



Weging WOON 2015

Erwin Vondenhoff

Abstract: In deze nota wordt de toekenning van gewichten aan de respondenten van WOON 2015 besproken. Aanpassingen ten opzichte van 2012 komen aan bod en de statistieken van de correctie- en eindgewichten worden gepresenteerd.

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.

Er zijn twee doelpopulaties gedefinieerd, één op het niveau van personen en één op het niveau van huishoudens. Op persoonsniveau hebben uitkomsten betrekking op alle personen die op 1 januari 2015 minstens 18 jaar zijn, deel uitmaken van particuliere huishoudens en ingeschreven zijn in de Basisregistratie Personen (BRP). Voor huishoudens is de doelpopulatie gedefinieerd als alle particuliere huishoudens die op 1 januari 2015 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.

Voor het WOON is een personensteekproef getrokken. Deze steekproef kan echter ook opgevat worden als huishoudenssteekproef. Er wordt onderscheid gemaakt tussen respondenten die tot de huishoudkern behoren en respondenten die niet tot de huishoudkern behoren. Alleen de eerste groep krijgt naast een persoonsgewicht een huishoudensgewicht, omdat alleen zij vragen beantwoorden over het huishouden. Hier komen we later in dit document op terug. Meer detail is te vinden in [1].

De methodiek van het wegen is veelal gelijk aan 2012. Voor details met betrekking tot de methode verwijzen we naar de documentatie van 2012 [1]. Hier zullen we ons beperken tot de hoofdlijnen. Daarnaast bespreken we veranderingen ten opzichte van 2012. Voor een gedetailleerde beschrijving van de programmatuur voor de weging verwijzen we naar [4].

2. Gebiedsindelingen

Bij de weging spelen een tweetal gebiedsindelingen een rol. In deze paragraaf bespreken we de indeling van Nederland in uitkruisgebieden en insluitgewichtgebieden. Deze gebiedsindelingen

hangen samen met regionale oversampling.

Bepaalde gemeenten en combinaties van gemeenten hebben deelgenomen aan de oversampling. De steekproef is zo ontworpen dat in deze (combinaties van) gemeenten meer respondenten worden behaald dan in andere gebieden. Hierdoor zijn nauwkeurigere uitspraken te doen voor deze (combinaties van) gemeenten. Voor de oversamplingdeelnemers zijn in het weegmodel een aantal termen opgenomen om er voor te zorgen dat binnen de betreffende gebieden de gewichten optellen tot het juiste aantal. De gebieden zijn uitgekruist met geslacht, leeftijd, een huur/koop variabele en het woontype. We zullen in deze notitie de term uitkruisgebieden gebruiken voor deze gebieden die deel hebben genomen aan de oversampling waarbinnen voldoende responsen zijn behaald om de genoemde uitkruisingen te kunnen maken. De restanten per COROP-gebied worden bij de indeling naar uitkruisgebieden als aparte klassen beschouwd. De exacte indeling naar uitkruisgebieden is terug te vinden in Appendix B.

Naast uitkruisgebieden zijn er insluitgewichtgebieden. Insluitgewichtgebieden zijn gebieden waarbinnen alle huishoudens dezelfde trekkingskans hebben gekregen. De insluitgewichtgebieden vormen een verfijning van de uitkruisgebieden in de zin dat wanneer twee adressen in hetzelfde insluitgewichtgebied vallen, ze ook tot hetzelfde uitkruisgebied behoren. In 2015 is de indeling in insluitgewichtgebieden als volgt. Ga uit van de uitkruisgebieden en beschouw de eerder besproken restanten per COROP-gebied. Uit de restanten worden de adressen die tot een krachtwijk van de G18 behoren als één extra klasse opgevat. Vervolgens is er nog een klasse bestaande uit de niet-krachtwijken van de G18 binnen die restanten.

3. De strata en de insluitgewichten

In de vorige paragraaf zijn de insluitgewichtgebieden besproken. Deze indeling legt, samen met het kenmerk of iemand een partner heeft, de insluitgewichten vast. Partners worden namelijk met een halve kans getrokken om ervoor te zorgen dat alle huishoudens dezelfde trekkingskans hebben. Huishoudens kunnen namelijk een huishoudenskern met omvang één of omvang twee hebben.

De combinatie van insluitgewichtgebied en de partner-variabele leggen de stratum-indeling vast. Per insluitgewichtgebied zijn er dus twee strata, namelijk één voor de partners en één voor de niet-partners. Bij de steekproeftrekking heeft het begrip stratum een iets andere definitie.

De insluitgewichten worden berekend door het aantal elementen in de populatie te delen door het aantal respondenten. Voor details over stratum-indeling en insluitgewichten verwijzen we naar [3]. In 2012 is met betrekking tot de stratum-indeling exact dezelfde methode gehanteerd. De methode is terug te vinden in [1]. De oversamplingdeelnemers wijken in 2015 echter af.

4. Het weegmodel, de persoonsgewichten en de huishoudensgewichten

In deze paragraaf beschrijven we het weegmodel dat gehanteerd is om op basis van de insluitgewichten de uiteindelijke eindgewichten te bepalen. De methode van lineair wegen [2] wordt hierbij gehanteerd.

De eindgewichten zijn persoonsgewichten en tellen daarmee op tot de eerder gedefinieerde

populatie op persoonsniveau. Op basis van deze persoonsgewichten kunnen echter ook de huishoudensgewichten berekend worden. De huishoudensgewichten zijn gelijk aan de persoonsgewichten vermenigvuldigd met de zogenaamde kwantificatie-variabele. Op deze manier kan de steekproef als personensteekproef maar ook als huishoudenssteekproef worden opgevat.

De respondenten worden opgedeeld in twee groepen, namelijk de huishoudkernleden en de zogenaamde “leden van het huishouden”. Alleen huishoudkernleden krijgen vragen over het huishouden. De leden van het huishouden kunnen bijvoorbeeld thuiswonende kinderen zijn. Leden van het huishouden krijgen kwantificatievariabele nul en daarmee huishoudensgewicht nul. Voor huishoudkernleden is de kwantificatievariabele gelijk aan één wanneer de huishoudkern uit één persoon bestaat. Bestaat de huishoudkern uit twee personen dan is de kwantificatievariabele gelijk aan 0.5.

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

Stratum(204)
+ Geslacht(2) × Leeftijd(15)
+ Geslacht(2) × Herkomst(7)
+ Leeftijd(7) × Herkomst(7)
+ Landsdeel(4) × Herkomst(7)
+ Provincie(12) × Inkomen(5) (kwantitatief)
+ WOZ-klasse(22) (kwantitatief)
+ Gemeente(393)
+ Positie in huishouden(5) × Leeftijd(11)
+ Aantal personen in huishouden (5)
+ Uitkruisgebied(108) × Leeftijd(7)
+ Uitkruisgebied(108) × Geslacht(2)
+ Uitkruisgebied(108) × Eigendom(3) (kwantitatief)
+ Uitkruisgebied(108) × Woontype(2) (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 2015. Voor een aantal variabelen zijn ten tijde van de weging geen data beschikbaar. Voor die variabelen zijn data uit voorafgaande jaren gekoppeld aan de populatie van 2015. Inkomen is afkomstig uit het Inkomensregister van 2013. WOZ-waarde en woontype komen uit het BAG. De variabele eigendom is berekend uit variabelen die afkomstig zijn uit zowel het Inkomensregister als het BAG.

Een viertal termen is op huishoudensniveau gewogen. Dat wil zeggen dat niet de persoonsgewichten, maar de huishoudensgewichten tot bepaalde randtotalen moeten optellen. Het is mogelijk om op deze manier simultaan te wegen naar zowel persoons- als huishoudenskenmerken, gebruikmakend van de kwantificatie-variabele. Bij het berekenen van populatietotalen worden eerst de betreffende variabelen aan respondenten gekoppeld. Vervolgens dient een kwantificatie-variabele voor de populatie berekend te worden. Dit gebeurt als volgt. Op basis van positie in het huishouden is bekend welke personen tot de huishoudkern behoren, zie

Tabel 1. Vervolgens wordt op basis van het huishoudnummer, binnen elk huishouden geteld hoeveel personen in de huishoudkern zitten. Dit kan alleen 1 of twee zijn. De kwantificatievariabele is nul voor leden van het huishouden en voor huishoudkernleden is deze één gedeeld door de berekende omvang van de huishoudkern, zie ook [1].

Positie in huishouden (9 cat)	Pos. in hh. (5 cat)	hhkern
thuiswonend kind	kind	0
alleenstaande	alleenstaande	1
partner niet-gehuwd zonder kinderen	partner	1
partner gehuwd zonder kinderen	partner	1
partner niet-gehuwd met kind(eren)	partner	1
partner gehuwd met kind(eren)	partner	1
ouder in eenouderhuishouden	eenouder	1
referentiepersoon in overig huishouden	overig	1
overig lid in huishouden	overig	0

Tabel 1: Afleiding van huishoudkern en kwantificatievariabele voor de populatie aan de hand van positie in huishouden.

De volgende wijzigingen zijn aangebracht ten opzichte van de ophoogmethode in 2012:

- WOZ wordt kwantitatief opgehoogd op huishoudensniveau. In 2012 is WOZ opgehoogd op persoonsniveau.
- Er zijn geen respondenten op Vlieland. Vlieland is daarom, voor de weegterm “gemeente”, samen met Terschelling opgevat als één gemeente. Verder zijn alle gemeenten opgevat als aparte klassen.
- Bij de variabele Woontype is in 2012 onderscheid gemaakt tussen eengezinswoningen en meergezinswoningen. Verder vormden woningen waarvoor niet bekend was of het eengezins- of meergezinswoningen waren en aparte groep. In 2015 is deze onbekend-groep samengevoegd met de eengezinswoningen, zowel voor de respons als de populatie, omdat het een kleine groep is.

5. Ophogen van enquête naar register voor positie in huishouden

Net als in 2009 en 2012 is er in 2015 voor gekozen om leeftijd en positie in huishouden op te hogen van enquêtewaarde naar registerdata. Voor leeftijd is dit goed te beargumenteren. Bij positie in huishouden gaat het mis in de overig-categorie. Van personen die volgens enquête tot de categorie overig behoren, wordt in de enquête een relatief groot deel tot de huishoudkernleden gerekend, in verhouding tot de populatie. Dit leidt tot veel negatieve gewichten die na een groot aantal iteraties pas te begrenzen zijn. Negatieve gewichten komen vooral voor bij de oververtegenwoordigde groepen. De negatieve gewichten komen vooral voor bij personen in de overig-categorie van “positie in huishouden” die zichzelf tot de huishoudkern rekenen.

Net als bij positie in huishouden en leeftijd wordt ook de variabele die aangeeft of iemand tot de huishoudkern behoort en de kwantificatievariabele niet uit het register gehaald maar bepaald aan de hand van de antwoorden van de respondent. Zoals gezegd, in het register worden deze afgeleid aan de hand van positie in het huishouden, zie Tabel 1. In het responsbestand geeft de variabele “respkern” aan of de respondent zichzelf tot de huishoudkern rekent. Het probleem bij de overig categorie is dat veel respondenten vinden dat ze tot de huishoudkern behoren terwijl de groep in

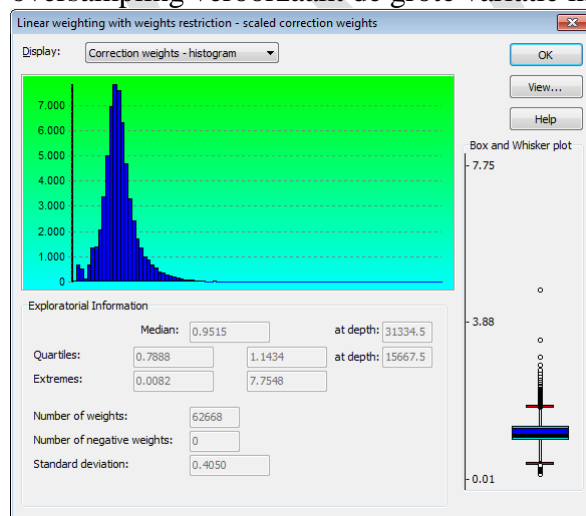
de populatie relatief klein is. De omvang van het huishouden wordt voor de populatie afgeleid zoals eerder uitgelegd. In het responsbestand wordt echter uitgegaan van de huishoudensfactor (hhfac). Deze variabelen in het responsbestand komen niet altijd overeen met de variabele volgens het register.

6. Resultaat van de weging

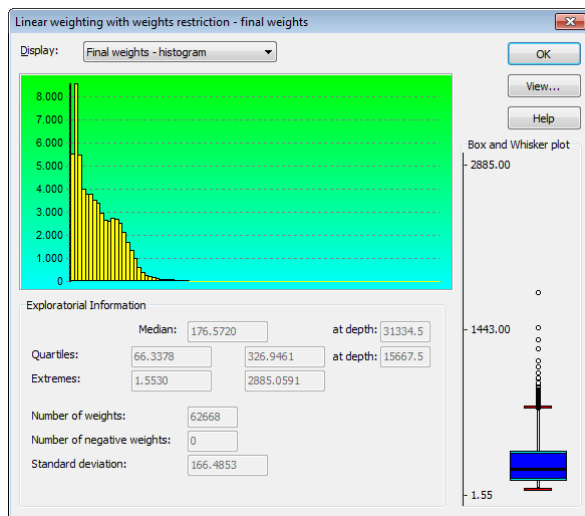
Lineair wegen met Bascula [2] leidt tot 541 negatieve gewichten. Deze zijn vooral voor respondenten in de overig categorie van de hulpvariabele “positie in huishouden” die zichzelf tot de huishoudkern rekenen. Deze gewichten zijn te begrenzen met tien iteraties van het Huang Fuller algoritme [3]. De 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 $V_w^2 = 0,621$. 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 62,1%. Vaak correleren hulpvariabelen met de doelvariabele. In dat geval is de factor waarmee de variantie toeneemt lager. Wanneer de correlatiecoëfficiënt tussen doel- en hulpvariabelen 0,619 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 trekkingkans 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

7. Conclusies

De weging van WOON is succesvol uitgevoerd. Negatieve gewichten zijn met behulp van het algoritme van Huang Fuller begrensd.

In tegenstelling tot 2012 is WOZ-klasse kwantitatief op huishoudensniveau gewogen en de onbekend-klasse bij de variabele woontype is samengevoegd met eengezinswoningen.

Net als in 2012 zijn er problemen ontstaan doordat positie in huishouden is opgehoogd van enquête- naar registervariabele. Hierdoor zijn er in de respons in de overig-groep relatief veel huishoudkernleden ten opzichte van de populatie. Het is aan te bevelen om voor uitvoeringen van WOON in de toekomst op zoek te gaan naar een methode waarbij dit probleem omzeild kan worden.

8. Referenties

- [1] Erwin Vondenhoff, Weging WOON 2012. Centraal Bureau voor de Statistiek, Heerlen.
- [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] Erwin Vondenhoff Bestandsverwerking Weging WOON 2015. Centraal Bureau voor de Statistiek, Heerlen.

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): Koopwoning, sociale huurwoning, particuliere huurwoning;

Geslacht(2): man, vrouw;

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

Uitkruisgebied: De gebieden met 600 gewenste respondenten door oversamplingdeelnemers zijn aparte categorieën; het restant buiten de oversampling van elk COROP-gebied wordt ook als apart gebied beschouwd; voor details, zie Appendix B;

Inkomen(5): Huishoudensinkomen in 5 categorieën: minder dan 21.030, 21.031-30.587, 30.588-41.283, 41.284-55.081, meer dan 55.081 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: Eengezinswoning, meergezinswoning (onbekend-klasse is samengevoegd met eengezinswoning);

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 de variabele uitkruisgebied worden de onderstaande gebieden onderscheiden. Voor elk COROP-gebied vormt het restant na weglaten van ondergenoemde gebieden ook een apart gebied.

- Den Haag: personen in koopwoning, buiten de Vogelaarwijken
- Den Haag: personen in sociale huurwoning (huurwoning van een toegelaten instelling), buiten de Vogelaarwijken
- Den Haag: personen in een particuliere huurwoning, buiten de Vogelaarwijken
- Den Haag: personen in koopwoning, in de Vogelaarwijken
- Den Haag: personen in sociale huurwoning (huurwoning van een toegelaten instelling), in de Vogelaarwijken
- Den Haag: personen in een particuliere huurwoning, in de Vogelaarwijken
- Lelystad: Zuiderzeewijk
- Lelystad: Atolwijk
- Lelystad: Boswijk
- Lelystad: Waterwijk/Landerijen
- Lelystad: Bolder
- Lelystad: Kustwijk
- Lelystad: Havendiep
- Lelystad: Lelystad-Haven
- Lelystad: Warande/Stadshart/Buitengebied
- Nijmegen: ISV gebieden buiten de Vogelaarwijken
- Nijmegen: ISV-gebieden in de Vogelaarwijken
- Nijmegen: niet-ISV gebieden (en deze blijken automatisch allemaal buiten de Vogelaarwijken te liggen)
- Enschede: binnen Vogelaarwijken
- Enschede: buiten Vogelaarwijken
- Arnhem: binnen Vogelaarwijken
- Arnhem: buiten Vogelaarwijken
- Utrecht: binnen Vogelaarwijken
- Utrecht: buiten Vogelaarwijken
- Maastricht: binnen Vogelaarwijken
- Maastricht: buiten Vogelaarwijken
- Zeeuws-Vlaanderen (de gemeenten Hulst (0677), Sluis (1714) en Terneuzen(0715))
- Walcheren (de gemeenten Veere (0717), Middelburg (0687) en Vlissingen (0718))
- Bevelanden (de gemeenten Borsele (0654), Goes (0664), Kapelle (0678), Noord-Beveland (1695) en Reimerswaal (0703))
- Schouwen-Duiveland (1676)
- Tholen (0716)
- Arnhem e.o.: Renkum, Rheden, Rozendaal, Overbetuwe, Lingewaard
- Nijmegen e.o.: Beuningen, Groesbeek, Heumen, Millingen aan de Rijn, Mook en Middelaar, Ubbergen, Wijchen
- Liemers: Doesburg, Duiven, Montferland, Rijnwaarden, Westervoort, Zevenaar
- Achterhoek: Aalten, Berkelland, Bronckhorst, Doetinchem, Oost Gelre, Oude IJsselstreek, Winterswijk
- Noord-Veluwe: Elburg, Ermelo, Hattum, Heerde, Nunspeet, Oldebroek, Putten
- Rivierenland: Buren, Culemborg, Druten, Geldermalsen, Lingewaard, Maasdriel, Neder-Betuwe, Neerijnen, Tiel, West Maas en Waal, Zaltbommel

- Harderwijk
- Delft
- Pijnacker-Nootdorp
- Midden-Delfland
- Leidschendam-Voorburg
- Rijswijk
- Wassenaar
- Westland
- Zoetermeer
- Breda
- Gouda
- Waddinxveen
- Zuidplas
- Bergen op Zoom
- Steenbergen
- Moerdijk
- Halderberge
- Woensdrecht
- Kampen
- Goeree-Overflakkee
- Maassluis
- Vlaardingen
- Teylingen
- Katwijk
- Food Valley: Barneveld, Ede, Nijkerk, Scherpenzeel, Wageningen, Rhenen, Renswoude, Veenendaan
- Stedendriehoek: Apeldoorn, Brummen, Deventer, Epe, Lochem, Voorst, Zutphen
- Rotterdam: ISV-gebieden
- Rotterdam: niet-ISV-gebieden
- Amsterdam: ISV-gebieden
- Amsterdam: niet-ISV-gebieden
- Almere