

Propensity Weighting: een nieuwe methode voor het wegen van onderzoeksgegevens

Mariola de Gouw
Edward Groenland

Blauw Research Rotterdam

De laatste jaren is er sprake van een sterk stijgende non-respons in het marktonderzoek. Waar deze systematisch is, in plaats van at random, kunnen vertekeningen optreden in de gerapporteerde uitkomsten. Als gevolg hiervan kan de praktische waarde van het onderzoek sterk afnemen. Het is dan ook van toenemend belang om methodieken beschikbaar te hebben om te corrigeren voor vertekeningen in de onderzoeksgegevens als gevolg van vermoede, systematische non-respons. Een veelgebruikte strategie om dit probleem het hoofd te bieden is het toepassen van de compensatiemethode Poststratificatie. De ervaring leert dat deze methodiek niet optimaal voldoet.

Propensity Weighting is een alternatieve compensatiemethode, die verschillende basisprincipes van de bestaande compensatiemethoden samenvoegt. Deze methode is ontwikkeld door Harris Interactive, een Amerikaans onderzoeksbureau dat zich een leidende positie heeft verworven op het gebied van online marktonderzoek. Het gebruik van Propensity Weighting vergt een belangrijke validatie-stap alvorens deze methodiek in de praktijk kan worden toegepast. Blauw Research, affiliate partner van Harris Interactive, heeft Propensity Weighting geschikt gemaakt voor gebruik voor de Nederlandse populatie, waar ze met name wordt toegepast bij het herwegen van onderzoeksgegevens verkregen door middel van online onderzoek. In dit kader werd een uitgebreide, vergelijkende studie uitgevoerd, teneinde de performance van Propensity Weighting te bepalen ten opzichte van de klassieke compensatiemethode Poststratificatie.

Dit artikel bespreekt de representativiteitsproblematiek bij het uitvoeren van telefonisch onderzoek en online onderzoek. Vervolgens vindt een beschrijving plaats van de uitgangspunten en de werkwijze van de wegingsmethodiek Propensity Weighting, mede in vergelijking tot de klassieke compensatiemethode Poststratificatie.

Tenslotte worden de resultaten van een vergelijkend onderzoek met betrekking tot het toepassen van Poststratificatie en Propensity Weighting gerapporteerd en besproken.

De uitkomsten van deze vergelijkende studie ondersteunen de opvatting dat Propensity Weighting een aantal unieke voordelen biedt bij het wegen van onderzoeksgegevens, en dat deze nieuwe methode, over het geheel genomen, tot betere resultaten kan leiden dan klassieke herwegingmethoden.

Representativiteit : een vergelijking van telefonisch onderzoek en online onderzoek

Bij het uitvoeren van onderzoek is het optreden van selectieve, of systematische non-respons een van de oorzaken die de kwaliteit van de gegevens kunnen aantasten. Dit is het geval wanneer bepaalde categorieën respondenten (jongeren, hoog opgeleiden, tweeverdieners, heavy of light users, klagende consumenten, tevreden productgebruikers) meer, of juist minder in de steekproef voorkomen dan in de achterliggende populatie. Het niet corrigeren voor dergelijke vormen van vertekening in de steekproef leidt uiteindelijk tot onjuiste projecties naar de feitelijke populatie.

Er zijn veel mogelijke factoren aan te wijzen die bijdragen aan het ontstaan van selectieve non-respons, en aldus afbreuk doen aan de representativiteit van het onderzoek. Vele ervan hebben te maken met de kwaliteit van het veldwerk, en de kwaliteit van het steekproefkader. Onderzoekers, maar ook opdrachtgevers besteden met recht aandacht aan deze problematiek.

Een nieuwe dimensie in het representativiteitsvraagstuk heeft te maken met de introductie van online onderzoek in Nederland en de rest van de wereld. Online onderzoek maakt momenteel een krachtige groei door, en zal, naar verwachting, op termijn het telefonisch onderzoek in belangrijke mate verdringen. In de huidige overgangsfase beschikt al ongeveer de helft van de huishoudens over een

internetaansluiting, ofwel in de woning, ofwel via het werk. Onderzoek toont aan dat de internetpopulatie steeds meer gaat lijken op de algemene populatie, hoewel er nog steeds verschillen zijn met betrekking tot demografie, attitude en gedrag.

De vraag kan nu worden gesteld of conventioneel telefonisch onderzoek tot representatiever uitkomsten leidt dan online onderzoek. Bezien we de respons rate bij telefonisch onderzoek dan blijkt dat deze op ongeveer 30 % ligt. Dit heeft te maken met het gebruik van mobiele telefoons, die soms de plaats innemen van de reguliere telefoonaansluiting, en het gebruik van antwoordapparaten. Een belangrijke oorzaak is ook de gestaag afnemende bereidheid om interviewers telefonisch te woord te staan, een fenomeen dat door sommigen in verband wordt gebracht met de activiteiten van de telemarketingbranche. In dit licht bezien kunnen telefonische steekproeven bij wijze van spreken beschouwd worden als convenience samples.

Online onderzoek biedt een aantal voordelen ten opzichte van telefonisch onderzoek. Zo zijn respondenten gemakkelijker te bereiken, met name low incidence-respondenten. Ook is er een voordeel in termen van efficiency. Ten opzichte van telefonisch onderzoek zijn respondenten in staat tot 35 % meer vragen te beantwoorden. Ze doen dit op de momenten die hen uitkomen. In het algemeen is binnen 48 uur een substantieel deel van de vragenlijsten beantwoord. De vragenlijsten kunnen beelden en foto's bevatten, bij voorbeeld verpakkingen, print advertising, die online kunnen worden beoordeeld. Tenslotte is er sprake van beter respons-management: het selecteren van non-respondenten en timen van reminders verloopt gemakkelijker. In het algemeen geldt dat bij online onderzoek de kosten per eenheid respons lager zijn dan die voor telefonisch interviewen.

Online onderzoek op basis van een database met respondenten die zich eerder principieel bereid hebben verklaard vragenlijsten in te vullen leidt tot een response-percentages dat tot 60-70 % kan oplopen. Al eerder werd aangegeven dat de internetpopulatie echter in een aantal opzichten afwijkt van de algemene populatie.

Uit het bovengaande moge blijken dat zowel het telefonisch onderzoek als het huidige online onderzoek problemen opleveren wat betreft hun vermogen projecties te maken naar de algemene populatie. Gegeven de eerder beschreven, evidente voordelen mag worden verwacht dat online onderzoek op termijn een dominante gegevensverzamelingsmethode zal worden. In deze fase doet zich de behoefte gevoelen aan een wegingsmethode die gegevens afkomstig uit beide methoden vergelijkbaar maakt. Propensity Weigting voorziet in deze behoefte.

Omgaan met non-respons

Non-respons kan optreden op het niveau van individuele vragen, of items. Van item non-respons is sprake als de antwoorden op bepaalde items niet in het analysebestand aanwezig zijn. Hierbij is de persoon uit de steekproef bijvoorbeeld wel bereid om mee te werken, maar heeft, om vaak onduidelijke redenen, niet alle vragen beantwoord.

Er kan ook sprake zijn van zog. Eenheid-non-respons. Van eenheid non-respons is sprake als de gegevens van een hele statistische eenheid (respondent) ontbreken. Een voorbeeld hierbij is een persoon die wel deel uitmaakt van de getrokken steekproef, maar die niet bereid is om mee te werken.

Als er sprake is van non-respons dan kan de onderzoeker in principe op vier manieren reageren: Op de eerste plaats kan zij er voor kiezen om deze niet te corrigeren. Zoals gezegd kan dit leiden tot misleidende onderzoeksresultaten en –aanbevelingen. Een tweede mogelijkheid is om te proberen om de omvang van de non-respons te reduceren. Hierbij kan men denken aan het terugbrengen van het aantal weigeraars en het terugbrengen van het aantal niet-gecontacteerde personen. Dit vergt een extra inspanning met betrekking tot het veldwerk die, binnen de grenzen van tijd en budget niet altijd mogelijk is. Een derde mogelijkheid bestaat uit het beoordelen van de aard en de grootte van de vertekening. Hiervoor zijn in principe aanvullende gegevens nodig. De laatste mogelijkheid wordt gevormd door de wegings- en imputatiemethoden. In dit artikel zal ingegaan worden op de wegingsmethoden.

Wegingsmethoden

Little en Rubin (1987) geven een aantal methoden aan om systematische non-respons te compenseren, opdat de effecten van non-respons worden geminimaliseerd. Wij bespreken hier de wegingsprocedures, en wel in termen van mogelijkheden en beperkingen in de setting van de praktijk van het marktonderzoek.

Wegingprocedures worden gebruikt voor de correctie van eenheid non-respons. Ze leiden tot het kopiëren van cases in een stratum dat onder-gerepresenteerd is. Het eindresultaat is dat alle soorten cases in de juiste verhouding voorkomen in de steekproef. Hiertoe worden gewichten uitgerekend, en toegepast op de steekproef. Volgens De Leeuw en Hox (1998) hangt deze wegingstrategie echter in belangrijke mate af van de beschikbaarheid van aanvullende informatie. Deze kan zijn gebaseerd op de achterliggende populatie, op de steekproef, of op het model. Voorbeelden van wegingprocedures zijn Poststratificatie, multiplicatief wegen en additief wegen. Wegingprocedures zijn vaak eenvoudig uit te voeren, zeker als gebruik gemaakt wordt van Poststratificatie. Nadelig is echter dat wegenen kunnen leiden tot relatief hoge gewichten, waardoor onstabiele schattingen kunnen worden geproduceerd. Dergelijke gewichten mogen dan ook niet worden toegepast. In de praktijk is dit een belemmering voor de toepassing van Poststratificatie.

Als een snelle procedure uitgevoerd dient te worden, die praktisch is in het gebruik, kan gedacht worden aan Poststratificatie. Echter, de procedures moeten met de nodige voorzichtigheid worden uitgevoerd. Controles achteraf zijn altijd noodzakelijk.

Propensity Weighting

Blauw Research heeft onlangs een nieuwe compensatiemethode gevalideerd die niet in één van bovenstaande methoden is in te delen. Deze methode, Propensity Weighting, kan met name zinvol worden toegepast bij het herwegen van gegevens verkregen bij een online steekproef die ondervraagd is met behulp van vragenlijsten aangeboden via het internet. Propensity Weighting is een techniek waarbij logistische regressie wordt gebruikt. Het principe is dat herwegingvariabelen worden gebruikt als predictorvariabelen in een regressiemodel. De criteriumvariabele is een dichotome, afhankelijke variabele. Deze geeft aan tot welke van de twee groepen een respondent behoort, bij voorbeeld de groep van online ondervraagden en telefonisch ondervraagden. De logistische regressie-analyse levert een schatting op van de kans dat een waarneming tot een van deze beide groepen behoort. In de context van herweging kunnen deze kansen worden uitgedrukt in zogenaamde propensity scores. Herweging van het doelbestand vindt plaats door de propensity-score-verdeling van het doelbestand gelijk te maken aan de propensity-score-verdeling van het referentiebestand. Hierbij dient de propensity-score-verdeling van het referentiebestand te worden opgesplitst in 5 gelijke groepen, ofwel quintielen. De randwaarden die bij deze quintielen horen, worden overgebracht op het doelbestand, waarna de percentages van de ontstane groepen in het doelbestand bepaald kunnen worden. De laatste stap is het maken van een weegfactor voor elk quintiel, welke het percentage van het referentiebestand, gedeeld door het percentage van het doelbestand is. Propensity Weighting is geschikt als compensatiemethode voor eenheid non-respons.

Een vergelijking tussen Propensity Weighting en de andere compensatiemethoden leert dat Propensity Weighting is gebaseerd op de basistechnieken van multi-pele imputatie, wegingprocedures en modelgebaseerde procedures. Het voordeel van deze methode is, dat deze zeer praktisch is, en relatief snel kan worden uitgevoerd, dit in tegenstelling tot bovengenoemde andere herwegingmethoden.

Poststratificatie versus Propensity Weighting

In een speciaal voor dit doel opgezet, vergelijkend onderzoek, heeft Blauw Research een performance-vergelijking uitgevoerd tussen de klassieke wegingmethode Poststratificatie en Propensity Weighting. Bij Poststratificatie wordt de steekproef, na het verzamelen van de data,

gestratificeerd in strata met een verschillende responskans aan de hand van de herwegingvariabelen. Vervolgens wordt bepaald welke (aantallen) waarnemingen in welke strata zijn te vinden. Weegfactoren worden nu bepaald door vergelijking van de aantallen waarnemingen in de overeenkomstige strata uit de beide bestanden. Voor deze methode is informatie over alle strata nodig. Het voordeel van Poststratificatie is dat het een zeer eenvoudig uit te voeren methode is. Poststratificatie kan een bijdrage leveren aan het reduceren van de vertekening ten gevolge van non-respons. Een nadeel van Poststratificatie is dat informatie over alle strata nodig is. Ook moeten er voldoende waarnemingen zijn in alle strata. Toepassing van deze methode geeft dan ook als zodanig geen zekerheid dat vertekeningen daadwerkelijk zijn gereduceerd. Een ander nadeel is dat grote gewichten niet mogen worden toegepast.

Theoretische en praktische voordelen van Propensity Weighting ten opzichte van Poststratificatie omvatten onder meer het feit dat Propensity Weighting uitgevoerd kan worden voor categoriale en continue variabelen, waar Poststratificatie slechts uitgaat van categoriale variabelen. Bij toepassing van Propensity Weighting is het hiermee mogelijk om herwegingvariabelen te gebruiken in de vorm van intervalgeschaalde variabelen. De methode Propensity Weighting kan daarenboven gebruik maken van een veel groter aantal herwegingvariabelen in de analyse dan Poststratificatie.

De validatiestap van Propensity Weighting die Blauw Research heeft uitgevoerd voor de Nederlandse markt bestaat uit het ontwikkelen en testen van effectieve herwegingsvariabelen. Dit zijn variabelen die een hoog voorspellend vermogen hebben als predictorvariabelen in de logistische regressievergelijking. Blauw Research heeft, in een systematische, wetenschappelijke opzet, een zeer groot aantal variabelen getest op hun vermogen tot adequaat herwegen. De variabelen met bewezen herwegingspotentie worden momenteel in de Propensity Weighting-procedure gebruikt, ten behoeve van onderzoek voor opdrachtgevers. De set van herwegingsvariabelen wordt voortdurend verfijnd en nader geoptimaliseerd.

Een nadeel bij het gebruik van Poststratificatie is dat er vaak te veel strata ontstaan, waardoor de inhoud van de strata te klein wordt. Bij Propensity Weighting is standaard sprake van 5 quintielen. Nadelen van Propensity Weighting zijn dat ook hier relatief grote gewichten niet mogen worden toegepast, en dat er sprake kan zijn van lege quintielen. Echter, deze nadelen zijn, zoals gezegd, ook terug te vinden bij Poststratificatie.

Politieke voorkeur: een vergelijking van Poststratificatie met Propensity Weighting

In een vergelijkende studie heeft Blauw Research gegevens ten aanzien van politieke voorkeur verzameld door middel van telefonische enquëtering en door middel van online onderzoek. Vervolgens werden de wegingsmethoden Poststratificatie en Propensity Weighting toegepast met als doel de online gegevens te wegen naar het gegevensbestand dat via telefonische enquëtering werd verkregen (het referentiebestand).

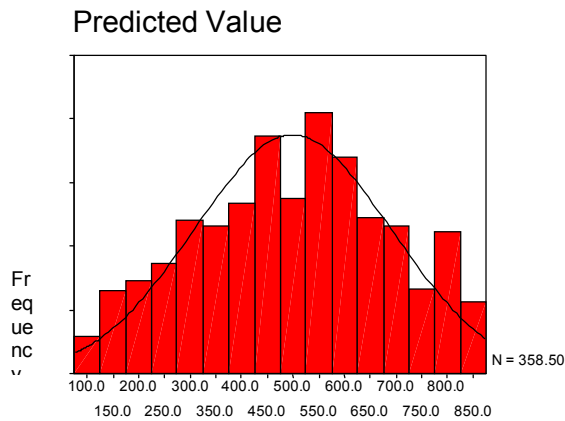
We laten eerst de algemene effecten zien van Propensity Weighting, en wel in de vorm van de onder weergegeven grafieken.

De predicted value in deze grafieken geeft voor elke respondent de kans aan (tussen 0 en 1.0) dat deze respondent behoort tot degenen die telefonisch zijn ondervraagd of online zijn ondervraagd. En dit op basis van de logistische regressie-analyse, met als predictoren de herwegingsvariabelen. Deze waarden zijn gebaseerd op de geoptimaliseerde set herwegingsvariabelen, en gelden voor alle doelvariabelen.

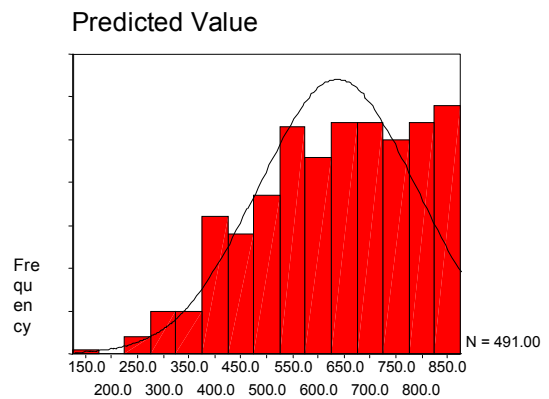
Wat het gevolg is van een succesvolle toepassing van Propensity Weighting is te zien in onderstaande grafieken, waarbij gelet dient te worden op de vorm van de grafieken. In het ideale geval zijn de verdelingen van het telefonisch bestand en het online bestand identiek. De eerste grafiek laat de frequentieverdeling zien van het telefonisch bestand. De tweede grafiek toont deze verdeling van het ongewogen online bestand. Wanneer de vorm van deze twee grafieken wordt vergeleken, zijn de verschillen tussen beiden duidelijk zichtbaar. De derde grafiek geeft de verdeling weer van het met

Propensity Weighting gewogen bestand. Door de herweging vertoont deze grafiek een veel grotere overeenkomst met de grafiek van het telefonische bestand dan het ongewogen online bestand.

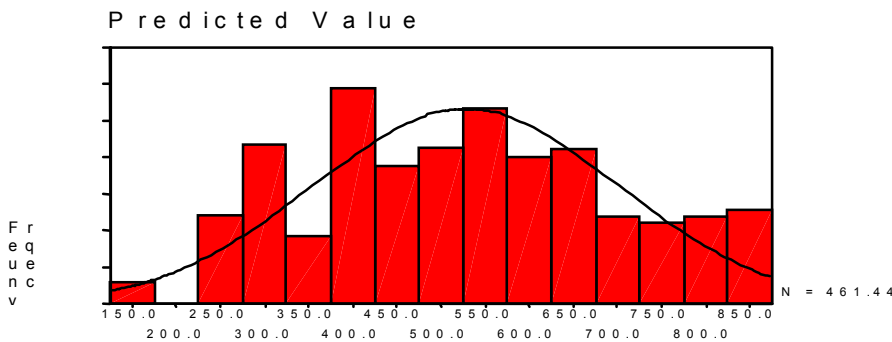
Grafiek 1 telefonisch bestand



Grafiek 2 online bestand ongewogen



Grafiek 3 online bestand PW-gewogen



Vervolgens hebben we een directe performance-vergelijking uitgevoerd tussen Propensity Weighting en Poststratificatie met betrekking tot politieke voorkeur. De relatieve performance van Propensity Weighting ten opzichte van Poststratificatie blijkt uit een analyse van de gemiddelde absolute verschillen in percentages met betrekking tot politieke voorkeur tussen het referentiebestand en het online bestand.

Tabel 1: Effecten van weegstrategieën voor politieke voorkeur

Referentiebestand	online ongewogen	online pw-gewogen	online ps-gewogen
0.0*	3.375	2.550	3.625

*uitgedrukt in percentagepunten

Uit de tabel blijkt dat Propensity Weighting leidt tot een significant grotere reductie van de bias dan Poststratificatie, en dat Propensity Weighting een resultaat kan opleveren dat minder vertekend is, vergeleken met het referentiebestand, dan een ongewogen versie. Ook bevestigde dit onderzoek de bevindingen uit de wetenschappelijke literatuur met betrekking tot het feit dat Poststratificatie in veel gevallen een meer vertekend bestand oplevert, in plaats van een minder vertekend bestand. In deze gevallen is het verschil tussen het referentie- en het doelbestand groter geworden, in plaats van kleiner, en had achteraf derhalve beter niet herwogen kunnen worden!

Conclusies

Op grond van de resultaten van het vergelijkend onderzoek concluderen wij dat, wanneer een compensatiemethode toegepast dient te worden, Propensity Weighting in het algemeen de voorkeur verdient boven Poststratificatie. Dit geldt met name wanneer online verkregen gegevens herwogen dienen te worden naar telefonisch verkregen gegevens, of naar een algemeen referentiebestand, zoals Minicensus. Blauw Research is onlangs overgegaan tot de ingebruikname van de methode van Propensity Weighting voor deze doeleinden.

Referenties

Leeuw, E.D. de, & Hox, J.J. (1998). Course given at the Centralarchiv fur empirische Socialforschung in Cologne, *On issues in handling missing data*. 4 augustus 2000, gevonden op het WWW: <http://www.fss.uu.nl/ms/jh/papers/papers.htm>

Little, R.J.A. & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: Wiley.