



Centraal Bureau voor de Statistiek

Weging Sociaal Fysiek 2009

Erwin Vondenhoff, Kees van Berkel en Joep Burger

Abstract: In opdracht van het Ministerie van Volkshuisvesting Ruimtelijke Ordening en Milieubeheer (VROM) is het onderzoek Sociaal Fysiek 2009 als vervolgonderzoek van de Woningmarktmodule 2009 door het CBS uitgevoerd. In deze nota wordt de weging besproken waarmee populatieschattingen kunnen worden gemaakt op basis van de verzamelde gegevens. Er zijn drie wegingen uitgevoerd op persoonsniveau. Bij de eerste weging wordt het startgewicht bepaald op basis van de strata in de steekproef van Sociaal Fysiek en de bijbehorende populatietotalen. De tweede en de derde weging zijn twee-fasen wegingen waarbij naast de stratumindeling gebruik wordt gemaakt van de eindgewichten van de Woningmarktmodule. Behalve persoonsgewichten worden gewichten voor huishoudens en woningvoorraad bepaald. Hoewel een aantal personen binnen de G31 zonder woningopname-deel is geïnterviewd, is uiteindelijk besloten om geen gewichten aan ze toe te kennen. Om de informatie van deze groep respondenten niet verloren te laten gaan zijn in een eerste fase gewichten toegekend aan alle respondenten waarmee schattingen zijn gemaakt voor een aantal variabelen. Deze schattingen worden beschouwd in het weegmodel aan de hand waarvan de uiteindelijke gewichten bepaald worden, waarbij de G31 zonder woningopname-deel is weggelaten.

1. Inleiding

De module Sociaal Fysiek van het onderzoek WoON is uitgevoerd in de periode van februari tot en met juni 2009. De beoogde doelpopulatie is de Nederlandse bevolking van 18 jaar of ouder. Personen die tot institutionele huishoudens behoren worden niet meegeteld in de populatie. Er blijven dan 12.762.876 personen over. De steekproef voor Sociaal Fysiek is in twee fasen getrokken uit deze populatie. Aan de respondenten van de Woningmarktmodule is gevraagd of ze bereid waren om mee te doen aan een vervolgonderzoek. Uit degenen die deze vraag met “ja” beantwoordden is de uiteindelijke steekproef getrokken [1]. In deze nota wordt de toekenning van gewichten aan de 9.060 respondenten van Sociaal Fysiek besproken.

Bij het trekken van de steekproef zijn ten eerste alle respondenten uit Vogelaarwijken of ISV gebieden getrokken die hebben toegestemd in een vervolgonderzoek (zie categorieën 1 tot en met 5 in paragraaf 2). Uit personen die noch in een Vogelaarwijk noch in een ISV gebied wonen zijn steekproefelementen getrokken met ongelijke kansen naar leeftijd, herkomst en inkomen. Hierbij is rekening gehouden met responskansen die op historische data berusten. Bij de bepaling van de strata en de startgewichten is daarom rekening gehouden met de indeling van de populatie naar deze kenmerken. Daarnaast worden personen die in probleemwijken wonen onderscheiden.

De kans om in de steekproef van Sociaal Fysiek terecht te komen hangt af van de trekkingskans bij de Woningmarktmodule. Het steekproefontwerp voor dat onderzoek bestaat uit 655 strata. Omdat de uiteindelijke respons voor Sociaal Fysiek uit 9.060 elementen bestaat is het geen optie om deze strata opnieuw te gebruiken laat staan verder in te delen op basis van het steekproefontwerp van Sociaal Fysiek [1]. Er kan voor gekozen worden om in eerste instantie, bij de bepaling van het insluitgewicht, de scheefheid in de respons van de Woningmarktmodule te negeren en pas bij de bepaling van het correctiegewicht hiervoor te corrigeren. Er kan echter ook gebruik gemaakt worden van eindgewichten die bepaald zijn bij de weging van de Woningmarktmodule om direct in het startgewicht voor Sociaal Fysiek de trekkingskans en de responskans bij de Woningmarktmodule te verwerken.

Oorspronkelijk bestond de respons van Sociaal Fysiek uit 10.438 personen. Later heeft het CBS in overleg met VROM ervoor gekozen om respondenten binnen de G31 zonder woning-opname deel (WOpname) niet mee te tellen, zie Tabel 1, waardoor 9.060 respondenten overbleven. Uiteindelijk moeten gewichten toegekend worden aan deze 9.060 respondenten.

Tabel 1: Respondenten Sociaal Fysiek 2009 naar wel/niet WOpname en wel/niet G31

WOpname	G31		totaal
	ja	nee	
ja	6865	0	6865
nee	1378	2195	3573
totaal	8243	2195	10438

Om de 1.378 responsen binnen de G31 zonder WOpname niet verloren te laten gaan wordt informatie van deze groep verwerkt in de gewichten van de overige 9.060 respondenten. Dit wordt gedaan door te wegen in twee fasen. Eerst worden persoonsgewichten toegekend aan de oorspronkelijke groep van 10.438 respondenten, zie paragrafen 2 en 3. Vervolgens worden populatietotalen geschat voor een viertal variabelen op basis van de gevonden gewichten, zie paragraaf 6. Ten slotte worden de uiteindelijke persoonsgewichten toegekend aan de kleinere groep van 9.060 respondenten waaruit de G31 zonder WOpname is weggelaten. Het model dat hierbij gebruikt wordt bestaat ten eerste uit de termen van het model voor de eerste weging en ten tweede uit de termen met de vier variabelen waarvan de populatietotalen geschat zijn bij de eerste weging. De variabelen die bij de eerste weging geschat worden dienen belangrijk te zijn voor Sociaal Fysiek maar ze mogen niet gebaseerd zijn op resultaten van WOpname. Wanneer schattingen voor de geselecteerde variabelen gemaakt worden op basis van de tweede set gewichten, voor 9.060 respondenten, zullen deze schattingen beter worden op deze manier.

Behalve een persoonsgewicht wordt per respondent een huishoudensgewicht en een woningvoorraadgewicht bepaald. De huishoudensgewichten worden in paragraaf 4 bepaald voor respondenten met huisvestingssituatie 1 t/m 5 door het gevonden persoonsgewicht te delen door het aantal 18-plussers in de huishoudkern van het huishouden waartoe de respondent behoort. Respondenten met huisvestingssituatie 6 (leden van het huishouden) krijgen huishoudensgewicht 0. Vervolgens wordt gecorrigeerd met een constante factor, die zo gekozen is dat de som van de huishoudensgewichten gelijk is aan het aantal huishoudens in de populatie. De startgewichten voor woningvoorraad worden bepaald op basis van de huishoudensgewichten van respondenten met huisvestingssituatie 1. Met behulp van een weegmodel worden de eindgewichten voor

woningvoorraad hieruit bepaald, zie paragraaf 5.

2. De strata en de startgewichten

Bij de stratumindeling is gekeken naar een aantal regionale en persoonlijke kenmerken. De populatie is opgedeeld in acht categorieën gebaseerd op regionale kenmerken. Vervolgens is elke categorie opgesplitst in twaalf groepen waarbij gekeken is naar persoonlijke kenmerken. In paragrafen 2 tot en met 5 wordt steeds uitgegaan van de gehele groep van 10.438 respondenten waartoe ook de 1.378 personen binnen de G31 zonder WOpname behoren. Pas in paragraaf 6 wordt deze groep weggelaten.

In het steekproefontwerp voor Sociaal Fysiek is de volgende regionale indeling gemaakt:

- a. vogelaarwijken,
- b. ISV gebieden binnen de G4,
- c. ISV gebieden buiten de G4 maar binnen de G31,
- d. de G18 zonder vogelaarwijken en ISV gebieden,
- e. de G19-31 zonder ISV gebieden (per definitie geen vogelaarwijken),
- f. de niet-G31 (per definitie geen vogelaarwijken of ISV gebieden).

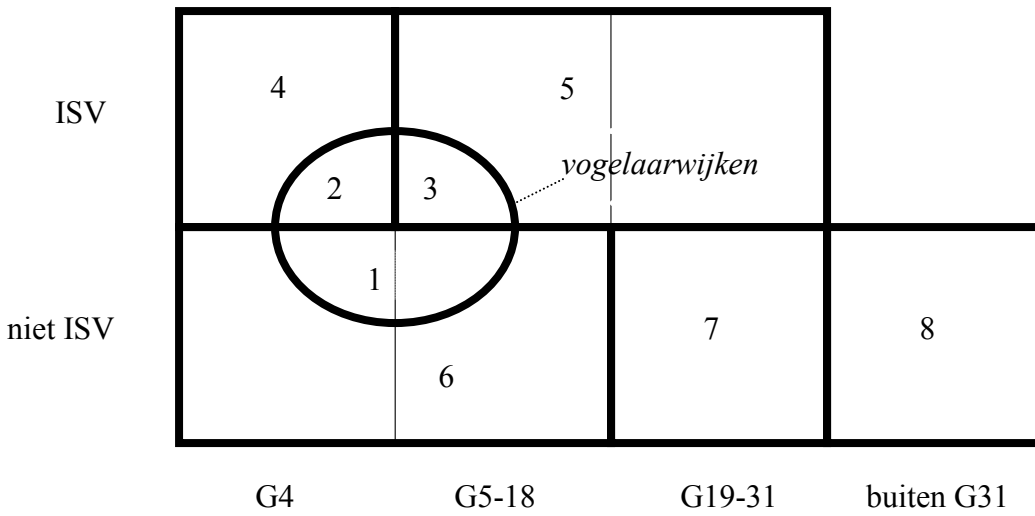
Merk op dat deze zes groepen niet disjunct zijn. Groep a heeft een niet-lege doorsnee met zowel groep b als groep c. De stratumindeling kan daarom niet gebaseerd worden op deze groepsindeling. Definieer de volgende acht disjuncte categorieën gebaseerd op de bovenstaande zes groepen:

1. alle elementen in groep a die niet in groep b of c zitten,
2. alle elementen die zowel in groep a als in groep b zitten,
3. alle elementen die zowel in groep a als in groep c zitten,
4. alle elementen die niet in groep a maar wel in groep b zitten,
5. alle elementen die niet in groep a maar wel in groep c zitten,
6. groep d,
7. groep e,
8. groep f.

Het woord “categorie” zal in het vervolg alleen gebruikt worden om naar deze indeling te verwijzen. In Figuur 1 is deze indeling schematisch weergegeven.

Behalve deze regionale indeling wordt bij de bepaling van de strata onderscheid gemaakt tussen zes klassen die gebaseerd zijn op leeftijd (65-plusser, niet-65-plusser), herkomst (westers, niet westers) en inkomen (laag, niet laag). In de definitie van de variabele S_2 in de volgende paragraaf is vastgelegd wanneer het inkomen laag dan wel niet laag is. Omdat er weinig niet-westerlingen zijn die 65 jaar of ouder zijn wordt geen onderscheid naar herkomst gemaakt voor deze groep. De zes klassen zijn dus als volgt gekozen:

1. jonger dan 65, niet westers, laag inkomen,
2. jonger dan 65, niet westers, niet-laag inkomen,
3. jonger dan 65, westers, laag inkomen,
4. jonger dan 65, westers, niet-laag inkomen,
5. 65 plusser, laag inkomen,
6. 65 plusser, niet-laag inkomen.



Figuur 1: Schema met categorieën 1 t/m 8. De ISV gebieden staan in het bovenste gedeelte van het schema en de ellips in het midden stelt de vogelaarwijken voor. Van links naar rechts zien we de indeling G4, G5t/m18, G19t/m31 en de rest.

In het steekproefontwerp zijn gelijksoortige klassen gedefinieerd, waarbij de klassen 1, 2, 3 en 4 overeenkomen met de klassen 1, 2, 3 en 4 in deze nota. Klassen 5 en 7 in het steekproefontwerp zijn hier samengevoegd tot één klasse 5. Klassen 6 en 8 zijn samengevoegd tot klasse 6.

Ten slotte wordt onderscheid gemaakt tussen personen met en zonder een partner. De zes bovenstaande klassen worden verfijnd tot 12 subklassen 1A, 1B, 2A, 2B, etc., waarbij de letter “A” voor “zonder partner” en de letter “B” voor “met partner” staat. Bij de trekking van de steekproef voor de Woningmarktmodule hebben partners een kleinere trekkingskans gekregen dan mensen zonder partner. Bij de steekproeftrekking van Sociaal Fysiek zijn deze kansen gelijk. Toch wordt deze variabele beschouwd in de stratumindeling, omdat zo de uiteindelijke correctiegewichten een kleinere standaardafwijking hebben en beter gecentreerd zijn rond het getal één. Om niet al te veel strata te krijgen zijn andere scheefheden ten gevolge van de steekproeftrekking van de Woningmarktmodule buiten beschouwing gelaten voor de stratumindeling. Hiervoor kan immers gecorrigeerd worden door de betreffende variabelen in het weegmodel op te nemen.

Uiteindelijk zijn er $8 \times 6 \times 2 = 96$ strata, gebaseerd op categorie, klasse en de variabele partner/niet-partner. In Tabellen 2a t/m 2d zien we zowel de responsaantallen als de populatieaantallen per stratum voor partners en niet-partners. Merk op dat in deze tabellen ook de 1.378 personen zitten die weliswaar binnen de G31 wonen, maar geen WOopname hebben.

Tabel 2a: Aantallen respondenten zonder partner naar categorie en klasse

categorie	klasse						totaal
	1A	2A	3A	4A	5A	6A	
1	135	106	186	164	74	33	698
2	148	115	166	116	76	21	642
3	31	20	86	45	18	1	201
4	103	63	138	96	58	15	473
5	71	55	330	265	103	23	847
6	128	99	394	592	134	89	1436
7	45	20	150	230	71	28	544
8	31	27	224	426	166	54	928
totaal	692	505	1674	1934	700	264	5769

Tabel 2b: Aantallen respondenten met partner naar categorie en klasse

categorie	klasse						totaal
	1B	2B	3B	4B	5B	6B	
1	32	51	49	154	24	33	343
2	43	77	33	90	14	25	282
3	8	10	24	34	6	12	94
4	22	44	35	119	14	10	244
5	26	56	95	333	49	59	618
6	29	79	105	825	47	118	1203
7	11	31	68	414	30	64	618
8	13	48	151	814	105	136	1267
totaal	184	396	560	2783	289	457	4669

Tabel 2c: Aantallen niet-partners in populatie naar categorie en klasse

categorie	klasse						totaal
	1A	2A	3A	4A	5A	6A	
1	30011	23179	36837	33267	15923	4476	143693
2	36666	27751	28534	22445	15812	3778	134986
3	6010	3689	13619	8477	4261	1124	37180
4	20202	16507	24351	22264	11710	2695	97729
5	16564	12256	63204	52621	24388	6279	175312
6	79811	70955	289931	331797	107401	46407	926302
7	18690	17541	101983	148828	49942	19140	356124
8	92339	91934	614376	1179262	419351	165716	2562978
totaal	300293	263812	1172835	1798961	648788	249615	4434304

Tabel 2d: Aantallen partners in populatie naar categorie en klasse

categorie	klasse						totaal
	1B	2B	3B	4B	5B	6B	
1	22190	27313	16335	48143	9691	9416	133088
2	27132	32184	11663	29812	9149	7902	117842
3	4337	5250	6523	15221	3342	2956	37629
4	12172	17672	10448	35997	6321	6831	89441
5	13088	18495	33473	110850	18703	19153	213762
6	47857	89220	152190	740877	67031	121631	1218806
7	15567	27568	75425	420793	43148	64379	646880
8	80678	160065	705135	3865560	437466	622220	5871124
totaal	223021	377767	1011192	5267253	594851	854488	8328572

De respondenten worden aangegeven met het symbool i , de strata met s en het stratum waartoe i behoort met $s(i)$. Laat verder N_s het aantal elementen in de populatie zijn dat tot stratum s behoort en laat r_s het aantal respondenten in stratum s van Sociaal Fysiek zijn. Definieer N als het aantal elementen in de populatie zodat

$$N = \sum_s N_s$$

en definieer $n:=10.438$ als het totale aantal respondenten. Voor elk element i zal met drie methoden een startgewicht bepaald worden. Deze methoden zullen in het algemeen drie verschillende getallen per respondent opleveren. De definitieve keuze voor één van de drie methoden wordt pas gemaakt nadat de verdeling van de bijbehorende correctiegewichten bekend is.

Bij de eerste methode is het stratum waartoe het element behoort de enige informatie die gebruikt wordt. Het startgewicht is de reciproke van de insluitkans, die alleen afhangt van het stratum waartoe de respondent behoort. Deze kans is het quotiënt van r_s en N_s , zodat voor elke respondent i geldt

$$d_i = \frac{N_{s(i)}}{r_{s(i)}}$$

Er wordt hier geen gebruik gemaakt van de gewichten die gevonden zijn bij de Woningmarktmodule. Bij het steekproefontwerp van dit voorafgaande onderzoek zijn bepaalde deelpopulaties over/ondervertegenwoordigd die beschouwd zijn in de stratumindeling. Behalve ongelijke trekkingskansen is er selectiviteit ten gevolge van non-respons. De steekproef is immers getrokken uit de selectieve groep van respondenten van de Woningmarktmodule. Voor beide vormen van selectiviteit kan in zekere mate gecorrigeerd worden met het weegmodel dat later in deze nota besproken wordt. Toch stellen we een tweede methode voor waarbij gebruik gemaakt wordt van de eindgewichten \tilde{w}_i van de Woningmarktmodule bij de bepaling van alternatieve startgewichten Δ_i voor Sociaal Fysiek:

$$\Delta_i := \lambda \tilde{w}_i,$$

waarbij λ een schaalfactor is die zodanig wordt gekozen dat de som van startgewichten van alle respondenten voor Sociaal Fysiek gelijk is aan het aantal elementen in de populatie:

$$N = \sum_{i \in u} \Delta_i,$$

waarbij u de verzameling van alle respondenten van Sociaal Fysiek is. Uit de laatste twee formules volgt nu

$$\lambda = \frac{N}{\sum_{i \in u} \tilde{w}_i}.$$

In tegenstelling tot de eerste methode tellen de gewichten per stratum niet op tot het populatieaantal voor dat stratum. Voor vertekening ten gevolge van oververtegenwoordiging van bepaalde groepen bij de steekproeftrekking van Sociaal Fysiek is dus niet gecorrigeerd. Dit moet dus gebeuren bij de bepaling van de correctiegewichten. Omdat het misschien beter is om dit bij de startgewichten te doen stellen we een derde methode voor ter bepaling van startgewichten δ_i :

$$\delta_i := \lambda_{s(i)} \tilde{w}_i,$$

waarbij $\lambda_{s(i)}$ schaalfactoren zijn die zodanig worden gekozen dat de som van startgewichten van alle respondenten voor Sociaal Fysiek in stratum $s(i)$ gelijk is aan het aantal elementen in dat stratum in de populatie:

$$N_s = \sum_{i \in s} \delta_i.$$

Elk stratum krijgt dus een andere schaalfactor. Uit de laatste twee formules volgt

$$\lambda_s = \frac{N_s}{\sum_{i \in s} \tilde{w}_i}.$$

Op deze manier wordt voor selectiviteit bij beide onderzoeken, Woningmarktmodule en Sociaal Fysiek gecorrigeerd in de startgewichten. Merk op dat deze startgewichten een grotere variantie kunnen hebben dan de andere twee typen gewichten omdat er meer kenmerken van steekprofelementen zijn meegenomen. In de volgende paragraaf worden correctiegewichten bepaald op basis van één weegmodel. Daarna wordt een definitieve keuze gemaakt voor één van de drie sets startgewichten en bijbehorende correctiegewichten.

3. Het weegmodel en de correctiegewichten

Het weegmodel dat wordt gebruikt voor de bepaling van de eindgewichten ziet er als volgt uit:

$$\begin{aligned} & (A_5 \times H_2 \times L_2 \times S_2) \\ & + (L_7 \times P_4) + (L_{15} \times G_2) + (I_6 \times P_4) + (A_5 \times W_5) + (P_4 \times H_3 \times I_3 \times L_2) \\ & + (L_{15} \times P_2) + (P_4 \times H_3) + (P_4 \times G_2) + (L_7 \times H_3) \\ & + K_3 + C_{40}. \end{aligned}$$

Dit lijkt in veel opzichten op het model dat door ABF Research is gebruikt voor Sociaal Fysiek 2006 [4]. De subindices geven aan hoeveel waarden de variabelen kunnen aannemen. De variabelen zijn als volgt gedefinieerd:

- A_5 : (aandacht) maakt binnen de G31 onderscheid naar wel/niet vogelaarwijk en wel/niet ISV gebied; personen buiten de G31 vormen een vijfde groep; preciezer gezegd, categorie 1 zoals gedefinieerd in de vorige paragraaf vormt een groep, categorieën 2 en 3 vormen de tweede groep, categorieën 4 en 5 vormen de derde, categorieën 6 en 7 de vierde en categorie 8 de vijfde groep,
- G_2 : (geslacht) man of vrouw,
- H_2 : (herkomst) westers of niet,
- H_3 : (herkomst) deelt de westerlingen van H_2 op in autochtonen en westers allochtonen; van 109 personen in de populatie is de herkomst niet bekend; deze worden behandeld alsof ze in de grootste groep zitten, namelijk de autochtonen,
- I_3 : (indeling) geeft aan of de persoon in de G4, de G27 of buiten de G31 woont,
- I_6 : (indeling) maakt binnen de G4 onderscheid tussen de gemeenten Amsterdam, Rotterdam, Den Haag en Utrecht,
- L_2 : (leeftijd) 65-plusser of niet; hier wordt de leeftijd van de respondent op de datum van enquêtering genomen,
- L_7 : (leeftijd) 18-24, 25-34, ..., 65-74, 75+ (tussenliggende intervallen van tien jaar) op de datum van enquêtering,
- L_{15} : (leeftijd) 18-19, 20-24, ..., 70-74, 75-79, 80-84, 85+ (tussenliggende intervallen van vijf jaar) op de datum van enquêtering,
- P_2 : (positie in huishouden) partner of niet-partner op datum van enquêtering,
- P_4 : (positie in huishouden) deelt de niet-partners op in drie categorieën: kinderen, alleenstaanden en eenouders; de 251.630 “overigen” worden bij de “niet-partners” gerekend omdat ze bij de steekproeftrekking van de Woningmarktmodule de trekkingskans van een niet-partner hebben gekregen; binnen de niet-partners worden ze bij de “alleenstaanden” geteld omdat dit de grootste groep is binnen de niet-partners; hetzelfde geldt voor de “overigen” in de steekproef en de elementen in de steekproef waarvoor deze variabele onbekend is,
- S_2 : (inkomen) maakt onderscheid tussen de 40 procent laagste inkomens en de overige 60 procent niet-lage inkomens; dit inkomen is gebaseerd op gegevens uit de Fiscale Database

(FiBase) van 2006; een persoon heeft een laag inkomen als het huishoudensinkomen bij de 40 procent laagste inkomens van de Fiscale Database hoort,

W_5 : (woonmilieu) centrum stedelijk, buiten centrum, groen stedelijk, centrum dorps en landelijk; van 6128 personen in de populatie is het woonmilieu (nog) niet bekend; deze worden behandeld alsof ze in de grootste groep zitten, namelijk “buiten centrum”; hetzelfde geldt voor de personen in de steekproef waarvan het woonmilieu niet bekend is,

K_3 : (koop/huur) maakt onderscheid tussen koopwoningen, huurwoningen en elementen waarvoor dit onbekend is,

C_{40} : (COROP-gebied) geeft aan in welk COROP-gebied de persoon woont.

Bij de weging worden de kenmerken van de respondenten vergeleken met populatietotalen op 1 januari 2009. Uitzondering hierop zijn de populatietotalen voor K_3 . Deze zijn geschat bij de Woningmarktmodule.

Merk op dat de eerste term in het weegmodel de variabelen bevat van het steekproefontwerp voor Sociaal Fysiek. De tweede tot en met de tiende term komen vrijwel overeen met de weging van Sociaal Fysiek 2006 [4]. De koop/huur variabele is toegevoegd opdat schattingen op basis van gewichten bij Sociaal Fysiek en de Woningmarktmodule consistent zijn. De variabele COROP-gebied wordt toegevoegd om te corrigeren voor ondervertegenwoordiging van bepaalde gebieden in Nederland. Zo zijn er maar 81 Zeeuwen in de steekproef (0,78%) terwijl er 295.352 Zeeuwen zijn in de populatie (2,31%).

Voor alle drie typen startgewichten wordt per respondent een correctiefactor c_i bepaald zodat het eindgewicht w_i het product is van c_i en het bijbehorende startgewicht d_i , Δ_i of δ_i . Dit wordt gedaan met behulp van het programma Bascula [3]. De verdeling van de correctiegewichten is te zien in Figuren 2, 3 en 4.

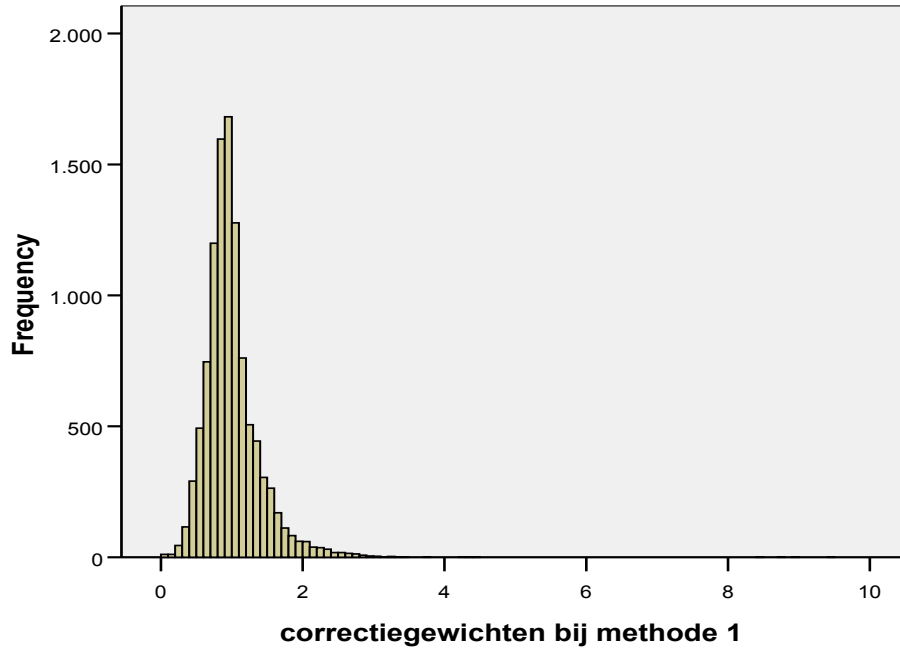
Uitgaande van het eerste type startgewichten $(d_i)_{i=1}^n$ krijgen we door lineair te wegen een set correctiegewichten waarvan er negen negatief zijn. Met het Huang-Fuller algoritme [2] worden in één iteratie alle negatieve gewichten begrensd (positief gemaakt). De mediaan van de correctiegewichten is 0,9393 en de standaardafwijking 0,4112. Het eerste en derde kwartiel zijn 0,7790 en 1,1440 en de extrema zijn 0,0692 en 9,4954.

Bij het tweede type startgewichten $(\Delta_i)_{i=1}^n$ vinden we maar liefst 105 negatieve correctiegewichten. Dit keer zijn twee iteraties nodig om te begrenzen. De mediaan is 0,5350 en de standaardafwijking is 1,0125. Het eerste en derde kwartiel zijn 0,3380 en 1,1180 en de extrema 0,0172 en 12,1067. We adviseren om deze gewichten niet te gebruiken, ten eerste omdat het aantal negatieve gewichten voor begrenzen erg groot is. Verder is de spreiding van de gewichten groot en de mediaan is erg laag. Een mogelijke verklaring voor de grote spreiding is de scheefheid ten gevolge van ongelijke trekkingskansen bij Sociaal Fysiek die niet verwerkt is in de insluitgewichten en gecorrigeerd dient te worden.

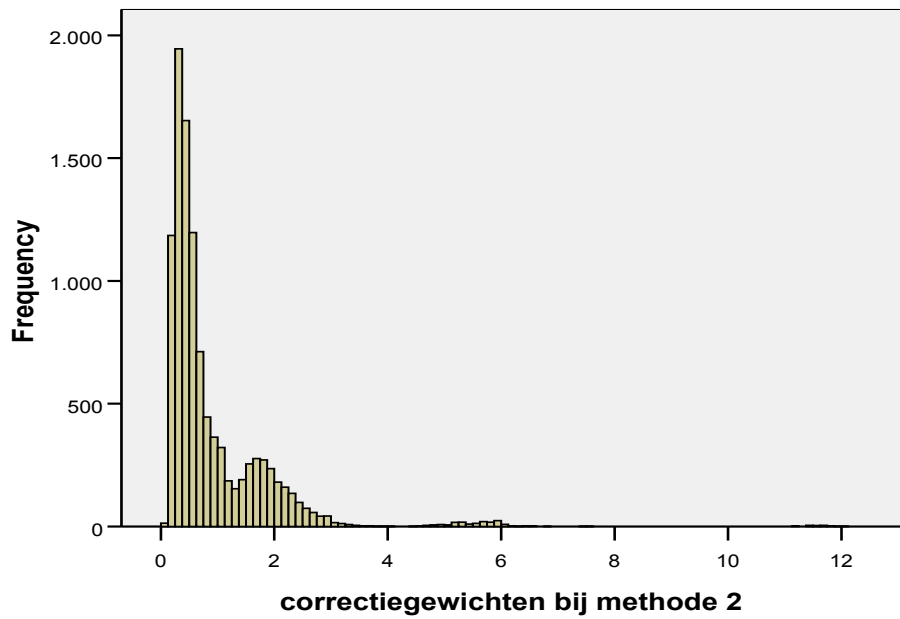
Voor het derde type startgewichten $(\delta_i)_{i=1}^n$ krijgen we tien negatieve correctiegewichten die in twee iteraties begrensd kunnen worden. De mediaan is 0,9653 en de standaardafwijking 0,5511.

Het eerste en derde kwartiel zijn 0,7938 en 1,2391 en de extrema zijn 0,0130 en 6,2712.

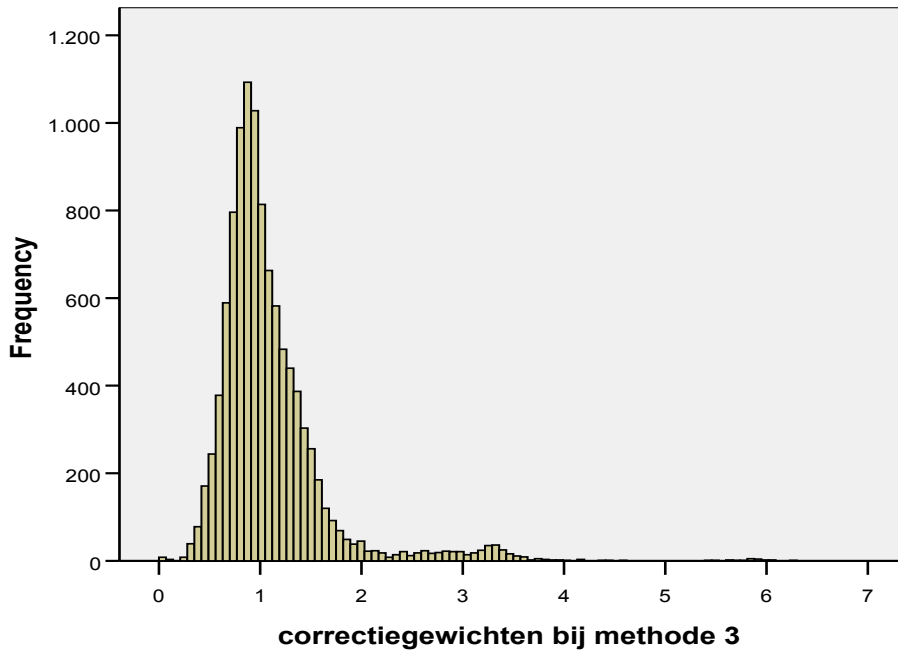
Figuur 2: Verdeling correctiegewichten bij methode 1



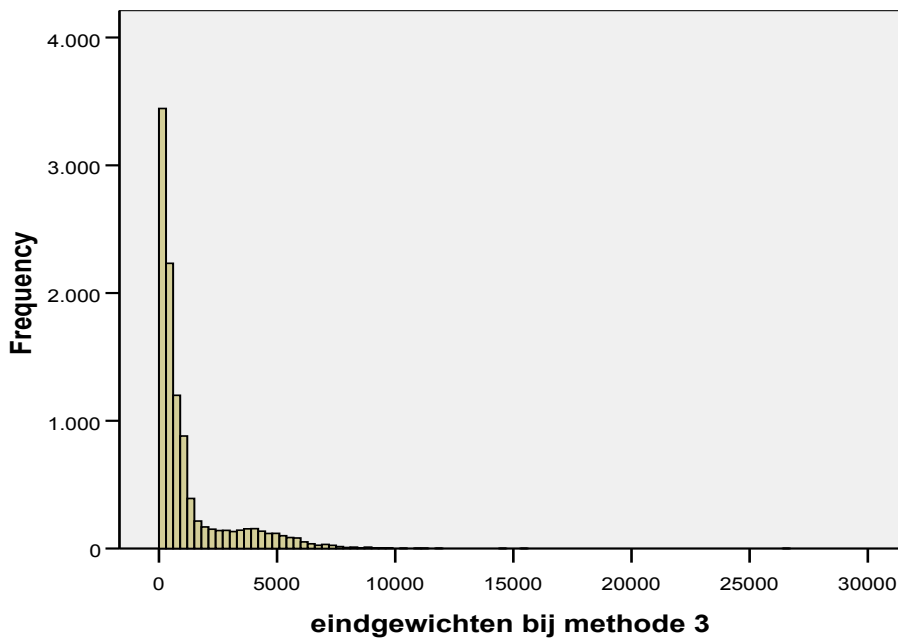
Figuur 3: Verdeling correctiegewichten bij methode 2



Figuur 4: Verdeling correctiegewichten bij methode 3



Figuur 5: Verdeling eindgewichten bij methode 3



Nu het tweede type gewichten ongeschikt is gebleken moet een keuze gemaakt worden tussen het eerste en het derde type. Onze keuze gaat uit naar het derde type vanwege de volgende redenen:

1. Ongelijke kansen bij de steekproeftrekking van de Woningmarktmodule zijn hier al in de startgewichten verwerkt.
2. Er is minder scheefheid met betrekking tot de kenmerken waarnaar gewogen is bij de Woningmarktmodule.

3. De mediaan van de correctiegewichten ligt dichterbij 1 dan bij de eerste set correctiegewichten.

We adviseren dus om de derde set gewichten te gebruiken en definiëren

$$w_i := c_i \delta_i.$$

De mediaan van de eindgewichten is 518 en de standaardafwijking 1.676. Het eerste en derde kwartiel zijn 225 en 1.239 en de extrema zijn 0,64 en 26.661, zie Figuur 5.

Om de plausibiliteit van de eindgewichten te testen wordt de samenstelling van het huishouden en de grootte van enkele gemeenten geschat aan de hand van deze gewichten en vergeleken met de juiste populatiegegevens. De relatieve fout is als volgt gedefinieerd:

$$\text{relatieve fout} = \frac{\text{geschat}}{\text{exact}} - 1.$$

Tabel 3: Populatieschattingen voor samenstelling huishouden

samenstelling huishouden	exact aantal	geschat aantal	relatieve fout
Eenpersoonshuishouden	2611245	2423421	0,005
Paar	4260181	4538494	0,040
Paar met kinderen	5020298	4894891	-0,035
Eenoudergezin	738927	689116	-0,077
Overig of niet koppel	132225	216953	0,371

We zien een grote fout bij de categorie “overig of niet koppel”. Voor de andere categorieën liggen schattingen en realiteit dichterbij elkaar.

Tabel 4: Populatieschattingen voor de grootste tien gemeenten

Gemeente	exact aantal	geschat aantal	relatieve fout
Amsterdam	611071	611071	0
Rotterdam	464743	464743	0
's-Gravenhage	376212	376212	0
Utrecht	239610	239610	0
Eindhoven	170909	189953,4	0,111
Tilburg	160532	176936,5	0,102
Groningen	154398	171309,5	0,110
Breda	134633	147677,7	0,097
Enschede	123658	119745,5	-0,032
Nijmegen	130878	114850,4	-0,122

We adviseren om geen uitspraken te doen op gemeenteniveau als het over kleinere gemeenten gaat, omdat de relatieve fout hiervoor groter kan zijn. Een mogelijke verklaring is dat het aantal steekproefelementen in deze gemeenten te klein is.

4. Gewichten voor huishoudens

Naast persoonsgewichten wordt per respondent een huishoudensgewicht bepaald voor het huishouden waar de respondent tot behoort. Bij de bepaling van deze gewichten speelt de variabele huisvestingssituatie een rol. Deze variabele kan de volgende zes waarden aannemen:

1. woning,
2. zelfstandige wooneenheid,
3. onzelfstandige wooneenheid,
4. bewoonde andere ruimte,
5. inwonend huishouden,
6. lid huishouden.

De verzameling van respondenten waarvoor de variabele huisvestingssituatie waarde j aanneemt wordt aangegeven met u_j . De verzameling van elementen waarvoor de huisvestingssituatie ongelijk is aan j (het complement van u_j) geven we aan met u_j^C .

Om tot een formule te komen voor de huishoudensgewichten dient rekening gehouden te worden met de volgende feiten:

1. Huishoudens met veel leden in de doelpopulatie hebben een grotere kans om in de steekproef te komen dan huishoudens met weinig leden omdat de steekproef getrokken is op persoonsniveau.
2. Vragen over huishoudens zijn alleen gesteld aan personen die tot de huishoudkern behoren. Dit zijn de personen in u_6^C .
3. De gewichten moeten sommeren naar het aantal huishoudens in Nederland. Op 1 januari 2009 was dit gelijk aan $NH = 7.312.579$.

Om het huishoudensgewicht te bepalen dienen de persoonsgewichten w_i gedeeld te worden door het aantal 18-plussers in de huishoudkern van het huishouden waartoe respondent i behoort. Dit aantal geven we aan met η_i (Bij de weging van de Woningmarktmodule wordt dit aantal $hhfac_i$ genoemd). De definitie voor de huishoudensgewichten w_i^H is dus als volgt:

1. Indien de respondent i een lid van het huishouden is, i.e. $i \in u_6$,

$$w_i^H := 0.$$

2. Indien de respondent i niet een lid van het huishouden is, i.e. $i \in u_6^C$,

$$w_i^H := \mu \frac{w_i}{\eta_i}.$$

De schalingsfactor μ wordt op dusdanige wijze gekozen dat

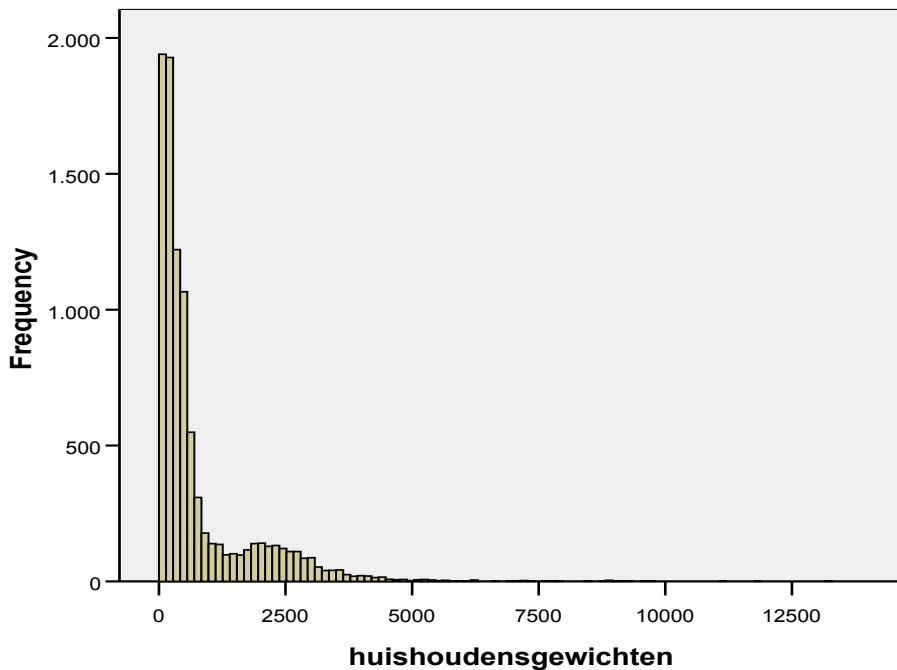
$$\sum_{i \in u_6^C} w_i^H = NH.$$

Hieruit volgt de formule

$$\mu = \frac{NH}{\sum_{i \in u_6^c} \frac{w_i}{\eta_i}}$$

We vinden $\sum_{i \in u_6^c} \frac{w_i}{\eta_i} \approx 7.335.306$. Dit ongeveer gelijk aan NH . Dus μ is vrijwel gelijk aan 1. De precieze waarde is $\mu = 0,997$. De verdeling van de gewichten is weergegeven in Figuur 6.

Figuur 6: Verdeling huishoudensgewichten



5. Gewichten voor woningvoorraad

Behalve gewichten voor personen en huishoudens worden gewichten voor woningvoorraad w_i^V bepaald. Eerst worden startgewichten berekend op basis van de huishoudensgewichten voor de respondenten met huisvestingssituatie 1. Vervolgens wordt een weegmodel met variabelen voor woningen gebruikt om de eindgewichten te bepalen. De startgewichten voor woningvoorraad d_i^V zijn als volgt gedefinieerd:

1. Indien de huisvestingssituatie niet 1 is (2 t/m 6)

$$d_i^V := 0 \quad \text{als } i \in u_1^c.$$

2. Indien de huisvestingssituatie 1 is

$$d_i^V := \nu w_i^H \quad \text{als } i \in u_1.$$

De waarde van ν wordt op dusdanige wijze gekozen dat

$$\sum_{i \in u_1} d_i^V = NV,$$

waar $NV=7.028.606$ staat voor het aantal woningen in Nederland. Hieruit volgt

$$\nu = \frac{NV}{\sum_{i \in u_1} w_i^H}.$$

Omdat $\sum_{i \in u_1} w_i^H \approx 7.051.964$ krijgen we $\nu \approx 0,997$.

De correctiegewichten worden bepaald op basis van het volgende weegmodel:

$$(K_2 \times \Gamma_2 \times B_9) + (I_2 \times B_9) + (I_2 \times K_2 \times \Gamma_2).$$

De variabelen zijn als volgt gedefinieerd:

B_9 : (bouwjaarklasse) geeft aan in welke periode de woning gebouwd is; we onderscheiden 1906 of eerder, 1907-1930, 1931-1944, 1945-1959, 1960-1970, 1971-1980, 1981-1990, 1991-2000, 2001 of later,

I_2 : (indeling) onderscheidt de G31 van de rest,

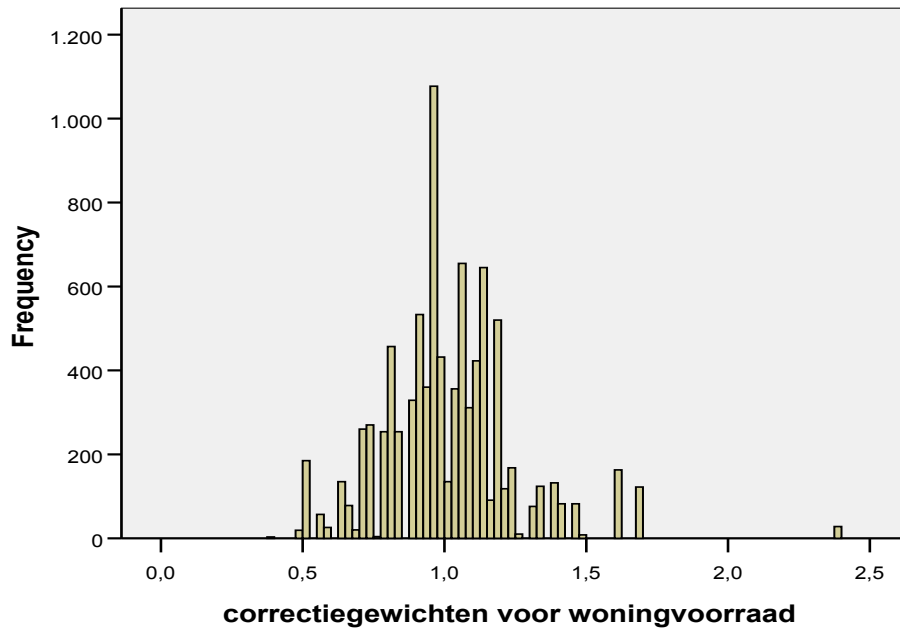
K_3 : (koop/huur) onderscheidt koop- en huurwoningen,

Γ_2 : (eengezins/meergezins) onderscheidt eengezins- en meergezinswoningen.

Dit weegmodel is vrijwel identiek aan het model dat door ABF Research is gebruikt voor de bepaling van woningvoorraadgewichten voor Sociaal Fysiek 2006 [4]. De populatietotalen die gebruikt zijn, zijn afkomstig uit SysWov 2008 (Systeem Woningvoorraad).

De verdeling van de correctiegewichten c_i^V wordt weergegeven in Figuur 7. De mediaan is 0,9843, de standaardafwijking 0,2317, de kwartielen 0,8978 en 1,1268 en de extrema 0,3754 en 2,3840.

Figuur 7: Verdeling correctiegewichten voor woningvoorraad

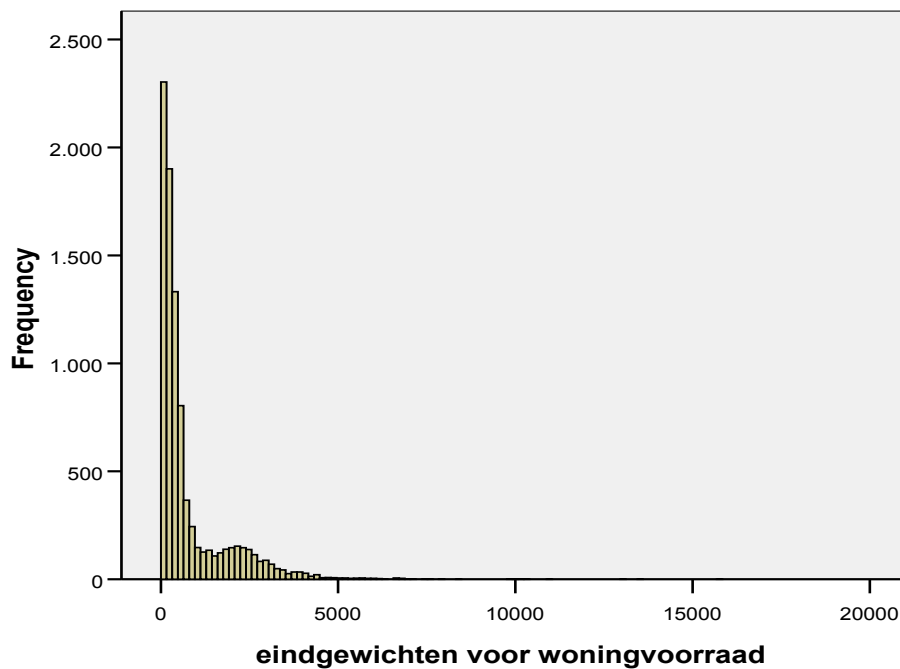


De eindgewichten w_i^V zijn het product van de startgewichten en de correctiegewichten:

$$w_i^V = c_i^V d_i^V .$$

De mediaan van deze eindgewichten is 354, de standaardafwijking is 1.075, de kwartielen zijn 157 en 829 en de extrema zijn 0,76 en 15.809, zie Figuur 8.

Figuur 8: Verdeling eindgewichten voor woningvoorraad



6. Tweede fase waarin de G31 zonder woningopname-deel niet is meegeteld

In de vorige paragrafen zijn wegingen uitgevoerd voor de gehele groep van 10.438 respondenten waarin ook de G31 zonder woningopname-deel is vertegenwoordigd. Uiteindelijk is ervoor gekozen om respondenten zonder WOopname alleen mee te tellen wanneer deze buiten de G31 wonen. Op deze manier blijven 9.060 personen over, zie Tabel 1. Er is voor gezorgd dat informatie van de 1.378 personen, waaraan uiteindelijk geen gewicht mag worden toegekend, toch verwerkt is in de gewichten van de groep van 9.060 respondenten die overblijft. De toekenning van de uiteindelijke persoons- huishoudens- en woningvoorraadgewichten aan deze groep wordt beschreven in deze paragraaf.

Ten eerste worden de respondenten die overblijven opnieuw in strata ingedeeld. De stratumindeling is anders dan in paragraaf 2. In de nieuwe indeling wordt geen onderscheid gemaakt tussen personen met partner en personen zonder partner. Hierdoor zijn er geen 96 strata, maar slechts naar 48 strata. De reden voor deze samenvoeging is het feit dat het stratum dat gevormd wordt door categorie 3 en subklasse 6A leeg is. In de nieuwe indeling zitten toch 11 personen zowel in categorie 3 als in klasse 6, zie Tabel 5. Het stratum waartoe respondent $i \in \{1, 2, \dots, 9.060\}$ behoort wordt aangegeven met $s(i)$.

Tabel 5: Aantallen respondenten met partner naar categorie en klasse

categorie	Klasse						totaal
	1	2	3	4	5	6	
1	167	157	201	272	82	45	924
2	191	192	159	176	85	34	837
3	39	30	83	72	22	11	257
4	124	107	146	186	64	24	651
5	97	110	343	527	151	60	1288
6	148	171	445	935	165	127	1991
7	43	50	186	489	88	61	917
8	44	75	375	1240	271	190	2195
Totaal	853	892	1938	3897	928	552	9060

De startgewichten worden bepaald op dezelfde manier als de startgewichten van het derde type in paragraaf 2. Laat $N_{s(i)}$ het aantal personen in de populatie zijn in het stratum waartoe respondent i volgens de nieuwe indeling behoort. Het startgewicht δ_i is dan

$$\delta_i := \frac{N_{s(i)}}{\sum_{j \in s(i)} \tilde{w}_j} \tilde{w}_i.$$

De populatie waarnaar wordt opgehoogd is hier exact hetzelfde als in paragraaf 2. Op basis van de gewichten $(w_i)_{i=1}^{10.438}$ die bepaald zijn in paragraaf 3 worden populatieschattingen gemaakt voor de volgende vier variabelen: tevredenheid buurt, betrokkenheid, vertrouwen buurt, vertrouwen algemeen. Deze schattingen worden vervolgens toegevoegd aan het in paragraaf 3 besproken

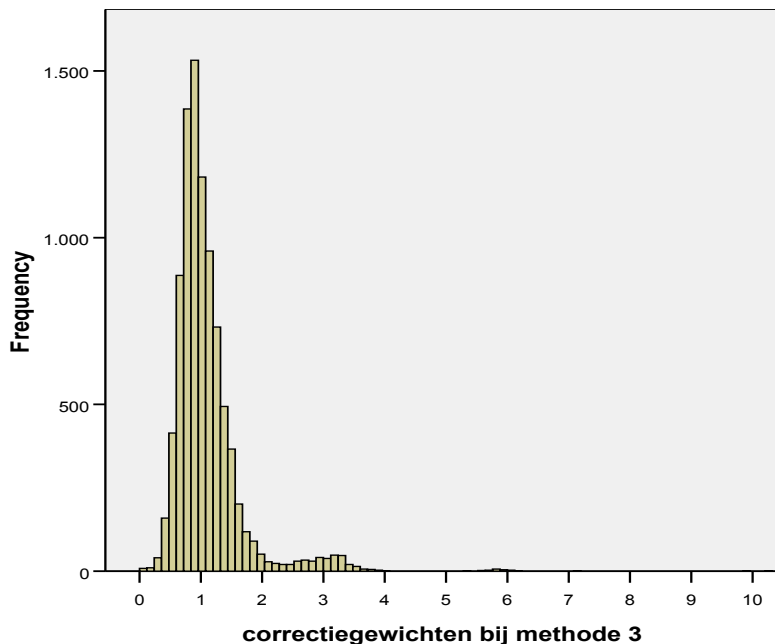
weegmodel om nieuwe correctiegewichten τ_i te bepalen op basis van de nieuwe startgewichten. De uiteindelijke eindgewichten w_i worden verkregen door τ_i met δ_i te vermenigvuldigen.

De huishoudensgewichten w_i^H worden uit w_i bepaald op dezelfde wijze als in paragraaf 2 is gedaan. Leden van het huishouden (huisvestingssituatie 6) krijgen gewicht 0. Voor de overige respondenten wordt eerst w_i gedeeld door η_i . Vervolgens wordt met een vaste factor vermenigvuldigd zodat de som van de huishoudensgewichten gelijk is aan NH .

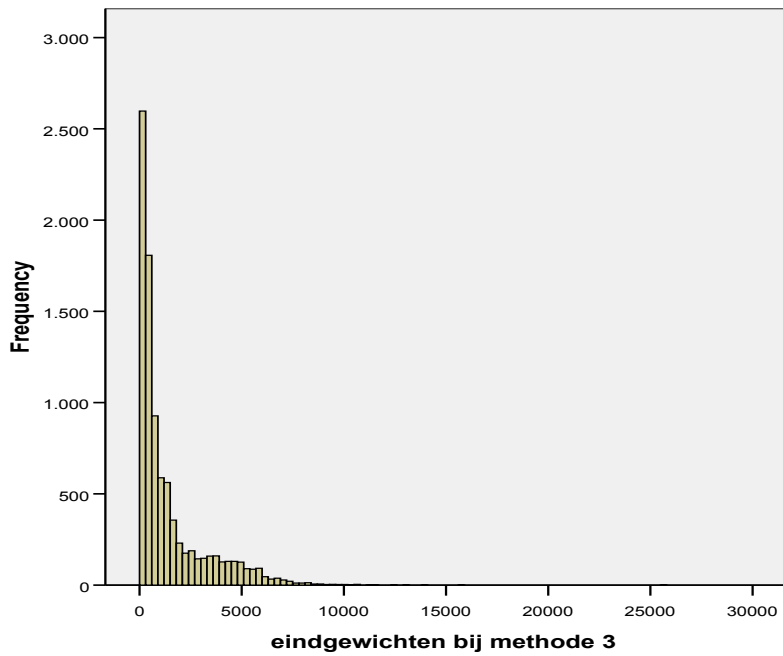
Ook voor de bepaling van de woningvoorraadgewichten wordt dezelfde methode gebruikt als eerder is gedaan voor de gehele groep van 10.438 respondenten. De startgewichten worden bepaald door de huishoudensgewichten voor huisvestingssituatie 1 zo te schalen dat ze sommeren naar NV . Vervolgens worden de correctiegewichten en de eindgewichten w_i^V bepaald aan de hand van exact hetzelfde weegmodel als in paragraaf 5.

In Figuren 6 en 7 is de verdeling van de correctiegewichten en eindgewichten $(w_i)_{i=1}^{9,060}$ te zien. Voor begrenzen zijn er 16 negatieve correctiegewichten die in drie iteraties begrensd kunnen worden. De mediaan is 0,9690 en de standaardafwijking 0,5710. Het eerste en derde kwartiel zijn 0,7887 en 1,2282 en de extrema zijn 0,0060 en 10,2824. De mediaan van de eindgewichten is 632 en de standaardafwijking 1.758. De kwartielen zijn 261 en 1.745 en de extrema zijn 0,71 en 25.784.

Figuur 9: Verdeling uiteindelijke correctiegewichten

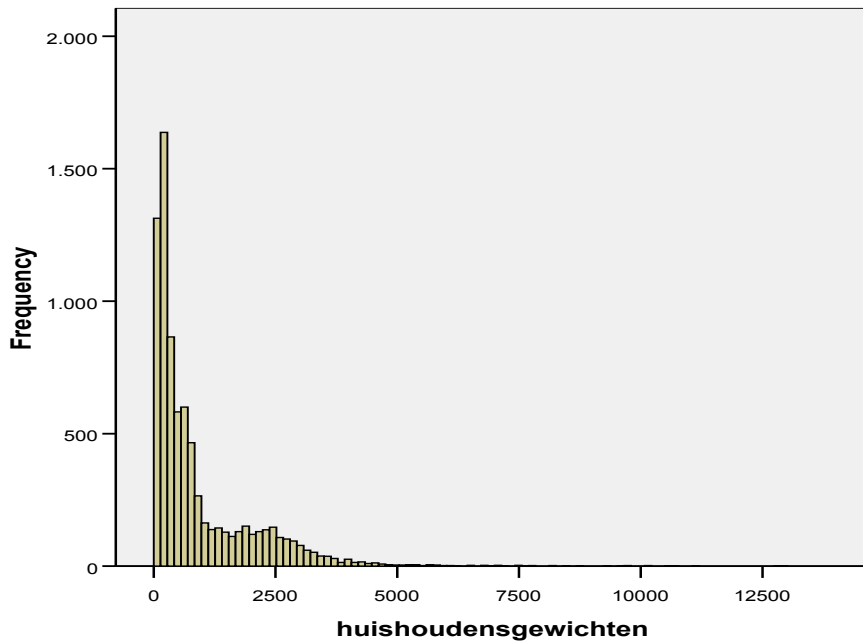


Figuur 10: Verdeling uiteindelijke persoonsgewichten



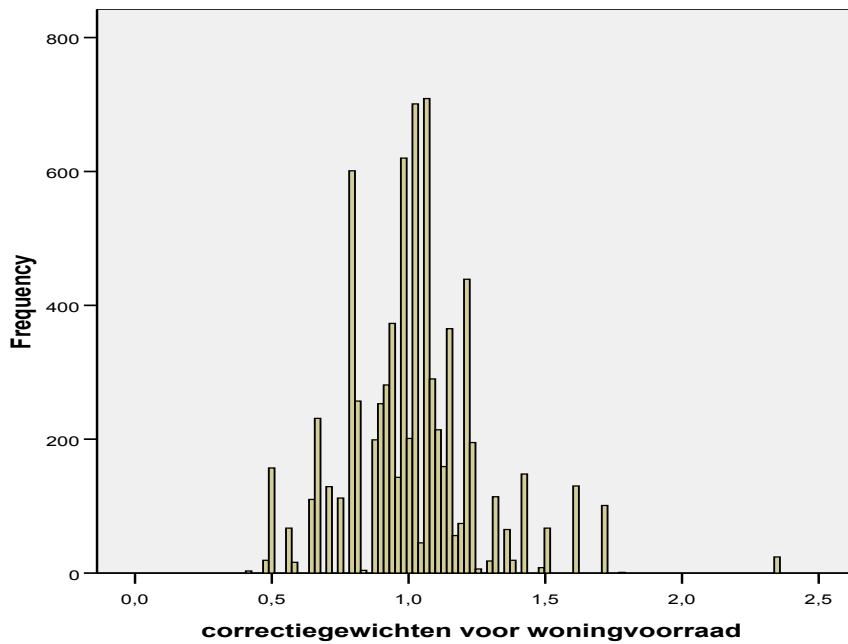
In Figuur 11 is de verdeling van de huishoudensgewichten weergegeven. De mediaan is 461, de standaardafwijking 1.117, de kwartielen 188 en 1.221 en de extrema 2,30 en 12.957.

Figuur 11: Verdeling uiteindelijke huishoudensgewichten

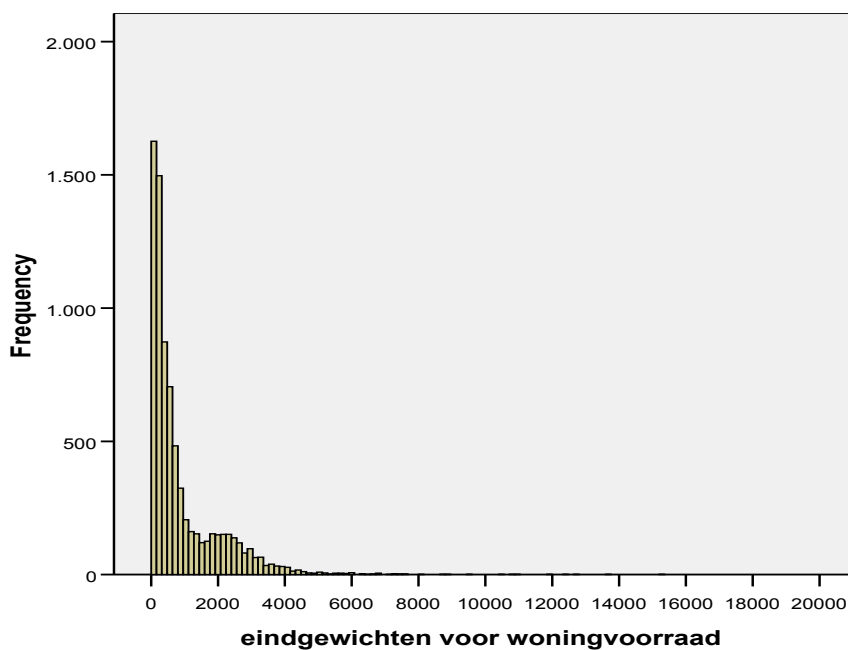


De gewichten voor woningvoorraad zijn weergegeven in Figuren 12 en 13. De mediaan van de correctiegewichten is 1,0169, de standaardafwijking 0,2367, de kwartielen 0,8964 en 1,1194 en de extrema 0,4086 en 2,3550. De mediaan van de eindgewichten is 451, de standaardafwijking is 1.135, de kwartielen 186 en 1.200 en de extrema 1,15 en 15.340.

Figuur 12: Verdeling uiteindelijke correctiegewichten woningvoorraad



Figuur 13: Verdeling uiteindelijke woningvoorraadgewichten



De plausibiliteitsanalyses in paragraaf 3 voor de persoonsgewichten w_i zijn ook uitgevoerd aan de hand van de nieuwe persoonsgewichten \bar{w}_i . In Tabellen 6 en 7 zijn de resultaten weergegeven.

Tabel 6: Populatieschattingen voor samenstelling huishouden

samenstelling huishouden	exact aantal	geschat aantal	relatieve fout
Eenpersoonshuishouden	2611245	2636350	0.009614
Paar	4260181	4429338	0.039707
Paar met kinderen	5020298	4838759	-0.03616
Eenoudergezin	738927	682977	-0.07572
Overig of niet koppel	132225	175452	0.326918

Tabel 7: Populatieschattingen voor de grootste tien gemeenten

gemeente	exact aantal	geschat aantal	relatieve fout
Amsterdam	611071	611071	0
Rotterdam	464743	464743	0
's-Gravenhage	376212	376212	0
Utrecht	239610	239610	0
Eindhoven	170909	192307.6	0.125
Tilburg	160532	181928.5	0.133
Groningen	154398	171101.1	0.108
Breda	134633	146404	0.087
Enschede	123658	116736.4	-0.056
Nijmegen	130878	110758.7	-0.154

7. Dankwoord

Speciale dank gaat uit naar José Gouweleeuw, Saskia Janssen-Jansen, Frank Meessen en Harm Jan Boonstra.

8. Referenties

- [1] Van Berkel, C. A. M. en Janssen B. Steekproefontwerp Sociaal Fysiek 2009, versie 4.
- [2] Huang, E.T. and Fuller, W.A. (1978), Nonnegative regression estimation for sample survey data, Proceedings of the Social Statistics Section, American Statistical Association, pp. 300-305.
- [3] Nieuwenbroek, N. en H.J. Boonstra (2002). Bascula 4.0 Reference Manual. Centraal Bureau voor de Statistiek, Heerlen.
- [4] Onderzoeksdocumentatie WoON (2007); module Sociaal Fysiek. ABF Research.