26 DE STEEKPROEFOPHOGING BIJ EEN GEDETAILLEERDE INTERAKTIEMATRIX: ENKELE PROBLEMEN EN MOGELIJKE OPLOSSINGEN

P.H.L. Bovy

Instituut voor Stedebouwkundig Onderzoek, Delft

Samenvatting

De steekproefophoging bij een gedetailleerde interaktiematrix: enkele problemen en mogelijke oplossingen.

Een kenmerk van het steeproefresultaat van enigermate fijne interaktiematrix (van b.v. vervoersrelaties, woon-werk-relaties, verhuisbewegingen, etc) is het zeer frekwente voorkomen van zeer kleine aantallen: bij het overgrote deel der relaties bedraagt het steekproefresultaat 0, 1 of 2. Zelfs bij omvangrijke steekproeven blijft dit verschijnsel bestaan. Bij de gebruikelijke ophoging van steekproeven leidt dit tot populatieschattingen met typische kenmerken. De paper beschrijft de nadelen van de gebruikelijke ophoging bij een dergelijke steekproefmatrix en geeft enkele mogelijke andere benaderingen hiervoor aan.

Summary

Sample expansion in the case of a detailed interaction matrix: some problems and possible solutions.

The sample data of a detailed interaction matrix (of for example trip or migration patterns) is characterized by high frequencies of small numbers: a large proportion of the interrelations has a sampled interaction frequency of 0, 1 or 2. Even with large samples such figures will occur. Applying standard sample expansion procedures will lead in this case to population estimates having typical and undesired characteristics. The paper deals with the drawback of these procedures and suggests some alternative procedures.

INHOUDSOPGAVE

1. INLEIDING

In vele rapporten waarin verslag wordt gedaan van onderzoek naar ruimtelijke interaktiepatronen, zoals ritten, woon-werk-relaties, verhuisbewegingen, etc. wordt vaak de *interaktiematrix waarmee is gewerkt*, als bladvullende afbeelding bijgevoegd. De matrix bevat meestal, zoals bij elk vervoersonderzoek, de opgehoogde steekproefaantallen. Men kan hierbij vaak de volgende opvallende waarneming doen: zeer veel cellen zijn leeg d.w.z.: nul, vaak meer dan de helft, en de frequentieverdeling van de aantallen *is zeer onregelmatig* en veeltoppig zij het dat de frequentie pieken juist voorkomen bij die aantallen die al gauw als veelvouden van de ophoogfaktor worden herkend.

i \ j	1	2	3	4	5					n
1	36	0	0	0	25	0	0	0		
2	0	13	11	13	24	0	11	0		
3	0	0	12	15	12	12	0	0		
4	0	0	0	11	0	0	12	0		
	11	0	0	0	0	0	0	12		
	0	12	0	0	11	0	0	0		
n										

Wat is hiervoor de verklaring?

In zo een geval blijkt de steekproefmatrix voor een zeer groot gedeelte te bestaan uit zeer kleine aantallen: het merendeel der relaties bedraagt 0, 1, 2, of iets meer. De ophoging is een rechttoe-rechtaan vermenigvuldiging met de voor alle zones vrijwel gelijke steekproeffaktor. Het aantal situaties waar men met dergelijke observaties van het interaktiepatroon werkt is legio. Zeker nu het steeds meer tot de goede toon lijkt te behoren om de interakties ruimtelijk en ook anderszins, zeer gedetailleerd te inventariseren, en te onderzoeken.

Zo een situatie is niet zonder *problemen:* trekt men uit een zeer gedetailleerd *interaktiepatroon een steekproef*, en *hoogt* de aantallen hieruit op de rechttoe-rechtaan manier op, dan leidt dit tot geheel verkeerde schattingen van de kleine relaties (ruwweg tussen 0 en 10 ritten). Het aantal nul-relaties wordt bv. zwaar overschat. Daarenboven vormen deze kleine relaties in veel gevallen niet alleen het merendeel van al-

le relaties, maar ze omvatten tezamen vaak wel bijna de helft van alle interakties. Het verrichten van verder onderzoek, zoals het schatten van interaktie - en andere modellen, op basis van dergelijke data zal dan ook zonder twijfel de ongewenste sporen hiervan dragen.
In het kader van een onderzoek naar de effekten van het ruimtelijk detailniveau bij vervoersplanologisch onderzoek, zie hierover de bijdragen op beide voorgaande Colloquia [1, 2], worden wij op zeer indringende wijze met dit probleem gekonfronteerd. Zo zal er onder meer een vervoersberekening worden uitgevoerd (voor Eindhoven) met een gebiedsindeling van meer dan 1000 zones (bouwblokken) voor de ca. 20.000 autoritten in de avondspits. De verwachting is dat de ware rittenmatrix (de populatie) voor 90% uit nullen zal bestaan en dat de overige relaties alle zeer klein (< 10) zullen zijn. We staan voor de taak deze rittenmatrix te schatten op basis van een 20% steekproef (de zg. VODAR-enquête). De vraag is: wat voor een schattingsmethode kun je het beste volgen in zo een situatie?
De inhoud van deze paper is gericht op deze ophoog problematiek bij kleine aantallen. We zullen eerst enige empirische gegevens presenteren met betrekking tot dergelijke situaties en zullen trachten aan te geven waarmee het optreden van deze situatie zoal kan samenhangen.
De aard van de problemen die erdoor ontstaan zal wat verder worden uitgesponnen.

Na deze beschrijving van het probleem zal een poging worden ondernomen tot "oplossing". Verschillende alternatieve benaderingswijzen voor zo een schattingsprobleem zullen worden geschetst en ter diskussie gesteld.

2. OVERZICHT VAN HET OPHOOG VRAAGSTUK BIJ KLEINE AANTALLEN

2.1. Algemeen

We gaan in het vervolg uit van de volgende veronderstellingen c.q. beperkingen:
- het interaktiepatroon in het onderzoekgebied wordt gemeten via een ruimtelijk gestratificeerde steekproef uit de huishoudens;

- alleen ophoging vanwege de steekproef sec wordt beschouwd, en niet verdere "ophogingen" via bv. passeerlijn- of kordongegevens.
- de beschouwing is geheel toegespitst op ritten. Ze is echter even goed van toepassing op andere vormen van interaktie (migratie, woon-werk-relaties, etc.), en zelfs op elk nominaal verdeeld verschijnsel.

2.2. Empirische gegevens

a. Een goede introduktie in de hier ter diskussie staande ophoogproblematiek kan men krijgen door een aantal rittenmatrices nader te analyseren. Een willekeurige greep uit een aantal binnen- en buitenlandse onderzoekingen levert het volgende beeld op. Het wemelt van de nullen, en de kleine cijfers (1, 2 en 3) zoekt men veelal tevergeefs. De frequentieverdeling van de ritaantallen per relatie (zone-paar) zijn diskontinue, veeltoppig, en vertonen uitermate hoge percentages nullen (zie fig. 1).
De meest opvallende *diskontinuïteit* is de "gap" direkt na ritaantal "nul". De omvang van dit gebied varieert nogal.
Ofschoon er wèl veel nul-relaties zijn, lijkt het erop alsof er geen enkele zone-relatie met 1, of 2, of 3, etc. ritten zou voorkomen! De toppen en dalen van de verdelingen zijn alleen in het gebied van de kleine aantallen duidelijk aanwezig; met toenemende ritaantallen "verwatert" deze topvorming snel. De dalen lijken op duidelijke scheidingen te wijzen, die de verdeling in dat gebied in mootjes opsplitsen. De toppen blijken veelal een veelvoud te zijn van de gemiddelde ophoogfaktor die in het betreffende onderzoek is gehanteerd. Het duidelijkst is dat in fig. 1 te zien in het geval van Utrecht, waar een 100% enquête met een gemiddelde responsie van ca. 47% leidde tot een gemiddelde ophoogfaktor van 2,1.
Hoe is dit typische verloop te verklaren?
Het veeltoppige verloop duidt erop dat we te maken hebben met een samengestelde verdeling. In feite is de frekwentieverdeling van de opgehoogde aantallen opgebouwd uit een stel afzonderlijke frekwentieverdelingen, namelijk verdelingen van veelvouden van

de ophoogfaktor(en). De spreiding van deze verdelingen is te verklaren door verschil in ophoogfaktor tussen de zones (o.a. als gevolg van verschil in non-response, en soms steekproefpercentage. Ook afrondingen spelen een rol).

Elk van deze afzonderlijke verdelingen komt nu tot stand doordat één en hetzelfde steekproefaantal steeds wordt vermenigvuldigd met andere ophoogfaktoren, al naar gelang de relatie waarbij dit ene aantal is waargenomen. In beginsel korrespondeert met elk steekproef aantal, te beginnen met nul, een verdeling van met dit aantal vermenigvuldigde ophoogfaktoren. En al deze verdelingen tezamen leiden tot de frekwentieverdeling van opgehoogde ritaantallen.

Dit verklaart zowel de topvorming aan het begin als ook de "verwatering", n.l. door vereffening van elkaar overlappende ophoogfaktorverdelingen. Het verklaart ook waarom de kleine aantallen niet voorkomen. Dit mechanisme is aan de hand van een typisch geval geïllustreerd in fig. 2.

In sommige gevallen, zoals in dat van Utrecht, is het dan ook niet moeilijk om uit een gegeven matrix van opgehoogde ritaantallen de frekwentieverdeling van de steekproefaantallen, althans in het bereik van ca. < 5 ritten, vrij redelijk te rekonstrueren. Op deze manier zou men wat aan de weet kunnen komen over het voorkomen van kleine ritaantallen in de steekproef, al is het niet de meest voor de hand liggende methode.

b. Hoewel de *steekproefresultaten* van rittenonderzoek vrijwel nooit in de verslaggeving worden aangetroffen, kan men in gevallen van de rechttoe-rechtaan ophoging dus toch wel tot op zekere hoogte, via de frekwentieverdeling van de opgehoogde aantallen en de gemiddelde ophoogfaktor een indruk krijgen van de frekwenties van de kleine ritaantallen (⩽ 5) in de steekproef. Wij hebben in een aantal hiervoor geschikte gevallen eens een dergelijke rekonstruktie toegepast (ook vanwege gebrek aan het oorspronkelijke materiaal) om te kunnen illustreren hoe de kleine aantallen in de steekproef bij de benadering zijn verdeeld en welke fraktie zij op het totaal vormen.

Tabel 1a: Enkele statistische gegevens over het voorkomen van kleine aantallen in herkomst-bestemmings-onderzoek.

	studie-gebied	jaar onderzoek	aantal inwoners	aantal zones	steekproef % (incl. non-resp)	vervoer-middel	ritmotief	periode	totaal aantal ritten (opgehoogd)	percentage relaties in de steekproef met ritaantal 0	1	2	⩽5	% ritten in rel.met⩽5 stk.pr.ritn.
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Apeldoorn	1965	90.000	27	20	motorv.	alle	7-10	10.125	38	20	11	90	57
2	"					fiets	alle	7-10	21.154	24	15	13	70	22
3	"					alle	alle	7-10	32.856	16	13	9	58	14
4	"					motorv.	alle	16-19	15.905	38	17	7	82	42
5	"					fiets	alle	16-19	36.622	18	15	9	59	14
6	"					alle	alle	16-19	54.541	15	4	6	40	6
7	Hilversum	1965	107.000	27	20	motorv.	alle	7-10	11.526	27	22	13	85	57
8	"					fiets	alle	7-10	31.562	14	14	9	55	14
9	"					alle	alle	7-10	44.627	7	11	8	47	11
10	"					motorv.	alle	16-19	14.971	24	22	14	85	58
11	"					fiets	alle	16-19	38.290	10	12	8	51	14
12	"					alle	alle	16-19	55.192	4	8	6	37	9
13	Utrecht	1962	260.000	38	47	bus	alle	0-24	55.015	14	11	6	47	4
14	Leicester	1963	452.000	149	8	alle	alle	8-9	158.000	76	12	4	96	63
15	Waco(Texas)	1965	132.000	206	12,5	motorv.	alle	0-24	217.615	79	9	5	98	±50

In tabel 1a zijn deze bevindingen opgenomen. Uit de frekwentieverdelingen blijkt (kolom 11-14) dat de nul-klasse er niet meer zo scherp uitsteekt als bij de opgehoogde verdeling. In sommige gevallen is het ook niet meer de hoogste frekwentie. Alle relaties < 5 ritten tezamen vormen in alle gevallen een achtenswaardig percentage van het totaal aantal relaties, variërend van ca. 40-95% (kolom 14). Verder vertegenwoordigt deze groep soms wel 50% van alle ritten (kolom 15). Kortom, deze cijfers zijn voldoende aanleiding om nader op dit verschijnsel in te gaan.

c. Het voorgaande geeft nog geen afdoend inzicht in de feitelijke situaties waarin de ophoogproblematiek speelt. Alle cijfers zijn immers verkregen via steekproeven. Het is niet zonder meer mogelijk uit dergelijke gegevens iets af te leiden over de verdeling van de kleine aantallen in de *populatie*. Nochtans, bij gebrek aan beter, lijkt mij het percentage nul-relaties in de steekproef een bruikbare indikator van de orde van grootte van de kleine aantallen ($\leqslant$ 5) in de populatie. Tabel 1a laat zien (kolom 11) dat dit kan variëren van 10 tot 90%.
Maar eigenlijk zou men dit probleem willen bestuderen op basis van een 100% steekproef, m.a.w. met kennis van de hele populatie van ritten. Er zijn twee van dergelijke onderzoekingen bekend. Thomas [4] heeft zicht verdiept in het verschijnsel van de nulrelaties aan de hand van 100% onderzoek in een tweetal Duitse steden. Hij gaat na hoe het percentage nulrelaties varieert met de steekproefomvang. Zijn bevindingen staan vermeld in tabel 1b. Er blijkt vooral uit dat wanneer de populaties klein zijn (1- of 4-uurs perioden) en het aantal zones groot het aantal nulrelaties in de populatie al erg groot is.
Beperkt Thomas zich tot de nullen, het andere onderzoek besteedt juist veel aandacht aan de kleine relaties van 0-10 ritten. Hiertoe is in 1974 in San Antonio, Texas, een zeer beperkt 100% onderzoek gehouden, t.w. in een drietal zones van een onderzoekgebied van ca. 1560 zones [3].

Tabel 1b: Percentage nulrelaties in populatie en steekproef in twee Duitse steden [4]

	aantal inwoners	aantal zones	periode onderzoek	populatie 100%	percentage nulrelaties in steekproef van				
					30%	20%	10%	5%	3%
Stadt A	240.000	35	12-13	*32*	52	62	73	93	96
			10-14	*12*	29	38	51	65	71
			17-18	*21*	37	45	60	74	80
			6-24	*1*	5	9	18	29	40
Stadt B	80.000	48	7- 8	*61*	80	85	91	96	
			6-10	*26*	47	58	69	80	
			17-18	*56*	78	84	89	95	
			15-19	*21*	43	52	65	79	

In deze 3 kleine aan elkaar grenzende zones zijn alle huishoudens geënquêteerd. Alle ritten gemaakt door huishoudens van deze zones naar en van alle overige zones zijn bekend.
In tabel 2 is de relatieve frekwentieverdeling van de ritaantallen weergegeven, gespecificeerd naar zone en motief. Tabel 3 bevat apart het bereik van 0 tot en met 10 ritten, voor de "home based person trips" en de kombinatie van de 3 zones.
De vergelijkbaarheid met de eerder geïnventariseerde studies (tabel 1) is helaas beperkt: het gaat hier om 24-uurs home-

Ritmotief (huisgebonden) en produktiezone	Aantal ritten	Aantal relaties met ritt	aantal ritten per relatie																		
			0	1 t/m 10			11 t/m 20			21 t/m 50			51 t/m 100			101 t/m 150			151 t/m 182		
			a	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. personenritten																					
Zone A	860	167	89,4	9,8	92,2	56,4	0,3	3,0	7,6	0,3	3,0	15,5	0,2	1,8	20,6	0	0	0	0	0	0
Zone B	1180	234	85,1	13,7	91,9	53,2	0,5	3,0	9,4	0,6	3,9	22,0	0,2	1,3	15,4	0	0	0	0	0	0
Zone C	1180	219	86,0	12,6	90,0	48,6	0,6	4,6	11,4	0,6	4,1	24,9	0,2	1,4	15,1	0	0	0	0	0	0
Kombinatiezone	3220	399	74,5	22,1	86,7	37,9	1,7	6,8	11,5	0,8	3,3	13,1	0,6	2,3	19,6	0,1	0,5	7,2	0,1	0,5	10,7
2. autoritten (3+4)																					
Zone A	563	159	89,9	9,6	94,3	66,6	0,3	2,5	10,3	0,3	3,1	23,1	0	0	0	0	0	0	0	0	0
Zone B	764	212	86,5	12,8	94,3	67,0	0,4	2,8	11,9	0,4	2,8	21,1	0	0	0	0	0	0	0	0	0
Zone C	763	205	86,9	12,1	92,7	59,1	0,6	4,9	20,3	0,3	2,4	20,6	0	0	0	0	0	0	0	0	0
Kombinatiezone	2090	372	76,3	21,6	90,9	48,0	0,8	3,5	8,9	1,0	4,0	21,6	0,3	1,3	12,4	0,1	0,3	9,1	0	0	0
3. woon-werk autoritten																					
Zone A	173	70	93,5	4,4	98,6	87,9	0	0	0	0,1	1,4	12,1	0	0	0	0	0	0	0	0	0
Zone B	261	109	93,1	6,9	99,1	94,6	0,1	0,9	5,4	0	0	0	0	0	0	0	0	0	0	0	0
Zone C	247	99	93,7	6,3	98,0	86,6	0	0	0	0,1	1,0	13,4	0	0	0	0	0	0	0	0	0
Kombinatiezone	681	216	86,2	13,5	98,1	84,4	0,2	1,4	5,6	0	0	0	0,1	0,5	10,0	0	0	0	0	0	0
4. overige autoritten																					
Zone A	390	107	93,2	6,4	93,5	62,1	0,2	2,8	11,0	0 ,3	3,7	26,9	0	0	0	0	0	0	0	0	0
Zone B	503	133	91,5	7,8	92,5	57,7	0,3	3,8	15,3	0,3	3,8	27,0	0	0	0	0	0	0	0	0	0
Zone C	516	142	91,0	8,1	91,6	56,4	0,6	6,3	25,3	0,2	2,1	18,4	0	0	0	0	0	0	0	0	0
Kombinatiezone	1409	253	83,9	14,5	89,7	40,7	0,6	3,6	10,1	0,8	5,1	27,5	0,2	1,2	13,8	0,1	0,4	7,8	0	0	0

Tabel 2: Frekwentieverdelingen van de ritaantallen per relatie in
100% onderzoek San Antonio, Texas (600.000 inw., 1569 relaties)

a = % relaties in aangegeven ritaantalgroep

b = % relaties > 0 in aangegeven ritaantalgroep

c = % ritten in aangegeven ritaantalgroep

Zone A = 96 huishoudens (324 pers.)
Zone B = 164 huishoudens (490 pers.)
Zone C = 164 huishoudens (463 pers.)
Kombinatie = 424 huishoudens (1277 pers.)

<u>Tabel 3</u>: Frekwentieverdeling van de kleine ritaantallen in San Antonio, Texas (100% enquête)

rit-aantal	aantal abs.	relaties %	aantal abs.	ritten %
1	2	3	4	5
0	1168	74,5	-	-
1	61	3,9	61	1,9
2	122	7,8	244	7,6
3	29	1,8	87	2,7
4	39	2,5	156	4,8
5	20	1,3	100	3,1
6	25	1,6	150	4,7
7	15	1,0	105	3,3
8	14	0,9	112	3,5
9	10	0,6	90	2,8
10	12	0,8	120	3,7
>10	52	3,3	1995	62,0
Σ	1567	100,0	3220	100,0

based ritten, zeer kleine zones (alleen de kombinatie van de drie zones heeft een enigszins vergelijkbare gebiedsgrootte) en slechts onvolledig geïnventariseerd interzonale relaties. Zelfs indien men hiermee rekening houdt is het duidelijk dat het aandeel nulrelaties en kleine ritaantallen aan het totaal aantal relaties en totaal aantal ritten zeer hoog is.

d. Veel bruikbaarder voor ons doel zijn de *steekproefsimulaties* die met dit materiaal zijn uitgevoerd. Bij diverse steekproef-frakties zijn een groot aantal random steekproeven getrokken uit de populatie van huishoudens. Op grond hiervan zijn de ritaantallen van de 1567 relaties geschat. Een overzicht hiervan, beperkt tot de kleine relaties, geven de tabellen 4A en 4B. De over 1000 steekproeven gemiddelde frekwentieverdelingen zijn hier weergegeven. Het gaat dus in wezen om frekwentieverdelingen van verwachtingswaarden.
De tabellen demonstreren overduidelijk hoe de steekproefaantallen en de schattingen tot stand komen, en hoe ze in relatie staan tot de werkelijke aantallen. De vergelijking steekproef versus populatie laat het volgende zien:
- het aantal nulrelaties wordt systematisch overschat;
- ook de overige relaties die een veelvoud van de ophoogfaktor als ritaantal hebben worden overschat;
- de kleine aantallen, voor zover geen veelvoud van de ophoogfaktor worden zwaar onderschat (omdat het in dit onderzoek slechts om één zone met slechts één ophoogfaktor gaat, komen zelfs geen andere waarden voor dan veelvouden van de ophoogfaktor).

De verklaring hiervoor is eenvoudig. Nulrelaties in de populatie blijven nulrelaties in de steekproef èn na ophoging. Echter, trekt men uit een populatie van huishoudens een steekproef van bv. 1 op 10, teneinde een relatie van bv. 1 rit te detekteren, dan is de kans deze rit in de steekproef aan te treffen 1/10. In 90% van de gevallen (1-relaties) zal een nul resulteren in de steekproef en dus ook in de schatting. In 10% van de 1-relaties wordt inderdaad de ene rit in de steekproef aangetroffen. In de opgehoogde matrix verschijnen deze echter als 10

Tabel 4 : Frekwentieverdelingen van schattingen van kleine ritaantallen bij steekproefsimulaties (1000 steekproeven). Bron: San Antonio, Texas.

Tabel 4A : 5% steekproef (ophoogfaktor = 20,2)

werkel. ritaant.	aant. rel.	Opgehoogd ritaantal resp. steekproefritaantal (..). 0 (0)	20 (1)	40 (2)	61 (3)	81 (4)	101 (5)	121 (6)	141 (7)	162 (8)	182 (9)	202 (10)	gemidd. ritaant. schatt.
0	1168	100											0
1	61	95,3	4,7										0,94
2	122	94,9	0,4	4,7									1,96
3	29	92,7	2,5	2,8	2,0								2,85
4	39	91,4	1,2	5,3	0,2	1,9							4,02
5	20	89,7	3,6	3,2	0,4	1,9	1,2						4,99
6	25	88,9	2,2	4,6	0,4	1,9	0,1	1,9					6,40
7	15	86,4	2,3	7,0	0,5	1,5	1,0	0,8	0,5				7,43
8	14	85,2	4,5	5,9	1,6	1,7	0,0	0,9	0,1	1,0			7,97
9	10	81,7	5,6	5,4	1,1	3,4	0,1	1,8	0,7	0,2	0,0		10,30
10	12	82,8	2,5	9,1	1,2	2,6	0,7	0,8	0,0	0,0	0,0	0,3	9,20
gemidd. ritaant. frekw. verdel.	1515	1488	7	14	1	3	1	1	0	0	0	0	

Tabel 4B : 12,5% steekproef (ophoogfaktor = 8,0)

werkel ritaant.	aant. rel.	0 (0)	8 (1)	16 (2)	24 (3)	32 (4)	40 (5)	48 (6)	56 (7)	64 (8)	72 (9)	80 (10)	gemidd. ritaant. schatt.
0	1168	100											0
1	61	87,7	12,3										0,98
2	122	86,7	1,1	12,2									2,04
3	29	81,6	5,9	6,9	5,6								2,92
4	39	80,1	2,4	11,5	0,6	5,4							3,90
5	20	76,5	6,6	7,4	1,9	3,8	3,8						4,90
6	25	75,5	3,8	9,6	1,3	5,6	0,6	3,7					5,96
7	15	70,1	4,9	12,2	2,9	3,7	2,9	2,3	1,0				7,05
8	14	64,3	5,8	16,6	4,4	4,9	0,5	2,2	0,2	1,1			7,82
9	10	63,2	7,6	11,5	4,3	6,5	1,2	2,9	2,0	0,5	0,3		9,09
10	12	60,1	5,2	18,7	2,8	6,4	1,9	3,3	0,3	0,6	0,0	0,7	9,57
gemidd. ritaant. frekw. verd.	1515	1449	17	33	4	7	2	2	0	0	0	0	

ritten. Overeenkomstig loopt het bij de ritaantallen 2, 3, 4, etc. Kleine ritaantallen in de populatie worden dus ōf nul ōf veel groter dan zichzelf, n.l. veelvouden van de ophoogfaktor. De nauwkeurigheid van schattingen van de kleine aantallen is dienovereenkomstig: foutenmarges van honderden procenten liggen voor de hand.

e. Het voorgaande overziende kunnen we het volgende stellen:

1. Bij de gebruikelijke opzet (gebiedsindeling, ritdefinitie, e.d.) van de verkeersonderzoekingen mag men een *rittenpopulatiepatroon* veronderstellen waarvan de ritaantalfrekwentieverdeling de volgende kenmerken heeft (fig. 2):
   - de hoogste frekwenties hebben relaties met 0,1 of soms 2 ritten.
   - vrijwel alle relaties hebben een ritaantal ⩽10, een zeer groot gedeelte ritaantallen ⩽5;
   - de groep van relaties met ritaantal ⩽5 vormt een niet te verwaarlozen percentage van alle ritten (vaak meer dan 50%).

2. Bij de gebruikelijke wijze en omvang van *steekproeftrekken* bij verkeersonderzoekingen wordt de groep van kleine relaties, ruwweg tussen 0 en 10 ritten, slechts zeer ten dele gedetekteerd; het merendeel ervan verschijnt als nul in de steekproef.

3. Bij de gebruikelijke *ophoging* worden alle steekproefaantallen vermenigvuldigd met de ophoogfaktor. Vaak varieert deze enigszins van zone tot zone. De opgehoogde aantallen hebben daardoor de volgende kenmerken (zie fig. 2):
   - het percentage nul-relaties is vrij hoog (tussen 40 en 80%);
   - kleine aantallen, tussen 0 en 10, komen vrijwel niet voor;
   - de veelvouden van de (gemiddelde) ophoogfaktor vertonen opvallend hoge frekwenties.

   De *schattingen* van de populatie-aantallen zijn gekenmerkt door (zie fig. 2):
   - een systematische overschatting van het aantal nul-relaties;

- een systematische vertekening bij de kleine aantallen: het meredeel wordt zwaar onderschat, de rest zwaar overschat.

2.3. Waarvan is het optreden van kleine aantallen afhankelijk?

a. *Een eerste benadering* van de in de kop gestelde vraag is mogelijk met behulp van de waarschijnlijkheidstheorie.
Stel, we verdelen een verzameling van N ritten willekeurig over R klassen, de cellen van de herkomstbestemmingsmatrix. Welke frekwentieverdeling van ritaantallen mag men dan verwachten?
M.a.w., welke verdeling zou resulteren na herhaaldelijk "at random" toedelen van N aan R? Dit is gelijk aan de vraag: wat is de kan op 0,1 , 2 of x ritten in een willekeurige cel?
Deze kans kan worden afgeleid met behulp van de multinomiale formule [5] en bedraagt:

$$p\,(x) = \frac{N!}{x!\,(N-x)!} \cdot (R-1)^{N-x} \cdot \frac{1}{R^N} \qquad x \geqslant 0$$

Is het aantal zones gelijk aan n, dan geldt:

$$R = n^2$$

zodat: $$p\,(x) = \frac{N!}{x!\,(N-x)!} \cdot (n^2-1)^{N-x} \cdot \frac{1}{n^{2N}}$$

De kans op nul ritten in een willekeurige cel (relatie) is dan:

$$p\,(0) = \left(\frac{n^2-1}{n^{2N}}\right)^N = \left(\frac{n^2-1}{n^2}\right)^N = \left(1-\frac{1}{n^2}\right)^N$$

De kans op een ritaantal ⩽ 5 is gelijk aan:

$$p\,(\leqslant 5) = \sum_{x=0}^{5} p(x)$$

De formule voor p(0) maakt in één oogopslag duidelijk welke invloed de faktoren N en n hebben. Het is volledig in overeenstemming met wat ons "gezonde boerenverstand" zegt.
Het blijkt verder dat in de gebruikelijke situaties, waarbij het aantal ritten N in de populatie veel groter is dan het aantal relaties n^2 (zoals de gevallen in tabel 1), het percentage nulrelaties dat bij "toevalstoedelingen" zou resulteren verwaarloosbaar klein is (< 1%).
Voor de twee gevallen met een zeer fijne zone-indeling (Leicester en Waco) is nagegaan welk percentage relaties met een ritaantal ⩽ 5 men zou mogen verwachten in de populatie bij een toevalstoedeling van alle ritten aan alle matrixcellen. Dit percentage beschouwen wij als redelijk vergelijkbaar met het percentage nul-relaties in de steekproef. In Leicester zou dit 28% zijn (tegen 76% in werkelijkheid) en in Waco zelfs meer dan 50% (tegen 79% in werkelijkheid).
Hiermeè is aangetoond dat op zich niet zo verwonderlijk is wanneer bij fijne zone-indelingen het overgrote deel der relaties ⩽ 5 is.
Verder mogen wij uit de vergelijking van de "toevalspercentages" met de empirische gegevens konkluderen dat de populatieverdelingen kennelijk een hoge mate van ruimtelijke koncentratie hebben en niet gelijkmatig of toevals verdeeld zijn.

b. De weinige *empirische gegevens* die kunnen worden verstrekt (zie de tabellen 1 t/m 4) geven ook duidelijke aanwijzingen omtrent de faktoren die het optreden van nulrelaties en kleine aantallen in de rittenmatrices beïnvloeden. Maar zelfs zonder deze data is het niet moeilijk de belangrijkste faktoren op te sommen:
1. de omvang van de populatie van ritten;
2. de mate van ruimtelijke spreiding resp. koncentratie van deze ritten;
3. aantal en ordening van de zones;
4. omvang en aard van de steekproef.

Er geldt ceteris paribus: hoe groter de populatie, hoe kleiner de kans op kleine relaties. De figuren 3a-d geven een illustratie van dit effekt. De omvang van de populatie is op zijn beurt vooral een funktie van de ritdefinitie (zoals periode, vervoermiddel, motief, etc.).

Ook de mate van ruimtelijke koncentratie van de rittenpopulatie is vooral afhankelijk van het rittype (periode, vervoermiddel, motief, etc.) en van de verdeling van het grondgebruik en het vervoersnet in het onderzoekgebied. Het is bv. aannemelijk dat het percentage kleine ritrelaties bij de busritten veel groter zal zijn dan bij autoritten, al het overige gelijk zijnde.
Er geldt ceteris paribus: hoe groter de ruimtelijke koncentratie van de ritten, hoe groter de kans op kleine relaties.
In de zeer goed vergelijkbare gevallen van Apeldoorn en Hilversum is de mate van ruimtelijke koncentratie berekend met de maatstaf [6]:

$$H = 1 - \frac{\Sigma p_i \log \frac{1}{p_i}}{\log n}$$

$i = 1,n$

p_i = relatieve ritfrekwentie in klasse i

n = aantal klassen.

$$0 \leqslant H \leqslant 1$$

H = 0 is: minimale koncentratie (maximale spreiding)

H = 1 is: maximale koncentratie.

Het blijkt bv. dat de rittenpatronen in Apeldoorn duidelijk meer gekoncentreerd zijn dan die in Hilversum. H varieert in Apeldoorn tussen 0,07 en 0,11; in Hilversum tussen 0,04 en 0,07. De verschillen in de gevonden percentages nulrelaties tussen beide steden die resteren na uitschakeling van de faktor "populatieomvang" (fig. 3a) zouden hiermee kunnen worden verklaard. (Het is overigens ook denkbaar dat deze H-waarden door verschil in zone-indeling mede worden bepaald.) Verder blijkt de ochtendspits meer gekoncentreerd te zijn dan de avondspits; en de fietsritten meer dan de autoritten.

Ten aanzien van de zoneïndeling geldt ceteris paribus: hoe meer zones, hoe groter de kans op kleine relaties.
De invloed van het zone-aantal is sterk afhankelijk van de omvang van de populatie, zoals uit de gegeven formules blijkt.
Het effekt van de zoneïndeling is het zuiverst zichtbaar in tabel 2: bij opsplitsing van de kombinatiezone in 3 aparte zones neemt het percentage nulrelaties sprongsgewijs toe.

Het effekt van de steekproefomvang is onderzocht door Benson et al. [3] en Thomas [4]. Een duidelijke illustratie hiervan geeft een vergelijking van tabel 4A en 4B. De bevindingen van Thomas zijn weergegeven in fig. 3c-d. Het is duidelijk dat hoe kleiner de steekproef is, hoe groter het percentage kleine relaties in de steekproef zal zijn. Dit effekt is sterk afhankelijk van het bovenbehandelde drietal kenmerken van de populatie.

2.4. Problemen

Heeft men bij ruimtelijk onderzoek te maken met een interaktiematrix met veel kleine aantallen dan komt een reeks problemen naar voren:

- de geijkte ophoogprocedure ("schatter") is niet adequaat. De steekproeffaktor als ophoogfaktor levert weliswaar een zuivere schatting, doch is verre van nauwkeurig.
- de onnauwkeurigheid van de schatting van de interaktiematrix is groot: per kleine relatie loopt de fout in de honderden procenten, zeer veel relaties hebben een dergelijke fout, en ondanks de kleine aantallen waar het om gaat wordt een belangrijk deel van de interakties (ritten) verkeerd toegedeeld (totale absolute fout).
- het opstellen en toepassen van modellen (van het zwaartekrachtstype) van het interaktiepatroon is diskutabel: zowel methodologisch als wiskundig. Zwaartekrachtsmodellen zijn geschikt voor massaverschijnselen waarvan de afhankelijk variabele veel vari-

atie heeft en als kontinue kan worden verondersteld. Bij het overheersen van kleine aantallen gaat dit niet meer op. Bij afleidingen via bv. entropie-maximalisatie ontstaat verder het probleem dat enkele zeer belangrijke benaderingen, waaronder die van Stirling, bij kleine aantallen zeer onnauwkeurig worden (zie [7]).

- het nummerieke "rekenen" is zeer gevoelig: afrondingen bv. hebben grote gevolgen.
- etc.

De meest voor de hand liggende oplossing zou zijn kleine aantallen te vermijden door een grotere steekproef te nemen of een andere onderzoekopzet (gebiedsindeling, e.d.) te kiezen. Maar wat te doen als toch met een steekproef-interaktiematrix "vol" nullen en kleine aantallen moet worden verder gewerkt? We zullen hierna ingaan op enkele alternatieve ophoogprocedures voor dergelijke situaties.

3. SUGGESTIES VOOR OPHOOGPROCEDURES EN SCHATTINGSVERBETERING

3.1. Algemeen

Zuiver statistisch gezien, d.w.z. wanneer er buiten de steekproefresultaten geen informatie over het onderzochte verschijnsel beschikbaar is, is direkte ophoging met de reciproke van de steekproeffraktie de beste schatter voor de populatieaantallen. De schattingen zijn zuiver en hebben de kleinste variantie, hoe vreemd dit misschien op eerste gezicht ook moge lijken voor wat betreft de nul- en kleine relaties |8|.
Verbetering van de schattingen is alleen mogelijk als extra informatie beschikbaar komt over het verschijnsel in kwestie. In het volgende gaan we dan ook uit van deze situatie.
Er zullen een drietal verschillende benaderingen worden beschreven om een steekproefinteraktiematrix op te hogen in geval van hoge percentages relaties met zeer kleine aantallen. Het zijn slechts suggesties bedoeld om een diskussie over deze problematiek

los te maken.

a. De eerste methode is gebaseerd op principes uit de statistische testtheorie.
b. De tweede mogelijkheid is de modelmethode: een model van het interaktiepatroon gemaakt op basis van de steekproef levert de uiteindelijke populatieschattingen.
c. Al meer numeriek en minder statistisch van aard is een disaggregatiebenadering op basis van entropiemaximalisatie.

Uitgaande van deze principes zijn weer allerlei variaties en kombinaties mogelijk.
Eén van de eisen die men aan deze methoden zal moeten stellen is dat nullen in de steekproef door de ophoging in principe in positieve getallen moeten kunnen veranderen.
Het is de vraag of men moet eisen dat de schattingen van de ritaantallen diskreet zijn, i.c. nullen of positieve gehele getallen, of dat ook niet-gehele ritaantallen als zinvolle schattingen kunnen worden beschouwd. Dit belangrijke inhoudelijke facet hangt nauw samen met de afrondingsproblematiek die bij kleine aantallen zo een grote rol speelt. We zullen in het bestek van deze paper echter niet hierop kunnen ingaan.
Bij schattingsverbetering verdienen nog twee punten de aandacht.
Allereerst zou, welke methode ook wordt gebruikt, altijd maximaal gebruik moeten worden gemaakt van de informatie die in de steekproef zit. In konkreto wordt bedoeld dat gebruik zou moeten worden gemaakt van de kennis van de nauwkeurigheidsmarges die men heeft bij elke afzonderlijke schatting. Dit geldt overigens niet alleen voor de ophoging of schattingsverbetering maar voor elk gebruik van de steekproefuitkomsten, zoals bv. bij het schatten van modellen.
Ten tweede, voordat men aan de ophoging begint zou men moeten trachten aan te geven, op grond van additionele informatie, welke nulrelaties in de steekproef gegarandeerd echte nulrelaties in de populatie zijn.
Afgezien van speciale ophoogprocedures voor de kleine aantallen is het zinvol ook nog even de aandacht te vestigen op de ophoogprocedure überhaupt, toe te passen op alle relaties. In de prak-

tijk wordt vrijwel uitsluitend opgehoogd via direkte schatters: de ophoogfaktor is de reciproke van de steekproeffraktie.

$$\hat{T}_{ij} = t_{ij} \frac{N_i}{n_i}$$

$\hat{T}_{ij}$ = schatting van populatieaantal ritten van i naar j gemaakt door huishoudens van zone i.
t_{ij} = aantal ritten in steekproef.
N_i = totaal aantal huishoudens in i.
n_i = aantal steekproefhuishoudens in i.

In veel gevallen zal het echter beter zijn gebruik te maken van zg. ratio- en regressieschatters [8]. In deze gevallen wordt gebruik gemaakt van een hulpvariabele die bij de steekproefeenheden (huishoudens) en eventueel populatie wordt gemeten; deze is gemakkelijk en nauwkeurig meetbaar en heeft een hoge korrelatie met de te schatten variabele. In plaats van op te hogen met het quotient van de aantallen huishoudens in populatie en steekproef wordt dan bv. opgehoogd met het quotient van de aantallen personen in de populatie en in de steekproefhuishoudens.

ratioschatter: $\hat{T}_{ij} = t_{ij} \frac{\sum_{k=1}^{N} x_{ik}}{\sum_{k=1}^{n} x_{ik}} = t_{ij} \frac{X_i}{x_i}$ x_{ik} = hulpvariabele van huishouden k in zone i (bv. aantal personen).

regressieschatter: $\hat{T}_{ij} = t_{ij} + b_i (X_i - x_i)$ b_i = regressiekoëffieciënt voor zone i.

Hoewel deze methode dan in het algemeen tot betere schattingen zal leiden, biedt deze toch geen soelaas voor het probleem van de kleine aantallen. Er wordt immers weer met een vaste faktor vermenigvuldigd, nullen blijven nullen.

3.2. Testen op statistische signifikantie

Thomas [4] doet een suggestie met betrekking tot de nulrelaties, uitgaande van principes uit de statistische testtheorie. Hij onderscheidt 3 groepen van nulrelaties in de steekproefmatrix:
a. relaties die in de populatie ook nul zijn (echte nulrelaties)
b. relaties die statistisch niet signifikant afwijken van het populatieaantal (negatieve nulrelaties);
c. relaties die wel signifikant verschillen van het ritaantal in de populatie (positieve nulrelaties).

De beslissing bij welk ritaantal populatie en steekproef signifikant van elkaar afwijken vereist de (subjektieve) keuze van een betrouwbaarheidniveau.
Thomas wil de positieve nulrelaties veranderen in een klein positief getal. Welke relaties hiervoor in aanmerking komen is normaal in praktijksituaties niet bekend. Daarom zou via een toevalstrekking zoveel nulrelaties met een klein positief getal bezet moeten worden als er vermoedelijk positieve nulrelaties in de matrix zijn. Uit daartoe geëigend onderzoek wordt afgeleid welk percentage van de nulrelaties positieve nulrelaties zijn. Dit hangt van een aantal faktoren af, zoals vooral van de populatieomvang (totaal aantal ritten), de steekproeffraktie en de gekozen betrouwbaarheidswaarschijnlijkheid. Thomas heeft over deze samenhang uitgebreid onderzoek gedaan.
Op het eerste gezicht vraagt men zich bij deze procedure af of het middel niet erger is dan de kwaal, of het wel tot informatieverbetering leidt. Het risiko bestaat immers dat echte nulrelaties met een getal worden bezet, of dat relaties nul blijven die in de populatie niet nul zijn. De kans op zo een fout is echter berekenbaar; deze is vooral een funktie van het gekozen betrouwbaarheidsniveau. Op grond hiervan kan men beslissen of men het risiko verbonden met deze schattingsverbetering aksepteert. Ook de keuze van het nul vervangend getal gebeurt in afhankelijkheid van het risiko dat men wil lopen eventueel een fout te maken.
De kans op echte verbeteringen is het grootst bij matrices met hoge percentages positieve nulrelaties. Dit treedt op als de

populatie omvang groot is en de steekproef daaruit klein (zie fig. 4). Men zou het percentage positieve nulrelaties in een bepaalde situatie ook in die situatie zelf kunnen schatten met behulp van een 100% enquête in een steekproef van de zones. De methode van Thomas kan in principe ook worden uitgebreid tot de andere kleine relaties.

Voor een beoordeling van Thomas' suggestie is het nodig er op te wijzen dat de populatieparameter waarop deze procedure is gericht het aantal ritten van een individuele relatie is.

Andere populatieparameters, zoals het aantal ritten per zone, of het totaal aantal ritten in het studiegebied worden door de steekproef doorgaans goed geschat en behoeven geen verbetering. Beperkt men zich dus tot het verbeteren van de nulrelaties dan ontstaan er inkonsistenties tussen de schattingen van de verschillende parameters, t.w. individuele relaties en totalen daarvan. Deze kunnen echter worden verholpen door het binnenwerk van de matrix op statistisch verantwoorde wijze aan te passen aan de gegeven randtotalen, nl. rekening houdend met de nauwkeurigheid van elke individuele schatting (zie bv. [9]).

3.3. De modelmethode

Bij de modelmethode wordt gebruik gemaakt van kennis die men heeft over de samenhang tussen de gezochte variabele, t.w. de zonale interaktie, en verklarende variabelen. Het aantal ritten tussen twee zones kan bv. in verband worden gebracht met kenmerken van beide zones, zoals aantal inwoners en onderlinge afstand. Deze samenhang is gegoten in de vorm van een wiskundig model, zoals bv. het zwaartekrachtsmodel.

De werkwijze bij deze schattingsverbetering is dan als volgt. De steekproefuitkomsten worden voorlopig volgens een standaardprocedure opgehoogd (direkte of ratio-resp. regressieschatters). Van deze schattingen worden de nauwkeurigheidsmarges bepaald. Met behulp van deze voorlopige schattingen wordt het interaktiemodel gekalibreerd. Het is te overwegen of de nauwkeurigheidsmarges hierbij als wegingsfaktor dienst kunnen doen. De uitkomsten van dit model, bij dezelfde gegevens voor de verklarende variabelen, zijn het uitgangspunt voor schattingsverbetering. Ligt de modeluitkomst binnen de nauwkeurigheidsmarges van de voorlopige schatting dan wordt deze als nieuwe schatting genomen. In feite is deze methode een vorm van "smoothing", het vereffenen van (steekproef) onnauwkeurigheden door berekening van het "gemiddelde".

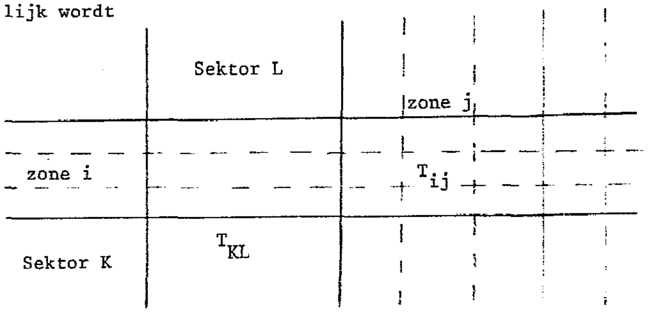
Voorwaarde voor sukses is dat het model een geringe specifikatiefout heeft en dat de verklarende variabelen nauwkeurig meetbaar zijn.

In hoeverre de voorgestelde methode daadwerkelijk tot betere schattingen leidt, en onder welke omstandigheden, is alleen goed mogelijk op grond van steekproefsimulaties uit een 100% enquête. Onderzoek hierover is niet bekend. Een klein aanknopingspunt biedt een studie van Long [10] die speciaal het gedrag van het zwaartekrachtsmodel in het bereik van de kleine aantallen heeft geanalyseerd voor uiteenlopende ritkategorieën. De op zichzelf belangwekkende analyse laat echter geen konklusies toe t.a.v. het vraagstuk van de schattingsverbetering.

Wel blijkt er t.a.v. de zonale ritproduktie uitvoerig studie te zijn gedaan naar de mogelijkheden van schattingsverbetering met behulp van modellen [11]. Dit is nagegaan bij diverse steekproeffrakties, verschillende modellen, en enkele ritkategorieën. In alle gevallen werd een aanzienlijke schattingsverbetering bereikt, met uitzondering van de "non-home-based" ritten. Bij deze ritkategorie is de eigenlijke steekproefschatting steeds het beste.

3.4. Matrixdisaggregatie

Bij deze methode is de basis het gegeven dat grote aantallen nauwkeurig, "hard", zijn. Bij de interaktiematrices kunnen bv. de randtotalen, geschat uit de steekproef, doorgaans als voldoende "hard" worden beschouwd. Voor veel interzonale relaties geldt dit echter niet omdat ze zeer kleine aantallen hebben. Pas door groepering van zones, laten we zeggen in sektoren, ontstaan interakties van zodanige omvang dat de steekproef-onnauwkeurigheid verwaarloosbaar wordt. Om betere schattingen te krijgen van de kleine interzonale relaties zou men de harde relaties tussen sektoren kunnen disaggregeren over de bijbehorende zoneparen overeenkomstig één of ander kriterium en rekening houdemd met een aantal randvoorwaarden zoals de zonale randtotalen en de intersektorale relatieaantallen. In dit kriterium en eventuele additionele randvoorwaarden is de extra-statistische informatie vervat waarmee een verbetering t.o.v. de steekproef mogelijk wordt



Stel we nemen als kriterium "de meest waarschijnlijke verdeling", d.w.z. we maximaliseren het aantal mikroverdelingen. Mathematisch ziet deze benadering er dan als volgt uit:

t_{ij} = steekproefaantal
T_{ij} = opgehoogd steekproefaantal = voorlopige populatieschatting.
$\hat{T}_{ij}$ = verbeterde populatieschatting
T = totaal aantal ritten populatie.

$$\text{maximaliseer} \quad \frac{T!}{\prod_{ij} \hat{T}_{ij}!}$$

onder de randvoorwaarden:

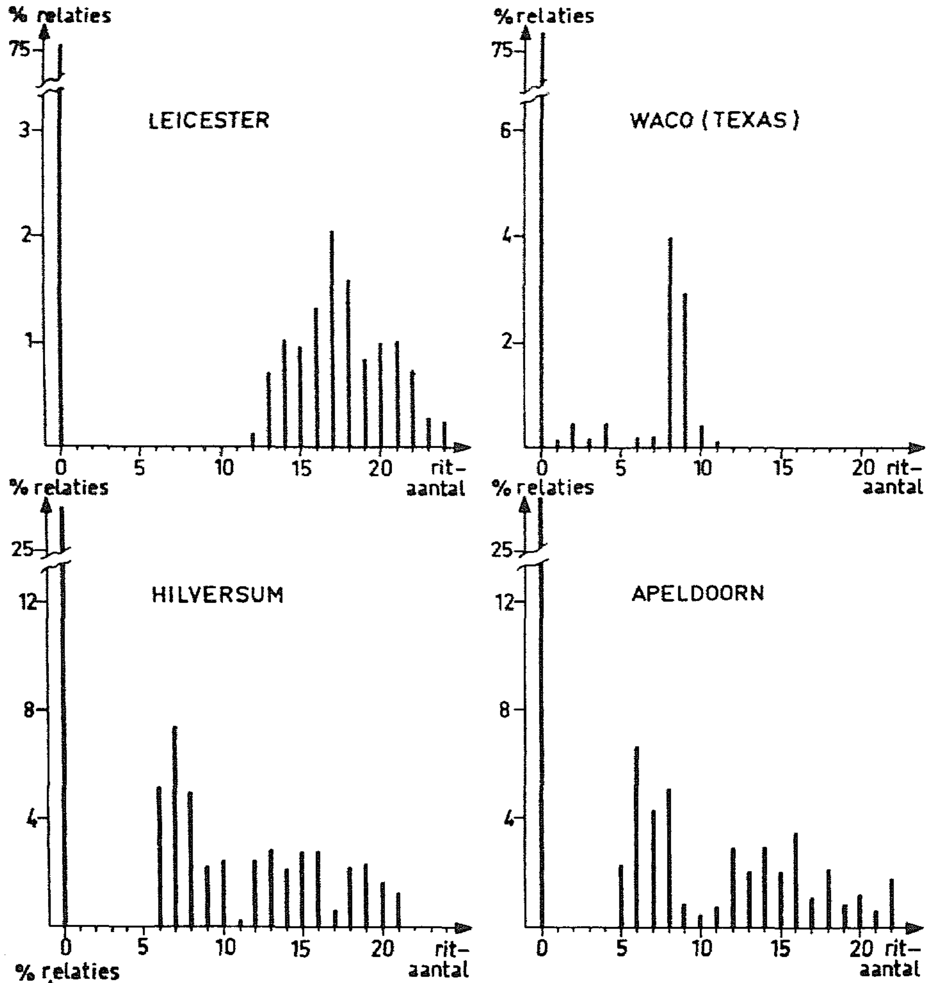
$$\hat{T}_{ij} \geq t_{ij} \qquad \forall j \; \forall i$$

$$\sum_i \hat{T}_{ij} = \sum_i T_{ij} \qquad \forall j$$

$$\sum_j \hat{T}_{ij} = \sum_j T_{ij} \qquad \forall i$$

$$\sum_{\forall i \in K} \sum_{\forall j \in L} \hat{T}_{ij} = T_{KL} \qquad \forall L \; \forall K$$

Dit zijn de minimale randvoorwaarden. Naar behoefte kunnen andere worden toegevoegd.
Het beschreven disaggregatieprincipe kan worden uitgebreid tot het disaggregeren naar bv. ritkategoriën, perioden, etc..



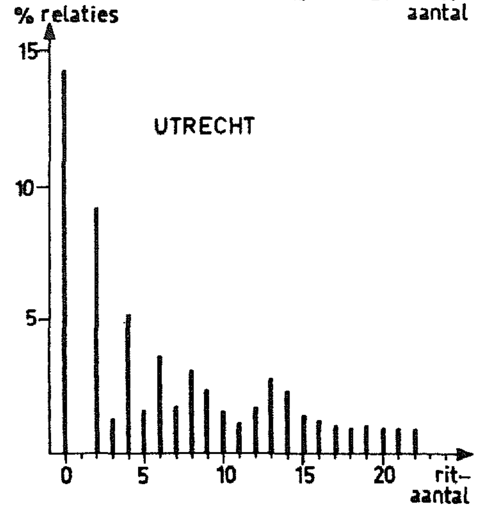


Fig. 1: Enkele praktijkvoorbeelden van frekwentieverdelingen van ritaantallen in herkomst-bestemmingsmatrices.

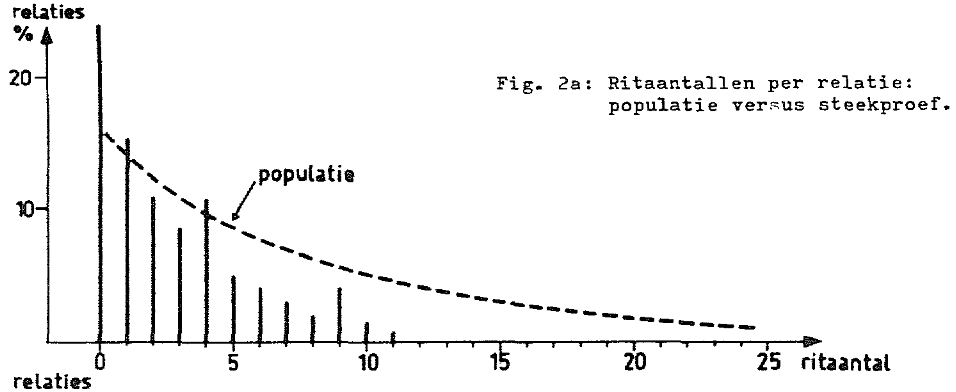


Fig. 2a: Ritaantallen per relatie: populatie versus steekproef.

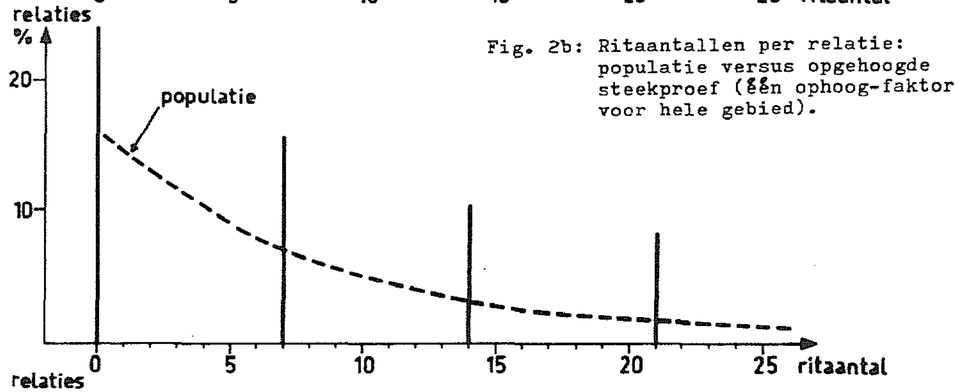


Fig. 2b: Ritaantallen per relatie: populatie versus opgehoogde steekproef (één ophoog-faktor voor hele gebied).

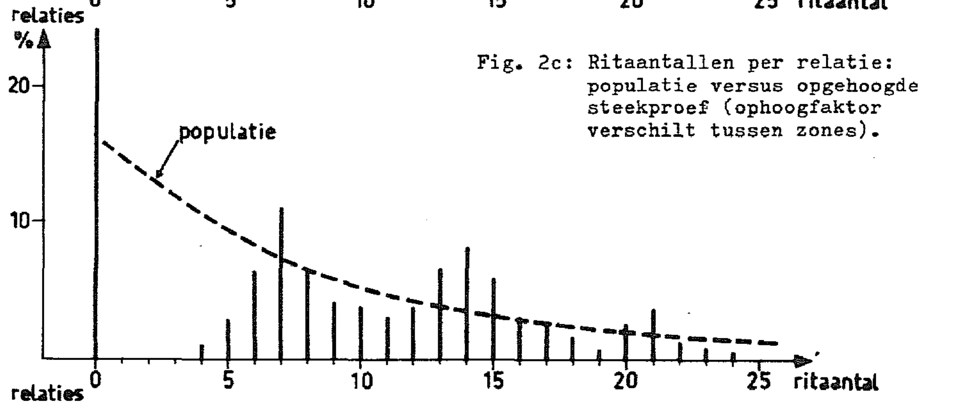


Fig. 2c: Ritaantallen per relatie: populatie versus opgehoogde steekproef (ophoogfaktor verschilt tussen zones).

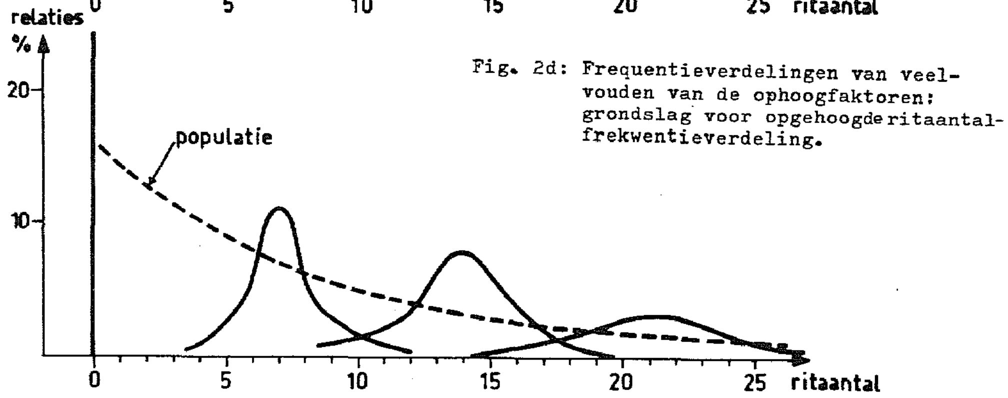


Fig. 2d: Frequentieverdelingen van veelvouden van de ophoogfaktoren: grondslag voor opgehoogde ritaantalfrekwentieverdeling.

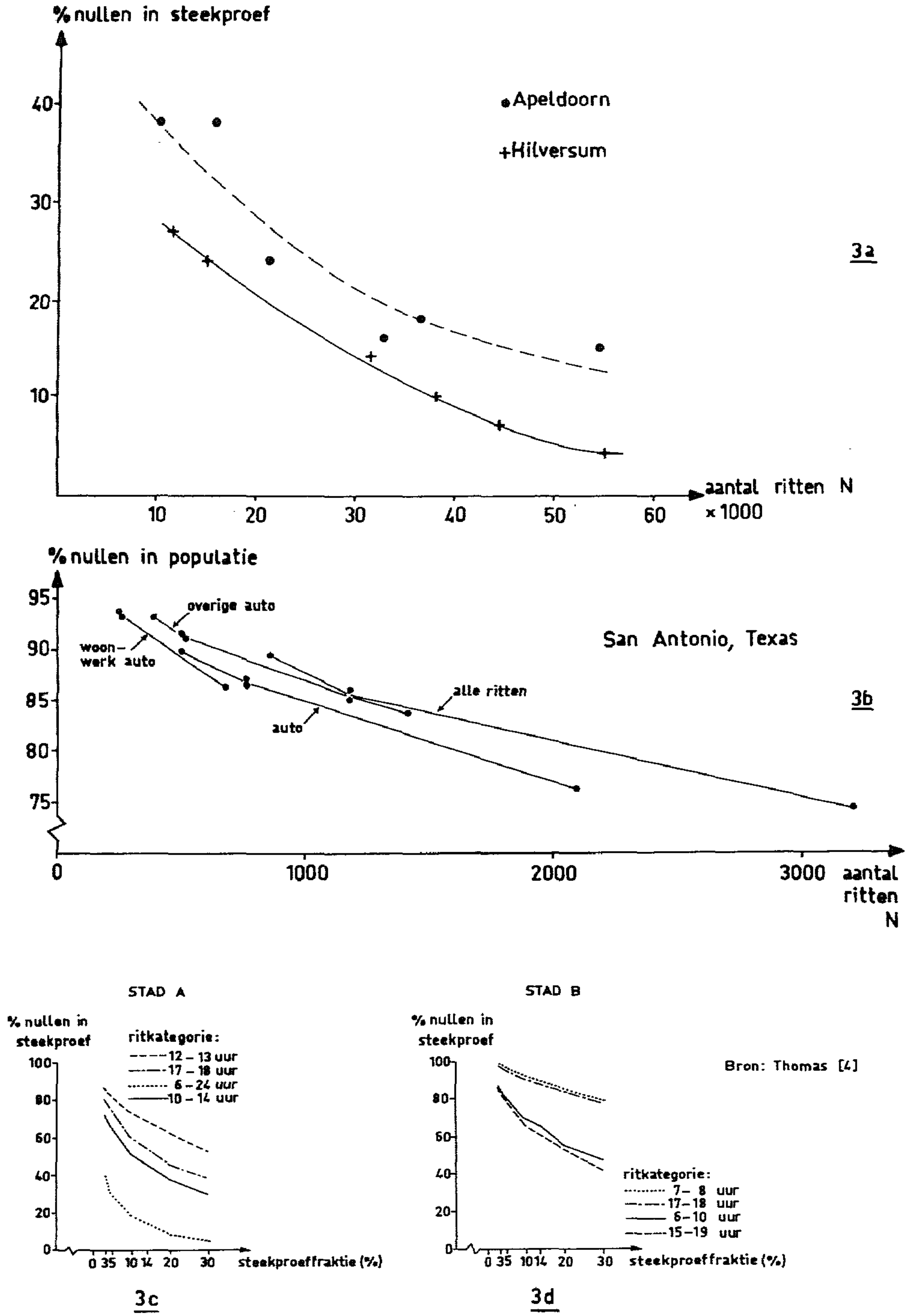


Fig. 3

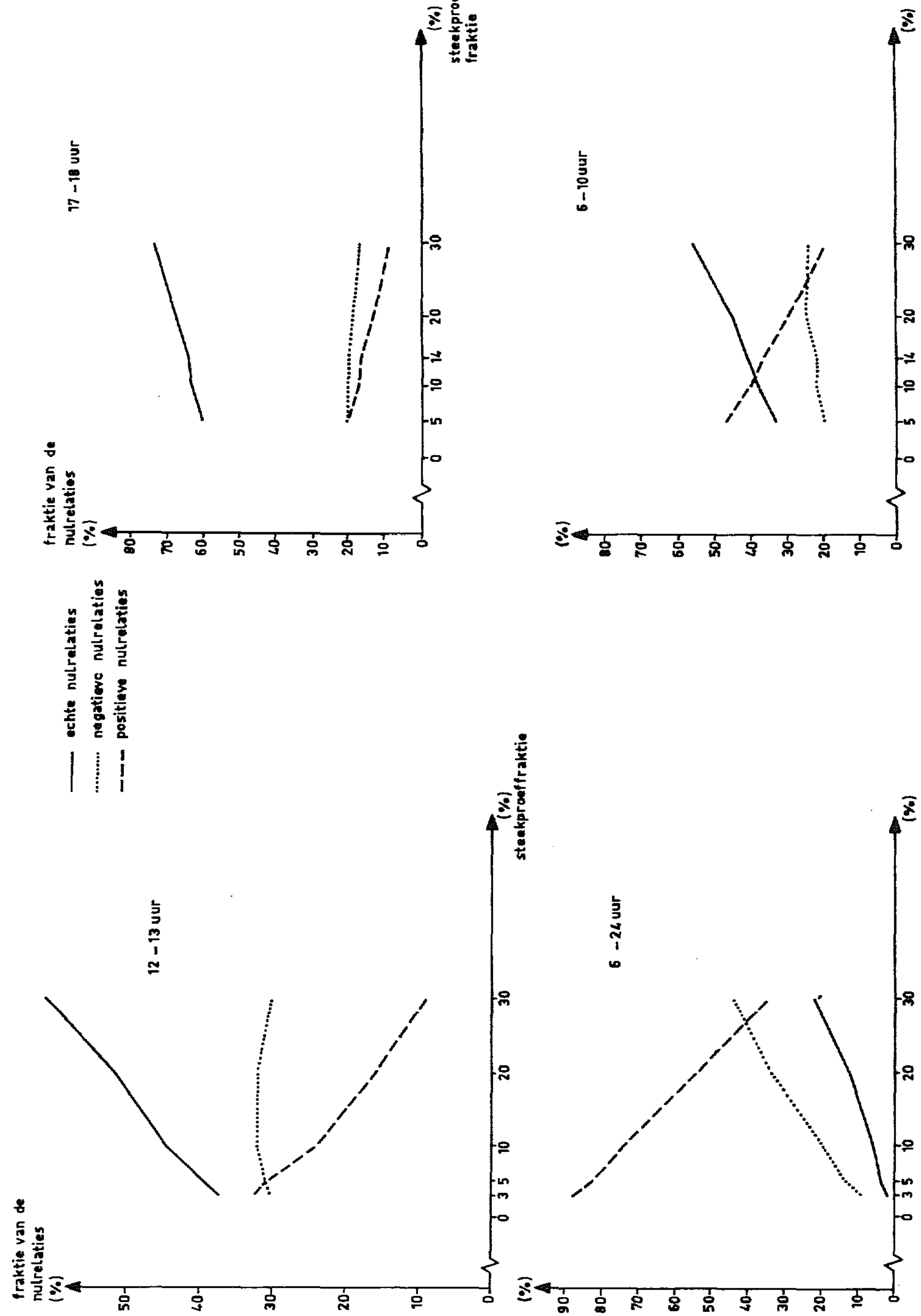


Fig. 4 Bron: Thomas [4]

LITERATUUR

1. *Bovy, P.H.L. & Jansen, G.R.M.*
"Ruimtelijke abstraktie in de vervoersplanning: een probleemanalyse en opzet voor initieel onderzoek".
in: P.H.L. Bovy et al. (red.)."Colloquium vervoersplanologisch speurwerk -1974-. Modellen en methoden in de vervoersplanologie". Delft, 1974, pp. 18.1-18.42.

2. *Bovy, P.H.L. & Jansen, G.R.M.*
"Het effekt van het ruimtelijk detailniveau op de rekentijden van het assignment"
in: F. le Clercq et al. (red.). "Colloquium vervoersplanologisch speurwerk -1975- praktijk en model in de vervoersplanning".
Delft, 1975, pp. 545-

3. *Benson, J.D., Pearson, D.F. & Stover, V.G.*
"Accuracy of travel pattern estimates from the home interview survey".
College Station (Tex.), Texas A&M Univ., Texas Transportation Institute, March 1974, Research Report 167-8.

4. *Thomas, W.*
"Sensitivitätsanalyse eines Verkehrsplanungsmodells".
Karlsruhe, Technische Universität, Institut für Verkehrswesen 1975.

5. *Adler, I.*
"Waarschijnlijkheidsrekening en statistiek".
Utrecht, Spectrum, 1966.

6. *Pluym, W.K.*
"De statistische informatietheorie in het ruimtelijk onderzoek".
Delft, T.H., Instituut voor Stedebouwkundig Onderzoek, 1976 (in voorbereiding).

7. *Chilton, R. & Poet, R.R.W.*
"An entropy maximising approach to the recovery of detailed migration patterns from aggregate census data".
Environment & Planning, Vol. 5 (1973) pp. 135-146.

8. *Moors, J.J.A. & Muilwijk, J.*
"Steekproeven, een inleiding tot de praktijk".
Amsterdam, Agon Elsevier, 1975.

9. *Terhal, P.H.J.J.*
"Het schatten van het binnenwerk van een matrix bij gegeven randtotalen".
Statistica Neerlandica, Vol. 24 (1970) No. 3, pp. 125-126.

10. *Long, G.D.*
"An evaluation of the gravity model trip distribution".
College Station (Texas), A&M University, Texas Transportation Institute, August 1968, Research Report No. 60-13.

11. *Parsonson, P.S. & Cribbins, P.D.*
"Estimating urban trip production and attraction".
ASCE Journal of the Highway Division, 1967, No. HW2, pp. 115-154.