

Foutenbronnen bij statistisch onderzoek

10

Jelke Bethlehem

Statistische Methoden (10004)



Verklaring van tekens

.	= gegevens ontbreken
*	= voorlopig cijfer
**	= nader voorlopig cijfer
x	= geheim
–	= nihil
–	= (indien voorkomend tussen twee getallen) tot en met
0 (0,0)	= het getal is kleiner dan de helft van de gekozen eenheid
niets (blank)	= een cijfer kan op logische gronden niet voorkomen
2008–2009	= 2008 tot en met 2009
2008/2009	= het gemiddelde over de jaren 2008 tot en met 2009
2008/'09	= oogstjaar, boekjaar, schooljaar enz., beginnend in 2008 en eindigend in 2009
2006/'07–2008/'09	= oogstjaar, boekjaar enz., 2006/'07 tot en met 2008/'09

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

Colofon

Uitgever
Centraal Bureau voor de Statistiek
Henri Faasdreef 312
2492 JP Den Haag

Prepress
Centraal Bureau voor de Statistiek - Grafimedia

Omslag
TelDesign, Rotterdam

Inlichtingen
Tel. (088) 570 70 70
Fax (070) 337 59 94
Via contactformulier: www.cbs.nl/infoservice

Bestellingen
E-mail: verkoop@cbs.nl
Fax (045) 570 62 68

Internet
www.cbs.nl

ISSN: 1876-0333

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen, 2010.
Verveelvoudiging is toegestaan, mits het CBS als bron wordt vermeld.

Inhoudsopgave

1.	Inleiding	4
1.1	Algemene beschrijving en leeswijzer	4
1.2	Afbakening en relatie met andere thema's	4
1.3	Plaats in het statistisch proces.....	4
2.	Systematische en toevallige fouten	6
3.	Een taxonomie van fouten.....	8
3.1	Steekproeffouten.....	9
3.2	Niet-steekproeffout.....	9
3.3	Literatuur	12

1. Inleiding

1.1 Algemene beschrijving en leeswijzer

Bij het uitvoeren van een survey-onderzoek krijgt een onderzoeker met allerlei verschijnselen te maken die de kwaliteit van de uitkomsten kunnen beïnvloeden. Sommige verstoringen kunnen worden voorkomen door ermee rekening te houden bij de opzet en uitvoering van een onderzoek. Het optreden van sommige andere verstoringen is echter niet te voorkomen. Inspanningen zullen er dan vooral op zijn gericht de invloed ervan beperkt te houden.

De foutenbronnen zullen, indien aanwezig, leiden tot een grotere onzekerheid ten aanzien van de juistheid (validiteit) van de uitkomst. Deze onzekerheid kan zich op twee manieren manifesteren: als een systematische afwijking (vertekening) of als een grotere variatie in de uitkomsten (de mogelijke uitkomsten fluctueren meer om de werkelijke waarde). Dit wordt in meer detail besproken in hoofdstuk 2.

Al die verstoringen kunnen leiden tot afwijkende schattingen van populatie-grootheden. Verschijnselen die tot dit type verstoringen aanleiding kunnen geven, worden aangeduid met *foutenbronnen* en de afwijking die ze samen veroorzaken heet de *totale fout*. Hoofdstuk 3 geeft een taxonomie van fouten, waarbij de totale fout wordt ontleed in diverse specifieke foutensoorten

Er zijn verschijnselen die zich alleen bij steekproefonderzoek kunnen voordoen. Andere verschijnselen kunnen ook optreden bij integraal onderzoek of bij het gebruik van gegevens uit registers.

1.2 Afbakening en relatie met andere thema's

Verstorende verschijnselen zullen zich altijd en overal voordoen. Daarom is het belangrijk om bij de opzet van het onderzoek hier al rekening mee te houden. Veel onderdelen van het statistisch proces zijn erop gericht die verstoringen binnen de perken te houden of achteraf te corrigeren. Immers, er kunnen alleen cijfers van voldoende kwaliteit worden gepubliceerd indien de gebruikte gegevens niet al te veel zijn aangetast.

1.3 Plaats in het statistisch proces

Het probleem van de verstoringen en de fouten die ze kunnen veroorzaken, komt overal in het statistisch proces terug. Hierbij een aantal voorbeelden:

- Het ontwerpen van vragenlijsten is erop gericht dat de vragen correct worden gesteld en beantwoord.
- Het steekproefontwerp moet zodanig zijn dat precieze schattingen kunnen worden gemaakt.
- Het gaafmaken is bedoeld om allerlei fouten uit de gegevens te halen.

- Met imputatietechnieken wordt getracht te corrigeren voor de gevolgen van ontbrekende antwoorden op vragen.
- Met wegen wordt geprobeerd te corrigeren voor vertekeningen als gevolg van unit non-respons.

2. Systematische en toevallige fouten

Laat Z een populatiegrootte zijn die moet worden geschat. Voorbeelden van populatiegrootheden zijn het populatietotaal, het populatiegemiddelde en het populatiepercentage. Laat z de schatter zijn die wordt gebruikt om de populatiegrootte te schatten op basis van de beschikbare gegevens. De voorkeur gaat uit naar een *zuivere schatter*, dat wil zeggen

$$E(z) = Z. \quad (2.1)$$

Met andere woorden betekent dit dat als het onderzoek een groot aantal malen zou worden herhaald onder dezelfde omstandigheden, de schatting gemiddeld op de correcte waarde (de waarde van de populatiegrootte) uitkomt.

Is de schatter niet zuiver, dan is er sprake van een vertekening (Engels: bias). Die vertekening wordt aangegeven met

$$B(z) = E(z) - Z. \quad (2.2)$$

De variantie van de schatter moet klein zijn, dat wil zeggen dat

$$V(z) = E(z - E(z))^2 \quad (2.3)$$

klein moet zijn. Als de variantie klein is, wordt de schatter *precies* genoemd. De variantie zegt niet veel over de kwaliteit van de schatter als deze onzuiver is. Een betere maat hiervoor, en dus ook voor de totale fout, is de *gemiddelde kwadratische fout*. Deze is gedefinieerd als

$$G(z) = E(z - Z)^2. \quad (2.4)$$

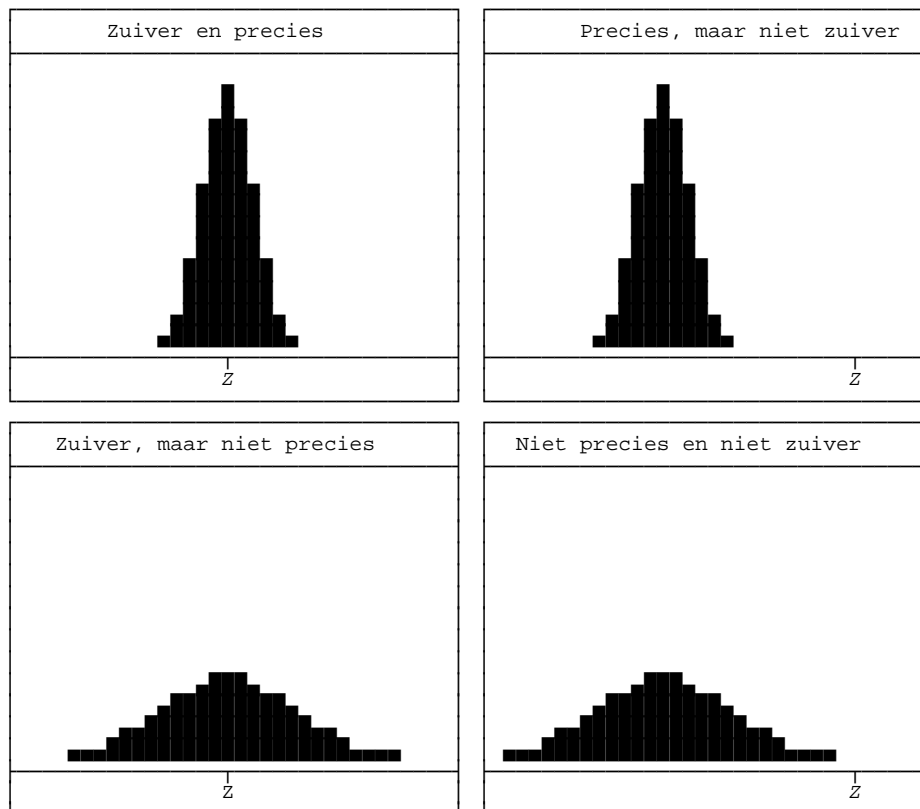
Uitwerken van deze definitie leidt tot een andere uitdrukking voor de gemiddelde kwadratische fout:

$$G(z) = V(z) + B^2(z). \quad (2.5)$$

Aan deze uitdrukking is te zien dat de gemiddelde kwadratische fout is opgebouwd uit een variantieterm en een vertekeningsterm (in het kwadraat). Is de vertekening van de schatter 0, dan reduceert de gemiddelde kwadratische fout tot de variantie. Voor een kleine totale fout moeten dus zowel de variantie als de vertekening klein zijn. In figuur 1 worden vier situaties onderscheiden.

De situatie links boven beschrijft de ideale situatie: een zuivere schatter die bovendien nauwkeurig is. Iets minder aantrekkelijk is de situatie links onder. De schatter is nog wel zuiver maar niet zo nauwkeurig. Dat is jammer, maar op basis van de berekende betrouwbaarheidsintervallen zal geen verkeerde conclusie worden getrokken. Heel anders ligt de situatie rechts boven. De schatter heeft een grote vertekening, maar een kleine variantie. Het is bijna zeker dat op basis van de berekende betrouwbaarheidsintervallen verkeerde conclusies worden getrokken.

Figuur 1. De relatie tussen totale fout, vertekening en precisie



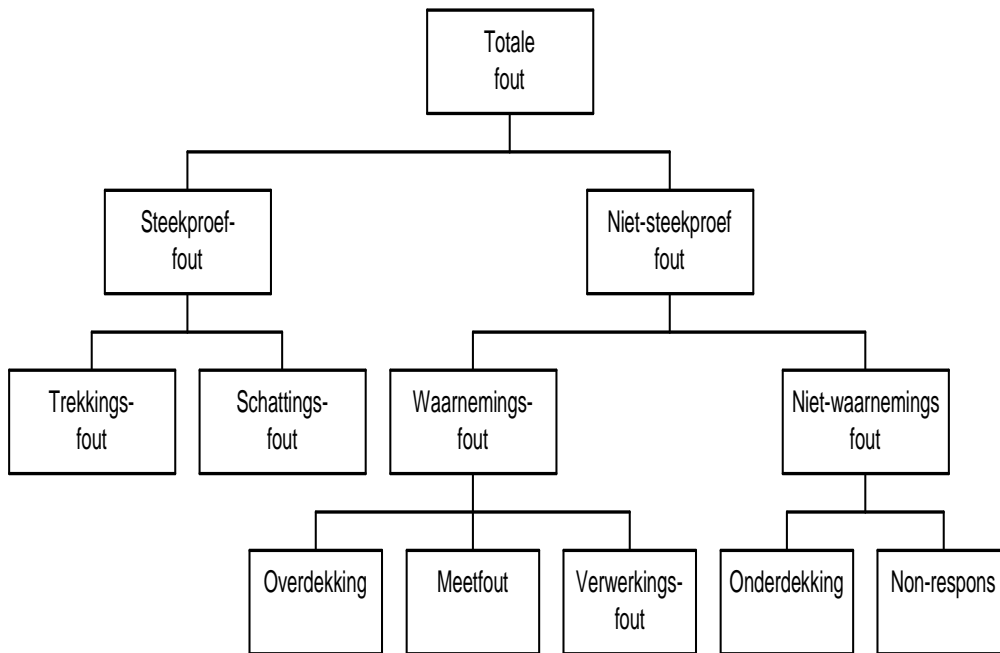
De situatie rechts onder levert de grootste onzekerheid op. De gemiddelde kwadratische fout is hier het grootst.

Uit de grafiek wordt duidelijk dat het onverstandig is om alleen af te gaan op de variantie bij het beoordelen van de kwaliteit van de uitkomsten als het vermoeden bestaat dat er wel eens sprake zou kunnen zijn van vertekening.

3. Een taxonomie van fouten

Het maken van een indeling van mogelijke foutenbronnen kan behulpzaam zijn bij het onderzoeken en lokaliseren van allerlei verschijnselen die de kwaliteit van de uitkomsten van steekproefonderzoek kunnen aantasten. De in figuur 2 gegeven indeling heeft een hiërarchische structuur, waarbij een steeds grotere verfijning wordt aangebracht in de omschrijving van de fouten. De taxonomie is afkomstig van Bethlehem (1999). Het is een uitgebreide versie van een schema dat eerder door Kish (1967) is beschreven.

Figuur 2. De ontleding van de totale fout



De totale fout wordt eerst gesplitst in de twee hoofdcategorieën, die worden aangeduid met steekproeffout en niet-steekproeffout. De *steekproeffout* is de fout die ontstaat doordat niet de gehele populatie wordt onderzocht maar slechts een beperkt deel (een steekproef) daarvan. De steekproeffout treedt niet op bij een integraal onderzoek of bij registers die de populatie volledig dekken.

De *niet-steekproeffout* is de afwijking die ontstaat door verschijnselen die ook zouden zijn opgetreden indien een integraal onderzoek was gehouden (met dezelfde onderzoekstechnieken). Dus ook in een register kan een niet-steekproeffout optreden. Het is duidelijk dat bij een register geen sprake is van een steekproef. De term niet-steekproeffout (Eng: non-sampling error) is zo ingeburgerd dat hij toch wordt gehanteerd. De niet-steekproeffout kan vanzelfsprekend ook optreden bij onderzoeken waarbij wel met steekproeven wordt gewerkt.

3.1 Steekproeffouten

De steekproeffout wordt onderverdeeld in twee componenten: de trekkingsfout en de schattingsfout. De *trekkingsfout* is de fout die wordt geïntroduceerd doordat de werkelijke trekkingskansen van de elementen afwijken van de bij het steekproefontwerp geanticiperde trekkingskansen waarvan wordt uitgegaan bij de berekening van de schatting. Bij het trekken van de steekproef is het mogelijk dat door het gebruiken van een verkeerde trekkingsprocedure of door de opbouw van het steekproefkader, sommige elementen een andere trekkingskans krijgen dan was bedoeld. Een veel voorkomende oorzaak is het dubbel of nog vaker voorkomen van elementen in het steekproefkader.

Wanneer onder gelijk blijvende omstandigheden het (op correcte wijze) trekken van een steekproef een aantal malen wordt herhaald, dan zullen de schattingen op grond van deze steekproeven niet steeds aan elkaar gelijk zijn. Dit wordt veroorzaakt door het feit dat de steekproeven verschillende elementen kunnen bevatten en de daarbij behorende waarden van de doelvariabelen zullen ook niet steeds dezelfde zijn. De fout die ontstaat doordat toevallig andere elementen in de steekproef worden getrokken, heet de *schattingfout*. Deze fout wordt gekwantificeerd in de variantie of standaardfout van de schatter.

Een foutenbron die verband houdt met de trekkingsfout en het trekken van steekproeven uit steekproefkaders is het feit dat elementen soms niet met de juiste gegevens in het steekproefkader zijn opgenomen. Men spreekt in dit verband van *kaderfouten*. Bij kaderfouten kan worden gedacht aan het feit dat bevolkingsadministraties soms achterlopen en mutaties zoals bijvoorbeeld verhuizingen nog niet verwerkt zijn. Wordt uit zo'n bevolkingsadministratie een personensteekproef getrokken, dan kunnen personen in de steekproef terecht komen die niet meer op het door de gemeente opgegeven adres wonen. Kaderfouten kunnen niet worden beschouwd als alleen een onderdeel van de steekproeffout; zij kunnen ook optreden bij integraal onderzoek.

3.2 Niet-steekproeffout

De niet-steekproeffout kan worden gesplitst in een waarnemingsfout en een niet-waarnemingsfout. De *waarnemingsfout* is dat gedeelte van de niet-steekproeffout dat ontstaat door het op incorrecte wijze verzamelen, vastleggen en verwerken van de gegevens. De waarnemingsfout kan ontstaan *doordat* waarnemingen worden verricht. Een *niet-waarnemingsfout* ontstaat doordat het niet gelukt is om waarnemingen te verrichten.

De waarnemingsfout kan worden gesplitst in drie componenten: overdekking, meetfout en verwerkingsfout. *Overdekking* kan zich voordoen als het steekproefkader of register elementen bevat die niet tot de doelpopulatie behoren. Het ten onrechte meenemen van dit soort elementen in het onderzoek kan ook tot verkeerde conclusies leiden. Overdekking kan vrij eenvoudig worden ontdekt in het veld, door voor elk element in de steekproef eerst te verifiëren of het wel voldoet aan de definitie van de doelpopulatie.

De *meetfout* is dat gedeelte van de waarnemingsfout dat ontstaat doordat de gegevens die van de elementen na waarneming en verwerking beschikbaar komen, niet met de werkelijkheid overeenkomen. Een meetfout kan door een groot aantal verstoringen worden veroorzaakt. Deze verstoringen kunnen verschillende invloeden op de uitkomsten hebben. Een onduidelijke vraagstelling kan leiden tot misverstanden, en daardoor tot verkeerde antwoorden. Het stellen van irrelevante vragen (door het volgen van de verkeerde route) kan irritatie opwekken, met alle gevolgen van dien voor de beantwoording van de vragen. Er kunnen verstoringen optreden als gevolg van interacties tussen respondent, interviewer en andere aanwezigen bij het vraaggesprek, en door de situatie (omgeving) waarin het vraaggesprek plaatsvindt. Factoren die hierbij een rol spelen zijn huidkleur, geslacht, leeftijd, opleiding en sociale klasse van interviewer en/of respondent. Waar de onderwerpen gevoelig liggen, bestaat de mogelijkheid dat de respondent soms sociaal-wenselijke antwoorden geeft.

Problemen kunnen ook ontstaan bij het stellen van vragen die betrekking hebben op het verleden van de respondent. Gebeurtenissen kunnen worden vergeten of op een verkeerd moment in de tijd worden geplaatst. Dit soort verschijnselen, die ook wel *geheugeneffecten* worden genoemd, doen zich voor bijvoorbeeld voor bij vragen over aankopen van goederen, bioscoopbezoek en raadplegen van een arts.

De derde component van de waarnemingsfout is de verwerkingsfout. De *verwerkingsfout* is dat gedeelte van de waarnemingsfout dat ontstaat door verstoringen bij de verwerking van de gegevens. Hierbij kan worden gedacht aan fouten bij het interpreteren, controleren en coderen van de antwoorden, het intypen van de codes, het programmeren, het uitvoeren van wegingprocedures en analysetechnieken.

Net als de waarnemingsfout, is de niet-waarnemingsfout opgebouwd uit een aantal componenten, namelijk onderdekking en non-respons. *Onderdekking* doet zich voor als niet alle elementen uit de doelpopulatie terug te vinden zijn in het steekproefkader of het register. Dit verschijnsel treedt, bijvoorbeeld, op ten aanzien van illegaal in ons land verblijvende buitenlanders, terwijl er een steekproef uit het persoonsregister van de gemeenten wordt getrokken. De gevolgen van onderdekking kunnen ernstig zijn, vooral als een selecte groep uit de doelpopulatie daardoor niet wordt meegenomen in het onderzoek, waardoor de resultaten een vertekend beeld kunnen opleveren. Daarbij komt bovendien dat onderdekking niet eenvoudig valt waar te nemen als men er niet echt heel hard naar zoekt.

Onder *non-respons* verstaat men het verschijnsel dat van elementen die tot de doelpopulatie behoren en die in de steekproef zijn getrokken niet alle gegevens kunnen worden verkregen. Het probleem van de non-respons wordt uitgebreid beschreven in andere delen van de Methodenreeks.

Sommige foutenbronnen kunnen heel goed onder controle worden gehouden. Zo kan door toepassing van een geschikt steekproefontwerp, een goede schatter en een grote steekproefomvang de schattingsfout klein worden gehouden. Ook is er soms iets te doen aan de problemen die ontstaan als gevolg van non-respons. Verder kan het

gebruik van computergestuurd enquêteren helpen bij het detecteren en corrigeren van meetfouten en verwerkingsfouten.

Als echter met traditionele papieren vragenlijsten wordt gewerkt, kan worden verwacht dat er nog fouten in de gegevens zitten. Enerzijds kunnen dat inconsistenties zijn (een gehuwd kind van vier jaar) en anderzijds kan dat *item non-respons* zijn. Bij dit laatste zijn één of meer (maar niet alle) vragen ten onrechte niet beantwoord. Dat betekent dat achteraf nog een uitgebreid controle- en correctieproces zal moeten plaatsvinden.

In het vorige hoofdstuk is beschreven dat de hierboven beschreven fouten verschillende effecten kunnen hebben. Er kan sprake zijn van een toevallig of een systematisch effect. Ook is het denkbaar dat sommige fouten zowel een toevallig als een systematisch effect kunnen hebben.

3.3 Literatuur

Bethlehem. J.G. (1999), Cross-sectional Research. In: H.J. Adèr and G.J. Mellenbergh, *Research Methodology in the Social, Behavioural & Life Science*. Sage Publications, London, pp.110-142.

Bethlehem. J.G. (2009), *Applied Survey Methods, A Statistical Perspective*. John Wiley & Sons, Hoboken, NJ, USA.

Kish, L. (1967), *Survey Sampling*. John Wiley & Sons, New York, USA.

Groves, R.M. (1989), *Survey errors and survey costs*. Wiley, New York.