provincie Limburg

PROVINCIALE HOGESCHOOL LIMBURG

DEPARTEMENT ARCHITECTUUR

ONDERZOEKSCEL

Architectuur **M**obiliteit **O**mgeving

ONDERZOEK VERPLAATSINGSGEDRAG VLAANDEREN 2 (januari 2000 - januari 2001)

DEEL 1: METHODOLOGISCHE ANALYSE

Onderzoek in opdracht van het
ministerie van de Vlaamse Gemeenschap
departement Leefmilieu en Infrastructuur, Mobiliteitscel

ENID ZWERTS - ERIK NUYTS
Diepenbeek, Brussel, januari 2004
D/2004/3241/016

Inhoud

1 Lijst van tabellen

2 Inleiding

Tijdens de periode januari 2000 tot januari 2001 werden er gegevens verzameld over een aantal mobiliteitskenmerken van gezinnen en personen vanaf 6 jaar in Vlaanderen waaronder het verplaatsingsgedrag van personen. De steekproef voor deze studie bestond uit 3.028 gezinnen uit Vlaanderen.

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 3.028 huishoudens die de formulieren ingevuld hadden. Omdat niet alle huishoudens meedoen aan het onderzoek, begonnen we met een steekproef van 5.000 huishoudens gestratifieerd volgens de leeftijd van het gezinshoofd. Deze steekproef werd getrokken uit het Rijksregister in december 1999. Een tweede steekproef van 1.215 huishoudens werd eind juni 2000 bezorgd, en half augustus 2000 werd opnieuw een grotere steekproef van 5.000 huishoudens geleverd. Deze laatste steekproef werd niet volledig opgebruikt.

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 van het aantal personenvragenlijsten dat er naar het huishouden moet opgestuurd worden. Indien er geen vaste telefoon was (of in geval 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 en Beeldende Kunst).

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.

Vervolgens bespreken we een aantal *onderzoeksresultaten i.f.v. de gehanteerde contactprocedure.*

Tenslotte bespreken we nog de gehanteerde *statistische technieken en de betrouwbaarheid/nauwkeurigheid* die het gevolg is van deze technieken.

De bijlage bevat:

- een meer technische uitleg over de berekening van de gewichten, de ophoogfactor en het samenvoegen van gegevens;

- de vragenlijsten.

3 Non-respons

3.1 Bereidwilligheid om vragenlijsten in te vullen (unit non-respons)

Een aantal huishoudens weigerden mee te werken aan de enquête, of stuurden onvoldoende formulieren terug. In Tabel 1 geven we de responspercentages van OVG Vlaanderen en OVG Gent dat tegelijkertijd en met dezelfde methodiek werd uitgevoerd.

Tabel 1. Responspercentages van OVG Vlaanderen en OVG Gent

	Vlaanderen	**Gent**
Periode van uitvoering	**Januari 2000- januari 2001**	**Januari 2000- januari 2001**
Telefonisch		
Aantal telefoonnummers gevonden	100%	100%
Aantal telefonische contacten	84%	78%
Deelname recruteringsvragenlijst	80%	74%
Akkoord verdere deelname	66%	61%
Aantal bundeltjes terug	41%	38%
Aantal goedgekeurde bundeltjes*	30%	29%
Aantal goedgekeurde bundeltjes**	4%	3%
Postaal		
Verzonden	100%	100%
Aantal bundeltjes terug	22%	19%
Aantal goedgekeurde bundeltjes*	15%	14%
Aantal goedgekeurde bundeltjes**	2%	1%

* 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 %.

We bespreken eerst de '**telefonische gezinnen'**. Voor huishoudens met telefoon is de eerste contactmogelijkheid de telefonische vraag voor medewerking, waarbij ook reeds enkele vragen over het huishouden gesteld worden (= recruteringsvragenlijst).

Bij 16% van de Vlaamse huishoudens die wel telefoon hebben, werd nooit opgenomen. 4% andere huishoudens werden wel bereikt, maar weigerden medewerking. Er bleef dus 80% van de huishoudens over die meewerkten aan de telefonische enquête (voor details van de absolute aantallen, zie Nuyts & Zwerts (2001b)). Uiteindelijk stuurt 41 % van de huishoudens de bundeltjes terug.

34% (=30% + 4%) van de huishoudens sturen hun bundeltjes terug met een volledig aantal of een voldoende aantal personenvragenlijsten .

In vergelijking met Gent zien we dat er in Vlaanderen meer telefonische contacten kunnen gelegd worden (84 % i.p.v. 78 %, een verschil van 6 %). De rest van het proces verloopt evenwel iets slechter dan in Gent vermits we uiteindelijk 34 % huishoudens hebben die hun bundeltjes volledig of met een voldoende aantal personenvragenlijsten terugsturen en in Gent 32 % (een verschil van nog maar 2 %).

Huishoudens die telefonisch niet bereikt kunnen worden (de 'postale gezinnen') reageren veel minder dan telefonisch wel bereikte huishoudens. Uiteindelijk stuurt 22% van de gezinnen de bundeltjes terug. Het responspercentage van gezinnen met volledige of voldoende personenvragenlijsten ligt hier op 17% (=15%+2%). In vergelijking met Gent zien we dat in Vlaanderen de bundeltjes iets beter werden teruggestuurd (22 % tegenover 19 % voor Gent, een verschil van 3 %). De kwaliteit van de bundeltjes in Vlaanderen is evenwel iets minder goed vermits we uiteindelijk 17 % huishoudens hebben die hun bundeltjes volledig of met voldoende aantal personenvragenlijsten terugsturen en in Gent 15 % (een verschil van 2 %).

3.2 Profiel van huishoudens die (uiteindelijk) niet meewerkten

Voor alle huishoudens uit de steekproef beschikken we over de leeftijd en het geslacht van het gezinshoofd, bijgevolg ook voor de huishoudens die weigerden 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 recruteringsvragenlijst hebben deelgenomen (en dus per definitie ook niet aan het schriftelijke deel van de enquête) als degenen die **wel** aan de recruteringsvragenlijst hebben deelgenomen maar die niet aan het schriftelijke deel wensten deel te nemen.

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

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

Frequency Cell Chi-Square Percent Row Pct Col Pct	GESLACHT=man						
	17-24	25-34	35-44	45-54	55-64	65+	Total
0	2	28	90	95	99	309	623
	0.4521	27.369	17.113	13.266	1.7894	145.66	
	0.04	0.63	2.01	2.12	2.21	6.91	13.93
	0.32	4.49	14.45	15.25	15.89	49.60	
	8.70	5.37	9.04	9.61	12.18	27.32	
1	21	493	906	894	714	822	3850
	0.0732	4.4288	2.7692	2.1467	0.2896	23.57	
	0.47	11.02	20.25	19.99	15.96	18.38	86.07
	0.55	12.81	23.53	23.22	18.55	21.35	
	91.30	94.63	90.96	90.39	87.82	72.68	
Total	23	521	996	989	813	1131	4473
	0.51	11.65	22.27	22.11	18.18	25.29	100.00

TROK LFTKL

Frequency Cell Chi-Square Percent Row Pct Col Pct	GESLACHT=vrouw 17-24	25-34	35-44	45-54	55-64	65+	Total
0	1	9	14	19	24	300	367
	1.7135	15.137	23.704	16.65	11.923	62.714	
	0.08	0.76	1.19	1.61	2.03	25.42	31.10
	0.27	2.45	3.81	5.18	6.54	81.74	
	9.09	9.18	9.15	12.58	15.58	48.94	
1	10	89	139	132	130	313	813
	0.7735	6.8331	10.701	7.5162	5.382	28.31	
	0.85	7.54	11.78	11.19	11.02	26.53	68.90
	1.23	10.95	17.10	16.24	15.99	38.50	
	90.91	90.82	90.85	87.42	84.42	51.06	
Total	11	98	153	151	154	613	1180
	0.93	8.31	12.97	12.80	13.05	51.95	100.00

Bij het telefonisch contact weigeren vrouwen opvallend meer dan mannen: slechts 14% van de mannen weigert medewerking, tegenover 31% bij de vrouwen. Bij beide geslachten vinden we dat, hoe ouder het gezinshoofd, hoe groter de kans op een weigering. 65-plussers weigeren twee tot vijf keer zoveel als alle andere leeftijdgroepen.

Merk op dat meer dan 50% van de vrouwelijke gezinshoofden ouder is dan 65 jaar, terwijl dit bij de mannen een kwart is (25,29 %). 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 recruteringsvragenlijst). 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

(RQOK [Retour Questionnaire OK]=1 is teruggestuurd; RQOK=0 is niet-teruggestuurd)

RQOK LFTKL

GESLACHT=man

Frequency / Cell Chi-Square / Percent / Row Pct / Col Pct	17-24	25-34	35-44	45-54	55-64	65+	Total
0	8	207	340	303	212	342	1412
	0.0115	3.7938	0.1794	1.8876	9.4943	5.4486	
	0.21	5.38	8.83	7.87	5.51	8.88	36.68
	0.57	14.66	24.08	21.46	15.01	24.22	
	38.10	41.99	37.53	33.89	29.69	41.61	
1	13	286	566	591	502	480	2438
	0.0067	2.1972	0.1039	1.0932	5.4988	3.1556	
	0.34	7.43	14.70	15.35	13.04	12.47	63.32
	0.53	11.73	23.22	24.24	20.59	19.69	
	61.90	58.01	62.47	66.11	70.31	58.39	
Total	21	493	906	894	714	822	3850
	0.55	12.81	23.53	23.22	18.55	21.35	100.00

RQOK LFTKL

GESLACHT=vrouw

Frequency / Cell Chi-Square / Percent / Row Pct / Col Pct	17-24	25-34	35-44	45-54	55-64	65+	Total
0	4	46	70	49	44	169	382
	0.1039	0.4182	0.3366	2.7341	4.7773	3.2708	
	0.49	5.66	8.61	6.03	5.41	20.79	46.99
	1.05	12.04	18.32	12.83	11.52	44.24	
	40.00	51.69	50.36	37.12	33.85	53.99	
1	6	43	69	83	86	144	431
	0.0921	0.3707	0.2983	2.4233	4.2342	2.8989	
	0.74	5.29	8.49	10.21	10.58	17.71	53.01
	1.39	9.98	16.01	19.26	19.95	33.41	
	60.00	48.31	49.64	62.88	66.15	46.01	
Total	10	89	139	132	130	313	813
	1.23	10.95	17.10	16.24	15.99	38.50	100.00

Deze tabel heeft enkel betrekking op die huishoudens die meegewerkt hebben aan de telefonische recruteringsvragenlijst en *gezegd hebben* dat zij ook aan het schriftelijke deel wilden deelnemen. De in grijs aangeduide kolompercentages wijken significant af van het kolomgemiddelde. 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 47%. Deze tweede, schriftelijke, uitval is groter dan de eerste (namelijk degenen

die aan de telefoon reeds, al dan niet na deelname aan de recruteringsvragenlijst, weigerden verder mee te werken), en opnieuw meer uitgesproken bij vrouwelijke gezinshoofden dan bij mannelijke gezinshoofden: bij mannelijke gezinshoofden weigert 14% ogenblikkelijk (Tabel 2), en van degenen die overblijven haakt nog eens 37% af (Tabel 3). Bij vrouwelijke gezinshoofden weigert 31% ogenblikkelijk (Tabel 2), en van degenen die overblijven haakt nog eens 47% af (Tabel 3). Ook in dit stadium is het afhaken afhankelijk van de leeftijd maar niet meer lineair: soms werken oudere leeftijdsklassen mee, soms jongere leeftijdsklassen (Tabel 3).

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

Tabel 4. Terugsturen van de vragenlijst volgens wagenbezit

(RQOK [Retour Questionnaire OK]=1 is teruggestuurd; RQOK=0 is niet-teruggestuurd)

RQOK WAGAANT

Frequency Percent Row Pct Col Pct	0	1	2+	Total
0	285 6.12 15.91 52.29	959 20.58 53.55 36.94	547 11.74 30.54 36.01	1791 38.43
1	260 5.58 9.06 47.71	1637 35.13 57.06 63.06	972 20.86 33.88 63.99	2869 61.57
Total	545 11.70	2596 55.71	1519 32.60	4660 100.00

Deze tabel heeft enkel betrekking op die huishoudens die meegewerkt hebben aan de telefonische recruteringsvragenlijst en die gezegd hebben dat zij ook aan de schriftelijke wilden deelnemen.

Personen die uiteindelijk toch afhaken hebben relatief vaker (52% i.p.v. de verwachte 38%) geen wagen dan personen die wel verder meewerken (Tabel 4). Er is een duidelijk verband met geslacht en leeftijd van het gezinshoofd (zie ook Tabel 2 en Tabel 3): huishoudens met vrouwelijke en/of oudere gezinshoofden hebben minder wagens (data hier niet weergegeven).

4 Bepalen van gewichten

Van een aantal huishoudens uit de steekproef hebben we dus geen gegevens. Zulke non-respons veroorzaakt bijna altijd een vertekening van de gegevens t.o.v. de populatie. In dit geval bevat de uiteindelijk gerealiseerde steekproef een overaanbod aan mannen en aan gehuwden en een tekort aan personen jonger dan 25 jaar en aan alleenstaanden. Deze tekorten doen zich voor zowel bij het onderzoek naar de gezinshoofden en hun huishoudens, als bij de individuele personen. Deze tekorten zijn gedeeltelijk te wijten aan de steekproeftrekking uit het Rijksregister en gedeeltelijk aan een beperktere medewerking van bepaalde bevolkingsgroepen (Nuyts & Zwerts, 2001b).

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 en omgekeerd. Alle analyses van dit onderzoek zijn gebaseerd op de gewogen gerealiseerde steekproef.

Daarnaast is er nog een scheeftrekking voor de spreiding over de maanden, met een tekort voor de zomermaanden, en een ongelijke verdeling over de dagen van de week (Nuyts & Zwerts 2001b). Ook hiervoor corrigeren we door aan de maanden en de dagen gewichten toe te kennen.

Voor de berekening van de gewichten, die erg technisch is, verwijzen we naar bijlage 9.1.

4.1 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 (gezinsgrootte, leeftijd, burgerlijke staat en geslacht) nog significant af van de verdelingen van de populatie nadat we het produkt 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	1.49	25-	1.90	**ongehuwd**	0.86	**man**	0.98
1 vrouw	1.65	25-34	1.21	**gehuwd**	0.98	**vrouw**	1.05
2 personen	0.87	35-44	1.02	**gescheiden**	1.20		
3 personen	0.90	45-64	0.80	**weduwe(naar)**	1.09		
4 personen	0.84	65+	1.18				
5 personen	0.84						
6 personen	0.93						
7 personen	1.58						
8 personen	1.93						

De exacte interpretatie van de cijfers in Tabel 5 is alleen in combinatie zinvol. Bijvoorbeeld: het gewicht van '1 man alleen' van 1.49 wordt afgezwakt omdat het gezinshoofd per definitie dan een man is en dus als geslacht-gewicht 0.98 heeft. Wat samen toch nog neer komt op 1.49 * 0.98=1.46, en dus op een ondervertegenwoordiging van alleenstaande mannen in de steekproef. Ander voorbeeld: een gezin dat bestaat uit 1 ongehuwde man tussen 25 en 34 jaar krijgt gewicht **1.49 * 1.21 * 0,86 * 0,98** = **1.52.**

Merk op dat, indien men vergeet dat de gewichten *alleen als combinatie zinvol zijn*, men zelfs schijnbare tegenstellingen kan vinden. De 35-44-jarige gezinshoofden zijn in de steekproef oververtegenwoordigd: 22.2% van de steekproef tegenover 21.3% in de populatie. Indien we echter énkel kijken naar het gewicht van deze groep: 1.02, dan lijkt het of ze licht ondervertegenwoordigd zijn. Doordat deze gezinnen vaak met twee of meer zijn, wordt de schijnbare ondervertegenwoordiging steeds opgeheven.

4.2 Effectief gebruikte gewichten personen

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

0.97	**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*	*0.61*	*0.97*	*0.89*	*0.76*	*0.73*	*0.92*	*1.17*	*1.92*
Gescheiden	*1*	*1*	*2*	*1.83*	*1.58*	*1.50*	*1.89*	*2.41*	*1*
Ongehuwd	*1.09*	*0.99*	*1.58*	*1.45*	*1.25*	*1.19*	*1.50*	*1.91*	*3.15*
Weduwenaar	*1*	*1*	*1.13*	*1.03*	*0.89*	*0.85*	*1.07*	*1.34*	*2.24*
1.03	**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*	*0.67*	*0.93*	*0.81*	*0.74*	*0.76*	*1.08*	*1.75*	*2.42*
Gescheiden	*1*	*1*	*1.47*	*1.28*	*1.16*	*1.20*	*1.69*	*2.75*	*1*
Ongehuwd	*1.15*	*0.94*	*1.30*	*1.14*	*1.04*	*1.06*	*1.51*	*2.45*	*3.38*
Weduwe	*1*	*0.82*	*1.14*	*1.00*	*0.91*	*0.93*	*1.32*	*2.14*	*2.96*

Voor de berekening van de gewichten in Tabel 6 gebruiken we twee stappen. Eerst corrigeren we voor een lichte scheeftrekking van de geslachtsverhouding: voor mannen x 0.97 en voor vrouwen x 1.03. Daarna vermenigvuldigen we deze waarden met de *schuine getallen* uit Tabel 6. Voorbeeld: gewicht van een gehuwde man van 47 jaar = **0.97 * 0.76 = 0.74**.

4.3 Effectief gebruikte gewichten verplaatsingen

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

Dag ® Maand ¯	maandag	dinsdag	woensdag	donderdag	vrijdag	zaterdag	zondag
januari	*1.51*	*1.77*	*0.95*	*0.90*	*0.75*	*0.66*	*0.80*
februari	*0.94*	*1.17*	*0.80*	*0.97*	*1.06*	*1.11*	*0.91*
maart	*1.06*	*1.16*	*0.86*	*1.16*	*1.07*	*1.16*	*1.08*
april	*0.91*	*0.93*	*0.66*	*0.76*	*0.84*	*0.81*	*1.01*
mei	*0.94*	*0.95*	*0.92*	*1.24*	*1.00*	*0.78*	*0.87*
juni	*1.54*	*1.39*	*1.08*	*1.24*	*0.99*	*1.02*	*1.39*
juli	*1.63*	*1.70*	*1.34*	*1.19*	*1.54*	*1.20*	*1.43*
augustus	*2.38*	*2.13*	*0.97*	*1.10*	*1.19*	*1.02*	*1.17*
september	*0.91*	*0.99*	*0.97*	*1.39*	*1.53*	*1.10*	*0.97*
oktober	*0.97*	*0.97*	*0.85*	*0.84*	*0.81*	*0.81*	*1.04*
november	*0.75*	*0.99*	*0.95*	*0.71*	*0.63*	*1.03*	*0.79*
december	*0.96*	*0.73*	*1.23*	*0.83*	*0.88*	*0.76*	*1.01*

De gewichten waarmee we verplaatsingen willen vermenigvuldigen zijn berekend op personenniveau. Dit wil zeggen dat we nagaan of er in elke maand en op elke weekdag 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 en per weekdag maar worden maanden waarin mensen zich minder verplaatsen niet kunstmatig opgehoogd. Het gewicht voor een maandag in januari is 1.51 (Tabel 7), niet omdat er te weinig verplaatsingen waren op maandagen in januari, maar omdat er minder personen 'ondervraagd' zijn op die dagen.

5 Bij de analyse maken we geen gebruik meer van de 2e invuldag

Bij alle tot op heden uitgevoerde OVG's in Vlaanderen (Vlaanderen april 1994-april 1995, Vlaanderen januari 2000-januari 2001, de stadsgewesten Antwerpen april 1999-april 2000, Hasselt-Genk april 1999-april 2000 en Gent januari 2000-januari 2001) hebben we aan de respondenten gevraagd om gedurende twee dagen hun verplaatsingen te noteren. Hoewel er enkele conceptuele kanttekeningen bij deze werkwijze te maken waren, leek het grote voordeel dat men dubbel zoveel verplaatsingsgegevens had voor slechts een kleine prijsstijging. We hebben nu voor de vier vermelde OVG's analyses gedaan op het verschil tussen de eerste en de tweede invuldag en die verschillen blijken groter dan verwacht.

5.1 Onderrapporteringen op de 2e invuldag

Tabel 8. Aantal respondenten per invuldag

	1e invuldag	2e invuldag	aantal 2e dag t.o.v. 1e dag
Vlaanderen 2000-2001	7413	7287	98%
Gent	6617	6484	98%
Antwerpen	5445	5329	98%
Hasselt-Genk	6708	6578	98%

2% van de respondenten haakt af na de 1e invuldag.

Tabel 9. Gemiddeld aantal verplaatsingen per persoon per dag

	1e invuldag	2e invuldag	aantal 2e dag t.o.v. 1e dag
Vlaanderen 2000-2001	2.76	2.49	90%
Gent	2.81	2.52	89%
Antwerpen	2.82	2.61	93%
Hasselt-Genk	2.90	2.70	93%

Voor de personen die de tweede dag hebben ingevuld vinden we een duidelijk kleiner aantal verplaatsingen voor de 2e invuldag tegenover de 1e invuldag. In werkelijkheid is er, gemiddeld, echter geen enkel verschil tussen de 1e en de 2e invuldag: de proportie weekdagen is dezelfde, de geslachtsverhouding, de verdeling over de maanden, de proportie studenten, enzovoorts. Het enige verschil is dat het voor de respondent de 2e invuldag is. M.a.w. de respondenten die ook de 2e dag invullen, worden slordiger en noteren hun verplaatsingen minder exact.

Tabel 10. Gemiddeld aantal kilometer per persoon per dag

	1e invuldag	2e invuldag	aantal 2e dag t.o.v. 1e dag
Vlaanderen 2000-2001	32.69	30.60	94%
Gent	31.50	30.87	98%
Antwerpen	30.95	26.35	85%
Hasselt-Genk	32.24	30.95	96%

Het percentage kilometers van de 2e dag (ruwweg 95%) is hoger dan het percentage verplaatsingen van de 2e dag (ruwweg 92%). De lange afstanden worden op dag 2 dus beter ingevuld dan de korte afstanden.

5.2 Onderrapporteringen op de 2e invuldag volgens hoofdvervoermiddel, motief, afstand en verplaatsingstijd

Indien we dieper inzoomen op de data dan merken we dat de verdelingen van een aantal kenmerken niet verschillen voor verplaatsingen van dag 2 of de verplaatsingen van dag 1. Dit is het geval voor b.v. de proportie weekdagen, de geslachtsverhouding…. Voor een bepaald aantal kenmerken is dit evenwel wèl het geval n.l. met de afstand, de verplaatsingstijd, de hoofdvervoerswijze en het motief (zie Tabellen 11, 12, 13 en 14). In de volgende tabellen staan de geteste variabelen opgesplitst per klasse.

Tabel 11. Significante verschillen volgens afstand tussen het aantal verplaatsingen van de 1e en de 2e invuldag

	1e invuldag	2e invuldag
0.1-0.2 km	A*	
0.3-0.5 km		
0.6-1 km	G*	
1.1-2 km	G**	
2.1-3 km	A* V* H*	
3.1-5 km	G*	H*
5.1-7.5 km		V*
7.6-10 km		V* G***
10.1-15 km		
15.1-25 km		
25.1-40 km		
> 40 km		G*

Tabel 11 geeft aan op welke dag bepaalde afstandsklassen relatief meer gerapporteerd worden. De afkorting van het OVG is geplaatst bij de dag waar de relatieve frequentie het grootste is. A = Antwerpen, G = Gent, H = Hasselt-Genk en V = Vlaanderen 2. * = P<0,05; ** = P<0,01 en *** = P<0,005 waarbij P = de kans is dat we dit resultaat zuiver door het toeval bekomen.

De eenvoudigste wijze om Tabel 11 te interpreteren is als volgt: klassen met een symbool in de eerste kolom worden duidelijk ondergerapporteerd op de 2e dag, waardoor klassen met een symbool in de laatste kolom relatief gesproken meer gerapporteerd worden.

Hoewel het in elk onderzoek apart niet opvalt, zien we een duidelijke tendens als we alle onderzoeken combineren. Korte verplaatsingen worden de 2e invuldag vaker vergeten. Dit komt overeen met het resultaat van Tabel 10 dat langere verplaatsingen beter gerapporteerd worden tijdens de eerste invuldag.

Tabel 12. Significante verschillen volgens verplaatsingstijd tussen het aantal verplaatsingen van de 1e en de 2e invuldag

	1e invuldag	2e invuldag
0-5 min		
6-10 min		
11-15 min		H*
16-20 min		G*
21-30 min	V*** G ***	
31-60 min		V*** G*
61-120 min	V***	H*
> 120 min		

De afkorting van het OVG is geplaatst bij de dag waar de relatieve frequentie het grootste is. A = Antwerpen, G = Gent, H = Hasselt-Genk en V = Vlaanderen 2. * = P<0,05; ** = P<0,01 en *** = P<0,005 waarbij P = de kans is dat we dit resultaat zuiver door het toeval bekomen.

Aangezien verplaatsingen over korte afstand de 2e dag minder gerapporteerd worden, kunnen we veronderstellen dat dit ook zo zal zijn voor verplaatsingen over een korte tijd. Dat blijkt niet uit Tabel 12. Het zijn verplaatsingen van 21-30 minuten en van 1 à 2 uur die minder vermeld zijn ten voordele van enkele klassen die daar tussen of daar rond liggen. Waar zich bij de afstanden een duidelijk patroon aftekende (zie vorige tabel), geldt dit niet bij de verplaatsingstijden.

Tabel 13. Significante verschillen volgens hoofdvervoerswijze tussen het aantal verplaatsingen van de 1e en de 2e invuldag

	1e invuldag	2e invuldag
autobestuurder	A *	
autopassagier		
te voet	V***	
fietser	H*	
brom-snorfiets		
motor		
(pre)metro		V*
bus		
tram		
trein		G* H*
andere/onbepaald		V*** A ***H*** G*

De afkorting van het OVG is geplaatst bij de dag waar de relatieve frequentie het grootste is. A = Antwerpen, G = Gent, H = Hasselt-Genk en V = Vlaanderen 2. * = P<0,05; ** = P<0,01 en *** = P<0,005 waaarbij P = de kans is dat we dit resultaat zuiver door het toeval bekomen.

Tabel 13 laat maar één duidelijke conclusie toe: mensen laten de 2e invuldag vaker het vakje 'gebruikt vervoermiddel' open. Trein of metro laat misschien iets meer indruk na, zodat de respondenten hierin minder slordig worden. Indien we data samenvoegen, zoals alle openbaar vervoer samen, of alle lokaal openbaar vervoer (lijnbus, tram, metro), of autobestuurders en –passagiers, verschijnen er geen duidelijker patronen.

Tabel 14. Significante verschillen volgens motief tussen het aantal verplaatsingen van de 1e en de 2e invuldag

	1e invuldag	2e invuldag
werken		G*
zakelijk bezoek		
onderwijs volgen		
iemand/iets brengen/halen		A***
winkelen	V*** G***	
diensten		G*** H***
wandelen/rondrijden		
iemand bezoeken	A*** H***	V*
ontspanning/sport/cultuur		H**
andere		
onbepaald		V***

De afkorting van het OVG is geplaatst bij de dag waar de relatieve frequentie het grootste is. A = Antwerpen, G = Gent, H = Hasselt-Genk en V = Vlaanderen. * = P<0,05; ** = P<0,01 en *** = P<0,005 waarbij P = de kans is dat we dit resultaat zuiver door het toeval bekomen.

Winkelen en iemand bezoeken worden minder gerapporteerd op de 2e invuldag (Tabel 14), maar er zit geen lijn in de motieven die wel goed gerapporteerd worden.

5.3 Besluit: weglaten van de 2e invuldag

De 2e dag worden de verplaatsingen slordiger ingevuld dan de 1e: er zijn minder mensen die de 2e dag nog invullen, degenen die hem invullen, vergeten een aantal verplaatsingen en bij de verplaatsingen die wel ingevuld worden, vergeet men vaker om bepaalde velden in te vullen. Dit op zich maakt het reeds moeilijk om op basis van de twee invuldagen samen extrapolaties te maken voor de populatie. Ten opzichte van de 1e invuldag veroorzaakt de 2e invuldag ruis.

Veel erger is dat er op de 2e invuldag systematische fouten zitten. Korte verplaatsingen, winkelen en bezoekjes worden veel gemakkelijker vergeten dan andere verplaatsingen. Trein wordt misschien iets beter onthouden dan andere vervoermiddelen. We krijgen dus niet alleen *ruis* door het gebruik van de 2e invuldag, we krijgen waarschijnlijk ook *vertekeningen*.

Daarom hebben we beslist om voor de analyses de 2e invuldag niet te gebruiken.

6 Vergelijking telefonisch/postaal bevraagden en enkel postaal bevraagden

Er zijn drie groepen van mensen die enkel postaal bevraagd 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.

Elk van deze groepen kan een eigen, specifiek mobiliteitsgedrag hebben. Bij de analyse van hun verplaatsingen bestuderen we dus een combinatie van drie deelgroepen, die we echter op geen enkele wijze met onze data kunnen scheiden.

Tabel 15. VMB-index[1] volgens postale en telefonische bevraging

Frequency Percent Row Pct Col Pct	> 2 wagens	2 wagens	1 wagen	Motor/bromfiets	Enkel fietsen	Geen/overig	Total
beige / Vlaanderen postaal	5.9949 0.20 1.39 6.97	47.011 1.55 10.90 6.41	246.37 8.14 57.10 13.87	11.69 0.39 2.71 30.58	56.568 1.87 13.11 26.17	63.834 2.11 14.79 35.93	431.47 14.25
roze / Vlaanderen telefonisch	79.966 2.64 3.08 93.03	686.45 22.67 26.44 93.59	1529.8 50.53 58.92 86.13	26.535 0.88 1.02 69.42	159.61 5.27 6.15 73.83	113.8 3.76 4.38 64.07	2596.1 85.75
Total	85.9608 2.84	733.466 24.23	1776.14 58.66	38.2251 1.26	216.175 7.14	177.638 5.87	3027.61 100.00

Deze gemengde postale groep bezit gemiddeld minder vervoermiddelen (Tabel 15).

Tabel 16. Gemiddeld aantal verplaatsingen per persoon per dag volgens postale en telefonische bevraging

KLEUR	Frequency
beige / Vlaanderen postaal	2.187783
roze / Vlaanderen telefonisch	2.831702

Deze gemengde postale groep verplaatst zich ook minder vaak (Tabel 16).

[1] Voor een definitie van 'VMB-index: zie bladzijde 13 in deel 2

Tabel 17. Verdeling van het gemiddeld aantal verplaatsingen per persoon per dag volgens hoofdvervoerswijze en volgens postale of telefonische bevraging

KLEUR(Kleur van de vragenlijst) HFDVM

Frequency Row Pct	(pre)metro	andere/onbep.	autobestuurder	autopassagier	brom/snorfietser	lijnbus	Total
beige / Vlaanderen postaal	0 0.00	0.1767 8.08	0.9927 45.37	0.3133 14.32	0.0214 0.98	0.0316 1.44	2.1878
roze / Vlaanderen telefonisch	0.0024 0.09	0.2062 7.28	1.2479 44.07	0.5086 17.96	0.0314 1.11	0.0447 1.58	2.8317

(Continued)

KLEUR(Kleur van de vragenlijst) HFDVM

Frequency Row Pct	fietser	motorrijder	te voet	tram	trein	Total
beige / Vlaanderen postaal	0.3203 14.64	0.0087 0.40	0.2692 12.30	0.0116 0.53	0.0424 1.94	2.1878
roze / Vlaanderen telefonisch	0.4146 14.64	0.0096 0.34	0.3116 11.00	0.0103 0.36	0.0443 1.57	2.8317

Uit Tabel 17 blijkt dat, hoewel postaal bevraagde huishoudens veel minder auto's hebben (30.61% (2.71+13.11+14.79) tegenover 11.55% bij de telefonisch bevraagde huishoudens, zie Tabel 15), zij relatief gesproken bijna evenveel de wagen gebruiken als personen uit huishoudens die telefonisch bevraagd waren (Tabel 17: 59.7% (45.37 + 14.32) tegenover 62%). Dit wil zeggen dat de postaal bevraagde huishoudens met wagen, hun wagen gemiddeld vaker gebruiken dan de huishoudens die telefonisch bevraagd waren.

Tabel 18. Gemiddeld aantal kilometer per persoon per dag volgens postale en telefonische bevraging

KLEUR	Frequency
beige / Vlaanderen postaal	29.78934
roze / Vlaanderen telefonisch	33.05749

De postaal bevraagden leggen per dag iets minder kilometers af (Tabel 18). Maar ze maken dan ook minder verplaatsingen per dag (Tabel 16). Per verplaatsing leggen de postaal bevraagden zelfs meer km af: postaal bevraagden leggen gemiddeld per verplaatsing 13.6 km af en telefonisch bevraagden 11.7 km.

Tabel 19. Verdeling van het gemiddeld aantal kilometer per persoon per dag volgens hoofdvervoerswijze en volgens postale of telefonische bevraging

KLEUR(Kleur van de vragenlijst) HFDVM

Frequency Row Pct	(pre)metro	andere/onbep.	autobestuurder	autopassagier	brom/snorfietser	lijnbus	Total
beige / Vlaanderen postaal	0 0.00	4.6463 15.60	16.185 54.33	4.6797 15.71	0.1746 0.59	0.2908 0.98	29.789
roze / Vlaanderen telefonisch	0.0269 0.08	2.1394 6.47	17.837 53.96	7.6792 23.23	0.184 0.56	0.6029 1.82	33.057

(Continued)

KLEUR(Kleur van de vragenlijst) HFDVM

Frequency Row Pct	fietser	motorrijder	te voet	tram	trein	Total
beige / Vlaanderen postaal	1.3974 4.69	0.1563 0.52	0.4634 1.56	0.0413 0.14	1.7548 5.89	29.789
roze / Vlaanderen telefonisch	1.9259 5.83	0.1591 0.48	0.4912 1.49	0.0765 0.23	1.9359 5.86	33.057

In aantal kilometers per dag gebruiken postaal bevraagden minder de auto als passagier (16% tegenover 23%). Ze geven veel vaker als antwoord andere/onbepaald .

7 Technische aspecten i.v.m. de statistische verwerking

7.1 Gebruikte technieken: frequentietabellen en regressie

De bekomen data werden uitgezuiverd aan de hand van strikte regels (Nuyts & Zwerts 2001b), 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 regressie gebruikt.

Voor de meeste analyses gebruiken we dus 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 één of meerdere (zgn. onafhankelijke) variabelen elk afzonderlijk hebben op één andere (zgn. afhankelijke) 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[2]. Afhankelijk van de mogelijke waarden van de afhankelijke variabele gebruiken we een ander 'type' regressie: *lineaire* regressie of *logistische* regressie.

Bij een lineaire regressie drukken we een bepaalde variabele die veel getalwaarden kan aannemen, bijvoorbeeld aantal dagen carpoolen per jaar, uit als een lineaire functie van andere variabelen, bv. 'vrouw zijn' en 'leeftijd tussen 25 en 35 jaar'.

Dan is de regressie van de vorm:

$$Y = aX1 + bX2 + c$$

met Y het aantal dagen dat men met iemand meerijdt, X1 en X2 onafhankelijke variabelen, hier 'vrouw zijn' en 'leeftijd tussen 25 en 35 jaar', en a, b en c door een statistisch pakket[3], berekende constanten.

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:

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

P is dan de kans dat iemand een rijbewijs heeft en net zoals bij lineaire regressie, X1 en X2 onafhankelijke variabelen, hier 'vrouw zijn' en 'leeftijd tussen 25 en 35 jaar', en a, b en c door een statistisch pakket berekende constanten.

We kunnen deze vergelijking ook schrijven als:

$$De\,kans\,op\,een\,rijbewijs = \frac{1}{1+e^{-(aX1+bX2+c.)}}.$$

Dit maakt het (iets) eenvoudiger om de getallen te interpreteren.

Tabel 20. Fictief voorbeeld van een logistische regressie om de begrippen uit te leggen. Afhankelijke variabele is rijbewijsbezit

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	4.5595	0.2098	472.1404	0.0001	.	.
VROUW	1	-1.2984	0.1815	51.2019	0.0001	-0.351515	0.273
LFT1624	1	-0.1966	0.3123	0.3962	0.5291	-0.022571	0.822

[2] Multi-variate regressie in tegenstelling tot *univariate* regressie waar men 1 afhankelijke variabele probeert te begrijpen met behulp van 1 onafhankelijke variabele.

[3] In deze documenten is dat SAS.

Voor wie niet echt geïnteresseerd is in de exacte getalwaarde, maar enkel in het feit of iemand meer of minder kans heeft op een rijbewijs volstaat volgende vuistregel. Als de '*Parameter Estimate*' positief is dan stijgt de kans op een rijbewijs; indien de Parameter Estimate negatief is dan daalt de kans op een rijbewijs.
Voorbeeld uit Tabel 20: De 'parameter estimate' van vrouw is "-1.2984". Deze parameter estimate is negatief, dus daalt de kans op een rijbewijs indien de persoon in kwestie een vrouw is.

De volledige betekenis van deze logistische regressie is:

$$De\ kans\ op\ een\ rijbewijs = \frac{1}{1+e^{-(4.5595-1.2984\ 'indien\ vrouw'-1.966\ 'indien\ \ tussen\ 16\ en\ 24\ \ jaar')}}$$

Als variabelen niet in de regressie voorkomen, wil dit zeggen dat ze geen toegevoegde waarde meer hebben *bij alle variabelen die reeds in het model zitten*. Zo blijkt dat een aanzienlijk aantal respondenten geen treinstation in de buurt van hun huis hebben, maar ook niet in de buurt van hun werk. Vaak is dit een reden om een ander vervoermiddel te nemen. Het is van belang dat er op één plaats geen station is. Maar het extra probleem dat er op de andere plaats ook geen station is, is erg beperkt. Daardoor verschijnt dit vaak niet meer in de regressies. Welke afstand (halte thuis of halte op het werk) er in het model opgenomen wordt, is zuiver bepaald door de statistische berekeningen. De meest significante variabelen blijven over.
We hebben dus meestal veel meer variabelen uitgeprobeerd, dan dat er uiteindelijk in de regressie overblijven.

Ook het aantal variabelen dat in de regressies is opgenomen, is zuiver statistisch bepaald. De variabelen met een significantie kleiner dan 5% (P < 5%) zijn opgenomen in het model. Dit heeft tot gevolg dat er soms variabelen verschijnen die we niet verwacht hadden, en die we zelfs bij nader inzien niet kunnen verklaren[4]. Het is ook mogelijk om meer 'gericht' modellen te maken. We kunnen bijvoorbeeld enkel variabelen met P<5% weerhouden als we kunnen begrijpen waarom ze relevant zijn, en variabelen waarvan we niet begrijpen waarvan hun impact komt, weerhouden we enkel bij P<1%, of P<0.1%. Dit is duidelijk minder wetenschappelijk, maar het levert een model op waarvan we alle aspecten (denken te) begrijpen, en een model dat ook eenvoudiger uit te leggen is aan de buitenwereld, bv. voor het sturen van beleidsbeslissingen. Anderzijds kunnen we ook op voorhand bepaalde variabelen in het model dwingen, om hun P-waarde te kennen. Het kan voor een overheidsbeslissing van belang zijn of de afstand tot een bushalte met 7% kans irrelevant is voor het nemen van het openbaar vervoer, of met 60% kans geen impact heeft op een stijging van het gebruik van het openbaar vervoer. In het geval van P =7% is het de moeite om andere analyses te doen, of zelfs ander onderzoek te verrichten om meer zekerheid te krijgen of er nu wel of niet een invloed is, in het geval dat P=60% is er gewoonweg geen verband.
We hebben er heel bewust niet voor gekozen om op het eerste zicht vreemde variabelen weg te laten of andere variabelen in de regressie te dwingen. We hebben nu op vier onafhankelijke OVG's deze regressies gebruikt, en het blijkt namelijk dat enkele variabelen die we op het eerste gezicht niet kunnen verklaren, toch systematisch in de regressies van de vier OVG's voorkomen. Bv. onafhankelijk van hun geslacht, leeftijd, opleiding of inkomen hebben ongehuwde personen en

[4] Strikt gezien kan dit een gevolg zijn van het gebruikte statistische criterium. De variabelen met een significantie kleiner dan 5% zijn behouden in het model. Dit wil zeggen dat, als er maar 5% kans is dat een variabele bij wijze van pech door de steekproef relevant lijkt, maar in werkelijkheid toch niet relevant is, dat we dan de variabele behouden. De redenering daarachter is: '5% kans is zo klein, dat kan geen toeval meer zijn'.
We kunnen dit ook minder positief formuleren: indien we 100 variabelen proberen om een model te maken, dan kunnen er door zuiver pech 5 geselecteerd worden die significant lijken, maar het eigenlijk niet zijn. Welnu, voor deze modellen hebben we ongeveer 60 variabelen uitgeprobeerd. Normaal gezien worden alle relevante variabelen geselecteerd, maar we moeten er rekening mee houden dat er ook variabelen geselecteerd zijn die toch niet relevant zijn. Het is verleidelijk om te stellen dat dit de variabelen zijn waarvan we de impact niet begrijpen.
Met behulp van meer geavanceerde statistische, maar helaas ook meer arbeidsintensieve, technieken is het mogelijk om overfitting (zie voetnoot 5) met grotere zekerheid uit te sluiten.

weduwen/weduwnaars minder vaak een rijbewijs. Een resultaat dat vier maal terugkomt in vier onafhankelijk onderzoeken kan geen artefact zijn van de data. Het is dus de moeite om deze informatie te bewaren. Daarnaast zijn er een aantal vreemde resultaten die slechts in één van de vier OVG's terugkomen. Misschien zijn die resultaten erg streekgebonden ofwel zijn het artefacten van die specifieke dataset.

De 'parameter estimate' bij het intercept geeft de waarde van de regressie in de referentiesituatie. Dit impliceert dan ook dat er een referentiesituatie bepaald wordt. Ook dit hebben we aan de statistiek overgelaten: we hebben de statistiek de significant afwijkende variabelen laten zoeken. Wat niet afwijkt is dan de referentiesituatie. Hiervoor geldt dezelfde opmerking als hierboven. We hadden zelf kunnen ingrijpen, maar zolang we niet zeker weten hoe en waarom, is het voorzichtig om dit niet te doen. Dat blijkt bijvoorbeeld uit het feit dat de referentiesituatie voor de bus-, tram en treinhaltes verschilt van regressie tot regressie.

In principe is regressie uitgevonden om continue variabelen te vergelijken met continue variabelen, b.v. lengte van armen i.f.v. totale lengte van een persoon. 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 referentieklasse 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 men 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 verhoopt 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[5] te voorkomen, hebben we enkel groepen met meer dan 25 waarnemingen gebruikt. 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.

Voor elk van de variabelen i.v.m. de afstand tot de trein-, tram- en bushalte zijn meerdere mogelijkheden: halte dichter dan 200 m, dichter dan 500 m, dichter dan 1 km, dichter dan 2 km[6]. In principe kunnen we modellen bouwen waarin zowel voorkomt 'bushalte dichter dan 500 m" en "bushalte dichter dan 2 km". Dan is de variabele "dichter dan 500 m" een soort extra toevoeging op de variabele "dichter dan 2 km". We willen echter de belangrijkste impact van de afstand tot de bushalte in kaart brengen. Daarom hebben we voor elk vervoermiddel slechts één afstand tot de halte in het model toegelaten, en nooit combinaties zoals hierboven.

7.2 Niet gebruikte technieken: multinomiale logit analyse

Bij de keuze van een vervoermiddel moet men kiezen tussen meer dan twee objecten zonder intrinsieke volgorde. Er is n.l. geen enkele eenduidige manier om te zeggen dat een auto 'meer' of 'minder' is dan een fiets, die op zijn beurt 'meer' of 'minder zou zijn dan een bus, enz. Voor het opstellen van keuzemodellen met meerdere ongeordende mogelijkheden geeft men in de literatuur drie alternatieven.

1. Men kan logistische regressie gebruiken waarbij één vervoermiddel (bv. de fiets) wordt vergeleken met de groep van alle andere vervoermiddelen samen (Stokes et al. 1995). Dit model analyseert dus eigenlijk 'fiets' tegenover 'niet-fiets'. We vinden dan wanneer mensen de fiets nemen en wanneer men de fiets laat staan. We kunnen dan niet zien wélk vervoermiddel er in de plaats van de fiets gebruikt wordt. Bij deze analysemethode kunnen we voor elk vervoermiddel alle data gebruiken.

2. Men kan logistische regressie gebruiken waarbij men slechts twee vervoermiddelen selecteert (Agresti 1990), bv. de fiets en de bus. Het voordeel hiervan is dat men weet wanneer men van de fiets overschakelt naar de bus en omgekeerd. Het nadeel is dat men voor elke analyse slechts een deel van de data kan gebruiken. In dit voorbeeld wordt het grootste aantal verplaatsingen, die met de auto, niet gebruikt.

3. Via multinomiale logit analyse kan men in één model alle data gebruiken en toch zien wanneer men van het ene vervoermiddel overgaat op een ander. Deze werkwijze wordt in de literatuur het meeste aangeraden (McFadden 1974, Agresti 1990, Stokes et al. 1995, en verscheidene verwijzingen in het literatuuroverzicht van van Wee 1994). Deze techniek vergt een groot aantal data (Agresti 1990), en minimaal 20 à 30 waarnemingen van elk vervoermiddel.

In de OVG's beschikken we over ruim 3000 werkende personen. Niet al deze personen hebben alle vragen beantwoord, maar we hebben toch steeds 1500 of meer personen die alle vragen, die relevant zijn voor deze analyse, hebben ingevuld. Dit aantal leek ruim voldoende om een multinomiale logit analyse te proberen om de vervoermiddelenkeuze beter te begrijpen. Daarom hebben we een eerste test uitgevoerd. In praktijk hebben we voor deze test de data van OVG Gent gebruikt (waarvan het veldwerk dus gelijktijdig met OVG Vlaanderen 2 werd uitgevoerd). Het resultaat viel echter tegen. Reeds bij het gebruik van twee variabelen, geslacht en leeftijd, kregen we onaangename resultaten.

In de getoonde analyse werd de variabele leeftijd opgesplitst in 5 leeftijdsklassen. Als referentiepersoon gebruikten we : man, autogebruiker en ouder dan 55 jaar.

[5] Overfitting vindt 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.

[6] In praktijk werden ook de variabelen 'verder dan 200 m', 'verder dan 500 m, enz. gebruikt. In eerste instantie lijken de ja-nee variabelen 'dichter dan 500 m' en 'verder dan 500 m' erg op elkaar. Wat bij de ene variabele 'ja' is, is bij de andere 'nee'. Het verschil ligt in het toewijzen van de 'weet niet'. Bij elk van deze variabelen komt dit bij 'nee' terecht.

Tabel 21. Resultaten van een multinomiale logit analyse op het niveau van vervoermiddelengebruik als één groot geheel.

MAXIMUM-LIKELIHOOD ANALYSIS-OF-VARIANCE TABLE

Source		DF	Chi-Square	Prob
INTERCEPT	5	154.37	0.0000	
VROUW		5	39.23	0.0000
LFT1624		5	1.83	0.8727
LFT2534		5	16.40	0.0058
LFT3544		5	10.04	0.0741
LFT4554		5	4.41	0.4915

Elke variabele ('source' in de tabel) heeft 5 vrijheidsgraden (Df= 5) omdat alle resultaten van de aparte vervoermiddelen gegroepeerd worden. Dat de leeftijdsklasse 25-34 significant is (P = 0.0058) kunnen we in min of meer gewoon Nederlands vertalen als 'deze leeftijdsklasse verklaart gedeeltelijk het vervoermiddelengebruik'.

Tabel 22. Resultaten van een multinomiale logit analyse op het niveau waarbij elk vervoermiddel apart beschouwd wordt.

ANALYSIS OF MAXIMUM-LIKELIHOOD ESTIMATES

Effect	Parameter	Estimate	Standard Error	Chi-Square	Prob
INTERCEPT	voet	-3.6334	0.4196	74.97	0.0000
	fiets	-1.8431	0.3167	33.87	0.0000
	bus	-3.2791	0.4968	43.57	0.0000
	tram	-4.2927	1.0541	16.58	0.0000
	trein	-2.1219	0.3831	30.67	0.0000
VROUW	voet	-0.1467	0.1195	1.51	0.2197
	fiets	-0.0148	0.0680	0.05	0.8281
	bus	-0.6070	0.1342	20.47	0.0000
	tram	-0.6092	0.2265	7.24	0.0071
	trein	0.2222	0.0773	8.25	0.0041
LFT1624	voet	-0.0378	0.2467	0.02	0.8782
	fiets	0.00755	0.2045	0.00	0.9706
	bus	-0.3166	0.2712	1.36	0.2431
	tram	-0.0583	0.6414	0.01	0.9275
	trein	0.1476	0.2635	0.31	0.5755
LFT2534	voet	0.5565	0.1946	8.18	0.0042
	fiets	0.0417	0.1377	0.09	0.7618
	bus	0.5807	0.2484	5.47	0.0194
	tram	-0.2572	0.4473	0.33	0.5652
	trein	-0.2082	0.1556	1.79	0.1809
LFT3544	voet	0.5851	0.1965	8.87	0.0029
	fiets	0.0378	0.1374	0.08	0.7831
	bus	0.2913	0.2305	1.60	0.2065
	tram	0.1085	0.4708	0.05	0.8177
	trein	0.0761	0.1615	0.22	0.6375
LFT4554	voet	0.3930	0.1963	4.01	0.0452
	fiets	0.0931	0.1442	0.42	0.5188
	bus	0.0245	0.2273	0.01	0.9141
	tram	0.2364	0.5077	0.22	0.6414
	trein	-0.00197	0.1651	0.00	0.9905

Voor elke variabele ('effect' in de tabel) wordt voor elk vervoermiddel één waarde (= 'Estimate') geschat. Dat de waarde bij de leeftijdsklasse 25-34 voor de modus 'te voet' significant is (P = 0.0042)

kunnen we in min of meer gewoon Nederlands vertalen als ‘deze leeftijdsklasse verklaart gedeeltelijk wanneer iemand te voet gaat’.

Op het niveau waarop het vervoermiddelengebruik gegroepeerd is, is enkel ‘vrouw’ en ‘leeftijd 25-34 jaar’ significant (Tabel 21). Kijken we echter op het diepere niveau van de aparte vervoermiddelen (Tabel 22) dan blijkt ‘vrouw’ wel significante verschillen te geven voor de drie vormen van openbaar vervoer, maar niet voor ‘te voet’ of ‘fiets’. In zekere zin wordt het model dus wat zwaar doordat we twee parameters moeten meenemen die eigenlijk niet significant zijn. Wegens technische redenen, waar we hier niet verder op in gaan, moeten we ze echter wel berekenen.

Veel erger is dat variabelen die op het niveau van ‘vervoermiddelengebruik als geheel’ niet significant zijn, op het diepere niveau van één enkel vervoermiddel wel significant kunnen zijn. In dit voorbeeld is ‘ leeftijd 45-54’ niet significant op het algemene niveau (Tabel 21 : χ^2=4.41, Df=5, P=0.4915). Maar als we naar de verschillende vervoermiddelen apart kijken, blijkt dat personen van 45-54 wel meer te voet gaan (Tabel 22: χ^2=4.01, Df=1, P=0.0452). Voor alle andere vervoermiddelen is deze leeftijdsklasse echter insignificant. Vervelend genoeg hebben we exact hetzelfde probleem bij leeftijdsklasse 34-44, en dit voor exact hetzelfde vervoermiddel. Welke werkwijze we in de volgende paragraaf ook kiezen, de negatieve effecten van deze keuze zullen zich in dit voorbeeld zeker verdubbelen.

We hebben twee mogelijkheden.
Ofwel proberen we voorzichtig te zijn en nemen we elk effect dat significant is voor een of ander vervoermiddel mee op in het model , ook al is het globaal niet significant. Dan wordt het model heel log, want we krijgen een groot aantal ‘overbodige’ parameters voor de andere vervoermiddelen. Uit de analyses van de stadsgewesten Hasselt-Genk en Antwerpen (Nuyts et al. 2001, Zwerts et al. 2001) blijkt dat slechts weinig variabelen voor alle vervoermiddelen relevant zijn. Indien we alle variabelen die voor één of meerdere vervoermiddelen relevant zijn, zouden meenemen, dan eindigen we met een groot kluwen waar we uiteindelijk niets meer in terug vinden[7].
Ofwel nemen we enkel variabelen mee die op globaal niveau significant zijn. Maar dan verliezen we informatie over de aparte vervoermiddelen.

Geen van beide oplossingen is gunstig. Daarom hebben we gekozen om geen multinomiale logit analyse te gebruiken, maar wel logistische regressie waarbij elk vervoermiddel vergeleken wordt met de groep van alle andere vervoermiddelen. Indien we over meer data beschikken, b.v. bij het samenvoegen van de data van Vlaanderen 1994-1995, Vlaanderen 2000-2001 en eventuele stadsgewesten, dan kunnen we elk van de vervoermiddelen vergelijken met elk van de andere vervoermiddelen, wat duidelijk de meest gewenste informatie oplevert.

Het gebruik van logistische regressies heeft nog een ander voordeel, alleszins bij het statistische pakket SAS. In de regressie-analyse is het mogelijk om het statistisch pakket zelf variabelen aan het model te laten toevoegen, of ze er weer uit te nemen, afhankelijk of deze variabele nog significant is samen met andere variabelen (door de option=stepwise aan het programma op te leggen). Dergelijke optie is er niet bij het opstellen van een multinomiale logit analyse. Dat wil zeggen dat alle mogelijke combinaties van variabelen één voor één door de onderzoeker zelf uitgeprobeerd moeten worden, wat een enorm tijdrovend karwei is. Het opstellen van een basismodel indien men ± 50 variabelen a-priori relevant acht, duurt bij logistische regressie enkele minuten. Bij multinomiale logit analyse duurt dit met ± 50 a-priori variabelen eerder een dag.

[7] Een ruwe schatting na het uitvoeren van aparte analyses per vervoermiddel op de data van Gent leert het volgende. Er zijn te weinig data over tramgebruik om tram apart te modelleren. We houden dus nog 4 vervoermiddelen over die geen ‘auto’ zijn. Deze analyses leveren samen 26 verschillende variabelen op. Deze methode zou ons een model opgeleverd hebben met ± 100 parameters (26 variabelen x 4 vervoermiddelen), waarvan de meerderheid niet significant is, maar wel noodzakelijk om het model te maken.

7.3 Betrouwbaarheid van de resultaten.

7.3.1 Betrouwbaarheid en nauwkeurigheidsintervallen bij proporties

Bij proporties kan men 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 variabele met 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 ± 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.

7.3.2 Betrouwbaarheid en nauwkeurigheidsintervallen bij regressies

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 σ wordt door SAS mee berekend en ook getoond in de output. De schatting van de parameter heeft ongeveer een normale verdeling. Om bijvoorbeeld een betrouwbaarheid te krijgen van 95%, neemt men de schatting van de parameter ± 1.96*σ.

7.3.3 Vervangingsvariabelen.

De beperktheid van de gebruikte variabelen is waarschijnlijk één van de belangrijkste oorzaken van afwijkingen tussen statistische modellen en de werkelijkheid (Anas 1982). Het probleem dat de gewenste variabele niet beschikbaar is, probeert men vaak op te lossen door het gebruik van vervangingsvariabelen. Maar het gebruik van vervangingsvariabelen i.p.v. de echte variabele kan resulteren in niet-causale modellen met oninterpreteerbare resultaten (van Wee 1994).

In de OVG's gebruiken we vaak 'afstand tot de meest nabije bushalte' als vervangingsvariabele voor 'afstand tot de noodzakelijke bushalte'. De bus die de mensen nodig hebben voor hun woon-werk of woon-school verkeer stopt soms wel, maar zeker niet altijd aan de meest nabije bushalte. Hetzelfde geldt voor de tramhaltes, en in beperkte mate zelfs voor de treinstations. Dit kan een verklaring zijn waarom afstanden tot de meest nabije halte contra-intuïtieve resultaten geeft in de regressies.

7.3.4 Significantietoetsen

Met significantietoetsen gaan we na of een verschillend getal dat we vinden tussen 2 groepen (bv. tussen mannen en vrouwen of tussen personen in 1995 en 2001), al dan niet 'toevallig' is.

Wanneer de toets aangeeft dat het verschil niet 'significant' is dan is het verschil 'toevallig". Dit betekent dan dat het verschil dat gevonden werd in de steekproef tussen de groepen (of voor eenzelfde groep tussen verschillende steekproeven) puur toeval is en zich in de realiteit (= populatie) waarschijnlijk (minimaal met een betrouwbaarheid van 95 % en soms meer) niet voordoet. De waarden zijn dus waarschijnlijk ongeveer dezelfde gebleven (rekening houdend met de betrouwbaarheidsgrenzen, zie punt 7.3.1).

Wanneer de toets aangeeft dat het verschil wel 'significant' is dan is het verschil niet toevallig. Dit betekent dan dat het verschil dat gevonden werd in de steekproef tussen de groepen (of voor eenzelfde groep tussen verschillende steekproeven) geen toeval is en zich in de realiteit (= populatie) waarschijnlijk (minimaal met een betrouwbaarheid van 95 % en soms meer) wel voordoet.

8 Bibliografie

* Agresti, A. (1990). 'Categorical Data Analysis'. New York: Wiley. 558p.

* Anas, A. (1982). "Residential location and urban transportation, economic theory, econometrics, and policy analysis with discrete choice models". New York, Academic Press.

* 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.

* McFadden, D. (1974). 'Conditional logit analysis of qualitative behaviour'. In 'Frontiers in Econometrics', ed. by P. Zarmebka. New York: Academic Press, p. 105-142.

* Nationaal Instituut voor Statistiek. (2000) "Bevolkingsstatistieken. Totale en Belgische bevolking op1.1.2000". N.I.S., Brussel, 274p.

* Nationaal Instituut voor Statistiek. (2001a) "Bevolkingsstatistieken. Huishoudens en familiekernen op1.1.2000, Tabel 00.13: Referentiepersonen van de private huishoudens naar geslacht, leeftijdsklasse en burgerlijke staat". N.I.S., Brussel. Elektronisch opvraagbaar.

* Nationaal Instituut voor Statistiek. (2001b) "Bevolkingsstatistieken. Huishoudens en familiekernen op1.1.2000, Tabel 00.07 A: Particuliere huishoudens naar grootte van het huishouden - collectieve huishoudens". N.I.S., Brussel. Elektronisch opvraagbaar.

* Nuyts, E. & Zwerts, E. (2000) "Onderzoek Verplaatsingsgedrag Stadsgewest Hasselt-Genk, Stadsgewest Antwerpen, Controle en begeleidingdsopdracht, eindverslag". Onderzoekscel Architectuur en Mobiliteit, Provinciale Hogeschool Limburg, Departement Architectuur , Diepenbeek, 126p.

* Nuyts, E. & Zwerts, E. (2001a) "Onderzoek Verplaatsingsgedrag Stadsgewest Hasselt-Genk (april 1999-april 2000). Deel 1: Methodologische analyse". Onderzoekscel Architectuur en Mobiliteit, Provinciale Hogeschool Limburg, Departement Architectuur , Diepenbeek, 23 p.

* Nuyts, E. & Zwerts, E. (2001b) "Onderzoek Verplaatsingsgedrag Vlaanderen en Stadsgewest Gent, Januari 2000 - januari 2001. Controle en begeleidingdsopdracht". Onderzoekscel Architectuur en Mobiliteit, Provinciale Hogeschool Limburg, Departement Architectuur , Diepenbeek, 153p.

* Nuyts, E., Zwerts, E. & Miermans, W., (2001). "Onderzoek Verplaatsingsgedrag Stadsgewest Hasselt-Genk (april 1999-april 2000). Deel 3A: analyse personenvragenlijst." Provinciale Hogeschool Limburg, Departement Architectuur en Beeldende Kunst, Diepenbeek, 225 p.

* Stokes, M., Davis, C. & Koch, G. (1995). 'Categorical Data Analysis Using the SAS System'. Cary, NC : SAS Institute Inc. 499p.

* 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.

* van Wee, B. (1994). “Werklocaties, woonlocaties en woon-werkverkeer. Literatuurstudie”. Rijksinistituut voor volksgezondheid en Milieuhygiëne. Bilthoven, 135p.

* Zwerts, E. & Nuyts, E. (2001) “Onderzoek Verplaatsingsgedrag Stadsgewest Antwerpen (april 1999-april 2000). Deel 1: Methodologische analyse”. Onderzoekscel Architectuur en Mobiliteit, Provinciale Hogeschool Limburg, Departement Architectuur , Diepenbeek, 21 p.

* Zwerts, E. & Nuyts, E. (2001b) “Onderzoek Verplaatsingsgedrag Gent (januari 2000-januari 2001). Deel 1: Methodologische analyse”. Onderzoekscel Architectuur en Mobiliteit, Provinciale Hogeschool Limburg, Departement Architectuur , Diepenbeek.

* Zwerts, E., Nuyts E. & Miermans, W., (2001). “Onderzoek Verplaatsingsgedrag Stadsgewest Antwerpen (april 1999-april 2000). Deel 3A: Analyse personenvragenlijst,” Provinciale Hogeschool Limburg, Departement Architectuur, Diepenbeek, 224 p

9 Bijlage

9.1 Berekening van de gewichten

9.1.1 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 gezamelijke 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 Vlaanderen 2000-2001 hebben we voor de huishoudens 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. 2001a, 2001b), en voor de personen de gezamenlijke verdeling van geslacht en burgerlijke staat, en de gezamenlijke verdeling van geslacht en leeftijdsklassen (N.I.S. 2000). We nemen de data van gans Vlaanderen zonder te kijken naar de verdeling over de verschillende gemeenten. 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 gewichten niet te laten afhangen van de gemeenten.

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 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 Weight 2.1 Ook voor de eerste variabele V1, want diens 'ideale' gewicht Weeg1 is verschoven door het toevoegen van Weeg2. Bemerk 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 chi² 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.

Enkele conclusies van de techniek zijn dat:

- In praktijk blijkt dat de iteratie altijd vrij snel lukt (Nuyts & Zwerts 2001a, Zwerts & Nuyts 2001, dit document, Zwerts & Nuyts 2001b).
- Bij elke stap kan de iteratie weer verslechteren (Nuyts & Zwerts, 2001a).
- Het beste resultaat bekom je niet per se op het einde van een 'ronde' (Nuyts & Zwerts, 2001a).

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.

9.1.2 Huishoudens: vier relevante variabelen

Via de publicaties van het N.I.S. (2001a, 2001b) beschikken we over vier variabelen die relevant zijn. 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. Na één 'ronde', waarbij elke variabele zijn factor aan het uiteindelijke gewicht toevoegde, week geen enkele van de marginale verdelingen nog significant af van de marginale verdelingen van de populatie.

Het eindresultaat van deze berekening zijn vier series gewichten (één serie per variabele) die met elkaar vermenigvuldigd moeten worden om het uiteindelijke gewicht van een huishouden te bekomen. De resultaten zijn ook zo getoond in Tabel 5.

9.1.3 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. 2000). 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.

9.1.4 Verplaatsingen: een verdeling van één variabele

We willen dat de invuldagen gelijkmatig verspreid zijn over de weekdagen, en dat ze evenredig verdeeld zijn over de maanden. We hebben één variabele gemaakt die maand en weekdag combineert, met 12*7=84 mogelijke antwoorden (maandag in januari, dinsdag in januari, …zondag in december). Gewichten berekenen voor één enkele variabele kan het handigste met de standaard module van Weight 2.1, die speciaal hiervoor ontworpen is.

9.2 Berekening van de ophoogfactor

De gebruikte ophoogfactor = populatie aantal vanaf 6 jaar/gewogen steekproef aantal.
Men kan eventueel delen door steekproef aantal i.p.v. door gewogen 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.

9.3 Samenvoegen van gegevens

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

9.3.1 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=  'opvoedster'          wordt          7=‘bediende’
14=  'leraar'                             7=‘bediende’
15=  'militair'                           12=‘andere, WEL beroepsactief’
16=  'ambtenaar'                          7=‘bediende’
17=  'houtbewerker'                       6=‘arbeider’
18=  'verzorgende'                        6=‘arbeider’
19=  'rijkswachter'                       12=‘andere, WEL beroepsactief’
20=  'verpleegkundige'                    7=‘bediende’
21=  'docent'                             7=‘bediende’
22=  'onthaalmoeder'                      12=‘andere, WEL beroepsactief’
23=  'muzikant'                           12=‘andere, WEL beroepsactief’
24=  'chauffeur'                          6=‘arbeider’
25=  'magistraat'                         8=‘kader’
26=  'luchtverkeersleider'                12=‘andere, WEL beroepsactief’
27=  'politie in opleiding'               12=‘andere, WEL beroepsactief’
28=  'geestelijke'                        7=‘bediende’
29=  'zelfstandige helper'                6=‘arbeider’
30=  'vertegenwoordiger'                  7=‘bediende’
31=  'voorzitter'                         8=‘kader’
32=  'sportman'                           12=‘andere, WEL beroepsactief’
33=  'meewerkende echtgenoot'             10=‘zelfstandige’’
```

```
 34=  'kunstenaar'                          12='andere, WEL beroepsactief'
 35=  'officier'                            8='kader'
 36=  'acteur'                              12='andere, WEL beroepsactief'
 37=  'doctoraat'                           7='bediende'
 38=  'onderzoeker'                         7='bediende'
 39=  'stewardess'                          12='andere, WEL beroepsactief'
 40=  'horeca'                              12='andere, WEL beroepsactief'
 41=  'bursaal'                             7='bediende'
 42=  'toezichter'                          6='arbeider'
 43=  'stadswacht'                          12='andere, WEL beroepsactief'
 44=  'brandweer'                           12='andere, WEL beroepsactief'
 45=  'assistent'                           7='bediende'
 46=  'kabinetsmedewerker'                  12='andere, WEL beroepsactief'
 47=  'leercontract'                        12='andere, WEL beroepsactief'
 48=  'professor'                           8='kader'
 49=  'stagiair'                            1=scholier, student''
 50=  'animator'                            12='andere, WEL beroepsactief'
 51=  'inspecteur'                          7='bediende'
 71=  'zwanger- moederschapsverlof'         11='andere, NIET beroepsactief'
 72=  'invalide'                            5='arbeidsongeschikt'
 73=  'opleiding VDAB'                      11='andere, NIET beroepsactief'
 74=  'loopbaanonderbreking'                11='andere, NIET beroepsactief'
 75=  'vrijwilligerswerk'                   11='andere, NIET beroepsactief'
 76=  'tbs'                                 11='andere, NIET beroepsactief'
 77=  'meewerkende echtgenoot'              10='zelfstandige''
 78=  'sociaal plan'                        11='andere, NIET beroepsactief'
 79=  'rentenier'                           11='andere, NIET beroepsactief'
 80=  'ziekenkas'                           5='arbeidsongeschikt'
 81=  'overlevingspensioen'                 4='gepensioneerd''
 82=  'zonder papieren'                     6='arbeider'
 83=  'zelfstandige helper'                 6='arbeider'
 84=  'uitstapregeling'                     11='andere, NIET beroepsactief';
```

9.3.2 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 = 'stage'                          wordt   6  ='onderwijs volgen'
13 = 'mijn auto halen'                        3  ='winkelen'
14 = 'gaan stempelen'                         11  ='andere'
15 = 'naar de kerk\mosquee gaan'              11  ='andere'
16 = 'gaan tanken\benzine station'            3  ='winkelen'
17 = 'examen\proclamatie'                     6  ='onderwijs volgen'
18 = 'kot'                                    11  ='andere'
19 = 'gaan eten(restaurant)'                  9  ='ontspanning/sport/cultuur'
20 = 'vergadering'                            11  ='andere'
21 = 'verblijfplaats'                         11  ='andere'
22 = 'bijscholing\cursus\opleiding'           11  ='andere'
23 = 'werken nieuw gebouw'                    11  ='andere'
24 = 'verhuis van gerief'                     11  ='andere'
25 = 'begrafenis'                             11  ='andere'
26 = 'receptie\drink'                         9  ='ontspanning/sport/cultuur'
27 = 'ziekenhuis'                             11  ='andere'
```

```
 28 = 'stadsbezoek'                              9  ='ontspanning/sport/cultuur'
 29 = 'babysitten\onthaalmoeder\opvang'          11  ='andere'
 30 = 'containerpark'                            10  ='diensten (dokter, bank)'
 31 = 'hond uitlaten'                            9  ='ontspanning/sport/cultuur'
 32 = 'tweede verblijf\buitenverblijf'           11  ='andere'
 33 = 'koffietafel'                              11  ='andere'
 34 = 'kerkhof'                                  11  ='andere'
 35 = 'solliciteren'                             11  ='andere'
 36 = 'naar garage'                              3  ='winkelen'
 37 = 'op vakantie\op reis\in buitenland'        9  ='ontspanning/sport/cultuur'
 38 = 'bloed geven'                              11  ='andere'
 39 = 'auto rijlessen'                           3  ='winkelen'
 40 = 'hout hakken'                              11  ='andere'
 41 = 'internaat'                                11  ='andere'
 42 = 'scouts\jeugd-\jongerenbeweging'           9  ='ontspanning/sport/cultuur'
 43 = 'demonstratie\opendeur'                    3  ='winkelen'
 44 = 'autokeuring'                              3  ='winkelen'
 45 = 'carwash\auto wassen'                      3  ='winkelen'
 46 = 'VDAB'                                     10  ='diensten (dokter, bank)'
 47 = 'iemand vergezellen'                       5  ='iemand een bezoek brengen'
 48 = 'feest'                                    9  ='ontspanning/sport/cultuur'
 49 = 'vrijwilligerswerk'                        5  ='iemand een bezoek brengen'
 50 = 'verzorging dieren'                        9  ='ontspanning/sport/cultuur'
 51 = 'iemand helpen'                            5  ='iemand een bezoek brengen'
 52 = 'schoolreis'                               6  ='onderwijs volgen'
 53 = 'klussen'                                  11  ='andere'
 54 = 'veiling'                                  11  ='andere'
 55 = 'beurs'                                    11  ='andere'
 56 = 'plaats van vertrek'                       7  ='wandelen / rondrijden'
 57 = 'hondenschool'                             9  ='ontspanning/sport/cultuur'
 58 = 'kamp'                                     9  ='ontspanning/sport/cultuur'
 59 = 'catechese\godsdienst'                     11  ='andere'
 60 = 'hotel'                                    11  ='andere'
 61 = 'kippen slachten'                          11  ='andere'
 62 = 'uitstap'                                  9  ='ontspanning/sport/cultuur'
 63 = 'bijberoep'                                2  ='werken'
 64 = 'voorbereiding'                            11  ='andere'
 65 = 'parkeerschijf verzetten'                  10  ='diensten (dokter, bank)'
 66 = 'voordracht'                               9  ='ontspanning/sport/cultuur'
 67 = 'politiek'                                 9  ='ontspanning/sport/cultuur'
 68 = 'proefrit'                                 3  ='winkelen'
 69 = 'laatste groet'                            11  ='andere'
 70 = 'gaan logeren'                             5  ='iemand een bezoek brengen'
 71 = 'informatie vragen'                        10  ='diensten (dokter, bank)'
 72 = 'stemmen\verkiezingen'                     11  ='andere'
 73 = 'betogen'                                  9  ='ontspanning/sport/cultuur'
 74 = 'seminarie'                                11  ='andere'
 75 = 'repetitie'                                9  ='ontspanning/sport/cultuur'
 76 = 'huizen bekijken'                          11  ='andere'
 77 = 'bouwgrond bekijken'                       11  ='andere'
 78 = 'instelling'                               11  ='andere'
 79 = 'strafstudie'                              6  ='onderwijs volgen'
 80 = 'rusten'                                   9  ='ontspanning/sport/cultuur'
 99 = 'onbepaald' ;
niet ingevuld                                    99 = 'onbepaald'
```

9.4 Vragenlijsten