

VAN STEEKPROEF TOT STEEKSEL

(antwoord aan collega Blokdijk)

*door M. Vermaas***1. 'De handschoen opgenomen'**

Collega Blokdijk heeft - zoals te verwachten was - gereageerd¹⁾ op mijn in het mei-nummer 1979 verschenen artikel „Uitbreiden van de steekproef”.

Blokdijk heeft waardering voor de „verpakking” van mijn artikel, maar niet voor de inhoud. Daar staat tegenover dat de reactie van Blokdijk mij geen enkele aanleiding geeft mijn visie op het aan de orde gestelde vraagstuk te wijzigen. Het wordt dus een tornooi, ook wel „steekspel” geheten. De bedoeling van een tornooi is, dat vooral de toeschouwers er plezier aan beleven. Voor de lezers zal ik daarom trachten duidelijk te maken waar het om gaat.

2. De kern van de zaak

De kern van de zaak is het begrip „betrouwbaarheid”, n.l. de betrouwbaarheid van een uitspraak welke op grond van een steekproef wordt gedaan omtrent de kwaliteit van een populatie.

Het complement van betrouwbaarheid is - zoals Blokdijk ook zegt - het begrip „risico”. Indien er bijv. 5% risico wordt gelopen een foute uitspraak te doen, dan is de uitspraak voor 95% betrouwbaar. Het risico wordt bepaald door kansrekening.

We zullen ook het begrip „(fout) grensgeval” tegenkomen. Indien de maximaal toelaatbare foutenfractie voor een populatie bijv. op 4% is gesteld, dan noemen we de populatie waarin (net iets meer dan) 4% fouten schuilt het „(foute) grensgeval”. Terloops merken we op dat een accountant die er bij toeval achter komt een „fout grensgeval” - ten onrechte - te hebben goedgekeurd, daar als regel geen slapeloze nacht van zal hebben.

In de paragraaf „kans en betrouwbaarheid” van Blokdijk's reactie blijkt, dat hij vasthoudt aan een begrip betrouwbaarheid dat gebaseerd is op de acceptatiekans van het „grensgeval”. Ik zal dat toelichten met het voorbeeld dat de collega's Tuitjer en Zuydervliet op blz. 492 van hun verslag van de NIVRA-najaarsconferentie 1974 gebruiken:

- maximaal toelaatbare foutenfractie: 1%
- aanvaardbaar risico: 1%, d.w.z. een vereiste betrouwbaarheid van 99%
- steekproefomvang: 1005 trekkingen
- aantal fouten in de steekproef waarbij de populatie kan worden goedgekeurd: maximaal 3.

Waarom is nu, indien de populatie wordt goedgekeurd, volgens Blokdijk c.s. de betrouwbaarheid van de op grond van de steekproef gedane uitspraak 99%? Dat wordt als volgt beredeneerd.

¹⁾ Zie MAB november 1979.

Het „(foute) grensgeval”, d.w.z. de populatie waarin (net iets meer dan) 1% fouten schuilt, heeft een kans van (net iets minder dan) 1% om te worden geaccepteerd. Het risico is dus (net iets minder dan) 1% en de betrouwbaarheid (net iets meer dan) 99%.

Als men bovenbedoelde inhoud aan het begrip betrouwbaarheid geeft, dan behoeft het geen verbazing te wekken dat hoogleraren in de statistiek en in de beslistkunde fulmineren tegen „uitbreiden van de steekproef”. Zouden in het weergegeven voorbeeld 4 fouten worden gevonden, dan zou bij uitbreiden van de steekproef een herkansing plaats vinden; de acceptatiekans van het „grensgeval” zou groter dan 1% worden en daarmee de betrouwbaarheid kleiner dan de vereiste 99%.

Het door Blokdijk c.s. aangehouden begrip betrouwbaarheid is echter een ondoelmatig begrip. Bij de in het voorgaande voorbeeld gehanteerde cijfers valt dat nog niet zo dadelijk op; daarom zullen we het voorbeeld van andere - extreme - cijfers voorzien.

3. Een extreem geval

Door extreme cijfers te nemen verandert uiteraard het principe niet, maar de waarheid treedt beter aan het licht. Het voorbeeld wordt nu als volgt:

- maximaal toelaatbare foutenfractie: 1%
- aanvaardbaar risico: 96,8%, d.w.z. een vereiste betrouwbaarheid van 3,2%
- steekproefomvang: 500 trekkingen
- aantal fouten in de steekproef waarbij de populatie kan worden goedgekeurd: maximaal 9.

Stel dat 0 fouten worden gevonden, zodat de populatie wordt goedgekeurd. Blokdijk c.s. stellen dan dat de betrouwbaarheid van de op grond van de steekproef gedane uitspraak 3,2% is. De acceptatiekans van het „grensgeval” is nl. 96,8%.

Het is nu echter wel zonneklaar dat de betrouwbaarheid van de op grond van de steekproef gedane uitspraak niet moet worden afgeleid uit de acceptatiekans van het „grensgeval”. Zonder berekening is al duidelijk, dat bij een steekproef van 500 trekkingen waarbij 0 fouten worden gevonden, met een zeer grote betrouwbaarheid de uitspraak kan worden gedaan dat de maximaal toelaatbare foutenfractie (in dit voorbeeld 1%) niet is overschreden en dat derhalve de populatie wordt goedgekeurd.

De uitspraak van Blokdijk c.s., welke wordt gegrond op de omstandigheid dat, als de populatie 1% fouten bevat, zij 96,8% kans heeft gekregen te worden geaccepteerd, is wel juist, maar zinloos, nu het praktisch uitgesloten moet worden geacht dat er 1% fouten in de populatie zit.

Met dit extreme geval heb ik duidelijk laten uitkomen, dat de acceptatiekans van het „grensgeval” en de betrouwbaarheid van de uitspraak over de kwaliteit van een populatie op grond van een steekproef, verschillende zaken zijn, die men beter niet met elkaar in verband kan brengen.

4. Hoe bepalen we de betrouwbaarheid wel?

Bij de door mij bepleite werkwijze wordt de betrouwbaarheid van de op grond van de steekproef gedane uitspraak omtrent de kwaliteit van de onderzochte po-

pulatie niet afgeleid uit de acceptatiekans van het „grensgeval”, maar zij wordt berekend uit de steekproefomvang en *het aantal daadwerkelijk in die steekproef geconstateerde fouten*.

We zullen de rekenwijze eerst demonstreren voor het in paragraaf 3 vermelde extreme geval. Dus 500 trekkingen en 0 fouten gevonden. Indien de maximaal toelaatbare foutenfractie van 1% overschreden zou zijn, dan was er minder dan 0,657% kans niet meer dan 0 fouten te vinden. We keren de zaak nu om en zeggen: omdat er niet meer dan 0 fouten gevonden zijn, is er slechts minder dan 0,657% kans dat de maximaal toelaatbare foutenfractie is overschreden. De betrouwbaarheid is dus 99,343% in plaats van - volgens het door Blokdijk c.s. vastgehouden begrip - 3,2%! Maar dit was ook een extreem geval.

We nemen nu het voorbeeld van Tuitjer en Zuydervliet, door mij vermeld in paragraaf 2. Stel dat 0 fouten werden gevonden. De situatie is dan 1005 trekkingen en 0 fouten. Indien de maximaal toelaatbare foutenfractie van 1% overschreden zou zijn, dan was er minder dan 0,0041% kans niet meer dan 0 fouten te vinden. We keren de zaak weer om en zeggen: omdat er niet meer dan 0 fouten gevonden zijn, is er slechts minder dan 0,0041% kans dat de maximaal toelaatbare foutenfractie is overschreden. De betrouwbaarheid is dus 99,9959% in plaats van - volgens Blokdijk c.s. - 99%. We zien bij een „gewoon geval” hetzelfde verschijnsel als bij een „extreem geval”; het loopt alleen wat minder in de gaten.

In het voorbeeld van Tuitjer en Zuydervliet hadden we ook 3 fouten kunnen vinden. Dan was de populatie ook nog goedgekeurd en - volgens Blokdijk c.s. - nog met dezelfde betrouwbaarheid van 99%.

Bij de door mij voorgestane werkwijze zou ik in het gestelde geval niet met een steekproef van 1005 trekkingen beginnen. Ik zou het volgende steekproefschema volgen:

- | | |
|---------------|-------------------------------|
| 1e steekproef | 460 trekkingen: |
| | 0 fout, populatie goedkeuren |
| | 1 fout, steekproef uitbreiden |
| | etc. ²⁾ |
| 2e steekproef | 204 trekkingen: |
| in totaal | 1 fout, populatie goedkeuren |
| in totaal | 2 fout, steekproef uitbreiden |
| | etc. ²⁾ |
| 3e steekproef | 177 trekkingen: |
| in totaal | 2 fout, populatie goedkeuren |
| in totaal | 3 fout, steekproef uitbreiden |
| | etc. ²⁾ |
| 4e steekproef | 164 trekkingen: |
| in totaal | 3 fout, populatie goedkeuren |
| in totaal | 4 fout, steekproef uitbreiden |
| | etc. ²⁾ |

²⁾ dat ik niet alsmaar door blijf gaan, leze men in mijn artikel in het mei nummer 1979 van dit maandblad.

Zou ik in de eerste 460 trekkingen 0 fouten aantreffen, dan was ik klaar met mijn steekproef (en Blokdijk nog niet half!).

Maar we nemen het geval dat ik door het vinden van fouten aan de 4e steekproef toekom en dan de populatie kan goedkeuren op grond van het resultaat $460 + 204 + 177 + 164 = 1005$ trekkingen en 3 fouten gevonden.

Volgens de door Blokdijk c.s. gehuldigde opvatting is nu de betrouwbaarheid van de uitspraak geen 99%, maar slechts 97,6%, want de acceptatiekans van het „grensgeval” (een populatie met 1% fouten) is door het „herkansen” gestegen van 1% tot 2,4%.

Ik ga voor de berekening van de betrouwbaarheid weer uit van het *feit* dat bij in totaal 1005 trekkingen slechts 3 fouten werden gevonden. De redenering is nu bekend: indien de maximaal toelaatbare foutenfractie van 1% overschreden zou zijn, dan was er minder dan 1% kans niet meer dan 3 fouten te vinden. We keren de zaak om en zeggen: omdat er niet meer dan 3 fouten gevonden zijn, is er slechts minder dan 1% kans dat de maximaal toelaatbare foutenfractie is overschreden. Nu het zò onwaarschijnlijk is dat we met een „fout grensgeval” te doen hebben, laat de acceptatiekans daarvan ons koud!

De betrouwbaarheid is volgens mij dus op het vereiste peil van 99% en derhalve keur ik de populatie goed. Blokdijk c.s. zouden dat – in verband met „hun betrouwbaarheid” van 97,6% – niet kunnen doen. Daarom bepalen zij de steekproefomvang direct op 1005 trekkingen; daarmee beroven zij zichzelf van het mogelijke voordeel, dat met een kleinere steekproef van 460, 664 of 841 trekkingen kan worden volstaan.

5. Naar eigen hand gezet?

In de laatste paragraaf van zijn reactie doet Blokdijk een vervaarlijke uitval. Hij werpt mij voor de voeten 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.

Ik pareer die uitval als volgt:

- De in paragraaf 4 toegepaste omkeermethode (zie de bij herhaling gebruikte woorden „we keren de zaak om”) is de kansrekenaars niet vreemd.
- De door mij bepleite werkwijze heb ik niet zelf bedacht; het is mij bekend dat Prof. Dr. P. de Wolff, al jaren geleden, die werkwijze aanbeval. Deze hoogleraar is toch statisticus. Ik ben in het krijt getreden om de bewering van Blokdijk c.s., dat bedoelde werkwijze onjuist zou zijn, te weerleggen, en om aan te tonen dat de thans door Blokdijk c.s. aangehangen opvatting verwarrend is en tot een ondoelmatige werkwijze leidt.

Overigens is het stuk kansrekening dat we voor onze steekproefproblemen nodig hebben, voor accountants niet zo onbegrijpelijk, dat we met de (eigen) hand voor ogen, aan de leiband van statistici moeten lopen. Als een statisticus de door mij bepleite werkwijze zou willen bestrijden, dan zal hij mij het onlogische in wat ik *doe*, moeten aanwijzen. Dat lijkt mij voor ons beroep ook beter, want als er dan nog eens een bewindsman is die een controle-opdracht heeft te vergeven waarbij een steekproefsgewijze aanpak voor de hand ligt, dan zal hij zich toch tot de accountants kunnen wenden, in plaats van · zoals Blokdijk hem adviseert · tot de statistici.

6. Een interessant geval

Eveneens in de laatste paragraaf van zijn reactie zegt Blokdijk dat het misschien interessant is te vermelden dat de accountants A tot en met D uit het verhaal van mijn „mei-artikel”, resp. 53, 33, 35 en 13% (afgerond) kans zouden hebben gehad méér dan het voor ieder van hen aanvaardbare aantal fouten te ontdekken, indien de populatie 1% fouten zou hebben bevat.

Dat is inderdaad hoogst interessant. Hier reikt Blokdijk mij als het ware zelf de lans aan, waarmee ik zal trachten hem uit het zadel te werpen.

Bedacht moet worden dat de 4 accountants allen te werk gingen volgens de door Blokdijk c.s. gehuldigde opvatting en dat de maximaal toelaatbare foutenfractie op 4% was gesteld. Dat er voor accountants die zich houden aan de „opvatting Blokdijk c.s.”, bij een zò goede populatie (1% fouten bij 4% toelaatbaar) toch zò veel kans is dat zij van goedkeuring moeten afzien, is nu juist een van mijn grote bezwaren tegen die „opvatting Blokdijk c.s.”.

De 4 accountants beëindigen de steekproef - in verwarring gebracht door de acceptatiekans van het „grensgeval” - veel te snel. Daarom is mijn parool: „steekproef uitbreiden!” Laten we eerst weer eens een extreem voorbeeld nemen. Het zal voor een ieder wel duidelijk zijn, dat bij uitbreiding van de steekproef tot b.v. 10.000 trekkingen (!) uit deze populatie met 1% fouten, het zonneklaar wordt dat er niet meer dan 4% fouten inzitten (en dat behoeft de accountants slechts, met 95% betrouwbaarheid, te verklaren).

De uitbreiding behoeft echter helemaal niet zo extreem groot te zijn. We zullen de situatie schetsen bij uitbreiding tot in totaal 500 trekkingen. Er is dan bij deze populatie met 1% fouten 96,8% kans dat er minder dan 10 fouten worden gevonden. Maar laten de 4 accountants een gelukkige greep doen en toch wel 10 fouten vinden. Dan kunnen zij nog met een goed geweten verklaren, dat de populatie niet meer dan 4% fouten bevat; aan de eis van 95% betrouwbaarheid wordt ruim voldaan; zij is zelfs - op mijn manier berekend - 99%. Volgens Blokdijk c.s. zou de vereiste betrouwbaarheid van 95% nooit meer kunnen worden bereikt, zelfs niet bij 10.000 of meer trekkingen!

7. Een wens

Ik zou wensen dat Blokdijk, en ook de geïnteresseerde lezer, tegen de achtergrond van de hier gegeven beschouwingen, mijn artikel „Uitbreiden van de steekproef” in het mei-nummer 1979 nog eens zouden willen doornemen, alsmede mijn inmiddels in het M.A.B. (november-nummer 1979) geplaatste artikel „Steekproeven bij controles op ernstige fouten”. Men zal dan zien dat ik echt foute populaties beslist niet *altijd* goedkeur, zoals Blokdijk beweert, maar dat ik een populatie eerst goedkeur, indien ik slechts een zo gering aantal fouten in mijn totale steekproef heb aangetroffen, dat het hoogst onwaarschijnlijk is dat we te doen hebben met een populatie waarin de maximaal toelaatbare foutenfractie is overschreden. Op dat moment is de acceptatiekans van het - eveneens hoogst onwaarschijnlijke - „grensgeval” niet interessant meer. Bovendien zal men zien dat echt goede populaties bij mij meer kans krijgen te worden geaccepteerd dan bij Blokdijk c.s. Van Blokdijk hoop ik, dat hij dan zal toegeven dat mijn „verhaal” tòch geen „sprookje” was.