



## Weging WOON 2012

**Erwin Vondenhoff**

*Abstract: In deze nota wordt de toekenning van gewichten aan de respondenten van WOON 2012 besproken. Een aantal verbeteringen ten opzichte van de ophoging van WOON 2009 komen aan de orde. In tegenstelling tot 2009 wordt in 2012 een aantal kenmerken op huishoudensniveau in plaats van op persoonsniveau gewogen. Dit wordt gedaan door kwantitatief te wegen, waarbij het aantal leden in de huishoudkern de kwantificatievariabele is. Daarnaast is in 2012 gebruik gemaakt van registers voor hulpvariabelen waar CBS in 2009 geen beschikking over had. Ten derde zijn een aantal gebieden aangewezen waarvoor de gewichten optellen tot het juiste aantal inwoners of huishoudens. Dit kunnen wijken, combinaties van wijken, ISV-gebieden, hele gemeenten of soms hele COROP-gebieden of provincies zijn. De oversamplingdeelnemers hebben aangegeven dat dit belangrijke gebieden zijn, waarover uitspraken gedaan moeten worden. De variabele die deze gebieden onderscheidt is gekruist met leeftijd, geslacht, eigendom (huur/koop) en woontype (eengezins/meergezins).*

### 1. Inleiding

Het doel van de Woningmarktmodule, het basisonderzoek van het WOON, is het verschaffen van informatie over de actuele situatie op de woningmarkt. Belangrijke uitkomsten van dit onderzoek gaan over de eigendom van woningen, dat wil zeggen het feit of men een huurwoning dan wel een koopwoning heeft, tevredenheid over de woning en de woonomgeving en verhuisgedrag.

Aangezien er onderzoek gedaan wordt op zowel het niveau van personen als het niveau van huishoudens worden er twee doelpopulaties gedefinieerd. Op persoonsniveau hebben uitkomsten betrekking op alle personen die op 1 januari 2012 minstens 18 jaar zijn, deel uitmaken van particuliere huishoudens en ingeschreven zijn in de Gemeentelijke Basisadministratie (GBA). Voor huishoudens is de doelpopulatie gedefinieerd als alle particuliere huishoudens die op 1 januari 2012 een huisvesting in Nederland hebben en waarbinnen ten minste één persoon 18 jaar of ouder is. In dit document wordt zowel de toekenning van de gewichten op persoonsniveau als op het niveau van huishoudens beschreven.

Bij WOON wordt onderscheid gemaakt tussen huishoudkernleden en leden van het huishouden. Alleen de huishoudkernleden krijgen vragen over het huishouden. Wanneer een huishoudkernlid een partner heeft, bestaat de huishoudkern uit twee personen. Heeft een huishoudkernlid geen partner dan bestaat de huishoudkern uit één persoon. Uit de doelpopulatie is een steekproef getrokken waarbij personen die volgens de GBA een partner hebben een trekkingskans hebben gekregen die twee keer zo klein is als de trekkingskans van iemand die geen partner heeft.

Hierdoor heeft een huishouden met één lid in de huishoudkern een even grote trekkingskans als een huishouden met twee huishoudkernleden. In bepaalde gebieden in Nederland is de steekproeffractie groter dan in andere gebieden, voor details zie [4].

Bij de weging wordt voor deze ongelijkheid in trekkingskansen gecorrigeerd. Ook voor selectiviteit in de respons wordt gecorrigeerd. Ten slotte wordt door te wegen afgedwongen dat bepaalde randtotalen waarnaar gepubliceerd wordt exact overeenkomen met de opgehoogde respons.

In paragraaf 2 wordt de toekenning van de insluitgewichten besproken. Insluitgewichten corrigeren voor scheefheid ten gevolge van ongelijke trekkingskansen in de steekproef. In paragraaf 3 wordt de toekenning van de correctie- en eindgewichten besproken. Correctiegewichten corrigeren voor het feit dat bepaalde personen beter responderen dan anderen. Deze correctie wordt gedaan aan de hand van een aantal achtergrondkenmerken van de respondenten. Deze kenmerken zijn samengevat in een zogenaamd weegmodel. In paragraaf 4 bespreken we hoe de variabelen zijn afgeleid die van belang zijn voor de weging. In paragraaf 5 worden de uiteindelijke gewichten besproken en in paragraaf 6 worden conclusies getrokken.

## **2. De strata en de insluitgewichten**

Het gewicht van een respondent geeft weer hoeveel elementen in de populatie vertegenwoordigd worden door deze ene respondent. De eerste stap bij het wegen is het vastleggen van de insluitgewichten, ook wel startgewichten genoemd. Bij de steekproeftrekking zijn bepaalde groepen met opzet over- of ondervertegenwoordigd. Insluitgewichten corrigeren voor selectiviteit die het steekproefontwerp met zich meebrengt. De populatie wordt opgedeeld in de groepen waarop het steekproefontwerp is gebaseerd. Vervolgens wordt, ter bepaling van de insluitgewichten, voor elke groep het aantal personen in de populatie gedeeld door het bijbehorende aantal respondenten. Op deze wijze wordt niet alleen gecorrigeerd voor de ongelijke trekkingskansen, maar er wordt ook een eerste correctie voor selectieve non-respons aangebracht. Als een bepaalde groep namelijk slecht respondeert, worden de insluitgewichten groot, zodat ze, ondanks het tegenvallende aantal respondenten, optellen tot het juiste aantal personen in de populatie. Dit is echter niet voldoende om volledig te corrigeren voor selectiviteit ten gevolge van non-respons. De insluitkansen zijn immers gebaseerd op een beperkt aantal variabelen, terwijl het responsgedrag afhangt van veel meer achtergrondkenmerken. Daarom worden in een volgende stap meer achtergrondkenmerken gebruikt om de zogenaamde correctiegewichten te bepalen. Dit is beschreven in paragraaf 3.

De insluitgewichten zijn bepaald op basis van de strata waartoe de respondenten behoren. Het stratum waartoe men behoort is namelijk bepalend voor de insluitkans die gehanteerd is bij de steekproeftrekking. Wanneer men in een gebied woont waarin relatief veel steekproefpersonen getrokken worden heeft men een grote insluitkans en daarom een klein insluitgewicht. In een gebied waar weinig steekproefpersonen getrokken worden is de insluitkans juist klein en het insluitgewicht groot. Een afzonderlijke waarneming telt als het ware sterker mee als de ondervraagde in een gebied woont waar de steekproeffractie relatief klein is. Eén waargenomen steekproefelement in een gebied met een kleine steekproeffractie vertegenwoordigt immers een groter aantal elementen in de populatie dan een waargenomen element in een gebied met een grote steekproeffractie. Het insluitgewicht wordt berekend door in de populatie het totale aantal

elementen in het stratum waartoe het steekprofelement behoort, te delen door het totale aantal respondenten in dat stratum. Deze stratumindeling is bij WOON gebaseerd op twee kenmerken, namelijk een regionaal kenmerk en het feit of iemand volgens de registratie een partner heeft of niet.

Bij de regionale indeling onderscheiden we zogenaamde primaire eenheden. De primaire eenheden zijn per definitie de gebieden die bij de steekproeftrekking als apart zijn beschouwd. In andere woorden, een primaire eenheid is een gebied waarbinnen iedereen dezelfde insluitkans gekregen heeft, afgezien van het feit dat er onderscheid wordt gemaakt tussen partners en niet-partners.

Bij de indeling in primaire eenheden worden onder andere de gemeenten in de G18 onderscheiden van de rest. Binnen deze gemeenten wordt onderscheid gemaakt tussen een oversamplingsgebied en een regulier gebied. Daarnaast is door oversamplingsdeelnemers gewenst om binnen bepaalde gebieden een bepaald aantal respondenten te halen. Deze gebieden, die overeen kunnen komen met wijken, (delen van) gemeenten, COROP-gebieden of provincies, worden dus opgevat als aparte primaire eenheden. Ten slotte wordt per COROP-gebied het restant opgevat als een aparte primaire eenheid.

Binnen elke primaire eenheid wordt onderscheid gemaakt tussen personen met of zonder partner, zodat het aantal strata gelijk is aan het aantal primaire eenheden vermenigvuldigd met twee. Iemand zonder partner heeft bij WOON namelijk een twee keer zo grote kans gekregen om in de steekproef terecht te komen als iemand met partner in dezelfde primaire eenheid. Hierdoor hebben per primaire eenheid, alle huishoudens dezelfde insluitkans.

### **3. Het weegmodel, de persoonsgewichten en de huishoudensgewichten**

In deze paragraaf beschrijven we de bepaling van de uiteindelijke eindgewichten op basis waarvan schattingen voor doelvariabelen gemaakt kunnen worden. De insluitgewichten dienen hierbij als uitgangspunt. Deze worden met een zo klein mogelijke correctie aangepast tot de eindgewichten. Voor details van dit zogenaamde lineaire wegen verwijzen we naar [2].

De eindgewichten dienen te voldoen aan een aantal eisen die vastgelegd worden in het weegmodel. Deze eindgewichten zijn persoonsgewichten. Dat wil zeggen dat ze weergeven hoeveel personen in de populatie worden vertegenwoordigd door één respondent.

De respondenten worden opgedeeld in twee groepen, namelijk de huishoudkernleden en de zogenaamde “leden van het huishouden”. In een gezin met twee ouders en twee kinderen ouder dan 18 jaar, vormen de ouders de huishoudkern en worden de kinderen opgevat als leden van het huishouden. Alleen huishoudkernleden hebben vragen beantwoord over het huishouden.

De steekproef van WOON kan niet alleen als een personensteekproef maar ook als een huishoudenssteekproef worden opgevat. De personensteekproef is de gehele steekproef, terwijl de huishoudenssteekproef gevormd wordt door die personen in de steekproef die geen leden van het huishouden zijn. De overgebleven huishoudkernleden vertegenwoordigen als het ware hun huishouden in de huishoudenssteekproef.

Naast persoonsgewichten worden huishoudensgewichten gedefinieerd voor alle respondenten die

huishoudkernlid zijn. Deze geven weer hoeveel huishoudens in de populatie vertegenwoordigd worden door het huishouden waartoe de betreffende respondent behoort. De leden van het huishouden krijgen geen huishoudensgewicht omdat ze geen deel uitmaken van de huishoudenssteekproef.

De huishoudkernleden krijgen een huishoudensgewicht dat gelijk is aan het persoonsgewicht gedeeld door het aantal huishoudkernleden in het huishouden waartoe de respondent behoort. Een persoon in een huishouden waarvan de huishoudkern uit twee personen bestaat, vertegenwoordigt als het ware slechts de helft van zijn/haar huishouden. Daarom worden de betreffende persoonsgewichten door twee gedeeld. Op deze wijze tellen de huishoudensgewichten van personen met twee huishoudkernleden op tot het aantal huishoudens met twee huishoudkernleden. Het aantal personen verschilt immers een factor twee van het aantal huishoudens.

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

Stratum(402)  
+ Geslacht(2) × Leeftijd(15)  
+ Geslacht(2) × Herkomst(7)  
+ Leeftijd(7) × Herkomst(7)  
+ Landsdeel(4) × Herkomst(7)  
+ Provincie(12) × Inkomen(5) (kwantitatief)  
+ WOZ-klasse(22)  
+ Gemeente(418)  
+ Positie in huishouden(5) × Leeftijd(11)  
+ Aantal personen in huishouden (5)  
+ Gebied(122) × Leeftijd(7)  
+ Gebied(122) × Geslacht(2)  
+ Gebied(122) × Eigendom(3) (kwantitatief)  
+ Gebied(122) × Woontype(3) (kwantitatief)

In Appendix A is de betekenis van alle variabelen weergegeven.

Bij de weging worden de kenmerken van de respondenten vergeleken met populatietotalen op 1 januari 2012. Voor een aantal variabelen is ten tijde van de weging geen data beschikbaar. Voor die variabelen is data uit voorafgaande jaren gekoppeld aan de populatie van 2012. In paragraaf 4 komen we hierop terug.

In veel opzichten lijkt het bovenstaande weegmodel op het model dat in 2009 voor WOON 2009 is gehanteerd, zie [1]. Het weegmodel van 2009 is dan ook het uitgangspunt voor het model van 2012. De verschillen zijn als volgt.

Voor een drietal termen is op huishoudensniveau gewogen. Bij alle andere termen is op persoonsniveau gecorrigeerd. Zo tellen de gewichten van mannen en vrouwen per leeftijdsklasse op tot het aantal mannen en vrouwen in de populatie. De termen met eigendom, inkomen en woontype zijn daarentegen kwantitatief gewogen waarbij de kwantificatievariabele het aantal personen is dat tot de huishoudkern behoort. Dit houdt in dat niet de persoonsgewichten, maar de

huishoudensgewichten optellen tot het aantal huishoudens in de populatie voor elk type eigendom, inkomen en woontype. Hierbij is er ook voor gezorgd dat de leden van het huishouden, die niet tot de huishoudenskern behoren, geen huishoudensgewichten krijgen. In 2009 is niet kwantitatief naar inkomen en eigendom op huishoudensniveau gewogen, maar, net als bij alle andere achtergrondkenmerken, op persoonsniveau. De huishoudensgewichten zijn toen berekend door net als in 2012 voor de leden van de huishoudkern het persoonsgewicht te delen door het aantal leden in de huishoudkern. Dit laatste gaat dus op dezelfde manier. Vervolgens is echter in 2009 een kleine correctie aangebracht om ervoor te zorgen dat de gewichten precies optellen tot het totale aantal huishoudens in de doelpopulatie. Dit is gedaan door alle huishoudensgewichten met dezelfde factor te vermenigvuldigen. Dit is bij bovenstaand weegmodel niet meer nodig omdat de verkregen huishoudensgewichten na het toepassen van het weegmodel automatisch optellen tot het juiste aantal huishoudens.

In tegenstelling tot 2009 zijn de kenmerken eigendom en inkomen niet meer met elkaar gekruist. Er wordt namelijk niet naar de randtotalen van deze uitkruising gepubliceerd. Anderzijds kan een uitkruising leiden tot klassen waartoe weinig respondenten behoren en dit is niet wenselijk.

De termen leeftijd, geslacht, eigendom en woontype zijn gekruist met “gebied(122)”. Dit is een geografische indeling. In 2009 is eigendom naast inkomen alleen gekruist met provincie. De variabele “gebied(122)” is gedefinieerd, door bij elke oversamplingsdeelnemer de gebieden te onderscheiden waarin ten minste 600 respondenten gehaald dienden te worden. Soms zijn dit aparte wijken binnen één gemeente. In Den Haag zijn bijvoorbeeld zes aparte gebieden aangewezen. Buiten de oversampling worden de restanten van elk COROP-gebied opgevat als een apart gebied. In totaal vinden we op deze manier 122 gebieden. Soms behoort een COROP-gebied in zijn geheel tot de oversampling. Zo is de gehele provincie Zeeland één oversamplingsgebied. Voor de exacte gebiedsindeling bij “gebied(122)” verwijzen we naar Appendix B.

De weegterm “stratum(402)” is toegevoegd om de stratumindeling, op basis waarvan de insluitgewichten in de vorige paragraaf bepaald zijn, te behouden. Binnen elke primaire eenheid tellen de insluitgewichten op tot de populatie binnen de primaire eenheid. Na de weging zouden de eindgewichten deze eigenschap niet meer hebben, als deze stratumindeling niet in het weegmodel zou zijn opgenomen. Daarom is deze term toegevoegd.

Bij de weging van WOON 2009 zijn de termen geslacht(2)  $\times$  leeftijd(15) en geslacht(2)  $\times$  herkomst(7) gekruist met de variabele “oversampled”. Deze variabele neemt twee waarden aan door de gemeenten waarin oversampling heeft plaatsgevonden te onderscheiden van de gemeenten waarin dit niet het geval was. In 2009 is voor deze uitkruising gekozen omdat het onderscheid naar wel of niet oversampling op geen andere manier in het weegmodel zat. Omdat in het nieuwe model het stratum en gebied(122) zijn opgenomen is de uitkruising met oversampling overbodig.

In tegenstelling tot het weegmodel van 2009 is er voor gekozen om alle gemeenten te onderscheiden. In 2009 zijn de gemeenten in 123 groepen ingedeeld. Nu wordt elke gemeente afzonderlijk beschouwd. Er is één uitzondering. De respons van 2012 bevat geen enkele waarneming in de gemeente Schiermonnikoog. Daarom is in de populatie de gemeente Schiermonnikoog samengevoegd met de gemeente Ameland. De gewichten van respondenten op

Ameland tellen dus op tot de populatie van Ameland en Schiermonnikoog samen.

Net als in 2009 is ervoor gekozen om bij de variabelen leeftijd en positie in het huishouden, de enquêtewaarde op te hogen naar de registerdata. Wanneer iemand volgens het register een partner heeft, maar zelf zegt op de datum van enquêtering geen partner te hebben, dan telt deze persoon bij de weging mee bij de niet-partners. Ook bij de leeftijd, is de leeftijd op de enquêtedatum leidend. Wanneer de peildatum exact in het midden van de waarneemperiode zou vallen, zouden de aantallen respondenten, waarvoor de leeftijd op enquêtedatum hoger dan wel lager is dan de leeftijd op de peildatum, ongeveer aan elkaar gelijk zijn. Omdat de peildatum niet exact in het midden van de waarneemperiode ligt is de gemiddelde leeftijd van de respondent op de enquêtedatum te hoog of te laag. Door op te hogen van enquête naar register wordt dit probleem opgelost.

#### **4. De afleiding van de weegvariabelen voor populatie en respons**

De meeste achtergrondkenmerken voor respondenten zijn afgeleid uit de Gemeentelijke Basisadministratie van 1 januari 2012. Dit geldt ook voor de populatietotalen. In deze paragraaf bespreken we de uitzonderingen.

Zoals eerder gezegd, vormen de variabelen leeftijd en positie in huishouden een uitzondering. Net als in 2009, is de waarde op de datum van enquêtering opgehoogd naar de populatie. Meestal wordt voor respondenten de peildatum gebruikt, omdat ook in de populatie vergeleken wordt met leeftijd en positie in huishouden op deze peildatum. Eerder is uitgelegd waarom dit tot problemen kan leiden.

De WOZ-waarde is afkomstig uit het Woonruimtere register van 1 januari 2011. Voor een aantal personen in de populatie is de WOZ-waarde van de woning onbekend. Deze groep vormt een aparte categorie. De WOZ-klasse wordt aan de respons gekoppeld en gebruikt bij de ophoging. Onbekenden in de respons worden opgehoogd naar onbekenden in de populatie.

De variabele inkomen komt overeen met het besteedbaar huishoudensinkomen uit het Inkomensregister van 2010. Voor een aantal personen in de populatie van 1 januari 2012 kon het inkomen niet uit dit bestand gehaald worden. Deze personen zijn toegevoegd aan de laagste inkomensklasse. Omdat de groep klein is, kan deze namelijk niet als aparte categorie opgevat worden. De inkomensklassen zijn aan de respons gekoppeld en gebruikt bij de ophoging.

De weegvariabele “eigendom(3)” is voor alle personen afgeleid op basis van twee registervariabelen. Ten eerste is de variabele EHALG uit het Inkomensbestand van 2010 gebruikt. Daarnaast is een eigendom-variabele uit het Woonruimtere register van 1 januari 2012 gebruikt. De variabele EHALG geeft aan of het een huur of koopwoning betreft. Wanneer EHALG “koopwoning” is, is de weegvariabele “eigendom(3)” ook “koopwoning”. Is EHALG “huurwoning”, dan neemt de weegvariabele “eigendom(3)” de waarde “huurwoning” aan. Bij de weegvariabele “eigendom(3)” wordt bij huurwoningen ook onderscheid gemaakt naar corporatieve verhuur en overige verhuur. Dit gebeurt op basis van de variabele “eigendom” uit het Woonruimtere register, zie Figuur 1. Het komt voor dat er volgens de variabele EHALG sprake is van een huurwoning, terwijl we volgens de eigendom-variabele uit het Woonruimtere register met iemand met een koopwoning te maken hebben. Daarnaast kan het zo zijn dat EHALG of

“eigendom” onbekend is. De definitie van de weegvariabele “eigendom(3)” is voor deze gevallen af te lezen uit Tabel 1.

EHALG (Inkomensregister)	Eigendom(Woonruimteregeister)			
	koopwoning	Huur corporatie	Huur overig	onbekend
Eigen woning	1	1	1	1
Huur zonder toeslag	3	2	3	2
Huur met toeslag	3	2	3	3
Onbekend	3	2	3	3

*Figuur 1 : De afleiding van de weegvariabele eigendom(3). Deze weegvariabele neemt drie waarden aan: 1. koopwoning, 2. huurwoning corporatie, 3. huurwoning overig. De waarden van deze variabele hoeven niet noodzakelijkerwijs overeen te komen met de waarden van de variabele EHALG uit het Inkomensregister of eigendom uit het Woonruimteregeister.*

Ten slotte is het aantal huishoudkernleden afgeleid om de kwantificatievariabelen te kunnen bepalen en populatietotalen op huishoudensniveau te kunnen berekenen. Hiertoe is eerst voor alle personen in de populatie afgeleid of ze huishoudkernleden zijn of niet. Dit is gedaan op basis van de positie in het huishouden volgens registratie op 1 januari 2012. De alleenstaanden, partners en eenouders worden tot de huishoudkern gerekend. Kinderen behoren niet tot de huishoudkern. Voor de categorie “overig” is in de registratie van 1 januari 2012 bekend of ze tot de huishoudkern behoren of niet. Sommigen in deze groep behoren wel tot de huishoudkern, anderen niet. Wanneer voor de hele populatie bekend is wie tot de huishoudkern behoort, wordt per huishouden, op basis van het huishoudenskermerk “hhnr”, in de populatie het aantal leden in de huishoudkern geteld. Dit aantal is altijd één of twee en wordt aan elke persoon in de populatie als achtergrondkenmerk meegegeven.

De drie populatietabellen op huishoudensniveau in het weegmodel zijn eerst op persoonsniveau berekend, waarbij verder is uitgekruist naar twee extra variabelen, namelijk de variabele die aangeeft of iemand tot de huishoudkern behoort en de variabele die aangeeft hoeveel huishoudkernleden in het huishouden van de persoon zitten. Om de tweedimensionale tabel gebied(122)×eigendom(3) op huishoudensniveau te verkrijgen, maken we eerst de vierdimensionale tabel met gebied(122), eigendom(3) en deze twee extra variabelen op persoonsniveau. De tweedimensionale tabel op huishoudensniveau verkrijgen we door de aantallen in deze vierdimensionale tabel bij de leden van het huishouden (buiten de huishoudensker) niet mee te tellen, de aantallen bij huishoudkernleden in een huishouden met een huishoudkern bestaande uit twee personen voor de helft mee te tellen, en de aantallen bij huishoudkernleden in een huishouden met slechts één persoon in de huishoudkern één keer mee te tellen.

Vervolgens wordt opgehoogd naar de verkregen populatietotalen op huishoudensniveau. Hiertoe wordt afgedwongen dat per klasse de persoonsgewichten van de respondenten in de huishoudkern gedeeld door het aantal huishoudkernleden gelijk is aan het corresponderende populatietotaal. Het aantal huishoudkernleden dat hierbij gehanteerd is, is niet gelijk aan het aantal volgens de registratie, maar aan het aantal dat de respondent zelf opgeeft. Ook het gegeven of de respondent

tot de huishoudkern behoort wordt niet afgeleid op de wijze zoals dat voor de populatie is gedaan, zoals besproken in de voorafgaande paragrafen, maar het wordt uit de enquête gehaald. Het aantal personen in de huishoudkern volgens de respondent wordt “hhfac” genoemd. In bijlage C is opgenomen hoe het huishoudgewicht (hweegwon) is afgeleid (SPSS-syntax). In deze bijlage staat ook de afleiding van het gewenste huishoudgewicht (gweegwon).

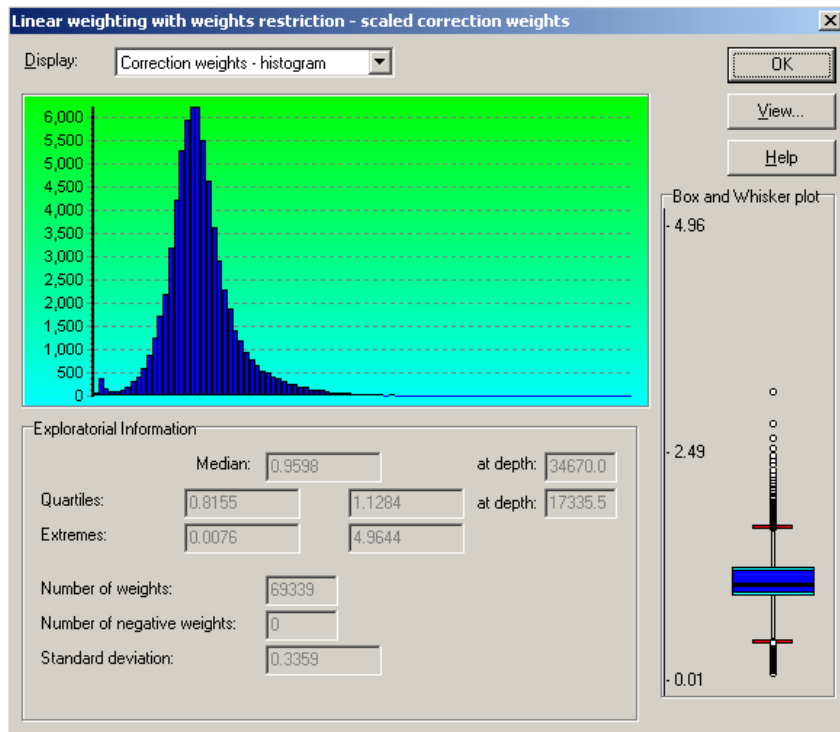
## 5. Resultaat van de weging

Lineair wegen met Bascula [2] leidt tot 295 negatieve gewichten. Deze zijn te begrenzen met het Huang Fuller algoritme. De statistieken van de gewichten na begrenzen zijn hieronder weergegeven in Figuur 2 en 3.

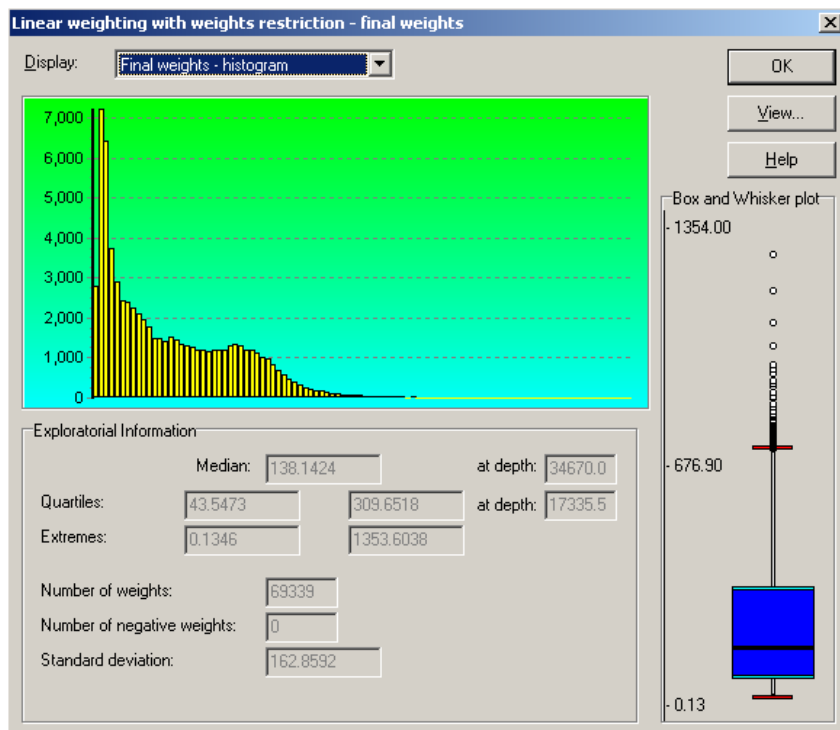
Door te wegen wordt in het algemeen de vertekening van schatters verkleind. Ook de variantie van schatters kan worden beïnvloed door gebruik te maken van ongelijke gewichten. De variantie van de regressieschatter voor doelvariabelen verandert ongeveer met een factor  $(1 + V_w^2)(1 - R^2)$ . Hierin is  $R \in [-1,1]$  de correlatie-coëfficiënt tussen de betreffende doelvariabele en de hulpvariabelen van het weegmodel. Verder is  $V_w^2$  de variatiecoëfficiënt van de gewichten. Deze coëfficiënt is per definitie het kwadraat van de standaardafwijking van de gewichten (de variantie van de gewichten) gedeeld door het kwadraat van het gemiddelde van de gewichten. Voor WOON geldt  $V_w^2 = 0,749$ . Als er geen verband is tussen de doelvariabele en de hulpvariabelen van het weegmodel, dan is  $R$  gelijk aan nul. In dit geval neemt de variantie van de regressieschatter toe met 74,9%. Vaak correleren hulpvariabelen met de doelvariabele. In dat geval is de factor waarmee de variantie toeneemt lager. Wanneer de correlatiecoëfficiënt tussen doel- en hulpvariabelen 0,654 of hoger is, zal deze factor lager dan één zijn. In dat geval wordt niet alleen vertekening maar ook variantie gereduceerd door te wegen.

Ook in Figuur 3 is te zien dat er veel variatie tussen de eindgewichten is. Dit heeft te maken met het feit dat de gehanteerde trekkingskans over de regio's sterk varieert. In andere woorden, de oversampling veroorzaakt de grote variatie in de gewichten.





*Figuur 2 : Correctiegewichten na het toepassen van het Huang-Fuller algoritme*



*Figuur 3 : Eindgewichten na het toepassen van het Huang-Fuller algoritme*

## 6. Conclusies

In vergelijking met 2009 is de weging in een aantal opzichten verbeterd. Ten eerste zijn een aantal variabelen direct op huishoudensniveau gewogen. Net als in 2009 zijn de huishoudensgewichten berekend door de persoonsgewichten te delen door het aantal personen in de huishoudkern. Echter, in 2009 moesten de gewichten met een schaalfactor vermenigvuldigd worden om af te dwingen dat de gewichten optellen tot het juiste aantal huishoudens in de populatie volgens registratie. Dit laatste is in 2012 niet meer nodig omdat het aantal huishoudens exact klopt vanwege de kwantitatieve weegtermen.

Een ander verbeterpunt is dat de gewichten binnen bepaalde delen van de oversamplingsgebieden optellen tot de juiste aantallen. Voor elke oversamplingsdeelnemer komen bepaalde randtotalen overeen met gebruikte registers. Hiervoor is gebruik gemaakt van de variabele gebied(122).

## 7. Referenties

- [1] Joep M.S. Burger, José M. Gouweleeuw & Bart Buelens, Weging Woononderzoek Nederland 2009, basismodule Woningmarkt met oversampling
- [2] Nieuwenbroek, N. en H.J. Boonstra (2002). Bascula 4.0 Reference Manual. Centraal Bureau voor de Statistiek, Heerlen.
- [3] 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.
- [4] José Gouweleeuw en Kees van Berkel, Derde aanzet tot een steekproefontwerp voor WoON 2012, CBS-nota, 4 april 2011

## **Appendix A. Categorieën van de weegvariabelen**

Aantal personen in huishouden (5): 1, 2, 3, 4, 5 of meer personen;

COROP-gebied(40): Een fijnere indeling van de twaalf provincies in 40 gebieden. De naam COROP komt van Coördinatie Commissie Regionaal Onderzoeksprogramma. Dit was de naam van de commissie die in 1971 de indeling van Nederland in COROP-gebieden ontwierp.

Eigendom(3): hulpvariabele die is afgeleid zoals weergegeven in Figuur 1;

Geslacht(2): man, vrouw;

Herkomst(7): Nederlands, Marokkaans, Turks, Surinaams, Antilliaans en Arubaans, Overig niet-westers, overig westers;

Gebied(122): Regionale indeling in 122 gebieden. De gebieden met 600 gewenste respondenten door oversamplingdeelnemers zijn aparte categorieën; ook de overgebleven gemeenten G31 vormen aparte categorieën; het restant buiten de oversampling en G31 van elk COROP-gebied wordt als apart gebied beschouwd; voor details, zie Appendix B;

Inkomen(5): Huishoudensinkomen in 5 categorieën: minder dan 16.800, 16.800-23.900, 23.900-33.100, 33.100-46.100, meer dan 46.100 euro. Dit is het CBS besteedbaar huishoudinkomen, afgeleid uit belastinggegevens.

Landsdeel(4): noord (Groningen, Friesland en Drenthe), west (Utrecht, Noord-Holland en Zuid-Holland), oost (Overijssel, Flevoland en Gelderland), zuid (Noord-Brabant en Limburg);

Leeftijd(7): 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+ jaar;

Leeftijd(11): 18-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65+ jaar;

Leeftijd(15): 18-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+ jaar;

Positie in huishouden(5): kind, alleenstaand, met partner, eenouder, overig;

Woontype(3): Eengezinswoning, meergezinswoning, onbekend;

WOZ-klasse(22): WOZ-waarde in 22 categorieën: 0-25.000, 25.000-50.000, ..., 475.000-500.000, 500.000+ euro, onbekend.

## **Appendix B. Gebiedsindeling bij “gebied(122)”**

Bij gebied(122) worden de onderstaande gebieden onderscheiden. Voor elk COROP-gebied vormt het restant na weglaten van ondergenoemde gebieden ook een apart gebied. Een zestal COROP-gebieden valt volledig binnen de oversampling. De gebiedsindeling is ook terug te vinden in de contracten waarin de minimaal te behalen responsaantallen staan voor de oversamplingdeelnemers. Verder is per gemeente op het internet te vinden welke postcodes tot het ISV-gebied behoren.

Alkmaar

Almelo

Amersfoort

Amstelveen

Amsterdam

Arnhem wijken 3, 4, 6, 7, 10, 16, 17, 18, 19, 21 en 22

Arnhem wijken 1, 2, 5, 8, 9, 11, 12, 13, 14, 15, 20, 23 en 24

Gemeenten Lingewaard en Overbetuwe

Gemeenten Groesbeek, Millingen a/d Rijn, Ubbergen en Mook en Middelaar

Gemeenten Beuningen, Heumen en Wijchen

Bergen op Zoom Bergen op Zoom stadsdeel A

Bergen op Zoom Bergen op Zoom stadsdeel B

Steenbergen

Woensdrecht

Breda

Den Haag Laakkwartier

Den Haag Schildersbuurt, Stationsbuurt en Transvaalkwartier

Den Haag Vinex

Den Haag overig niet particuliere huur

Den Haag overig particuliere huur

Dordrecht

Renkum

Gemeenten Rheden en Rozendaal

Duiven

Gemeenten Rijnwaarden, Doesburg, Westervoort, Zevenaar en Montferland

Eindhoven

Emmen

Enschede

Gelderland Achterhoek: Gemeenten Aalten, Berkelland, Bronckhorst, Doetinchem, Oost Gelre, Oude IJsselstreek en Winterswijk

Elburg

Ermelo

Harderwijk

Hatter

Heerde

Nunspeet

Oldenbroek

Putten

Gelderland Rivierenland: Gemeenten Buren, Culemborg, Druten, Geldermalsen, Lingewaal, Maasdriel, Neder-Betuwe, Neerijnen, Tiel, West Maas en Waal en Zaltbommel

Gelderland Vallei: Gemeenten Barneveld, Ede, Nijkerk, Scherpenzeel en Wageningen

Gouda

Midden-Holland: Gemeenten Bergambacht, Bodegraven-Reeuwijk, Boskoop, Nederlek, Ouderkerk, Schoonhoven, Vlist, Waddinxveen en Zuidplas

Groningen

Delft

Leidschendam-Voorburg

Midden Delftland

Pijnacker-Nootdorp

Rijswijk

Wassenaar

Westland

Zoetermeer

Haarlem

Heerlen

Helmond

Hengelo

Kampen wijk 016600

Kampen wijk 016603 t/m 016608

Katwijk  
Leeuwarden  
Leiden  
Lelystad  
Maassluis  
Maastricht  
Nijmegen ISV-gebieden (wijken 0, 1, 20, 21, 23, 24, 14, 16, 32, 33, 35 en 36 plus 2.366  
aangeleverde postcodegebieden)  
Nijmegen niet ISV-gebieden (overig Nijmegen)  
Rotterdam  
Schiedam  
s-Hertogenbosch  
Sittard-Geleen  
Teylingen  
Tilburg  
Utrecht ISV-gebieden  
Utrecht niet-ISV-gebieden  
Venlo  
Vlaardingen gebied 1  
Vlaardingen gebied 2  
Vlaardingen gebied 3  
Vlaardingen gebied 4  
Vlaardingen gebied 5  
Vlaardingen gebied 6  
Vlaardingen gebied 7  
Zaanstad  
Zeeland zonder de gemeenten Schouwen Duiveland en Tholen  
Schouwen Duiveland  
Tholen  
Zwolle ISV-gebieden  
  
Zwolle niet-ISV-gebieden

## Appendix C. Afleidingen van HWEEGWON en GWEEGWON

```
COMPUTE HHFAC=$SYSMIS.
* HHFAC = aantal 18-plussers in huishoudkern.
COMPUTE HHFAC=1. /*Steekproefpersoon behoort per definitie tot huishoudkern.
IF ((hhkern=1 OR hhkern=2 OR hhkern=3 OR hhkern=4) AND (lftpa>=18 OR MISSING(lftpa) OR
any(lftpa,0,1) )) HHFAC=2.
IF (hhkern=7) HHFAC=MIN(aantalpp,10).
VARIABLE LABELS HHFAC 'Aantal 18-plussers in huishoudkern'.

IF (hvs=6) HWEEGWON=0. /*lid huishouden wordt niet meegeteld.
IF (hvs<>6) HWEEGWON=Weegpers/HHFAC. /*persoonsgewicht gedeeld door het aantal 18-plussers in de
huishoudkern.
VARIABLE LABELS HWEEGWON 'Huishoudensgewicht = WEEGPERS/HHFAC'.
EXECUTE.

COMPUTE GHHFAC=$SYSMIS.

* GHHFAC = aantal 18-plussers in het huishouden na verhuizing.
IF (samhhnv=1) GHHFAC=HHFAC. /*samenstelling van het huishouden na verhuizing is hetzelfde.
IF (samhhnv<>1 AND tposhh=1) GHHFAC=1. /*alleen.
IF (samhhnv<>1 AND tposhh=2) GHHFAC=2. /*hoofd/partner zonder kinderen.
IF (samhhnv<>1 AND tposhh=3) GHHFAC=2. /*hoofd/partner met kinderen.
IF (samhhnv<>1 AND tposhh=4) GHHFAC=1. /*hoofd eenoudergezin.
IF (samhhnv<>1 AND tposhh=5) GHHFAC=0. /*kind tweeoudergezin.
IF (samhhnv<>1 AND tposhh=6) GHHFAC=0. /*kind eenoudergezin.
IF (samhhnv<>1 AND tposhh=7) GHHFAC=0. /*overig lid.
IF (samhhnv<>1 AND tposhh=8) GHHFAC=2. /*niet-gezinshuishouden.
VARIABLE LABELS GHHFAC 'Aantal 18-plussers in het huishouden na gewenste verhuizing'.
EXECUTE.

COMPUTE gweegwon=$sysmis.
* GWEEGWON = huishoudensgewicht na verhuizing van verhuisgeneigde respondenten.
IF (verh=4 OR ghvs=4 OR GHHFAC=0) GWEEGWON=0. /*niet verhuisgeneigd, of na verhuizing lid van
huishouden, of na verhuizing geen 18+.
IF (verh~4 AND ghvs~4 AND GHHFAC>0) GWEEGWON=WEEGPERS/GHHFAC. /*persoonsgewicht gedeeld door
het aantal 18-plussers in het huishouden na verhuizing.
VARIABLE LABELS GWEEGWON 'Huishoudensgewicht na verhuizing'.
```