

het aantal waarnemingen in de betrokken groep. Kleine groepen hebben minder kans om significant te zijn. Om mogelijke overfitting⁴ 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 σ 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 ± 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.

⁴ 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-index⁵ 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

⁵ 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 bevrooging.

	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 bevrooging.

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).