



# **ONDERZOEK VERPLAATSINGSGEDRAG STADSGEWEST HASSELT - GENK (april 1999 - april 2000)**

---

## DEEL 1: METHODOLOGISCHE ANALYSE

## Inhoud

<b>1</b>	<b><i>Inleiding</i></b> .....	<b>3</b>
<b>2</b>	<b><i>Non-respons</i></b> .....	<b>4</b>
2.1	<b>Bereidwilligheid om vragenlijsten in te vullen</b> .....	<b>4</b>
2.2	<b>Profiel van huishoudens die (uiteindelijk) niet meewerkten</b> .....	<b>5</b>
<b>3</b>	<b><i>Bepalen van gewichten</i></b> .....	<b>7</b>
3.1	<b>Effectief gebruikte gewichten huishoudens</b> .....	<b>8</b>
3.2	<b>Effectief gebruikte gewichten personen</b> .....	<b>9</b>
3.3	<b>Effectief gebruikte gewichten verplaatsingen</b> .....	<b>10</b>
<b>4</b>	<b><i>Technische aspecten i.v.m. de statistische verwerking</i></b> .....	<b>10</b>
4.1	<b>Frequentietabellen en regressie</b> .....	<b>10</b>
4.2	<b>Betrouwbaarheid van de resultaten</b> .....	<b>12</b>
<b>5</b>	<b><i>Vergelijking telefonisch/postaal bevroegden en enkel postaal bevroegden</i></b> .....	<b>12</b>
<b>6</b>	<b><i>Minder antwoorden op de 2<sup>e</sup> invuldag</i></b> .....	<b>15</b>
<b>7</b>	<b><i>Bibliografie</i></b> .....	<b>16</b>
<b>8</b>	<b><i>Bijlagen</i></b> .....	<b>17</b>
<b>8.1</b>	<b>Berekening van de gewichten</b> .....	<b>17</b>
8.1.1	Stappenplan voor meerdere marginale verdelingen.....	17
8.1.2	Huishoudens: vier relevante variabelen.....	18
8.1.3	Personen: tweemaal een gezamenlijke verdeling van vier variabelen.....	20
8.1.4	Verplaatsingen: een verdeling van één variabele.....	21
<b>8.2</b>	<b>Berekening van de ophoogfactor</b> .....	<b>21</b>
<b>8.3</b>	<b>Samenvoegen van gegevens</b> .....	<b>21</b>
8.3.1	Statuut.....	21
8.3.2	Doel.....	22
<b>8.4</b>	<b>Vragenlijsten</b> .....	<b>24</b>

# 1 Inleiding

Tijdens de periode april 1999 tot april 2000 werden er gegevens verzameld over een aantal mobiliteitskenmerken van gezinnen en personen (vanaf 6 jaar) in het stadsgewest Hasselt-Genk<sup>1</sup> waaronder het verplaatsingsgedrag van personen. De steekproef voor deze studie bestond uit 2.500 gezinnen uit het stadsgewest Hasselt-Genk.

Dit onderzoek gebeurde via een enquête waarbij:

- 1) een vragenlijst moest ingevuld worden met gegevens over het gezin (gezinsvragenlijst)
- 2) een vragenlijst moest ingevuld worden met gegevens over de gezinsleden vanaf 6 jaar met daarbij ook een deel over hun verplaatsingen tijdens een opgegeven periode van 2 dagen (personenvragenlijst met verplaatsingendeel).

We wilden dus analyses doen op gegevens van 2.500 gezinnen. Omdat niet alle huishoudens meedoen aan het onderzoek, werden 5000 huishoudens aselekt getrokken uit het Rijksregister. Een tweede steekproef van 5000 huishoudens werd in de maand oktober 1999 bezorgd. De tweede steekproef werd niet volledig opgebruikt. Uiteindelijk werden voor het stadsgewest Hasselt-Genk 6.674 huishoudens gecontacteerd.

De contactprocedure was ofwel telefonisch/postaal ofwel uitsluitend postaal. De huishoudens werden indien mogelijk op voorhand telefonisch gecontacteerd. Dit verhoogt de kans op respons, en het geeft een beter beeld op het aantal personenvragenlijsten dat er naar het huishouden moet opgestuurd worden. Indien er geen vaste telefoon was (of ingeval van een geheim nummer), werden 1 huishouden- en 5 personenvragenlijsten opgestuurd.

De verzameling van deze gegevens (= veldwerk) werd uitgevoerd door het onderzoeksbureau Dimarso. De begeleiding en controle van het veldwerk werd uitgevoerd door de Onderzoekscel Architectuur en Mobiliteit van de Provinciale Hogeschool Limburg (departement Architectuur).

De rapportage van deze analyse bestaat uit 3 delen die verwerkt zijn in 3 overeenkomstige en afzonderlijke rapporten:

1. een methodologische analyse
2. een analyse van de huishoudenvragenlijst
3. een analyse van de personenvragenlijst

Het voorliggend document is het rapport met de methodologische analyse.

In dit document geven we een overzicht van de *non-respons* problematiek, waarbij we o.a. nagaan wat het profiel is van mensen die initieel telefonisch wel wilden meewerken, maar uiteindelijk toch afhaakten. We berekenen de *gewichten* om de scheeftrekking ten gevolge van de non-respons te minimaliseren.

---

<sup>1</sup> Als omschrijving van dit stadsgewest werd de afbakening van het Ruimtelijk Structuurplan Vlaanderen (RSV) gehanteerd waarbij wel voor de gemeenten, die volgens dit RSV slechts gedeeltelijk in het stadsgewest gelegen zijn, de totale gemeente in aanmerking werd genomen. Het stadsgewest Hasselt-Genk in dit onderzoek bestaat uit de gemeenten Hasselt, Genk, Diepenbeek en Zonhoven.

Vervolgens bespreken we de gehanteerde *statistische technieken en de betrouwbaarheid* die het gevolg is van deze technieken.

Tenslotte bespreken we nog een aantal *onderzoekresultaten i.f.v. de gehanteerde methodologie*.

De bijlage bevat een meer technische uitleg over de berekening van de gewichten en de ophoogfactor, het samenvoegen van gegevens en de vragenlijsten.

## 2 Non-respons

### 2.1 Bereidwilligheid om vragenlijsten in te vullen

Tabel 1. Responspercentages van OVG Hasselt-Genk en OVG Antwerpen

	Hasselt-Genk	Antwerpen	Hasselt-Genk	Antwerpen
Aantal telnr's gevonden	100%	100%		
Aantal telefonische contacten	86%	82%		
Deelname recruiteringsvragenlijst	80%	77%	100%	100%
Akkoord verdere deelname	70%	65%	87%	85%
Aantal bundeltjes terug	44%	41%	55%	53%
Aantal goedgekeurde bundeltjes*	30%	29%	37%	37%
Aantal goedgekeurde bundeltjes**	8%	7%	10%	9%
<b>Postaal</b>				
Verzonden	100%	100%		
Aantal bundeltjes terug	24%	20%	100%	100%
Aantal goedgekeurde bundeltjes*	15%	13%	64%	65%
Aantal goedgekeurde bundeltjes**	4%	4%		

\* waarbij alle gezinsleden de personenvragenlijsten hebben ingevuld

\*\* waarbij voldoende gezinsleden de personenvragenlijst hebben ingevuld

De percentages zijn steeds berekend t.o.v. de bovenstaande 100%.

Een aantal huishoudens weigerden mee te doen aan de enquête of stuurden onvoldoende formulieren terug. In Tabel 1 geven we de responspercentages van het stadsgewest Hasselt-Genk en ter vergelijking die van het stadsgewest Antwerpen<sup>2</sup>. Op alle mogelijke niveaus is de respons van Hasselt-Genk iets beter dan die van Antwerpen. Het verschil is echter vrij klein.

Voor huishoudens met telefoon is de eerste contactmogelijkheid de telefonische vraag voor medewerking, waarbij ook reeds enkele vragen over het huishouden gesteld worden (= recruiteringsvragenlijst). Bij 14% van de huishoudens die wel telefoon hebben, werd nooit opgenomen. 6% weigerde medewerking. Er bleef dus 80% van de huishoudens over die meewerkten aan de telefonische enquête (voor details van de absolute aantallen, zie Nuyts & Zwerts (2000)).

Uiteindelijk stuurt 38% (=30% + 8%) van de huishoudens bundeltjes met een volledig aantal of een voldoende aantal personenlijsten terug.

<sup>2</sup> Gelijktijdig met het veldwerk van het OVG in het stadsgewest Hasselt-Genk heeft een analoog onderzoek plaatsgehad in het stadsgewest Antwerpen.

Huishoudens die telefonisch niet vooraf geïnformeerd kunnen worden (de 'postale gezinnen'), en die tijdens de enquête niet mondeling aan deze afspraak herinnerd kunnen worden, reageren veel minder dan telefonisch wel bereikte huishoudens. Het responspercentage ligt dan op 19% (+15% + 4%).

## 2.2 Profiel van huishoudens die (uiteindelijk) niet meewerkten.

Voor alle huishoudens die getrokken zijn uit het Rijksregister beschikken we over de leeftijd en het geslacht van het gezinshoofd. Dus ook voor de huishoudens die weigerden om mee te werken. Hierdoor kunnen we in beperkte mate nagaan welk profiel de weigeraars hebben. De weigeraars waarover we het hier hebben zijn zowel respondenten die **niet** aan de recruiteringsvragenlijst hebben deelgenomen (en dus per definitie ook niet aan het schriftelijke deel van de enquête) als degenen die **wel** aan de recruiteringsvragenlijst hebben deelgenomen maar die niet aan het schriftelijke deel wensten deel te nemen).

**Tabel 2. Reacties bij eerste contact van huishoudens die telefonisch bereikt zijn volgens geslacht en leeftijd van het gezinshoofd**

(TROK [Telefonische recruitering OK]=1 is positieve reactie; TROK=0 is weigering (verdere medewerking)

TROK BY LFTKL

TROK	LFTKL						Total
	17-24	25-34	35-44	45-54	55-64	65+	
0	0	29	54	74	72	173	402
	0.00	0.77	1.42	1.95	1.90	4.56	10.61
	0.00	7.21	13.43	18.41	17.91	43.03	
	0.00	7.32	5.59	8.09	10.40	21.23	
1	6	367	912	841	620	642	3388
	0.16	9.68	24.06	22.19	16.36	16.94	89.39
	0.18	10.83	26.92	24.82	18.30	18.95	
	100.00	92.68	94.41	91.91	89.60	78.77	
Total	6	396	966	915	692	815	3790
	0.16	10.45	25.49	24.14	18.26	21.50	100.00

TROK	LFTKL						Total
	17-24	25-34	35-44	45-54	55-64	65+	
0	1	5	13	13	18	188	238
	0.10	0.50	1.30	1.30	1.81	18.86	23.87
	0.42	2.10	5.46	5.46	7.56	78.99	
	33.33	8.62	9.03	9.35	12.24	37.15	
1	2	53	131	126	129	318	759
	0.20	5.32	13.14	12.64	12.94	31.90	76.13
	0.26	6.98	17.26	16.60	17.00	41.90	
	66.67	91.38	90.97	90.65	87.76	62.85	
Total	3	58	144	139	147	506	997
	0.30	5.82	14.44	13.94	14.74	50.75	100.00

Bij het telefonisch contact weigeren vrouwen opvallend meer dan mannen (Tabel 2). Slechts 11% van de mannen weigert medewerking, tegenover 24% bij de vrouwen. Bij beide sexen is de weigering van 65-plussers het grootst (resp. 21% en 37%). Het verschil tussen de andere leeftijdsklassen is niet significant.

Bemerk dat 51% van de vrouwelijke gezinshoofden ouder is dan 65 (Tabel 2) terwijl dit bij de mannen ongeveer 20% is. Dit wil zeggen dat een vrouw al bijna weduwe moet zijn om gezinshoofd te kunnen worden. Daaruit kunnen we afleiden dat bij analyses op 'geslacht gezinshoofd' er bij vrouwelijke gezinshoofden ook andere factoren spelen dan enkel het geslacht van het gezinshoofd.

Bij het telefonisch contact haken dus een aantal huishoudens af (ofwel onmiddellijk ofwel na het beantwoorden van de telefonische recruiteringsvragenlijst). De anderen hebben gesteld dat zij verder aan het schriftelijk deel van het onderzoek wensten mee te werken en deze gezinnen hebben enquêteformulieren ontvangen. Hiervan zijn er een aantal die geen formulieren terugsturen (zie Tabel 3).

**Tabel 3. Terugsturen van de schriftelijke vragenlijst volgens geslacht en leeftijd van het gezinshoofd**

Deze tabel heeft enkel betrekking op die huishoudens die meegewerkt hebben aan de telefonische recruiteringsvragenlijst, en die *gezegd hebben* dat ze ook aan de schriftelijke wilden meedoen.

RQOK (Retour Questionnaire)=0 is negatieve reactie; RQOK=1 is positieve reactie

		LFTKL						
RQOK		CONTROLLING FOR GESLACHT=man						
Frequency	Percent	17-24	25-34	35-44	45-54	55-64	65+	Total
Row Pct	Col Pct							
0	5	164	358	303	193	233	1256	
	0.15	4.84	10.57	8.94	5.70	6.88	37.07	
	0.40	13.06	28.50	24.12	15.37	18.55		
	83.33	44.69	39.25	36.03	31.13	36.29		
1	1	203	554	538	427	409	2132	
	0.03	5.99	16.35	15.88	12.60	12.07	62.93	
	0.05	9.52	25.98	25.23	20.03	19.18		
	16.67	55.31	60.75	63.97	68.87	63.71		
Total	6	367	912	841	620	642	3388	
	0.18	10.83	26.92	24.82	18.30	18.95	100.00	
		CONTROLLING FOR GESLACHT=vrouw						
Frequency	Percent	17-24	25-34	35-44	45-54	55-64	65+	Total
Row Pct	Col Pct							
0	2	27	53	51	44	169	346	
	0.26	3.56	6.98	6.72	5.80	22.27	45.59	
	0.58	7.80	15.32	14.74	12.72	48.84		
	100.00	50.94	40.46	40.48	34.11	53.14		
1	0	26	78	75	85	149	413	
	0.00	3.43	10.28	9.88	11.20	19.63	54.41	
	0.00	6.30	18.89	18.16	20.58	36.08		
	0.00	49.06	59.54	59.52	65.89	46.86		
Total	2	53	131	126	129	318	759	
	0.26	6.98	17.26	16.60	17.00	41.90	100.00	

Uit Tabel 3 blijkt duidelijk dat niet alle personen die zeggen dat ze zullen meedoen aan de schriftelijke enquête, ook effectief meedoen. Bij de mannelijke gezinshoofden haakt 37% af, en bij de vrouwelijke gezinshoofden zelfs 46%. Deze tweede, schriftelijke, uitval is groter dan de eerste (degenen die aan telefoon reeds weigerden al dan niet na deelname aan de recruiteringsvragenlijst), en opnieuw meer uitgesproken bij vrouwelijke gezinshoofden dan bij mannelijke gezinshoofden: bij mannelijke gezinshoofden weigert 11% ogenblikkelijk (Tabel 2), en van degenen die overblijven haakt nog eens 37% af (Tabel 3). Bij vrouwelijke gezinshoofden weigert 24% ogenblikkelijk (Tabel 2), en van degenen die overblijven haakt nog eens 46% af (Tabel 3). In dit stadium is het afhaken wel afhankelijk van de leeftijd: hoe ouder het gezinshoofd (maar nog niet ouder dan 65) hoe kleiner de kans dat hij / zij afhaakt (Tabel 3). Oudere mannelijke gezinshoofden haken evenveel af als het gemiddelde, oudere vrouwelijke gezinshoofden haken vaker af dan het gemiddelde.

Telefonisch werd bij de recruiteringsvragenlijst reeds gevraagd hoeveel wagens het gezin in bezit heeft. Dat maakt dat van degenen die geen formulieren terugstuurden maar wel aan de recruiteringsvragenlijst hadden deelgenomen, we toch weten hoeveel wagens ze hebben.

**Tabel 4. Terugsturen van de schriftelijke vragenlijst volgens wagenbezit**

Het gaat hier enkel om die huishoudens die meegewerkt hebben met de telefonische vragenlijst, en *gezegd hebben* dat ze ook aan de schriftelijke wilden meedoen.

RQOK (Retour Questionnaire)=0 is negatieve reactie; RQOK=1 is positieve reactie  
RQOK WAGAANT

Frequency Percent Row Pct Col Pct	WAGAANT			Total
	0	1	2+	
0	223 5.40 13.95 49.01	904 21.89 56.57 38.81	471 11.40 29.47 34.99	1598 38.69
1	232 5.62 9.16 50.99	1425 34.50 56.28 61.19	875 21.19 34.56 65.01	2532 61.31
Total	455 11.02	2329 56.39	1346 32.59	4130 100.00

Personen die uiteindelijk toch afhaken hebben relatief vaker (bijna 50% i.p.v. de verwachte kleine 40%) geen wagen dan personen die wel verder reageren (Tabel 4).

### 3 Bepalen van gewichten

Een aantal huishoudens heeft niet gereageerd op de enquête. Zulke non-respons veroorzaakt bijna altijd een vertekening van de steekproef t.o.v. de populatie. In dit geval bevat de steekproef een overaanbod aan mannen en aan gehuwden, een tekort aan personen jonger dan 25 jaar en ouderen en een tekort aan alleenstaanden en aan grote gezinnen. Deze tekorten doen zich voor zowel bij het onderzoek naar de gezinshoofden met hun huishoudens, als bij de individuele personen. Opvallend is dat we dit vinden bij het stadsgewest Hasselt-Genk (dit document), maar ook voor het parallel onderzoek dat uitgevoerd werd in het stadsgewest Antwerpen (Zwerts & Nuyts, 2001).

Om deze vertekening zo goed mogelijk op te vangen, zijn aan personen en huishoudens gewichten toegekend. Indien we bijvoorbeeld van een bepaalde groep huishoudens er 10 in de steekproef verwacht hadden, en er hebben slechts 5 huishoudens gereageerd, dan krijgt elk van deze huishoudens een gewicht 2. Een gewicht groter dan 1 betekent dus een ondervertegenwoordiging van een groep in de steekproef. Alle analyses van dit onderzoek zijn gebaseerd op de gewogen gerealiseerde steekproef.

Tijdens de lente- en zomermaanden hebben er minder personen gereageerd dan verwacht. Om toch tot het vereiste totale aantal te komen zijn er in de herfst en in de winter meer personen ondervraagd. Voor meer details verwijzen we naar Nuyts & Zwerts (2000). Hierdoor hebben we in de totale steekproef echter een overaantal personen in de herfst- en wintermaanden. Ook hiervoor corrigeren we door aan de maanden gewichten toe te kennen.

Voor de berekening van de gewichten, die erg technisch is, verwijzen we naar de bijlage 8.1.

### 3.1 Effectief gebruikte gewichten huishoudens

We hebben vier reeksen van gewichten berekend om de afwijkingen te corrigeren. Door de wijze waarop we de gewichten berekend hebben, wijkt geen van de verdelingen van de vier variabelen nog significant af van de verdelingen van de populatie nadat we het product maken van de gewichten. De gewichten zijn de volgende:

**Tabel 5. Gewichten die aan de huishoudens zijn toegekend om de steekproef representatiever te maken**

		Gezinshoofd					
Gezinsleden		leeftijd		burg.staat		geslacht	
1 man	1.321	< 25	14.624	ongehuwd	0.822	man	0.967
1 vrouw	1.508	25-34	1.503	gehuwd	1.072	vrouw	1.128
2 personen	0.840	35-44	0.893	gescheiden	1.215		
3 personen	1.065	45-64	0.822	weduwe(naar)	0.701		
4 personen	0.815	65+	1.145				
5 personen	0.956						
6 personen	1.608						
7 personen	5.469						
8 personen	1.706						

De exacte interpretatie van de cijfers uit Tabel 5 is alleen 'in combinatie' zinvol. Bijvoorbeeld: het gewicht van '1 man alleen' van 1.321 wordt afgezwakt omdat het gezinshoofd per definitie dan een man is, en dus als geslacht-gewicht 0.967 heeft. Wat samen neer komt op  $1.321 \cdot 0.967 = 1.28$ , en dus op een ondervertegenwoordiging van alleenstaande mannen in de steekproef. Ander voorbeeld: een



gezin dat bestaat uit een gehuwd koppel met als gezinshoofd een man tussen 25 en 34 jaar krijgt gewicht  $0.840 \cdot 1.503 \cdot 1.072 \cdot 0,967 = 1.308$ .

Bemerk dat, indien men vergeet dat de gewichten *alleen als combinatie zinvol zijn*, men zelfs schijnbare tegenstellingen kan vinden. De gehuwden zijn in de steekproef oververtegenwoordigd: 71% van de steekproef tegenover 65% in de populatie. Indien we echter énkél kijken naar het gewicht van gehuwden: 1.072, dan lijkt het of ze ondervertegenwoordigd zijn. Doordat gezinnen van gehuwden echter steeds minstens met twee zijn, en meestal met drie of vier, wordt de schijnbare ondervertegenwoordiging over de ganse steekproef heen wel opgeheven.

### 3.2 Effectief gebruikte gewichten personen

**Tabel 6. Gewichten die aan de personen zijn toegekend om de steekproef representatiever te maken**

<b>0.9788</b>	Man	Man	Man	Man	Man	Man	Man	Man	Man
	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85+
Gehuwd	1	0.71	1.24	0.84	0.79	0.79	0.89	0.92	2.05
Gescheiden	1	1	1.46	1	0.94	0.94	1.05	1	1
Ongehuwd	1.13	1.05	1.82	1.24	1.17	1.17	1.31	1.36	3.08
Weduwnaar	1	1	1.27	0.86	0.81	0.81	0.91	0.94	2.09
<b>1.0217</b>	Vrouw	Vrouw	Vrouw	Vrouw	Vrouw	Vrouw	Vrouw	Vrouw	Vrouw
	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85+
Gehuwd	1	0.77	1.12	0.81	0.79	0.82	1.01	1.45	1
Gescheiden	1	1	1.03	0.74	0.73	0.76	0.93	1.34	1
Ongehuwd	1.18	0.99	1.57	1.13	1.10	1.15	1.42	2.04	6.52
Weduwe	1	1	1.35	0.97	0.95	0.99	1.22	1.75	5.60

De berekening van de gewichten voor de personen gebeurt in twee stappen (Tabel 6). Eerst corrigeren we voor een lichte scheef trekking van de geslachtsverhouding: voor mannen x **0.9788** en voor vrouwen x **1.0217**. Daarna vermenigvuldigen we deze waarden met de *schuin gedrukte* getallen uit Tabel 6.

Voorbeeld: gewicht van een gescheiden man van 31 jaar =  $0.9788 \cdot 1.46 = 1,43$ .

### 3.3 Effectief gebruikte gewichten verplaatsingen

Tabel 7. Gewichten die aan de invuldagen en maanden zijn toegekend om de steekproef representatiever te maken

Label Maand van Invuldag	Sample freq.	Pop.freq.	Pop.prop.	Expected freq.	Chi-2	Weight
1	593	31	0.0846995	587.052	0.0602668	0.989969
2	704	29	0.079235	549.178	43.647	0.780082
3	620	31	0.0846995	587.052	1.8492	0.946858
4	516	30	0.0819672	568.115	4.78063	1.101
5	531	31	0.0846995	587.052	5.35186	1.10556
6	562	30	0.0819672	568.115	0.0658146	1.01088
7	545	31	0.0846995	587.052	3.01228	1.07716
8	532	31	0.0846995	587.052	5.1626	1.10348
9	556	30	0.0819672	568.115	0.258341	1.02179
10	589	31	0.0846995	587.052	0.00646458	0.996693
11	548	30	0.0819672	568.115	0.712186	1.03671
12	635	31	0.0846995	587.052	3.91621	0.924491

De gewichten waarmee we verplaatsingen willen vermenigvuldigen, zijn berekend op personenniveau. Dit wil zeggen dat we nagaan of er in elke maand een voldoende aantal personen ondervraagd is die zich hadden kunnen verplaatsen. Deze personen krijgen elk hun gewicht mee, zoals bepaald bij 'gewichten personen'. Op deze wijze bekomen we voldoende gegevens per maand, maar worden maanden waarin mensen zich minder verplaatsen niet kunstmatig opgehoogd. Het gewicht voor de maand februari is 0.78 (Tabel 7), niet omdat er te veel verplaatsingen waren in februari, maar omdat in die maand we meer personen ondervraagd hebben. De verschillen tussen 1<sup>e</sup> en 2<sup>e</sup> invuldag zijn niet zo groot dat er andere daggewichten voor nodig zijn.

## 4 Technische aspecten i.v.m. de statistische verwerking

De bekomen data werden uitgezuiverd aan de hand van strikte regels (Nuyts & Zwerts 2000), en verwerkt met behulp van het statistische pakket SAS. De meeste resultaten zijn voorgesteld als frequentietabellen. Voor sommige vragen waarbij we willen weten op welke wijze een variabele afhangt van meerdere andere variabelen is lineaire of logistische regressie gebruikt.

### 4.1 Frequentietabellen en regressie

Voor de meeste analyses gebruiken we enkel frequentietabellen. Hiermee kunnen we het verband tonen tussen twee variabelen of, indien handig geschikt, eventueel tussen drie variabelen. In bepaalde gevallen willen we echter de invloed kennen die meerdere variabelen hebben op één andere variabele. Bijvoorbeeld: hoe hangt het aantal personenwagens af van het geslacht van het gezinshoofd, de leeftijd van het gezinshoofd, het gezinsinkomen en het aantal gezinsleden. Dit doen we via regressie.

In de OVG's passen we *multivariate* regressie toe: we proberen steeds een verband te leggen tussen 1 afhankelijke variabele en verscheidene onafhankelijke variabelen<sup>3</sup>. Afhankelijk van de mogelijke waarden van de afhankelijke variabele gebruiken we een ander 'type' regressie: *lineaire* regressie of *logistische* regressie.

<sup>3</sup> Multi-variate regressie in tegenstelling tot *univariate* regressie waar men 1 afhankelijke variabele probeert te begrijpen met behulp van 1 onafhankelijke variabele.

Lineaire regressie is de meest 'klassieke' regressie. Deze regressie is van de vorm:

$Y = aX_1 + bX_2 + \dots + cX_n + d$ , met Y de afhankelijke variabele en X1 tot Xn n onafhankelijke variabelen. De regressie heet lineair omdat alle variabelen, zowel de afhankelijke als de onafhankelijke, lineair gebruikt worden (=zonder er kwadraten of andere functies op toe te passen).

Indien we een regressie willen berekenen voor een variabele die enkel 'ja' of 'nee' kan zijn, zoals het bezit van een rijbewijs, dan kunnen we geen gewone lineaire regressie toepassen, maar wel een logistische regressie. De logistische regressie lijkt op een gewone regressie, maar op de afhankelijke variabele wordt eerst een logistische transformatie toegepast.

De regressie is van de vorm:

$$\ln\left(\frac{P}{1-P}\right) = aX_1 + bX_2 + \dots + cX_n + d,$$

P is dan de kans dat iemand een rijbewijs heeft.

In principe is regressie uitgevonden om continue variabelen te vergelijken met continue variabelen, b.v. lengte van armen i.f.v. totale lengte. Bij dit onderzoek zijn er echter verscheidene variabelen die opgedeeld zijn in klassen. We kunnen hier op verschillende manieren mee omgaan in de regressie.

- Indien de geklasseerde waarden oorspronkelijk continu waren, dan kunnen we de klassen vervangen door hun midden. Dit levert een (oplosbaar) praktisch probleem voor de laatste klasse, daar die in principe geen midden heeft. Een ander nadeel is dat men ervan uitgaat dat elke onafhankelijke variabele een lineaire invloed heeft op de afhankelijke variabele. Dit is in praktijk niet waar. Een stijging van het inkomen van 40.000 BEF per maand, indien men 20.000 BEF per maand verdient, heeft een totaal andere invloed op de mobiliteit van deze persoon dan een stijging van het inkomen van 40.000 BEF per maand, indien men reeds 200.000 BEF per maand verdient. Dit is dan wel op te vangen door de variabelen te transformeren, maar dan vermindert het inzicht in het uiteindelijke model.
- Een andere mogelijkheid is de klassen vervangen door een zelfbepaalde waarde. De waarde wordt zo bepaald, dat een univariate regressie van deze variabele zo performant mogelijk is. Indien de waarden handig gekozen worden verhoogt dit de performantie van het uiteindelijke model. Het nadeel is dat de bepaling van de waarden steeds iets arbitrairs heeft, en dat het uiteindelijke model moeilijker te interpreteren is.
- Een derde mogelijkheid is de k klassen vervangen door k-1 dummy variabelen (=ja/nee of 0/1 variabelen). Dit wil zeggen dat we één referentieklassen kiezen, en voor elke andere klasse een variabele die zegt of de waarneming ertoe behoort of niet. Dit heeft twee nadelen. Het aantal variabelen kan snel oplopen. Indien men echter genoeg waarnemingen heeft, en het regressiemodel eerder manueel opbouwt dan SAS de variabelen te laten kiezen, dan kan dit probleem omzeild worden. Een ander nadeel is dat soms alle onafhankelijke variabelen dummy variabelen zijn. Lineaire regressie veronderstelt dat alle variabelen samen een multivariate normaalverdeling hebben. Dat is erg onwaarschijnlijk indien alle variabelen dummy variabelen zijn. Het gevolg hiervan is dat de schatting van de coëfficiënten iets minder correct is dan verwacht kon worden. Het grote voordeel van het gebruik van dummy variabelen is dat elke klasse zijn eigen coëfficiënt krijgt, en dat het uiteindelijke regressiemodel erg inzichtelijk is.

We willen de regressiemodellen zo overzichtelijk mogelijk houden, en hebben daarom gekozen voor het gebruik van dummy variabelen.

Bij regressie kan men rekening houden met verschillende variabelen, bijvoorbeeld vrouw zijn, of jonger dan 25 jaar, maar ook met combinaties van dergelijke variabelen: vrouwen jonger dan 25 jaar. Hoe meer men combineert, hoe kleiner het aantal waarnemingen in de doelgroep. In principe houdt SAS bij de berekening van de relevantie (= significantie) van een bepaalde combinatie rekening met

het aantal waarnemingen in de betrokken groep. Kleine groepen hebben minder kans om significant te zijn. Om mogelijke overfitting<sup>4</sup> te voorkomen, hebben we op voorhand reeds groepen met minder dan 25 waarnemingen uitgesloten. We hebben eveneens getracht zoveel mogelijk beïnvloedende factoren te betrekken alhoewel dit niet steeds mogelijk is (b.v. de afstand tot een bepaalde bushalte is opgenomen in de regressie, de ritfrequentie van de bus(sen) evenwel niet). In die zin moeten we de regressieresultaten enigszins relativeren.

## 4.2 Betrouwbaarheid van de resultaten.

De fout bij de berekening van de parameters bij regressie kan men berekenen door een veelvoud te nemen van de standaarderror op deze parameter. De standaarderror  $\sigma$  wordt door SAS mee berekend en ook getoond in de output. De schatting van de parameter heeft ongeveer een normaal verdeling. Om bijvoorbeeld een betrouwbaarheidsinterval te krijgen van 95%, neemt men de schatting van de parameter  $\pm 1.96 \cdot \sigma$ .

Bij proporties geldt een vergelijkbaar resultaat. Men kan dan de standaarderror berekenen aan de hand van de bekomen proportie en het gebruikte aantal in de steekproef. Voor een proportie  $p$  en een steekproefaantal  $n$  vindt men met een betrouwbaarheid van 95% :

$$p \pm 1.96 * \sqrt{\frac{p * (1 - p)}{n}}.$$

Bij een steekproefaantal van 2000 waarnemingen, en een proportie van 0.10 vinden we dan

$$0.10 \pm 1.96 * \sqrt{\frac{0.10 * (0.90)}{2000}} = 0.10 \pm 0.013.$$

We hebben bij de steekproef dus een proportie gevonden van 0.10, en we zijn 95% zeker dat de proportie voor de populatie ligt tussen 0.087 en 0.113.

De onzekerheid stijgt hoe dichter de waarde van de proportie bij 0.5 ligt. Bij  $p=0.5$  vinden we  $0.5 \pm 0.022$ . De onzekerheid is dubbel zo groot als bij een proportie van 0.10. Vanzelfsprekend stijgt de fout ook bij een kleiner aantal waarnemingen.

Als globale regel kunnen we stellen dat we, zolang we 2000 waarnemingen hebben, voor alle proporties de onzekerheid van de proportie in de buurt ligt van 2%. Stijgt het aantal waarnemingen, of is de proportie verder verwijderd van 0.5, dan daalt de fout.

## 5 Vergelijking telefonisch/postaal bevroagden en enkel postaal bevroagden

Er zijn 3 groepen van mensen die enkel postaal bevroagd zijn:

- (1) Huishoudens waarmee geen telefonisch contact mogelijk is omdat ze geen vaste telefoonaansluiting hebben.
- (2) Huishoudens waarmee geen telefonisch contact mogelijk is omdat een fout nummer aan het betrokken gezin gekoppeld was.
- (3) Huishoudens waarmee geen telefonisch contact mogelijk is omdat ze weliswaar een vaste telefoonaansluiting hebben maar enkel een geheim nummer.

---

<sup>4</sup> Overfitting heeft plaats als men een variabele toevoegt die belangrijk lijkt, maar het eigenlijk niet is. Het is een variabele die voor deze steekproef significant is, maar dat bij een andere steekproef niet meer zou zijn. Door deze variabele toe te voegen lijkt het dus alsof de regressie verbetert, maar in werkelijkheid weten we daardoor niets meer over de populatie.

**Tabel 8. Gemiddeld aantal verplaatsingen per persoon per dag (gavppd) volgens postale en telefonische bevraging**

KLEUR	Frequency
oranje / Hasselt/Genk postaal	2.411036
groen / Hasselt/Genk telefonisch	2.850661

Deze gemengde postale groep verplaatst zich minder vaak (Tabel 8).

**Tabel 9. VMB-index<sup>5</sup> volgens postale en telefonische bevraging**

Frequency Percent Row Pct Col Pct	KLEUR(Kleur van de vragenlijst)						VMB(OVG VervoermiddelenIndex)	
	> 2 wagens	2 wagens	1 wagen	Motor/bromfiets	Fiets	Geen/overig	Total	
oranje / Hasselt /Genk postaal	8.6868 0.31 2.06 8.87	75.477 2.70 17.92 10.03	247.96 8.88 58.86 15.41	4.4516 0.16 1.06 25.48	51.979 1.86 12.34 25.71	32.711 1.17 7.77 28.64	421.26 15.08	
groen / Hasselt/Genk telefonisch	89.287 3.20 3.76 91.13	676.83 24.23 28.53 89.97	1361.4 48.74 57.39 84.59	13.018 0.47 0.55 74.52	150.17 5.38 6.33 74.29	81.515 2.92 3.44 71.36	2372.2 84.92	
Total	97.9739 3.51	752.302 26.93	1609.37 57.61	17.4696 0.63	202.153 7.24	114.226 4.09	2793.49 100.00	

**Tabel 10. Verdeling van het gavppd volgens hoofdvervoerswijze en volgens postale of telefonische bevraging**

Frequency Row Pct	KLEUR(Kleur van de vragenlijst)						HFDVM	
	(pre)metro	andere/onbep.	autobestuurder	autopassagier	brom/snorfietser	bus	Total	
oranje / Hasselt /Genk postaal	0.0007 0.03	0.2057 8.53	1.2308 51.05	0.4025 16.69	0.0159 0.66	0.0537 2.23	2.411	
groen / Hasselt/Genk telefonisch	0.001 0.03	0.1941 6.81	1.3341 46.80	0.5778 20.27	0.0254 0.89	0.0692 2.43	2.8507	

(Continued)

Frequency Row Pct	fietser	motorrijder	te voet	tram	trein	Total
oranje / Hasselt /Genk postaal	0.277 11.49	0.0086 0.36	0.1899 7.88	0.0016 0.07	0.0246 1.02	2.411
groen / Hasselt/Genk telefonisch	0.3511 12.31	0.0077 0.27	0.2656 9.32	0.0006 0.02	0.024 0.84	2.8507

<sup>5</sup> VMB-index = vervoermiddelenbezit-index; voor een definitie: zie Deel 2, bladzijde 12.

Uit Tabel 9 blijkt dat, hoewel de postaal bevroogden minder auto's hebben (21.16% (1,06 + 12,34 + 7,77) heeft geen auto tegenover 10.12% bij de telefonisch bevroogde huishoudens), ze relatief gesproken evenveel de wagen gebruiken als huishoudens die telefonisch bevroogd waren (Tabel 10: 67.74% (51,05 + 16,69) tegenover 67.07%). Dit wil zeggen dat de postaal bevroogde huishoudens met wagens, hun wagen gemiddeld meer gebruiken dan de huishoudens die telefonisch bevroogd zijn.

**Tabel 11. Gemiddeld aantal afgelegde kilometer per persoon per dag (gaakppd) volgens postale en telefonische bevroaging.**

	KLEUR	Frequency
oranje / Hasselt/Genk postaal		29.52969
groen / Hasselt/Genk telefonisch		31.87067

De postaal bevroogden leggen per dag minder kilometers af (Tabel 11), maar dit is volkomen te wijten aan het feit dat ze zich minder vaak verplaatsen. Per verplaatsing leggen ze gemiddeld zelfs iets meer kilometers af.

**Tabel 12. Verdeling van het gaakppd volgens hoofdvervoerswijze en volgens postale of telefonische bevroaging.**

Frequency Row Pct	KLEUR(Kleur van de vragenlijst)						Total
	(pre)metro	andere/onbep.	autobestuurder	autopassagier	brom/snorfietser	bus	
oranje / Hasselt /Genk postaal	0.0532 0.18	2.4642 8.34	17.356 58.77	4.8608 16.46	0.0663 0.22	0.6059 2.05	29.53
groen / Hasselt/Genk telefonisch	0.03 0.09	2.0348 6.38	17.094 53.63	7.6859 24.12	0.1311 0.41	1.0296 3.23	31.871

(Continued)

Frequency Row Pct	KLEUR(Kleur van de vragenlijst)					Total
	fietser	motorrijder	te voet	tram	trein	
oranje / Hasselt /Genk postaal	1.2718 4.31	0.2763 0.94	0.4076 1.38	0.0047 0.02	2.1632 7.33	29.53
groen / Hasselt/Genk telefonisch	1.3904 4.36	0.12 0.38	0.5097 1.60	0.0037 0.01	1.8417 5.78	31.871

Inzake de verdeling van de kilometers per dag over de vervoerswijzen is er geen verschil tussen de postale huishoudens en de telefonische huishoudens (Tabel 12).

## 6 Minder antwoorden op de 2<sup>e</sup> invuldag

Tabel 13. Verschillen tussen 1<sup>e</sup> en 2<sup>e</sup> invuldag

	1 <sup>e</sup> invuldag	2 <sup>e</sup> invuldag	2 <sup>e</sup> dag als % van 1 <sup>e</sup> dag
Aantal respondenten	6708	6578	98%
Gavppd	2.90	2.70	93%
Gaakppd	32.24	30.95	96%

De respondenten moeten gedurende 2 dagen hun verplaatsingen noteren. De 1<sup>e</sup> invuldag is ingevuld door 6.708 personen, de 2<sup>e</sup> door 6.578 personen, dat is 98% van het aantal dat de eerste dag heeft ingevuld. 2% van de respondenten haakt dus af na de 1<sup>e</sup> invuldag. Hetzelfde resultaat is gevonden voor Antwerpen (Zwerts et al., 2001).

Er zijn gemiddeld 2.90 verplaatsingen per persoon per dag op basis van de data van de 1<sup>e</sup> invuldag, er zijn gemiddeld slechts 2.71 verplaatsingen per persoon per dag voor de 2<sup>e</sup> invuldag, dit is 93% van het aantal van de 1<sup>e</sup> invuldag. Het gemiddeld aantal kilometer per persoon per dag van de 2<sup>e</sup> invuldag is 96% van dat van de 1<sup>e</sup> invuldag. In werkelijkheid is er, gemiddeld, echter geen enkel verschil tussen de 1<sup>e</sup> en de 2<sup>e</sup> invuldag: de proportie wekdagen is dezelfde, de geslachtsverhouding, de verdeling over de maanden, de proportie studenten, enzovoorts. Het enige verschil is dat het voor de respondent de 2<sup>e</sup> invuldag is. M.a.w. de respondenten die ook de 2<sup>e</sup> dag invullen worden slordiger en noteren hun verplaatsingen minder exact.

Dat het aantal mensen van de tweede invuldag kleiner is, is niet erg. In de berekeningen worden gemiddelden genomen voor het totaal aantal personen van de 1<sup>e</sup> invuldag samen met het totaal aantal van de 2<sup>e</sup> invuldag. De slordigheid van het invullen die we hierboven concluderen, kunnen we echter niet goed corrigeren. Willen we corrigeren voor verplaatsingen, dan moeten we rekening houden met 93%, voor kilometers echter met 96%, en indien we andere variabelen zouden nagaan, dan vinden we ongetwijfeld nog andere percentages. De percentages kunnen bovendien vrij ver uit elkaar liggen. Voor het stadsgewest Antwerpen vonden we voor verplaatsingen een factor 93%, maar voor het aantal kilometers een factor 85% (Zwerts et al., 2001). Onze conclusies, gebaseerd op beide invuldagen, zijn waarschijnlijk onderschattingen.

Anderzijds is deze 'slordigheid' toch wel van belang voor de resultaten van het onderzoek. Stel dat we in dit onderzoek slechts één invuldag hadden genomen, dan was het gemiddeld aantal verplaatsingen ruwweg 10% groter geweest. In het Onderzoek Verplaatsingsgedrag van België van 1999 (Toint et al., 2000) werd slechts één invuldag gebruikt (cijfer gevonden van 3,04). Vergelijking met deze cijfers is dan eigenlijk niet meer mogelijk. Bij het eerste OVG Vlaanderen werd de keuze gemaakt om twee invuldagen te nemen, om op deze wijze voldoende gegevens over de verplaatsingen te bekomen zonder het aantal geënquêteerden drastisch te moeten verhogen. Op basis van deze resultaten moeten we overwegen of het gebruik van deze tweede invuldag wel opportuun is.

## 7 Bibliografie

- \* Hajnal, I. (1995) "Weight 2.1 voor Windows. Een programma voor het herwegen van steekproeven". Bulletin 1995/58 van het centrum voor dataverzameling en analyse. Leuven, 15 p.
- \* Hajnal I. & Miermans W. (1995) "Onderzoek Verplaatsingsgedrag Vlaanderen. Controle en Begeleidingsopdracht. Eindverslag". Provinciaal Hoger Architectuur Instituut Diepenbeek, Hogeschool voor Verkeerskunde, 53p.
- \* Hajnal, I. & Miermans, W. (1996a) "Onderzoek Verplaatsingsgedrag Vlaanderen. Analyse opdracht. Eindverslag". Provinciale Hogeschool Limburg, Departement Architectuur en Hogeschool voor Verkeerskunde, Diepenbeek, 168p.
- \* Hajnal, I. & Miermans, W. (1996b) "Onderzoek Verplaatsingsgedrag Vlaanderen. Analyse opdracht. Documentatie van het onderzoek verplaatsingsgedrag Vlaanderen 1994-1995. Versie 2.0". Provinciale Hogeschool Limburg, Departement Architectuur en Hogeschool voor Verkeerskunde, Diepenbeek, 143p.
- \* Nationaal Instituut voor Statistiek. (1999a) "Bevolkingsstatistieken. Huishoudens en familiekeren op 1.1.1999". N.I.S., Brussel, 163p.
- \* Nationaal Instituut voor Statistiek. (1999b) "Bevolkingsstatistieken. Totale en Belgische bevolking op 1.1.1999". N.I.S., Brussel, 237p.
- \* Nuyts, E. & Zwerts, E. (2000). "Onderzoek Verplaatsingsgedrag Stadsgewest Hasselt-Genk, Stadsgewest Antwerpen, Controle en begeleidingsopdracht, eindverslag". Onderzoeksceel Architectuur en Mobiliteit, Provinciale Hogeschool Limburg, Departement Architectuur, Diepenbeek, 126p.
- \* Toint, P., Barette, P. & Dessy, A. (2000). "Enquête nationale sur la mobilité des ménages (1998-1999). Résumé méthodologique." In congresdocument van Studiedag Duurzame Mobiliteit 30 maart 2000.
- \* Zwerts, E. (2000). "Bijkomende Analyses Onderzoek Verplaatsingsgedrag Vlaanderen. 1994-1995". Onderzoeksceel Architectuur en Mobiliteit, Provinciale Hogeschool Limburg, Departement Architectuur, Diepenbeek, 54p.
- \* Zwerts, E. & Nuyts, E. (2001). "Onderzoek Verplaatsingsgedrag Stadsgewest Antwerpen, Methodologische analyse, eindverslag". Onderzoeksceel Architectuur en Mobiliteit, Provinciale Hogeschool Limburg, Departement Architectuur, Diepenbeek.



## 8 Bijlage

### 8.1 Berekening van de gewichten

#### 8.1.1 Stappenplan voor meerdere marginale verdelingen

Hajnal (1995) heeft een programma geschreven dat gewichten berekent voor een steekproef indien er gewogen wordt (a) o.b.v. één variabele, of (b) o.b.v. een willekeurig aantal variabelen waarvan zowel voor populatie als voor steekproef de volledige gezamenlijke verdeling bekend is, of (c) o.b.v. twee variabelen waarvan enkel de marginale verdelingen bekend zijn voor de populatie, en de gezamenlijke verdeling voor de steekproef. Bij de analyse van OVG Hasselt-Genk hadden we echter vier variabelen die van belang waren, waarvan enkel de marginale verdelingen bekend zijn voor de populatie: geslacht, leeftijd en burgerlijke staat van het gezinshoofd, en het aantal gezinsleden van het huishouden. We hebben dus te weinig data over de gezamenlijke verdeling van de populatie om (b) te gebruiken, en te veel variabelen om op (c) terug te kunnen vallen. Daarom hebben we de methode van Hajnal moeten aanpassen tot volgende werkwijze.

- 1 Zoek de verdeling van de populatie op in de publicaties van het Nationaal Instituut van de Statistiek. We vinden voor de huishoudens de marginale verdelingen per burgerlijke staat, geslacht, leeftijdsklasse van het gezinshoofd en ledenaantal (N.I.S. 1999a), en voor de personen de gezamenlijke verdeling van geslacht en burgerlijke staat, en de gezamenlijke verdeling van geslacht en leeftijdsklassen (N.I.S. 1999b). Sommeer in Excel de gemeenten per stadsgewest. Door deze sommatie kunnen we niet meer zorgen dat de steekproef evenredig over alle gemeenten verdeeld is. Indien we echter én de verdeling van de gemeenten, én de sociologische verdelingen van hierboven willen bekomen, dan vinden we in de tabellen erg veel cellen met een verwachte celfrequentie kleiner dan 5. Hierdoor worden de statistische testen onbruikbaar. Aangezien we het aantal variabelen moeten beperken, en aangezien we menen dat de mobiliteit van huishoudens eerder door de sociologische kenmerken dan door de gemeentegrenzen bepaald wordt, kiezen we ervoor om de gemeenten samen te voegen.
- 2 Bereken voor de steekproef de marginale verdelingen voor betrokken variabelen, b.v. met SAS.
- 3 Combineer de marginale verdelingen van populatie en steekproef in Excel tot een bruikbare input voor Weight 2.1.
- 4 Bereken voor elke variabele apart de  $\chi^2$  van de verdeling van de steekproef t.o.v. de populatie via Weight 2.1 (Hajnal 1995).
- 5 Neem de variabele V1 met de kleinste P-waarde. Bepaal hiervoor de gewichten via Weight 2.1. Gebruik deze gewichten als een eerste benadering Weeg1 van de uiteindelijke gewichten in SAS.
- 6 Bepaal voor de steekproef via SAS o.b.v. deze gewichten de nieuwe marginale verdelingen voor de andere variabelen.
- 7 Voer voor elk van de variabelen die nieuwe marginale verdelingen in in Weight 2.1. Dit is het eenvoudigste via een tussenstap via een Excel omschrijving (zie stap 3) .
- 8 Je krijgt voor alle variabelen nu opnieuw de  $\chi^2$  en de P-waarde van de vergelijking tussen de marginale steekproefverdeling en de marginale populatieverdeling. Bepaal de gewichten van de variabele V2 die nu de kleinste P-waarde heeft. Dit geeft je gewichten Weeg2.
- 9 In SAS bereken je opnieuw voor alle variabelen een nieuwe marginale verdeling deze keer o.b.v. gewichten Weeg1\*Weeg2.

- 10 Deze nieuwe reeks verdelingen geef je weer via Excel in in Weight 2.1 Ook voor de eerste variabele V1, want diens 'ideale' gewicht Weeg1 is verschoven door het toevoegen van Weeg2. Bemerkt dat de laatst gewijzigde variabele, hier V2, niet extra hoeft ingegeven te worden, want die heeft een 'ideaal' gewicht  $Weeg1 * Weeg2$ . Je berekent opnieuw voor elke variabele apart de  $\chi^2$  en de P-waarde van het verschil tussen de marginale steekproefverdeling en de marginale populatieverdeling.
- 11 Zo blijf je bezig tot voor alle variabelen er geen significant verschil is tussen de populatie en steekproefverdelingen.

Het is niet vanzelfsprekend, maar wel waarschijnlijk dat na verloop van tijd de wegingen convergeren naar niet-significante verschillen. Intuïtief zou ik zeggen dat dit moet lukken als de afwijkingen tussen steekproef en populatie tussen de variabelen onderling ofwel niet-, ofwel positief gecorreleerd zijn. In het laatste geval helpt een aanpassing van de gewichten van de ene variabele om dichter bij de populatie te komen voor de andere variabele.

Ook al vind je in het begin een variabele die niet significant afwijkt, dan moet je die variabele toch meenemen in het proces, omdat die door de wijziging in gewichten voor andere variabelen toch kan beginnen afwijken.

Indien je maar twee variabelen hebt, dan kan je via IPF in Weight 2.1 dit proces automatisch laten lopen. Het kost je veel minder werk, en het resultaat is nauwkeuriger. Jammer genoeg hadden we voor de huishoudens minstens vier relevante variabelen.

## **8.1.2 Huishoudens: vier relevante variabelen**

### *8.1.2.1 Berekening van de huishouden gewichten*

Via de publicatie van het N.I.S. (1999a) beschikken we over vier variabelen die relevant (zouden kunnen) zijn. Het lijkt voorzichtig om die, tot we zeker zijn dat ze niet relevant zijn, allemaal mee te nemen. Deze gegevens combineren we niet met gegevens uit andere publicaties dan die van het NIS, omdat we dan niet zeker zijn dat die over dezelfde populatie handelen. Zouden we dat toch doen, dan trekken we de verdeling misschien nog schever i.p.v. ze representatiever te maken.

Elk van de vier variabelen vertoonden in het stappenplan zoals hierboven beschreven eenmaal de meest afwijkende verdeling t.o.v. de populatieverdeling had. Na één 'ronde', waarbij elke variabele zijn factor aan het uiteindelijke gewicht toevoegde, week geen van de marginale verdelingen nog significant af van de marginale verdelingen van de populatie. De iteratie ging sneller dan gehoopt, want we hadden gevreesd dat we verscheidene 'rondes' zouden moeten maken.

Het eindresultaat van deze berekening zijn vier series gewichten die met elkaar vermenigvuldigd moeten worden om het uiteindelijke gewicht te bekomen. De resultaten zijn ook zo getoond in Tabel 5.

### *8.1.2.2 Test van een tweede ronde - conclusies*

Uit nieuwsgierigheid<sup>6</sup> hebben we ook nog een tweede 'ronde' uitgewerkt, om te zien hoeveel deze ronde de resultaten verbeterde (Tabel 14).

---

<sup>6</sup> Uit nieuwsgierigheid, want voor het onderzoek was het strikt gezien niet meer nodig, omdat we geen significante afwijkingen meer vonden.

**Tabel 14. P-waarden van de afwijkingen tussen de marginale verdelingen van de vier variabelen om de gewichten te bepalen voor de steekproef en de populatie voor Hasselt-Genk.**

	Na 1 <sup>e</sup> volledige 'ronde'	Na 2 <sup>e</sup> volledige 'ronde'
Geslacht	0.98	0.89
Leeftijdsklasse	0.61	1
Burgerlijke staat	0.16	0.29
Leden aantal klasse	0.90	0.57

De gewichten zijn bepaald in 'rondes' waarbij elke variabele in één ronde een factor aan het gewicht kan toevoegen. Vanuit een globaal standpunt is de kleinste van de vier P-waarden de bepalende. Indien die kleiner is dan 5%, of 10% indien we conservatief zijn, dan moeten de gewichten nog aangepast worden. Hoe groter de kleinste P-waarde, hoe beter de steekproef de populatie benadert. Voor dit globale standpunt is het niet van belang bij welke variabele een bepaalde P-waarde hoort. Een P-waarde van 0.01 bij 'geslacht' is even storend als een P-waarde van 0.01 bij 'leeftijdsklasse'. In Tabel 14 zien we de P-waarden na de 1<sup>e</sup> en na de 2<sup>e</sup> volledige ronde. Na de 2<sup>e</sup> ronde is de kleinste P-waarde verschoven van 0.16 naar 0.29. De andere P-waarden zijn ruwweg gelijk gebleven: 0.61 -> 0.57; 0.90 -> 0.88 en 0.98-> 0.99 (weliswaar niet bij dezelfde variabele, maar dat is hier niet van belang). We hebben dus winst geboekt met deze tweede ronde. Maar deze winst is beperkt, en de vereiste procedure is erg arbeidsintensief. Al bij al is een tweede ronde voor de analyse van de data de extra moeite niet waard.

Hieronder volgen de P-waarden per tussenstap, om een idee te geven van het verloop van de convergentie.

De P-waarde van de laatst gewijzigde variabele is niet altijd berekend. In theorie zou die steeds 1 moeten zijn, in praktijk ligt die door afrondingsfouten steeds boven 0.95.

Stap 1: geen gewichten

geslacht Probability: 5.85623e-014

leeftkl Probability: **2.67197e-035**

bssamen<sup>7</sup> Probability: 2.90142e-011

ledenkl Probability: 1.13376e-028

Start van de eerste ronde.

Stap 2 na aanpassing leeftijdsklasse

geslacht Probability: 4.56545e-011

ledenkl Probability: **1.78907e-026**

bssamen Probability: 1.1415e-008

Stap 3 na aanpassing leden aantal klasse

geslacht Probability: 0.0638474

bssamen Probability: **8.11541e-011**

leeftkl Probability: 0.755581

Stap 4 na aanpassing burgerlijke staat

geslacht Probability: **0.000989587**

leeftkl Probability: 0.436558

ledenkl Probability: 0.622823

<sup>7</sup> Bssamen = Burgerlijke staat (ongehuwd en samenwonend samen genomen)

Stap 5 na aanpassing geslacht (cfr. Tabel 14: na 1<sup>e</sup> volledige 'ronde')

geslacht	Probability: 0.975596
leeftkl	Probability: 0.610053
bssamen	Probability: <b>0.159459</b>
ledenkl	Probability: 0.902992

Einde van de eerste ronde.

Begin tweede ronde.

Stap 6: aanpassing bssamen

geslacht	Probability: <b>0.132169</b>
leeftkl	Probability: 0.481567
bssamen	Probability: 0.966902
ledenkl	Probability: 0.491564

Bemerk dat we op het einde van de eerste ronde reeds hadden:  $P=0.98 \ 0.61 \ 0.16 \ 0.90$   
Na de eerste stap van de volgende ronde, is dit niet verbeterd. We vinden  $P=0.13 \ 0.48 \ 0.97 \ 0.49$   
De minimale P-waarde is nu kleiner, namelijk 0.13 i.p.v. 0.16.

Stap 7 na aanpassing geslacht

Leeftkl	Probability: 0.572875
bssamen	Probability: 0.80868
ledenkl	Probability: <b>0.489446</b>

Bemerk dat dit het beste resultaat is dat we zullen bekomen: alle P-waarden liggen rond de 50% of hoger. Indien we consequent de ronde verder af werken dan bereiken we geen minimale P-waarde meer rond de 50%.

Stap 8 na aanpassing ledenkl

geslacht	Probability: 0.976083
leeftkl	Probability: <b>0.250615</b>
bssamen	Probability: 0.599859

Stap 9 na aanpassing leeftkl (cfr. Tabel 14: na 2<sup>e</sup> volledige 'ronde')

geslacht	Probability: 0.888683
leeftkl	Probability: 0.999998
bssamen	Probability: <b>0.294637</b>
ledenkl	Probability: 0.571828

Conclusies i.v.m. de iteraties:

- De iteraties gaan sneller dan verwacht, na een eerste ronde kunnen al heel behoorlijke resultaten bekomen worden.
- Bij elke stap kan de iteratie weer verslechteren (stap 6 slechter dan stap 5, stap 8 slechter dan stap 7).
- Het beste resultaat bekom je niet per se op het einde van een 'ronde'.
- Moraal: stop maar zodra je een degelijk resultaat hebt, waarbij alle P-waarden groter zijn dan 0.10.

### **8.1.3 Personen: tweemaal een gezamenlijke verdeling van vier variabelen**

We beschikken voor de populatie over de gezamenlijke verdeling van geslacht en burgerlijke staat, en de gezamenlijke verdeling van geslacht en leeftijdsklassen (N.I.S. 1999b). Indien we dit herschikken tot een mannelijke en een vrouwelijke deelpopulatie, dan hebben we voor deze deelpopulaties twee marginale verdelingen, n.l. die van burgerlijke staat en die van leeftijdsklassen. Voor de steekproef beschikken we per deelpopulatie over de gezamenlijke verdeling van burgerlijke staat en leeftijdsklassen. Zodoende beschikken we per deelpopulatie over alle gegevens om de Iterative Proportional Fitting –module van Weight 2.1 te gebruiken (Hajnal 1995). De output hiervan

zijn gewichten voor elke combinatie van leeftijdsklasse en burgerlijke staat. Indien we die corrigeren voor de vertekening van geslacht in de steekproef, dan bekomen we de uiteindelijke gewichten.

Bemerk dat de gewichten van cellen die intrinsiek een frequentie nul hebben, zoals bijvoorbeeld het aantal gehuwde kinderen onder de 14 jaar bij deze berekening steeds gelijk blijven aan 1. Het programma corrigeert geen gewichten van lege cellen.

#### **8.1.4 Verplaatsingen: een verdeling van één variabele**

We willen dat de invuldagen gelijkmatig verspreid zijn over de weekdays en dat ze evenredig verdeeld zijn over de maanden. De gelijkmatige verdeling over de weekdays was in orde. Dit blijkt in praktijk vrij spontaan te verlopen, want dit was zonder correcties reeds in orde voor Antwerpen, Hasselt-Genk (maar ook voor OVG Vlaanderen 1 1994-1995, waar er toch een andere enquêteur was).

We moeten alleen corrigeren om een evenredige verdeling over de maanden te bekomen. Dit kan het handigste met de standaard module van Weight 2.1, die bedoeld is voor het berekenen van gewichten indien slechts één variabele van belang is.

### **8.2 Berekening van de ophoogfactor**

In bepaalde situaties willen we de gegevens van de steekproef ophogen naar de populatie. De gebruikte ophoogfactor = populatie aantal vanaf 6 jaar/ steekproef aantal.

Men kan eventueel delen door 'gewogen steekproef aantal' i.p.v. door 'steekproef aantal'. Zonder afrondingsfouten bij de berekeningen zou het gewogen steekproef aantal en het gewone steekproef aantal hetzelfde moeten zijn. De verschillen tussen beide zijn hoe dan ook beperkt.

### **8.3 Samenvoegen van gegevens**

De antwoorden van de respondenten zijn voor statuut en doel teruggebracht naar de oorspronkelijke categorieën.

#### **8.3.1 Statuut**

1='scholier, student'  
2='werkzaam in het eigen huishouden'  
3='werkloos'  
4='gepensioneerd'  
5='arbeidsongeschikt'  
6='arbeider'  
7='bediende'  
8='kader'  
9='vrij beroep'  
10='zelfstandige'  
11='andere, NIET beroepsactief'  
12='andere, WEL beroepsactief'

13= 'leraar'	7=bediende
14= 'tbs'	11='andere niet beroepsactief'
15= 'verpleegkundige'	7=bediende
16= 'ingenieur'	8=kader
17= 'vertegenwoordiger'	7=bediende
18= 'meewerkende echtgenote'	10=zelfstandige
19= 'ambtenaar'	7=bediende
20= 'musicus'	12=andere, wel beroepsactief
21= 'opvoeder'	7=bediende
22= 'rijkswachter/politieagent'	7=bediende

23= 'chauffeur'	6=arbeider
24= 'wetenschappelijk onderzoeker'	7=bediende
25= 'doctoraatstudent'	7=bediende
26= 'beperkte tewerkstelling'	12=andere, wel beroepsactief
27= 'onthaal moeder'	12=andere, wel beroepsactief
28= 'hoofd magazijnier'	6=arbeider
29= 'poetsvrouw'	6=arbeider
30= 'tuinier'	6=arbeider
31= 'pastoor'	7=bediende
32= 'bestuurder vennootschap'	8=kader
33= 'journalist'	12=andere, wel beroepsactief
34= 'verzorgender'	6=arbeider
35= 'militair'	12=andere, wel beroepsactief
36= 'werkster ocmw'	6=arbeider
37= 'interim'	12=andere, wel beroepsactief
38= 'beheerder bedrijf'	8=kader
39= 'zelfstandige helper'	6=arbeider
40= 'militair (zie code 35)'	12=andere, wel beroepsactief
41= 'huisbewaarster'	6=arbeider
42= 'kelner'	6=arbeider
43= 'magistraat'	8=kader
44= 'bedienaar'	12=andere, wel beroepsactief
45= 'sportman'	12=andere, wel beroepsactief
46= 'kunstenaar'	12=andere, wel beroepsactief
47= 'helper'	6=arbeider
48= 'opleiding VDAB'	11=andere, niet beroepsactief
49= 'directeur'	8=kader
50= 'leercontract'	1=student
51= 'bijverdienste'	11=andere, niet beroepsactief
52= 'freelance'	10='zelfstandige'
53= 'luchtvaart'	12=andere, wel beroepsactief
54= 'beschermde\beschuttende werkplaats'	12=andere, wel beroepsactief
55= 'beenhouwer'	10='zelfstandige'
56= 'hostess'	12=andere, wel beroepsactief
57= 'geco-statuut'	7=bediende
71= 'sociaal plan'	11=andere, niet beroepsactief
72= 'ouderschap verlof'	11=andere, niet beroepsactief
73= 'zwangerschap verlof'	11=andere, niet beroepsactief
74= 'loopbaan onderbreking'	11=andere, niet beroepsactief
75= 'invalide'	5='arbeidsongeschikt'
76= 'disponibiliteit'	11=andere, niet beroepsactief
77= 'cao + IBSSS'	11=andere, niet beroepsactief
78= 'prestatie vrije opzet';	10='zelfstandige'

### 8.3.2 Doel

1='naar huis'
2='werken'
3='winkelen'
4='zakelijk bezoek'
5='iemand een bezoek brengen'
6='onderwijs volgen'
7='wandelen/rondrijden'
8='iemand brengen/halen'
9='ontspanning/sport/cultuur'
10='diensten (dokter, bank)'
11='andere'

12 = 'caravan'	9 = 'ontspanning/sport/cultuur'
13 = 'vrijwilligerswerk'	5 = 'iemand een bezoek brengen'
14 = 'benzine station\gaan tanken'	3 = 'winkelen'
15 = 'verblijfplaats'	11 = 'andere'
16 = 'iets opbrengen\afhalen'	11 = 'andere'
17 = 'camping'	9 = 'ontspanning/sport/cultuur'
18 = 'iets gaan eten'	9 = 'ontspanning/sport/cultuur'
19 = 'naar de kerk\mosquee'	11 = 'andere'
20 = 'velt 25j. antwerpen'	11 = 'andere'
21 = 'containerpark'	10 = 'diensten (dokter, bank)'
22 = 'kot'	11 = 'andere'
23 = 'naar garage'	3 = 'winkelen'
24 = 'wassalon'	3 = 'winkelen'
25 = 'hond uitlaten'	9 = 'ontspanning/sport/cultuur'
26 = 'iemand helpen'	5 = 'iemand een bezoek brengen'
27 = 'huis van vertrek'	11 = 'andere'
28 = 'vergadering'	11 = 'andere'
29 = '2de verblijf'	11 = 'andere'
30 = 'carwash'	3 = 'winkelen'
31 = 'monitor sportkamp'	9 = 'ontspanning/sport/cultuur'
32 = 'babysit'	11 = 'andere'
33 = 'op vakantie'	9 = 'ontspanning/sport/cultuur'
34 = 'verzorging dieren'	10 = 'diensten (dokter, bank)'
35 = 'vergadering (zie code 28)'	11 = 'andere'
36 = 'feest'	9 = 'ontspanning/sport/cultuur'
37 = 'rijles'	3 = 'winkelen'
38 = 'naar luchthaven'	11 = 'andere'
39 = 'begrafenis'	11 = 'andere'
40 = 'werken in de nieuwe woning/naar nieuw huis'	11 = 'andere'
41 = 'gaan doppen'	10 = 'diensten (dokter, bank)'
42 = 'naar een banket'	9 = 'ontspanning/sport/cultuur'
43 = 'wagen halen'	3 = 'winkelen'
44 = 'scouts'	9 = 'ontspanning/sport/cultuur'
45 = 'naar feitelijk adres'	11 = 'andere'
46 = 'adres van bezoek'	5 = 'iemand een bezoek brengen'
47 = 'tuinklus'	11 = 'andere'
48 = 'verkiezingen'	11 = 'andere'
49 = 'stage'	6 = 'onderwijs volgen'
50 = 'bloed geven'	11 = 'andere'
51 = 'naar school (examen\proclamatie)'	6 = 'onderwijs volgen'
52 = 'autocontrole'	3 = 'winkelen'
53 = 'cursus\opleiding'	11 = 'andere'
54 = 'kerkhof'	11 = 'andere'
55 = 'sollicitatie'	11 = 'andere'
56 = 'begeleiding'	11 = 'andere'
57 = 'iemand vergezellen'	5 = 'iemand een bezoek brengen'
58 = 'verhuizen'	11 = 'andere'
59 = 'funerarium'	11 = 'andere'
60 = 'koffietafel'	11 = 'andere'
61 = 'instelling'	11 = 'andere'
62 = 'catechese\godsdienst';	11 = 'andere'

## **8.4 Vragenlijsten**