold: "Forecasting Economic Time Series", Academic Press, New York, 1977.
- P. Newbold en C.W.J. Granger: "Experience with Forecasting Univariate Time Series and the Combination of Forecasts". Journal of the Royal Statistical Society, Vol. 137, Part 3, 1974.
- P. Newbold en G.V. Reed: "The Implications for Economic Forecasting of Time Series Model Building Methods". O.D. Anderson, ed., "Forecasting", North-Holland 1979.
- A.G.M. Steerneman, B. Bos, M.A. Kooyman, W. Voorhoeve, R.H. Ketelapper: "Onzekerheid van voorspellingen en realisaties van economische variabelen als gevolg van voortloepig cijfermateriaal; een analyse m.b.v. GRECON 79-9". Economisch Instituut Groningen, SE-68/8002, 1980.
- E.G.F. van Winkel: "ARIMA-processen: De betekenis van univariate methoden voor de econometrie". Proefschrift, V.U., Amsterdam 1979.

EEN ANALYSE VAN VERKIEZINGSUITSLAGEN MET BEHULP VAN CORRESPONDENTIEANALYSE

Peter van der Heijden *)

Samenvatting: Met behulp van correspondentieanalyse wordt een 3-weg matrix geanalyseerd. Het betreft hier verkiezingsuitslagen: een matrix van jaren (2) bij stembedistricten (58) bij partijen (10 en 20). Hierbij worden verschillende methoden gebruikt om de 3-weg matrix als 2-weg matrix te analyseren: d.m.v. het bekijken van elk jaar afzonderlijk, d.m.v. concateneren, en d.m.v. het analyseren van een zgn. 'bimarginale' matrix. De resultaten van de analyses worden gepresenteerd.

Trefwoorden: Correspondentieanalyse, categorische data, longitudinaal onderzoek, verkiezingen.

1. Inleiding en probleemstelling

Een verkiezingsuitslag kan weergegeven worden in een frequentietabel, waarbij de kolommen staan voor partijen, en de rijen voor bijvoorbeeld provincies, steden, districten, wijken.

De kolomtotalen leveren dan het aantal stemmen per partij op, de rijtotalen het aantal stemmen dat per district, stad etcetera is uitgebracht.

Indien men een bepaalde rij wil vergelijken met andere rijen, wordt dit gemakkelijker als de frequentietabel is omgezet in een tabel met percentages, waarbij de percentages van elke rij optellen tot 100. Hetzelfde kan natuurlijk, vice versa, gedaan worden voor de kolommen. Men vermenigvuldigt dan partijen met elkaar.

*) Peter van der Heijden is als student-assistent werkzaam bij de vakgroep Datatheorie FSW/RUL, Hooigracht 15, Leiden. Hij wil hierbij de andere leden van de vakgroep danken voor het waardevolle commentaar op eerdere versies.

bekend onder de naam 'reciprocal averaging' (Richardson (1933), genoemd in Horst (1935)) en 'le principe baricentrique' (Benzécri (1973)), en wordt ook gebruikt bij principale componenten analyse.

In het hierbovenstaande mogen rijen en kolommen verwisseld worden. Dat de rijpunten dan in het gewogen gemiddelde komen te liggen van de kolommen, is slechts een kwestie van normalisatie. In de analyses is gekozen voor een zgn. symmetrische aanpak: op elke dimensie liggen de rijen (kolommen) in het gewogen gemiddelde van de kolommen (rijen) gedeeld door de bijbehorende singuliere waarden. De symmetrie van de aanpak blijkt hieruit, dat, indien de matrix met frequenties gekanteld wordt, dezelfde oplossing wordt gevonden.

Het bovenstaande leidt tot het volgende: een rijpunt wordt dicht bij een kolompunt gelegd naarmate de frequentie in cel f_{ij} groter is dan op grond van de marginales van die cel verwachte waarde $e_{ij} = f_{i.} \cdot f_{.j} / n$. Is f_{ij} kleiner dan e_{ij} , dan worden de punten van rij i en kolom j uiteen geleid. Lijken patronen van twee rijen veel op elkaar, dan zullen deze punten in het plaatje dicht bij elkaar worden geplaatst. Hetzelfde geldt voor de kolommen. Is er sprake van een unieke relatie tussen een rij en een kolom, dan worden deze niet alleen dicht bij elkaar geplaatst, maar ook meer naar de periferie.

Ligt een rijpunt in het centrum, dan is f_{ij} voor de gehele rij ongeveer gelijk aan e_{ij} . Hetzelfde geldt uiteraard voor een kolompunt.

Een standaardwerk op het gebied van correspondentieanalyse is geschreven door Benzécri e.a. (1973). Door Hill (1974) is dit samengevat. Een recentere samenvatting van correspondentieanalyse en gerelateerde technieken is te vinden in Gifi (1981). Eveneens van de hand van Gifi (1982) is een User's Guide bij het programma ANACOR, waarin de methode en enkele voorbeelden worden beschreven.

In KM werd correspondentieanalyse kort beschreven door A.Z. Israëls e.a. (1981).

3. Leidse stembedistricten 1981 (TK) en 1982 (PS)

Er zijn twee frequentietabellen, weergegeven in tabel 1 (uitslag 1981) en tabel 2 (uitslag 1982). Beide zijn verkrijgbaar bij het Bureau Statistiek van het stadhuis van de gemeente Leiden.

In de tabellen is het aantal kolommen verschillend: in 1981 (TK) deden achttien politieke partijen mee, in 1982 slechts acht partijen. In beide

indien men uitslagen van verschillende verkiezingen heeft, kan men op soortgelijke wijze veranderingen in kolommen en rijen over die verkiezingen bekijken. Hoe groter de afmetingen van de frequentietabel echter, en hoe meer verkiezingen men tegelijk wenst te bekijken, des te moeilijker is het een globaal overzicht te krijgen van de verkiezingsuitslagen.

In dit artikel willen we laten zien hoe correspondentieanalyse, uitgevoerd middels het computerprogramma ANACOR dat ontwikkeld is door de vakgroep Data-theorie (FSW) te Leiden, een oplossing biedt voor dit probleem.

Beschreven wordt één voorbeeld: de uitslagen van de Tweede Kamerverkiezingen van 1981 en de Provinciale Statenverkiezingen van 1982 in de stembedistricten van de gemeente Leiden. De nadruk komt hierbij te liggen op de gehanteerde methode. Allereerst wordt beschreven wat correspondentieanalyse inhoudt.

2. Correspondentieanalyse

Correspondentieanalyse is de Nederlandse naam voor wat in het Frans wordt aangeduid met Analyse des Correspondances. Het betreft een techniek die de relatie tussen rijen en kolommen van een matrix zichtbaar maakt. De rijen en kolommen worden afgebeeld als punten in een p-dimensionele ruimte met behulp van zgn. χ^2 -afstanden. Dit gebeurt als volgt:

als f_{ij} de frequentie is van de cel behorende bij rij i en kolom j , en de rijmarginales worden weergegeven door $f_{i.}$, de kolommarginales door $f_{.j}$, en $n = f_{..}$ is de som van de frequenties van de matrix, dan is de χ^2 voor een matrix

$$\chi^2 = \sum_i \sum_j f_{ij} \cdot \frac{(f_{ij}/f_{i.} - f_{.j}/n)^2}{f_{.j}/n}$$

Deze χ^2 is te splitsen voor de rijen: $\chi^2 = \sum_i f_{i.} \chi_i^2$. χ_i^2 is de χ^2 -afstand tussen rij i en de kolommarginales. Dit brengt ons tot de χ^2 -afstand δ_{ik} tussen elk paar van de rijen i en k in de matrix:

$$\delta_{ik}^2 = \sum_j \frac{(f_{ij}/f_{i.} - f_{kj}/f_{k.})^2}{f_{.j}/n}$$

Hierbij geldt dat $\delta_{ik} \geq 0$, $\delta_{ii} = 0$ en $\delta_{ik} = \delta_{ki}$.

Als rijen i en k proportioneel zijn, dan is $\delta_{ik} = 0$.

De coördinaten van de rijpunten krijgen we door de singuliere waarden decompositie van de matrix van δ_{ik} 's te nemen. Vervolgens plaatsen we de kolompunten in het gewogen gemiddelde van de rijpunten. Dit principe staat

Tabel 1: Tweede Kamer 1981. Verkiezingsuitslagen per stembedistrict. Aantal op elke partij uitgebrachte stemmen.

STEM-DISTRICHT	PPDA	CDA	VVD	D.66	SP	GPV	ON-AFW	TOTAAL
1	179	544	229	28	18	1	1	145
2	242	149	7	66	103	103	103	151
3	159	103	9	20	5	1	1	157
4	147	154	108	117	154	154	154	191
5	184	108	8	95	11	1	1	187
6	184	108	8	95	11	1	1	187
7	184	108	8	95	11	1	1	187
8	184	108	8	95	11	1	1	187
9	184	108	8	95	11	1	1	187
10	184	108	8	95	11	1	1	187
11	184	108	8	95	11	1	1	187
12	184	108	8	95	11	1	1	187
13	184	108	8	95	11	1	1	187
14	184	108	8	95	11	1	1	187
15	184	108	8	95	11	1	1	187
16	184	108	8	95	11	1	1	187
17	184	108	8	95	11	1	1	187
18	184	108	8	95	11	1	1	187
19	184	108	8	95	11	1	1	187
20	184	108	8	95	11	1	1	187
21	184	108	8	95	11	1	1	187
22	184	108	8	95	11	1	1	187
23	184	108	8	95	11	1	1	187
24	184	108	8	95	11	1	1	187
25	184	108	8	95	11	1	1	187
26	184	108	8	95	11	1	1	187
27	184	108	8	95	11	1	1	187
28	184	108	8	95	11	1	1	187
29	184	108	8	95	11	1	1	187
30	184	108	8	95	11	1	1	187
31	184	108	8	95	11	1	1	187
32	184	108	8	95	11	1	1	187
33	184	108	8	95	11	1	1	187
34	184	108	8	95	11	1	1	187
35	184	108	8	95	11	1	1	187
36	184	108	8	95	11	1	1	187
37	184	108	8	95	11	1	1	187
38	184	108	8	95	11	1	1	187
39	184	108	8	95	11	1	1	187
40	184	108	8	95	11	1	1	187
41	184	108	8	95	11	1	1	187
42	184	108	8	95	11	1	1	187
43	184	108	8	95	11	1	1	187
44	184	108	8	95	11	1	1	187
45	184	108	8	95	11	1	1	187
46	184	108	8	95	11	1	1	187
47	184	108	8	95	11	1	1	187
48	184	108	8	95	11	1	1	187
49	184	108	8	95	11	1	1	187
50	184	108	8	95	11	1	1	187
51	184	108	8	95	11	1	1	187
52	184	108	8	95	11	1	1	187
53	184	108	8	95	11	1	1	187
54	184	108	8	95	11	1	1	187
55	184	108	8	95	11	1	1	187
56	184	108	8	95	11	1	1	187
57	184	108	8	95	11	1	1	187
58	184	108	8	95	11	1	1	187
59	184	108	8	95	11	1	1	187
60	184	108	8	95	11	1	1	187
61	184	108	8	95	11	1	1	187
62	184	108	8	95	11	1	1	187
63	184	108	8	95	11	1	1	187
64	184	108	8	95	11	1	1	187
65	184	108	8	95	11	1	1	187
66	184	108	8	95	11	1	1	187
67	184	108	8	95	11	1	1	187
68	184	108	8	95	11	1	1	187
69	184	108	8	95	11	1	1	187
70	184	108	8	95	11	1	1	187
71	184	108	8	95	11	1	1	187
72	184	108	8	95	11	1	1	187
73	184	108	8	95	11	1	1	187
74	184	108	8	95	11	1	1	187
75	184	108	8	95	11	1	1	187
76	184	108	8	95	11	1	1	187
77	184	108	8	95	11	1	1	187
78	184	108	8	95	11	1	1	187
79	184	108	8	95	11	1	1	187
80	184	108	8	95	11	1	1	187
81	184	108	8	95	11	1	1	187
82	184	108	8	95	11	1	1	187
83	184	108	8	95	11	1	1	187
84	184	108	8	95	11	1	1	187
85	184	108	8	95	11	1	1	187
86	184	108	8	95	11	1	1	187
87	184	108	8	95	11	1	1	187
88	184	108	8	95	11	1	1	187
89	184	108	8	95	11	1	1	187
90	184	108	8	95	11	1	1	187
91	184	108	8	95	11	1	1	187
92	184	108	8	95	11	1	1	187
93	184	108	8	95	11	1	1	187
94	184	108	8	95	11	1	1	187
95	184	108	8	95	11	1	1	187
96	184	108	8	95	11	1	1	187
97	184	108	8	95	11	1	1	187
98	184	108	8	95	11	1	1	187
99	184	108	8	95	11	1	1	187
100	184	108	8	95	11	1	1	187

Tabel 2: Provinciale Staten 1982. Aantal op elke partij uitgebrachte stemmen

STEM-DISTRICHT	PPDA	CDA	VVD	D.66	SP	GPV	ON-AFW	TOTAAL
1	193	131	107	127	7	1	1	143
2	198	131	107	127	7	1	1	143
3	198	131	107	127	7	1	1	143
4	198	131	107	127	7	1	1	143
5	198	131	107	127	7	1	1	143
6	198	131	107	127	7	1	1	143
7	198	131	107	127	7	1	1	143
8	198	131	107	127	7	1	1	143
9	198	131	107	127	7	1	1	143
10	198	131	107	127	7	1	1	143
11	198	131	107	127	7	1	1	143
12	198	131	107	127	7	1	1	143
13	198	131	107	127	7	1	1	143
14	198	131	107	127	7	1	1	143
15	198	131	107	127	7	1	1	143
16	198	131	107	127	7	1	1	143
17	198	131	107	127	7	1	1	143
18	198	131	107	127	7	1	1	143
19	198	131	107	127	7	1	1	143
20	198	131	107	127	7	1	1	143
21	198	131	107	127	7	1	1	143
22	198	131	107	127	7	1	1	143
23	198	131	107	127	7	1	1	143
24	198	131	107	127	7	1	1	143
25	198	131	107	127	7	1	1	143
26	198	131	107	127	7	1	1	143
27	198	131	107	127	7	1	1	143
28	198	131	107	127	7	1	1	143
29	198	131	107	127	7	1	1	143
30	198	131	107	127	7	1	1	143
31	198	131	107	127	7	1	1	143
32	198	131	107	127	7	1	1	143
33	198	131	107	127	7	1	1	143
34	198	131	107	127	7	1	1	143
35	198	131	107	127	7	1	1	143
36	198	131	107	127	7	1	1	143
37	198	131	107	127	7	1	1	143
38	198	131	107	127	7	1	1	143
39	198	131	107	127	7	1	1	143
40	198	131	107	127	7	1	1	143
41	198	131	107	127	7	1	1	143
42	198	131	107	127	7	1	1	143
43	198	131	107	127	7	1	1	143
44	198	131	107	127	7	1	1	143
45	198	131	107	127	7	1	1	143
46	198	131	107	127	7	1	1	143
47	198	131	107	127	7	1	1	143
48	198	131	107	127	7	1	1	143
49	198	131	107	127	7	1	1	143
50	198	131	107	127	7	1	1	143
51	198	131	107	127	7	1	1	143
52	198	131	107	127	7	1	1	143
53	198	131	107	127	7	1	1	143
54	198	131	107	127	7	1	1	143
55	198	131	107	127	7	1	1	143
56	198	131	107	127	7	1	1	143
57	198	131	107	127	7	1	1	143
58	198	131	107	127	7	1	1	143
59	198	131	107	127	7	1	1	143
60	198	131	107	127	7	1	1	143
61	198	131	107	127	7	1	1	143
62	198	131	107	127	7	1	1	143
63	198	131	107	127	7	1	1	143
64	198	131	107	127	7	1	1	143
65	198	131	107	127	7	1	1	143
66	198	131	107	127	7	1	1	143
67	198	131	107	127	7	1	1	143
68	198	131	107	127	7	1	1	143
69	198	131	107	127	7	1	1	143
70	198	131	107	127	7	1	1	143
71	198	131	107	127	7	1	1	143
72	198	131	107	127	7	1	1	143
73	198	131	107	127	7	1	1	143
74	198	131						

tabellen is er sprake van een kolom 'ongeldig', voor ongeldig gemaakte stembiljetten, en een kolom 'afwezig', voor de mensen die niet van hun stemrecht gebruik hebben gemaakt.

De verklaring van de voor de politieke partijen gebruikte afkortingen staat naast tabel 2.

Deze tabellen hebben achteenvijftig stembedistricten. Geografisch gezien horen de volgende stembedistricten bij elkaar: 1 t/m 9; 10 t/m 14; 15 en 16; 17 t/m 27; 28 t/m 38; 39 t/m 50; 51 t/m 58. Deze clusters van stembedistricten zullen in de resultaten van de analyses in grote lijnen zijn terug te vinden.

Achtereenvolgens wordt behandeld de analyse van 1981 en 1982 apart, de analyse van 1981 en 1982 tezamen, en de berekening van betrouwbaarheidsintervallen voor de kolompunten (partijen).

3.1 De analyse van de verkiezingen 1981 en 1982 afzonderlijk

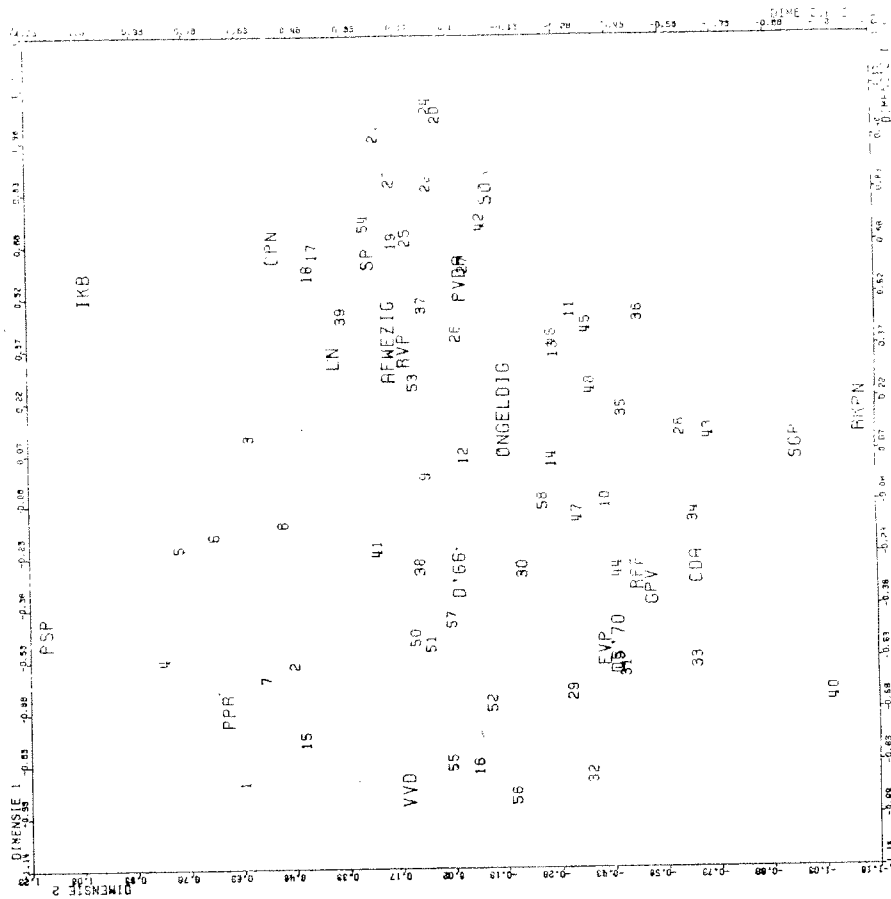
Indien we met het programma ANACOR correspondentieanalyse doen op tabel 1 (1981) krijgen wij het volgende plaatje (zie figuur 1). De stembedistricten hebben hier dezelfde nummers als in de tabel. Hetzelfde geldt voor de labels van de partijen.

De stembedistricten 1 t/m 9, liggend in de binnenstad, zijn dicht te vinden bij de PSP en de PPR. Dit ligt waarschijnlijk voor een belangrijk deel aan de hier wonende studenten. Districten 15 en 16, horend bij een ietwat goede buurt, liggen dicht bij de VVD. Districten 17 t/m 27, horend bij een verkrotte buurt, liggen voornamelijk rond de PVDA en de CPN. Districten 28 t/m 38, die in een oudere gegoede buurt liggen, is te vinden bij de christelijke partijen. Stembedistricten 51 t/m 58, horend bij een nieuwbouwwijk met voornamelijk duurdere woningen, liggen voor het grootste deel in de buurt van D'66 en de VVD. Districten 39 t/m 50, ook liggend in een nieuwbouwwijk, zijn zo'n beetje bij alle grote partijen in de buurt te vinden. De niet-stemmers wonen vooral in districten waar veel op de PVDA wordt gestemd, en wonen relatief minder in wijken waar veel CDA en VVD wordt gestemd.

We kunnen ook individuele districten bekijken. Zo ligt district 1 in de buurt van de VVD en PPR, PSP. Als we naar tabel 1 kijken, zien we dat in dit district op deze partijen hoger dan het stedelijk gemiddelde is gescoord.

De verkiezingen van 1982 laten wat de plaats van de stembedistricten onder-

Figuur 1: Verkiezingen Tweede Kamer 1981



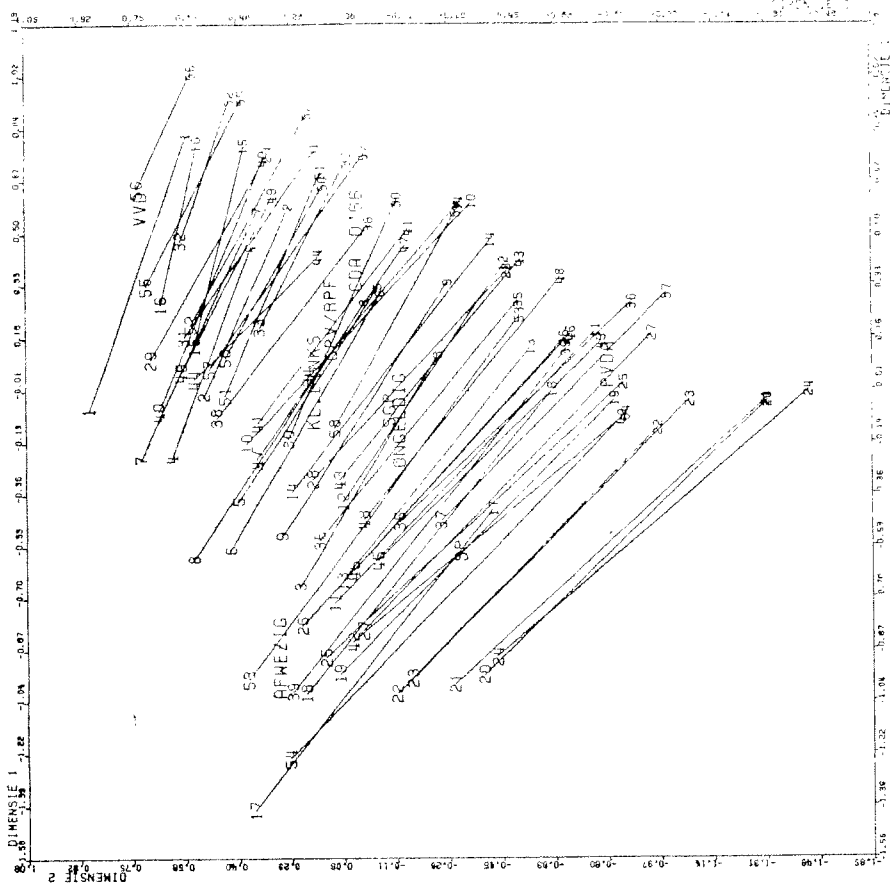
ling betreft een vrijwel identiek beeld zien (figuur 2). Dit geldt niet voor de partijen: het punt afwezig ligt dicht bij de stembedistricten uit de verkrotte buurt, de PVDA ligt iets verder van dezelfde districten af. VVD en D'66 liggen dicht bij elkaar. D'66 lijkt verloren te hebben in de verkrotte buurt, of gewonnen in de VVD-districten (is namelijk vanaf de verkrotte buurt naar de 'VVD-achtige' districten bewegen).

Tot zover de analyses van de verkiezingen 1981 en 1982 afzonderlijk.

3.2.1 Tabellen onder elkaar

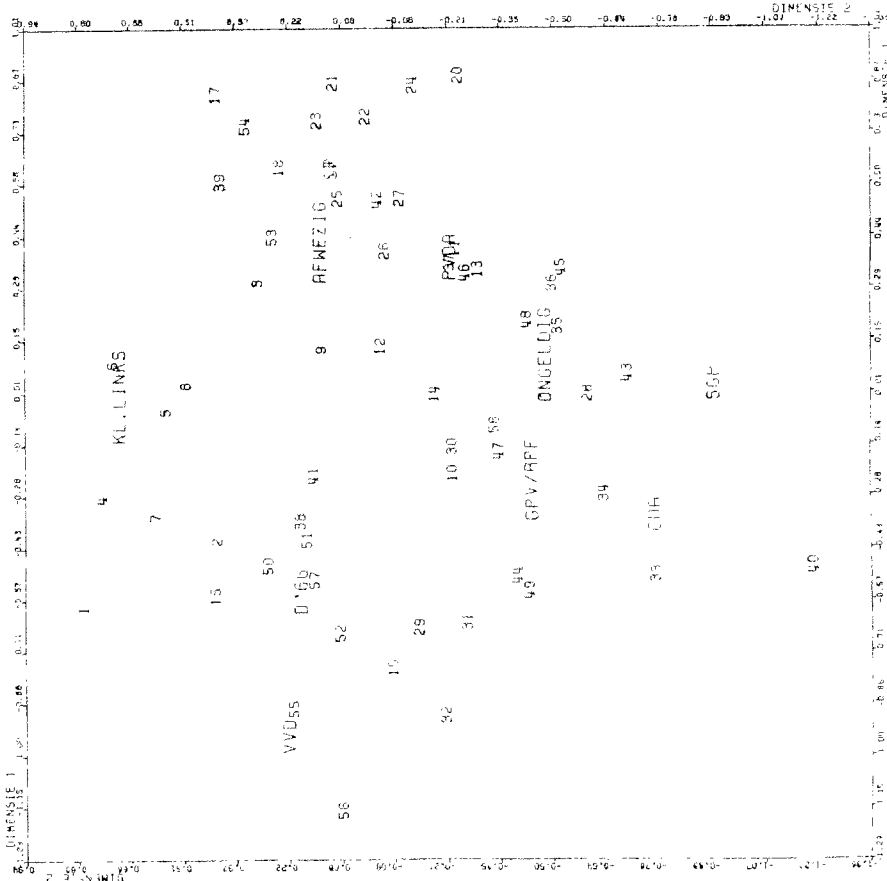
Het is mogelijk de tabel van 1982 onder die van 1981 te zetten. Vóór we dit doen, moeten we zorgen dat deze tabellen een gelijk aantal kolommen, met identieke betekenis per kolom, hebben. Hiertoe nemen we voor 1981 de partij- en PSP, CPN en PPR tesamen, evenals RPF en de GPV. De kolommen RVP, DS'70, RKN, LN, IKB, EYP en SOS laten we weg. Dit is bij elkaar 2,2% van de stemmen van 1981.

Figuur 3: Verkiezingen Tweede Kamer 1981 en Provinciale Staten 1982



STEMMESTRUKTUREN (GELUMMERD) IN 1981, GROOT (1981) EN PARTIJEN (AFGEORD.)

Figuur 2: Verkiezingen Provinciale Staten 1982



STEMMESTRUKTUREN (GELUMMERD) EN PARTIJEN (AFGEORD.)

3.2 De analyse van de verkiezingen 1981 en 1982 tesamen

Hoe kunnen we bekijken of er verschuivingen zijn opgetreden in de uitslagen van 1982 ten opzichte van die van 1981? Hier hebben we drie manieren voor. Twee behelzen het aan elkaar plakken van de tabellen. De derde manier is de bimarignale tabel te analyseren. Deze drie manieren zullen we hieronder achtereenvolgens beschrijven.

De nieuwe tabel bestaat nu uit 116 districten (ieder district twee maal) en tien partijen. We kunnen nu bekijken hoe voor elk district het punt van 1981 zich verhoudt tot dat van 1982. Dit is te zien in figuur 3. De districten zijn met lijnen verbonden.

Alle districten maken een sprong in dezelfde richting, namelijk vanaf de PVDA naar AFWEZIG. Dit komt doordat in alle districten de PVDA in 1981 veel minder stemmen heeft gekregen dan in 1982, en AFWEZIG veel meer stemmen. Indien er een lijn wordt getrokken van het punt PVDA naar het punt afwezig, zien we dat de lijnen die de bij elkaar horende stemdistricten verbinden, hieraan parallel lopen. De oorzaak hiervan is dat wijken zowel van de PVDA af moeten bewegen als naar AFWEZIG toe. Dit kan alleen door de lijn tussen 1981 en 1982 van een stemdistrict parallel aan de lijn van de PVDA naar AFWEZIG te laten lopen.

Indien alle partijen op deze lijn geprojecteerd worden, zien we de partij- en ongeveer in de volgorde winst verlies hierop liggen: AFWEZIG - (VVD-KLEIN LINKS) - (SGP-RPF/GPV-ONGELDIG) - CDA - D'66 - PVDA. Bovendien zien we de PVDA, AFWEZIG en de VVD extreem liggen: zij halen hun stemmen vooral uit een specifieke stemdistrictengroep. Dit in tegenstelling tot het CDA en KLEIN LINKS.

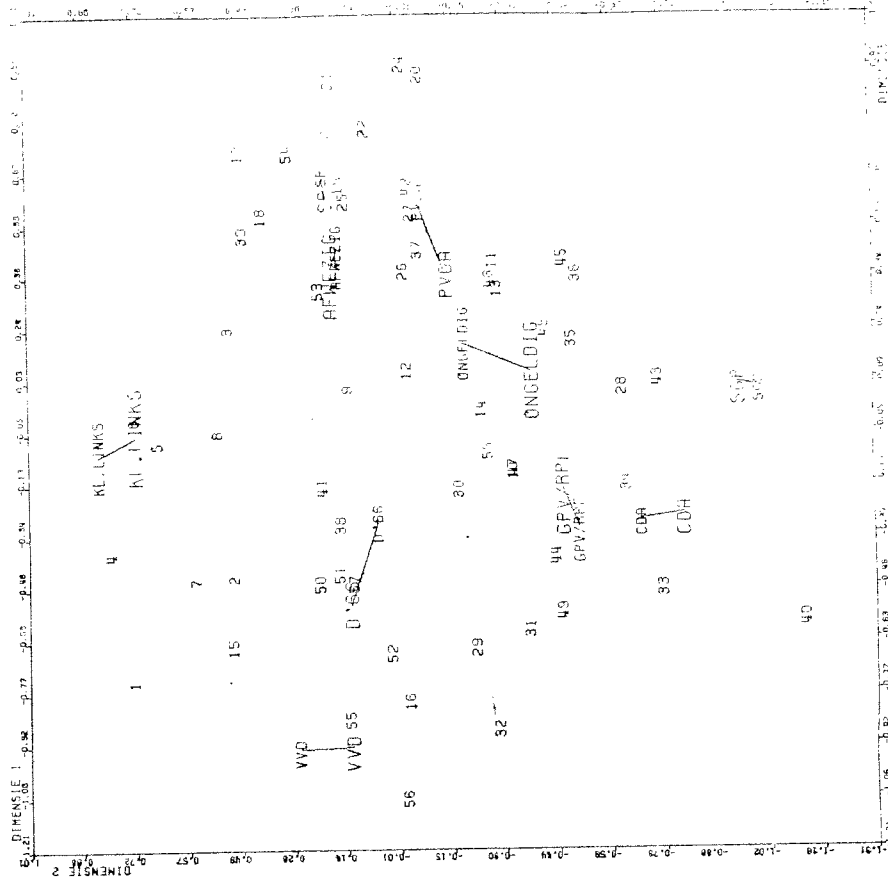
Vergelijken we figuur 3 met figuur 1 en 2, dan blijkt door de lengte van de lijnen dat de verandering in de stemdistricten waar veel PVDA, en veel niet gestemd wordt, veel groter is dan in de districten waar veel VVD wordt gestemd.

3.2.2 Tabellen naast elkaar

Plaatsen we de tabel van 1981 met 10 partijen naast die van 1982, en analyseren we deze tabel, dan krijgen we een plaatje waarin we partijpunten met elkaar kunnen verbinden (figuur 4). We zien dat de PVDA haar stemmen in 1982 minder specifiek uit de verkrotte buurt heeft gekregen, en omdat we weten dat de PVDA minder stemmen heeft gekregen, duidt dit op verlies van de PVDA in deze buurt.

Klein links is meer naar het centrum van het plaatje verschoven, en D'66 meer naar de VVD: daar D'66 heeft verloren duidt dit erop dat zij afhanke-lijker is geworden van de stemmen uit 'VVD-achtige' districten. Het punt AFWEZIG is vrijwel blijven liggen: dit betekent dat het frequentiepatroon van de thuisblijvers per district voor de twee verkiezingen redelijke overeenkomst vertoont, en dus dat in de meeste districten een rond de twee en een half maal zo groot aantal thuisblijvers woont in 1982 als

Figuur 4: Verkiezingen Tweede Kamer 1981 en Provinciale Staten 1982



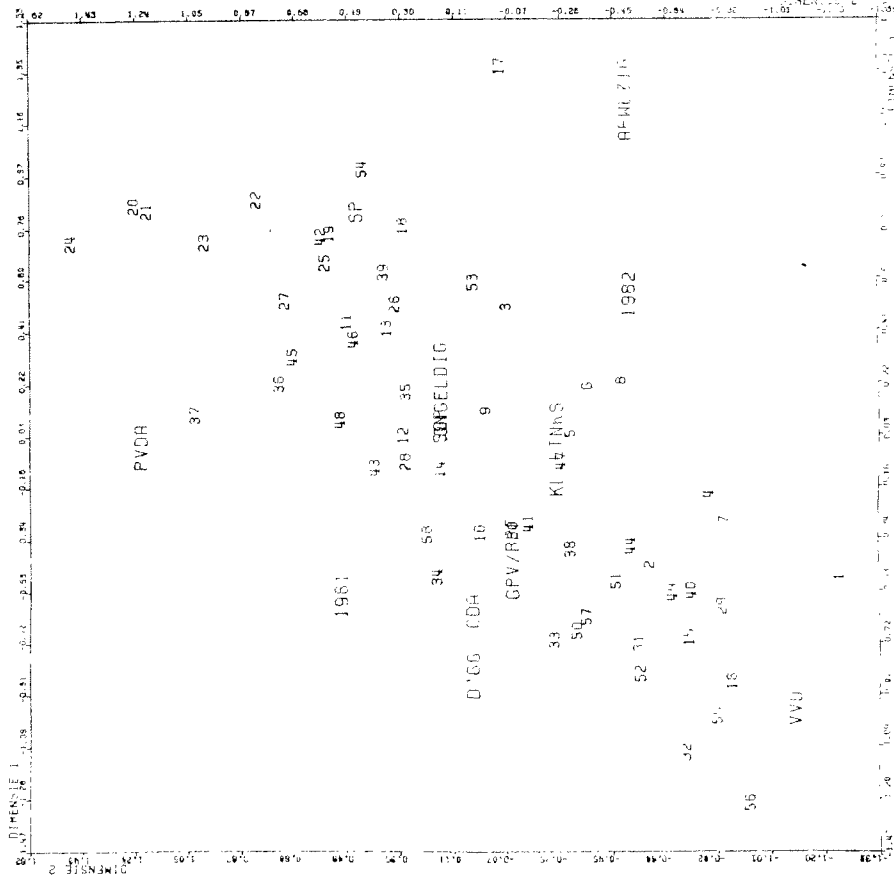
in 1981 (zie tabellen 1 en 2). Daar thuisblijven vooral gebeurt in de verkrotte buurt (daar ligt het punt immers bij), en de PVDA daar hoog scoorde, verklaart dit gedeeltelijk het verlies van de PVDA. Het CDA is wat haar stemmen betreft meer afhankelijk geworden van haar specifieke stemdistricten, aangezien het CDA-punt van 1982 meer naar de periferie ligt dan het CDA-punt van 1981. ONGELDIG heeft de grootste verschuiving ondergaan. Deze verschuiving zegt echter niet zoveel, daar dit slechts op een klein aantal stemmers betrekking heeft.

3.2.3 Bimarginale matrix

Een zogenaamde bimarginale matrix construeert men door als kolommen en rijen identieke categoriën te nemen, en wel de jaren (2), de partijen (10) en de stemdistricten (58).

De bimarginale matrix heeft dus 70 rijen en 70 kolommen, namelijk zowel voor de rijen als de kolommen de categoriën 1981, 1982, PVDA t/m AFWEZIG, district 1 t/m 58. De oorspronkelijke 3-weg matrix kun je op drie manieren optellen

Figuur 5: Verkiezingen 1981 en 1982: de bimarginale matrix



STEMMINGSRESULTATEN VERKIEZINGEN 1981 EN 1982 (ZIE FIGUUR 5)

tot een 2-weg matrix: over de jaren, over de districten en over de partijen. Alle drie op deze wijze verkregen 2-weg matrices zijn terug te vinden in de bimarginale matrix. Er zitten ook lege cellen in de matrix: bijvoorbeeld de cel behorend bij rij 1981 en kolom 1982: een stem kan niet tegelijk zowel in 1981 als in 1982 uitgebracht zijn.

Het resultaat van de analyse van deze 70 bij 70 matrix is te zien in figuur 5 (NB. SGP en ONGELDIG overlappen elkaar).

Bij 1982 ligt het punt afwezig en KL. LINKS, deze hebben in 1982 meer stemmen gekregen dan in 1981. Het omgekeerde geldt voor de PVDA, D'66 en het CDA: deze hebben in 1982 minder stemmen gekregen. De PVDA en het punt AFWEZIG liggen verder uiteen dan op andere plaatjes.

In figuur 5 kunnen we zien dat in bijvoorbeeld district 37 de PVDA relatief minder heeft verloren dan in andere districten die ook in de verkrotte buurt liggen, en méér mensen naar de stembus zijn gegaan. In districten 17, 18, 53 en 54 is relatief méér niet gestemd, en tegelijkertijd is het aantal PVDA-stemmen relatief meer gedaald dan in andere wijken.

3.3 Betrouwbaarheidsintervallen

Indien we aannemen dat de data zijn op te vatten als een willekeurige steekproef uit een onbekende populatie, dan volgen de celfrequenties een multivariate verdeling. Onder aanname van dit model is het mogelijk de momenten (gemiddelde, variantie) van de parameters te berekenen.

Het computerprogramma ANACOR kan de varianties van de singuliere waarden en de rij- en kolomscores berekenen, en de correlaties tussen scores van dezelfde parameter in verschillende dimensies. Zo bestaat de mogelijkheid de plaatjes van rij- en kolompunten van betrouwbaarheidsintervallen te voorzien: er kunnen ellipsen getekend worden die het gebied aangeven waarbinnen elk populatiepunt met 95% zekerheid ligt. (Het programma levert slechts de scores met behulp waarvan de ellipsen zijn te construeren, niet de ellipsen zelf).

Wat betreft de aanname van een willekeurige steekproef uit een onbekende populatie, moeten we ons iets voorstellen als dat de districten staan voor een grotere groep vergelijkbare districten uit bijvoorbeeld Nederland.

We hebben de betrouwbaarheidsgebieden van de partijpunten van figuur 4 als voorbeeld genomen (zie figuur 6).

Wat men kan zien is dat grotere partijen kleinere ellipsen hebben dan de kleinere partijen, en dus stabielier op hun plaats liggen. Ook van de ver-

3.4 Naar welk plaatje te kijken?

Figuur 1 t/m 5 zijn te beschouwen als vijf verschillende weergaven van de 3-weg matrix. Alle vijf leggen ze echter andere accenten. Deze zullen we kort aanstippen.

Figuur 1 en 2 geven de verkiezingsuitslag van 1981 en 1982 afzonderlijk weer. Indien men slechts in deze afzonderlijke uitslagen geïnteresseerd is, zijn deze figuren voldoende.

Figuren 3, 4 en 5 geven verschuivingen weer.

Indien men interactie veronderstelt tussen wijken en jaren, kijkt men naar figuur 3. Men kan bijvoorbeeld zien dat sommige wijken meer verschoven zijn dan andere, of in een andere richting, en daar conclusies uit trekken.

Hetzelfde kan gezegd worden over de interactie tussen partijen en jaren.

Indien men deze veronderstelt, kijkt men naar figuur 4.

In figuur 5 is de afbeelding van de variabelen te vinden zonder dat a priori interactieeffecten zijn verondersteld. Indien men niet zozeer geïnteresseerd is in verschuivingen van wijken en partijen, kan men deze figuur bekijken.

Correspondentieanalyse is op te vatten als een exploratieve dataanalysestechniek. Het is aan te bevelen terug te vallen op de oorspronkelijke frequentietabellen in geval van twijfel bij de interpretatie van de plaatjes.

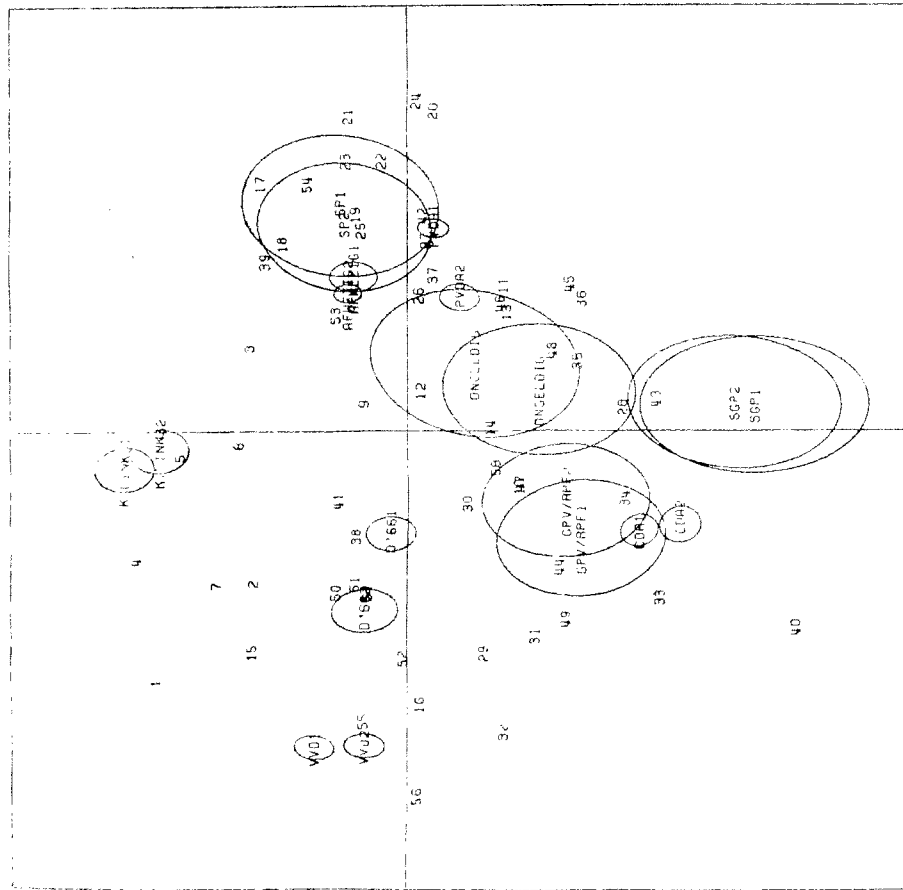
4. Verder onderzoek

De hier gehanteerde methode kan op veel terreinen voor de analyse van 2-weg of 3-weg matrixen worden toegepast. Zo is het bij verkiezingsonderzoek interessant om meer dan twee jaren in te voeren. Men kan ook iets anders als rijen nemen, bijvoorbeeld plaatsen, provincies. Ook zou men slechts één of twee partijen kunnen nemen, en vervolgens verschuivingen van wijken per verkiezing kunnen bestuderen.

Een theoretischer tak van verder onderzoek is de relatie tussen loglineaire analyse en correspondentieanalyse: bij loglineaire analyse veronderstelt men interactie, en vervolgens worden de verwachte frequenties vergeleken met de waargenomen frequenties, en gekeken hoe goed de fit van het model is; bij correspondentieanalyse veronderstelt men óók interactie (bij een matrix van meer dan twee dimensies) en onderzoekt men vervolgens de (on)afhankelijkheid in een geconstrueerde 2-weg matrix.

De exacte relatie tussen loglineaire analyse en correspondentieanalyse is nog niet duidelijk. Aan het uitzoeken hiervan wordt momenteel gewerkt.

Figuur 6: Figuur 4 opnieuw, ellipsen rondom de partijen



schuivingen van deze partijen kunnen we dus zekerder zijn: de verschuivingen van de PVDA, het CDA, de VVD en D'66 zijn alle significant: de betrokken gebieden overlappen niet.

Literatuur

- Benzécri, J.P. & Collaborateurs,
1973, L'analyse des Données, II L'analyse des Correspondances. Dunod
Paris-Bruxelles-Montréal
- Gifi, A.,
1981, Non-linear multivariate analysis, Department of Datatheory,
University of Leyden
- Gifi, A.,
1982, ANACOR User's guide, Department of Datatheory, University of Leyden
- Hill, M.O.
1974, Correspondence analysis: a neglected method. Applied Statistics 23,
pp. 340-354
- Horst, P
1935, Measuring complex attitudes. J. Soc. Psychoï. 6, pp. 369-374
- Israëls, A.Z.; Bethelam, J.P.; van Driel, J.; Janssen, M.E.; Pannekoek, J.;
de Ree, S.J.M.; Sikkel, D.,
1981, Multivariate analysemethoden voor discrete variabelen.
Kwantitatieve methoden 2, pp. 128-133

STMPLE RANDÓMIZED RESPONSE PROCEDURES WITH BOUNDED RESPONDENT
RISK FOR QUANTITATIVE DATA

W. Albers (1)

SUMMARY: For the case of quantitative data we introduce a class of randomized response procedures which are simple, have bounded risk for the respondent, and satisfy a certain admissibility condition. We also indicate how approximately optimal procedures in this class can be found in a simple way.

1. Introduction

Randomized response procedures (rrp's) were introduced by Warner (1965) to reduce respondent bias in surveys of human populations involving embarrassing questions. Since then various types of rrp's have been proposed, among others by Greenberg et al. (1971) and by Warner (1971). Below we shall by way of introduction discuss two examples which will also be useful in the sequel.

In this paper we consider the case of quantitative data. Let X denote the answer of an arbitrarily chosen respondent to a sensitive question, then we want to estimate $\mu = EX$.

In this situation, Greenberg et al. (1971) suggest to select an unrelated and innocent question such that the range of its answer Z is approximately the same as that of X and moreover the distribution of Z is known. Then the respondent is asked to give not X but merely the randomized response

$$Y = VX + (1-V)Z, \quad (1.1)$$

where V , X and Z are independent and $P(V=1) = 1-P(V=0) = c$. Hence the randomization in the response is due to the presence of V . If c is not too close to 1, it can be hoped that, while the respondent might refuse to give X or might lie about its value, he will not mind to give Y .

(1) Mathematics Department, Twente University of Technology.

Key Words: Randomized response; Respondent risk; Quantitative Data.