

## UITBREIDING VAN DE STEEKPROEF

door J. H. Blokdijk

In het mei nummer van dit maandblad heeft collega M. Vermaas over het in de titel aangeduide onderwerp een artikel (4) geschreven waarin hij veelvuldig verwijst naar de opvattingen van de subcommissie Controle en Steekproeven (CCS), ingesteld door de Commissie van Advies inzake Controlevraagstukken van het NIVRA. Als voorzitter van genoemde subcommissie kan ik de verleiding niet weerstaan op dit artikel te reageren.

De probleemstelling is als volgt met een cijfervoorbeeld te illustreren. Indien men in een steekproef van 75 stuks geen fouten aantreft, kan men met 95% betrouwbaarheid beweren dat de populatie hoogstens 4% fouten bevat; indien men in een steekproef van 119 stuks niet meer dan één fout vindt, kan men met dezelfde betrouwbaarheid hetzelfde beweren. Kan men nu, indien men in een steekproef van 75 stuks één fout aantreft, de steekproef vervolgens met 44 stuks uitbreidt en daarin geen fout vindt, met 95% betrouwbaarheid beweren dat de populatie hoogstens 4% fouten bevat?

Vermaas onderkent twee „opvattingen” hierover bij de CCS. De „oude opvatting” ziet hij belichaamd in een zin uit het interimrapport van de CCS, gepubliceerd in juli/augustus 1974 (2); deze zin luidt: „Als in de steekproef slechts één fout gevonden wordt kan men het vereiste zekerheidsniveau toch nog bereiken door de steekproef uit te breiden tot een zodanige omvang dat bij het aantreffen van één fout tolerantie en risico gelijk zijn aan die van de oorspronkelijke steekproef indien geen fout wordt aangetroffen.” Met „risico” is hier bedoeld het complement van de betrouwbaarheid, met „tolerantie” de maximaal toelaatbare foutenfractie.

De „nieuwe opvatting” vindt Vermaas in het verslag van E. B. Tuitjer en J. M. Zuydervliet (3) van de NIVRA-najaarsconferentie 1974 over steekproeven, alsmede in een artikel van schrijver dezes (1) uit 1976. In beide publikaties wordt de in de probleemstelling gestelde vraag ontkennend beantwoord.

Vermaas is daardoor niet overtuigd. Hij verduidelijkt het probleem met een eigentijds sprookje. Het gerucht gaat dat in circa  $1/3$  van de aangiften voor de inkomstenbelasting opzettelijke telfouten voorkomen die niet door de belastingdienst ontdekt worden. De verantwoordelijke bewindsman laat weten dat dit in hoogstens 4% van de aangiften het geval is. Hij organiseert een televisieprogramma waarin deze bewering op dramatische wijze „bewezen” moet worden. Dartoefnodigt hij vier accountants uit, die ieder een steekproefprocedure uitwerken waarbij met 95% betrouwbaarheid de gevraagde bewering gedaan kan worden indien ten hoogste het hieronder aangegeven aantal telfouten wordt aangetroffen. De vier steekproefprocedures zijn de volgende:

- accountant A: 75 stuks, 0 telfouten;
- accountant B: 119 stuks, 1 telfout;
- accountant C: 88 stuks, 0 telfouten; indien 1 telfout wordt aangetroffen: nog 44 stuks, 0 telfouten;
- accountant D: 194 stuks, 3 telfouten.

Het welhaast onvermijdelijke geschiedt: het 60e en het 125e geselecteerde biljet bevatten een telfout. De accountants B en D kunnen de gevraagde uitspraak doen, de accountants A en C niet (maar zij kunnen ook het tegendeel niet beweren!). De bewindsman is nogal boos op A en C en vindt dat accountants er maar vreemde beroepsopvattingen op na houden.

Dit was dan het „unhappy end” van het door Vermaas veel levendiger vertelde sprookje. Het nare slot is volgens hem te wijten aan de zogeheten „nieuwe opvatting” van de CCS.

### **De geschiedenis van de „opvattingen”**

Nu eerst een klein stukje historie. Om de NIVRA-najaarsconferentie 1974 zo goed mogelijk te doen verlopen, diende het interimrapport van de CCS medio 1974 gereed te zijn. Dit rapport ging met name over de aanvaardbaarheid en de toepasbaarheid van de steekproef in de accountantscontrole; de praktische vraagstukken waarmee de accountant bij de toepassing van dit statistische hulpmiddel geconfronteerd wordt zijn daarin bewust nog slechts zeer ten dele aan de orde gesteld.

Dit geldt ook voor het onderhavige probleem. De CCS was zich het vraagstuk wel bewust, maar kon het niet meer volledig uitdiepen en meende met de eerder geciteerde zin voldoende ruimte voor een nadere uitwerking geschapen te hebben.

Het rapport behoorde tot de conferentiestukken en kwam daardoor al snel onder ogen van Prof. Dr. W. Molenaar, hoogleraar statistiek en van Prof. J. Kriens, hoogleraar besliskunde, die op de conferentie de inbreng op statistisch gebied hebben geleverd (de conferentie is tweemaal gehouden). Zij reageerden in niet mis te verstane termen op de eerder geciteerde zin. De CCS moest, na beraad, haar fout erkennen en schrijver dezes heeft dit op beide conferenties ter kennis van de deelnemers gebracht. Dit is ook in het verslag van Tuitjer en Zuydervliet (3) vermeld en deze passage is niet aan de aandacht van Vermaas ontsnapt.

Het is dus wat zwaar om van twee „opvattingen” te spreken, in de zin dat de CCS zou zijn „omgegaan”; er is veeleer sprake van een ongelukkige formulering en daarop volgende pogingen om de schade zoveel mogelijk te beperken.

### **Kans en betrouwbaarheid**

Deze kleine geschiedenis vertel ik om duidelijk te maken dat het hier niet om beroepsopvattingen van *accountants* gaat: het is een statistisch probleem. De gekweldde bewindsman in het sprookje van Vermaas had voor zijn televisieshow dan ook geen accountants moeten vragen, maar statistici! En hoewel ik niet goed gekwalificeerd ben om hierover een voorspelling te doen, meen ik dat de zaak dan een geheel andere wending gekregen zou hebben.

De statistici zouden, naar ik meen, de bewindsman duidelijk hebben gemaakt dat het onverstandig is vier steekproeven met gedeeltelijk dezelfde trekkingen te laten uitvoeren, omdat daarmee de *kans* op uiteenlopende uitspraken geschapen wordt. De kern van het statistische vraagstuk is gelegen in deze kans, in samenhang met de betrouwbaarheid van de uitspraak. Vermaas vindt de kern van het verschil van inzicht dan ook terecht terug in de definitie van het begrip betrouw-

baarheid. Hij geeft hiervan twee verschillende omschrijvingen die ik volledig citeer:

„In de „nieuwe opvatting” betekent een betrouwbaarheid van 95% dat, *als* de onderzochte en geaccepteerde populatie een populatie is, waarin meer dan het maximaal toelaatbare foutenpercentage voorkomt, zij hoogstens 5% kans heeft gekregen te worden geaccepteerd.

In de „oude opvatting” betekent een betrouwbaarheid van 95%, dat er maar 5% kans is dat de onderzochte en geaccepteerde populatie een populatie is, waarin meer dan het maximaal toelaatbare foutenpercentage voorkomt.”

Nu kan er in de situatie dat een bestaande populatie onderzocht wordt, geen sprake meer zijn van een *kans* dat de populatie een ontoelaatbaar foutenpercentage bevat: de populatie bevat een zeker foutenpercentage en dat *is* al dan niet hoger dan het toelaatbare. Dat de accountant niet *weet* of het feitelijke percentage al dan niet toelaatbaar is, doet daaraan niet af. Hij heeft een *kans* daarover de juiste of de onjuiste conclusie te trekken en bij toepassing van een steekproef hangt die kans af van de trekking van de te onderzoeken elementen. Dit betekent dat Vermaas’ omschrijving van betrouwbaarheid in de „oude opvatting” een onbestaanbaar begrip omvat!

De kans treedt dus op bij de trekking. Men kan uit een populatie verschillende steekproeven van een bepaalde omvang trekken, met uiteenlopende resultaten: 0, 1, 2, 3 etc. fouten. Als de steekproefomvang en het foutenpercentage in de populatie bekend zijn, zijn de kansen op die verschillende resultaten te berekenen. Vermaas geeft hiervan een aantal voorbeelden.

Maar het foutenpercentage in de populatie is nu juist niet bekend; men wil weten of de populatie aan een gestelde eis voldoet (toetsen). Dan moet een procedure ontwikkeld worden met een *beheersbare* kans op een onjuiste uitspraak. Bekend is dat die kans *maximaal* is in wat Vermaas het „grensgeval” noemt, namelijk in de situatie dat de populatie net niet aan de gestelde eis voldoet. Voor iedere denkbare steekproefomvang is die maximale kans te berekenen en dus is er een steekproefomvang vast te stellen waarbij dit maximum een gewenst geachte hoogte niet overschrijft. Een andere variabele is het aantal fouten dat in de steekproef mag worden aangetroffen zonder dat een positieve uitspraak over de populatie moet worden geweigerd; zo laat Vermaas de accountants A, B en D verschillende goedkeuringscriteria hanteren, waarbij zij dan ook tot verschillende steekproeflengten komen.

Vorenbedoelde maximale kans belichaamt bij toetsing de onbetrouwbaarheid van de steekproefprocedure; de betrouwbaarheid van de steekproef is het complement van die kans.

#### **Uitbreiding van de steekproef**

Wat gebeurt er nu indien een steekproef met een tevoren vastgestelde onbetrouwbaarheid achteraf uitgebreid wordt omdat de uitkomsten in de eerste fase niet aan het gestelde criterium voor aanvaarding van de populatie voldeden?

Iedere uitbreiding is in feite een herkansing: er komt een nieuwe, extra kans op aanvaarding bij. Daarmee neemt ook de maximale kans op ten onrechte aanvaarding toe en vermindert dus de betrouwbaarheid. Dat is ook zonder berekeningen niet onbegrijpelijk: indien men de steekproef niet voortzet als men geen

fout in de eerste fase vindt en wel als er wel een fout in zit, dan vermijdt men in het eerste geval de kwade kans in de verdere fasen „te veel” fouten aan te treffen. De procedure van Vermaas, op basis van de zogeheten „oude opvatting”, leidt in theorie, althans bij trekking met terugleggen, *altijd* tot goedkeuring, ook van onaanvaardbare populaties; toegegeven zij dat men dan soms wel erg lang moet doorgaan! De betrouwbaarheid van de procedure is dan echter vrijwel tot nihil gereduceerd.

De kwade kans in de verdere fasen „te veel” fouten aan te treffen komt ook in het sprookje van Vermaas om de hoek kijken. Accountant B heeft zijn goedkeuring reeds gegeven wanneer er een tweede fout ontdekt wordt, die accountant C, met de wat grotere steekproef, verhindert dezelfde uitspraak te doen. Indien de bewindsman het aantal accountants uit kostenoverwegingen tot drie beperkt zou hebben, zou het sprookje van Vermaas een geheel andere moraal gehad hebben.

Het is wel mogelijk een steekproef met een vaststaand aantal fasen te ontwerpen waarvan de betrouwbaarheid aan een tevoren gestelde eis voldoet: de procedure van accountant C is daarvan een voorbeeld. Vermaas geeft op blz. 209 en 210 een steekproef in 4 fasen; uit de aldaar gegeven berekening blijkt zeer duidelijk dat iedere verdere fase de betrouwbaarheid vermindert. Zolang deze echter aan de gestelde eis voldoet, is een dergelijke steekproefprocedure stellig verantwoord.

#### **Een verantwoorde oplossing**

Het vraagstuk waarvoor Vermaas een oplossing zoekt houdt eigenlijk weinig verband met de kans op het ten onrechte aanvaarden van een slechte populatie, met de betrouwbaarheid van de steekproef; het gaat veeleer om de kans een goede populatie ten onrechte niet te accepteren.

In dit verband voert Vermaas een miljoen populaties ten tonele, met een frequentieverdeling naar foutenpercentage. Zo deze frequentieverdeling van de populaties die aan accountantscontrole onderworpen worden al bekend zou zijn, dan zou dit gegeven nog volmaakt oninteressant zijn. Bij de controle van een jaarrekening worden hooguit een gering aantal populaties met behulp van een steekproef gecontroleerd en op die controles wordt de verklaring (mede) gebaseerd.

Bij de controle van de jaarrekening is het van belang in een populatie fouten met een gezamenlijke omvang die het beeld van de jaarrekening aantast, te ontdekken. De maximale kans op niet-ontdekking moet beperkt blijven; de betrouwbaarheid van de steekproef moet daarop afgestemd worden.

Anderzijds kan bij een goed ingerichte en bemande administratieve organisatie verwacht worden dat door de werking van de interne controle het werkelijk niveau van de fouten, in de zin van onjuiste informatie, aanmerkelijk lager zal zijn dan hetgeen als aantasting van het beeld van de jaarrekening aangemerkt moet worden. De steekproef moet dan zo ingericht worden dat de kans om de populatie niet te accepteren gering is *indien* deze het verwachte foutenniveau zou hebben. Deze kans noemt men het onderscheidingsvermogen van de toets.

Het is mogelijk een steekproef te ontwerpen die (ongeveer) aan vooraf te stellen eisen van betrouwbaarheid en onderscheidingsvermogen voldoet.

Het vorenstaande impliceert een laatste punt van kritiek op het door Vermaas

beschreven televisieprogramma. Statistici zouden de bewindsman vermoedelijk hebben aangeraden in de eerste plaats het gerucht ( $1/3$  van de populatie fout) te ontcrachten en niet zijn eigen verwachting (4% fout) te bewijzen. De steekproef zou dan zo zijn ingericht dat de uitspraak over het onaanvaardbaar geachte percentage ( $33\frac{1}{3}\%$ ) met grote betrouwbaarheid gedaan kon worden, terwijl de kans op een ongunstige steekproefuitkomst indien de populatie niet meer dan het verwachte foutenpercentage (4%) zou bevatten, beperkt gehouden zou worden. Dit zou bijvoorbeeld bereikt kunnen worden met een steekproef van 30 stuks; indien minder dan 4 fouten aangetroffen worden, kan dan met bijna 99,7% betrouwbaarheid worden gesteld dat het gerucht niet op waarheid berust, terwijl de kans op meer dan 3 fouten indien de populatie slechts 4% fouten bevat, slechts 3,4% zou zijn.

Bij de populaties die in het kader van de controle van de jaarrekening aan een onderzoek moeten worden onderworpen, liggen het onaanvaardbare en het verwachte foutenpercentage beide meestal veel lager; de steekproeven moeten dan ook groter zijn.

Het is misschien interessant te vermelden welke kans de accountants A tot en met D zouden hebben gehad méér dan het voor ieder van hen aanvaardbare aantal fouten te ontdekken, indien de populatie belastingaangiften 1% onontdekte telfouten zou hebben bevat. Deze kansen zijn:

- voor A: 52,8%;
- voor B: 33,3%;
- voor C: 34,9%;
- voor D: 13,2%.

Uit het voorgaande moge blijken hoe een verantwoorde oplossing voor het door Vermaas gesignaleerde probleem kan worden verkregen. Het gaat echter niet aan een door een andere wetenschap, in casu de statistiek, ontwikkelde techniek te hanteren en dan een kernelement daarvan, het begrip betrouwbaarheid, naar eigen hand te zetten.

De levendige en indringende wijze waarop Vermaas een lastig vraagstuk aan de orde heeft gesteld verdient de dank van degenen die zich met de toepassing van de statistische steekproef in de accountantscontrole bezighouden.

#### Literatuurverwijzingen

1. J. H. Blokdijk, Steekproeven, in: Handboek Accountancy, Samsom, aflevering april 1976, blz. 3084.
2. Commissie Controle en Steekproeven van het NivRA, De statistische steekproef in de accountantscontrole (Interimrapport), De Accountant, juli/augustus 1974, blz. 487.
3. E. B. Tuitjer en J. M. Zuydervliet, De steekproef (Verslag NivRA-najaarsconferentie 1974), De Accountant, april 1975, blz. 491 en 492.
4. M. Vermaas, Uitbreiden van de steekproef, M.A.B., mei 1979, blz. 203 tot en met 215.