

M. Vermaas

# Vertrouwen in de uitgebreide steekproef

## 1 Inleiding

In de nummers 10 en 11/1986 van dit maandblad geeft Drs. J. W. Nool een uitvoerige beschouwing over statistische steekproeven bij accountantscontroles. Hij vond daartoe aanleiding in de discussie die in de jaren 1979-1982 in dit blad is geplaatst. Een discussie die - zoals Drs. Nool stelt - door mij werd aangezwengeld.

Met het betoog van Drs. Nool kan ik mij geheel verenigen, maar er is naar mijn mening nog wel iets aan toe te voegen. Ik zal er naar streven dat zodanig te doen dat het ook voor degenen die met de voorgaande discussie niet goed (meer) op de hoogte zijn, te volgen is.

## 2 De nieuwe vorm

De discussie ging om het op een bepaalde wijze uitbreiden van de steekproef. Volgens mij is dat toelaatbaar; volgens mijn opponenten niet. Ik verdedigde mijn standpunt door een andere definitie van het in de statistiek veelvuldig gebruikte begrip betrouwbaarheid te geven. Mijn tegenstanders vielen over mij heen met het verwijt dat ik een onbestaand kansbegrip zou hanteren. Ook Drs. Nool wrijft mij dat nog aan op blz. 10-442 van zijn betoog. Op blz. 11-497 wordt Van Heerden (een naam die geen toelichting behoeft voor degenen die zich interesseren voor steekproeven bij de accountantscontrole) letterlijk hetzelfde verweten. Laat ik thans volstaan met op te merken dat we blijkbaar bij het volgen van het spraakgebruik de kans(?) lopen een voor statistici onbestaand kansbegrip te hanteren.

Op blz. 10-442 zegt Drs. Nool: 'In een iets andere vorm gegoten blijkt het door Vermaas gedefinieerde begrip wel degelijk goed bruikbaar'.

Het gaat hierbij om het begrip betrouwbaarheid en de 'iets andere vorm' houdt in dat:

- we ons houden aan het jargon der statistici en dat we dus geen voor hen onbestaand kansbegrip hanteren;
- het door mij gedefinieerde begrip 'betrouwbaarheid' moet worden omgedoopt tot 'mate van vertrouwen'.

Drs. Nool legt uit dat het begrip betrouwbaarheid in de statistiek is gereserveerd voor de kans dat de gevolgde statistische methode een juist resultaat oplevert.

Mijn gedefinieerd begrip is volgens Drs. Nool een soort betrouwbaarheid

van het resultaat. Na de in het voorgaande genoemde vorm-verandering noemt hij het een mate van vertrouwen in een genomen beslissing welke aangeeft hoe sterk het resultaat van de steekproeftrekking naar goed- of afkeuring van de populatie wijst.

Ik wil Drs. Nool in dit alles volgen, maar ik kom dan tot de conclusie dat de bezwaren die werden ingebracht tegen het uitbreiden van de steekproef op de door mij bepleite wijze, komen te vervallen. Drs. Nool doet daar echter geen enkele uitspraak over; daarin ligt de reden van mijn reactie.

Het zou mij overigens niet verbazen indien er lezers zijn die opmerken dat de betrouwbaarheid van de steekproefmethode en de mate van vertrouwen in de op grond van het steekproefresultaat genomen beslissing toch gelijk aan elkaar zouden moeten zijn. Ik vind dat eigenlijk ook, maar we kunnen de statistici niet het recht ontzeggen begrippen op hun wijze te definiëren.

### **3 Het voorbeeld van Drs. Nool**

We knopen aan bij het door Drs. Nool op blz. 11-490 e.v. uitgewerkte voorbeeld van een goed-afkeurtoets.

De tevoren gekozen waarden zijn:

- 1 de onacceptabele foutenfractie = 0,06
- 2 de maximale kans op onterecht goedkeuren = 5%  
NB de minimale mate van vertrouwen in de goedkeurbeslissing is dan 95%
- 3 een waarde kleiner dan de onacceptabele foutenfractie, de ten hoogste verwachte foutenfractie = 0,03
- 4 de maximale kans op (onterecht) afkeuren indien de foutenfractie < 0,03 = 10%  
NB de minimale mate van vertrouwen in de afkeurbeslissing is dan 90%.

Met deze gekozen waarden komt Drs. Nool tot een steekproefomvang van 425 trekkingen. Bij 18 of meer gevonden fouten wordt afgekeurd; bij minder dan 18 gevonden fouten wordt goedgekeurd.

De betrouwbaarheid is 95% want een populatie met precies de onacceptabele foutenfractie van 0,06 heeft slechts 5% kans te worden goedgekeurd. Worden bij 425 trekkingen 17 fouten gevonden, dan volgt goedkeuring. De mate van vertrouwen in de goedkeurbeslissing is 95%. Volgens de Poissonverdeling is er namelijk bij een populatie met een foutenfractie van 0,06 en 425 trekkingen 95% kans op het vinden van meer dan 17 fouten. In dit geval is de mate van vertrouwen dus gelijk aan de betrouwbaarheid.

Maar stel dat er niet 17 maar 10 fouten in de 425 trekkingen worden gevonden. Dat is heel goed mogelijk bij de ten hoogste verwachte foutenfractie van 0,03. De betrouwbaarheid blijft dan 95%. De mate van vertrouwen in de goedkeurbeslissing is dan echter 99,96%. Volgens de Poissonverdeling is er namelijk bij een populatie met een foutenfractie van 0,06 en 425 trekkingen 99,96% kans op het vinden van meer dan 10 fouten. In dat geval is de mate van vertrouwen belangrijk groter dan minimaal geëist. Is er dan geen controletijd verspild? We zullen zien dat zulks inderdaad het geval kan zijn.

#### 4 Was de steekproefomvang te groot?

Indien Drs. Nool - onder handhaving van de tevoren gestelde waarden - de steekproefomvang op 105 in plaats van 425 trekkingen had bepaald, dan had er reeds goedgekeurd kunnen worden indien er 2 of minder fouten waren gevonden; bij 6 of meer fouten had er dan reeds afgekeurd kunnen worden.

De betrouwbaarheid zou weer 95% zijn geweest, want een populatie met precies de onacceptabele foutenfractie van 0,06 zou slechts 5% kans hebben gehad te worden goedgekeurd.

Bij 2 fouten in 105 trekkingen zou de mate van vertrouwen in de goedkeurbeslissing 95% zijn geweest. Volgens de Poissonverdeling is er namelijk bij een populatie met een foutenfractie van 0,06 en 105 trekkingen 95% kans op het vinden van meer dan 2 fouten.

Aan de tevoren gestelde waarden zou voldaan zijn en het controlewerk was tot een kwart gereduceerd. Bij de ten hoogste verwachte foutenfractie van 0,03 is dat een situatie die tot de reële mogelijkheden behoort.

Het weerwoord van Drs. Nool is wel te voorspellen: als er in de 105 trekkingen meer dan 2, maar minder dan 6 fouten worden gevonden, kan er geen beslissing worden genomen!

Hoe staan de zaken indien we in deze situatie de steekproef uitbreiden van 105 tot 425 trekkingen, waarbij we dan weer bij 18 of meer gevonden fouten afkeuren en bij minder dan 18 gevonden fouten goedkeuren?

#### 5 Uitbreiden van de steekproef

Bij de zojuist bedoelde wijze van uitbreiden van de steekproef daalt de betrouwbaarheid tot 92,1%, want een populatie met precies de onacceptabele foutenfractie van 0,06 heeft de volgende kans te worden goedgekeurd:

bij de eerste 105 trekkingen	5 %
bij de uitbreiding met 320 trekkingen	<u>2,9%</u>
totaal	<u>7,9%</u>

NB Voor de manier waarop zo'n acceptatiekans wordt berekend, moge ik verwijzen naar mijn artikel 'Uitbreiden van de steekproef' in dit maandblad, mei 1979, blz. 209/210.

Aan de door Drs. Nool tevoren gestelde 2e waarde wordt bij uitbreiden van de steekproef niet meer voldaan. De acceptatiekans van het 'grensgeval' (populatie met een foutenfractie van net 0,06) mag immers niet groter zijn dan 5%.

De tevoren gestelde minimale waarde voor de mate van vertrouwen in de goed- of afkeurbeslissing wordt daarentegen bij uitbreiden van de steekproef niet aangetast. Dat hangt samen met het volgende feit: als we een steekproef nemen uit een populatie met precies de onacceptabele foutenfractie van 0,06 dan is de kans dat we in 425 trekkingen meer dan 17 fouten vinden 95%; bovendien is de kans dat we in de eerste 105 trekkingen meer dan 2 fouten vinden ook 95%. Op grond hiervan kunnen we met een mate

van vertrouwen van 95% de populatie goedkeuren indien in 425 trekkingen slechts 17 fouten gevonden zijn; maar we kunnen - eveneens met een mate van vertrouwen van 95% - de populatie ook reeds goedkeuren indien in de eerste 105 trekkingen slechts 2 fouten gevonden zijn. Bij een aantal gevonden fouten van minder dan 17, respectievelijk 2 is de mate van vertrouwen groter dan 95%.

Ik blijf de aan het eind van paragraaf 4 bedoelde wijze van uitbreiden van de steekproef bepleiten en dat betekent dat mijn tevoren gestelde waarden iets afwijken van die van Drs. Nool. De door mij tevoren gestelde waarden luiden - om bij ons voorbeeld te blijven - als volgt:

- 1 de onacceptabele foutenfractie = 0,06
- 2 de minimale mate van vertrouwen in de goedkeurbeslissing = 95%
- 3 een waarde kleiner dan de onacceptabele foutenfractie, de ten hoogste verwachte foutenfractie = 0,03
- 4 de minimale mate van vertrouwen in de afkeurbeslissing = 90%.

Om de mogelijkheid open te houden de steekproef - zo nodig - uit te breiden, stel ik tevoren geen waarde voor de maximale acceptatiekans van de populatie met een onacceptabele foutenfractie. We zullen hierna nog zien dat die acceptatiekans niet relevant is indien we tot een goedkeurbeslissing komen.

Op grond van de door mij gestelde waarden komen we tot het volgende steekproefschema:

		<i>Goedkeuren</i> <i>bij</i>	<i>Afkeuren</i> <i>bij</i>
eerste	105 trekkingen	2 fouten	6 fouten
volgende	<u>320</u> trekkingen		
totaal	425 trekkingen	17 fouten	18 fouten

We nemen een steekproef en vinden in de eerste 105 trekkingen bijv. 3 fouten. We vervolgen dan de steekproef met nog 320 trekkingen. Vinden we er 9 fouten bij, dan zijn de feiten: 425 trekkingen met 12 fouten. We keuren de populatie goed. Er bestaat 99,76% vertrouwen dat de goedkeurbeslissing juist is, oftewel dat de foutenfractie in de populatie kleiner is dan 0,06. Bij een populatie met een foutenfractie van 0,06 is er namelijk volgens de Poissonverdeling 99,76% kans om in 425 trekkingen meer dan 12 fouten te vinden.

Vinden we in de eerste 105 trekkingen slechts 1 fout, dan kunnen we op dat moment reeds goedkeuren met een mate van vertrouwen in de goedkeurbeslissing van 98,66%. Immers bij een populatie met een foutenfractie van 0,06 is er volgens de Poissonverdeling 98,66% kans in de eerste 105 trekkingen meer dan 1 fout te vinden.

Zelfs als we het maximale aantal fouten vinden waarbij we nog goedkeuren, dat wil zeggen 2 fouten in de eerste 105 trekkingen of 17 fouten in het totaal van 425 trekkingen, bestaat er nog 95% vertrouwen dat de goedkeurbeslissing juist is, oftewel dat de foutenfractie kleiner is dan 0,06.

Is het dan te boud gesproken wanneer ik zeg dat bij een goedkeurbeslissing

de acceptatiekans van de populatie met een foutenfractie van net (iets meer dan) 0,06 niet relevant is?

In het voorgaande heb ik mij - korthedshalve - beperkt tot de situaties bij goedkeuren. Voor de situaties bij afkeuren geldt een analoog verhaal.

## 6 Een extreem geval

Wellicht zijn er die toch betwijfelen of bij uitbreiden van de steekproef de mate van vertrouwen in de goed- of afkeurbeslissing bepaald mag worden zoals ik dat gedemonstreerd heb.

Laten we daarom eens het geval bezien waarbij tevoren de volgende waarden worden gesteld:

- 1 de onacceptabele foutenfractie = 0,06
- 2 de minimale mate van vertrouwen in de goedkeurbeslissing = 80% (zo laag gesteld om er een sprekend voorbeeld van te maken)
- 3 een waarde kleiner dan de onacceptabele foutenfractie, de ten hoogste verwachte foutenfractie = 0,03
- 4 de minimale mate van vertrouwen in de afkeurbeslissing = 90%.

We kunnen bij deze waarden een steekproef nemen van 27 trekkingen; vinden we geen fouten dan kan goedgekeurd worden. De mate van vertrouwen in de goedkeurbeslissing is 80%. Volgens de Poissonverdeling is er namelijk bij een populatie met een foutenfractie van 0,06 en 27 trekkingen 80% kans op het vinden van meer dan 0 fouten.

Vinden we echter 1 fout in de 27 trekkingen dan laat de gestelde minimale mate van vertrouwen niet toe dat we goedkeuren. Als we nu de steekproef (extreem) uitbreiden met nog 973 trekkingen en we vinden er 36 fouten bij dan zijn de feiten: 1000 trekkingen en 37 gevonden fouten. Zonder becijfering is nu wel voor een ieder duidelijk dat er een grote mate van vertrouwen bestaat dat de foutenfractie in de populatie kleiner is dan 0,06 en dat een goedkeurbeslissing juist is.

Volgens mij is de mate van vertrouwen in dit geval 99,90% omdat er bij een populatie met een foutenfractie van 0,06 volgens de Poissonverdeling 99,90% kans is om in 1000 trekkingen meer dan 37 fouten te vinden.

Wie de juistheid hiervan ontkent, nodig ik uit kenbaar te maken hoe hij of zij de mate van vertrouwen in dit geval dan wel becijfert.

## 7 Gebruik van de tabel

Bij het voorbeeld in paragraaf 5 heb ik de steekproef slechts eenmaal uitgebreid, namelijk van 105 trekkingen tot 425 trekkingen. Het is ook mogelijk de steekproef - zo nodig - in meer dan 2 fasen uit te voeren. Bovendien kunnen we beginnen met een minimale steekproefomvang waarbij goedgekeurd kan worden indien geen enkele fout gevonden wordt.

We kunnen daarbij gebruik maken van de in accountantskringen tamelijk bekende tabel die ook weer bij de artikelen van Drs. Nool werd afgedrukt. De tabel is gebaseerd op de Poissonverdeling, die benaderende uitkomsten

geeft. Zij is goed bruikbaar zolang de steekproefomvang niet uitgaat boven ca. 10% van de populatieomvang. Bij een populatie met een foutenfractie van 0,06 en een omvang van bijv. 100.000 eenheden is de kans om meer dan 0 fouten te vinden in een steekproef van 27 trekkingen volgens de Poisson-verdeling 80,21%. Exact berekend - we gebruiken dan de hypergeometrische verdeling - is die kans 81,18%. Een afwijking die nog wel te accepteren is.

Om het gebruik van de tabel te demonstreren gaan we weer uit van de 4 tevoren gestelde waarden zoals vermeld in paragraaf 5.

In kolom 0,05 (complement van 0,95) zien we onder 'upper' het getal 3,00 op regel 0. Delen we het getal 3,00 door 0,06 (de onacceptabele foutenfractie) dan is de uitkomst 50. Op deze wijze lezen we uit de tabel af dat bij een foutenfractie van 0,06 en 50 trekkingen er 95% kans is om meer dan 0 fouten te vinden. We kunnen dus goedkeuren indien we niet meer dan 0 fouten vinden in 50 trekkingen.

Om de beslissende punten voor afkeuren te vinden, kijken we in de kolom 0,10 (complement van 0,90) onder 'lower'. De getallen op de regels 1, 2, 3 en 4 delen we door 0,03 (de hoogstens verwachte foutenfractie) en we vinden dan dat we kunnen afkeuren bij 1 fout in de eerste 3 trekkingen, bij 2 fouten in een totaal van 17 trekkingen, bij 3 fouten in een totaal van 36 trekkingen, en bij 4 fouten in een totaal van 58 trekkingen. In feite lezen we uit de tabel af dat bij een foutenfractie van 0,03 er 90% kans is niet meer dan 0 fouten te vinden in de eerste 3 trekkingen; er 90% kans is niet meer dan 1 fout te vinden in een totaal van 17 trekkingen; etc.

Blijkt de 10e trekking een fout op te leveren en blijft het daarbij tot de 50e trekking gedaan is, dan staan we voor de noodzaak de steekproef uit te breiden.

Uit de tabel valt af te lezen dat de steekproef minstens tot 79 trekkingen moet worden uitgebreid om alsnog te kunnen goedkeuren ( $4,74 / 0,06 = 79$ ). Worden er dan 1 of meer fouten bij gevonden, dan ... etc., etc.: tot we eventueel toekomen aan de steekproefomvang van 425 trekkingen waarbij in ieder geval wordt goedgekeurd of afgekeurd.

## **8 'Geen feiten verdonkeremanen'**

Om misverstand te voorkomen wijs ik erop dat we ons bij het nemen van een beslissing op grond van een steekproefresultaat moeten baseren op alle in de steekproef geconstateerde feiten. Bij uitbreiden van de steekproef moeten we daarom de aantallen trekkingen en de aantallen gevonden fouten cumuleren. We mogen uiteraard geen feiten verdonkeremanen door onwelgevallige steekproefresultaten in de prullenmand te deponeren en opnieuw te beginnen. Dan ontlopen we (kwade) kansen en gaat de kansrekening, waarop het steekproefstelsel is gebaseerd, niet meer op.

## **9 Slotbeschouwing**

Indien we bij een steekproefcontrole de steekproefomvang bepalen uitgaande van een tevoren als eis gestelde mate van betrouwbaarheid, mogen

we - bij het vinden van (net) te veel fouten om te kunnen goedkeuren - de steekproef niet uitbreiden om te bezien of de populatie eventueel alsnog kan worden goedgekeurd. Dit berust op de bij statistici gangbare definitie van het begrip betrouwbaarheid.

In dit maandblad (mei 1979) heb ik indertijd het begrip betrouwbaarheid op een andere wijze geformuleerd en - daarvan uitgaande - bedoelde uitbreiding van de steekproef toch bepleit. Dat werd echter bestreden met de opmerking dat het niet aangaat 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 (J. H. Blokdijk, *MAB* november 1979, blz. 541).

Nu stelt Drs. Nool, deskundige op het gebied van de statistiek, dat het door mij gedefinieerde begrip wel degelijk goed bruikbaar is. Het moet echter in een andere vorm gegoten worden en hij noemt het dan de mate van vertrouwen in de goed- of afkeurbeslissing.

In aansluiting daarop heb ik in het voorgaande laten zien dat, indien we de steekproefomvang bepalen uitgaande van een tevoren geëiste mate van vertrouwen, het - zo nodig - uitbreiden van de steekproef wel verantwoord is. Het voordeel daarvan is dat we met een minimale steekproefomvang kunnen beginnen, waarbij de mogelijkheid bestaat dat we daarmee kunnen volstaan.

## Literatuur

- 1 Vermaas, M., Uitbreiden van de steekproef, *MAB* mei 1979, blz. 203-215.
- 2 Blokdijk, J. H., Uitbreiding van de steekproef, *MAB* november 1979, blz. 537-541.
- 3 Vermaas, M., Steekproeven bij controle op ernstige fouten, *MAB* november 1979, blz. 525-536.
- 4 Vermaas, M., Van steekproef tot steekspel, *MAB* januari 1980, blz. 36-40.
- 5 Blokdijk, J. H., Terug naar de steekproef, *MAB* maart 1980, blz. 127-131.
- 6 Vermaas, M., De steekproef nogmaals uitgebreid, *MAB* juni 1980, blz. 255-256.
- 7 Muilwijk, J., Kansrekening, statistiek en accountantscontrole, *MAB* december 1981, blz. 628-669.
- 8 Kriens, J. en Dekkers, A. C., Misverstanden over het gebruik van steekproeven in de accountantscontrole, *MAB* april 1982, blz. 177-192.
- 9 Nool, J. W., Betrouwbaarheid van en vertrouwen in steekproefcontroles, *MAB* november 1986, blz. 441-458.
- 10 Nool, J. W., Toepassing van goed-afkeursoetsen, *MAB* december 1986, blz. 487-503.

## **Naschrift**

*door Prof. J. H. Blokdijsk*

Het is aardig weer eens aan een al wat weggezakte discussie te worden herinnerd, zoals de heer Vermaas in vorenstaand artikel doet. Met prof. drs. J. Muilwijk behoorde ik namelijk tot de opponenten waaraan Vermaas in het begin van paragraaf 2 refereert.

In de context van het cijfervoorbeeld van Nool en Vermaas zou ik in de praktijk stellig de door Vermaas gehuldigde gedragslijn volgen indien in totaal bijv. 9 of 13 fouten zouden worden aangetroffen. Dan zou immers de 'mate van vertrouwen', zoals door Nool gedefinieerd, niet of nauwelijks afwijken van de geëiste betrouwbaarheid zoals door de statistici gedefinieerd.

De hedendaagse (externe) accountantspraktijk laat echter veelal niet de ruimte voor deze gedragslijn. Men start uit kostenoverwegingen met een procedure waarbij slechts bij 0 fouten wordt goedgekeurd, en probeert bij het aantreffen van 1 fout zijn problemen met zo min mogelijk extra werk op te lossen. Dat betekent toepassing van een (onjuiste) procedure waarbij men bij aantreffen van 1 fout nog kan goedkeuren. Dan komt men terecht in de situatie die Vermaas aan het begin van paragraaf 5 beschrijft, namelijk verlaging van de betrouwbaarheid. Aangezien er tussen 0 en 1 fout geen ruimte zit, helpt de berekening van de 'mate van vertrouwen' dan ook niet. Toch ben ik met Vermaas van mening dat Nools bijdrage in sommige gevallen van nut kan zijn.