

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 χ^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 χ^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 χ^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 nieuwsgierigheid⁶ hebben we ook nog een tweede 'ronde' uitgewerkt, om te zien hoeveel deze ronde de resultaten verbeterde (Tabel 14).

⁶ 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 ^e volledige 'ronde'	Na 2 ^e 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^e en na de 2^e volledige ronde. Na de 2^e 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**

bssamen⁷ 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

⁷ Bssamen = Burgerlijke staat (ongehuwd en samenwonend samen genomen)

Stap 5 na aanpassing geslacht (cfr. Tabel 14: na 1^e volledige 'ronde')

geslacht	Probability: 0.975596
leeftkl	Probability: 0.610053
bssamen	Probability: 0.159459
ledenkl	Probability: 0.902992

Einde van de eerste ronde.

Begin tweede ronde.

Stap 6: aanpassing bssamen

geslacht	Probability: 0.132169
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: 0.489446

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: 0.250615
bssamen	Probability: 0.599859

Stap 9 na aanpassing leeftkl (cfr. Tabel 14: na 2^e volledige 'ronde')

geslacht	Probability: 0.888683
leeftkl	Probability: 0.999998
bssamen	Probability: 0.294637
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