Drs. A.C. Dekkers en Prof. Drs. J. Kriens

Misverstanden over het gebruik van steekproeven in de accountantscontrole

1 Inleiding

Ons in het voorjaar 1979 verschenen boek 'Steekproeven in de accountantscontrole' (6) heeft commentaar opgeroepen in recensies, terwijl het tevens onderwerp van bespreking is geweest in recente MAB-discussies.

Uit de boekbesprekingen en uit de discussies, gevoerd over de toepassingsmogelijkheden van statistische methoden en steekproeftypen, komen verschillen in opvatting naar voren, die deels aan onduidelijkheden in de terminologie en deels aan misverstanden te wijten zijn.
Teneinde een verdere verwarring omtrent een aantal fundamentele begrippen uit de steekproefmaterie en de toepassing daarvan in de accountantscontrole te voorkomen, respectievelijk op te helderen, willen wij in dit artikel ons standpunt hierover nader verduidelijken en verder uitwerken. Hierbij zal, voor zover nodig, op enkele kritische geluiden worden ingegaan.

De belangrijkste misverstanden hebben betrekking op de interpretatie van:
- het begrip 'ernstige fout';
- de toetsingsmethode;
- het begrip 'tolerantie';
- de mogelijkheid tot 'uitbreiden van de steekproef'.

Een centrale rol in de discussie speelt het begrip 'ernstige fout' waarop in paragraaf 2 uitvoerig wordt teruggekomen.

In dit artikel worden achtereenvolgens in kort bestek behandeld:
par. 2 foutdefinities;
par. 3 toetsen of schatten;
par. 4 tolerantie en risico;
par. 5 enkel- en meervoudige steekproeven of 'uitbreiden van de steekproef?'.
De bij paragraaf 5 behorende statistische bewijsvoering is in een appendix opgenomen.

2 Foutdefinities

De accountant die tot een goedkeuring van een verantwoording wil komen, zal zijn werkprogramma zodanig inrichten, dat hij op grond van zijn onder-

zoek kan vaststellen, dat de verantwoording voldoet aan de eraan te stellen eisen (zie GBR-artikel 13 lid 1).
De verklaring die de accountant omtrent de uitkomst van zijn controlearbeid bij een verantwoording afgeeft, garandeert geen volledige exactheid, doch geeft binnen redelijke nauwkeurigheidsgrenzen aan, dat de gecontroleerde verantwoording aanvaardbaar is. De accountant dient bij zijn onderzoek aan te geven welke fouten hij in het kader van de gestelde nauwkeurigheidstolerantie nog acceptabel vindt; het meest wezenlijke criterium voor het onderscheiden van fouten in een verantwoording is dan ook gelegen in de mate waarin zij de aanvaardbaarheid van een verantwoording beïnvloeden.

Fouten in een verantwoording, die de accountant in zijn attestfunctie onaanvaardbaar vindt, zal hij bij zijn controle beslist willen ontdekken, omdat het niet ontdekken van dergelijke fouten tot een ondeugdelijke oordeelvorming over de verantwoording zal leiden.
Daarnaast kunnen fouten in de verantwoording optreden, die op zichzelf niet onaanvaardbaar zijn, doch waarvan de accountant in het kader van zijn attestfunctie of van de daaruit voortvloeiende adviesfunctie toch kennis wil nemen.
Om het verschil tussen beide typen fouten goed te kunnen onderscheiden hebben wij de begrippen 'ernstige fouten' en 'niet-ernstige fouten' ingevoerd.
De term 'ernstige fouten' is als een verzamelterm te beschouwen van fouten die de getrouwheid van een verantwoording aantasten. Het is een *theoretisch* begrip, dat in *elke concrete situatie* door de accountant op basis van zijn deskundigheid moet worden ingevuld (zie GBR - artikel 11). Hierbij dient hij de beroepsregels in acht te nemen, doch deze laten ruimte voor een persoonlijk waarde-oordeel over de ernst van de fouten, aangezien het accountantsberoep hiervoor geen algemene richtlijnen heeft ontwikkeld.

Tuitjer (9) noemt twee foutenonderscheidingen die bij de controle van een verantwoording kunnen worden gemaakt, namelijk die in:
- materiële en formele fouten;
- fouten met en zonder directe financiële gevolgen;

waarbij zijn voorkeur uitgaat naar laatstgenoemde onderscheiding.
Wij menen dat de onderscheiding in ernstige en niet-ernstige fouten een beter criterium voor de accountantscontrole is omdat er een rechtstreekse relatie wordt gelegd met het al dan niet aanvaardbaar zijn van de verantwoording. In paragraaf 3 zal bovendien blijken dat het onderscheid in ernstige en niet-ernstige fouten een goed hulpmiddel is bij de keuze van de toe te passen statistische methode.
Bij onze foutenonderscheiding kunnen materiële fouten die de getrouwheid van de te onderzoeken verantwoording niet aantasten, als niet-ernstige fouten worden beschouwd en zijn formele fouten die wèl invloed op de getrouwheid van de verantwoording uitoefenen, als ernstige fouten aan te merken.
Zo is het denkbaar dat een materiële fout niet als ernstige fout wordt

gerekend, omdat de financiële consequentie in een bepaalde situatie niet groter is dan bijvoorbeeld *f* 500 of *f* 1.000.

Er kunnen zich ook controlesituaties voordoen, waarin de materiële juistheid van bepaalde posten in zeer sterke mate bepaald wordt door het formele aspect, bijvoorbeeld door de aanwezigheid van een paraaf op bepaalde documenten van daartoe door de organisatie van de gecontroleerde verantwoordelijk gestelde functionarissen.
Het ontbreken van zo'n paraaf kan een dusdanige inbreuk op de organisatie van het bedrijf betekenen, dat dit door de accountant als ernstige fout wordt beschouwd.
Samengevat kan worden gesteld dat het invullen van de begrippen 'ernstige fouten' en 'niet-ernstige fouten', is gebaseerd op een weging van de mogelijke fouten naar de mate waarin zij wel of geen invloed op de getrouwheid van de verantwoording uitoefenen. De specifieke controle-omstandigheden en het deskundige inzicht van de accountant spelen hierbij een belangrijke rol.

In de beschreven steekproeftoepassingen, die ook door Vermaas zijn aangehaald, (11) blz. 528/529, hebben wij een aantal materiële en formele fouten genoemd die in bepaalde controlesituaties ernstige fouten kunnen zijn. Dit betekent uiteraard niet dat deze fouten onder alle omstandigheden ernstige fouten zullen zijn. Indien Vermaas een dubbele betaling van een kostennota, een decimaalfout in een factuur etc., die niet door de interne controle worden ontdekt, in een bepaalde controlesituatie niet ernstig vindt, dan heeft hij die vrijheid binnen onze methode.
Wanneer de accountant bij het ontdekken van een dubbele betaling van één kostennota niet reeds tot afkeuring van de te controleren verantwoording wil overgaan en bereid is (en het ook controletechnisch verantwoord acht) een beperkt aantal dubbele betalingen in de steekproef te accepteren, kan hij dit bewerkstelligen door een ruimere weging in de foutdefinitie (bijvoorbeeld 4 dubbele betalingen = één ernstige fout) toe te passen. Deze mogelijkheid van weging is vermoedelijk ook Beek (1) blz. 141, ontgaan.

3 Toetsen of schatten

3.1 Beknopte samenvatting van de statistische technieken toetsen en schatten

Bij een steekproefonderzoek kan men een uitspraak kiezen, die gebaseerd is op een toetsingstechniek of op een schattingstechniek. Na korte beschrijvingen van beide technieken zal in paragraaf 3.2 nader op de relatie tussen deze technieken en de accountantscontrole worden ingegaan.
Er doen zich tal van situaties voor waarin men na wil gaan of een gedane bewering of een ontstane verwachting juist is of niet. Onderzoekt men dit met behulp van een aselecte steekproef, dan spreekt men in de statistiek van het *toetsen van een hypothese.* De te onderzoeken veronderstelling wordt de nulhypothese (= H_0) genoemd. Deze kan bijvoorbeeld luiden: in een debiteurenadministratie bedraagt de fractie posten (= p), die minder

dan een maand oud is 0,6 (of 60%). Treft men in de steekproef een fractie posten (= f), minder dan een maand oud, aan, die veel van 0,6 afwijkt, dan zal men de nulhypothese verwerpen. Men kan bijvoorbeeld besluiten de nulhypothese te verwerpen wanneer f minder is dan 0,5 of meer dan 0,7. De waarden waarbij men verwerpt vormen te zamen het zogenaamde kritieke gebied van de toets. In figuur 3.1 is dit schematisch aangegeven, met arcering van het kritieke gebied.

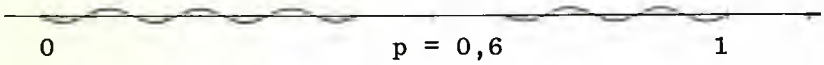


Fig. 3.1 Nulhypothese en kritiek gebied bij de toets $H_0 : p = 0,6$

Nu doet zich de complicatie voor dat ook wanneer H_0 juist is, men toch een steekproeffractie kan vinden die in het kritieke gebied ligt. De nulhypothese wordt dan ten onrechte verworpen; men noemt dit een *fout van de eerste soort*.
De kans op zo'n fout kan uitgerekend worden wanneer de steekproefomvang en de grenzen van het kritieke gebied gekozen zijn; deze kans is de onbetrouwbaarheid (= α) van de toets. Meer voor de hand ligt het echter reeds bij voorbaat een maximum aan deze onbetrouwbaarheid op te leggen, de zogenaamde onbetrouwbaarheidsdrempel (= α_0). Kiest men in het gegeven voorbeeld voor α_0 de waarde 0,05 (of 5%) en neemt men een steekproef van 100 posten, dan kan men uitrekenen dat de bovengrens van het linkerdeel van het kritieke gebied 0,504 bedraagt en de ondergrens van het rechter deel 0,696.
Wordt de nulhypothese verworpen, dan doet men dit ten gunste van de alternatieve hypothese H_1 die in ons voorbeeld luidt: p is ongelijk aan 0,6, kortweg $p \neq 0,6$.

Behalve een fout van de eerste soort kan men ook een *fout van de tweede soort* maken. Dit gebeurt wanneer H_0 niet juist is, terwijl men toch een steekproefresultaat vindt dat niet in het kritieke gebied ligt; de nulhypothese wordt dan ten onrechte niet verworpen. De kans op een dergelijke fout wordt aangegeven met β of met $\beta(p_1)$ omdat de waarde van β afhangt van de werkelijke waarde p_1 van p. Is men in een speciale waarde p_1 geïnteresseerd, dan kan men de steekproef zo inrichten dat $\beta(p_1)$ hoogstens gelijk is aan een van te voren gekozen drempel β_0.

Wanneer bij een toetsingsprocedure het steekproefresultaat *niet* tot de conclusie leidt dat de nulhypothese verworpen moet worden, dan luidt de conclusie: de nulhypothese wordt niet verworpen. Dit is iets anders dan de nulhypothese aanvaarden, immers, 'niet verwerpen' is zwakker dan 'aanvaarden'.
De statisticus moet dit wel doen, want ook al vindt hij een steekproefuitkomst welke dicht bij 0,6 ligt, dan kan hij wel de conclusie trekken dat p = 0,6 niet onmogelijk is, dus niet verworpen behoeft te worden, maar hij

zal toch dicht bij 0,6 liggende waarden van p niet kunnen uitsluiten. De conclusie, H_0 wordt aanvaard, gaat derhalve te ver.
De beschreven toets is een eenvoudig voorbeeld van een tweezijdige toets. Vaak is het realistischer om als nulhypothese te nemen $H_0 : p \leqq p_0$, bijvoorbeeld de fractie fouten in een populatie is hoogstens $p_0 = 0{,}02$ en dan als alternatieve hypothese $H_1 : p > p_0 = 0{,}02$, d.w.z. de fractie fouten is groter dan 0,02. De nulhypothese wordt nu alleen verworpen bij grote waarden van de fractie fouten in de steekproef. Er is dus alleen een rechter kritiek gebied en men spreekt van een *rechtseenzijdige toets*. De gedachtengang bij de uitvoering van de eenzijdige toets is in grote lijnen dezelfde als bij een tweezijdige toets. Toetst men de nulhypothese $H_0 : p \geqq p_0$ dan is er alleen een linker kritiek gebied en spreekt men van een *linkseenzijdige toets*.

Wil de accountant informatie over een grootheid verwerven, terwijl er geen uitspraak (door iemand anders) wordt gedaan en erover ook geen verwachting bestaat, dan kan hij deze grootheid schatten. Doet hij dit met een aselecte steekproef, dan kan hij op grond van de in de steekproef gevonden resultaten een interval berekenen, waarin de werkelijke waarde van de onbekende grootheid vermoedelijk ligt. Een dergelijk interval heet een *betrouwbaarheidsinterval*. Ook hier loopt men het risico een onjuiste conclusie te trekken, hetgeen dan betekent dat een betrouwbaarheidsinterval wordt opgegeven dat de werkelijke waarde van de te schatten grootheid niet bevat. Van een reeks uitspraken zal een bepaalde fractie dus onjuist zijn. De gemiddelde waarde van deze fractie is de onbetrouwbaarheid ($= \alpha$) van de methode. Wederom kiest men voordat met het onderzoek begonnen wordt een bovengrens, voor α, de onbetrouwbaarheidsdrempel ($= \alpha_0$), die niet mag worden overschreden en bepaalt men de formules waarmee de grenzen van het interval berekend worden dan zodanig, dat aan deze voorwaarde wordt voldaan. Naast tweezijdig begrensde intervallen is het ook mogelijk te werken met eenzijdig naar boven of eenzijdig naar beneden begrensde intervallen.

3.2 Algemene overwegingen bij de keuze tussen toetsen en schatten in de accountantscontrole

De accountantscontrole is doorgaans gericht op het goedkeuren - en niet op het afkeuren - van een verantwoording, waarbij de accountant in het algemeen uitgaat van de goede trouw van de verantwoordingsplichtige. De controle van een verantwoording is in wezen een toetsingsvraagstuk; de accountant wil in zijn attestfunctie toetsen of de te onderzoeken verantwoording een getrouwe weergave van de werkelijkheid is. De (statistische) toets sluit goed aan bij deze gedachtengang, wanneer de accountant de veronderstelde getrouwheid van een verantwoording door middel van een steekproef bevestigd wil zien.
Schatten is het aangewezen middel, wanneer de accountant over bepaalde kenmerken geïnformeerd wil worden. Deze kenmerken kunnen van velerlei aard zijn en ook niet-ernstige fouten betreffen. De (statistische) schattingstechnieken zijn op zichzelf bruikbaar te maken voor het toetsingsvraagstuk in de accountancy doch zij liggen minder voor de hand.[1]

Zoals wij reeds hiervoor opmerkten sluit de steekproefcontrole op ernstige fouten eveneens goed aan op het gebruik van de statistische toetsingstechniek. Er is namelijk een duidelijke grens tussen de waarde p van de fractie ernstige fouten onder de nulhypothese ($p = p_0 = 0$) en de desbetreffende waarde onder de alternatieve hypothese ($p > 0$), terwijl ook de interpretatie van de toets bevredigender is dan die van een eenzijdig naar boven begrensd betrouwbaarheidsinterval (vergelijk (6) blz. 37 - 38).
Vermaas gaat op onze argumentatie *waarom* wij bij de controle op ernstige fouten de voorkeur geven aan toetsen boven het berekenen van betrouwbaarheidsintervallen, niet in, maar berekent voor twee mogelijke steekproefuitkomsten betrouwbaarheidsintervallen. Uit deze alternatieve berekeningswijze concludeert Vermaas dat het door ons gegeven steekproefschema onlogisch is: (11) blz. 527 - 528. Het is duidelijk dat hier een ondeugdelijke bewijsvoering wordt gegeven.
Bij een steekproefcontrole op niet-ernstige fouten is het voor de accountant veelal moeilijk - en is de noodzaak eigenlijk ook niet zo groot - om exact de grenzen aan te geven van het foutenpercentage waarbij hij de populatie wil accepteren respectievelijk verwerpen.

Vanuit statistisch gezichtspunt gezien hebben Gill en Van Gasselt (4) blz. 469, echter gelijk, dat ook bij ernstige fouten in principe een schattingsmethode toegepast kan worden. Onze argumenten om dit met betrekking tot deze fouten niet te doen, hebben wij in het voorgaande reeds aangegeven. Acht de accountant een bepaald type fouten in beperkte mate toelaatbaar, dan valt dit type niet onder onze categorie ernstige fouten; in die gevallen wordt de keuze tussen toetsingstechniek of schattingstechniek inderdaad bepaald door het antwoord op de vraag of de accountant in staat is de kritieke grens van het percentage fouten te formuleren dat voor hem nog acceptabel is (4).
Dat wij ons hoofdzakelijk geconcentreerd hebben op de combinaties 'ernstige fouten - toetsen' en 'niet-ernstige fouten - schatten' is overigens niet alleen voortgekomen uit ons streven naar eenvoud, maar tevens gebaseerd op de beslissing ons te beperken tot die typen problemen, waarmee wij zelf ruime ervaring hadden met steekproefsgewijze controles.

3.3 Uitwerking met betrekking tot ernstige fouten
De in paragraaf 2 gegeven definitie van een ernstige fout leidt bij steekproefsgewijze controle tot de toetsingsprocedure

(3.1) $H_0 : p = p_0 = 0$

tegen

(3.2) $H_1 : p > 0,$

waarin p de fractie ernstige fouten in de populatie is. Het is dus een rechtseenzijdige toets.

Verder is een logische consequentie van de definitie dat alleen bij $k = 0$ fouten in de steekproef de nulhypothese niet verworpen wordt. De goedkeurgrens k_0 is dus 0. Het voorkomen van één of meer ernstige fouten in de steekproef heeft verwerping van de nulhypothese ten gevolge. Doordat H_0 alleen wordt verworpen wanneer de steekproef één of meer ernstige fouten bevat, hetgeen impliceert dat de fractie fouten p in de populatie groter is dan 0, is de kans α op een fout van de eerste soort gelijk aan 0: het is niet mogelijk dat de nulhypothese ten onrechte wordt verworpen.
Doordat de waarden $k_0 = 0$ en $\alpha = 0$ reeds uit de probleemstelling voortvloeien, wordt de *minimaal vereiste steekproefomvang* door twee grootheden bepaald, te weten, de waarde van p, zeg p_1, waarbij men slechts een kleine kans wil hebben geen enkele in de populatie voorkomende ernstige fout in de steekproef te zien en de waarde β_0 van deze kans.

Deze bijzondere vorm van toetsen, die wij bij de controle op ernstige fouten gebruiken, heeft alleen zin indien de accountant op grond van andere bevindingen en informatie dan uit de te nemen steekproef, de stringente eis $p = p_0 = 0$ aan de populatie kan, respectievelijk moet stellen.
Uit het bovenstaande volgt dat wij het niet eens zijn met Muilwijk (7), blz 655, wanneer hij stelt: 'Daarom vind ik het stel hypothesen 3, waarbij $c = 0$ (bij ons $k_0 = 0$), minder geschikt voor de praktijk'.
De bedoelde hypothesen zijn (3.1) en (3.2). Waarden $k_0 > 0$ geven weliswaar meer keuzemogelijkheden, maar deze zijn per definitie in strijd met ons begrip 'ernstige fout'. Het toelaten van waarden $k_0 > 0$ is wel mogelijk bij onze definitie van 'niet-ernstige fouten'; zie echter paragraaf 3.4.[2].

Is de accountant bereid van een bepaald type fouten een gering aantal in de populatie en daarmede ook in de steekproef te accepteren, dan kan hij deze toetsingsvorm met $k_0 = 0$ onder dezelfde voorwaarden en dus met dezelfde steekproefomvang blijven hanteren door de ene fout zwaarder te wegen dan een andere fout. Een gebrek A in de populatie wordt dan bij het invullen van het begrip 'ernstige fout' bijvoorbeeld zwaarder gewogen dan een gebrek B door zowel het aantreffen van eenmaal gebrek A als viermaal gebrek B als een ernstige fout te rekenen.
De accountant richt bij een controle-opdracht zijn werkzaamheden zodanig in, dat hij in principe tot goedkeuring van de te onderzoeken populatie kan komen. Bij afkeuring van een populatie op grond van de steekproefuitkomsten moet hij zich daarom beraden op aanvullende controlemaatregelen om alsnog tot een goedkeuring van de verantwoording te komen.

Leidt het steekproefonderzoek, uitgevoerd als toets, tot afkeuring, dan is er een mogelijkheid om met de steekproefresultaten alsnog door middel van een schatting een indicatie van de ernst van de situatie te geven.
De verkregen informatie kan als hulpmiddel dienen voor de accountant om de richting en de omvang van het aanvullende onderzoek nader te bepalen. Wij berekenen dan een voorwaardelijke betrouwbaarheidsbovengrens voor de aanwezige fractie fouten, respectievelijk het betrokken bedrag aan fouten. Dit kan op de gebruikelijke wijze gebeuren; de werkelijke onbetrouw-

baarheid van het opgegeven interval zal steeds kleiner zijn dan de gehanteerde onbetrouwbaarheidsdrempel α_0.
Voor een nadere uitwerking van deze methode verwijzen wij naar J. Kriens (5) blz 9.

Tenslotte merken wij in aansluiting op het gestelde in paragraaf 3.1 op, dat wanneer H_0 niet verworpen wordt, dit niet betekent dat wij de conclusie trekken dat de fractie fouten in de populatie 0 is. In dit opzicht maakt Vermaas een denkfout wanneer hij stelt: 'Het voorgaande betekent 'gewoon' dat de accountant bij het vinden van 0 fouten de hypothese aanvaardt, dat de foutenfractie in de populatie 0 is' (11) blz. 527.

3.4 Uitwerking met betrekking tot niet-ernstige fouten
Bij niet-ernstige fouten ligt de keuze tussen toetsen en schatten minder duidelijk dan bij ernstige fouten. Terzijde merken wij op dat in het volgende controle op niet-ernstige fouten door middel van het Average Outgoing Quality Limit (A.O.Q.L.)-systeem buiten beschouwing zal worden gelaten.

Overweegt men een *toetsingsmethode*, dan is het in tegenstelling tot de situatie bij ernstige fouten, niet meer vanzelfsprekend dat de goedkeurgrens k_0 gelijk aan 0 gekozen wordt, doch kunnen ook andere waarden van k_0 in aanmerking komen. De toets zal gewoonlijk dan ook niet de vorm (3.1) tegen (3.2) hebben, maar eerder bijvoorbeeld de algemene vorm

(3.3) $\quad H_0 : p \leqq p_0$

tegen

(3.4) $\quad H_1 : p \geqq p_1 \quad (p_1 > p_0).$

Voert men de toets in deze vorm uit, dan moeten behalve de waarden p_0 en p_1 ook de maxima α_0 en β_0 van de kansen op het ten onrechte verwerpen en het ten onrechte niet verwerpen van de nulhypothese gespecificeerd worden.
Een belangrijk probleem voor de accountant is bij deze toets het exact aangeven van twee grenzen, namelijk één, waarbij de fractie fouten nog aanvaardbaar is (bijvoorbeeld $p_0 = 0{,}02$) en één, waarbij hij de foutenfractie niet meer aanvaardbaar vindt (bijvoorbeeld $p_1 = 0{,}03$).
De vraag rijst of bij elk steekproefonderzoek een dergelijke verfijning van de normstelling der foutenfractie controletechnisch gezien noodzakelijk is.

Wanneer de doelstelling van het steekproefonderzoek in bepaalde situaties wel een verfijning van die grenzen zou vergen, kan men zich bovendien afvragen in hoeverre de accountant die twee grenzen duidelijk kan vaststellen.
Het is derhalve discutabel of het mogelijk is in het controlewerk een toets in de vorm (3.3) tegen (3.4) goed te formuleren. Wel kunnen er twee algemene opmerkingen over worden gemaakt:

- bij het toepassen van een toets van de vorm (3.3) tegen (3.4) zal de accountant geneigd zijn β_0 (H_0 ten onrechte niet verwerpen, dus een vaktechnisch risico) lager te kiezen dan α_0 (H_0 ten onrechte verwerpen, hetgeen extra kosten veroorzaakt en het functioneren tegenover de cliënt in het geding brengt);
- de accountant die een nuancering in de grootten van de waarden p_0 en p_1 wil aanbrengen, beïnvloedt daarmee de grootte van de steekproefomvang. Naarmate het verschil tussen de onaanvaardbaar en de aanvaardbaar geachte fractie fouten in de populatie ($= p_1 - p_0$) groter of kleiner wordt gekozen, neemt de steekproefomvang af of toe.

Kiest de acccountant voor een *schattingsmethode*: het op grond van de steekproefresultaten berekenen van een betrouwbaarheidsinterval, dan moet hij, voordat de steekproef genomen wordt, alleen de vereiste betrouwbaarheid ($= 1 - \alpha_0$) van het op te geven interval kiezen en zo mogelijk iets bepalen omtrent de gewenste breedte van het interval.
Hoewel ook het specificeren van deze grootheden in het algemeen lastig zal zijn, lijkt het toch minder moeilijk dan de voor de toets (3.3) tegen (3.4) vereiste vier grootheden p_0, p_1, α_0 en β_0. Daarom geven wij met betrekking tot niet-ernstige fouten bij de keuze tussen een schattingsmethode en een toetsingsmethode in de meeste gevallen de voorkeur aan de eerste.

4. Tolerantie en risico

4.1 Inleiding
De accountant die op grond van (aselecte) steekproeven uit populaties uitspraken over die populatie wil doen, moet rekening houden met toleranties in de gewenste uitspraken, en met risico's van onjuiste uitspraken.
De begrippen 'tolerantie' en 'risico' zijn ontleend aan het algemene spraakgebruik, maar sluiten niet steeds volledig aan bij de in de statistiek gehanteerde begrippen. Aan beide kan een meer specifieke inhoud worden gegeven al naar gelang het steekproefonderzoek gebaseerd is op toetsen of op schatten.

4.2 Tolerantie
Bij het *toetsen* van hypothesen wordt nagegaan of de te onderzoeken populatie naar alle waarschijnlijkheid aan een van te voren geuite verwachting of gestelde norm (bijvoorbeeld foutenfractie) voldoet. Voldoet de populatie aan de gestelde norm, dan wordt deze aanvaard; is dit niet het geval, dan wordt zij niet aanvaard.
Met betrekking tot ernstige fouten is een populatie aanvaardbaar, wanneer deze geen enkele ernstige fout bevat. Door deze wijze van definiëren is het, zoals wij reeds gezien hebben, noodzakelijk een populatie reeds af te keuren, zodra één ernstige fout in de steekproef wordt gevonden.

In een door ons gegeven en door Vermaas aangehaald voorbeeld (6) blz. 57, respectievelijk (11) blz. 525, moet een accountant 130.000 inkoopfacturen controleren op ernstige fouten, in dit geval ontbrekende ontvangstmeldin-

gen. Hij besluit dit met behulp van een steekproef te doen, waaraan hij de eis stelt dat de kans geen enkele fout in de steekproef aan te treffen hoogstens 2% bedraagt, als de fractie ontbrekende ontvangstmeldingen meer dan 0,5% bedraagt.
Vermaas reageert hierop als volgt: 'Om met een steekproef te kunnen werken is men gedwongen te bepalen welke tolerantie en welk risico men aanvaardbaar vindt. In het voorbeeld uit paragraaf 2 is bij goedkeuring de tolerantie 0,5% en het risico 2%'.

Zelf hebben wij in dit voorbeeld het woord tolerantie bewust niet gebruikt. Het woord tolerantie suggereert dat het aanvaardbaar is en zo stelt Vermaas het ook op blz. 526 van zijn artikel, waar hij spreekt over 'aanvaardbare tolerantie'. Ten aanzien van ernstige fouten is de aanvaardbare tolerantie naar onze mening nul, dus geen enkele ernstige fout, en niet 0,5%.
Of er geen enkele ernstige fout in een populatie aanwezig is, kan alleen door volledige detailcontrole worden vastgesteld. In de praktijk is dat op grond van het economisch motief veelal niet verantwoord, zodat in veel gevallen de voorkeur wordt gegeven aan een aselecte steekproef. De accountant aanvaardt bij een steekproefcontrole op ernstige fouten bewust het risico dat een populatie ten onrechte wordt goedgekeurd, maar hij kan dit risico kwantificeren en er grenzen aan stellen. In het genoemde voorbeeld hebben wij dit gedaan, door te eisen dat de kans geen enkele fout in de steekproef aan te treffen hoogstens 2% bedraagt, als de fractie ernstige fouten in de populatie meer dan 0,5% bedraagt. Voor fracties fouten kleiner dan 0,5% is de kans geen enkele fout in de steekproef aan te treffen groter dan 2%, voor fracties fouten groter dan 0,5% is deze kans kleiner dan 2%.
Het in formulevorm uitgedrukte verband tussen de fractie fouten p in de populatie, de steekproefomvang en de kans geen enkele fout in de steekproef te zien, is vermeld in (6), blz. 54 - 55.
Het bovenstaande betekent *niet*, dat 0,5% ernstige fouten een 'aanvaardbare tolerantie' is.

Samenvattend kunnen wij zeggen dat het algemene begrip 'tolerantie' in dit geval van toetsen inhoudt, dat in feite geen enkele ernstige fout in de populatie aanvaardbaar is, maar dat wij op rationele gronden hoogstens een kans β_0 accepteren de populatie ten onrechte goed te keuren wanneer de fractie ernstige fouten in de populatie p_1 of meer bedraagt.
De opmerking van Vermaas: 'Wat dat betreft is het tolerantie of geen tolerantie, dat wil zeggen steekproef of geen steekproef' is naar onze mening een zwart-wit stelling, waarmee men in praktische situaties niet veel kan beginnen.

Past men de meer algemene toets (3.3) tegen (3.4) toe, dan kan men bij $p_1 > p_0$ het begrip tolerantie interpreteren als het verschil tussen p_1 en p_0.
Berekent men met behulp van een aselecte steekproef een *schatting* in de vorm van een betrouwbaarheidsinterval voor een onbekende grootheid, dan kan men het verschil tussen de boven- en de ondergrens van het interval interpreteren als de aanwezige tolerantie.

4.3 Risico

Evenals het begrip tolerantie heeft ook het begrip risico bij toetsen en bij schatten een verschillende inhoud. Bij het *toetsen* van een hypothese loopt de accountant, zoals wij in paragraaf 3.1. gezien hebben, een tweeledig risico, namelijk:

- het alfarisico (α): de kans op het ten onrechte verwerpen van de nulhypothese;
- het bêtarisico (β): de kans op het ten onrechte niet verwerpen van de nulhypothese.

Bij *schatten* is het risico de onbetrouwbaarheid van de gehanteerde schattingsmethode.

5 Enkel- en meervoudige steekproeven, of 'Uitbreiden van de steekproef?'

Gezien de uitvoerige discussie over 'uitbreiden' van steekproeven in het MAB gedurende de afgelopen twee jaren (Vermaas (10) en (12), Blokdijk (2) en (3)), zullen wij ons in hoofdzaak beperken tot hetgeen Vermaas hierover zegt in verband met ons boek.

In de statistiek maakt men onderscheid tussen de volgende typen steekproeven:

1 de enkelvoudige steekproef, of steekproef met vaste omvang n: er wordt één steekproef van n exemplaren genomen en op grond daarvan wordt een populatie goedgekeurd of afgekeurd;
2 de dubbele steekproef, of steekproef in twee stappen: men neemt een steekproef van de omvang n_1, keurt op grond van het gevonden resultaat goed of af, of besluit nog een steekproef van de omvang n_2 te nemen en keurt dan de populatie goed of af op grond van de resultaten bij de $n_1 + n_2$ waarnemingen tezamen;
3 de meervoudige steekproef, of steekproef in meer dan twee stappen; deze werkt als de dubbele steekproef, maar nu kan ook na de tweede (eventueel derde, enzovoorts) stap besloten worden nog een steekproef te nemen;
4 de sequente steekproef, waarbij de elementen in principe één voor één worden aangewezen en na ieder gekeurd element wordt beslist: goedkeuren, afkeuren, of nog een controle verrichten.

De bedoeling van de onder de punten 2, 3 en 4 genoemde typen steekproeven is om, indien mogelijk, althans gemiddeld, tot een kleiner aantal te verrichten controles te komen voordat een beslissing wordt bereikt, dan bij de enkelvoudige steekproef is vereist. Daar staat tegenover dat de uitvoering van die steekproeven in het algemeen ingewikkelder is dan de toepassing van de enkelvoudige steekproef.

Men kan langs wiskundige weg bewijzen dat de gemiddeld vereiste steekproefomvang om de toets (3.1) tegen (3.2) uit te voeren het kleinst is wanneer men een sequente steekproef toepast. Dit betekent voor controles op ernstige fouten ($k_0 = 0$) dat men theoretisch gezien het beste sequent te

werk zou kunnen gaan: de elementen worden één voor één aangewezen, vervolgens gecontroleerd en men stopt met de steekproef zodra er een fout gevonden is.
De besparingen zijn echter gering in situaties waarin de fractie fouten p in de populatie klein is, hetgeen bij controles op ernstige fouten in het algemeen het geval zal zijn. De uitvoering daarentegen stuit op twee praktische bezwaren: ten eerste moet men met de controle wachten tot de populatie volledig aanwezig is en ten tweede moet ieder nieuw element uit de gehele populatie worden getrokken, waardoor de populatie dus niet in één gang van het begin tot het einde kan worden doorgewerkt. Op grond van deze overwegingen verdient bij controles op ernstige fouten de steekproef met vaste omvang de voorkeur.

Met betrekking tot niet-ernstige fouten is de methode van de betrouwbaarheidsintervallen het meest voor de hand liggend, zoals wij in paragraaf 3.4 opmerkten. In principe kunnen ook deze berekend worden op grond van de vier genoemde steekproeftypen. De formules worden echter al snel gecompliceerd, reden waarom de onder de punten 2, 3 en 4 genoemde steekproeven vrijwel niet worden gebruikt om betrouwbaarheidsintervallen uit te rekenen. Toepassingen van betrouwbaarheidsintervallen bij accountantscontroles, niet gebaseerd op enkelvoudige steekproeven, zijn ons noch uit de praktijk, noch uit de literatuur bekend.
Uit het bovenstaande blijkt dat er bij accountantscontroles tot nu toe weinig ruimte is voor andere steekproeven dan die met vaste omvang. Toch duiken deze andere methoden met betrekking tot keuringsproblemen regelmatig in de literatuur op, in het bijzonder voor situaties waarin op grond van het steekproefresultaat moet worden afgekeurd.

Beek (1) blz. 141, wil in een dergelijke situatie de objectieve statistische methode zonodig vervangen door een subjectieve, 'voor zover deze verandering niet in strijd is met hetgeen zijn geweten (het geweten van de accountant, D., K.) hem ingeeft'. Doet de accountant dit, dan moet hij wel bedenken dat op de mathematische statistiek gebaseerde uitspraken niet meer mogelijk zijn. Wij menen echter, dat invoering van de (aselecte) steekproef juist ten doel heeft wel zo veel mogelijk geobjectiveerde uitspraken te krijgen. Vandaar dat wij statistisch niet goed onderbouwde enkelvoudige en meervoudige steekproeven afwijzen.

Vermaas (11) blz. 532, zoekt de oplossing in 'uitbreiden van de steekproef'. Hoe dit in zijn werk gaat, illustreert hij aan de hand van het reeds in paragraaf 4.2 aangehaalde voorbeeld. Daar wordt de eis gesteld dat de kans geen enkele fout in de steekproef aan te treffen en dus de populatie goed te keuren hoogstens 2% bedraagt, wanneer het percentage ernstige fouten in de populatie groter is dan 0,5%.
In zijn toelichting vervangt Vermaas de kans van 2% door 2,5%, zodat de eisen in de terminologie van paragraaf 4.3 luiden: $p_1 = 0{,}005$ en $\beta_0 = 0{,}025$. De vereiste steekproefomvang is dan $n = 738$ omdat de kans op 0 fouten in

de steekproef bij een fractie 0,005 ernstige fouten in de populatie gelijk is aan 0,024972.
Voor de berekening van deze kans verwijzen wij naar de Appendix. Men moet dus een steekproef nemen van 738 posten, goedkeuren wanneer er 0 fouten worden aangetroffen en anders afkeuren.
Het eerste doet Vermaas ook, maar wanneer hij wel fouten vindt keurt hij pas dan af, wanneer het er 9 of meer zijn. Bij 1 t/m 8 fouten breidt hij de steekproef uit met een tweede stap, waarin hij $n_2 = 376$ posten controleert. Hij keurt dan goed wanneer er bij de 738 + 376 gecontroleerde posten slechts één fout gevonden wordt; hij keurt af wanneer het er 12 of meer zijn, en neemt nog een derde stap bij een aantal fouten tussen de 2 en 11. Vermaas schept zich door deze werkwijze nieuwe mogelijkheden om alsnog goed te keuren als het aantal fouten in de eerste stap tussen 1 en 8 ligt. In de tweede stap is dit het geval bij 0 fouten en slechts 1 gevonden fout in de eerste stap. In de derde stap doen zich twee extra mogelijkheden tot goedkeuren voor, te weten, wanneer er in de opeenvolgende stappen respectievelijk 1, 1 en 0 fouten gevonden worden, of wanneer dit er respectievelijk 2, 0, 0 zijn. De kans om bij $p_1 = 0{,}005$ in 3 stappen goed te keuren is dan 0,049, zoals in de Appendix wordt aangetoond.

Worden er verdere uitbreidingen toegepast, dan wordt de totale kans de populatie ten onrechte goed te keuren nog groter; bij 5 stappen is de goedkeurkans ruim 0,06 of ruim 6%. Het is duidelijk dat aan de gestelde eis, voor $p_1 \geq 0{,}005$ een goedkeurkans van hoogstens 0,025 niet meer is voldaan.
Terzijde merken wij op dat dit verschijnsel niet afhangt van de steekproefomvangen in de opeenvolgende stappen, of van de goed- en afkeurgrenzen. Het geldt algemeen omdat de gekozen waarde van β_0 reeds in de eerste stap verbruikt is door de wijze waarop n_1 bepaald is. Door de 'uitbreidingen' kan de kans op ten onrechte goedkeuren alleen maar groter worden.

Het is wel mogelijk 'uitbreidingen' toe te passen, of, anders gezegd, meer dan één stap in het steekproefschema op te nemen, zodanig dat wel aan de gestelde eis is voldaan. Een dubbel steekproefschema dat aan de eis voldoet, is eerst een steekproef nemen van de omvang $n_1 = 781$, goedkeuren bij 0 fouten, afkeuren bij 2 of meer fouten en nog een steekproef van de omvang $n_2 = 550$ nemen bij 1 fout.
Na de tweede steekproef kan dan alsnog goedgekeurd worden wanneer bij de 781 + 550 gecontroleerde posten slechts 1 fout gevonden is. In alle andere gevallen wordt nu afgekeurd. De toegestane kans om bij $p_1 = 0{,}005$ goed te keuren is verdeeld over de twee stappen en wel 0,020 in de eerste stap en 0,005 in de tweede stap. Ook deze afleiding is terug te vinden in de Appendix.

Het is niet moeilijk andere schema's te bedenken. Zij hebben echter alle als nadeel dat de steekproefomvang in de eerse fase groter moet zijn dan $n_1 = 738$ teneinde een kans op goedkeuren kleiner dan 0,025 te bereiken. Voor dit controleprobleem verdient de enkelvoudige steekproef derhalve de

voorkeur, hetgeen aansluit bij de in het begin van deze paragraaf gemaakte opmerking.

Nawoord

Dit artikel willen wij afsluiten met een woord van dank aan degenen die op ons boek gereageerd hebben. De reacties hebben ook ons inzicht in de materie verdiept.

Appendix

Wanneer men een steekproef neemt van de omvang n uit een grote populatie met een fractie p aan fouten, dan is het aantal fouten k in de steekproef bij benadering binomiaal verdeeld (k is onderstreept omdat het een variabele is met een kansverdeling).
De kans op k fouten wordt als volgt aangegeven.

$P\left[\underline{k} = k \mid n; p\right]$.

Voor kleine waarden van p kan deze kans goed benaderd worden met behulp van een Poissonverdeling. Er geldt dan

$$(1) \quad P\left[\underline{k} = k \mid n; p\right] \approx \frac{(np)^k}{k!} \times e^{-np}.$$

De gestelde eis luidt:

$P\left[\underline{k} = 0 \mid n; p_1 = 0{,}005\right] \leqq 0{,}025$

of

$e^{-n \cdot 0{,}005} \leqq 0{,}025.$

Hieruit volgt $n = 738$; de desbetreffende kans bedraagt dan 0,024972.

Geven wij de steekproefomvang in de i-de stap aan met n_i en het aantal in die stap gevonden fouten met k_i, dan ziet het door Vermaas voorgestelde steekproefschema er als volgt uit:

$n_1 = 738;$ $\quad k_1 = \begin{cases} 0 & \text{goedkeuren} \\ 1 \text{ t/m } 8 & \text{tweede stap} \\ \geqq 9 & \text{afkeuren} \end{cases}$

$n_2 = 376;$ $\quad k_1 + k_2 = \begin{cases} 1 & \text{goedkeuren} \\ 2 \text{ t/m } 11 & \text{derde stap} \\ \geqq 12 & \text{afkeuren} \end{cases}$

$$n_3 = 331; \qquad k_1 + k_2 + k_3 = \begin{cases} 2 & \text{goedkeuren} \\ 3 \text{ t/m } 13 & \text{vierde stap} \\ \geqq 14 & \text{afkeuren.} \end{cases}$$

Na de derde stap is desgewenst verdere uitbreiding mogelijk, maar in navolging van Vermaas beperken wij ons tot drie stappen (= twee uitbreidingen).

Er wordt in de volgende situaties goedgekeurd:

$k_1 = 0$;

$k_1 = 1$ en $k_2 = 0$;

$k_1 = 1$, $k_2 = 1$ en $k_3 = 0$;

$k_1 = 2$, $k_2 = 0$ en $k_3 = 0$.

De kansen hierop zijn respectievelijk, benaderd met formule (1),:

$$P\left[k_1 = 0 \mid n_1 = 738;\ p_1 = 0{,}005\right] \approx e^{-738\,.\,0{,}005} = 0{,}025;$$

$$P\left[k_1 = 1 \wedge k_2 = 0\right] = P\left[k_1 = 1 \mid n_1 = 738;\ p_1 = 0{,}005\right].$$

$$P\left[k_2 = 0 \mid n_2 = 376;\ p_1 = 0{,}005\right] = (3{,}69 \,.\, e^{-3{,}69}) \,.\, e^{-1{,}88} = 0{,}014;$$

$$P\left[k_1 = 1 \wedge k_2 = 1 \wedge k_3 = 0\right] = P\left[k_1 = 1 \mid n_1 = 738;\ p_1 = 0{,}005\right].$$

$$P\left[k_2 = 1 \mid n_2 = 376;\ p_1 = 0{,}005\right] . P\left[k_3 = 0 \mid n_3 = 331;\ p_1 = 0{,}005\right] =$$

$$= (3{,}69 \,.\, e^{-3{,}69}) \,.\, (1{,}88\, e^{-1\,88}) \,.\, e^{-1{,}655} = 0{,}005;$$

$$p\left[k_1 = 2 \wedge k_2 = 0 \wedge k_3 = 0\right] = P\left[k_1 = 2 \mid n_1 = 738;\ p_1 = 0{,}005\right].$$

$$P\left[k_2 = 0 \mid n_2 = 376;\ p_1 = 0{,}005\right] . P\left[k_3 = 0 \mid n_3 = 331;\ p_1 = 0{,}005\right] =$$

$$= \left(\frac{(3{,}69)^2}{2!}\, e^{-3{,}69}\right) . e^{-1{,}88} \,.\, e^{-1{,}655} = 0{,}005.$$

De totale goedkeurkans bedraagt derhalve:
$0{,}025 + 0{,}014 + 0{,}005 + 0{,}005 = 0{,}049$.

Een dubbel steekproefschema dat aan de eisen voldoet luidt:

$$n_1 = 781; \qquad k_1 = \begin{cases} 0 & \text{goedkeuren} \\ 1 & \text{tweede stap} \\ \geqq 2 & \text{afkeuren} \end{cases}$$

$$n_2 = 550; \qquad k_1 + k_2 = \begin{cases} 1 & \text{goedkeuren} \\ \geqq 2 & \text{afkeuren.} \end{cases}$$

De kans op goedkeuren, wanneer $p_1 = 0{,}005$, is

$$P\left[k_1 = 0 \mid n_1 = 781;\ p_1 = 0{,}005\right] + P\left[k_1 = 1 \mid n_1 = 781;\ p_1 = 0{,}005\right] .$$

$$. P\left[k_2 = 0 \mid n_2 = 550;\ p_1 = 0{,}005\right] = e^{-781\ 0{,}005} + (3{,}905 . e^{-3{,}905}) . e^{-2{,}75} =$$

$$= 0{,}020 + 0{,}005 = 0{,}025.$$

Men kan de hiervoor berekende kansen - via interpolaties - met behulp van tabel 4 uit de bij (8) behorende verzameling tabellen narekenen.

Literatuur

1 Beek, A., Boekbespreking: Steekproeven in de accountantscontrole (zie 6), Bedrijfskunde 52 (1980) 138 - 142.
2 Blokdijk, J. H., Uitbreiding van de steekproef, MAB 53 (1979) 537 - 541.
3 Blokdijk, J. H., Terug naar de steekproef, MAB 54 (1980) 127 - 131.
4 Gill, R. D. en Gasselt, L. G. P. van, Boekbespreking: Steekproeven in de accountantscontrole (zie 6), De Accountant 86 (1980) 469 - 470.
5 Kriens, J., Statistical Sampling in Auditing, Invited Paper for the 42nd Session of the International Statistical Institute, Manila, December 4 - 14, 1979, Reeks ter Discussie No. 79.087, Faculteit der Economische Wetenschappen, K.H.T., Tilburg (1979).
6 Kriens, J. en Dekkers, A. C., Steekproeven in de accountantscontrole, Uitgave H.E. Stenfert Kroese B.V., Leiden (1979).
7 Muilwijk, J., Kansrekening, Statistiek en Accountantscontrole, MAB 55 (1981) 628 - 669.
8 Statistiek en Kansrekening voor het V.W.O., Uitgave Stichting IVIO, Lelystad, 2e druk, (1974).
9 Tuitjer, E. B., Boekbespreking: Steekproeven in de accountantscontrole, (zie 6), MAB 53 (1979) 600 - 601.
10 Vermaas, M., Uitbreiden van de steekproef, MAB 53 (1979) 203 - 215.
11 Vermaas, M., Steekproeven bij controles op ernstige fouten, MAB 53 (1979) 525 - 536.
12 Vermaas, M., Van Steekproef tot steekspel, MAB 54 (1980) 36 - 40.

Noten

1 Statistisch gezien leiden het toetsen van hypothesen en het berekenen van betrouwbaarheidsintervallen immers tot dezelfde conclusies indien de nulhypothese duidelijk bekend is.

2 In het kader van dit artikel zullen wij verder niet ingaan op de bijdrage van Muilwijk.