

## De voorkeur voor een politieke partij : het toetsen van loglineaire modellen\*

---

Jaak BILLIET,

Docent aan de Katholieke Universiteit te Leuven

★

De bedoeling van dit artikel is tweevoudig. In de eerste plaats worden twee hypothesen getoetst die gesuggereerd worden door A. Frognier (1) en K. Hill (2) maar die bij gebrek aan empirische gegevens nog niet getest werden. Ten tweede wil ik de mogelijkheden illustreren van het toetsen van loglineaire modellen om deze analyse-procedure meer toegankelijk te maken voor onderzoekers. Deze doorgedreven techniek van multivariabele analyse legt geen beperkingen op aan het meetniveau. De techniek is bruikbaar voor zgn. categorische variabelen (nominaal meetniveau) en men kan de analyse uitvoeren wanneer men over multidimensionele tabellen beschikt waarvan de verdeling tot stand kwam via a-selecte steekproeftrekking mits het aantal eenheden voldoende groot is. In een eerste paragraaf worden de hypothesen voorgesteld. Vervolgens krijgt men enige informatie over de data-verzameling, meting en het geldigheidsveld. De derde paragraaf behandelt tenslotte uitvoerig de verschillende stappen bij het toetsen van loglineaire modellen.

### 1. Hypothesen over partij-voorkeur.

A. Frognier beschrijft de samenstelling van de verschillende politieke partijen naar twee kenmerken van de kiezers, met name de kerkelijke praktijk en het subjectief behoren tot een sociale klasse. Hiervoor wordt gebruik gemaakt van een maat van fragmentatie ontwikkeld door D. Rae

---

\* Met dank aan J. Teugels, J. Tack en G. Loosveldt voor hun waardevolle suggesties.

(1) A. FROGNIER, Vote, classe sociale et religion/pratique religieuse, in *Res Publica*, 17 (1975), 4, blz. 479-490.

(2) K. HILL, Belgium : Political Change in a Segmented Society, in R. Rose (ed.), *Electoral Behavior : A Comparative Handbook*, N.Y., 1974, blz. 29-108.

en M. Taylor (3). Vervolgens onderzoekt deze auteur de zwakke plekken van elke partij in het electoraat aan de hand van een door A. Lijphart ontwikkelde maat (4). Telkens wordt een onderscheid gemaakt tussen Vlaanderen, Wallonië en Brussel. Hier interesseert ons enkel de situatie in Vlaanderen, meer speciaal voor wat betreft de keuze voor de CVP en de BSP. Eerstgenoemde partij blijkt wat de sociale klasse betreft heterogeen te zijn, doch er is een overconcentratie van kerkelijk praktiserenden (voortaan « kerksen » genoemd). Voor de kerksen zou de relatie tussen het stemmen voor de CVP en het behoren tot een sociale klasse zelfs verdwijnen (5). De BSP daarentegen wordt gekenmerkt door een overwicht van zowel onkerksen als personen die zich tot de arbeidersklasse rekenen. De BSP heeft m.a.w. haar zwakste plekken bij de kerksen en de niet-arbeiders, terwijl de CVP enkel zwak is m.b.t. de onkerksen en in lichte mate bij degenen die zich tot de hoogste klasse rekenen, doch dit is kwantitatief de kleinste kiezersgroep (6). De bevindingen van Frogner hebben betrekking op 1974.

De keuze van vermelde variabelen is niet willekeurig. Die sluit aan bij de verschillende auteurs die handelen over de relaties tussen het bestaan van sociale conflictlijnen (« social cleavages ») en de vorming van politieke partijen. De godsdienstige (Kerk-Staat) en de sociaal-economische tegenstelling zijn twee van de « basic-cleavages » die hun oorsprong vinden in de nationale revolutie, het ontstaan van de staat, en de industriële revolutie (7). Laatstgenoemde tegenstelling heeft echter de eerste niet vervangen maar heeft er zich op geënt, vandaar een partijvorming waar enerzijds de tegenstelling « kerk-staat » domineert (katholieke partij versus liberale én socialistische) en anderzijds waar de sociaal-economische tegenstelling primeert (vleugels binnen de christen democratie, liberaal versus socialist) (8). De taaltegenstelling, alhoewel niet onbelangrijk wat het kiesgedrag betreft, laten we hier buiten beschouwing.

In de betreffende literatuur wordt gesignaleerd dat in België de arbeiders het minst van al in vergelijking met andere Westerse landen, stemmen

---

(3) A. FROGNIER, *art. cit.*, blz. 486-487 ; A. LIJPHART, *The Politics of Accommodation*, Berkeley, 1968.

(5) A. FROGNIER, *art. cit.*, blz. 485-486.

(6) *Ibid.*, blz. 487, 489.

(7) O.a. S.M. LIPSET and S. ROKKAN, Cleavage structures, Party Systems and Voter Alignments : An Introduction, in S.M. Lipset and S. Rokkan (eds.), *Party Systems and Voter Alignments : Cross National Perspectives*, New York, 1967, blz. 1-64 ; D. URWIN, Social Cleavages and Party Systems in Belgium ; Problems of Institutionalization, in *Political Studies*, 18 (1970) 3, blz. 320-330 ; V. LORWIN, Segmented Pluralism, Ideological Cleavages and Political Cohesion in the Smaller European Democracies, in K. McRae (ed.), *Consociational Democracy*, Toronto, 1974, blz. 34-35 ; V. LORWIN, Belgium : Conflict and Compromise, in K. McRae (ed.), *op. cit.*, blz. 178-179

(8) D. URWIN, *art. cit.*, blz. 325.

op basis van sociale klasse (9). De bevindingen van A. Frogner sluiten hierbij aan. De relatie tussen sociale klasse en partijvoorkeur wordt verzwakt doordat een groot aantal arbeiders de christen-democratie blijft steunen. De BSP daarentegen is (was?) het meest aantrekkelijk voor *onkerkse* arbeiders.

In tegenstelling tot A. Frogner sluit K. Hill wél rechtstreeks aan bij de vermelde theorievorming over « social cleavages » en stemgedrag (10). Na een historische benadering van de groei en het politieke systeem — met overwegende aandacht voor de partij- en verkiezingsproblematiek — wordt een analyse gemaakt van de relatie tussen stemgedrag, kerkelijkheid, beroep, leeftijd e.a., en van de « turnover » (11). De auteur steunt op de gegevens van het verkiezingsurvey uit 1968 (12) en op gegevens van de *European Communities study* van 1970.

Deze auteur repliceert de analyse van N. Delruelle *et al.* met behulp van AID (Automatic Interaction Detector) (13). De conclusies zijn de volgende. Er wordt geen sterke relatie vastgesteld tussen kerkelijkheid, beroep, regio en partijkeuze. Dit zou eigen zijn aan de onderwerping van de kiezers aan « crosscutting pressures » langs de drie dimensies van partijkeuze (religie, klasse en taal). De kerkelijkheid blijkt de voornaamste verklarende variabele voor het stemgedrag. De rol van de variabele « sociale klasse », en van de andere variabelen, wordt hierdoor afgezwakt (14). K. Hill vermeldt echter het belang van de zgn. « organisationele » variabele die o.i. verband houdt met de deelname aan zuilorganisaties (15). Deze belangrijke variabele kon echter wegens het gebrek aan informatie daarover, niet in het onderzoek betrokken worden. Alhoewel de organisationele variabele verband houdt met de kerk-staat conflictlijn is het toch noodzakelijk deze te onderscheiden van de variabele kerkelijkheid zoals ik reeds herhaaldelijk heb aangetoond (16). Het is juist die variabele die in de verdere analyse zal worden opgenomen.

(9) D. URWIN, *art. cit.*, blz. 321. Deze auteur verwijst naar M. Claeys-Vanhaegendoren, *Party and opposition Formation in Belgium*, in *Res Publica*, 9 (1967), blz. 413-456.

(10) K. HILL, *op. cit.*, blz. 29-51. Aanvankelijk probeert deze auteur de vier « cleavages » te betrekken, namelijk kerk-staat, stad-platteland, arbeid-kapitaal, centrum-periferie.

(11) K. HILL, *op. cit.*, blz. 69-103.

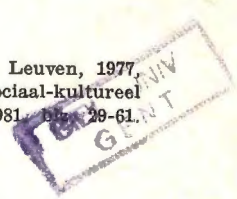
(12) N. DELRUELLE, R. EVALENKO en W. FRAYES, *Le comportement politique des élections belges*, Bruxelles, 1970.

(13) K. HILL, *op. cit.*, blz. 103. Voor A.I.D. zie J.A. SONQUIST, E.L. BAKER en J.N. MORGAN, *Searching for Structure*, Michigan, 1973.

(14) *Ibid.*, blz. 95-96, 100.

(15) *Ibid.*, blz. 42-43.

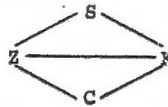
(16) O.a. J. BILLIET, *Secularisatie en verzuijing in het onderwijs*, Leuven, 1977, blz. 116-122; J. BILLIET, *Kenmerken en grondslagen van het sociaal-kultureel katholicisme*, in J. Servaes (ed.), *Van ideologie tot macht*, Leuven, 1981, blz. 29-61.



Het vorige laat toe twee hypothesen te formuleren m.b.t. de interrelaties tussen partijkeuze, kerkelijkheid, verzuimdheid en sociaal-professioneel niveau.

*Hypothese 1.* Er is een direct verband tussen de voorkeur voor de CVP (C) en respectievelijk de kerkelijkheid (K) en de mate van verzuimdheid (Z). Deze relaties zijn niet schijnbaar, d.w.z. dat ze stand houden wanneer de andere variabelen in het model onder controle worden gehouden. De richting van de relaties is de volgende: de kans op voorkeur voor de CVP is aanzienlijk groter bij kerksen dan bij niet-kerksen. Deze kans is eveneens groot bij degenen die geïntegreerd zijn in katholieke zuilorganisaties (afgezien van de mate van kerkelijkheid), ze neemt af bij de niet-verzuilden en is het geringst bij de geïntegreerden in socialistische organisaties.

Wegens de gekende samenhang (17) tussen kerkelijkheid, verzuimdheid en sociaal niveau (S) kan deze hypothese als volgt voorgesteld worden.



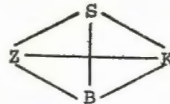
De relatie tussen voorkeur voor de CVP en kerksheid is vanzelfsprekend, de relatie met verzuimdheid op het eerste gezicht ook, doch er is tot nu toe geen toetsing gebeurd omdat deze organisationele variabele niet los van de kerkelijkheid werd beschouwd. Belangrijk is hier dat deze relatie niet schijnbaar is wegens de samenhang kerkelijkheid-verzuiling. Dit is van belang in het kader van onze theorievorming rond secularisatie en verzuiling waarin het integratief moment van de katholieke wereld veeleer gezien wordt in het sociaal-cultureel katholicisme, d.w.z. het georganiseerd katholicisme, dan in de kerkelijke binding (18). Van belang is ook in tegenstelling tot de tweede hypothese, het weglaten van elk direct verband tussen CVP-voorkeur en sociaal niveau.

*Hypothese 2.* Er is een direct verband tussen de voorkeur voor de BSP (B) en respectievelijk de kerkelijkheid (K), de verzuimdheid (Z) en het sociaal niveau (S). De kans op een voorkeur voor de BSP is het hoogst bij de onkerksen, degenen die geïntegreerd zijn in de socialistische organisaties en bij de handarbeiders. Niet alleen verschilt de

(17) J. BILLIET, *op. cit.*, blz. 150.

(18) J. BILLIET, Beschouwingen over het samengaan van secularisatie en verzuiling, in *De Nieuwe Maand*, 19 (1976), 7, blz. 243-258.

richting van de relaties in vergelijking met de eerste hypothese, opvallender is nog het verschil in model zoals blijkt uit volgend schema.



Bemerk dat de twee hypothesen symmetrisch geformuleerd zijn en dat er geen onderscheid werd gemaakt tussen afhankelijke en onafhankelijke variabelen. We willen dit zo houden om beter de werking van loglineaire modellen te kunnen uitleggen. Asymmetrische modellen kunnen eveneens getoetst worden doch het vergt eerst inzicht in het symmetrisch model.

## 2. Geldigheidsveld en dataverzameling.

Voor het toetsen van de hypothesen wordt gesteund op het materiaal van mijn onderzoek naar de schoolkeuze van 1974. Het geldigheidsveld betreft alle te Mechelen wonende vrouwen met een kind in het eerste leerjaar van alle Mechelse secundaire scholen in het schooljaar 1973-1974. Uit deze populatie van 817 eenheden werd een disproportionele gestratificeerde toevalssteekproef van 455 eenheden getrokken die echter bij de analyse met behulp van weegcoëfficiënten in de juiste verhoudingen werd teruggebracht. De non-respons bedroeg slechts 16 procent en kon nauwkeurig gelokaliseerd worden op basis van de verschillende stratificatievariabelen. De externe geldigheid is tot op grote hoogte gewaarborgd (19). Ten opzichte van de globale Vlaamse populatie gelden zeker de volgende restricties : het gaat uitsluitend om vrouwen met minstens een kind van 11-13 jaar ; 90 % van de vrouwen situeert zich binnen de leeftijdsgrenzen 30-50 jaar en naar verstedelijking van de actuele woonplaats is deze populatie homogeen. Ondanks de geringe variatiebreedte van de variabele leeftijd, blijkt nog een zwak verband met kerkelijkheid, doch van een rechtstreekse relatie met de overige variabelen is geen sprake. Hetzelfde geldt voor de mate van verstedelijking wanneer men het verleden van de ondervraagden in aanmerking neemt, alleen zijn de verbanden zo mogelijk nóg zwakker. We laten de variabelen leeftijd en verstedelingsgraad dan ook buiten beschouwing.

De variabele *kerkelijkheid* (K) is een index geconstrueerd op basis van een vraag naar de godsdienst als zodanig (katholiek of niet), de

(19) Voor de externe geldigheid van de steekproef wordt verwezen naar J. BILLIET, *op. cit.*, blz. 132-136.

seizoenritten (doopsel, vormsel, huwelijk), de kerkelijke praktijk (weekend liturgie e.a.) en het belang daaraan gehecht voor de kinderen. Alhoewel uiteindelijk voor de loglineaire analyse enkel sprake zal zijn van twee categorieën, namelijk kerksen ( $K_2$ , grotendeels overeenkomend met regelmatige praktijk) en onkerksen ( $K_1$ , niet praktiserende katholieken en niet-katholieken) is de meting scherper dan men doorgaans in de verkiezings-surveys aantreft. De studie werd immers overwegend vanuit godsdienstsociologische hoek opgezet. We menen te kunnen stellen dat onze twee categorieën vanuit theoretisch oogpunt betrekking hebben op twee onderscheiden types van kerkelijke normatieve integratie (het aanvaarden en naleven van de kerkelijke normen). Het ware beter geweest de niet-katholieken afzonderlijk te nemen, doch hun aantal is te gering (nl. slechts 70). Men vindt ze overwegend in de tussencategorie van de verzuiling (niet verzuilden), in alle sociale lagen doch iets sterker in de bovenste en met iets meer voorkeur voor de PVV. Naast enkele protestanten vindt men er overwegend katholiek gedoopten die zich om diverse redenen « niet-katholiek » noemen (20).

De variabele *verzuildheid* ( $Z$ ) is eveneens een index die geconstrueerd is op grond van vragen naar het lidmaatschap van mutualiteit, van beroeps-, culturele- en vrouwenverenigingen, het schoolnet tijdens de kinderjaren van de vrouw en het schoolnet van de kinderen, de tewerkstelling in zuilorganisaties en kernlidmaatschap. Bij deze variabele onderscheiden we drie categorieën: de geïntegreerden in de socialistische zuil ( $Z_1$ ); de geïntegreerden in de katholieke zuil ( $Z_2$ ) en de overigen ( $Z_3$ , grotendeels niet-verzuilden op enkele liberalen na).

De variabele *sociaal niveau* ( $S$ ) heeft eigenlijk betrekking op een gezinskenmerk vermits een belangrijk gedeelte van de vrouwen niet buitenhuis werkt. Er werd dus ook rekening gehouden met de beroeps-situatie van de echtgenoot. Echter, ook het subjectief behoren tot een sociale klasse werd bevraagd en vertoonde een dermate grote overeenstemming met onze « objectieve » indeling (Kendall tau van 0.54) zodat bleek dat de ondervraagde vrouwen bij hun subjectieve klassebepaling wel degelijk overwegend het beroep van hun echtgenoot in aanmerking namen. Met de uitzonderingen hierop werd zoveel mogelijk rekening gehouden. De variabele sociaal niveau heeft drie categorieën, het laag sociaal niveau ( $S_1$ , ongeschoolde en laaggeschoolde handarbeiders), het midden niveau ( $S_2$ , gespecialiseerde handarbeiders, zelfstandige arbeiders, lagere bedienden, kleine winkeliers) en het hoogste niveau ( $S_3$ ,

---

(20) Voor de constructie van de indexen wordt verwezen naar J. BILLIET, *op. cit.*, blz. 116-131, 140-141, 229-244.

hogere bedienden, gespecialiseerde en grote zelfstandigen, vrije beroepen, directies van scholen enz...).

De variabele *partijvoorkeur* heeft betrekking op de vraag naar de voorkeur voor de verschillende politieke partijen. Alleen m.b.t. de voorkeur voor de CVP (151, d.i. 34,09 %) en de BSP (110 vrouwen, d.i. 24,9 %) was het aantal voldoende om een multivariabelen analyse mogelijk te maken. Van de vraag naar de partijvoorkeur werden twee dummy-variabelen gemaakt voor respectievelijk de eerste en de tweede hypothese, namelijk de voorkeur voor de CVP met als categorieën geen voorkeur ( $C_1$ ) en wel voorkeur ( $C_2$ ) en de voorkeur voor de BSP met als categorieën geen voorkeur ( $B_1$ ) en wel voorkeur ( $B_2$ ). De vraag over de partijvoorkeur werd door 12 respondenten niet beantwoord zodat er voor de analyse 443 eenheden overbleven.

Het moet opgemerkt worden dat 107 vrouwen beweerden geen enkele voorkeur te hebben. Dit aantal is verhoudingsgewijs groter dan het aantal onthoudingen bij de feitelijke verkiezingen. Er wordt verondersteld dat deze mogelijks verdoken vorm van non-respons in verhouding voor de twee partijen ongeveer even groot is. In elk geval gaat het hier om de voorkeur en niet om feitelijk stemgedrag. Men mag wel aannemen dat de kans op het uitdrukken van de voorkeur groter is naarmate men een grotere voorkeur heeft.

### 3. Het loglineaire model.

Loglineaire modellen bieden een formeel kader voor analyse van relaties tussen dichotome en polytome variabelen die op nominaal niveau gemeten zijn waarbij interacties niet uitgesloten worden, er geen onderstellingen gemaakt moeten worden m.b.t. normaliteit en homoscedasticiteit (21).

Voor wat betreft de hier gebruikte procedure steunen we op het werk van Goodman (22), op de exegese van Goodman's papers door Davis (23) en op het werk van Upton (24). Het schatten en toetsen van de parameters in het model steunt op het « maximum likelihood-principle ».

(21) J.A.P. HAGENAARS en A.G.J.J. HEINEN, Analyse van contingentietabellen met behulp van het loglineaire model, in J.H.G. Segers en J.A.P. Hagenaars (eds.), *Sociologische onderzoeksmethoden, deel II*, Assen, 1980, blz. 186. Voor het begrip « homoscedasticiteit » zie W.L. Hays, *Statistics for the Social Sciences*, London, 1977, blz. 646.

(22) L.A. GOODMAN, The Multivariate Analysis of Qualitative Data: Interactions among Multiple Classifications, in *Journal of the American Statistical Association*, 65 (1970), blz. 225-256.

(23) J.A. DAVIS, Hierarchical models for significance tests in Multivariate Contingency Tables: An Exegesis of Goodman's recent papers, in *Sociological Methodology*, 1973-74, blz. 189-231.

(24) G.J.G. UPTON, *The Analysis of Cross-tabulated Data*, Chichester, 1978.

Dit principe van de grootste aannemelijkheid stelt dat gegeven een bepaalde waarde van de populatieparameter, men onder bepaalde veronderstellingen de kansen op diverse steekproefuitkomsten kan berekenen. Dit principe stelt voor, uitgaande van een geobserveerde steekproefuitkomst, die waarde voor de schatting van de populatiewaarde te kiezen, waarbij de kans op het gevonden steekproefresultaat het grootst is, of in geval van toetsing, die hypothese te aanvaarden die de steekproefuitkomst het meest aannemelijk maakt. Zo zullen we bijvoorbeeld zien dat in het gesatureerde model (zie verder) de geschatte verwachte frequenties  $\hat{F}_{ijkl}$  gelijk zijn aan de geobserveerde frequenties  $f_{ijkl}$ . Voor het schatten van de verwachte frequenties in niet gesatureerde modellen steunt men bovendien op het zgn. « collapsibility-theorema » (25). Een belangrijke assumptie bij het toepassen van het principe van de grootste aannemelijkheid is dat de data tot stand kwamen volgens de multinomiale verdeling (26), of in termen van steekproeftrekking dat er sprake is van enkelvoudige a-selecte steekproeftrekking uit een populatie van  $n$  klassen gevormd door combinaties van de categorieën van de discrete variabelen, ofwel van gestratificeerde a-selecte steekproeftrekking omdat men dan binnen elk stratum een multinomiale kansverdeling heeft (27). De data die voor de analyse gebruikt worden voldoen aan die voorwaarde.

Het loglineaire model is multiplicatief, d.w.z. dat de waarden van de te verklaren grootheden gezien worden als het produkt van een aantal verklarende grootheden i.p.v. hun gewogen som. De vergelijkingen van het multiplicatieve model kunnen echter omgezet worden in additieve vorm met behulp van het (natuurlijk) logaritme. Vandaar de benaming: loglineaire modellen. Welke zijn nu de te verklaren grootheden? Het gaat hier niet om individuele scores m.b.t. een afhankelijke variabele maar wel om celfrequenties of celkansen. De afhankelijke variabelen worden m.a.w. de celkansen  $P_{ijkl}$  of een functie ervan (de natuurlijke logaritme,  $G_{ijkl}$ ) voor een  $I \times J \times K \times L$  tabel. Men wil een model zoeken of toetsen dat de variatie in de celkansen verklaart (28). Om de verschillende stappen in de analyse te begrijpen is enige uitleg van drie centrale begrippen nodig, namelijk kansverhoudingen (« odds ratio's »), effecten en modellen.

(25) J.A.P. HAGENAARS en A.G.J.J. HEINEN, *op. cit.*, 187, blz. 228-229.

(26) J.A.P. HAGENAARS en A.G.J.J. HEINEN, *op. cit.*, blz. 187-188.

(27) W.L. HAYS, *op. cit.*, blz. 270-272.

(28) H.T. REYNOLDS, *Analysis of Nominal Data*, Beverley Hills, 1977, 57.



## 3.1. Kansverhoudingen.

De kansverhouding is de ratio van twee klassefrequenties van een variabele, bijvoorbeeld de verhouding tussen al of niet voorkeur voor de CVP. Neem bijvoorbeeld de marginale tabel met slechts twee van de vier variabelen, voorkeur voor de CVP en kerkelijkheid.

TABEL I  
Voorkeur voor de CVP volgens kerkelijkheid

	$C_1$	$C_2$	Totaal
$K_1$	231	30	261
$K_2$	61	121	182
Totaal	292	151	443

De (geschatte) kans op voorkeur voor de CVP is  $f_{.2}/f_{..} = 151/443 = 0,341$  (29). De kans op iemand zonder voorkeur voor de CVP is  $f_{.1}/f_{..} = 292/443 = 0,659$ . Er is dus een grote ongelijkheid van kansen. De kansverhouding pro CVP t.o.v. niet pro is  $0,341/0,659 = 0,518$  (d.i.  $f_{.2}/f_{.1} = 151/292$ ). De kans op pro CVP is m.a.w. slechts 0,518 maal de kans op niet pro CVP. Interessanter zijn de conditionele kansverhoudingen en hun verhouding. De kansverhouding wel versus geen CVP voorkeur, gegeven dat men niet kerkelijk is, vindt men als volgt :

$$\frac{f_{12}}{f_{1.}} / \frac{f_{21}}{f_{2.}} = f_{12}/f_{21} = 30/231 = 0.130$$

De kansverhouding is echter heel wat hoger bij de kerksen, namelijk

$$\frac{f_{22}}{f_{2.}} / \frac{f_{21}}{f_{2.}} = f_{22}/f_{21} = 121/61 = 1.984$$

Wanneer de conditionele kansverhoudingen voor de verschillende categorieën van een tweede variabele, bv. kerkelijkheid, identiek zouden zijn, d.w.z. wanneer de ratio van deze kansverhoudingen 1 is (of nul wanneer men de logaritme neemt) dan is er geen verband tussen de twee

(29) We gebruiken de volgende notatie :

$$f_{i.} = \sum_{j=1}^2 f_{ij} \quad f_{.j} = \sum_{i=1}^2 f_{ij}$$

$$f_{..} = \sum_{i=1}^2 \sum_{j=1}^2 f_{ij}$$

kenmerken. Dit heeft direct te maken met de begrippen « statistische onafhankelijkheid » en « associatie » (30). Wanneer twee variabelen niet onafhankelijk zijn, dan zijn de conditionele kansverhoudingen niet gelijk. Dit lijkt het geval te zijn in ons voorbeeld, al dient opgemerkt te worden dat het hier over steekproefresultaten gaat zodat de hypothese van onafhankelijkheid getoetst moet worden.

De kansverhouding van de tweede orde voor de kerksen t.o.v. de onkerksen is

$$\frac{f_{22}}{f_{21}} / \frac{f_{12}}{f_{11}} = \frac{f_{22} f_{11}}{f_{21} f_{12}} = 15.274$$

Dit wijst erop dat de verhouding wel of niet pro CVP meer dan 15 keer hoger ligt bij de kerksen dan bij de niet kerksen. Wanneer men werkt met celproporties i.p.v. celfrequenties ( $p_{ij} = f_{ij}/f_{..}$ ) dan krijgt men uiteraard dezelfde kansverhoudingen (31).

Bemerkt dat de relatie CVP voorkeur-kerkelijkheid in dit voorbeeld niet gecontroleerd is voor de overige variabelen. Dit zal wel gebeuren in de verdere analyse. Het partiële verband zal zwakker blijken.

Bij kansverhoudingen van de derde en hogere orde (d.w.z. met drie en meer variabelen) neemt het begrip interactie de plaats in van « associatie » en « onafhankelijkheid ». Kansverhoudingen van de derde orde kunnen bijvoorbeeld gelijk zijn aan 1 (of die benaderen), wanneer er bij elk paar variabelen wel degelijk associatie is. Dit betekent dan dat de associatie tussen twee variabelen gelijk is voor elke categorie van een derde variabele; er is m.a.w. geen interactie-effect (32).

Gegeven de omstandigheid dat bij de vier variabelen er twee zijn met drie categorieën zou een overzicht van alle kansverhoudingen tot en met de vierde orde te onoverzichtelijk zijn. Daarom beperken we ons ter illustratie tot een tabel met de variabelen CVP-voorkeur, kerkelijkheid en sociaal niveau (33).

(30) Zie hiervoor G.J.G. Upton, *op. cit.*, blz. 13-14.

(31) Dit is de zgn. *cross product ratio*. Uitgedrukt in celproporties i.p.v. frequenties krijgt men

(32) J.A. DAVIS, *art. cit.*, blz. 190-195.

(33) 2

C/KS betekent de kansverhouding wel CVP t.o.v. niet CVP gegeven een bepaalde categorie van kerkelijkheid en sociaal niveau.

22

CK/S betekent de kansverhouding wel-niet CVP van de kerksen t.o.v. de niet kerksen gegeven een bepaalde categorie van de derde variabele, sociaal niveau.

TABEL II

Voorkeur voor de CVP volgens kerkelijkheid en sociaal niveau

C				2	22	222	223	223
S	K	1	2	C/KS	C/KS	CKS	CKS	CKS
				1	11	111	111	112
1	1	122	18	0,148	13,514			
	2	17	34	2,000				
2	1	75	8	0,107	17,421	1,289	1,311	1,172
	2	22	41	1,864				
3	1	34	4	0,118	17,720			
	2	22	46	2,091				

We zien dat de kansverhoudingen van de derde orde 1 benaderen zodat waarschijnlijk van geen interactie sprake is. Voor het midden en hoogste sociaal niveau is de associatie tussen CVP-voorkeur en kerksheid ongeveer even groot. Telkens is de kansverhouding wel/niet CVP ongeveer 17 keer groter bij de kerksen dan bij de niet kerksen. Voor de arbeiders is die verhouding iets lager.

Tot zover een illustratie van kansverhoudingen van de derde orde. Bij het toetsen van de hypothesen zullen we te maken hebben met kansverhoudingen van de vierde orde. Wanneer deze gelijk zijn aan elkaar en hun verhouding dus gelijk aan 1 wordt, dan betekent dit dat de interacties tussen drie variabelen voor de verschillende categorieën van de vierde variabele, gelijk zijn.

### 3.2. Mogelijke effecten.

Zoals reeds gezegd wordt bij loglineaire analyse gepoogd de variatie in celfrequenties (of proporties) te verklaren. Zo kan bijvoorbeeld ter verklaring van een celfrequentie in de  $2 \times 2$  kruistabel m.b.t. CVP-voorkeur (C) en kerkelijkheid (K) het volgend multiplicatief model worden gestuleerd :

$$F_{ij} = n \tau_1^C \tau_j^K \tau_{ij}^{CK} \quad (34)$$

(34) De superscripts (C, K) verwijzen naar de variabelen, de suscripts (i, j) naar de categorieën van de overeenkomstige variabelen.

De waarde van elke celfrequentie  $F_{ij}$  wordt binnen dit model op de eerste plaats bepaald door een constante grootte  $\eta$ , een soort « overall » gemiddelde dat de weerspiegeling is van de omvang  $N$  van de steekproef. Daarnaast wordt de celfrequentie bepaald door zgn. hoofdeffecten van de marginale verdelingen van de verschillende variabelen. Wanneer er bijvoorbeeld aanzienlijk meer onkerksen zijn dan kerksen dan zullen de cellen  $F_{11}$  en  $F_{21}$  met een bepaalde factor  $\tau_1^K$  afwijken van het gemiddelde omdat ze betrekking hebben op categorie 1 (de onkerksen) van variabele  $K$  (kerkelijkheid). Op gelijkaardige wijze kan het hoofdeffect  $\tau_1^C$  omschreven worden. Tenslotte worden m.b.t. deze  $2 \times 2$  tabel de celfrequenties nog bepaald door de samenhang van de twee variabelen, met name het effect  $\tau_{ij}^{CK}$ . Associatie duidt immers op het systematisch meer en minder voorkomen van combinaties van categorieën van de twee variabelen, bv.  $C_2K_2$  (kerksen hebben relatief meer voorkeur voor de CVP) (35). Zoals gezegd kan dit multiplicatief model omgezet worden in een additief model door de logaritme van de vergelijking zodat een loglineaire vergelijking ontstaat :

$$G_{ij} = \theta + \lambda_i^C + \lambda_j^K + \lambda_{ij}^{CK} \quad (36)$$

Bij een gesatureerd model waarin alle mogelijke effecten zijn opgenomen, worden als schatters  $\hat{F}_{ij}$  op basis van het « maximum likelihood » principe de geobserveerde frequenties genomen of voor  $\hat{G}_{ij}$ , de logaritme van deze frequenties. Men krijgt dan echter een stelsel van 4 vergelijkingen met meer onbekenden dan bekenden (9 onbekenden tegenover 4 bekenden in dit voorbeeld). Oplossing is mogelijk omdat een aantal restricties aan de parameters kunnen gesteld worden.

Men is niet geïnteresseerd in het effect van bijvoorbeeld  $\lambda_1^C$  en  $\lambda_2^C$  op zich maar wél in de vraag wat het gevolg is van het feit dat de betreffende frequentie behoort tot categorie 1 van de variabele partijvoorkeur *in plaats van* tot categorie 2 van dezelfde variabele. We zijn m.a.w. geïnteresseerd in de mate waarin de effecten  $\lambda_1^C$  en  $\lambda_2^C$  van elkaar afwijken. Naar analogie met variantie-analyse kan men dit weergeven als afwijking van het gemiddelde effect. Voor de loglineaire vergelijkingen houdt dit in dat de afwijkingen, aangezien het om afwijkingen gaat van het rekenkundig gemiddelde, tot nul sommeren. Vandaar dat de volgende restricties kunnen aangebracht worden :

(35) J.A.P. HAGENAARS en A.G.J.J. HEINEN, *op. cit.*, blz. 194-195,

(36)  $G_{ij} = \ln F_{ij}$ ,  $\theta = \ln \eta$ ,  $\lambda_i^C = \ln \tau_i^C$ ,  $\lambda_j^K = \ln \tau_j^K$ ,  $\lambda_{ij}^{CK} = \ln \tau_{ij}^{CK}$

$$\sum_i \lambda^C = 0, \quad \sum_j \lambda_j^K = 0$$

Voor dichotome variabelen geldt dus  $\lambda_1^C = -\lambda_2^C$

$$\lambda_1^K = -\lambda_2^K$$

Evenzeer geldt  $\sum_i \lambda_{ij}^{CK} = 0, \quad \sum_j \lambda_{ij}^{CK} = 0$

en dus  $\lambda_{12}^{CK} = \lambda_{21}^{CK} = -\lambda_{11}^{CK} = -\lambda_{22}^{CK}$

Er blijven dan vier vergelijkingen met vier onbekenden zodat de loglineaire parameters te berekenen zijn (37).

Bij het oplossen van de vergelijkingen (zie bijlage 1) blijkt dat de parameters  $\tau_{ij}^{CK}$  en  $\lambda_{ij}^{CK}$  direct te relateren zijn aan het begrip statistische onafhankelijkheid en de relatieve kansverhoudingen (38).

Tot zover de effecten voor een  $2 \times 2$  tabel. In het geval van vier variabelen zijn nog andere effecten werkzaam naast de hoofdeffecten van C, Z, K en S en de associaties CK, CZ, CS, KZ, KS en ZS. De variatie in de celfrequenties kan ook toegeschreven worden aan de verschillen in associatie tussen twee variabelen voor de verschillende categorieën van een derde variabele, d.w.z. de interactie-effecten CZK, CZS, CKS en ZKS. Tenslotte is er een verschil in interactie-effecten tussen drie variabelen mogelijk voor de verschillende categorieën van een vierde variabele, het effect CZKS. De kansverhoudingen zijn de concrete uitdrukking van deze effecten en ze komen te voorschijn in de parameters van het model. De volgende vergelijking heeft betrekking op het gesatureerd model voor een  $2 \times 3 \times 2 \times 3$  tabel met de vier variabelen CVP-voorkeur, verzuiling, kerkelijkheid en sociaal niveau.

$$G_{ijkl} = \theta + \lambda_i^C + \lambda_j^Z + \lambda_k^K + \lambda_l^S + \lambda_{ij}^{CZ} + \lambda_{ik}^{CK} + \lambda_{il}^{CS} + \lambda_{jk}^{ZK} + \\ \lambda_{jl}^{ZS} + \lambda_{kl}^{KS} + \lambda_{ijk}^{CZK} + \lambda_{ijl}^{CZS} + \lambda_{ikl}^{CKS} + \lambda_{jkl}^{ZKS} + \lambda_{ijkl}^{CZKS}$$

waarin alle mogelijke effecten zijn opgenomen.

(37) J.A.P. HAGENAARS en A.G.J.J. HEINEN, *op. cit.*, blz. 196-198.

(38) Zie hiervoor J.A.P. HAGENAARS en A.G.J.J. HEINEN, *op. cit.*, blz. 188-192, 196-200. Voor het gesatureerd model in de  $2 \times 2$  kruistabel geldt bijvoorbeeld

$$\tau_{11}^{CK} = \left( \frac{f_{11} f_{22}}{f_{21} f_{12}} \right)^{\frac{1}{4}}$$

dit is de vierde machtswortel uit de « cross product ratio ».

Zie ook G.J.G. UPTON, *op. cit.*, blz. 49-55.

### 3.3. Modellen.

De centrale operatie in het systeem van Goodman bestaat erin dat modellen geconstrueerd en getest worden. Men zal telkens hypothetisch bepaalde effecten weglaten. Van deze modellen kunnen de verwachte frequenties berekend worden en deze worden dan vergeleken met de geobserveerde frequenties. De discrepantie tussen geobserveerde en verwachte frequenties gegeven een bepaald model, laat toe om tot de geldigheid van het model te besluiten. Bij het gesatureerde model, d.i. het model waar alle mogelijke effecten in opgenomen zijn, is de afwijking per definitie nul gezien geobserveerde en verwachte frequenties samenvallen. Het zal er dus om te doen zijn eenvoudiger modellen op te stellen waarbij verondersteld wordt dat sommige effecten 1 respectievelijk 0 zijn (in de loglineaire variant) zodat bepaalde parameters weggelaten worden. Op basis van een model wordt dus een geheel van semi-artificiële data gecreëerd die wel enkele overeenkomsten vertonen met de geobserveerde data, bijvoorbeeld dezelfde steekproefomvang  $N$  en dezelfde marginale verdelingen, doch met enkele opvallende verschillen omdat bepaalde associaties en interacties gelijk worden gesteld aan 0, d.w.z. de hypothese is dat deze geen effect hebben. Men berekent dan de kans waarmee een afwijking van waargenomen grootte (of groter) zou optreden onder de hypothese dat het model geldt voor de populatie. Indien de afwijking significant is, d.w.z. een geringe kans heeft, dan moet het model verworpen worden. Modellen echter waarbij de afwijking tussen verwachte en waargenomen frequenties dusdanig gering zijn dat de kans op optreden onder de assumptie van het model groot is, worden behouden (39). Als toetsstatistiek wordt gebruik gemaakt van de zgn. « Likelihood ratio Chisquare » (afgekort: « LR Chisq »). Deze laat partitie toe, d.w.z. dat de bijdragen tot de LR Chisq voor de verschillende effecten afzonderlijk kunnen berekend worden. Deze statistiek is voor grote  $a$ -selecte steekproeven onder assumptie van het model chiquadraat verdeeld (40).

Bij de beslissing kunnen twee bijkomende principes in acht genomen worden. Vooreerst kan gesteld worden dat de modellen hiërarchisch zijn, d.w.z. dat in het model een parameter wordt opgenomen die betrekking heeft op een verzameling variabelen  $V$ , dan moet het model alle parameters bevatten die betrekking hebben op elke deelverzameling

(39) J.A. DAVIS, *art. cit.*, blz. 194.

(40) LR Chisq =  $2 \sum_i \sum_j \sum_k \sum_l [f_{ijkl} \ln (f_{ijkl} / \hat{F}_{ijkl})]$

$\hat{F}_{ijkl}$  = geschatte verwachte frequenties.

van V. Concreet, wanneer het interactieëffect  $\lambda_{ijk}^{CZK}$  wordt opgenomen, dan eveneens de associatieeffecten  $\lambda_{ij}^{CZ}$ ,  $\lambda_{ik}^{CK}$ ,  $\lambda_{jk}^{ZK}$  en de hoofdeffecten  $\lambda_1^C$ ,  $\lambda_j^Z$  en  $\lambda_k^K$  (41).

Vervolgens is er het principe van de spaarzaamheid. Er wordt gepoogd zo een eenvoudig mogelijk model te behouden. Dit is vooral van belang wanneer uit een veelheid van « goede » modellen voor een moet gekozen worden. In ons geval ligt het accent echter niet op het zoeken naar modellen doch op het *toetsen* van twee welomschreven modellen op basis van vorig onderzoek en theorievorming.

De stappen die gezet worden zijn de volgende: 1. Veronderstel een model dat opgaat voor de geobserveerde data, d.w.z. een hypothese m.b.t. de distributies en interrelaties tussen de variabelen in een multidimensionale kruistabel. 2. Formuleer een aantal verwachtingen onder de assumptie dat het model juist is, d.w.z. hoe zouden de data er uit zien indien dit model juist is. Aangezien we enkel de steekproefgegevens kennen moeten deze verwachtingen geschat worden op basis van de geobserveerde data. Voor deze schattingen zijn computerprogramma's beschikbaar. Er wordt gebruik gemaakt van het programma BMDP 3F van het computerpakket « *Biomedical Computer Package* » (42). 3. Vergelijk de verwachte frequenties met de geobserveerde data en beslis of het model aanvaardbaar is. Indien de geobserveerde kruisclassificatie werkelijk uit een populatie afkomstig is waarbinnen het model geldt, dan zijn de discrepanties enkel afkomstig van steekproeffouten. De vraag is dan of afwijkingen tussen geobserveerde en verwachte waarden redelijk aan het toeval te wijten zijn dan wel of ze zo groot zijn dat het model als zodanig verworpen moet worden. 4. Indien de discrepanties voldoende klein zijn, wordt het model behouden, zoniet moet een ander model voorgesteld worden en moet men terug naar de eerste stap. 5. Als het model behouden wordt, dan moeten de parameters geschat worden (43).

We zullen nu overgaan tot het doorlopen van deze stappen voor de twee hypothesen. Er wordt niet uitvoerig ingegaan op de bepaling van de vrijheidsgraden (44), tenzij terloops, en op het schatten van de te verwachten frequenties (45). De hierboven geschetste logica volstaat om de toetsings- en beslissingsprocedure te begrijpen.

(41) G.J.G. UPTON, *op. cit.*, blz. 59.

(42) BMDP. *Biomedical Computer Package*, Berkeley, 1977, blz. 297-325.

(43) H.T. REYNOLDS, *op. cit.*, blz. 57-58.

(44) *Ibid.*, blz. 78.

(45) Zie hiervoor G.J.G. UPTON, *op. cit.*, blz. 46-69. De numerieke techniek voor het schatten van de frequenties in de zgn. « iterative proportional fitting ».

#### 4. Het toetsen van de hypothesen.

Voortgaand op de in de eerste paragraaf geformuleerde theoretische hypothesen kunnen we deze formuleren in operationele termen. De data vindt men in tabel III en IV.

*Hypothese 1* (CVP-voorkeur) : het te toetsen model is :

$$CZ, CK, ZK, ZS, KS$$

Of in termen van de verwachte logfrequenties en loglineaire parameters wegens het hiërarchisch principe

$$G_{ijkl} = \theta + \lambda_i^C + \lambda_j^Z + \lambda_k^K + \lambda_l^S + \lambda_{ij}^{CZ} \\ + \lambda_{ik}^{CK} + \lambda_{jk}^{ZK} + \lambda_{jl}^{ZS} + \lambda_{kl}^{KS}$$

Opvallend hierbij is dat niet alleen alle interacties en hogere effecten gelijk aan nul gesteld worden maar evenzeer de associatie tussen voorkeur voor de CVP en sociaal niveau (CS).

*Hypothese 2* (BSP-voorkeur) : het te toetsen model is :

$$BZ, BK, BS, ZK, ZS, KS$$

In termen van de loglineaire parameters

$$G_{ijkl} = \theta + \lambda_i^B + \lambda_j^Z + \lambda_k^K + \lambda_l^S + \lambda_{ij}^{BZ} \\ + \lambda_{ik}^{BK} + \lambda_{il}^{BS} + \lambda_{jk}^{ZK} + \lambda_{jl}^{ZS} + \lambda_{kl}^{KS}$$

Met als opvallend verschil t.o.v. de eerste hypothese dat de associatie BSP-sociaal niveau (BS) *niet* gelijk gesteld wordt aan nul.

##### 4.1. Het gesatureerd model als gids (zie bijlage 2).

De computer analyse neemt het gesatureerd model als gids om een overzicht te geven van de verschillende effecten (46). Het gesatureerd model bevat alle mogelijke associaties en interacties en dus zoveel onafhankelijke parameters als er cellen zijn. Er kunnen niet meer parameters aan toegevoegd worden en het gesatureerd model fit steeds de geobserveerde data in die zin dat zoals reeds gezegd, de verwachte geschatte frequenties onder dit model steeds gelijk zijn aan de geobserveerde frequenties.

Met dit model als gids wordt nagegaan welke effecten mogelijk aan nul kunnen gelijkgesteld worden (zie bijlage 2). Zo zou bijvoorbeeld m.b.t. de CVP-voorkeur het weglaten van alle associaties en interacties een « Likelihood ratio Chisquare » opleveren van 374,31. De kans op zo'n afwijking gegeven 29 vrijheidsgraden is quasi gelijk aan nul zodat

(46) G.J.G. UPTON, *op. cit.*, blz. 64.



de hypothese dat alle tweede-orde en hogere effecten nul zijn, moet verworpen worden. Voor de tabel m.b.t. de voorkeur voor de BSP geldt hetzelfde, ook daar mogen alle tweede-orde en hogere relaties niet weggelaten worden (LR Chisq = 348,72,  $p = 0,0$ ). Echter, in beide gevallen lijken de interacties (derde-orde relaties en hoger) wel gelijkgesteld te mogen worden aan nul. (Voor de CVP-tabel : LR Chisq : 14,21,  $p = 0,58$  ; voor de BSP-tabel : LR Chisq : 9,73,  $p = 0,88$ .)

Het is dus aannemelijk dat minstens sommige derde-orde relaties en ook het vier-variabeleneffect gelijk zijn aan nul. Eventueel is het mogelijk dat ook enkele associaties (tweede orde), nul mogen zijn. Een meer gedetailleerd beeld krijgt men in de volgende test waar effect per effect wordt getest (eveneens bijlage 2). Daar ziet men m.b.t. de voorkeur voor de CVP dat het weglaten van het effect CS (partijkeuze — sociaal niveau) een zo kleine afwijking teweegbrengt, namelijk slechts 0,48 voor 2 vrijheidsgraden, dat deze nulhypothese zeker niet kan verworpen worden ( $p = 0,785$ ). Het is m.a.w. zeer aannemelijk dat de relatie CS geen effect heeft op de variaties in de celfrequenties. Bij de voorkeur voor de BSP echter mag de relatie BS geenszins weggelaten worden want dit brengt een grote afwijking teweeg, nl. LR Chisq = 29,15 met 2 vrijheidsgraden. De kans dat BS nul is in de populatie is zeer gering ( $p = 0,0$ ) en deze hypothese moet dus verworpen worden. Ook blijkt dat het weglaten van het interactie effect BZS lang niet zo aannemelijk is dan het weglaten van de overige effecten (LR Chisq = 4,90,  $df = 4$ ,  $p = 0,298$ ) alhoewel de hypothese dat BZS nul is strikt genomen, niet kan verworpen worden op significantieniveau ( $\alpha = 0,01$ ).

Bemerk echter dat de LR Chisq-waarden berekend worden op basis van het gesatureerd model waarbij telkens elementen aan nul gelijk gesteld worden. Daardoor krijgt men slechts een eerste aanwijzing van de parameters die men eventueel kan weglaten. Uiteraard zullen de waarden van de overgebleven parameters en de Chi Square-bijdragen in het te toetsen model wijzigen t.o.v. het gesatureerde model.

#### 4.2. Toetsen van de hypothesen.

Er wordt nu overgegaan tot het toetsen van de twee specifieke modellen.

BMDP3F geeft eerst informatie over de waarschijnlijkheid van het getoetste model. Vervolgens worden de onder het model berekende verwachte frequenties gegeven en de gestandaardiseerde residuen, d.w.z. de verschillen tussen elke geobserveerde en verwachte frequentie, gedeeld door de vierkantswortel uit de verwachte frequenties.

## 4.2.1. Toetsing van hypothese 1.

Model 1	D.F.	LR Chisq	Prob	iteraties
CZ, CK, ZK, SZ, KS	18	14,69	0,683	7

Dit model kan niet verworpen worden, het is zelfs zeer aannemelijk ( $p = 0,68$ ).

TABEL III

Geobserveerde frequenties ( $f_{ijkl}$ ), de verwachte frequenties ( $\hat{F}_{ijkl}$ ) en gestandaardiseerde residuen onder het geteste model

C	Z	K	S	$f_{ijkl}$	$\hat{F}_{ijkl}$	Resi- duelen	C	Z	K	S	$f_{ijkl}$	$\hat{F}_{ijkl}$	Resi- duelen
1	1	1	1	32	33.27	-0.134	2	1	1	1	1	1.25	0.228
1	2	1	1	77	72.84	0.546	2	2	1	1	10	11.50	-0.294
1	3	1	1	13	16.35	-0.704	2	3	1	1	7	7.80	-0.109
1	1	2	1	1	1.17	0.309	2	1	2	1	0	0.32	0.311
1	2	2	1	11	9.07	0.806	2	2	2	1	4	10.60	-1.874
1	3	2	1	5	7.24	-0.647	2	3	2	1	30	25.60	0.970
1	1	1	2	38	37.73	0.126	2	1	1	2	1	1.41	0.074
1	2	1	2	30	32.56	-0.361	2	2	1	2	4	5.14	-0.282
1	3	1	2	7	6.20	0.521	2	3	1	2	3	2.96	0.313
1	1	2	2	3	3.80	-0.156	2	1	2	2	0	1.05	-0.540
1	2	2	2	11	11.67	-0.049	2	2	2	2	16	13.63	0.776
1	3	2	2	8	7.91	0.212	2	3	2	2	25	27.94	-0.461
1	1	1	3	18	19.59	-0.247	2	1	1	3	1	0.73	0.895
1	2	1	3	12	13.79	-0.347	2	2	1	3	2	2.18	0.219
1	3	1	3	4	3.19	0.736	2	3	1	3	1	1.52	-0.017
1	1	2	3	5	4.43	0.507	2	1	2	3	0	1.23	-0.657
1	2	2	3	7	11.08	-1.076	2	2	2	3	17	12.95	-1.264
1	3	2	3	10	9.08	0.461	2	3	2	3	29	32.20	-0.475

De overeenstemming tussen de geobserveerde en de verwachte frequenties is over het algemeen goed met uitzondering van een drietal celfrequenties, namelijk  $f_{1223}$ ,  $f_{2221}$  en  $f_{2223}$ .

Om meer zekerheid te hebben over het behouden model is het goed dit te vergelijken met de meest aanverwante modellen. BMDP3F geeft een test van de aanverwante modellen, d.w.z. modellen die inbegrepen zijn in het bovenstaande model maar die één effect minder bevatten en modellen die bovenstaand model bevatten maar die één effect méér bevatten.

Dit is vooral van belang voor een exploratie naar het best passende model, doch in ons geval is dit eveneens interessant omdat men dan

een indruk krijgt van de effecten die eventueel nog mogen opgenomen of weggelaten worden.

Volgende modellen zijn begrepen in het getoetste model en hebben één effect minder.

TABEL III A

<i>Model</i>		<i>Effect</i>	<i>DF</i>	<i>LR Chlsq</i>	<i>Prob.</i>	<i>Iteraties</i>
CK, ZK, ZS, KS	verschil	CZ	20	55,75	0,000	7
	wegens	CZ	2	41,05	0,000	
CZ, ZK, ZS, KS	verschil	CK	19	80,90	0,000	7
	wegens	CK	1	66,21	0,000	
CZ, CK, ZS, KS	verschil	ZK	20	61,85	0,000	4
	wegens	ZK	2	47,15	0,000	
CZ, CK, ZK, KS	verschil	ZS	22	32,62	0,067	5
	wegens	ZS	4	17,93	0,001	
CZ, CK, ZK, ZS	verschil	KS	20	58,49	0,000	5
	wegens	KS	2	43,80	0,000	

Telkens wordt het betreffende effect gelijkgesteld aan nul en krijgt men de test van het nieuwe model. Geen van deze modellen is aanneemelijk (kleine kansen) omdat de afwijking te sterk toeneemt, bijvoorbeeld voor het weglaten van het effect CZ met een winst van 2 vrijheidsgraden stijgt de « Likelihood ratio Chisquare » met 41,05 en zo'n afwijking is heel onwaarschijnlijk zodat dit model moet verworpen worden.

Volgende modellen bevatten het getoetste model en ze hebben één effect meer.

TABEL III B

<i>Model</i>		<i>Effect</i>	<i>DF</i>	<i>LR Chlsq</i>	<i>Prob.</i>	<i>Iteraties</i>
CS, CZ, CK, ZK, ZS, KS	verschil	CS	16	14,21	0,583	8
	wegens	CS	2	0,48	0,785	
CZK, ZS, KS	verschil	CZK	16	13,46	0,639	7
	wegens	CZK	2	1,24	0,538	
ZKS, CZ, CK	verschil	ZKS	14	11,32	0,661	6
	wegens	ZKS	4	3,37	0,498	

Alhoewel deze modellen aanvaardbaar zijn blijft het geteste model de voorkeur genieten omwille van de grotere eenvoudigheid, en tevens omdat de hypothese dat CS, CZK en ZKS nul zijn niet kunnen verworpen worden. Geen van deze modellen lijkt trouwens een betere fit te geven dan het geteste.

## 4.2.2. Toetsing van hypothese 2.

Model 2	D.F.	LR Chisq	Prob.	Iteraties
BZ, BK, BS, ZK, ZS, KS	16	9,73	0,88	10

Ook dit model kan behouden worden. De aannemelijkheid is nog groter dan bij hypothese 1 ( $p = 0,8805$ ).

TABEL IV

Geobserveerde frequenties ( $f_{ijkl}$ ), de verwachte frequenties ( $\hat{F}_{ijkl}$ ) en de gestandaardiseerde residuen onder het geteste model

B	Z	K	S	$f_{ijkl}$	$\hat{F}_{ijkl}$	Resi- duelen	B	Z	K	S	$f_{ijkl}$	$\hat{F}_{ijkl}$	Resi- duelen
1	1	1	1	8	5.61	1.220	2	1	1	1	25	29.01	-0.651
1	2	1	1	46	48.55	-0.294	2	2	1	1	41	36.76	0.782
1	3	1	1	19	20.39	-0.196	2	3	1	1	1	2.71	-0.735
1	1	2	1	0	0.56	-0.082	2	1	2	1	1	0.85	0.701
1	2	2	1	13	15.29	-0.458	2	2	2	1	2	3.40	-0.489
1	3	2	1	34	32.61	0.331	2	3	2	1	1	1.27	0.200
1	1	1	2	17	17.97	-0.111	2	1	1	2	22	21.89	0.131
1	2	1	2	30	31.38	-0.157	2	2	1	2	4	5.60	-0.464
1	3	1	2	10	8.88	0.546	2	3	1	2	0	0.28	0.421
1	1	2	2	3	4.50	-0.473	2	1	2	2	2	1.61	0.700
1	2	2	2	27	24.74	0.555	2	2	2	2	0	1.30	-0.700
1	3	2	2	33	35.53	-0.341	2	3	2	2	0	0.33	0.302
1	1	1	3	8	9.33	-0.273	2	1	1	3	10	9.17	0.439
1	2	1	3	14	13.75	0.203	2	2	1	3	0	1.98	-1.051
1	3	1	3	4	4.64	-0.066	2	3	1	3	0	0.12	1.117
1	1	2	3	4	5.03	-0.235	2	1	2	3	1	1.45	0.041
1	2	2	3	24	23.30	0.249	2	2	2	3	0	0.99	-0.489
1	3	2	3	39	39.95	-0.072	2	3	2	3	0	0.30	0.373

De afwijkingen zijn ook hier zeer gering.

Terug worden de aanverwante modellen bekeken. Vooreerst wordt telkens één effect weggelaten (zie blz. 555).

Ook geen van deze modellen brengt een verbetering. Vervolgens wordt telkens één effect bijgevoegd.

Hier ligt het moeilijker. Twee nieuwe modellen geven een minstens even goede fit. Het lijkt zelfs aannemelijk voor het effect BZS, d.w.z. een interactie tussen de voorkeur voor de BSP en verzuiling binnen de categorieën van het sociaal niveau. De associatie BZ wijzigt m.a.w. naar-

TABEL IV A

Model		Effect	DF	LR Chlsq	Prob.	Iteraties
ZS, KS, BZ, BK, BS	verschil	ZK	18	80,46	0,000	9
	wegens	ZK	2	70,73	0,000	
ZK, KS, BZ, BK, BS	verschil	ZS	20	47,13	0,000	5
	wegens	ZS	4	37,41	0,000	
ZK, ZS, BZ, BK, BS	verschil	KS	18	42,14	0,001	9
	wegens	KS	2	32,42	0,000	
ZK, ZS, KS, BK, BS	verschil	BZ	18	86,40	0,000	6
	wegens	BZ	2	76,67	0,000	
ZK, ZS, KS, BZ, BS	verschil	BK	17	22,45	0,168	10
	wegens	BK	1	12,72	0,000	
ZK, ZS, KS, BZ, BK	verschil	BS	18	38,88	0,003	7
	wegens	BS	2	29,15	0,000	

TABEL IV B

Model		Effect	DF	LR Chlsq	Prob.	Iteraties
BZK, ZS, KS, BS	verschil	BZK	14	7,25	0,929	9
	wegens	BZK	2	2,48	0,289	
BZS, ZK, KS, BK	verschil	BZS	12	3,60	0,989	5
	wegens	BZS	4	6,12	0,190	
BKS, ZK, ZS, BZ	verschil	BKS	14	9,67	0,786	10
	wegens	BKS	2	0,05	0,973	
ZKS, BZ, BK, BS	verschil	ZKS	12	7,85	0,797	9
	wegens	ZKS	4	1,88	0,758	

gelang van het sociaal niveau (of evenzeer, de associatie BS wijzigt naargelang van de verzuiling vermits de relaties symmetrisch zijn). Strikt genomen mag men het effect BZS gelijk aan nul stellen wanneer men als significantiedrempel  $\alpha = 0,01$  zou nemen doch dit is niet zo heel aannemelijk ( $p = 0,19$ ). In elk geval, ook wanneer dit effect opgenomen wordt, dan blijft het zo dat in tegenstelling tot de voorkeur voor de CVP, bij de voorkeur voor de BSP het sociaal niveau invloed heeft. Voorlopig houden we het bij hypothese 2.

#### 4.4. Beschrijving van enkele effecten.

Rest nu nog de beschrijving van de voornaamste effecten aan de hand van de loglineaire parameters. Voor de berekening van deze parameters verwijzen we naar bijlage 3.

Zoals reeds meegedeeld bevatten de getoeste modellen de parameters zoals zij vermeld staan op blz. 556 bovenaan.

In het tweede model is er een parameter meer dan in het eerste, namelijk  $\lambda_{ij}^{BS}$ .

We hebben twee stelsels met elk 36 vergelijkingen (nl.  $I \times J \times K \times L = 2 \times 3 \times 2 \times 3 = 36$ ) doch er zijn meer onbekenden, met name 42

$$\text{model 1 : } \hat{G}_{ijkl} = \theta + \lambda_1^C + \lambda_j^Z + \lambda_k^K + \lambda_1^S + \lambda_{ij}^{CZ} + \lambda_{ik}^{CK} + \lambda_{jk}^{ZK} + \lambda_{j1}^{ZS} + \lambda_{k1}^{KS}$$

$$\text{model 2 : } \hat{G}_{ijkl} = \theta + \lambda_1^B + \lambda_j^Z + \lambda_k^K + \lambda_1^S + \lambda_{ij}^{BZ} + \lambda_{ik}^{BK} + \lambda_{11}^{BS} + \lambda_{jk}^{ZK} + \lambda_{j1}^{ZS} + \lambda_{k1}^{KS}$$

voor het eerste en 48 voor het tweede model. Zoals bij het gesatureerde model gelden ook hier een aantal restricties (zie bijlage 3).

Hierdoor wordt het aantal zelf te schatten parameters teruggebracht tot respectievelijk 18 en 20. Het aantal vrijheidsgraden is respectievelijk 18 en 16, dit is het aantal geobserveerde frequenties (cellen) min het aantal zelf te schatten parameters. Bij niet gesatureerde modellen zijn er dus vrijheidsgraden over om de houdbaarheid van deze modellen na te gaan, wat niet het geval is bij gesatureerde modellen (47).

De loglineaire parameters geven een idee van de sterkte van de verschillende effecten. Wanneer men de antilogaritme neemt van deze parameters wordt het belang van de effecten uitgedrukt in termen van kansverhoudingen.

Om te besluiten bekijken we even de voornaamste parameters die ons interesseren, met name de parameters  $\lambda_{ij}^{CZ}$ ,  $\lambda_{ik}^{CK}$ ,  $\lambda_{ij}^{BZ}$ ,  $\lambda_{ik}^{BK}$  en  $\lambda_{11}^{BS}$  horend bij de opgenomen associaties waarin telkens de partijkzeuge een der variabelen is.

Bekijken we eerst de parameters m.b.t. de eerste hypothese voor zover deze significant zijn. Bemerkt dat de multiplicatieve parameters  $\tau$  gevonden worden door de antilogaritme van de  $\lambda$  parameters, zo is (bijlage 3, tabel V) :

$$\lambda_{11}^{CZ} = 0.664 \quad \text{en} \quad \tau_{11}^{CZ} = e^{0.664} = 1.943$$

$$\lambda_{11}^{CZ} = 0.664 \quad \tau_{11}^{CZ} = 1.943, \quad \lambda_{21}^{CZ} = -0.664 \quad \tau_{21}^{CZ} = 0.5148$$

$$\lambda_{13}^{CZ} = -0.604 \quad \tau_{13}^{CZ} = 0.544, \quad \lambda_{23}^{CZ} = 0.609 \quad \tau_{23}^{CZ} = 1.839$$

$$\lambda_{11}^{CK} = \lambda_{22}^{CK} = 0.500 \quad \tau_{11}^{CK} = \tau_{22}^{CK} = 1.649$$

$$\lambda_{12}^{CK} = \lambda_{21}^{CK} = -0.500 \quad \tau_{12}^{CK} = \tau_{21}^{CK} = 0.607$$

Opgemerkt moet worden dat het hier telkens om partiële relaties of gemiddelde verhoudingen gaat. De twee overige variabelen worden telkens constant gehouden zodat men de gemiddelde effecten heeft. We zullen ons beperken tot de significante parameters. Nochtans geeft het feit dat

(47) J.A.P. HAGENAARS en A.G.J.J. HEINEN, *op.cit.*, blz. 219 ; H.T. REYNOLDS, *op. cit.*, blz. 78.

de parameters m.b.t. de niet verzuilden (categorie 2 van Z) niet significant zijn ook reeds informatie (bijlage 3, tabel V en VI), met name dat de hypothese dat de kansverhoudingen al of niet CVP-voorkeur, en al of niet voorkeur voor de BSP weinig verschillen voor de niet-verzuilden. De hypothese dat die verhouding 1 is kan niet verworpen worden. Bemerkt dat bij de « niet verzuilden » een twintigtal leden van liberale organisaties zitten. Wat de partiële samenhang tussen kerkelijkheid en partijvoorkeur betreft kan gesteld worden dat over alle categorieën van de overige variabele (verzuiling en sociaal niveau) de kerksen gemiddeld veel frequenter een voorkeur voor de CVP hebben dan de niet kerksen ( $\tau_{22}^{CK} = 1,649$ ). Gemiddeld komt de combinatie kerks-CVP 2,7 keer zo frequent voor als de combinatie onkerks-CVP (namelijk  $1,649/0,607 = 2,717$ ). Wanneer er variabelen zijn met meer dan twee categorieën is de interpretatie van de partiële samenhang moeilijker. Wat de samenhang tussen verzuiling en CVP-voorkeur betreft, stellen we bijvoorbeeld vast dat de combinatie katholiek verzuild — CVP-voorkeur gemiddeld 3,6 keer (d.i.  $1,839/0,5148$ ) zo frequent voorkomt als de combinatie socialistisch verzuild — CVP-voorkeur.

De significante parameters bij de tweede hypothese zijn de volgende (bijlage 3, tabel VI) :

$$\lambda_{11}^{BZ} = -0.930 \quad \tau_{11}^{BZ} = 0.395, \quad \lambda_{21}^{BZ} = 0.930 \quad \tau_{21}^{BZ} = 2.535$$

$$\lambda_{13}^{BZ} = 0.900 \quad \tau_{13}^{BZ} = 2.459 \quad \lambda_{23}^{BZ} = -0.900 \quad \tau_{23}^{BZ} = 0.407$$

$$\lambda_{11}^{BK} = \lambda_{22}^{BK} = -0.306 \quad \tau_{11}^{BK} = \tau_{22}^{BK} = 0.736$$

$$\lambda_{12}^{BK} = \lambda_{21}^{BK} = 0.306 \quad \tau_{12}^{BK} = \tau_{21}^{BK} = 1.358$$

$$\lambda_{11}^{BS} = -0.518 \quad \tau_{11}^{BS} = 0.596, \quad \lambda_{21}^{BS} = 0.518 \quad \tau_{21}^{BS} = 1.679$$

$$\lambda_{13}^{BS} = 0.312 \quad \tau_{13}^{BS} = 1.367, \quad \lambda_{23}^{BS} = -0.312 \quad \tau_{23}^{BS} = 0.732$$

Zowel voor de middencategorie van de variabele verzuiling als de variabele sociaal niveau zijn de parameters niet significant. Wat de partiële associatie kerkelijkheid en voorkeur voor de BSP betreft is deze in omgekeerde richting als bij de voorkeur voor de CVP. Onkerksen zijn gemiddeld frequenter pro BSP ( $\tau_{21}^{BK} = 1,358$ ). Gemiddeld komt de combinatie onkerks — BSP-voorkeur 1,8 keer zo frequent voor als de combinatie kerks — BSP-voorkeur (namelijk  $1,358/0,736$ ). Echter in vergelijking met de voorkeur voor de CVP is de relatie tussen partijvoorkeur en kerkelijkheid zwakker. De variabelen sociaal niveau en verzuiling laten weer een moeilijker interpretatie toe wat de

samenhang betreft. De combinatie socialistisch verzuild — BSP-voorkeur komt gemiddeld bijvoorbeeld 6,2 keer (d.i.  $2,535/0,407$ ) zo frequent voor dan de combinatie katholiek verzuild — BSP-voorkeur. De combinatie laag sociaal niveau — BSP-voorkeur komt gemiddeld 2,3 keer (d.i.  $1,679/0,732$ ) frequenter voor dan de combinatie hoog sociaal niveau — BSP-voorkeur (48).

### Besluit.

Op basis van de bestaande literatuur over de geïnstitutionaliseerde tegenstellingen in de Belgische politiek en voortgaande op onderzoeksresultaten m.b.t. factoren die de keuze en/of de voorkeur voor een politieke partij mede verklaren, werden twee hypothesen geformuleerd. Deze hypothesen hebben betrekking op de onderlinge samenhang tussen de individuele kenmerken partijvoorkeur, kerkelijkheid, sociaal niveau en mate van verzuildheid. Dit laatste kenmerk sluit aan bij de zgn. organisatorische variabele die door Hill als belangrijk werd voorgesteld maar die voordien nog niet in een analyse was opgenomen bij gebrek aan gegevens daarover. De gegevens in dit artikel zijn afkomstig uit een onderzoek over de schoolkeuze (Mechelen, 1974). De twee hypothesen hebben de toetsing doorstaan. Bij de voorkeur voor de CVP zijn kerkelijkheid en verzuiling van belang. De voorkeur voor de BSP daarentegen wordt bovendien nog beïnvloed door het sociaal niveau. We krijgen ook een bevestiging van de hypothese van K. Hill m.b.t. de organisatorische variabele. Het is zelfs zo dat lidmaatschap van de katholieke of de socialistische zuil de grootste verschillen in kansverhoudingen realiseert. Dit sluit aan bij de hypothese m.b.t. de « long lasting loyalty » die de verzuiling verschafft voor de twee politieke partijen, CVP en BSP (49).

Het was vooral de bedoeling om aan de hand van een vertrouwd onderzoeksontwerp, de bruikbaarheid van het toetsen van loglineaire modellen voor nominale kenmerken aan te tonen en een voorbeeld te geven van interpretatie van de analyse-resultaten. Als multivariate techniek voor categorische variabelen kan de loglineaire analyse een vergelijking met de variantie-analyse en multiple regressie voor metrische kenmerken doorstaan. Allerhande effecten zoals associaties in interacties kunnen getest worden en bovendien krijgt men een indruk van de relatieve sterkte

---

(48) Cfr. J.A.P. HAGENAAERS en A.G.J.J. HEINEN, *op. cit.*, blz. 201, 208-210 ; G.J.G. UPTON, *op. cit.*, blz. 83-84.

(49) R. STEININGER, Pillarization (verzuiling) and Political Parties, in *Sociologische Gids*, (1977) 4, blz. 249-250.



van alle effecten. Er zijn echter ook enkele nadelen. Vooreerst is de techniek, de interpretatie en de rapportering omslachtig. Eens men over de computer-output beschikt begint het werk pas. Door de omstandigheid dat men met celfrequenties werkt is het aantal variabelen beperkt tenzij men over zeer grote steekproeven beschikt. De parameters zijn moeilijk te interpreteren, zeker bij polytome variabelen waar per effect meerdere waarden bestaan. Men heeft het nadeel dat de grootte van indirecte effecten niet kan vastgesteld worden. Een moeilijkheid blijft steeds dat men met het oog op overzichtelijkheid en wegens een geringe steekproefomvang zoveel mogelijk categoriën samentrekt. Wanneer dit arbitrair gebeurt kan dit tot valse conclusies leiden (50).

---

(50) J.A.P. HAGENAARS en A.G.J.J. HEINEN, *op. cit.*, blz. 251-254; H.T. REYNOLDS, *op. cit.*, blz. 79.

## BIJLAGE 1

Oplossing van de loglineaire vergelijkingen van het gesatureerd model voor een  $2 \times 2$  tabel (tabel I).

De vergelijking  $G_{ij} = \theta + \lambda_i^C + \lambda_j^K + \lambda_{ij}^{CK}$  wordt uitgeschreven als volgt :

$$g_{11} = \theta + \lambda_1^C + \lambda_1^K + \lambda_{11}^{CK}$$

$$g_{12} = \theta + \lambda_1^C + \lambda_2^K + \lambda_{12}^{CK}$$

$$g_{21} = \theta + \lambda_2^C + \lambda_1^K + \lambda_{21}^{CK}$$

$$g_{22} = \theta + \lambda_2^C + \lambda_2^K + \lambda_{22}^{CK}$$

zoals gezegd zijn er meer onbekenden dan vergelijkingen, nl. 9 tegenover 4 (de logaritme van de 4 celfrequenties).

Wegens de restricties

$$\sum_i \lambda_i^C = \sum_j \lambda_j^K = \sum_i \lambda_{ij}^{CK} = \sum_j \lambda_{ij}^{CK} = \theta$$

behouden we slechts 4 onbekenden en kunnen de vergelijkingen als volgt herschreven worden.

$$g_{11} = \theta + \lambda_1^C + \lambda_1^K + \lambda_{11}^{CK}$$

$$g_{12} = \theta + \lambda_1^C - \lambda_1^K - \lambda_{11}^{CK}$$

$$g_{21} = \theta + \lambda_1^C + \lambda_1^K - \lambda_{11}^{CK}$$

$$g_{22} = \theta - \lambda_1^C - \lambda_1^K + \lambda_{11}^{CK}$$

$$g_{11} + g_{12} + g_{21} + g_{22} = 4 \cdot \theta \quad \rightarrow \quad \theta = \frac{g_{11} + g_{12} + g_{21} + g_{22}}{4}$$

$$g_{11} + g_{12} - g_{21} - g_{22} = 4 \lambda_1^C \quad \rightarrow \quad \lambda_1^C = \frac{g_{11} + g_{12} - g_{21} - g_{22}}{4}$$

$$g_{11} + g_{21} - g_{12} - g_{22} = 4 \lambda_1^K \quad \rightarrow \quad \lambda_1^K = \frac{g_{11} + g_{21} - g_{12} - g_{22}}{4}$$

$$g_{11} + g_{22} - g_{12} - g_{21} = 4 \lambda_{11}^{CK} + \lambda_{11}^{CK} = \frac{g_{11} + g_{22} - g_{12} - g_{21}}{4}$$

In termen van de oorspronkelijke frequenties kunnen de parameters als volgt herschreven worden.

$$\text{Vermits } \theta = \ln \eta, \lambda_1^C = \ln \tau_1^C, \lambda_1^K = \ln \tau_1^K, \lambda_{1j}^{CK} = \ln \tau_{1j}^{CK}$$

$$\text{geldt : } \eta = (f_{11} f_{12} f_{21} f_{22})^{\frac{1}{4}} = (231 \times 30 \times 61 \times 121)^{\frac{1}{4}} = 84.569$$

$$\tau_1^C = \left( \frac{f_{11} f_{12}}{f_{12} f_{22}} \right)^{\frac{1}{4}} = \left( \frac{231 \times 30}{61 \times 121} \right)^{\frac{1}{4}} = 0.9844$$

$$\tau_1^K = \left( \frac{f_{11} f_{21}}{f_{12} f_{22}} \right)^{\frac{1}{4}} = \left( \frac{231 \times 61}{30 \times 121} \right)^{\frac{1}{4}} = 1.4036$$

$$\tau_{11}^{CK} = \tau_{22}^{CK} = \left( \frac{f_{11} f_{22}}{f_{12} f_{21}} \right)^{\frac{1}{4}} = \left( \frac{231 \times 121}{61 \times 30} \right)^{\frac{1}{4}} = 1.9769$$

Wegens de restricties geldt

$$\tau_2^C = \frac{1}{\tau_1^C} = 1.0158 \quad \tau_2^K = \frac{1}{\tau_1^K} = 0.713$$

$$\tau_{12}^{CK} = \tau_{21}^{CK} = \frac{1}{\tau_{11}^{CK}} = \frac{1}{\tau_{22}^{CK}} = 0.5058$$

Bemerk dat het « overall gemiddelde » ongeveer 84 eenheden bedraagt.  $(\tau_{11}^{CK})^4$  is de zgn. « cross product ratio », de kansverhouding van de tweede orde, immers  $(1,9769)^4 = 15.274$  (vergelijk met blz. 544).

Men kan stellen dat de frequenties  $f_{11}$  en  $f_{22}$  gemiddeld ongeveer tweemaal zo groot zijn als men op grond van het « overall » gemiddelde zou verwachten ( $\tau_{11}^{CK} = 1,9769$ ).

De frequenties  $f_{21}$  en  $f_{12}$  zijn gemiddeld ongeveer de helft zo groot als men zou verwachten ( $\tau_{12}^{CK} = 0,5058$ ).

## BIJLAGE 2

## Overzicht van de voorafgaandelijke tests in het BMDP 3F-programma.

1. *Hypothese 1. CVP-voorkeur.*

Resultaat bij het fitten van alle k-factor marginalen. Dit is een simultane test dat alle  $k + 1$  en hogere factor associaties en/of interacties gelijk zijn aan nul.

K-factor	DF	LR Chsq	Prob.	Iteraties
0	35	488,95	0,000	
1	29	374,31	0,000	2
2	16	14,21	0,583	8
3	4	2,77	0,597	4

Vervolgens een simultane test dat alle k-factor interacties gelijk zijn aan nul. Men krijgt de verschillen met bovenstaande tabel.

K-factor	DF	LR Chsq	Prob.
1	6	114,64	0,000
2	13	360,10	0,000
3	12	11,44	0,492
4	4	2,77	0,597

Een test van de partiële associaties van de factoren. Dit wordt berekend als een verschil tussen het volledige model van de vierde orde en dat model dat enkel het gespecificeerde effect uitsluit.

Een test van marginale associatie van de factoren. De tabel wordt gesommeerd over de niet gespecificeerde indices en dan wordt het effect getest als zou het nul zijn.

Effect	DF	LR Chsq	Prob.	Iteraties	LR Chsq	Prob.	Iteraties
C	1	43,82	0,000				
Z	2	33,50	0,000				
K	1	13,60	0,000				
S	2	23,71	0,000				
CZ	2	41,20	0,000	8	118,14	0,000	2
CK	1	56,11	0,000	8	143,29	0,000	2
CS	2	0,48	0,785	7	11,70	0,003	2
ZK	2	47,58	0,000	6	118,47	0,000	2
ZS	4	18,07	0,001	7	13,00	0,011	2
KS	2	33,70	0,000	5	38,87	0,000	2
CZK	2	1,34	0,511	5	1,24	0,538	5
CZS	4	6,69	0,153	4	10,02	0,040	5
CKS	2	0,61	0,737	5	0,10	0,951	7
ZKS	4	1,09	0,895	5	3,37	0,498	7
CZKS	4	2,77	0,597	4			

## 2. Hypothese 2. Voorkeur voor de BSP.

Resultaat bij het fitten van alle k-factor marginalen. Dit is een simultane test dat alle  $k + 1$  en hogere factor associaties gelijk zijn aan nul.

<i>K-factor</i>	<i>DF</i>	<i>LR Chisq</i>	<i>Prob.</i>	<i>Iteraties</i>
0	35	531,34	0,000	
1	29	348,72	0,000	2
2	16	9,73	0,881	10
3	4	0,91	0,923	4

Vervolgens een simultane test dat alle k-factor interacties gelijk zijn aan nul. Men krijgt de verschillen met bovenstaande tabel.

<i>K-factor</i>	<i>DF</i>	<i>LR Chisq</i>	<i>Prob.</i>
1	6	182,63	0,000
2	13	338,99	0,000
3	12	8,81	0,719
4	4	0,91	0,923

Een test van de partiële associaties van de factoren. Dit wordt berekend als een verschil tussen het volledige model van de vierde orde en dat model dat enkel het gespecificeerde effect uitsluit.

Een test van marginale associatie van de factoren. De tabel wordt gesommeerd over de niet gespecificeerde indices en dan wordt het effect getest als zou het nul zijn.

<i>Effect</i>	<i>DF</i>	<i>LR Chisq</i>	<i>Prob.</i>	<i>Iteraties</i>	<i>LR Chisq</i>	<i>Prob.</i>	<i>Iteraties</i>
B	1	112,53	0,000				
Z	2	32,85	0,000				
K	1	12,26	0,001				
S	2	25,00	0,000				
BZ	2	76,67	0,000	7	109,85	0,000	2
BK	1	12,72	0,000	10	78,41	0,000	2
BS	2	29,15	0,000	7	27,23	0,000	2
ZK	2	70,73	0,000	9	114,41	0,000	2
ZS	4	37,41	0,000	5	13,48	0,009	2
KS	2	32,42	0,000	9	41,00	0,000	2
BZK	2	1,42	0,492	5	2,86	0,239	5
BZS	4	4,90	0,298	8	6,87	0,143	9
BKS	2	0,48	0,789	4	0,16	0,922	5
ZKS	4	0,95	0,917	6	2,64	0,619	7
BZKS	4	0,91	0,923	4			

## BIJLAGE 3

## De loglineaire parameters bij de twee geteste modellen.

De restricties bij model 1 :

$$\sum_i \lambda_i^C = \sum_j \lambda_j^Z = \sum_k \lambda_k^K = \sum_l \lambda_l^S = 0$$

$$\sum_i \lambda_{ij}^{CZ} = \sum_j \lambda_{ij}^{CZ} = 0 \quad \sum_j \lambda_{jk}^{ZK} = \sum_k \lambda_{jk}^{ZK} = 0$$

$$\sum_i \lambda_{ik}^{CK} = \sum_k \lambda_{ik}^{CK} = 0 \quad \sum_j \lambda_{jl}^{ZS} = \sum_l \lambda_{jl}^{ZS} = 0$$

$$\sum_k \lambda_{kl}^{KS} = \sum_l \lambda_{kl}^{KS} = 0$$

Voor de hoofdeffecten zijn er dus slechts 6 onafhankelijke parameters in plaats van 10, namelijk  $(I-1) + (J-1) + (K-1) + (L-1)$ .

Voor de opgenomen associaties heeft men 11 zelf te schatten parameters, namelijk  $(I-1)(J-1) + (I-1)(K-1) + (J-1)(K-1) + (J-1)(L-1) + (K-1)(L-1)$ . Tenslotte is er een zelf te schatten parameter voor  $\theta$ , het "overall gemiddelde".

De restricties bij model 2 :

$$\sum_i \lambda_i^C = \sum_j \lambda_j^Z = \sum_k \lambda_k^K = \sum_l \lambda_l^S = 0$$

$$\sum_i \lambda_{ij}^{BZ} = \sum_j \lambda_{ij}^{BZ} = 0 \quad \sum_j \lambda_{jk}^{ZK} = \sum_k \lambda_{jk}^{ZK} = 0$$

$$\sum_i \lambda_{ik}^{BK} = \sum_k \lambda_{ik}^{BK} = 0 \quad \sum_l \lambda_{il}^{BS} = \sum_l \lambda_{il}^{BS} = 0$$

$$\sum_j \lambda_{jl}^{ZS} = \sum_l \lambda_{jl}^{ZS} = 0 \quad \sum_k \lambda_{kl}^{KS} = \sum_l \lambda_{kl}^{KS} = 0$$

Dit brengt het aantal zelf te schatten parameters terug op 20, namelijk 6 voor de hoofdeffecten, 13 voor de associaties (BS komt erbij, met 2 zelf te schatten parameters) en 1 parameter voor  $\theta$ .

De gestandaardiseerde loglineaire parameters verkrijgt men door de loglineaire parameters te delen door de standaardfout. Men kan een schatting maken van de standaardfout vermits de steekproefgegevens tot stand zijn gekomen volgens een (product)multinomiale verdeling en mits deze

voldoende groot is. De gestandaardiseerde waarde volgt asymptotisch een standaardnormale verdeling onder de assumptie dat voor de populatie geldt dat de betreffende parameter nul is, dus onder de nulhypothese. Men kan dus deze hypothese toetsen.

Bij een verwerpingsgebied van  $\alpha = 0,05$  en tweezijdige toetsing zijn de kritische waarden dus  $+1,96$  en  $-1,96$ . Gestandaardiseerde  $\lambda$ 's die in absolute waarde groter zijn wijzen daarin op significante effecten. Wanneer men echter binnen eenzelfde model uitspraken doet over bijvoorbeeld tien parameters ( $\lambda_{11}^{CZ}$ ,  $\lambda_{12}^{CZ}$  enz...) en indien de toetsen onafhankelijk van elkaar zouden zijn, dan is de kans dat bij tenminste één van de tien toetsen de nulhypothese ten onrechte wordt verworpen geen 5 % maar wel 40 % [namelijk  $1 - (0,95)^{10}$ ]. De situatie is echter ingewikkelder want de toetsen zijn niet onafhankelijk van elkaar. Bij onverzadigde modellen komen er nog complicaties bij het schatten van de standaardfouten. Volgens Hagenaars en Heinen worden de standaardfouten uit het gesatureerd model doorgaans beschouwd als de bovengrenzen van de standaardfouten en het onverzadigde model (51). Het toetsen op significantie van effecten gebeurt trouwens niet door middel van de afzonderlijke loglineaire parameters maar door gebruik te maken van de « *likelihood ratio chisquare* » zoals in de vorige bladzijden gebeurde.

---

(51) J.A.P. HAGENAARS en A.G.J.J. HEINEN, *op. cit.*, blz. 213-214, 219; G.J.G. UPTON, *op. cit.*, blz. 62-64.

TABEL V  
De loglineaire parameters voor model 1

Zelf te schatten parameters		Gestandaardiseerd	Overige parameters		Gestandaardiseerd
$\theta$	1.888				
$\lambda_{11}^C$	0,478	5,047	$\lambda_{22}^C$	-0,478	-5,047
$\lambda_{11}^Z$	-0,848	-4,726			
$\lambda_{22}^Z$	0,600	5,279	$\lambda_{33}^Z$	0,249	2,094
$\lambda_{11}^K$	0,054	0,543	$\lambda_{22}^K$	-0,054	-0,543
$\lambda_{11}^S$	0,122	1,141			
$\lambda_{22}^S$	0,101	1,039	$\lambda_{33}^S$	-0,222	-2,237
$\lambda_{11}^{OZ}$	0,664	3,853	$\lambda_{21}^{OZ}$	-0,664	-3,853
$\lambda_{12}^{OZ}$	-0,055	-0,515	$\lambda_{22}^{OZ}$	0,055	0,515
			$\lambda_{13}^{OZ}$	-0,609	-5,435
			$\lambda_{23}^{OZ}$	0,609	5,435
$\lambda_{11}^{OK}$	0,500	5,285	$\lambda_{13}^{OK}$	-0,500	-5,285
			$\lambda_{21}^{OK}$	-0,500	-5,285
			$\lambda_{22}^{OK}$	0,500	5,285
$\lambda_{11}^{ZK}$	0,634	3,532	$\lambda_{12}^{ZK}$	-0,634	-3,532
$\lambda_{21}^{ZK}$	0,000	0,001	$\lambda_{22}^{ZK}$	-0,000	-0,001
			$\lambda_{31}^{ZK}$	-0,634	-5,345
			$\lambda_{32}^{ZK}$	0,634	5,345
$\lambda_{11}^{ZS}$	-0,474	-2,514			
$\lambda_{12}^{ZS}$	0,201	1,237	$\lambda_{13}^{ZS}$	0,273	1,714
$\lambda_{21}^{ZS}$	0,215	1,712	$\lambda_{23}^{ZS}$	-0,174	-1,412
$\lambda_{22}^{ZS}$	-0,041	-0,354	$\lambda_{31}^{ZS}$	0,259	1,986
			$\lambda_{32}^{ZS}$	-0,160	-1,255
			$\lambda_{33}^{ZS}$	-0,099	-0,725
$\lambda_{11}^{KS}$	0,487	4,563	$\lambda_{21}^{KS}$	-0,487	-4,563
$\lambda_{12}^{KS}$	-0,041	-0,429	$\lambda_{22}^{KS}$	0,041	0,429
			$\lambda_{13}^{KS}$	-0,445	-4,483
			$\lambda_{23}^{KS}$	0,445	4,483



TABEL VI  
De loglineaire parameters voor model 2

Zelf te schatten parameters		Gestandaardiseerd	Overige parameters		Gestandaardiseerd
$\theta$	1,617				
$\lambda_{11}^B$	0,933	7,255	$\lambda_{22}^B$	-0,933	-7,255
$\lambda_{12}^Z$	-0,055	-0,331	$\lambda_{33}^Z$	-0,537	-2,715
$\lambda_{21}^Z$	0,592	3,282	$\lambda_{31}^K$	-0,371	-2,883
$\lambda_{11}^K$	0,371	2,883	$\lambda_{32}^S$	-0,389	-1,985
$\lambda_{12}^S$	0,344	2,072	$\lambda_{21}^{BZ}$	0,930	5,616
$\lambda_{21}^S$	0,045	0,248	$\lambda_{22}^{BZ}$	-0,030	-0,167
$\lambda_{11}^{BZ}$	-0,930	-5,616	$\lambda_{13}^{BZ}$	0,900	4,547
$\lambda_{12}^{BZ}$	0,030	0,167	$\lambda_{23}^{BZ}$	-0,900	-4,547
			$\lambda_{12}^{BK}$	0,306	2,382
$\lambda_{11}^{BK}$	-0,306	-2,382	$\lambda_{21}^{BK}$	-0,306	-2,382
			$\lambda_{22}^{BK}$	0,306	2,382
$\lambda_{11}^{BS}$	-0,518	-3,119	$\lambda_{21}^{BS}$	0,518	3,119
$\lambda_{12}^{BS}$	0,205	1,126	$\lambda_{22}^{BS}$	-0,205	-1,126
			$\lambda_{13}^{BS}$	0,312	1,594
$\lambda_{11}^{ZK}$	0,653	3,942	$\lambda_{23}^{BS}$	-0,312	-1,594
$\lambda_{21}^{ZK}$	0,080	0,442	$\lambda_{12}^{ZK}$	-0,653	-3,942
			$\lambda_{22}^{ZK}$	-0,080	-0,442
$\lambda_{11}^{ZS}$	-0,817	-3,392	$\lambda_{31}^{ZK}$	-0,733	-3,703
$\lambda_{12}^{ZS}$	0,382	1,713	$\lambda_{32}^{ZK}$	0,733	3,702
$\lambda_{21}^{ZS}$	0,306	1,411	$\lambda_{13}^{ZS}$	0,436	1,825
$\lambda_{22}^{ZS}$	-0,095	-0,373	$\lambda_{23}^{ZS}$	-0,211	-0,733
			$\lambda_{31}^{ZS}$	0,511	2,085
$\lambda_{11}^{KS}$	0,433	2,612	$\lambda_{32}^{ZS}$	-0,286	-0,935
$\lambda_{12}^{KS}$	-0,025	-0,140	$\lambda_{33}^{ZS}$	-0,225	-0,748
			$\lambda_{21}^{KS}$	-0,433	-2,612
			$\lambda_{22}^{KS}$	0,025	0,140
			$\lambda_{13}^{KS}$	-0,408	-2,031
			$\lambda_{23}^{KS}$	0,408	2,081

**Summary : Political party preference in Belgium : Testing log-linear models.**

*Previous research has shown that there is a strong relationship between church involvement and preference for the Belgian Christian Democratic Party (CVP/PSC). When religious preference is controlled, the relationship between social class and preference for the CVP/PSC seems to disappear completely. Indeed, this party is very heterogeneous as regards the social origins of its adherents. The situation is completely different for preference for the Socialist Party (BSP/PSB) which has a strong relationship between party preference, on the one hand, church involvement and social class, on the other. The unchurched and the workers manifest more preference for the BSP/PSB than do other people. According to Hill the degree of involvement in the « zuil » (pillar organizations), the so-called « organizational variable », is also significant in the explanation of voting behavior. This could not be confirmed, however, due to a lack of data.*

*Two hypotheses are formulated : 1) pillar involvement in addition to social class and church involvement is significant in the preference for the BSP ; 2) there is no connection between preference for the CVP and social class so that only church and pillar involvement remain. These two models were tested by means of loglinear analysis and seem to fit the data.*

*The intention of this article, however, is primarily didactic : to show the possibilities and limitations of the analysis of contingency tables by means of Goodmans loglinear analysis procedure through the use of a simple and relevant example.*

