

Selectieve non-respons is lang niet altijd desastreus

Henk Elffers

Nederlands Studiecentrum Criminaliteit en Rechtshandhaving *

NSCR

Samenvatting

In dit artikel wordt getoond hoe men met eenvoudige selectiemodellen en assumpties de vertekening op grond van selectieve respons van een te schatten populatiefractie kan kwantificeren. Er wordt betoogd dat de impact van selectieve respons veelal wel meevalt.

1. Het probleem van selectieve respons

Wanneer we willen schatten hoeveel procent van een populatie een bepaald kenmerk bezit, dan kunnen we de machinerie van de mathematische statistiek alleen inzetten als we beschikken over een aselechte steekproef. In de praktijk zijn steekproeven echter zelden echt getrokken door een aselekt kansmechanisme, en er blijkt vaak dat ze dan op menigerlei kenmerk niet overeenkomen met bekende populatiegegevens. Dan zitten we met de handen in het haar. We moeten verwerpen dat de steekproef aselekt tot stand is gekomen, en we kunnen dan vermoeden dat ook het kenmerk van belang niet correct geschat wordt op grond van die steekproef. Veelal wordt dan tot allerlei vormen van herweging of post-stratificatie overgegaan.

Het probleem is het meest pregnant als we moeten aannemen dat de kans om in de steekproef terecht te komen samenhangt met het kenmerk van belang. Dan is herweging niet mogelijk: dat zou immers kennis veronderstellen van het voorkomen van het kenmerk van belang, dat we nou juist met het onderzoek in kwestie willen bepalen.

Dit probleem, selectieve (non-) respons, kan in theorie heel ernstig zijn. Als de responsbereidheid onder mensen met een bepaald kenmerk dramatisch verschilt van de responsbereidheid onder hen die dat kenmerk missen, dan is de steekproeffractie van hen die het kenmerk hebben uiteraard een erg slechte, vertekende, schatter voor de populatiefractie. Echter, dat iets in theorie heel bezwaarlijk kan zijn, betekent natuurlijk ook meevallen. Als de responsbereidheid niet zo erg verschilt, dan zal de vertekening ook niet zo erg zijn. Het loont daarom vaak de moeite om een indruk te trachten te verkrijgen van de mate van selectiviteit van de respons, om op die manier te achterhalen hoe zeer selectiviteit de resultaten in een concreet geval vertekent.

* NSCR, Postbus 792, 2300 AT Leiden
telefoon: 071-5278509, fax: 071-5278537, e-mail: elffers@nscr.nl

In deze notitie zal ik laten zien, voor dit geval van het schatten van een fractie van voorkomen van een kenmerk in een populatie, hoe men met op grond van een buitengewoon simpel selectiviteitsmodelletje selectieve respons kan kwantificeren, en vervolgens laten zien hoe men in de praktijk quick-and-dirty kan nagaan of in een concreet onderzoek de selectiviteit van de non-respons een serieus probleem is. Het zal blijken dat dat in de praktijk nogal eens erg meevalt. Tenslotte zal ik laten zien hoe men dit soort onzekerheden kan combineren met steekproefvariatiegegevens.

2. Een voorbeeld

Stel dat we onderzoek willen onder middelgrote bouwvakbedrijven (aannemers etc., met minstens 50 man personeel) naar de fractie bedrijven die bereid is –onder omstandigheden– hun klanten *zwart* te laten betalen, dus zonder BTW in rekening te brengen en af te dragen. Omdat zulke bedrijven het vooral van de zakelijke markt moeten hebben, komt de *'ik wil geen gedonder, en een goede naam, dat is op den duur lonender dan gesjoemel'*-mentaliteit daar vrij regelmatig voor, maar er zijn uiteraard ook ondernemers die wel *'zonder rekening'* willen leveren. Een steekproef van middelgrote aannemers wordt aangeschreven voor een schriftelijke enquête *'met het oog op het verkrijgen van inzicht in de problematiek rond de administratieve lasten -fiscaal, sociale lasten, ARBO, etc.- die het Midden- en KleinBedrijf tegenkomt'*. Te vrezen valt dat de responsbereidheid niet heel hoog zal zijn. Bovendien valt te verwachten dat zij die wel bereid zijn *zwart* betaald te krijgen, minder bereid zijn aan de enquête mee te doen, dan zij die dat niet zijn, dus dat er sprake zal zijn van selectieve respons.

Het ligt voor de hand de mate van selectieve respons uit te drukken in een grootheid, de *selectiviteitsratio* r , die gedefinieerd is als

$$r = \frac{\text{responspercentage onder zwartwerkers}}{\text{responspercentage onder witwerkers}}$$

Die ratio is uiteraard onbekend, omdat teller zowel als noemer onbekend zijn, maar toch kan men, zeker als men enige ervaring heeft in het te onderzoeken veld, wel een idee hebben over hoe groot r mogelijkster zal zijn. Als de hierboven uitgesproken vrees juist is, zal in ieder geval gelden dat $0 < r < 1$.¹ Ik vraag de lezer nu voor zich zelf een schatting te maken van wat hem of haar een aannemelijke waarde voor r lijkt, opdat hij of zij die verderop met mijn assumpties kan vergelijken.

Laten we ervan uitgaan dat het doel van het onderzoek is het percentage zwartwerkers te schatten, en vooral te bekijken of dat al of niet boven de $\frac{1}{2}$ ligt.

¹ Uiteraard kunnen we ook overmatige respons, $r > 1$, modelleren, met dezelfde formules. Ik laat dat hier buiten beschouwing. Door het kenmerk door zijn tegengestelde te vervangen zijn beide gevallen immers direct in elkaar over te voeren.

3. Een modelletje voor selectiviteit

Bekijk een populatie $Z \cup W$ (zwartwerkers en witwerkers in het voorbeeld), waaruit een steekproef s is getrokken, waarvoor geldt dat $S \cap W$ een aselechte steekproef uit W is, en $S \cap Z$ een aselechte steekproef uit Z .

$$\begin{aligned}
 \text{Zij} \quad q &= P(Z|S) = \text{de fractie } Z\text{'s in de steekproef (geobserveerde fractie)} \\
 \text{en} \quad c &= P(Z) = \text{de fractie } Z\text{'s in populatie (doelfractie)} \\
 \text{dan geldt} \quad q &= P(Z|S) = P(Z \cap S) / P(S) \\
 &= \{ P(S|Z) * P(Z) \} / \{ P(S \cap Z) + P(S \cap W) \} \\
 &= P(S|Z) * c / \{ P(S|Z) * P(Z) + P(S|W) * P(W) \} \\
 &= P(S|Z) * c / \{ P(S|Z) * c + P(S|W) * (1-c) \}
 \end{aligned}$$

dus

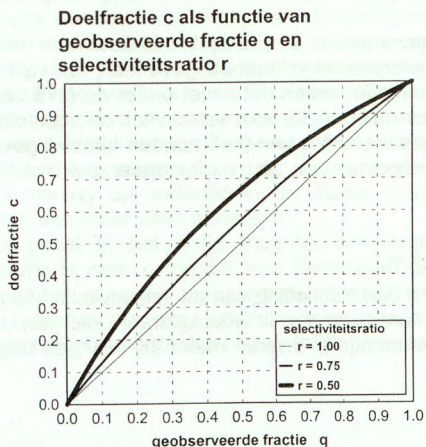
$$c = 1 / \{ P(S|Z) / P(S|W) * (1-q)/q + 1 \}$$

Nu is $P(S|Z) / P(S|W)$ juist de selectiviteitsratio r , dus

$$(\varphi) \quad c = 1 / \{ r * (1-q)/q + 1 \}$$

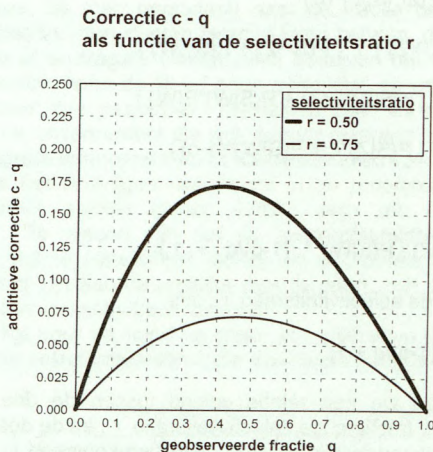
Met deze vergelijking hebben we een relatie gelegd tussen de drie betrokken grootheden, de geobserveerde fractie q , de selectiviteitsratio r , en de doelfractie c .

Merk op dat het selectiemodelletje waarop deze berekeningen is gebaseerd erg eenvoudig is. De enige assumptie die gemaakt wordt is dat binnen de beide strata (kenmerk aanwezig resp. afwezig) wel sprake is van aselechte steekproeftrekking.



In figuur 1 laten we het door (ϕ) gespecificeerde verband zien, voor twee waarden van de selectiviteitsratio r , te weten de waarden $r = 0.75$, en $r = 0.50$. In de volgende paragraaf zal ik betogen dat dat in de praktijk al vrij extreme waarden voor r zijn.

Uiteraard komt de doelfractie c hoger uit dan de geobserveerde fractie q , maar de verschillen zijn niet bovenmatig groot, met name niet voor de waarde $r = 0.75$. Voor de beschouwde waarden van r is de mate van vertekening die selectieve respons teweeg brengt, bescheiden te noemen. Dat blijkt nogmaals in figuur 2, waar we het verschil tussen c en q uitzetten, dus de correctie die men op de geobserveerde fractie moet optellen om voor selectiviteit te corrigeren.



Figuur 2

In figuur 2 zien we dat de maximale correctie voor selectiviteit kleiner is dan 0.075 als $r = 0.75$, en kleiner dan 0.175 als $r = 0.50$.

Hoe gebruiken we deze gegevens nou in de praktijk? Stel dat we in ons voorbeeld van de zwartwerkende aannemers zouden hebben gevonden dat $q = 0.40$. Dan durven we dus, als we de assumptie maken dat r niet onder de 0.75 zal liggen, de gevolgtrekking te maken dat ook na correctie voor selectiviteit de populatiefractie niet boven de $\frac{1}{2}$ zal liggen. Maar als we een q van 0.45 zouden hebben gevonden, zou die conclusie niet gerechtvaardigd zijn, en zou nader onderzoek naar selectiviteit geboden zijn.

4. Hoe groot is r in de praktijk?

Ik heb u hierboven gevraagd uw eigen schatting van de selectiviteitsratio in het geval van de zwartwerkende aannemers te geven. Ik weet uiteraard niet wat u daar heeft gekozen, maar ik heb dit experimentje al diverse malen bij lezingen uitgetprobeerd,

zowel onder methodologen als onder BTW-fiscalisten. Hun 'educated guesses' van de selectiviteitsratio liepen uiteen van 0.4 tot 1.0, met een mediane waarde van 0.8 en kwartielen 0.7 en 0.9. Ik vind het daarom niet onredelijk om te stellen dat de waarden die ik in de figuren 1 en 2 heb afgebeeld, $r = 0.75$ en $r = 0.5$, zeker niet te optimistisch zijn. Wie denkt dat een selectiviteitsratio onder de 0.75 ligt is al behoorlijk pessimistisch, en een waarde onder de 0.5 is extreem. Als u deze gedachte overneemt, ziet u dat het met het effect van selectieve respons meevalt.

Ook in een aantal feitelijke studies waarin ik heb geparticipeerd, bleek de selectiviteitsratio wel mee te vallen. Het gaat om studies waarin concrete informatie over de selectiviteitsratio kon worden verkregen. In alle gevallen was de resulterende r niet erg laag, ofschoon het vaak wel degelijk om gevoelige vormen van regel-overtreding ging. In een studie onder belastingontduikers en correcte belasting-betalers bleek $r = 0.84$ (Elffers, Weigel & Hessing, 1987), in een soortgelijke studie (Elffers, 1991) was r zelfs nagenoeg gelijk aan 1, in een studie naar fraude met werkloosheidsuitkeringen (Robben, Elffers & Verlind, 1989) was $r = 0.79$ bij het vergelijken van fraudeurs en reguliere uitkeringsgerechtigden.

Uiteraard betekent een en ander niet dat extremere waarden van de selectiviteitsratio niet zouden *kunnen* voorkomen, en dan ook heel grote effecten zouden kunnen hebben. Als we een geval zouden hebben waar de selectiviteitsratio 0.1 is (erg extreem, maar niet onmogelijk), dan zou de correctie volgens (ϕ) maximaal 0.52 bedragen.

Het is daarom een verstandig idee in voorkomende gevallen een indruk te verwerven over de geldende waarde van r . Dat kan veelal wel –quick and dirty– door een korte rondvraag bij enkele in het betreffende veld ingevoerde specialisten, door ze te laten schatten wat de responsrates onder de groepen met en zonder het kenmerk zullen zijn. Door die waarde op zo'n manier te benaderen, verkrijgt men een indruk in welke orde van grootte de selectiviteitsvertekening ligt.

5. Een reëel voorbeeld

Van Bijsterveld (1980) schokte politiek en fiscaal Nederland door te publiceren dat slechts bij 34% van de ondernemers bij wie een boekenonderzoek (dat is een grondige doorlichting van de administratie van een bedrijf door de belastinginspectie) was ingesteld géén poging tot fraude was vastgesteld. De steekproefgrootte is zodanig dat voor een aselechte steekproef een steekproefonzekerheid van plus of min 1% geldt. Uiteraard realiseerde ook Van Bijsterveld zich terdege dat de verzameling boekenonderzoeken geen aselechte steekproef van alle bedrijfs-belastingaangiften is, omdat de belastingdienst niet alleen routine-controles uitvoert, maar ook regelmatig controleert juist omdat men een vermoeden van incorrecte aangifte (niet noodzakelijk frauduleus) heeft. In zijn rapport bespreekt hij uitgebreid waarom hij niettemin gelooft dat die 34% ook ten naaste bij in de populatie van ondernemings-aangiften zal gelden, een mening die indertijd in de fiscale literatuur en de Tweede Kamer allerminst door een ieder werd aanvaard.

In dit geval is het goed mogelijk een onderbouwde schatting van de selectiviteitsratio te maken, omdat Van Bijsterveld in zijn rapport aangeeft dat de belastingdienst tien groepen ondernemers onderscheidt, waarin wel of geen a priori reden is om te vermoeden dat er mogelijksterwijs gecorrigeerd moet worden. Door het feitelijk percentage boekenonderzoeken in de 'onverdadichte' groepen te vergelijken

met dat percentage in de 'verdachte' groepen, krijgt men een schatting van de selectiviteitsratio $r = 0.92$.

Op grond van (φ) blijkt dan dat de gevonden q van 0.34 bij deze selectiviteitsratio duidt op een doelfractie $c = 0.36$, dus dat inderdaad zoals Van Bijsterveld stelde, slechts sprake is van een minieme vertekening. Mijn onderbouwing is echter een geheel andere dan die van Van Bijsterveld.

6. Selectiviteitsvertekening en steekproefomvang

Het is interessant selectiviteitsvertekening af te zetten tegen steekproefvariatie (onder assumptie van aselechte steekproeftrekking). Als we weer het geval $r = 0.75$ bekijken, dan is de vertekening blijkens figuur 2 kleiner dan 0.075. Zoals bekend is de steekproefon nauwkeurigheid (halve breedte van het 95%-betrouwbaarheids-interval van de schatting van een fractie) in het ongunstigste geval (als de te schatten fractie gelijk is aan $\frac{1}{2}$) ongeveer $n^{-1/2}$, dat is dus bij $n=178$ ook 0.075. Deze berekening geeft ons een richtlijn waarover men zich in een concreet geval meer moet bekreunen, over selectieve respons of over steekproefon nauwkeurigheid.

Uiteraard kan men ook beide onnauwkeurigheden optellen om te zien wat de maximale vertekening zou kunnen zijn. In verkiezingsonderzoek is het tegenwoordig gebruikelijk om op te geven dat de resultaten, op grond van steekproefon nauwkeurigheid, een bepaalde marge hebben. Het is goed denkbaar om daaraan de selectiviteitsonzekerheid toe te voegen, als men een assumptie over r durft te doen. Wanneer iemand steekproefonderzoek naar politieke voorkeur doet, en zich afvraagt hoe groot de aanhang van een nieuwe partij (bijvoorbeeld 'Leefbaar Nederland', LN) zal blijken, en men vindt in een steekproef van 300 Nederlanders 6% aanhangers van LN, hoe groot is dan de marge? De steekproefon nauwkeurigheid wordt dan grofweg op $2 \cdot \sqrt{(0.06 \cdot 0.94/300)} \approx 0.027$ geschat. De LN-marge zou dan op 3.3 tot 8.7% (5 tot 13 zetels) gesteld worden. Men mag evenwel ook aannemen dat er sprake is van selectieve respons, omdat aanhangers van LN, als nieuw-bekeerden, op dit moment wellicht eerder geneigd zullen zijn aan verkiezingsonderzoek mee te doen, om hun zegje te kunnen doen en hun favoriete partij de broodnodige steun te geven. Als we aannemen dat de responsbereidheid van niet-LN-ers t.o.v. van LN-ers weer $r=0.75$ zou zijn (of dat in dit geval een reële veronderstelling is, heb ik niet bij deskundigen gecheckt), dan leidt dat tot een vertekening: volgens figuur 2 of (φ) komen de 94% niet LN-stemmers in de steekproef overeen met 95.4% niet stemmers in de populatie onder deze assumptie, en dus navenant staan de 6% LN-ers in de steekproef voor 4.6% in de populatie. Pas op deze waarde dient men dan de steekproefon nauwkeurigheid van plus of min 2.7% toe te passen. Tezamen komt dat dan uit op een steun voor LN tussen de 1.9% en 7.3% (3 tot 11 zetels).

7. Besluit

Wat ik heb betoogd is dat het mogelijk is om op grond van een relatief eenvoudige aanname over selectiviteit de doorwerking ervan in ogenschouw te nemen, en ik heb aan de hand van enkele voorbeelden laten zien hoe dat in zijn werk kan gaan. Daarbij heb ik betoogd dat het de moeite loont om door rondvragen onder experts een indruk te krijgen van de geldende selectiviteitsratio.

Uiteraard is de hier behandelde situatie, schatting van de fractie van voorkomen van een kenmerk de eenvoudigst denkbare, en daarom ook zo goed te modelleren: het voorgestelde model is nagenoeg niet restrictief.

Als we meer weten of meer durven te veronderstellen kunnen we ook complexere situaties te lijf gaan. In Elffers(1982) behandel ik verkiezingsonderzoek met selectieve respons die verschillend is in de strata 'gaat feitelijk wel stemmen' en 'gaat feitelijk niet stemmen', zowel onder de aanhangers als onder de tegenstanders van de coalitie.

Over selectieve respons is uiteraard veel geschreven, zij het meestal in een wat andere toonzetting dan ik hier gekozen heb. Zonder enige aanspraak op volledigheid noem ik een drietal boeken en artikelen die mij geïnspireerd hebben, te weten Bethlehem en Kersten (1986), Heckman (1979) en Winship & Mare (1992), als titels die selectieve respons in een wat algemenere context behandelen.

8. Literatuur

- Bethlehem, J.G. & H.M.P. Kersten (1986), *Werken met non-respons*. Proefschrift Amsterdam.
- Bijsterveld, W.J. van (1980), *Aangepaste versie van het verslag van een onderzoek naar de aard en omvang van de belastingfraude, uitgebracht in april 1979 aan de Minister en aan de Staatssecretaris van Financiën door W.J. van Bijsterveld*. Den Haag: Ministerie van Financiën.
- Elffers, H. (1982), Non-randomness of samples and bold assumptions. In: C.P. Middendorp, B. Niemöller & W.E. Saris, *Sociometric Research, Proceedings of the 2nd Congress of the Sociometric Society at the Vrije Universiteit*, Amsterdam, p. 2-10.
- Elffers, H. (1991), *Income tax evasion. Theory and measurement*. Proefschrift EUR. Reeks Burger en Belastingen nr. 9, Deventer: Kluwer.
- Elffers, H., R.H. Weigel & D.J. Hessing (1991), The consequences of different strategies for measuring tax evasion behavior, *Journal of Economic Psychology*, 8, 311-337
- Heckman, J. (1979), Sample selection bias as a specification error. *Econometrica* 47, 153-161.
- Robben, H.S.J., H. Elffers & W.F.M. Verlind (1989), *Determinanten van individueel misbruik van sociale verzekeringen: het geval van WW-misbruik*. Rapport R89-02. Amsterdam: Federatie van Bedrijfsverenigingen.
- Winship, C. & R.D. Mare (1992), Models for sample selection bias. *Annual Review of Sociology*, 18, 327-350.

