



Essaybundel Statistical Auditing

Essaybündel Statistical Auditing

Colofon

Copyright © 2018

Uitgever: Publishing House Jacques de Swart

Redactie:

Hein Kloosterman, Paul van Batenburg en Kees Voorhoeve.

Coauteurs:

Ed Broeze, Roger Dassen, Ferry Geertman, Wouter Gerards, Arnold Heertje, Harrie Hendriks, Robbert Hoyng, Barbara Majoor, Niels van Nieuw Amerongen, Jacques de Swart, Ruud Veenstra en Jan Wille.

Eerste druk

Ontwerp en vormgeving:

wen-d | vormgeving-dtp

Economie, recht en bedrijfskunde algemeen

ISBN 978-90-816974-0-8

NUR 160

Niets uit deze uitgave mag worden verveelvoudigd, door middel van druk, fotokopieën, geautomatiseerde gegevensbestanden of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

Verantwoording

Deze bundel is op persoonlijke titel samengesteld door Hein Kloosterman en Paul van Batenburg.

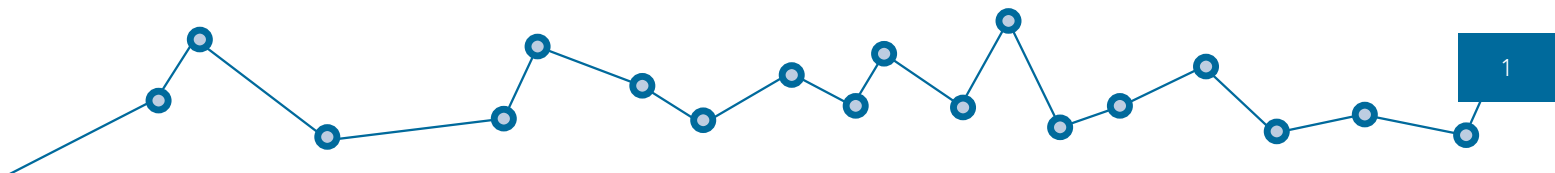
Aanzet tot deze bundel is genomen in de vergadering van de Stuurgroep Statistical Auditing van 22 november 2017. Doel ervan is een beeld te geven van het werk van de Stuurgroep door te verwijzen naar wetenschappelijk en praktisch onderzoek op het gebied van statistical auditing en de modellering van de beslissingen in het controleproces.

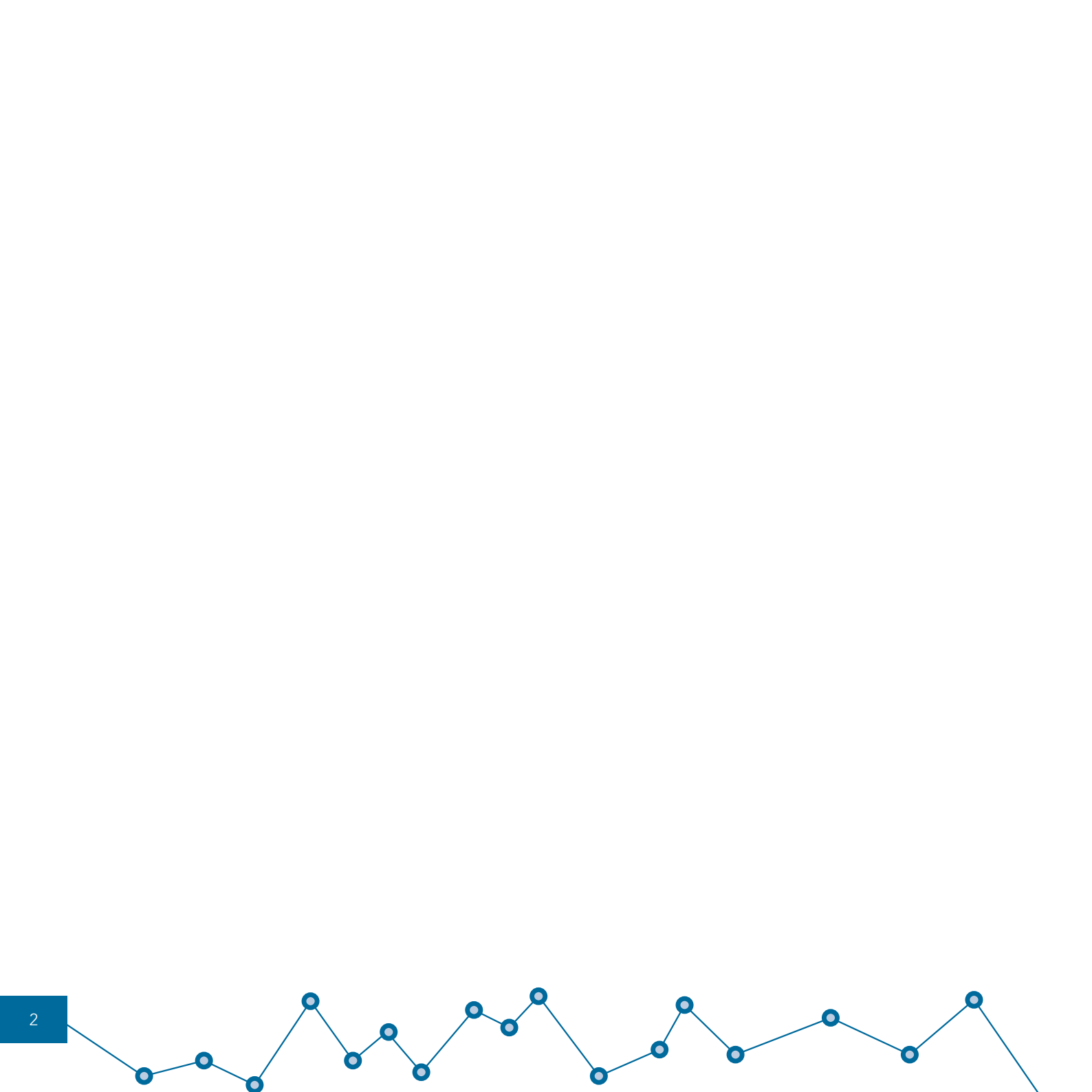
Voor de bundel zijn onder meer eerder geplaatste columns en artikelen gebruikt van (oud-)leden van de Stuurgroep (met hun toestemming). Sommige artikelen hebben een coauteur die niet lid is van de Stuurgroep. Dit zijn – in alfabetische volgorde – Roger Dassen, Arnold Heertje, Robbert Hoyng, Barbara Majoor, Niels van Nieuw Amerongen en Jan Wille. Voor zover van toepassing is aan de uitgever gevraagd of plaatsing in deze essaybundel kan plaatsvinden. Bij de betreffende artikelen is aangegeven waar zij eerder zijn gepubliceerd.

In sommige artikelen of essays worden begrippen als standaard, controlestandaard, ISA, COS of NV COS gebruikt. Deze begrippen moeten worden verstaan als synoniem.

Deze bundel is samengesteld om (als digitaal boek) beschikbaar te worden gesteld na afloop van het elfde symposium Statistical Auditing. Het motto van dat symposium was: “Auditkwaliteit: wanneer is er genoeg gecontroleerd?”

Mei 2018





Inhoudsopgave

Verantwoording	1	3.9 Data-analyse en statistical auditing	147
Voorwoord	4	3.10 Fraude en statistical audit	158
Leeswijzer	5	3.11 Postensteekproef om uitspraken in geld te doen	165
1 Inleidende artikelen		4 Gegevensgerichte cijferanalyses	
1.1 Inleiding Essaybundel Statistical Auditing	13	4.1 Gegevensgerichte cijferanalyse met behulp van regressierekening	171
1.2 Korte samenvatting Leslie, Teitlebaum & Anderson	34	5 Risicoanalyse in de accountantscontrole	
1.3 De NV COS in samenhang met Statistical Audit	43	5.1 Het Bayesiaanse model van Deloitte & Touche ter ondersteuning van de controlemix	175
1.4 Haat-liefdeverhouding tussen accountants en statistiek	50	5.2 Wat is eigenlijk risicoanalyse in de accountantscontrole	185
1.5 Materiality – Clarity?	54	5.3 Doeltreffende administratieve controle door steekproeven	201
1.6 Professional judgement, feilbaar denken	82	5.4 Samenvatting dissertatie 'Valideren van risico-inschattingen in accountantscontrole'	218
2 Interne beheersing		5.5 Rekenen met plaatjes: een proefschrift over een nieuwe kijk op steekproeven	225
2.1 Transactiemodel	89	5.6 Het 'push left'-principe als motor van data-analytics in de accountantscontrole	230
2.2 Interne beheersing en statistical audit; een schets	95	5.7 Het Bayesiaanse model en Tax Assurance	247
3 Detailcontroles		Auteurs	253
3.1 Steekproefselectie, de basis	103	Verklarende woordenlijst	259
3.2 Fundamenteel onderzoek naar wiskundige steekproeven	106	Statistical Auditing for dummies	267
3.3 Steekproeven bij volledigheidscntroles	111		
3.4 Isoleren: atypische of triviale afwijkingen	115		
3.5 Uitbreiden van steekproeven om beter te schatten	119		
3.6 Als ik niet kan goedkeuren: zes tips	122		
3.7 Het evalueren van steekproeven	125		
3.8 Goedkeuren, ondanks fouten	140		

Voorwoord

Deze bundel met essays, artikelen, is samengesteld naar aanleiding van de rondetafelconferentie over de Green Paper Auditkwaliteit in september 2017. Namens de Stuurgroep Statistical Auditing is er kritiek geleverd op dit Green Paper. De kern daarvan is dat het auditproces zelf een kritische succesfactor is voor de kwaliteit van een financial audit.

De Stuurgroep Statistical Auditing houdt zich bezig met een aantal belangrijke aspecten van de kwaliteit van het auditproces. Met het verzamelen van artikelen die licht werpen op die kwaliteit en op de mogelijkheden tot het verbeteren ervan zagen wij de mogelijkheid van een aanvulling op de onderzoeksprogramma's van de Foundation for Auditing Research (FAR).

Een aantal auteurs heeft speciaal voor deze bundel korte artikelen geschreven. Zoals in de verantwoording weergegeven hebben wij ook een aantal artikelen die al eens gepubliceerd zijn in deze bundel opgenomen. Verder zijn er columns Statistical Auditing opgenomen die zowel gelden ter lering als ter vermaak. Voorzover nodig zullen wij een clustering van columns, zoals bijvoorbeeld die over evalueren van steekproeven, kort gaan inleiden.

In het eerste artikel van deze bundel, Inleiding Essaybundel Statistical Auditing, is een raamwerk van de controle opgenomen. Aan dat raamwerk zijn de belangrijkste elementen van een financiële controle 'opgehangen'.

Namens de Stuurgroep,

Hein Kloosterman
Paul van Batenburg

Leeswijzer

Centraal thema van deze bundel is de kwaliteit van het auditproces. De Stuurgroep Statistical Auditing houdt zich bezig hoe die kwaliteit met behulp van methoden en technieken te verbeteren. Daartoe bevat deze bundel vijf soorten artikelen, namelijk 1) algemene artikelen over het auditproces, 2) over de beoordeling van interne beheersing, 3) over gegevensgerichte detailcontroles, 4) over gegevensgerichte cijferanalyses en tenslotte 5) over het risicoanalysemodel van de accountantscontrole.

1 Inleidende artikelen

1.1 Overzichtsartikel

Het overzichtsartikel behandelt de hoofdlijnen van het controleproces aan de hand van een netwerkmodel. Met behulp van dat netwerkmodel beschrijven Paul van Batenburg en Hein Kloosterman de beoordeling van de interne beheersing, inclusief de IT, de gegevensgerichte detailcontroles en de gegevensgerichte cijferanalyse. Daarnaast gaan zij in op de risicoanalyse. Zij geven aan hoe en waar door middel van (wetenschappelijk) onderzoek de praktijk kan leren van de theorie en andersom.

1.2 Korte samenvatting Leslie, Teitlebaum en Anderson

Deze samenvatting is als apart artikel opgenomen omdat deze auteurs als eersten het trekken en evalueren van geldsteekproeven praktisch toepasbaar hebben beschreven. Hun boek bestrijkt het totale gebied van de financiële controle.

1.3 NV COS in samenhang met Statistical Audit

In dit artikel heeft Hein Kloosterman geprobeerd met behulp van het netwerk van de controle zoals dat in het overzichtsartikel is gepresenteerd, de relatie van de verschillende elementen van de controle met de NV COS in verband te brengen met Statistical Auditing.

1.4 Haat-liefdeverhouding tussen accountants en statistiek

Jacques de Swart en Paul van Batenburg leggen in dit artikel een verband tussen de statistiek die accountants tijdens hun opleiding krijgen aangeboden en de statistiek die accountants in de praktijk als accountant zouden kunnen gebruiken. Het artikel laat zien dat de opleidingseisen geen plaats hebben ingeruimd voor de statistiek waarmee de detailcontroles kunnen worden gemodelleerd.

1.5 Materialiteit, Clarity?

Niels van Nieuw Amerongen en Paul van Batenburg leggen in dit artikel uit hoe de standaarden omgaan met de definiëring van de begrippen materialiteit, uitvoeringsmaterialiteit en toelaatbare afwijking. De auteurs geven de technische samenhang tussen deze begrippen. Zij laten daarnaast een aantal knelpunten zien.

1.6 Professional judgement, feilbaar denken

De auteur, Ed Broeze, maakt gebruik van de theorieën van Tversky en Kahnemann om een zeer kritische kanttekening te plaatsen bij de elementen van professional judgement bij het vaststellen van materialiteit, het toereikend achten van de interne beheersing en vooral bij risicoanalyse in de accountantscontrole.

2 Artikelen over de beoordeling van de interne beheersing

2.1 Transactiemodel

Hein Kloosterman heeft een artikel geschreven over het transactiemodel. In dit artikel legt hij het verband tussen de gebeurtenissen in de echte wereld (real world) en de vastleggingen ten behoeve van de interne beheersing en de verantwoording die het personeel, van hoog tot laag, moet kunnen afleggen.

2.2 Interne beheersing en statistical audit; een schets

Hein Kloosterman geeft een schets van hoe de interne beheersing met behulp van instrumentarium van statistical audit kan worden ondersteund, onderzocht en beoordeeld.

3 Artikelen over detailcontroles

3.1 Artikel Steekproefselectie, de basis

In de bundel is een artikel opgenomen over de praktische kant van het trekken van steekproeven (steekproefselectie).

3.2 Artikel Fundamenteel onderzoek naar wiskundige steekproeven

De theoretische kant van de wiskunde met betrekking tot geldsteekproeven is weergegeven in een artikel waarin wordt verwezen naar de complexe wiskunde die een aantal grensgebieden van het extrapoleren van de bovengrenzen verklaart. Dit artikel verwijst de belangstellenden naar het betreffende onderzoek zonder de formules en wiskundige bewijzen weer te geven die noodzakelijkerwijs bij deze materie horen.

3.3 Artikel Steekproeven bij volledigheidscntroles

Omdat vaak wordt gesteld dat volledigheidscntroles niet met behulp van steekproeven kunnen worden gecontroleerd geeft dit artikel aan onder welke omstandigheden dat wel mogelijk is.

3.4 Artikel Isoleren: atypische of triviale afwijkingen

In de praktijk van het toepassen van steekproeven wordt vaak gezocht naar een mogelijkheid om na het vinden van afwijkingen zo snel mogelijk toch nog te kunnen goedkeuren. Daarom hebben wij een aantal columns van Ed Broeze over ‘isoleren van fouten’ gebundeld tot een artikel.

3.5 Artikel Uitbreiden om beter te schatten

Als vervolg daarop is de column van Paul van Batenburg opgenomen waarin de praktische mogelijkheid wordt geboden om een steekproef niet uit te breiden om te gaan goedkeuren, maar om de geaggregeerde afwijkingen in de betreffende populatie beter, nauwkeuriger, te kunnen schatten en op grond van die verbeterde schatting de klant de mogelijkheid te bieden de afwijkingen te corrigeren.

3.6 Artikel: Als ik niet kan goedkeuren; zes tips.

In het verlengde daarvan is de column van Paul van Batenburg opgenomen die tips geeft wat te doen als goedkeuring niet, direct, mogelijk is.

3.7 Artikel Evalueren van steekproeven

Drie columns van Ferry Geertman en Hein Kloosterman, waarin zij een praktisch voorbeeld hebben uitgewerkt met verschillende mogelijkheden om een geldsteekproef te evalueren, zijn als afzonderlijke paragrafen gebundeld in een artikel. Dat artikel wordt afgesloten met een column van Paul van Batenburg over een aantal vuistregels die bij de evaluatie van verschillende soorten steekproeven in acht kunnen worden genomen.

3.8 Artikel Goedkeuren, ondanks fouten

Dit artikel bestaat uit twee paragrafen. De eerste paragraaf door Paul van Batenburg refereert aan de COS die beschrijven dat een accountant bij het constateren van afwijkingen aan het management van de gecontroleerde moet voorstellen de projectie van die afwijkingen te corrigeren.

In de tweede paragraaf zetten Wouter Gerards en Hein Kloosterman uiteen hoe men in een situatie waar men afwijkingen verwacht de controle kan inrichten, als er sprake is van heterogene deelpopulaties.

3.9 Artikel Data-analyse en statistical auditing

Dit artikel bestaat uit vijf paragrafen die een eerder leven als column hebben gehad. De eerste paragraaf van Paul van Batenburg ontkracht de stelling dat data-analyse steekproeven overbodig maakt.

De tweede paragraaf – ook van Paul van Batenburg – laat een manier van data-analyse zien die hij GEAUZOMOFO heeft genoemd: het geautomatiseerd zoeken naar mogelijke fouten. De stelling is dat de accountant met GEAUZOMOFO geen informatie krijgt over de niet-onderzochte elementen.

In de derde paragraaf laten Ferry Geertman en Hein Kloosterman een praktijkvoorbeeld zien waar een verzameling mogelijke fouten was geïnventariseerd en toch een steekproef was getrokken en onderzocht. Illustratief is de toegevoegde waarde van die steekproef.

In de vierde paragraaf geven Geertman en Kloosterman voorbeelden van data-analyses waarmee

in de praktijk discriminantanalyses kunnen worden gemaakt.

Paul van Batenburg schreef de vijfde en laatste paragraaf. Hij stelt dat als Soll-positie en Ist-positie in bestandsvorm beschikbaar zijn en ondubbelzinnig te koppelen, er integrale controle mogelijk is. Niet zou die alleen mogelijk zijn, maar zelfs verplicht moeten worden.

3.10 Artikel Fraude en statistical audit

Dit artikel behandelt fraudeonderzoek. Geertman en Kloosterman behandelen de mogelijkheden van statistical auditing bij fraudeonderzoek zowel vanuit de gegevensgerichte detailcontrole, als vanuit data-analyse.

3.11 Artikel Postensteekproef om uitspraken in geld te doen

In het artikel over postensteekproeven om uitspraken in geld te doen schrijven Ferry Geertman en Hein Kloosterman over de condities waaronder dergelijke steekproeven toepasbaar zijn. Zij melden ook dat het niet mogelijk is geldsteekproeven te trekken en die te evalueren met de technieken voor postensteekproeven en andersom.

4 Gegegevensgerichte cijferanalyse met behulp van regressierekening

Paul van Batenburg geeft in dit artikel een kort overzicht van het boek van Stringer & Stewart, *Statistical Techniques for Analytical Procedures* (1986). Dit boek bevat de meest instructieve litera-

tuur met betrekking tot gegevensgerichte cijferanalyse door middel van statistische technieken.

5 Risicoanalyse in de accountantscontrole

De hier opgenomen artikelen ondersteunen de paragraaf over risicoanalyse die is opgenomen in het overzichtsartikel. Zij laten zowel zien dat er behoefte is aan het leren van de praktijk ten behoeve de verbetering van de theorie en daardoor weer de praktijk.

5.1 Het Bayesiaanse model ter ondersteuning van de controlemix

Paul van Batenburg en Roger Dassen beschrijven in dit artikel de mogelijkheid om bestaande modellen voor risicoanalyse – door vermenigvuldigen van risico's – falsifieerbaar te maken door in plaats van het spelen met kansen, te spelen met de worstcaseverwachtingen.

5.2 Wat is eigenlijk risicoanalyse in de accountantscontrole?

Risicoanalyse in de accountantscontrole heeft een lange geschiedenis. In dit artikel heeft Hein Kloosterman die geschiedenis geschetst. De huidige stand van zaken is een volwassen model voor het gegevensgerichte deel van een financiële controle. Dat model is gebouwd met kennis uit het vakgebied Statistical Audit. Het artikel beschrijft ook dat een deel van de werkzaamheden van een financiële controle niet met Statistical Audit is te verklaren. Dat deel noemt Kloosterman voor-

waardelijke controles. Het audit risk dat daarmee gepaard gaat en met bijvoorbeeld IT-audit is tot nu toe nog ontoereikend beschreven. Het artikel doet een poging daarvoor een oplossingsrichting te bieden.

5.3 Doeltreffende administratieve controle door steekproeven

Ruud Veenstra en Arnold Heertje beschrijven methoden en middelen om tot een doeltreffende administratieve controle te komen. Administratieve controle door middel van steekproeven kan de doeltreffendheid, doelmatigheid en beheersing van waardenstromen en informatiestromen aanzienlijk verbeteren en de onaantastbaarheid van de verslaglegging daarover garanderen. Dit artikel draagt de bouwstenen daarvoor aan. Met deze bouwstenen kunnen de essentiële voorwaarden worden vervuld voor herstel van vertrouwen in de integriteit van het financieel-economische verkeer. Het Limpergiaanse concept van de volkomen controle en daarop gebaseerde steekproefmethoden voor accountantscontrole bieden hiervoor een stevige grondslag. Daarnaast is een belangrijke doorbraak bereikt door administratieve controle gebaseerd op de voor de industriële kwaliteitscontrole ontworpen AOQL-steekproefmethode (AOQL = Average Outgoing Quality Limit).

5.4 Valideren van risico-inschattingen in accountantscontrole (smv)

Ed Broeze heeft de samenvatting van zijn dissertatie over het valideren van risico-inschattingen

beschikbaar gesteld. Belangrijkste conclusie van de dissertatie is dat een organisatie alleen mag aannemen dat zijn risico-inschattingen valide zijn, als hij dit ook empirisch heeft vastgesteld. Broeze kon niet aannemelijk maken dat de tot nu toe gebruikte inschattingen van inherent risico, interne-beheersingsrisico en cijferanalyserisico het 'werkelijke risico' geven.

5.5 Rekenen met plaatjes: een nieuwe kijk op steekproeven

Paul van Batenburg schreef een samenvatting van de dissertatie van Trevor R. Stewart. Hij laat de essentie van het proefschrift zien met behulp van plaatjes. De essentie van het onderzoek van Stewart is dat de eerdere aanname dat deelpopulaties als homogeen kunnen worden beschouwd, losgelaten kan worden. Vanaf nu kan er ook gerekend worden met onafhankelijke, heterogene, deelpopulaties. Dat betekent niet dat alles eenvoudiger wordt, maar er zijn betere uitspraken mogelijk in het geval van onderzoeken naar afzonderlijke groepsonderdelen.

5.6 Het 'push left'-principe als motor voor data analytics in de accountantscontrole

De toepassing van data analytics volgens het 'push left'-principe biedt drie voordelen: Ten eerste heeft de accountant meer inzicht in de kans dat hij een goedkeurende verklaring afgeeft, terwijl de jaarrekening toch een materiële fout bevat.

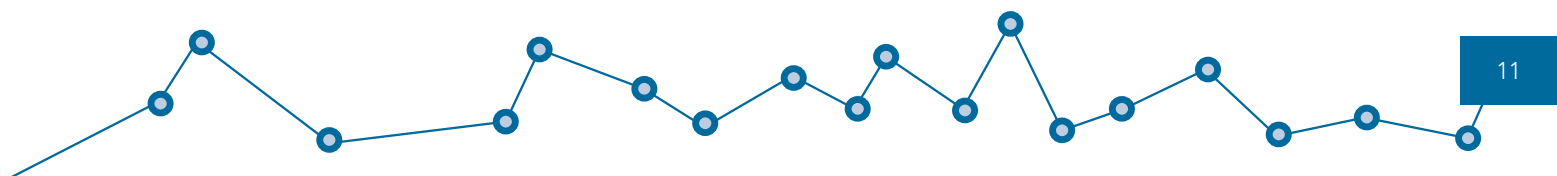
Ten tweede is hij in staat om te garanderen dat zijn controlemix optimaal is: gegeven de ver-

wachting die hij vooraf uitspreekt is het statistisch onmogelijk om met minder inspanning meer zekerheid te verkrijgen.

Ten derde vergroot het 'push left'-principe de relevantie van de accountant door meer dan alleen zekerheid te bieden. Het 'push left'-principe kan eenvoudig gekwantificeerd worden door een nieuwe Bayesiaanse methode genaamd Beta. Deze positionering past binnen de controlestandaarden, hoewel aanscherping van de standaarden soms wenselijk is.

5.7 Het Bayesiaanse model en Tax Assurance

In dit artikel maken Paul van Batenburg en Robbert Hoyng een praktische toepassing met het model dat door Trevor Stewart in zijn proefschrift is beschreven. Hun toepassing geldt fiscale controles. De auteurs verwijzen naar Excelmodellen waarmee het rekenwerk kan worden ondersteund.





Inleidende artikelen

Inleiding Essaybundel Statistical Auditing

Paul van Batenburg en Hein Kloosterman

Inleiding

Centraal thema van deze bundel is de kwaliteit van de accountantscontrole. Dit begrip staat volop in de belangstelling en is direct gekoppeld aan de vraag welke fouten of onregelmatigheden de accountant zou moeten zien in de accountantscontrole of wat hij misschien niet kan zien. De accountantscontrole vindt plaats met een bepaalde materialiteit, de omvang van de maximale fout die de jaarrekening kan hebben.

In dit artikel zetten wij het controleproces als geheel uiteen. Wij delen het proces op in interne beheersing, gegevensgerichte detailcontrole en de gegevensgerichte cijferbeoordeling. In aparte paragrafen besteden wij aandacht aan materialiteit en risicoanalyse in de accountantscontrole. Met name op dat laatste gebied kunnen de tot nu toe gebruikte modellen met behulp van wetenschappelijk onderzoek worden verbeterd.

De standaarden waaraan een controle dient te voldoen geven de consument in beperkte mate inzicht in wat er van de financiële controle als uitkomst – anders dan een abstracte verklaring – mag worden verwacht. De meeste studenten

accountancy hebben geleerd dat een goedkeurende verklaring bij een kwantitatief overzicht (zoals de jaarrekening er een is) vereist dat de accountant een *deugdelijke grondslag*, oftewel toereikend bewijs, heeft voor de stelling: “Ik heb voldoende werk gedaan om voldoende zeker te merken als de jaarrekening een materiële fout zou bevatten, maar dat gebeurde niet.” Vanuit dit perspectief hebben wij hierna het controleproces als geheel uiteengezet en a priori geredeneerd alsof de hoeveelheid gegevensgerichte detailcontroles niet gereduceerd zou hoeven te worden.

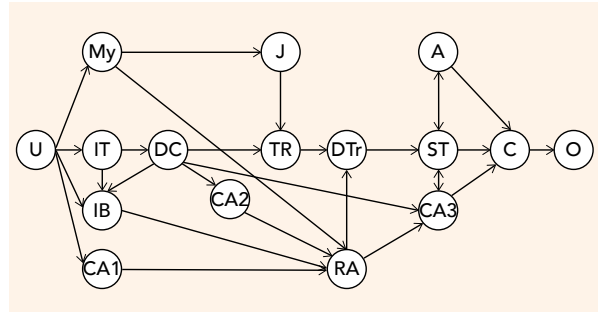
Dit artikel bevat veel verwijzingen naar wetenschappelijke literatuur. Daarmee proberen wij teksten met wiskundige formules te voorkomen. Wij proberen met die verwijzingen ook te laten zien wat de stand van zaken is met betrekking tot het wetenschappelijk onderzoek op het gebied van de kwaliteit van het proces van de accountantscontrole.

Daarom sluiten wij dit overzichtsartikel af met een samenvatting van wat er nog allemaal te leren valt op welk gebied van de controle en waartoe dus wetenschappelijk onderzoek te hulp kan schieten. Volgens de Stuurgroep Statistical Auditing kan de kwaliteit van het auditproces

pas worden verbeterd als uit (de feilen van) het auditproces lering wordt getrokken. Zoals ook bij industriële processen het geval is, dient een verbetering zich door te gaan aan wanneer er fouten, afwijkingen, onzuiverheden of omissies worden ontdekt: het product voldoet niet aan de eisen die door de belanghebbenden aan het product zijn gesteld.

Het controleproces als geheel

De aanpak van de jaarrekeningcontrole zoals de meeste kantoren toepassen, bevat in grote lijnen een serie beslissingen die tot de conclusie leidt (in het gunstige geval) dat de jaarrekening een 'getrouw beeld' geeft. In het huidige tijdgewricht houdt dat in dat er sprake was van kennis van de organisatie (U), kennisverwerving over in het oog springende aspecten (U plus CA1), het beoordelen van de interne beheersing (IB, IT, DC en CA2), inclusief de IT die daarbij een rol speelt, het verrichten van gegevensgerichte werkzaamheden zoals gegevensgerichte cijferanalyses (CA3) en gegevensgerichte detailcontroles (ST op een deeltrekking DTr van een grotere steekproef (TR)). Wij hebben het controleproces als volgt getekend.



Figuur 1

Met onder andere deze knooppunten kunnen de beslissingsprocessen van de meeste financiële controles worden beschreven.

In figuur 1 is een stel kritische gebieden aangegeven. Te weten: de beoordeling van de onvervangbare interne beheersing (IB en IT in onderlinge samenhang), het toepassen van risicoanalyse (RA is het gebruiken van voorinformatie), het trekken (TR en DTr), controleren en evalueren van gegevensgerichte steekproeven (ST), en tot slot het uitvoeren en evalueren van gegevensgerichte cijferanalyses (CA3).

Vaststelling materialiteit als onderdeel van het controleproces

Het vaststellen van de materialiteit voor de jaarrekening als geheel, de uitvoeringsmaterialiteit en de toelaatbare afwijking gaat gepaard met professional judgment. Dat impliceert een aanzienlijke hoeveelheid subjectiviteit. Op zich is dat geen

probleem, als maar die grootheid achteraf valt te valideren of falsifiëren.

Deze bundel bevat het artikel van Van Nieuw Amerongen en Van Batenburg (2010) over het vaststellen van materialiteit, uitvoeringsmaterialiteit en toelaatbare afwijking. In de paragraaf over het leereffect van de praktische toepassing van controleactiviteiten brengen wij naar voren dat materialiteit een grootheid is die, ten opzichte van het controleproces, exogeen is. Daarom is het nodig om van ‘het maatschappelijk verkeer’ te leren hoe de nauwkeurigheid van het – individuele – onderzoek zou moeten zijn.

Interne beheersing

Onvervangbare interne beheersing; toereikend?

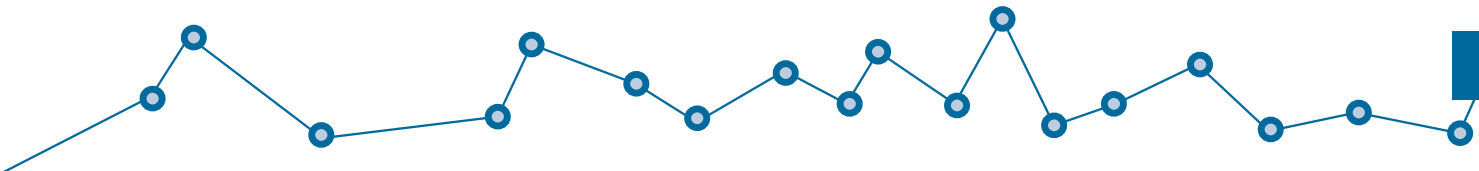
De onvervangbare interne beheersing is van wezenlijk belang voor de accountantscontrole. Als er geen conclusie is over de onvervangbare interne beheersing is er geen oordeel over de jaarrekening als geheel mogelijk. De opzet van die onvervangbare interne beheersing, het bestaan ervan én de werking, moeten ‘goed’ zijn. En het is nodig om daarna met een combinatie van gegevensgerichte detailcontroles en gegevensgerichte cijferanalyses tot een toereikend onderbouwd goedkeurend oordeel te kunnen komen. De uitruil van gegevensgerichte maatregelen en organisatiegerichte maatregelen is mogelijk, maar de ‘wisselkoers’ is subjectief (professional judgement) en de standaarden geven grenzen aan waarbinnen die uitruil kan plaatsvinden.

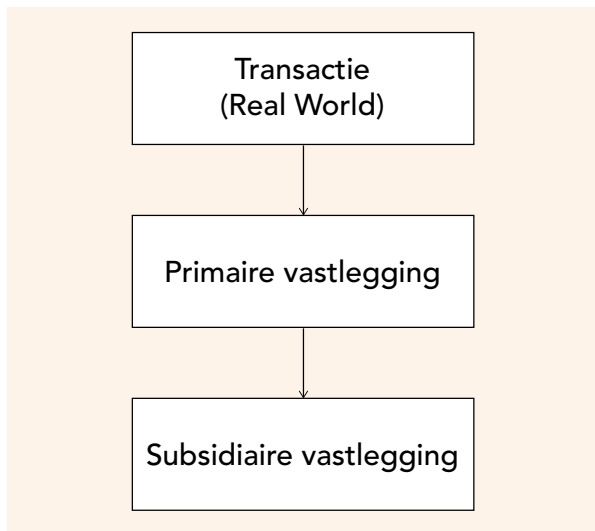
Voor de volledigheid: het vaststellen dat de opzet van de onvervangbare interne beheersing goed genoeg is, dat die opzet heeft bestaan en dat die heeft gewerkt kan leiden tot een reductie van de hoeveelheid gegevensgerichte detailcontroles. Maar het doen van meer gegevensgerichte detailcontroles mag geen enkel effect hebben op de werkzaamheden om opzet, bestaan en werking van de onvervangbare interne beheersing vast te stellen.

Blokdijk heeft over de onvervangbaarheid van de interne beheersing een groot aantal artikelen geschreven¹.

Een van de lastige elementen van de beoordeling van de interne beheersing is dat er geen, of misschien maar weinig, theorie beschikbaar is om de toereikendheid van de werkzaamheden vast te stellen. Wel zijn er oorzaken aan te wijzen waardoor het lastig is de conclusie te trekken dat de interne beheersing van een organisatie toereikend is of niet (vaak is de redenering intuïtief). Een van die oorzaken is de dynamiek in een organisatie, een andere is dat het moeilijk is één en slechts één Soll-positie voor de interne beheersing op een organisatie te plakken.

In figuur 2 hebben wij het transactiemodel weergegeven. Dat model schetst de interne beheersing van het ontstaan van een transactie, het primair vastleggen ervan en verwerken van de gegevens ervan in enige vorm van administratie.





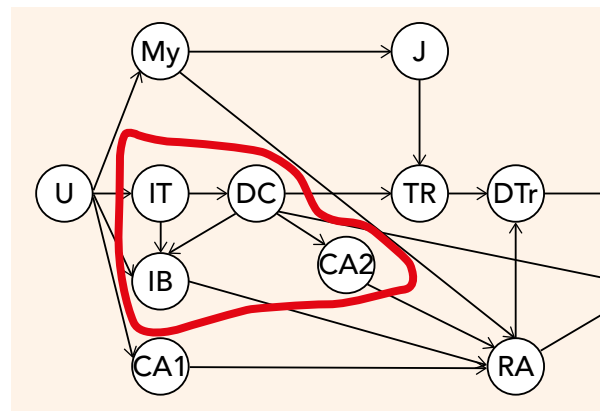
Figuur 2

Met dit model wordt de plaats van het onvervangbare deel van de interne beheersing duidelijk gemaakt. De overgang van de transactie die zich afspeelt in de echte wereld naar de primaire vastlegging van die transactie is kritisch. Achteraf kan niet meer op basis van (de individuele) vastleggingen – die zijn er immers niet meer – worden gereconstrueerd wat de primaire vastlegging had moeten zijn.

In de Nederlandse theorie is een belangrijke plaats ingeruimd voor de onderlinge verbanden van groepen van transacties (verbandcontroles)². Met die verbanden, meestal aangeduid als waar-denkringloop, kunnen omissies en inconsistenties worden aangetoond. De kracht van verbandscontroles is tevens gelegen in het kunnen kwantificeren van de mogelijke omissies.

4.2 Beoordeling interne beheersing binnen het controleproces

De beoordeling van de onvervangbare interne beheersing is als volgt geschetst.



Figuur 3

In het omliggende gebied staan de knooppunten, de elementen, die te maken hebben met de onvervangbare interne beheersing. Dit zijn het beoordelen van de interne beheersing (IB) en de interactie met de Informatietechnologie (het knooppunt IT) van de organisatie waarbij gegevens uit de IT worden verzameld (DC = datacollectie) om cijfer- of data-analyses (CA2) te verrichten.

De uitkomst van het gebied van de hele interne beheersing is informatie voor de risicoanalyse (het Audit Risk Model, al dan niet ondersteund door Bayesiaanse statistiek; hier aangeduid met RA). De accountant moet in dat gebied de IT (informatietechnologie) en de IB (interne beheersing)

in samenhang beoordelen. Die beoordeling moet opzet, bestaan en werking omvatten. De IT maakt datacollectie mogelijk (DC). Deze verzamelde data kunnen gebruikt worden voor geavanceerde cijfer- en data-analyses (CA2) en voor het trekken van steekproeven (TR en DTr) ten behoeve van de gegevensgerichte detailcontroles. Onder die CA2 begrijpen wij onder meer technieken voor data-analyse zoals process mining en verbandscontroles (NBA 2018). Wij vinden ook dat compliance tests – als dat geen verbandscontroles zijn – en een techniek als AOQL tot de technieken horen die bij het knooppunt CA2 kunnen worden gerekend. De betekenis van de informatie die volgt uit de beoordeling van IT en IB is verstrekkend. Het beslissingsproces ervan is nog weinig geëxpliciteerd. Verderop, bij de behandeling van risicoanalyse (RA) komen we daarop terug en op het lerend vermogen dat ook dit deel van de controle zou moeten hebben.

Interne beheersing en leereffect

Hiervoor hebben wij genoemd dat professional judgement deel uitmaakt van de beoordeling van het toereikend zijn van de interne beheersing (inclusief de IT die daarbij van toepassing is). Er is dus behoefte aan een expliciet en falsifieerbaar model om vast te stellen dat de mate van effectiviteit van de onvervangbare interne beheersing voldoende is. Er zijn ons geen onderzoeken bekend over het vaststellen van het toereikend zijn van het beoordelen van de interne beheersing met middelen als believe theory of iets dergelijks.

Ook kennen wij geen pragmatische stopregel voor onderzoeken naar IT, interne beheersing en de combinatie van beide. Wel kennen wij een oproep daartoe in bijvoorbeeld Kloosterman (2013)³.

De gegevensgerichte detailcontrole

Samenvatting en wat er te leren valt

Deze paragraaf gaat alleen over de gegevensgerichte controles zelf en niet over risicoanalyse in de zin van de reductie van de hoeveelheid gegevensgerichte controles op basis van risicoanalyse. Dat onderwerp komt separaat aan bod. Wij willen toewerken naar hetgeen te leren valt uit de hedendaagse toepassing van detailcontroles en de rol van wiskundige steekproeven daarbij.

De meeste elementen uit het overzichtsartikel van Ed Broeze et al. (2003) zijn inmiddels onderzocht en verbeterd. Ook zijn al veel elementen operationeel gemaakt en die worden in de praktijk toegepast. Zij het vooral bij bijzondere onderzoeken en in mindere mate bij de jaarrekeningcontrole.

Wat nog te leren valt is in de eerste plaats de keuze voor het te hanteren controlerisico, het audit risk. Accountants plegen dat in het algemeen te stellen op 5%.

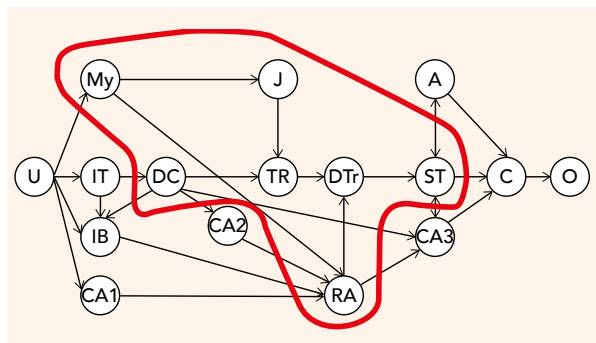
Het is naar ons beste weten nog niet onderzocht of er omstandigheden zijn waaronder een ander controlerisico gewenst is. Bijvoorbeeld bij fraudeonderzoek.

Het bepalen van de materialiteit gaat gepaard met veel professional judgement. Omdat materialiteit eigenlijk een kwaliteitseis voor de controle is die

ten opzichte van de controle exogeen is, moeten accountants willen leren hoe groot de omvang van de onnauwkeurigheid van het individuele onderzoek zou mogen zijn. Ten slotte valt er nog te leren over de kritieke grens tussen het toepassen van geldsteekproeven en het toepassen van postensteekproeven (om te schatten).

De elementen van de gegevensgerichte detailcontrole

Op het symposium van 2017 hebben we het toepassen van steekproeven (inclusief de risicoanalyse) door middel van een omlijning in het netwerkmodel aangegeven.



Figuur 4

Om de omvang van de hoeveelheid gegevensgerichte detailcontroles te bepalen is het vaststellen van de materialiteit (My) nodig⁴. Er moet (meestal) datacollectie (= DC) plaatsvinden en doorgaans past men risicoanalyse (ARM of beter: knooppunt RA) toe. Verder zal men de gegevensgerichte

steekproef trekken (met de knooppunten J, TR en DTr; waarbij J staat voor bepalen interval, TR voor trekking en DTr voor een deeltrekking n.a.v. risicoanalyse) en de elementen van de steekproef (substantive tests = ST) controleren. Op het symposium zijn de grote lijnen geschetst van de beslissingsmodellen die in dit gebied zijn toegepast. Die modellen gaan over het plannen en evalueren van steekproeven, de keuze van geldsteekproeven of postensteekproeven, het al dan niet hoeven toerekenen van materialiteit of toelaatbare afwijking⁵ en het combineren van verschillende soorten 'bewijs'. Met enige diepgang is op het symposium de evaluatie van geldsteekproeven vergeleken met die van postensteekproeven. Er is veel gepubliceerd over risicoanalyse (RA). Omdat deze term een homoniem is, levert die steeds weer misverstanden op. Soms worden alle mogelijke betekenissen door elkaar geklutst in een artikel. Verderop in dit artikel geven wij vanwege de complexiteit van dit onderwerp in een aparte paragraaf de huidige stand van zaken.

De gegevensgerichte (detail)controles met geldsteekproeven

De praktische kant van geldsteekproeven

De columns van de Stuurgroep gaan vaak over de gegevensgerichte detailcontroles. Meestal zijn die detailcontroles ingericht met behulp van wiskundige steekproeven, geldsteekproeven. Op een enkele column na gaan die columns over steekproeven, zonder dat er rekening is gehouden met een model voor risicoanalyse om de hoeveelheid

werk te beperken. Wij komen op het onderwerp risicoanalyse in een aparte paragraaf terug.

De stand van zaken van de Informatietechnologie (IT) laat toe om ervan uit te gaan dat de data van een organisatie (zoals hiervoor in het transactie-model weergegeven als subsidiaire vastleggingen) volledig elektronisch ter beschikking zijn voor de controle. Als gevolg daarvan kunnen die data bewerkt worden met behulp van audit software. Voor de gegevensgerichte controle is het trekken van geld- of postensteekproeven een relatief simpele bewerking. Omdat ook het uitvoeren van allerlei cijferbeoordelingen en data-analyses vrijwel steeds binnen de mogelijkheden ligt, worden steekproeven met enige regelmaat als overtoollig gezien, omdat immers alle data beschikbaar zouden zijn. In meerdere columns hebben wij aandacht aan die gedachte besteed.

Als de Soll-positie werkelijk in bestandsvorm beschikbaar is, zou een steekproef niet nodig moeten zijn en kan integraal worden gecontroleerd. De Stuurgroep is van mening dat integrale controle dan ook de norm zou moeten zijn. Steekproeven zouden in feite alleen nodig zijn als de Soll-positie (door tellen, meten, narekenen, raadplegen van medewerkers of het bij elkaar zoeken van papieren) niet digitaal beschikbaar is. Echter, bij een bestand met Soll-posities is natuurlijk wel de vraag waar, wanneer en door wie dat bestand gemaakt is.

Dat neemt niet weg dat data-analyse verdienstelijk is, bijvoorbeeld als er informatie is over

deelpopulaties met een grotere foutdichtheid. Met behulp van data-analyse kunnen dan die 'vervuilde' deelpopulaties worden ontdekt en daarna desgewenst⁶ aan een meer intensief onderzoek worden onderworpen.

De theoretische kant van geldsteekproeven

In het overzichtsartikel van Broeze et al. (2003) is een aantal prangende vragen over statistiek in relatie tot detailcontroles genoemd, en is aangeduid of en in hoeverre wetenschappelijk onderzoek die vragen heeft beantwoord.

De vragen die in dat artikel zijn genoemd zijn:

- (Hoe) moet men de controletolerantie⁷ aan onderdelen van een financieel overzicht toerekenen?
- Hoe moeten onderzoeksresultaten vertaald worden van delen naar het geheel van een financieel overzicht?
- Is de Stringer-bound wiskundig bewijsbaar conservatief?
- Zijn er methoden om het conservatisme van de Stringer-bound te reduceren?
- Onder welke voorwaarden kan men een uitspraak in geld doen op basis van een postensteekproef?

Op al deze vragen is in de loop der jaren antwoord gekomen. Wij nemen in deze bundel een aantal essays op (of een samenvatting ervan) waarin de antwoorden zijn beschreven. In deze inleiding volstaan wij met een paar uitkomsten – kort weergegeven – die de planning van een audit

aanzienlijk vereenvoudigen. De eerste vier vragen gaan over geldsteekproeven. De eerste vraag lijkt betrekkelijk eenvoudig. In onze modelbouw behandelen we die ook zo. Het wordt pas complex als er noodgedwongen meerdere verschillende nauwkeurigheden in de controle voor een financieel overzicht als geheel worden gebruikt.

Blokdijk (1996) heeft bewijs geleverd dat, gegeven steekproeven met nul verwachte fouten, de controletolerantie (in de zin van toelaatbare afwijking) niet hoeft te worden toegerekend aan de verschillende deelpopulaties die deel uitmaken van een financieel overzicht om een oordeel over dat financieel overzicht als geheel te krijgen. Men mag wel toerekenen, maar dan neemt de complexiteit van de berekeningen toe.

Achteraf kan dat toerekenen juist een verbetering zijn omdat met het toerekenen dilemma's als isolatie van fouten kunnen worden voorkomen. In Stewart (2013) is de evaluatie uitgewerkt van steekproeven op concernniveau. Daarbij is er sprake van niet-homogene deelpopulaties. Het niet-homogene heeft te maken met het oordelen over deelpopulaties. Er is op dit terrein dus sprake van voortschrijdend inzicht; vergelijk Popper (1979).

Wordt bij de planning van steekproeven vaak ervan uitgegaan dat vrijwel geen fouten zullen worden gevonden, bij de evaluatie blijkt dat nogal eens anders te zijn. Er is naarstig gezocht naar wiskundig bewijs dat de Stringer-bound – de meest gebruikte bovengrens voor uitkomsten van

geldsteekproeven – conservatief is. Onder conservatief verstaan we dat de bovengrens te hoog wordt berekend zodat te snel niet goedgekeurd kan worden. Uit het onderzoek dat Pap et al. (1995) hiernaar hebben gedaan bleek dat die stelling inderdaad geldt. Echter, als het steekproefrisico groter is dan 50% slaat het conservatisme om in onvoorzichtigheid.

In het overzichtsartikel van Broeze et al. (2003) zijn meerdere onderzoeken genoemd om het conservatisme van de Stringer-bound het hoofd te bieden. De meeste van die studies zijn gedaan met behulp van simulaties. Wij noemen de onderzoeken van Schapendonk (1998) en van Lucassen et al. (1996). Schapendonk (1998) heeft een groot aantal varianten op de Stringer-bound onderzocht. Lucassen et al. (1996) hebben onderzocht of het vervangen van de volgorde van afnemende taintings door de volgorde van afnemende absolute-foutbedragen houdbaar is.

De meer theoretische benadering die door De Jager et al. (1996) is geboden is vooral toepasbaar als een steekproef veel fouten bevat. Hij berekent dan een aftrek op de Stringer-bound gebaseerd op de centrale limietstelling.

Van de onderzoeken die erop zijn gericht om op een valide manier het conservatisme van de Stringer-bound te reduceren lijkt de methode van Lucassen et al. (1996) voorsnag het minst conservatief en bovendien nog steeds effectief: de minimaal vereiste 95% betrouwbaarheid wordt bereikt.

Ook Leslie et al. (1979) geven een minder conservatief alternatief voor de Stringer-bound namelijk de zogenaemde DUS Cell evaluation⁸.

Over het werken met postensteekproeven en het uitdrukken van de bevindingen in geld gaat de volgende paragraaf van dit artikel.

Postensteekproeven

De vraag wordt vaak gesteld: “Kan ik een postensteekproef trekken en de uitkomsten op een valide manier in geld vertalen?” Zoals uit het overzichtsartikel (Broeze et al., 2003) blijkt zijn er meerdere manieren bedacht om te proberen de uitkomsten van een steekproef op posten, naar een uitspraak in geld te vertalen. Als er sprake is van (deel-)populaties met een geringe hoeveelheid fouten worden doorgaans postensteekproeven als niet zinvol aangemerkt⁹. Zodra er sprake is van een populatie met veel fouten komen methoden die vaak worden aangeduid als ‘variables sampling’ of ‘estimation sampling’, in aanmerking. In het Nederlands heet deze soort steekproeven schattingssteekproeven.

Schattingssteekproeven kennen meerdere varianten, te weten: de directe schatter, de ratioschatter, de verschillenschatter en de regressieschatter. Deze laatste is eigenlijk de meest algemene vorm van de vier hier genoemde. Deze technieken worden daarom ook wel aangeduid als ‘de familie van de regressieschatter’.

Het onderzoek van Willemsen (1997) heeft laten zien dat als een populatie niet voldoet aan de

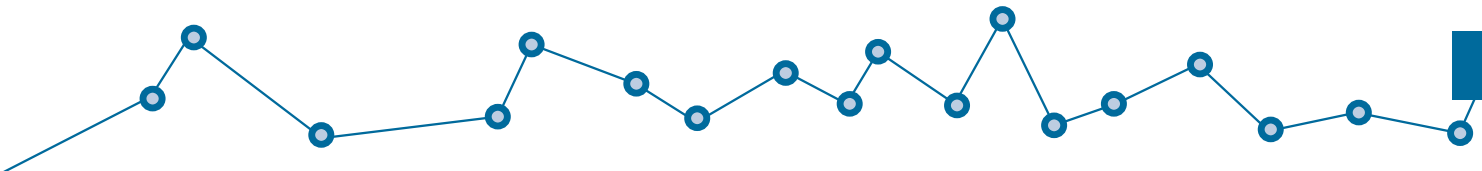
eisen die gelden voor het toepassen van een schattingssteekproef uit de familie van de regressieschatter, een andere manier van benaderen, in casu met behulp van bootstrapping-methodes wel betere resultaten levert dan zonder die techniek, maar dat desondanks meestal niet de nominale betrouwbaarheid wordt gerealiseerd.

Uit het voorgaande volgt dus dat bij een geringe verwachte foutdichtheid in een populatie een postensteekproef niet efficiënt gebruikt kan worden om de fout in de populatie te schatten.

Uit de literatuur over de toepassing van de regressieschatter¹⁰ komen restricties naar voren. Zo moet men een steekproef trekken van ongeveer 200 waarnemingen, 5% tot 10% van de waarnemingen moeten een afwijking laten zien en moet het patroon van de afwijkingen symmetrisch zijn.

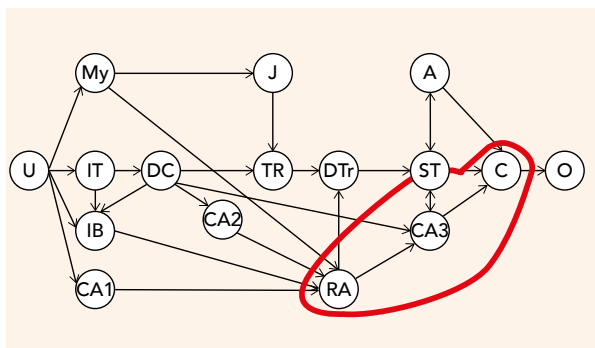
Hier zien we een belangrijk verschil met geldsteekproeven. Bij geldsteekproeven leidt een te kleine steekproef tot de conclusie dat er nog niet kan worden goedgekeurd ook al zijn de onderzochte elementen allemaal goed. Er moet dus ten minste zoveel worden gecontroleerd dat er als de steekproef geen enkele fout bevat, er ook goedgekeurd kan worden.

Bij de postensteekproef moet men ten minste die 200 elementen beoordelen, en moeten de uitkomsten aan de overige eisen voldoen, omdat anders de uitkomsten, de schattingen, geen of onvoldoende betekenis hebben.



Gegevensgerichte cijferanalyses

Gegevensgerichte cijferanalyses kennen maar een paar publicaties. Daarvan is de belangrijkste van Stringer & Stewart (1985). Die cijferanalyses zijn in de figuur aangeduid als CA3.



Figuur 4

Een goede manier waarop gegevensgerichte cijferanalyses uitgevoerd zouden kunnen worden, is uitgebreid beschreven in het boek van Stringer en Stewart (1985).

Verder is er weinig wetenschappelijke literatuur bekend over hoe gegevensgerichte cijferanalyses het best kunnen worden uitgevoerd. De manier van uitvoering zoals Stringer en Stewart die beschreven houdt impliciet een waardering in van de hoeveelheid evidence die de techniek met zich meebrengt.

Wij hebben gezocht naar literatuur die de techniek van gegevensgerichte cijferanalyse ondersteunt. Wij geven een paar voorbeelden van wat wij aantreffen.

Loebecke en Steinbart (1987) schrijven dat initiële cijferanalyse niet bruikbaar is voor de reductie van gegevensgerichte detailcontroles.

Kinney (1979) deed onderzoek in hoeverre een initiële cijferbeoordeling voorspelde dat een bijstelling van de te publiceren gegevens nodig was, en welke soort gegevens het betrof in de onderzochte gevallen. Kinney concludeert dat die cijferanalyse zowel leidt tot vals positieven als vals negatieven als de initiële cijferanalyse rekening houdt met de uitkomsten van de audit van het vorige jaar.

Burger et al. (2012) onderzochten de invloed van gegevensgerichte cijferanalyses op de keuze van de hoeveelheid te verrichten gegevensgerichte detailcontroles. Volgens het onderzoek passen accountants de hoeveelheid detailwaarnemingen aan als de cijferanalyses significante uitkomsten leveren. Onder significant moet dan worden verstaan dat cijferanalyse concludeert dat de te beoordelen populatie mogelijk een materiële fout bevat. Verwacht was dat bij niet-significante uitkomsten er voorinformatie aan de cijferanalyse zou worden ontleend.

Wij nemen in deze bundel een essay op over de manier op Stringer en Stewart de techniek hebben beschreven. De techniek past regressieanalyse toe op gecontroleerde gegevens; bij voorkeur in slechts geringe mate geaggregeerde gegevens. Daarmee wordt een reeds beredeneerd verband 'hard' gemaakt. Dat significante verband wordt

vervolgens gebruikt om datastromen in een nieuwe periode te onderzoeken. Als de waarden van de gepresenteerde nieuwe stroom liggen binnen het betrouwbaarheidsinterval van de oude periode, beschouwt men die waarden van de nieuwe periode als aanvaardbaar, rekening houdend met de uitvoeringsmaterialiteit voor de betreffende deelpopulatie. Liggen de nieuwere waarden buiten het betrouwbaarheidsinterval dan is de gegevensgerichte controle over deze geldstroom niet afgerond, en moet daarover alsnog een gegevensgerichte detailcontrole plaatsvinden.

Risicoanalyse en financial auditing

Risicoanalyse is een onderwerp dat vaak de gemoederen verhit. Dat heeft met verschillende zaken te maken. In de eerste plaats is risicoanalyse – net als risico – geen eenduidig begrip. In de tweede plaats is het Audit Risk Model (ARM) niet uit wetenschappelijke studie ontstaan maar uit pragmatiek, waardoor het model besmet is met emotionele uitingen en dito reacties daarop. Steele (1992) laat zien dat het ARM een slecht voorbeeld is van een Bayesiaans model. Van Batenburg en Dassen (1996) hebben beschreven hoe men het ARM zou kunnen valideren. Later herhaalde Kloosterman (2004) die beschrijving. Stewart (2013) laat zien dat het ARM alleen werkt als men uitgaat van steekproeven met nul fouten. De zoektocht naar verbetering van het Audit Risk Model door te verwijzen naar bedrijfsrisico's heeft veel te maken met het niet eenduidig hanteren

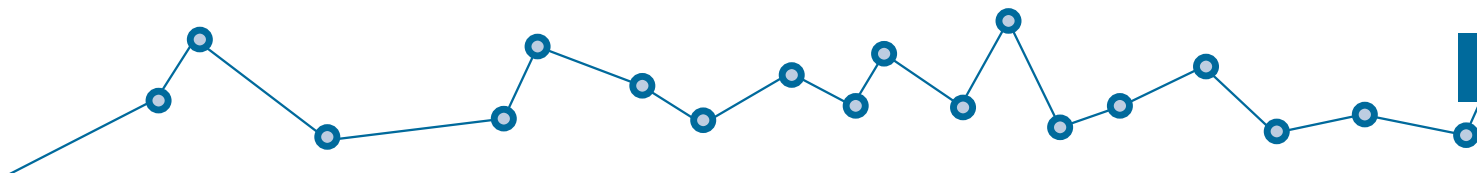
van begrippen als risico en risicoanalyse. Het Audit Risk Model (ARM) is door middel van Bayesiaanse vertalingen ervan rijp geworden om van de toepassing in de praktijk te leren. Er is onzes inziens onderzoek nodig om te laten zien wanneer het ARM leidt tot te optimistische risicoschattingen, en dus te weinig gegevensgerichte detailcontroles.

Het begrip risico

In deze inleiding hanteren wij het begrip 'risico' als een kansbegrip. En dan vooral als de kans op ten onrechte goedkeuren van financieel overzicht dat niet goed is. De term 'risico' is echter een homoniem. Soms betekent risico een kwade kans, een kans op een nadeel. Soms betekent risico het nadeel zelf. Soms betekent risico de aard van het mogelijke nadeel en soms duidt men risico aan als de kans op een soort effect maal het financiële gevolg van dat effect. Door deze slordige aanduiding van risico ontstaan er ook slordigheden in de definiëring van het werk van accountants.

(Ten minste) drie soorten risicoanalyse in financial auditing

In deze bundel gebruiken wij het begrip risicoanalyse in relatie tot de onzekerheid in de accountantscontrole. Hierover gaat de volgende subparagraaf. Binnen de toepassing van het Audit Risk Model zien we bij de meeste kantoren nog een type 'risicoanalyse': dat is het beoordelen of er sprake is van een laag inherent risico (op het aanwezig zijn van fouten in de jaarrekening). Bij veel kantoren bestaat de tendens geen separaat



oordeel uit te spreken over inherent risico en interne-beheersingsrisico.

Daarnaast gebruiken accountants nog twee typen risicoanalyses. De eerste variant is de analyse van de soorten gebeurtenissen die het nastreven van de doelen van een organisatie kunnen belemmeren. De tweede variant is de analyse van de bedreigingen ten aanzien van de opbrengstverantwoording en de verantwoording van de kosten. Deze variant wordt soms aangemerkt als een detaillering van de eerste.

Risicoanalyse om de hoeveelheid gegevensgerichte detailcontroles te reduceren

Het Audit Risk Model – ook wel het Audit Assurance Model genoemd – is bedoeld om de hoeveelheid gegevensgerichte detailcontroles te beperken. Het oorspronkelijke model heeft vooral een pragmatische grondslag. Desondanks – of juist daardoor – is het model behoorlijk in zwang geraakt. Het oorspronkelijke model is een simpele vermenigvuldiging van risico's. Het nadeel van die manier van modelbouw is dat dit vermenigvuldigingsmodel niet gevalideerd kan worden. De ervaring die met het model wordt opgedaan kan dus niet ter lering worden gebruikt¹¹.

Dat is ook Steele (1992) opgevallen. Hij stelt dat de vormgeving van het Audit Risk Model (ARM) gebaseerd is op de invalshoek dat het experiment vele malen kan worden herhaald. Nu het gaat om één (en slechts één) controle en daarover een uitspraak moet worden gedaan ligt het voor de hand een statistische methode te kiezen die een

dergelijke subjectiviteit aankan. Die methode van subjectieve kansen is de statistiek volgens Bayes (of: Bayesiaanse statistiek).

Eind tachtiger en begin negentiger jaren hebben Van Batenburg et al. (1991) in Nederland een aanzet gegeven tot het betrekken van Bayesiaanse statistiek in het model. Het Bayesiaanse model in die vorm is nog niet valideerbaar. Later – in 1996 – hebben Van Batenburg en Dassen (1996) de Bayesiaanse variant aangescherpt en worden de 'risico-inschattingen' omgezet in schattingen van de maximale fout (waarvoor men de hand in het vuur wil steken dat die niet wordt overschreden). Dit Bayesiaanse RisicoAnalyseModel (BRAM) biedt de mogelijkheid om de effecten van een audit die met een ARM is uitgevoerd om te rekenen tot de aannames die zijn gemaakt omtrent de maximale fouten waarvoor men bereid was de hand in het vuur te steken. BRAM is het eerste model dat toestaat dat ervaringen uit het verleden met het Audit Risk Model leerstof voor de toekomst kunnen worden.

De modellen voor reductie van de hoeveelheid gegevensgerichte controles gaan, zij het impliciet, uit van het vinden van geen enkele fout in de steekproef. Dat geldt zowel voor het pragmatische model¹² als voor de overige modellen (zie ook Stewart (2013)).

Onderzoek naar het toepassen van risicoanalyse in de praktijk

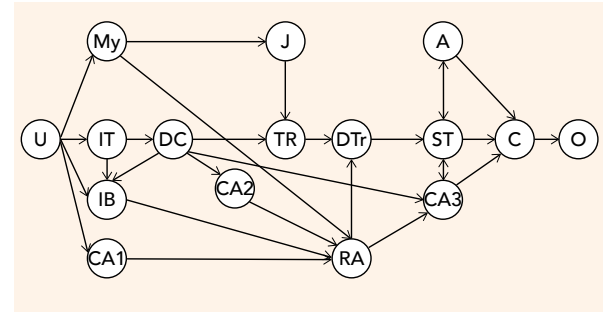
Broeze (2006) heeft onderzoek gedaan op het terrein van het toepassen van risicoanalyses in financial auditing. Een belangrijke conclusie van dat onderzoek is dat een organisatie alleen mag aannemen dat de risico-inschattingen valide zijn, als dit empirisch is vastgesteld. Een van de bevindingen is dat er reducties op de hoeveelheid gegevensgerichte detailcontroles worden toegepast, terwijl er in de gereduceerde steekproef toch fouten blijken te zitten. Broeze (2006) gaat er dus vanuit dat het toepassen van een risicoanalysemodel falsifieerbaar moet zijn (zie Popper (1959)).

Data-analyse en risicoanalyse

Data-analyse is een begrip dat voor meerdere uitleg vatbaar is. Dat maakt het lastig om het begrip in verband te brengen met de gegevensgerichte controles en met voorinformatie die daarbij kan worden toegepast. Zo wordt onder data-analyse verstaan het analyseren van (interne) gegevens van de gecontroleerde organisatie om daar betekenis aan te ontleen vooral ten behoeve van de (interne) besluitvorming. De technieken van data-analyse kennen bewerkingen zoals statistische analyse (zowel beschrijvend als voorspellend), het verrichten van logische analyse (zoals opvraging, query's, samenvattingen, draaitabellen), enzovoort.

Omdat de term data-analyse een homoniem is zit deze term het gebruik ervan voor de financiële controle ook in de weg. Want, hoe kan men de

verschillende technieken in de reeks van controlebeslissingen plaatsen?

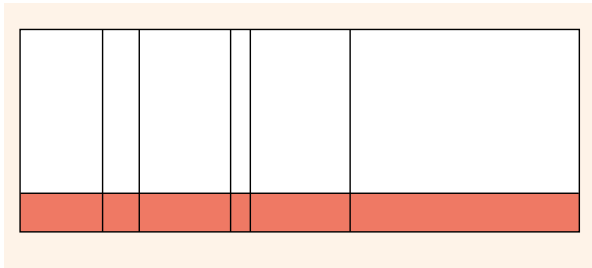


Figuur 5

In figuur 5 (een herhaling van figuur 1) zien we de knooppunten CA1, DC, CA2, CA3 en A. Die kunnen als verschillende soorten data-analyse worden aangemerkt. Knooppunt A is nog niet eerder behandeld: het is een analyse van de gegevens naar aanleiding van bevindingen uit een of meer steekproeven. Heel vaak worden technieken voor data-analyse ook gepromoot om zo snel mogelijk deelpopulaties aan te kunnen wijzen waarin waarschijnlijk fouten zullen zitten (CA2). Daarnaast zijn er technieken van data-analyse die bij uitstek geschikt zijn voor het beoordelen van de interne beheersing (zoals verbandcontroles en process mining; ook CA2).

Initiële cijferbeoordelingen plegen op geaggregeerde gegevens te worden verricht. Ze zijn bedoeld om de grote lijnen van een controle te kunnen sturen of bijsturen, dan wel enige informatie te geven over 'de stabiliteit van het administratieve proces', zoals in Veenstra en Heertje (2006).

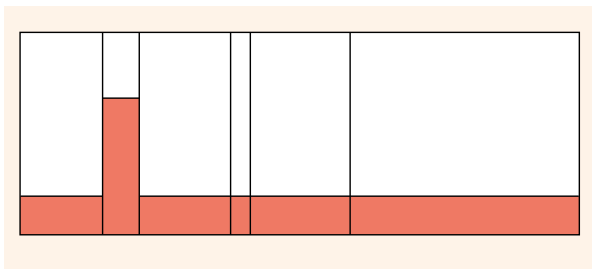
In figuur 6 hebben we de verdeling van de steekproeven geschetst over de jaarrekening als geheel. Het onderste, rode gedeelte stelt een steekproef voor die gelijkmatig over de deelpopulaties is verdeeld.



Figuur 6

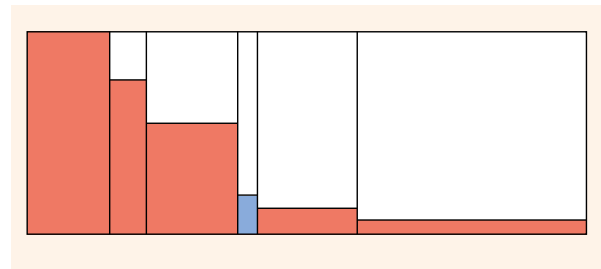
De deelpopulaties zijn bijvoorbeeld categorieën transacties, zoals posten in de jaarrekening.

Als er data-analyse wordt toegepast om zo snel mogelijk een deelpopulatie aan te wijzen waarin zich waarschijnlijk (een hogere concentratie) fouten bevindt, dan leidt dit tot een situatie als geschetst in de tweede kolom van figuur 7.



Figuur 7

Zo'n deelpopulatie vraagt dan om extra aandacht. Merk op dat wij in de figuur hebben aangeduid dat zo'n activiteit *geen* informatie geeft (of hoeft te geven) over de overige deelpopulaties.



Figuur 8

Meer algemeen is de verwachting dat sommige deelpopulaties intensievere controle krijgen (zoals in figuur 8) en sommige een minder intensieve (als gevolg van positieve voorinformatie). In figuur 8 wordt in de eerste, tweede en derde deelpopulatie meer gedaan, en in de vijfde en de zesde minder. In de vierde deelpopulatie is de 'normale' omvang getekend. De lagere omvang (deelpopulaties vijf en zes in figuur 8) van de hoeveelheid werk moet volgen uit bijvoorbeeld de kwaliteit van de interne beheersing. Er zijn ons geen technieken van data-analyse bekend die positieve voorinformatie genereren, anders dan de cijferanalyse volgens Stringer en Stewart (1985).

Risicoanalyse en Bayes

Als laatste paragraaf van dit hoofdstuk een korte uiteenzetting – eigenlijk een soort leeswijzer – van de rol van Bayesiaanse statistiek in de modelbouw voor risicoanalyse in de zin van het gebruik van voorinformatie voor het reduceren van gegevensgerichte detailcontroles.

Om minder werk te doen als de interne beheersing van een organisatie goed werkt, is al jaren ‘gefundenes Fressen’. Om de rechtvaardiging daarvoor te vinden is iets meer nodig dan het volgen van een intuïtief gevoel. Vanuit statistiek geredeneerd ligt het eerder in de rede om meer werk te doen als de waargenomen fractie fouten erg klein is: er zijn meer waarnemingen nodig om een 95%-betrouwbaarheidsinterval te verkrijgen en dan te toetsen of de uitvoeringsmaterialiteit daarbinnen ligt.

Steele (1992) verheldert de Bayesiaanse insteek. Het typische van de Bayesiaanse benadering verdient soms enige toelichting (b.v. te vinden in Iversen (1984)). Met zijn boek heeft Steele (1992) de kennis die besloten lag in een hele serie artikelen van een gevarieerd gezelschap schrijvers gebundeld. Hij wilde verder de accountancy-studenten laten zien dat een kwantitatief prior oordeel (de voorkennis) kan worden gelijkgesteld met een al verricht (steekproef)onderzoek¹³. Het kwantificeren dat daarmee plaatsvindt moet worden gezien als expliciet maken wat impliciet aan (voor-)kennis al aanwezig was.

Leren over risicoanalyse

Het Audit Risk Model (ARM) is door middel van Bayesiaanse vertalingen ervan rijp geworden om van de toepassing in de praktijk te leren. Er is onzes inziens onderzoek nodig om te laten zien wanneer het ARM leidt tot te optimistische risicoschattingen, en dus te weinig gegevensgerichte detailcontroles.

Samenvatting van wat er nog te leren valt

Eerder is al opgemerkt dat het verdienstelijk zou zijn als uit de ervaringen van accountants lering zou kunnen worden getrokken. En wel op een zo objectief mogelijke manier. Dat geldt voor alle elementen waar in meer of mindere mate ‘professional judgement’ aan kleeft.

Wij duiden met behulp van figuur 1 de knooppunten aan van een (kwantitatieve) controle waar dat het geval is.

Bepalen audit risk

Hoewel het toelaatbare audit risk van 5% inmiddels voor de hand lijkt te liggen zou het goed zijn als de audit professional zich blijft afvragen wat hij als audit risk, of controlerisico, zou moeten toepassen. Dus de vraag is of het audit risk van 5% voor iedere kwantitatieve controle geldt. Of misschien wil het publiek voor sommige soorten onderzoek – zoals bijvoorbeeld fraudeonderzoek – een hogere betrouwbaarheid (dat is een lager audit risk). Ook hier zullen we van het

maatschappelijk verkeer moeten leren waar deze (macro-)grens ligt en wanneer een andere omvang wenselijk is.

Bepalen materialiteit (My)

Bij het vaststellen van de materialiteit voor een onderzoek is 'professional judgement' aan de orde. Dat geldt, zij het in mindere mate, ook voor het daaruit afleiden van de nauwkeurigheid – zowel de uitvoeringsmaterialiteit als de toelaatbare afwijking – waarmee de controle zal worden uitgevoerd. Omdat materialiteit een grootheid is die, ten opzichte van het controleproces, exogeen is, is het nodig om van 'het maatschappelijk verkeer' te leren hoe de nauwkeurigheid van het – individuele – onderzoek zou moeten zijn. Het netwerkmodel (figuur 1) laat zien dat, in beginsel, de hoeveelheid gegevensgerichte controles uit de parameters audit risk en toelaatbare afwijking volgt.

Beoordeling interne beheersing en IT

De beoordeling van de interne beheersing (inclusief de IT die daarbij van toepassing is) kent tot nu toe geen expliciet model om vast te stellen dat de mate van effectiviteit van de onvervangbare interne beheersing voldoende is. Er zijn ons geen onderzoeken bekend van het vaststellen van het toereikend zijn van het beoordelen van de interne beheersing met middelen als believe theory of iets dergelijks. Ook kennen wij geen pragmatische stopregel voor onderzoeken naar IT, interne beheersing en de combinatie van beide. Wel kennen wij een oproep daartoe in bijvoorbeeld Kloosterman (2013)¹⁴.

Substantive tests en het schatten van afwijkingen

In de paragraaf over gegevensgerichte detailcontroles komen zowel geldsteekproeven als postensteekproeven om bedragen te schatten aan de orde. Bij de behandeling van beide technieken hebben wij laten zien dat doorgaans men bij de aanpak van de controle uitgaat van het controleren van (deel-)populaties met erg weinig fouten waardoor geldsteekproeven als a priori hulpmiddel dienen te worden toegepast. De toepassing van postensteekproeven om de afwijkingen in geld uit de drukken gaat gepaard met veel voorwaarden. Waaronder een minimum steekproefomvang, een minimumaantal afwijkingen en symmetrie in de afwijkingen.

We schreven ook dat een trekking met de ene methode niet met de andere methode kan worden geëvalueerd. Het zou de moeite waard zijn een kritieke grens te vinden tussen het toepassen van geldsteekproeven en het toepassen van postensteekproeven (om te schatten). Die grens zou bijvoorbeeld of in verwachte hoeveelheid afwijkingen of in de aard van de post kunnen worden gevonden.

Het Audit Assurance Model of Audit Risk Model (RA)

Het Audit Risk Model (ARM) in de meest gebruikte vorm (zonder expliciete toepassing van Bayesiaanse statistiek) is zoals gezegd niet valideerbaar en er kunnen met de meest gebruikte achtergrond (het vermenigvuldigen van kansen) dus geen lessen uit worden getrokken. In de essays over dat onderwerp zal worden aangegeven

dat wanneer het model wordt herschikt tot een Bayesiaans risicoanalysemodel het wel mogelijk is uit de toepassing lering te trekken. Een samenvatting uit de dissertatie van Ed Broeze (2006) geeft eveneens aan dat een model dat niets leert of kan leren uit het verleden de naam model niet waardig is.

Gegevensgerichte controle met cijferanalyse (CA3)

Deze heel specifieke manier van cijferanalyse – namelijk gegevensgerichte controle met behulp van regressieanalyse – is uitgebreid behandeld in Stringer en Stewart (1985). Andere schrijvers hebben naar deze methode verwezen (b.v. Vasarhelyi en Lin, 1988; hoofdstuk 12). Vasarhelyi en Lin (1988) zien technieken als ratioanalyse voornamelijk gebruikt worden in de initiële fase van de audit (CA1). Andere technieken worden toegepast bij een afsluitende cijferbeoordeling.

Verbandcontroles, zowel totaalverbandcontroles als gedetailleerde verbandcontroles, kunnen doorgaans als organisatiegerichte controles worden bestempeld. Zij drukken namelijk verbanden uit tussen verschillende inkomende en uitgaande geldstromen om daaruit eventuele ‘verdachte deelverzamelingen transacties’ af te leiden.

Wat over de soort gegevensgerichte cijferanalyses valt te leren is welke technieken, naast die van Stringer en Stewart (1985), bruikbaar zijn. Daarnaast welke technieken een reductie van de hoeveelheid gegevensgerichte detailcontroles rechtvaardigen.

Data-analyse

Zoals al eerder is geconstateerd is data-analyse een term die vele technieken impliceert (NBA 2018). Binnen de verzameling technieken die als data-analyse kan worden aangeduid wordt vaak gefocust op de interne beheersing. Process mining is daar een voorbeeld van. Het is zinvol te leren welke technieken allemaal beschikbaar zijn en die het etiket data-analyse of data-science krijgen. Aanvullend is het zinvol te onderzoeken of het toepassen van die technieken andere technieken, zoals steekproeven, gedeeltelijk overbodig maken dan wel beter toepasbaar. Dat laatste zou dan met name kunnen als de technieken van data-analyse krachtig zijn in het aanwijzen van subpopulaties waar extra werk zou moeten worden verricht om welke reden dan ook.

Het volgen van de waardenkringloop in een organisatie is degelijk beschreven in Elsas (1996). Ook deze methode kan worden aangemerkt als een lid van de familie die data-analyse heet.

Ook zou het in dit verband leerzaam zijn om te onderzoeken welke technieken in de praktijk welke informatie toevoegen over welke doelstellingen van een financiële controle. Daarbij kunnen ook in beginsel simpele technieken als verzamelingenleer behulpzaam zijn (bijvoorbeeld bij het onderzoek naar volledigheid van de verantwoording van transacties).

Literatuur

Batenburg, P.C. van, J. Kriens, W.M. Lammers van Bueren, R.H. Veenstra (1991); Audit Assurance Model and Bayesian Discovery Sampling, (*Research Memorandum FEW*), Tilburg University: Faculteit der Economische Wetenschappen.

Batenburg, P.C. van, R.J.M. Dassen (1996); Het Bayesiaanse model van Deloitte & Touche ter ondersteuning van de controlemix, in: *De Accountant*, jg. 103, no. 1, september, pp. 31-35.

Blokdijk, J.H. (1975); Een kernvraagstuk van de leer der accountantscontrole deel I en deel II; *Maandblad voor Bedrijfseconomie en Accountancy*, pp. 147-159 (deel I); pp. 190-207 (deel II).

Blokdijk, J.H. (1995); De betekenis van de onvervangbare interne controle voor de accountantscontrole; *Maandblad voor Bedrijfseconomie en Accountancy*, pp. 588-598.

Blokdijk J.H., F. Drieënhuizen en P. Wallage (1995); *Reflection on Audit Theory*; Kluwer, Limperg Instituut.

Blokdijk, J.H. (2001); De effectiviteit van de systeemgerichte aanpak in de accountantscontrole, *Maandblad voor Bedrijfseconomie en Accountancy*, pp. 71-80.

Blokdijk, J.H. (2004); Tests of Control in the Audit Risk Model: Effective? Efficient? *International Journal of Accountancy*; pp. 185-194.

Broeze, G.B., H.H.W. Kloosterman (2003); *Statistical Audit; overzicht van onderzoek over de afgelopen tien jaar* (niet gepubliceerd).

Broeze, G.B. (2006); Validation of Risk Assessment in Auditing, *proefschrift*, VU, Amsterdam.

Burger, D., A. Blumer, A. Muessig (2012); The use of substantive aggregate analytical procedures: do auditors particularly rely on nonsignificant outcomes, *Researchgate.net*.

Elsas, P.I. (1996); *Computational Auditing*, dissertatie, VU, Amsterdam.

Iversen, G.R. (1984); *Bayesian Statistical Inference*, *Sage University Paper*.

Jager, N.G. de, G. Pap, M.C.A. van Zuijlen (1996); Facts, Phantasies and a new Proposal Concerning the Stringer-bound; *Research memorandum*, Limperg Instituut.

Kinney, W. (1979); The Predictive Power of Limited Information in Preliminary Analytical Review: An Empirical Study, *Journal of Accounting Research*, pp. 148-165.

Kloosterman, H.H.W. (2004); Wat is eigenlijk risicoanalyse in de accountantscontrole? *Maandblad voor Bedrijfseconomie en Accountancy*, pp. 570-578.

Kloosterman, H.H.W. (2013); Audituitgangspunten en IT-audit (art 1400); *Handboek EDP-audit*, Kluwer, NBA, NOREA.

Kriens, J. en J.J.M. Peterse (1986); Toepassing van de regressieschatter in de accountantscontrole, *Statistische Methoden*.

Kriens, J., H. Timmermans, H. van den Wildenberg, J.P.C. Kleijnen (1989); Regression sampling in statistical auditing, *Statistica Neerlandica*, pp. 193-207.

Loebecke, J.K., P.J. Steinbart (1987); An Investigation of the Use of Preliminary Analytical Review to Provide Substantial Evidence, *Auditing: a journal of practice and theory*, pp. 74 – 89.

Lucassen, A., H. Moors en P.C. van Batenburg (1996); Modifications of the Stringer-bound: a simulation study on the performance of audit sampling evaluation methods, *Kwantitatieve methoden*, nr. 51. pp. 17-25.

NBA (2018); *Handreiking Data-analyse*.

Pap, G., M.C.A. van Zuijlen (1995); *On the asymptotic behaviour of the Stringer-bound*, Limperg Instituut.

Popper, K.R. (1959); *The Logic of Scientific Discovery*, Hutchinson Education.

Popper, K.R. (1979); *The Growth of Scientific Knowledge*, Klostermann, Frankfurt am Main.

Roberts Donald M. (1978); *Statistical Auditing*, AICPA.

Schapendonk-Maas, H. (1998); *Betrouwbaarheidsbovengrenzen en steekproeftechnieken in de accountancy*, Research Memorandum, Limperg Instituut.

Steele, A. (1992); *Audit Risk and Audit Evidence: the Bayesian Approach to Statistical auditing*; Academic Press & the Chartered Association of Certified Accountants (ACCA).

Stewart T. (2013); *A Bayesian Audit Assurance Model*, dissertation, VU.

Stringer, K.W., T. Stewart (1985); *Statistical Techniques for Analytical Review (STAR)*, John Wiley & Sons, Inc.

Stuurgroep Publiek Belang (2017); *Green Paper; De Definitie van Audit Kwaliteit*, NBA.

Vasarhelyi, M.A., T.W. Lin (1988); *Advanced Auditing, Fundamentals of EDP and Statistical Audit Technology*, Addison-Wesley Publishing Company Inc.

Veenstra, R.H., A. Heertje (2006); Doeltreffende
administratieve controle door steekproeven,
Maandblad voor Bedrijfseconomie en Accountancy,
pp. 611-619.

Willemsen, A. (1997); *On determining an upper
confidence bound for the total error amount in audit
populations*, Limperg Instituut.

Noot

- 1 Zoals in onder meer: Blokdijk (1975), Blokdijk (1995), Blokdijk (2001) en Blokdijk (2004).
- 2 Zoals in de paragrafen 2 en 3 van Blokdijk et al. (1995).
- 3 Kloosterman (2013, p. 2) stelt dat "op het gebied van de aanpak van IT-audits ... de verwachting (was) dat het rationaliseren van de aanpak van IT-audits ten opzichte van het jaar 2000 zou zijn toegenomen. De vakliteratuur over IT-auditing liet echter weinig of geen uitkomsten van onderzoek zien op het terrein van de toereikendheid van een IT-audit, van de objectivering van scoping van IT-audits en de samenhang van onvervangbare interne beheersing en IT-audit."
- 4 Het betreft zowel de materialiteit voor de financiële overzichten als geheel, als de uitvoeringsmaterialiteit voor samenstellende delen ervan. Men kan de vaststelling van de toelaatbare afwijking per relevant onderdeel er ook onder begrijpen.
- 5 De toelaatbare afwijking werd vroeger genoemd: controletolerantie.
- 6 De opmerking 'desgewenst' slaat op verschillende mogelijkheden voor nader onderzoek. Bijvoorbeeld een fraudeonderzoek, een extra onderzoek om een goede schatting van de fouten te maken, en een diepgaand onderzoek om mogelijke verbeteringen in de organisatie aan te brengen.
- 7 Controletolerantie is een wat oudere term. Inmiddels hanteren de controlestandaarden (NV COS) de termen materialiteit, uitvoeringsmaterialiteit en toelaatbare afwijking. De term 'toelaatbare afwijking' komt vrijwel overeen met het begrip 'controletolerantie'.
- 8 De DUS Cell Evaluation methode is uitgebreid beschreven in Appendix B van Leslie et al. (1979).
- 9 Onder meer Roberts (1978), hoofdstuk 6.
- 10 Onder meer Kriens et al. (1985) en Kriens et al. (1989).
- 11 Vergelijk K.R. Popper (1959): "...For I do not demand that every scientific statement must have been tested before it is accepted. I only demand that every such statement must be capable of being tested..." (p. 23).
- 12 De prille vorm van het Audit Risk Model kende drie niveaus. Uitgebreid niveau van gegevensgerichte detailcontroles met $R = 3$; een gereduceerd niveau met een $R = 2$ en een beperkt niveau met een $R = 1$. Deze waarden hebben alleen betekenis bij een steekproef met nul fouten.
- 13 Steele (1992) noemt dat Steekproef-equivalenten.
- 14 Kloosterman (2013, p. 2) stelt dat "op het gebied van de aanpak van IT-audits ... de verwachting (was) dat het rationaliseren van de aanpak van IT-audits ten opzichte van het jaar 2000 zou zijn toegenomen. De vakliteratuur over IT-auditing liet echter weinig of geen uitkomsten van onderzoek zien op het terrein van de toereikendheid van een IT-audit, van de objectivering van scoping van IT-audits en de samenhang van onvervangbare interne beheersing en IT-audit."

Korte samenvatting Leslie, Teitlebaum & Anderson

Hein Kloosterman

Voor de meeste mensen die affiniteit hebben met statistical auditing is het boek van Leslie, Teitlebaum en Anderson (1979) een standaardwerk. Zij waarderen in hoge mate de manier waarop deze auteurs hun artikelen over het onderwerp hebben gegroepeerd en tot een boek hebben gemaakt dat de praktijk van de accountant kan ondersteunen. Omdat de techniek sinds 1979 forse stappen heeft gemaakt is de toepasbaarheid van wat in het boek is vermeld vergroot. Wat overigens niet wegneemt dat in de dagelijkse praktijk statistische steekproeven voornamelijk bij bijzondere onderzoeken en in de sfeer van de overheidsaccountancy worden toegepast.

Inleiding

De schrijvers geven in hun inleiding de uitvinders van de geldsteekproef Van Heerden (1961) en Stringer (1961) de eer voor het concept om geld-eenheden te trekken in plaats van posten (in de zin van records of 'line items'). Omdat de hoofdstukindeling van het boek er al uitziet als een beknopte samenvatting beginnen we daar mee.

Hoofdstukindeling:

Inleiding:

Sufficient audit evidence

Audit testing concepts

Overall audit Evaluation

Dollar unit sampling:

Attribute sampling and upper error limits

Dollar unit sampling concepts

Dollar unit sample selection

Dollar unit sample evaluation

Planning and evaluation:

Compliance testing: Planning and evaluation

Substantive testing: Planning and evaluation

Applications and practical guidance:

The role of the computer in sampling

Efficiencies and short cuts

Applications

The future: directions for research.

Wij noemen nog enkele appendices van het boek die de samenvatting bijna compleet maken.

Appendix B: Dollar-unit sampling evaluation theory

Appendix C: Confidence levels: the Bayesian concept

Appendix D: Comparison with variables sampling methods

Appendix E: Materiality in audit sampling
Appendix F: Judgmental sampling: an approximation to DUS.

Grote lijnen aanpak geldsteekproeven

In grote lijnen geeft het boek uitleg over de toepasbaarheid van geldsteekproeven. Dat reikt van de bijdrage die een steekproef kan leveren aan de controle van een deelpopulatie zowel als de controle van een financieel overzicht als geheel (zoals een jaarrekening). Hoewel in het algemeen geldsteekproeven gebruikt zullen worden bij het uitvoeren van zogenoemde substantive tests, behandelt het boek ook compliance tests in relatie tot geldsteekproeven. Als er geldsteekproeven worden toegepast kan de accountant laten zien wat het financiële effect is van het niet goed naleven van in- of externe regelgeving. Als het gaat over het meten van de financiële effecten van externe regelgeving dan is een financieel verslag daarvan vergelijkbaar met een jaarrekening.

Het boek behandelt vrij uitgebreid de visie van de auteurs op het beslissingsproces in de financiële controle. Daarbij wordt uitgegaan van het Audit Risk Model (ARM). In appendix C staat men uitgebreid stil bij een Bayesiaanse uitleg van de praktijk van het ARM. Aardig is de samenvatting in deze appendix (p. 303):
“The joint risk models do not provide a reliable measure of the overall audit assurance when

sample results appear favourable. In most cases, the overall audit assurance (posterior confidence) will be higher, but in a few cases, with unfavourable prior probabilities, it will be lower. Bayesian inference provides a better reflection of the accumulated state of knowledge taking all sources of evidence into account.”

Grote lijnen materialiteit

Ook het omgaan met een fenomeen als materialiteit krijgt ruime aandacht (appendix E). De auteurs beschrijven zowel de rol van materialiteit om de financiële controle in te richten (uitvoeringsnorm) als de rol van materialiteit om vast te stellen of de controle toereikend is (beoordelingsnorm). De schrijvers geven een aantal richtsnoeren om de omvang van de materialiteit te bepalen. Zij stellen voor om voor de bepaling ervan een contingentiebenadering te kiezen. Daar geven zij een paar hints voor.

Grote lijnen trekking

Het boek besteedt ruim aandacht aan het trekken van steekproeven. In het huidige tijdsgewricht, dat wil zeggen met de huidige hulpmiddelen zoals IDEA en ACL is dat een vrij eenvoudige bezigheid. Men moet wel de aandacht erbij houden omdat er soms nog wel sprake is van een landschap met valkuilen.
De schrijvers behandelen zowel Random Dollar

Selection, als Cell Sampling, als Fixed Interval Selection (hoofdstuk 6). Elke trekkingstechniek heeft zijn beperkingen. De schrijvers bestempelen alle drie methoden als leidend tot een voldoende representatieve steekproef (van geldeenheden). Het random beginsel – aselechte trekking dus – moet de representativiteit van de steekproef garanderen.

De random selectie wordt om praktische redenen afgeraden¹. Alleen al de onmogelijkheid om al tijdens de te onderzoeken periode een deeltrekking te beoordelen is een groot nadeel. Het boek is in 1979 uitgegeven. Daardoor gaat een deel van het hoofdstuk met betrekking tot het trekken van een steekproef over het toepasbaar maken van een handmatige trekking. Uiteindelijk kiezen de schrijvers voor Cell-selectie en vast-intervalselectie. De Cell-selectie vinden zij het sterkst omdat in iedere Cell een random selectie plaatsvindt. Bij de vast-intervalselectie plaatsen zij de kanttekening dat een of andere vorm van systematiek in de populatie kan terugkeren in de steekproef, waardoor de representativiteit geweld kan worden aangedaan. Als de posten eerst random zouden zijn ‘geschud’ is die onvolkomenheid tenietgedaan. Bij Cell-selectie (dus random selecteren binnen iedere Cell) geldt het bezwaar niet dat aan de vast-intervalselectie kleeft.

De schrijvers noteren nog dat de noodzaak tot stratificeren als gevolg van verschillen in omvang van het bedrag per post, komt te vervallen. Iedere geldeenheid heeft bij het toepassen een geldsteekproef immers dezelfde trekkingskans.

Grote lijnen evaluatie

In hoofdstuk 7 wordt de evaluatie van geldsteekproeven uitgelegd. Het verschil tussen de ‘my-dollar-right-or-wrong-methode’ en de ‘tainted dollar’-methode wordt uitgelegd. De eerstgenoemde methode wordt afgeraden, vooral omdat er informatie pleegt te worden genegeerd. Een steekproef met erg weinig fouten kan daarvoor als foutloos worden aangemerkt terwijl er wel sprake is van bekende fouten. Dat is contra-intuïtief.

De tainted-dollar-methode maakt gebruik van de klassieke opstelling van de Poisson-verdeling. Per gevonden fout in de steekproef kiest men de entree in die tabel. Het accres van nul naar één wordt gebruikt voor de eerste fout, het accres van één naar twee, voor de tweede fout; en zo verder. De schrijvers gebruiken de nominale zekerheid die volgt uit de risicoanalyse (ARM)². De volgorde van de fouten in de opstelling is naar afnemende relatieve fout. Het is evident dat als alle fouten 100%-fouten zijn, de volgorde van de uitkomsten niet uitmaakt. De tainted dollar evaluation methode wordt in latere literatuur aangeduid als de ‘Stringer-bound’ (de bovengrens die K.W. Stringer heeft bedacht).

De schrijvers betogen dat met de methode zowel overstatements als understatements kunnen worden geëvalueerd. Naar onze mening is dat theoretisch niet valide. Hoe dat zit voert te ver voor deze samenvatting.

Doordat de schrijvers het ARM toelaten kan het zijn dat er resultaten moeten worden gecombineerd uit deelpopulaties die met een verschillend risico zijn benaderd en waar dientengevolge een verschillende omvang van het interval is toegepast. Het combineren is dan een probleem dat in dit boek niet wordt opgelost. Daar is later onderzoek naar gedaan en er zijn artikelen over verschenen.

In het boek wordt een voorbeeld gegeven waarin voor verschillende posten van een financieel overzicht met een interval van dezelfde omvang is gecontroleerd.

Het boek werkt een alternatief voor de Stringer-bound uit. Die methode heet 'DUS Cell Evaluation'. De DUS-Cell-methode is onder voorwaarden minder conservatief dan de Stringer-bound en vraagt iets meer rekenwerk dan de Stringer-bound.

Geldsteekproeven en compliance tests

De schrijvers adviseren om ook bij compliance tests geldsteekproeven in te zetten, althans dat te overwegen. De redenering is dat veel accountants de wens koesteren de grootste posten de grootste nadruk te geven en dat geldsteekproeven aan die wens tegemoetkomen.

De schrijvers geven geen oordeel over een volgorde in de werkzaamheden. Wel stellen zij dat voordat een accountant gaat werken met de (deel-)populaties die onderzocht moeten worden, hij de

samenstellende delen van alle delen op consistentie moet controleren; met andere woorden de aansluitingen moet vaststellen. Hoewel aansluitingen vaak als een cijferbeoordeling worden gezien, stellen de schrijvers dat hier sprake is van een randvoorwaarde voordat de eigenlijke controle kan plaatsvinden.

Geldsteekproeven en substantive³ tests

Het hoofdstuk over substantive tests geeft een gedetailleerde opsomming van de controle-activiteiten in deze fase. Er worden veel overwegingen gegeven hoe naar verschillende soorten fouten gekeken kan worden. Ook wordt gewaarschuwd voor een te groot pessimisme bij het inrichten van het onderzoek. Dat leidt immers tot (te) grote steekproeven. Wat hier duidelijk uit de doeken wordt gedaan is dat het onderzoek zal moeten worden uitgevoerd op deelpopulaties en dat die deelpopulaties meestal afzonderlijke posten⁴ van de financiële overzichten zijn. Ook moet rekening worden gehouden met afzonderlijke eenheden binnen de onderneming, die elk een afzonderlijke cultuur en dus ook interne beheersing kunnen hebben. Daar komt dan ook de inschatting van de risico's met betrekking tot de afzonderlijke audit-componenten om de hoek kijken.

Om de waarde van dit boek te illustreren hebben wij een praktijkvoorbeeld opgenomen van een evaluatie van een geldsteekproef.

Voorbeeld evaluatie met de Stringer-bound en met de DUS Cell methode

Dit voorbeeld is samengesteld met de gegevens van een praktijkvoorbeeld. De parameters in de controle waren als volgt:

- Audit Risk: 5%;
- Uitvoeringsmaterialiteit voor de onderzochte populatie: 135.000;
- Toelaatbare afwijking = controletolerantie = MP = 105.000
- Er is geen risicoanalyse toegepast (dus geen reductie van de steekproefomvang).

Uit de Poissonverdeling volgt dat de factor R drie bedraagt⁵ bij een Audit Risk van 5% en nul fouten in de steekproef. Uit de theorie volgt ook dat het interval (de intensiteit waarmee de controle plaatsvindt) wordt berekend als $MP/R = 105.000/3 = 35.000$.

In de aanpak is erin voorzien om als eerste altijd een grove berekening te maken.

De gegevens van de controle leiden tot het volgende:

Rek#	Geboekt bedrag	Conclusie na controle	Fout#	Fout bedrag	Tainting
47500	1.364,52	87,50 privé-uitgave, geen kosten	1	87,50	0,0641
41400	32.971,81	Onderhoud activa; 1/3 te activeren; afs.	2	9.354,10	0,2837
40250	159,20	19,11 privé-uitgave, geen kosten	3	19,11	0,1200
15230	3.250,00	Ten onrechte niet geactiveerd	4	3.250,00	1,0000
		Som relatieve fouten (taintings)			1,4678
Projectie	Interval = 35.000	Som Taintings * Interval	MLE		51.373,00

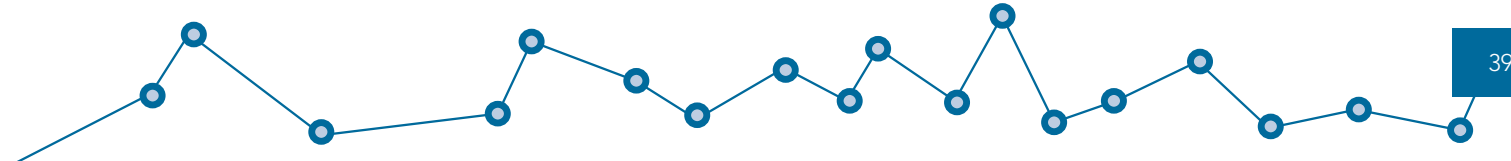
De Stringer-bound levert de volgende opstelling.
De taintings, relatieve fouten, zijn hier geplaatst in
volgorde van afnemende omvang.

Evaluatie met Stringer-bound:

Fout#	Foutbedrag	Tainting	Stringer	PGW	MF
4	3.250,00	1,0000	1	1,75	1,75
2	9.354,10	0,2837	2	1,55	0,439735
3	19,11	0,1200	3	1,46	0,1752
1	87,50	0,0641	4	1,40	0,08974
		1,4678	Nullast	3,00	3,00
				Totaal	5,454675
				In geld	190.914,00
				MLE	51.373,00
Grens					135.000,00

N.B.: de taintings zijn afgerond op vier decima-
len. Op de volgende pagina zijn de taintings niet
afgerond, alleen met 4 decimalen gepresenteerd.
Daardoor zijn er afrondingsverschillen te zien.

Hierna laten wij het effect van de DUS Cell
evaluatie zien.



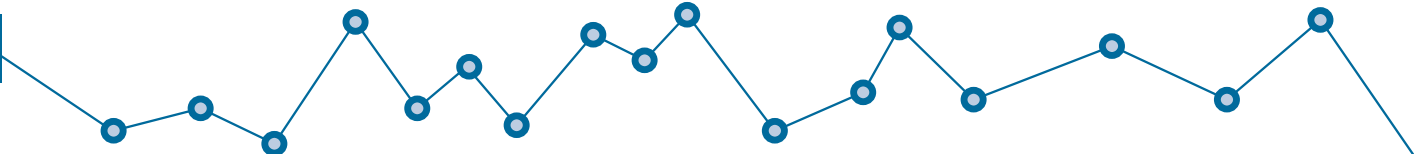
De toepassing van de DUS Cell methode levert de volgende uitkomsten:

Fout#	R(i)	Boekwaarde	Auditwaarde	Foutbedrag	Tainting	Cum tainting
0	2,995732274				1	
1	4,743864518	3.250,00	0,00	3.250,00	1,0000	1,0000
2	6,295793622	38.740,22	27.749,62	9.354,10	0,2837	1,2837
3	7,753656528	159,20	140,09	19,11	0,1200	1,4037
4	9,153519027	1.364,52	1.277,02	87,50	0,0641	1,4679

Cum tainting	Cum gemiddelde ti	UEL vorige	Load & Spread	Simple spread	UEL
					2,9957322740
1,0000	1	2,995732274	3,9957322736	4,7438645184	4,7438645184
1,2837	0,641849995	4,743864518	5,0275645077	4,0409551026	5,0275645077
1,4037	0,467912559	5,027564508	5,1476021961	3,6280332695	5,1476021961
1,4679	0,366965698	5,147602196	5,2117273097	3,3590274973	5,2117273097

Load & Spread is: UEL vorig + taint(i)
Simple spread is: R(i) * cumulatief gemiddelde taint(i)
UEL is: maximum (load & spread; simple spread).

Interval 35.000
Maximale fout Dus Cell methode 182.410
Projectie 51.373
Maximale fout Dus Cell methode na correctie PF 131.037



Conclusie

Dit praktijkvoorbeeld laat zien dat de methode om de bovengrens van de fout te berekenen die de schrijvers voorstellen, minder conservatief is (namelijk 182.410), dan de Stringer-bound (190.914).

Literatuur

Heerden, A. van, (1961); Steekproeven als middel van accountantscontrole, *Maandblad voor Accountancy en Bedrijfseconomie*, 1961, pp. 453 e.v.

Leslie, D.A., A.D. Teitlebaum and R.J. Anderson (1979), *Dollar Unit Sampling: A Practical Guide for Auditors*, Copp Clark Pitman, Toronto.

Stringer, K.W. (1961); Some Basic Concepts of Statistical Sampling in Auditing, *Journal of Accountancy*, pp. 63-69.

Noot

- 1 Er is ook nog wel een theoretisch bezwaar tegen random trekking in te brengen. Dat houdt verband met de mogelijkheid dat er veel geselecteerde geldeenheden in een post (boeking) kunnen voorkomen. Dat bezwaar wordt afgevangen door Cell-sampling (hooguit twee geselecteerde geldeenheden per post die niet als 'integraal gecontroleerd' moet worden aangemerkt) en systematic sampling (maar een getrokken geldeenheid per post die niet als 'integraal gecontroleerd' moet worden aangemerkt). Bij systematic sampling worden alle posten met een omvang groter dan of gelijk aan het interval (integraal) geselecteerd. Bij Cell sampling is dat het geval bij posten groter dan of gelijk aan twee keer het interval.
- 2 Dit doen zij ondanks het ARM, dat impliciet uitgaat van het vinden van nul fouten in de steekproef of door middel van de componenten (analyseren inherent risico, interne beheersing, cijferbeoordeling) die de vermenigvuldiging herbergt.
- 3 'Substantive test' is een term waarmee men een inhoudelijke beoordeling aanduidt. In het Nederlandse taalgebied wordt hiervoor vaak de term 'gegevensgerichte controle' gebruikt. Een andere term om de strekking van 'substantive test' duiden is 'materiële controle'. Voor de volledigheid: met 'gegevensgerichte controle' wordt niet bedoeld dat er gegevens of data worden gebruikt om tot een oordeel te komen, maar dat de informatie die wordt verstrekt op inhoudelijke juistheid wordt beoordeeld. Dus niet het proces van de totstandkoming van de informatie wordt beoordeeld, maar de gegevens zelf. En die gegevens worden geacht de informatie(waarde) te dragen die bedoeld was.
- 4 'Posten' is een niet-eenduidige term. Er pleegt zowel een rubriek in de jaarrekening mee te worden bedoeld als een afzonderlijke boeking, dan wel een 'line item' of een 'record'. In het Nederlandse taalgebied zijn we gewend om te spreken van 'postensteekproeven' als we 'record sampling' bedoelen.
- 5 Net als in het boek gaan wij uit van een heel geringe foutdichtheid, waardoor we de Poisson-verdeling mogen gebruiken om mee te rekenen. Bij 5% risico is volgens die verdeling de reliability-factor R ongeveer 3. Of: bij alfa is 5% en een hoeveelheid fouten, $k = \text{niel}$ is de uitkomst van de Poisson-verdeling bijna 3. Niet iedereen in de accountantspraktijk gebruikt hetzelfde symbool voor deze factor. Soms noemt men diezelfde factor S. Het audit risk is gedefinieerd als de 'kans op ten onrechte goedkeuren' (gegeven de toelaatbare fout in de te beoordelen populatie).

De NV COS in samenhang met Statistical Audit

Hein Kloosterman

De NV COS bevat een heleboel standaarden die het vakgebied Statistical Audit raken. In de inleiding bij de essays is toegelicht dat het vakgebied Statistical Audit inhoudt het toepassen van wiskundige methoden en technieken om de besluitvorming in de financiële controle te ondersteunen. Die ondersteuning kan vooral plaatsvinden bij cijferanalyses en detailcontroles. Hierna geven wij in het kort aan welke standaarden daar een rol spelen. Daarbij gaan wij niet de inhoud van al die standaarden herhalen, maar laten wij de begrippen zien van de standaarden en plaatsen die in de context van een controleaanpak zoals beschreven in de Inleiding Essaybundel Statistical Auditing.

Cijferanalyses

Standaard 520 behandelt de cijferanalyses. Deze worden onderverdeeld in initieel, gegevensgericht en afsluitend. De interesse van de Stuurgroep is vooral gericht op de gegevensgerichte cijferanalyses. Om die uit te kunnen voeren met behulp van wiskundige methoden, verwijst standaard 520 naar de standaarden 315 (risico-inschattingen), 330 (inspelen op geschatte risico's) en 320 (materialiteit en uitvoeringsmaterialiteit in samenhang

met de uitkomsten van cijferanalyses).

De cijferanalyses die het karakter hebben van gehele of gedeeltelijke beoordeling van een organisatie, zoals totaal-verbandscontroles worden gezien als instrument voor het beoordelen van risico's (daarover handelt standaard 315).

Standaard 520 verwijst bij het reageren op uitkomsten van risicoanalyses naar standaard 330 om eventueel cijferanalyses met het karakter van substantive tests in te schakelen.

Een mogelijke onnauwkeurigheid (een afwijking ten opzichte van een als verwachting gekwalificeerde Soll-positie) wordt hier niet 'toelaatbare afwijking' (zoals in standaard 530) genoemd.

Gesteld wordt dat een cijferanalyse een significant verschil kan laten zien. Die significantie is in 520 niet nader aangeduid. Wel wordt vermeld dat er rekening moet worden gehouden met het gewenste niveau van overtuigende controle-informatie. Standaard 520 verwijst *niet* naar standaard 450 om de evaluatie verder vorm te geven.

Gegevensgerichte controles (substantive tests of details)

Volgens standaard 500 moet er sprake zijn van gegevensgerichte controles. Wat daaronder kan worden verstaan is behandeld in standaard 530. Standaard 530 behandelt steekproeven in brede zin. Even afgezien van de niet-statistische steekproeven gaat het over steekproeven in relatie tot interne beheersing, balansposten en substantive tests. Standaard 530 behandelt geldsteekproeven, zowel als postensteekproeven.

Standaard 530 zoomt niet in op de verschillende mogelijke technieken en de informatie die daarvoor nodig is. Alleen voor wat betreft de substantive tests gaat de standaard enigszins de diepte in. Overeenkomstig hetgeen in de Inleiding op de Essays is gezegd, is de informatiebehoefte voor het plannen, trekken, controleren en evalueren van steekproeven als volgt te schetsen.

Plannen steekproeven

- De onnauwkeurigheid waarmee de steekproef zal worden ontworpen;
- In standaard 530 wordt deze aangeduid als 'toelaatbare afwijking'. In oudere geschriften wordt deze grootheid Monetary Precision of Controletolerantie genoemd.
- De toelaatbare afwijking is afgeleid van de uitvoeringsmaterialiteit (kleiner dan of gelijk daaraan). Standaard 320 omschrijft het begrip uitvoeringsmaterialiteit. De uitvoeringsmaterialiteit staat in relatie tot de materialiteit – ook in 320

gedefinieerd – voor de financiële overzichten als geheel. De uitvoeringsmaterialiteit wordt als ontwerpvariabele gedefinieerd voor een steekproef op een subpopulatie van de financiële overzichten als geheel (welke dan de gehele populatie is waarop het onderzoek zich richt).

- De onbetrouwbaarheid waarmee de steekproef zal worden ontworpen;
- In standaard 530 wordt verwezen naar standaard 315 en standaard 330 waar de uitkomst van de risicoanalyse leidt tot het vaststellen van het steekproefrisico. Het steekproefrisico is afgeleid uit het Audit Risk Model (verwoord in standaard 315). In paragraaf 30 gaat standaard 315 in op elementen van interne controle die met detailcontrole kunnen worden onderzocht. Het lijkt erop dat er sprake is van het onderzoeken van de onvervangbare interne controle of een deel ervan. Dat is niet erg waarschijnlijk omdat standaarden van de AICPA aan dit typisch Nederlandse aspect geen aandacht besteden.
- In standaard 315 wordt de stelling ingenomen dat de materialiteit en de risicoanalyse onderling afhankelijk zijn. Dat is merkwaardig omdat materialiteit ten opzichte van de audit een exogene factor is en de risicoanalyse een factor is die te maken heeft met de onzekerheid in de accountantscontrole en dus ten opzichte van de audit endogeen is.
- De inschatting van de (hoogte van de) risico's vindt plaats via standaard 315, en in standaard 330 wordt de actie die daarvan het gevolg is beschreven.

- De mogelijke kwantificering van het risico (de statistische onbetrouwbaarheid) wordt aangeduid met de term ‘waarschijnlijkheid’, wat het complement is van de statistische onbetrouwbaarheid en gelijk aan de statistische betrouwbaarheid.

De standaard beschrijft dat de steekproefomvang zodanig groot moet zijn dat de waarschijnlijkheid dat de deelpopulatie een (geaggregeerd niveau aan) fout(en) van materieel belang bevat tot een aanvaardbaar niveau is teruggebracht.

Het gangbare rekenwerk met betrekking tot steekproeven is niet zozeer gericht op het berekenen van de steekproefomvang, maar op het berekenen van de omvang van het steekproefinterval.

Voor het opzetten van controles bij groepsmaatschappijen wordt verwezen naar standaard 600. Daar wordt aangegeven hoe audits van leden van een groep kunnen inwerken op de audit van een groep als geheel. Bij de planning van de audit dient zowel met de eisen die te stellen zijn aan de afzonderlijke deelnemers als met de eisen voor het geheel rekening te worden gehouden.

Trekken steekproeven

De wijzen van trekking van een steekproef beschrijft standaard 530 in bijlage 4. Deze bijlage noemt aselechte trekking, systematische trekking (vast interval) en trekking van geldeenheden. De trekking van geldeenheden, de geldsteekproef, komt in bijlage 1 aan bod tegelijk met andere

vormen van stratificatie. Daar wordt weer naar bijlage 4 verwezen en de systematische trekking genoemd. Wij kennen naast trekking met een vast interval de zogenoemde trekking door middel van Cell-sampling (random trekking binnen een interval) en de zeefmethode. Er worden geen voorschriften voor het berekenen van steekproefomvang of interval gegeven. Best practice bij het toepassen van geldsteekproeven is Cell-sampling.

Het controleren zelf

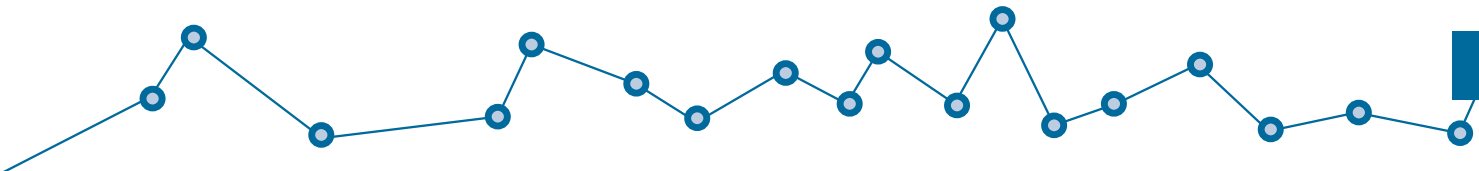
Standaard 530 noemt de noodzaak om een getrokken element dat blijkt niet te kunnen worden gecontroleerd, te vervangen door een ander element (paragraaf 10 en paragraaf A14). De noodzaak daarvan is niet toegelicht. Er is wellicht sprake van een aanbeveling hiertoe bij het beoordelen van de interne beheersing, bij geldsteekproeven kan daarvan geen sprake zijn.

Evalueren (geld-)steekproeven

Bij de evaluatie van de uitkomsten van de controle door middel van de steekproef is het van belang om de invloed van die uitkomsten voor de totale beoordeelde populatie te berekenen.

Daartoe zijn volgens standaard 530 nodig:

- De geconstateerde afwijkingen; zowel absoluut als relatief.
- Het gehanteerde interval.
- De maximaal toegestane onzekerheid (kans) ten aanzien van de beoordeelde populatie.



Veel accountants zijn gewend, stammend uit de tijd voor de zogenoemde 'clarity'-actie op de standaarden, om in plaats van te rekenen met de norm voor het risico, te rekenen met de norm voor de maximale fout. Beide berekeningen zijn equivalent. Er ligt een hogere moeilijkheidsgraad bij de interpretatie van de maximale-risiconorm, dan bij de interpretatie van de maximale-foutnorm. De standaard staat toe dat er fouten worden geïsoleerd. Die geïsoleerde fouten heten dan atypische fouten. Over de problematiek hieromtrent zijn enkele columns verschenen die wij in de bundel hebben opgenomen.

De te berekenen uitkomsten zijn dus een projectie van de afwijkingen en de kans dat de toelaatbare afwijking wordt overschreden door de werkelijke afwijking in de populatie. Die kans is dan het berekende steekproefrisico. Dat steekproefrisico (een Ist-positie) moet worden vergeleken met de van tevoren aangegeven maximale toegestane kans (dat de populatie een materiële fout bevat gegeven de toelaatbare afwijking; dat is de Soll-positie).

Het kan zijn dat waar in de laatste zin staat 'toelaatbare afwijking' eigenlijk de uitvoeringsmaterialiteit is bedoeld. In het andere geval zal iedere gevonden afwijking, hoe klein dan ook, de berekende kans hoger laten uitkomen dan de Soll-positie.

Evaluatie in standaard 450

De evaluatie van steekproeven wordt niet alleen maar in standaard 530 behandeld, maar (vooral) ook in standaard 450. Standaard 530 verwijst niet terug naar standaard 450. Standaard 450 wijst in de toelichting (A6) vooruit naar standaard 530. Het gaat dan om zaken als projectie van de afwijkingen en de evaluatie van de resultaten (van de steekproef).

Als een onderzoek is ingesteld als een steekproef zoals omschreven in standaard 530 heeft de uitvoering geleid tot het vaststellen, het detecteren, van afwijkingen¹. Dit vaststellen kan gevolgen hebben voor de controle. Bijvoorbeeld door de reeds vastgestelde materialiteit (van standaard 320) te herzien. De afwijkingen worden ter correctie voorgelegd aan de gecontroleerde organisatie. Na eventuele correctie daarvan resteren er zogenoemde niet-gecorrigeerde afwijkingen.

Standaard 450 noemt de mogelijkheid van triviale afwijkingen. De formulering is zodanig dat daaruit afgeleid kan worden dat daarbij sprake is van heel kleine verschillen die geen invloed kunnen hebben op de vraag of goedkeuring van de overzichten mogelijk is.

Het aggregeren van de afwijkingen (accumuleren noemt de standaard dat) leidt tot afwijkingen die door het management worden gecorrigeerd en een deel dat niet door het management wordt gecorrigeerd. De niet gecorrigeerde afwijkingen

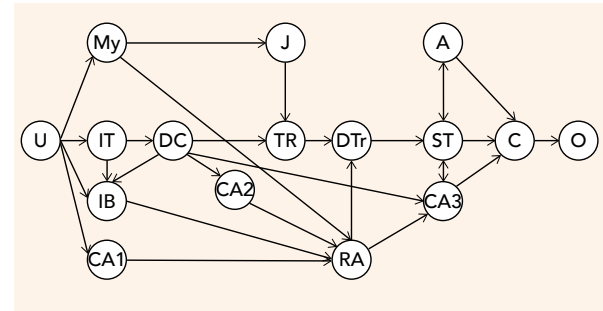
geven aanleiding tot het berekenen van de kans op ten onrechte goedkeuren gegeven de projectie van die afwijkingen.

Door de clarity-slag en de letterlijke vertaling van de Engelstalige standaard is niet helemaal helder wat men bedoelt.

In het essay over het standaardwerk van Leslie et al. en dat over evaluatie wordt de methode om te evalueren door het berekenen van maximale fouten met een voorbeeld uitgewerkt.

Standaard 450 voegt dus vooral als informatie-behoefte toe de correcties van de afwijkingen die door de leiding van de gecontroleerde organisatie worden aangebracht.

Schema controle-aanpak in relatie tot standaarden



Figuur 1

Planning en evaluatie in standaard 600

Standaard 600 gaat over groepscontroles. Deze standaard gaat uit van de mogelijkheid dat bij de controle van de groep meerdere accountants betrokken zijn. De standaard gaat voor wat betreft de substantive tests over de vaststelling van de materialiteit voor zowel de groep als geheel als voor de samenstellende delen. Standaard 600 verwijst naar de standaarden 315 en 330 voor wat betreft de risicoanalyses – met name de beoordelingen van de interne beheersing – die verwacht worden per onderdeel te worden gedaan.

Het effect van de aggregatie van de uitkomsten van de substantive tests wordt niet expliciet in standaard 600 behandeld. Er is een essay opgenomen over het onderzoek dat naar het aggregeren van bewijs is gedaan door Trevor Stewart (2013).

Verwijzing van elementen in figuur 1 (schema controleaanpak naar standaarden in NBA (2017)).

Knooppunt in het diagram	Standaarden
U = Understanding the business	200, 315
My = Vaststellen materialiteit	
Materialiteit overzicht als geheel	320
Uitvoeringsmaterialiteit	320
Toelaatbare afwijking	530
IT = Informatietechnologie (samen met IB)	
IB = Interne beheersing	
Schatten risico's (ook IB)	315
Effectueren IB-risico's	330
Beoordelen IB (details)	530
Beoordelen IB (cijferanalyse)	520
Beoordelen randvoorwaarden IT	-
DC = Datacollectie	-
CA1 = Initiële cijferanalyse	520
CA2 = Cijfer- en data-analyse	520
J = Vaststellen interval	-
TR = Trekking klassieke steekproefomvang	-
Dtr = Trekking te beoordelen steekproef	530
RA = Risicoanalyse (ARM of Bayesiaans)	315, 330, 530
ST = Substantive tests (detailcontroles)	530
CA3 = Substantive cijferanalyse	520
C = Conclusies uit substantive tests (evaluatie)	520, 530, 450, 600, 320
A = Foutenanalyse	520, 530, 450, 600, 320
O = Vormen oordeel	520, 530, 450, 600, 700

Literatuur

NBA (2017); *Handreiking Regelgeving Accountancy 2017*; geraadpleegd op NBA.nl.

Stewart, T. (2013); *A Bayesian Audit Assurance Model*, dissertation, VU.

Noot

- 1 De standaarden focussen op gevonden afwijkingen. De toegepaste methodiek echter focust op het aantal goede elementen van een trekking. De methodiek houdt in dat als er voldoende elementen zijn gecontroleerd en goed bevonden, de controle is voltooid en de beoordeelde populatie kan worden goedgekeurd.

Haat-liefdeverhouding tussen accountants en statistiek¹

Paul van Batenburg en Jacques de Swart

Zoals alle professionals die op basis van hun gegevensgerichte werkzaamheden statistisch verantwoorde conclusies willen trekken, is statistiek een mogelijk hulpmiddel voor accountants. Maar accountants zijn er niet altijd dol op. Hoe is het gesteld met de opleiding op dit gebied?

De International Standards for Auditing (ISA's) verplichten het gebruik van statistisch onderbouwde methoden niet, maar laten ook toe dat accountants bij hun gegevensgerichte controle en bij het uitvoeren van proceduretests op basis van vakkundige oordeelsvorming steekproeven opzetten, selecteren en evalueren.

Hetzelfde geldt voor cijferanalyses (Substantive Analytical Procedures). Naast de ISA's zijn er voor accountants relevante controleprotocollen voor subsidieregelingen en de sociale zekerheid die statistische steekproeven voorschrijven, dan wel adviseren, of zelfs voorwaarden stellen voor het gebruik van niet-statistische steekproeven.

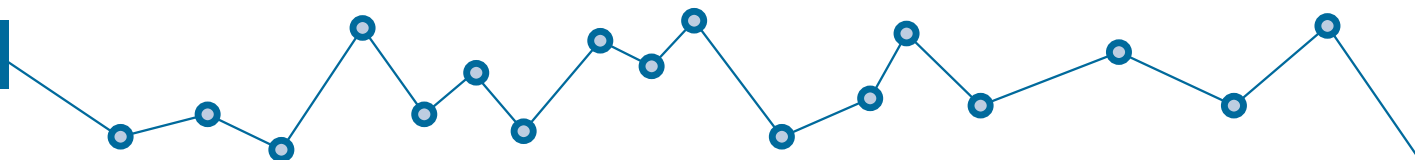
Accountants vinden statistiek moeilijk, maar kunnen zich ook vaak niet vinden in de uitkomsten van wiskundige berekeningen en denken bijvoorbeeld dat kleinere steekproeven voldoende zijn of

vinden de extrapolatie overdreven conservatief. Niet alle subjectieve voorkennis van de accountant vindt zijn weg in het statistische model zie Burgstahler, D., S.M. Glover & J. Jiambalvo (2000).

Maar wat weten accountants eigenlijk van statistiek? In de opleiding tot accountant is statistiek nodig om twee redenen:

- Studenten moeten een afstudeerscriptie schrijven op basis van empirisch onderzoek en moeten daarvoor artikelen uit de vakliteratuur bestuderen. Dit zijn eisen die ook voor andere academische opleidingen gelden.
- Specifiek voor het uitvoeren van accountantscontrole is er een mogelijkheid om statistiek te gebruiken. En dus ook een reden om dat deel in het opleidingsprogramma op te nemen.

De Commissie Eindtermen Accountancyopleidingen (CEA) benoemt noodzakelijke kennis van en ervaring met onderdelen van de statistiek, zonder bovengenoemde twee invalshoeken te scheiden. Omdat de eerste invalshoek, kennis van statistiek ten behoeve van het maken van de scriptie en het lezen van de literatuur, in alle opleidingen aan bod komt, spitst deze inventarisatie zich toe op het onderwijs in statistiek ten behoeve van



het uitvoeren van controles. Daarbij kan worden gedacht aan:

- Steekproefsgewijze controle - opzet, selectie en evaluatie van een deel van een populatie om een uitspraak over het geheel te doen.
- Aanpassingstoetsen - het vergelijken van een gevonden frequentieverdeling met verwachte kansverdelingen om zo patronen te gebruiken als norm voor plausibiliteit.
- Cijferanalyse - het opbouwen van een verwachting uit gecontroleerde gegevens en de confrontatie daarvan met realisaties.

Benodigde methoden en technieken voor de controle zijn bijvoorbeeld:

- Toetsen en schatten - en vooral ook het besef dat pas goedkeuren als daar voldoende bewijs voor is, iets anders is dan pas afkeuren als dat nodig is.
- De binomiale en poisson-verdeling - die in het model waarin waarnemingen kloppen of niet, alleen aannames vereisen over de omvang van de steekproef in relatie tot de omvang van de populatie.
- De normale verdeling - die bij schattingen van waarden wordt gebruikt, maar die wel een aanname behelzen die moet worden gevalideerd.
- Regressie - om bij cijferanalyses een verwachting op te bouwen en om een indicatie te geven welke afwijkingen ten opzichte van de realisatie daar te ver buiten vallen en nader moeten worden onderzocht.
- Bayesiaanse statistiek - als methodologie om

subjectieve informatie te combineren met steekproefgegevens, en zo de voorkennis van de accountant in het steekproefmodel op te nemen.

Behalve Nyenrode zijn alle masteropleidingen tot accountant in Nederland kopstudies als vervolg op een WO- of soms een HBO-opleiding. Wat in die vooropleiding aan statistiek wordt gedoceerd staat los van de CEA en zal dus daarom alleen de eerste invalshoek (1) raken. Nyenrode is de enige volledige opleiding waarbij in de bachelorfase statistiek wordt gegeven.

Daarbij is het boek van Touw en Hoogduin (2011) relevant genoemd te worden. Deze auteurs (als statisticus verbonden aan een HBO-opleiding, respectievelijk zowel accountant als statisticus, verbonden aan een Big-Four-kantoor) hebben een Nederlandstalig statistiekboek geschreven dat beide invalshoeken van het vak statistiek voor accountants dekt.

De helft

In de masteropleidingen wordt vanzelfsprekend statistiek behandeld voor de eerste invalshoek, maar hoe zit het met de tweede, de statistiek om de uitoefening van het vak van accountant te ondersteunen? Een informele rondgang langs contactpersonen van de zeven Nederlandse accountancyopleidingen leert ons dat ruim de helft daarvan in de masterfase aandacht aan dit onderwerp besteedt. Complicerende factor is dat studenten tijdens hun

studie van opleiding wisselen en zo onderwerpen mislopen die op de ene opleiding pas in een latere fase aan de orde komen en op de andere opleiding al in de basisopleiding zijn gedoceerd. Zo is op Nyenrode vijftig procent van de studenten in de masterfase nieuw.

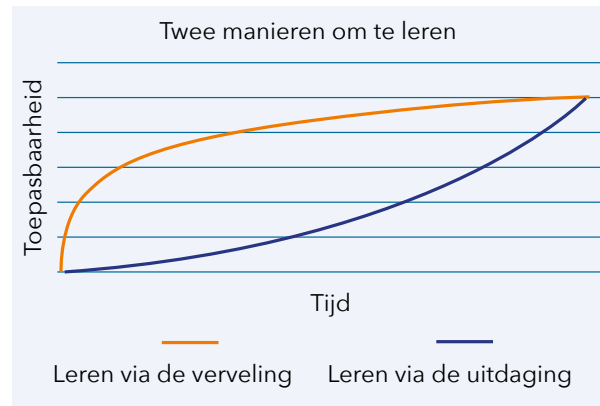
Vergeethoekje

Misschien geldt het ook wel voor andere vakken en bekwaamheden, maar voor statistiek geldt zeker, dat studenten snel vergeten. Wat in de bachelorfase wordt geleerd (en men heeft 'gekund') is in de masterfase compleet nieuw. Een ander aspect is dat studenten het nut van statistiek krijgen gedoceerd op een moment dat zij het in hun werk nog niet kunnen toepassen. Op het moment dat de stof hen wordt aangeboden, zijn zij nog niet in de positie in hun werkomgeving dat zij het geleerde in de praktijk kunnen brengen, maar slechts geacht worden te doen wat hoger-geplaatsten voorschrijven. En dan gaat, in dit geval, statistiek in het vergeethoekje.

Theoretisch fundament

Wat hier ook meespeelt is de onmogelijkheid om statistiek al toe te passen voordat er een theoretisch fundament is neergelegd. Basale kennis van wiskundige notatie, kansverdelingen, schattingen, toetsen, bias, ruis, overfitting, maar ook van datastructuren is nu eenmaal nodig voordat de studenten kunnen werken aan reallifetoepassingen.

Onderstaande figuur, die gebaseerd is op het ongepubliceerde gedachtegoed van Koos Ris, maakt dit duidelijk. Op de horizontale as staat de bestede tijd aan een statistische module. Op de verticale as staat de toepasbaarheid van het geleerde. De onderste curve toont het 'leren via de verveling' in de eerste kwantitatieve modules van een accountancyopleiding. Pas als er een theoretisch fundament enigszins gevormd is, kan de student beginnen met toepassingen in de accountancypraktijk.



Vervolgmodules kunnen dan via de bovenste curve, die het 'leren via de uitdaging' vertegenwoordigt, worden aangeboden. Reallifeproblemen worden al in het begin van zo'n vervolgmodule aangeboden, en vervolgens wordt uitgelegd hoe statistiek, wiskunde en datawetenschap aangewend kunnen worden om deze op te lossen.

Het is verrassend om te zien hoe snel studenten door kunnen stoten naar de toepassing van meer geavanceerde technieken als factoranalyse,

clusteranalyse, classificatie en machine learning als hun basiskennis over de kracht en significantie van modellen stevig is.

Materiaal

Er is materiaal beschikbaar voor de masterfase van accountancyopleidingen. De Stuurgroep Statistical Auditing van het Limperg Instituut heeft een cyclus van vier dagdelen ontwikkeld waarvan het materiaal rechtenvrij is, mits er een bekwame docent is om het materiaal te gebruiken bij het verzorgen van de colleges.

Deze stuurgroep (opgericht in 1992 onder voorzitterschap van prof. J.H. Blokdijk) is een zelfbenoemde groep van gebruikers en kenners van statistiek in de controle, komend vanuit de academische wereld, de big four, zbo's met een controle- of toezichtfunctie, en het Rijk. Voornaamste bekendheid van deze groep komt door de organisatie van het jaarlijkse symposium (vorig jaar 17 mei 2017, dit jaar 23 mei 2018) op de Vrije Universiteit over controle, data, en statistiek. Via de auteurs van dit artikel (die lid, respectievelijk voorzitter van deze Stuurgroep zijn) is dit materiaal in te zien.

In het lesmateriaal dat deze groep aanbiedt komt aan de orde:

- De plaats van gegevensgerichte controle in de standaarden;
- De plaats van data-analyse in de controle;
- Opzet en evaluatie van geldsteekproeven;
- Steekproefselectie;
- Postensteekproeven voor schattingen;

- Toerekenen van materialiteit;
- Het risicoanalysemodel en Bayesiaanse statistiek.

Literatuur

Burgstahler, D., S.M. Glover, J. Jiambalvo, (2000); Error Projection and Uncertainty in the Evaluation of Aggregate Error; *AUDITING: A Journal of Practice and Theory* Vol. 19 No. 1.

Touw, P., L.A. Hoogduin (2011); *Statistiek voor Audit en Controlling*, Academic Service.

Noot

- 1 Dit artikel is eerder geplaatst als column Statistical Auditing 61 op accountant.nl

Materiality - Clarity?¹

Niels van Nieuw Amerongen en Paul van Batenburg

Samenvatting

Over het onderwerp Materialiteit is gek genoeg relatief weinig onderzoek gepubliceerd, zeker als we dit vergelijken met de enorme hoeveelheid research die is gedaan op het terrein van de risicoanalyse. Voor het merendeel van deze paragraaf maken we dankbaar gebruik van een door Messier e.a. (2005) geschreven overzichtsartikel. Samenvattend passeren in deze paragraaf de volgende onderwerpen de revue:

- Begrippenkader
- Een Nederlandse studie naar het gebruik van planningsmaterialiteit
- De invloed van persoonskarakteristieken op materialiteit
- Foutevaluatie (foutprojectie)
- Foutevaluatie (roll-overeffect)
- Materialiteit en beslissingen van gebruikers
- Materialiteit en tekortkomingen in interne beheersing.

Inhoudsopgave

- 1 Inleiding
- 2 Samenvatting onderzoeksliteratuur voor clarity
 - 2.1 Begrippenkader
 - 2.2 Een Nederlandse studie naar het gebruik van materialiteit
 - 2.3 De invloed van persoonskarakteristieken op materialiteit
 - 2.4 Foutevaluatie (foutprojectie)
 - 2.5 Foutevaluatie (roll-over effect)
 - 2.6 Impact van materiële fouten op economische beslissingen van gebruikers
 - 2.7 Materialiteit en tekortkomingen in interne beheersing
- 3 Het materialiteitsbegrip in de nieuwe controlestandaarden
- 4 Materialiteit in de plannings- en uitvoeringsfase
 - 4.1 Uitvoeringsmaterialiteit en niet-onbeduidend
 - 4.2 Gegevensgerichte werkzaamheden
- 5 Materialiteit in de evaluatiefase
 - 5.1 Begrippen
 - 5.2 Communicatie en correctie van afwijkingen
 - 5.3 Niveau van evaluatie en samenstelling overzicht van controleverschillen
 - 5.4 Basisvragen in de evaluatiefase
- 6 Toekennen van materialiteit naar groepsonderdelen bij groepscontroles (ISA 600)
- 7 Terug naar de jaarrekening: uitvoeringsmaterialiteit en toegestane afwijking: Clarity?
- 8 Slotbeschouwing

Literatuur

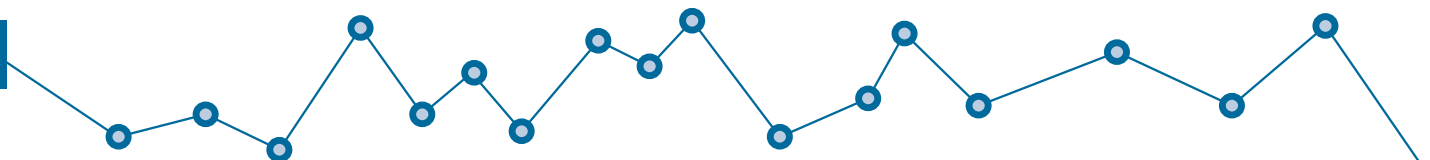
1 Inleiding

Het onderwerp Materialiteit staat regelmatig in de belangstelling. Zo komt het vaak voor dat accountants aansprakelijk worden gesteld voor niet-ontdekte of niet-gecorrigeerde fouten. Accountants hanteren in dit soort gevallen nogal eens het verweer dat de fouten niet materieel genoeg waren om te worden gecorrigeerd. Een van de meest bekende voorbeelden betreft een niet-gecorrigeerde fout van USD 51 miljoen in de door Arthur Andersen goedgekeurde jaarrekening 1997 van Enron. Indertijd fulmineerde de toenmalige SEC-topman Arthur Levitt zeer heftig tegen dergelijke verweren: 'Some companies misuse the concept of materiality. They intentionally record errors within a defined percentage ceiling. They then try to excuse that fib by arguing that the effect on the bottom line is too small to matter. If that's the case, why do they work so hard to create these errors? Maybe because the effect can matter, especially if it picks up that last penny of the consensus estimate. When either management or the outside auditors are questioned about these clear violations of GAAP, they answer sheepishly... It doesn't matter. It's immaterial.'

Zodra ondernemingen moeite gaan doen om bepaalde correcties niet door te voeren, heeft dit toch de schijn tegen dat het hier zou gaan om materiële correcties: gebruikers zouden hun economische beslissingen hierdoor laten leiden. In de praktijk is sprake van een nogal grote variatie in de omvang van de gehanteerde materialiteit. Een aantal

accountants mag in dat verband graag wijzen op het fenomeen 'professional judgement', waarmee onderlinge verschillen tussen accountants als vanzelfsprekend worden geaccepteerd.

In dit artikel beogen wij een analyse te geven van de hernieuwde controlestandaarden die het onderwerp 'materialiteit' behandelen. Een centrale vraag hierbij is in hoeverre de nieuwe controlestandaarden daadwerkelijk een 'clarification' zijn vergeleken met de bestaande controlestandaarden. Daarom zal, als aanloop naar de analyse van de nieuwe controlestandaarden, eerst een beeld worden geschetst hoe er in de onderzoeksliteratuur wordt aangekeken tegen het begrip materialiteit. Deze analyse heeft als doel om na te gaan hoe accountants omgaan met materialiteit in de planning van de controle alsmede hoe ze fouten in het licht van de materialiteit evalueren. In het tweede deel van dit artikel gaan wij nader in op ISA 320 (materialiteit in de plannings- en uitvoeringsfase) en 450 (materialiteit in de evaluatiefase) die in het kader van ISA Clarity onlangs zijn gewijzigd respectievelijk nieuw opgesteld. In dat kader wordt een basissituatie onderscheiden (controle van de enkelvoudige jaarrekening), alsook een tweetal specifieke aspecten, te weten: (a) het toekennen van de materialiteit aan groepsonderdelen, en (b) het materialiteitsbegrip bij de evaluatie van tekortkomingen in het stelsel van interne beheersing. In het nu voorliggende artikel zullen wij geen uitspraken doen over de juistheid van de absolute hoogte van de materialiteit. Wel richt onze analyse zich op het geven van handvatten voor het toepassen van



de materialiteit (zowel in de plannings- als in de evaluatiefase), waar mogelijk binnen de kaders van de geldende controlestandaarden.

2 Samenvatting onderzoeksliteratuur voor clarity

Over het onderwerp Materialiteit is gek genoeg relatief weinig onderzoek gepubliceerd, zeker als we dit vergelijken met de enorme hoeveelheid research die is gedaan op het terrein van de risicoanalyse. Voor het merendeel van deze paragraaf maken we dankbaar gebruik van een door Messier e.a. (2005) geschreven overzichtsartikel. Samenvattend passeren in deze paragraaf de volgende onderwerpen de revue:

- Begrippenkader
- Een Nederlandse studie naar het gebruik van planningsmaterialiteit
- De invloed van persoonskarakteristieken op materialiteit
- Foutevaluatie (foutprojectie)
- Foutevaluatie (roll-overeffect)
- Materialiteit en beslissingen van gebruikers
- Materialiteit en tekortkomingen in interne beheersing

2.1 Begrippenkader

Volgens Messier e.a. (2005) wordt in de planningsfase eerst een overall materialiteitsniveau bepaald. Dit bedrag wordt doorgaans planningsmaterialiteit of controlematerialiteit genoemd. In het algemeen, zo beweren zij, wordt een deel

van de planningsmaterialiteit toegekend aan balansposten of transactiestromen. Dit wordt, aldus de auteurs, toegestane fout genoemd. In de evaluatiefase dienen afwijkingen te worden geprojecteerd naar de jaarrekening. Uit een onderzoek van Allen en Elder (2005) komt naar voren dat foutprojectie in een groot aantal gevallen niet plaatsvindt, onder andere met een beroep op (im)materialiteit. In de evaluatiefase worden fouten gerelateerd aan de (definitieve) materialiteit (Braun 2001; Messier e.a. 2005; Burgstahler e.a. 2000). Een aanvullend begrip dat opgeld doet in de evaluatie is het zogenaamde 'roll over-' of 'turn-around effect'. Dat is het effect van in voorgaande jaren niet gecorrigeerde fouten.

2.2 Een Nederlandse studie naar het gebruik van planningsmaterialiteit

Het merendeel van de onderzoeken op het terrein van de materialiteit is gericht op de Amerikaanse accountantspraktijk. Een internationaal gepubliceerde studie naar de Nederlandse praktijken verscheen in 2003, uitgevoerd door Blokdijk e.a. In deze studie werden 108 jaarrekeningcontroles 1998/1999, uitgevoerd door (op dat moment) Big5 en non-Big5 accountantskantoren, betrokken. Samenvattend concludeerden de onderzoekers dat de materialiteit geen constant percentage van een basisbedrag bedroeg, maar dat dit toenam (afnemend in toename) met een toename van de cliëntgrootte. Dit is een opvallende onderzoeksbevinding. Het lijkt voor de hand te liggen dat een grotere omvang van de bedrijfsactiviteiten gepaard gaat met ruimere marges waarmee

wordt gecontroleerd. Anderzijds zou verwacht mogen worden dat grotere bedrijven zorgen voor hogere exposure voor de accountant (meer van belang voor de maatschappij, zeker van belang in geval de onderneming ook beursgenoteerd is), waardoor juist minder ruime marges worden geaccepteerd in het plannen van het vergaren van voldoende en geschikt bewijsmateriaal. Dit is ook wat we in de Nederlandse MKB-accountantspraktijk veelal zien (zie bijvoorbeeld de SRA-controlemethode). Een andere bevinding die in het onderzoek wordt gerapporteerd, betreft het resultaat dat Big5 accountantsfirma's werken met significant lagere niveaus van materialiteit vergeleken met non-Big5 accountantsfirma's. De onderzoekers zien dit resultaat in lijn liggen met andere onderzoeken die veelal rapporteren dat Big5 accountantsfirma's een hoger kwaliteitsniveau etaleren vergeleken met non-Big5 accountantsfirma's. Uit het onderzoek wordt echter niet direct duidelijk in hoeverre de basisbevinding – het niveau van materialiteit neemt toe met een toename van de ondernemingsgrootte – vooral gerealiseerd wordt in de categorie non-Big5 ondernemingen.

Andere interessante resultaten betreffen:

- het niveau van de materialiteit neemt toe bij een toename van de kwaliteit van de interne beheersingsomgeving van de cliënt ² ;
- het niveau van de materialiteit neemt toe bij een toenemend rendement op activa;
- het niveau van de materialiteit neemt af bij een toenemende complexiteit van de gecontroleerde onderneming;

- in geval de gerapporteerde resultaten het nul-niveau naderen, hanteren accountants doorgaans een lager niveau van materialiteit.

Deze bevindingen laten zien dat accountants bij het bepalen van de hoogte van de materialiteit niet alleen rekening houden met kwantitatieve factoren (percentages van grootheden), maar zeker ook met kwalitatieve factoren. Voor wat betreft break-evenniveaus van gerapporteerde resultaten is het logisch dat accountants scherper controleren; immers, voor gebruikers wordt de vraag naar het werkelijke resultaat belangrijker naarmate het in de buurt van 0 komt en de kans op mogelijke resultaatbeïnvloeding door de leiding is hier een reëel risico. Wat helaas niet uit het onderzoek blijkt, is in hoeverre hierin ook de verwachtingen van gebruikers zijn betrokken (banken, beleggers, enz.). Ook zou het interessant zijn geweest als in het onderzoek tevens op dit punt onderscheid zou zijn gemaakt naar bevindingen bij Big5 en bij non-Big5 ondernemingen alsmede naar bevindingen bij beursgenoteerde en niet-beursgenoteerde ondernemingen.

Zoals hiervoor weergegeven, heeft een adequate interne-beheersingsomgeving een positieve invloed op de hoogte van de materialiteit. Deze bevinding is overigens consistent gezien vanuit een ander onderzoekdomein, namelijk het onderzoek naar het constateren van fouten. Uit een onderzoek van Wallace en Kreutzfeldt (1995) komt naar voren dat wanneer sprake is van een onderneming met goede management controls

en goede controls op transactieniveau ('detailed controls') tegelijk ook sprake is van minder door de accountant gevonden fouten. Deze bevinding vormt overigens een bouwsteen voor het traditionele risicoanalysemodel, waarbij een goed raamwerk van interne beheersing leidt tot een lagere kans op het niet ontdekt en/of voorkomen worden van fouten door interne-beheersingsmaatregelen. Zie in dat verband ook Eilifsen en Messier (2000).

2.3 De invloed van persoonskarakteristieken op materialiteit

Uit verschillende onderzoeken komt naar voren dat verschillende accountants tot verschillende materialiteitsniveaus komen. Dit is wellicht onwenselijk maar is onontkoombaar omdat professional judgement gevraagd wordt bij het bepalen van het materialiteitsniveau. Een verschillend niveau van materialiteit heeft consequenties voor de effectiviteit van de geplande controlewerkzaamheden (worden voldoende werkzaamheden met voldoende diepgang gepland?) en de efficiency van de geplande controlewerkzaamheden (worden niet te veel werkzaamheden met te veel diepgang gepland?). De onderzoekers Estes en Reames (1988) onderzochten daarom welke persoonskarakteristieken van invloed zijn op materialiteitsinschattingen. Zij voerden een experiment uit bestaande uit twee basiscases: (1) het maken van inschattingen van oninbaarheid van een significant grote vordering, (2) het noemen van een drempelwaarde waarboven een inschatting van incurante voorraad – in geval geen correctie zou worden doorgevoerd – zou leiden tot een

niet-goedkeurende controleverklaring. Aan het onderzoek deden circa 600 CPA's mee. Naast materialiteitsinschattingen werd de respondenten tevens gevraagd aan te geven hoeveel vertrouwen ze hadden in hun eigen inschattingen.

De invloed van algemene ervaring

De onderzoekers concluderen dat een toename van algemene ervaring wel het vertrouwen in de eigen oordeelsvorming in positieve zin beïnvloedt, maar niet het niveau van de materialiteit zelf. Er is dus geen sprake van significante verschillen in materialiteitsinschattingen tussen relatief onervarenen en de meer ervarenen. De onderzoekers leiden hieruit de suggestie af dat materialiteitsinschattingen heel goed door assistent-accountants kunnen worden uitgevoerd. Hierin ligt de veronderstelling besloten dat materialiteitsbeslissingen goed middels academische studie geleerd kunnen worden, en in veel mindere mate door ervaring. De hier vermelde onderzoeksresultaten zijn overigens in tegenspraak met die van Messier (1983). Messier rapporteert juist significante ervaringseffecten. Een opvallend verschil tussen beide onderzoeken is het onderzochte functieniveau. Messier onderzocht vooral het functieniveau partners en concludeert dat minder ervaren partners doorgaans lagere materialiteitsniveaus hanteren dan meer ervaren partners. Dit ervaringseffect zou (maar dat is speculatie) te maken kunnen hebben met het verschijnsel dat minder ervaren partners doorgaans rechter in de leer zijn, terwijl meer ervaren partners meer focus hebben op commercie (waaronder aandacht voor audit efficiency).

De invloed van opleiding

De onderzoekers gebruikten verschillende maatstaven om invloeden van opleiding te meten: (1) college credits in accounting, (2) aantal jaren college gevolgd, (3) in hoeverre colleges controleleer gevolgd waren. Geen van deze factoren beïnvloedt de hoogte van materialiteitsinschattingen noch het vertrouwen in de eigen beslissingen. De gehanteerde maatstaven komen op ons wat merkwaardig over, aangezien we zouden verwachten dat CPA's een min of meer vergelijkbare vooropleiding zouden hebben genoten, dan wel dat ten minste het vak controleleer onderdeel zou uitmaken van het verplichte curriculum van de accountantsopleiding. Kijken we naar de beschrijvende statistieken dan zien we dit ook bevestigd (98% van de respondenten heeft colleges controleleer gevolgd). Relateren we de variabele 'opleiding' aan de uitkomsten op de variabele 'algemene ervaring' dan zou het ons inziens meer voor de hand hebben gelegen om, naast CPA's ook studenten in het onderzoek te hebben betrokken. Dat zou ons inziens een meer valide basis geven aan de juistheid van de veronderstelling dat het maken van materialiteitsbeslissingen vooral tijdens de opleiding wordt geleerd en minder in de praktijk.

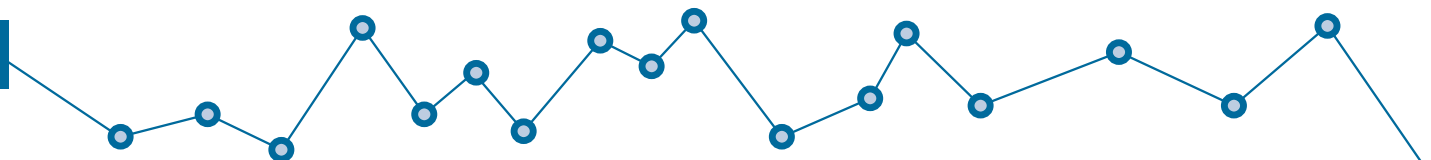
De invloed van huidige functie

Uit de beschrijvende statistiek van het onderzoek volgt dat 46% van de respondenten op het moment van het onderzoek werkzaam is in het openbare accountantsberoep en 54% buiten het openbare accountantsberoep. De onderzoekers

rapporteren dat openbaar accountants in de eerste case (inschatting oninbaarheid vorderingen) gemiddeld een 8% hogere waarschijnlijkheid van niet-inbaarheid zouden accepteren alvorens te komen tot een niet-goedkeurende controleverklaring. Aan de andere kant hebben openbaar accountants echter minder vertrouwen in hun eigen oordeelsvorming. Een dergelijk significant verschil troffen de auteurs niet aan bij de tweede case (incourante voorraad). Een verklaring voor de gevonden resultaten wordt hier niet gerapporteerd. Het zou interessant zijn om hiernaar specifiek onderzoek te doen. Zouden verschillen tussen openbaar accountants en niet-openbaar accountants mogelijk veroorzaakt kunnen worden door een verschil in focus bij materialiteitsinschattingen? Hier kan mogelijk gedacht worden aan een focus van de openbaar accountant op het commerciële belang van behoud van een controleopdracht (waardoor minder conservatief of vaktechnisch strikt wordt aangekeken tegen geconstateerde onzekerheden in jaarrekeningposten) respectievelijk de focus van de controller (met CPA-achtergrond) waarvoor geldt dat oninbaarheid direct leidt tot een lagere cash-inflow met een conservatiever inschatting tot gevolg.

De invloed van sekse en leeftijd

Uit het onderzoek van Estes en Reames volgt geen significante impact van sekse op de hoogte van materialiteitsbeslissingen. Wel trof men een verschil aan in vertrouwen in de eigen beslissingen. Bij de case over oninbaarheid vorderingen hadden vrouwen meer vertrouwen in de eigen beslissin-



gen, bij de case over incourantheid voorraden juist minder. Dit onderzoeksresultaat laat zich niet gemakkelijk interpreteren en verder onderzoek is daarom noodzakelijk op dit terrein. Voor wat betreft leeftijd vonden de onderzoekers dat oudere accountants meer conservatief zijn vergeleken met jongere accountants.

Als we de onderzoeken van Blokdijk e.a. (2003) en Estes en Reames (1988) met elkaar vergelijken dan lijken toch de cliënt-specifieke factoren en grootte van het accountantskantoor (Big5 versus non-Big5) de belangrijkste variabelen te zijn voor verschillen in materialiteitsinschattingen.

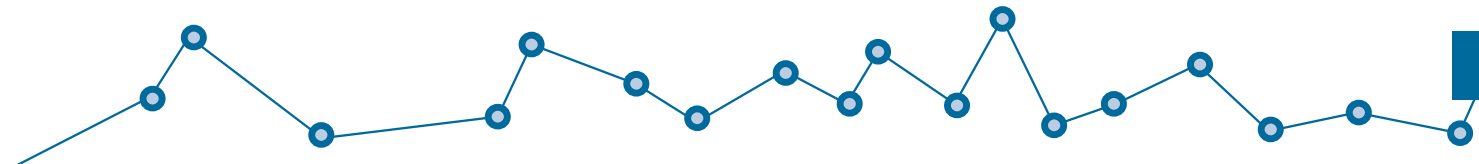
2.4 Foutevaluatie (foutprojectie)

Het evalueren van fouten wordt gezien als een complex proces van oordeelsvorming. Het vraagt om het (over)wegen van daadwerkelijk gevonden fouten, het projecteren van fouten, en het inschatten van het steekproefrisico voor de verschillende onderdelen van de controle. Burgstahler e.a. (2000) voerden een experiment uit waarbij bijzondere focus lag op het projecteren van fouten. In het onderzoek werd een casus gebruikt waarin sprake was van vier bekende fouten in jaarrekeningposten, en van twee aanvullende jaarrekeningposten werden de uitkomsten van de gegevensgerichte werkzaamheden vermeld. Ook in dit experiment werden de deelnemers gevraagd om bedragen te noemen waarboven ze (ervan uitgaande dat de huishouding niet zou corrigeren) geen goedkeurende controleverklaring zouden verstrekken. Een groep van accoun-

tants werd expliciet gevraagd om foutprojecties te maken, de andere groep niet. Een belangrijke onderzoeksuitkomst betreft het resultaat dat accountants de neiging hebben zowel de hoogte van de foutprojectie als de onzekerheid als gevolg van het steekproefrisico te laag in te schatten. De onderzoekers speculeren dat de onderzoeksresultaten het gevolg kunnen zijn van: (1) onwetendheid van accountants op het gebied van betrouwbaarheidsniveaus bij steekproeven, (2) de verwachting dat lagere niveaus van aanvaardbare fouten zouden leiden tot teleurstellingen aan cliëntzijde. Beide speculaties zijn zorgelijk voor het accountantsberoep. Een van de oplossingen die de onderzoekers suggereren is dat aan deze problematiek specifieke aandacht wordt besteed tijdens interne kwaliteitsinspecties.

2.5 Foutevaluatie (roll-over effect)

Accountantskantoren hanteren verschillende methoden of formats bij het evalueren van fouten. Een van die verschillen betreft het punt in hoeverre ongecorrigeerde fouten ontdekt in voorgaande controleperioden worden meegenomen in de foutevaluatie van het huidige controlejaar. Nelson e.a. (2005) deden hiernaar onderzoek door middel van een experiment waarbij de deelnemers werden onderverdeeld in twee groepen. De ene groep kreeg een format aangereikt waarin fouten uit voorgaande jaren waren opgenomen (de zogenaamde cumulatieve benadering), terwijl de andere groep een format kreeg aangereikt zonder fouten voorgaande jaren (de zogenaamde huidig jaar benadering). De resultaten suggereren dat ac-



countants sneller geneigd zijn om aanpassing van gevonden fouten te vereisen indien het cumulatieve effect van de fouten hoger is. Het format dat accountantskantoren hanteren doet er derhalve kennelijk toe. De onderzoekers bevelen de standard setters daarom aan om te vereisen dat fouten altijd gecorrigeerd moeten worden. Inmiddels is in ISA 450 hieraan opvolging gegeven.

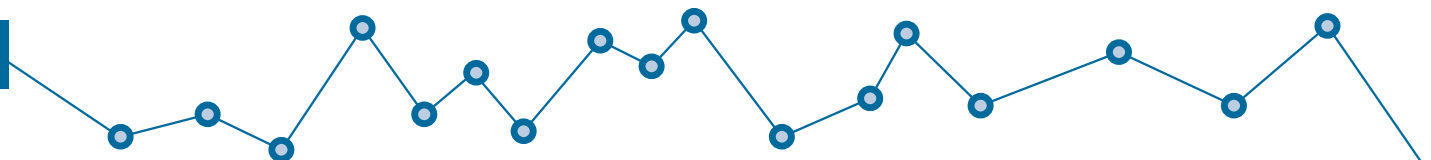
Braun (2001) deed onderzoek naar de oordeelsvorming van accountants bij het al dan niet vereisen van correctie van voorgestelde controleverschillen. Een van hun conclusies is dat accountants doorgaans minder geneigd zijn om correctie te vereisen indien de impact van de correcties het resultaat in positieve zin beïnvloedt. Dit onderzoeksresultaat wekt de suggestie dat accountants een rechttere rug tonen bij voorgestelde correcties die het resultaat negatief beïnvloeden. De cliënt heeft er doorgaans belang bij dat de verwachtingen van stakeholders worden gerealiseerd. Voorgestelde correcties met negatieve resultaatimpact dragen daaraan niet bij.

Onderzoek van Wright en Wright (1997) toont aan dat accountants meer staan op het corrigeren van objectief bepaalde controleverschillen vergeleken met subjectief bepaalde controleverschillen (bijvoorbeeld schattingsverschillen). Ook suggereren deze onderzoekers dat controleverschillen sneller worden 'gewaived' naarmate de omvang van de onderneming toeneemt. Deze tolerantiemarges bij grotere ondernemingen zijn daardoor ruimer vergeleken met kleinere ondernemingen. Deze

suggestie lijkt in tegenspraak met de resultaten uit het later gepubliceerde en eerder in dit artikel besproken onderzoek van Blokdijk e.a. (2005).

2.6 Impact van materiële fouten op economische beslissingen van gebruikers

Materialiteit is een begrip dat vooral wordt gedefinieerd vanuit gebruikersperspectief. In simpele bewoordingen: een fout is pas materieel als het economische beslissingen van jaarrekeninggebruikers beïnvloedt. Er is helaas nog erg weinig onderzoek gedaan naar dit gebruikersperspectief. Een onderzoek dat hierin wel voorziet, is dat van Tuttle e.a. (2002). De samenstelling van de gebruikerskring is divers te noemen. De onderzoekers hebben ervoor gekozen om een specifiek type gebruiker in het onderzoek te betrekken: beleggers. Hiertoe werden business students als 'surrogaat' voor beleggers gebruikt³. Het experiment maakte gebruik van computers op basis waarvan de deelnemers konden beleggen in korte tijdsperiodes van elk drie minuten. Tijdens de handelsperiode werden contract, bid en ask niveaus op een projectiescherm getoond. Voorafgaand aan elke handelsperiode werden de deelnemers voorzien van financiële informatie van het bedrijf waarin ze mochten handelen. In de eerste handelsperiodes was de gepresenteerde informatie fout-vrij. In latere sessies werden fouten gemanipuleerd in de verschillende groepen van participanten in het experiment. Op deze wijze werd nagebootst in hoeverre een verandering in de financiële cijfers van invloed was op concrete (des)investeringsbeslissingen. Ter benchmarking van de beslissingen



van gebruikers werd een tweetal materialiteitsniveaus door de onderzoekers gegeven, gebaseerd op gangbare uitgangspunten van accountantskantoren: een liberaal en een conservatief niveau van materialiteit. De uitkomsten van het onderzoek suggereren dat gebruikers hun beslissingen baseren op het meest liberale niveau van materialiteit: met andere woorden, vallen de fouten in het experiment binnen de benchmark range (liberaal-conservatief) dan leidt dit niet tot andere economische beslissingen. De onderzoekers concluderen daarom dat accountants in een aantal gevallen een te conservatief niveau van materialiteit hanteren bij het evalueren van fouten in de jaarrekeningcontrole. Tevens waarschuwen de onderzoekers rechtbanken ervoor om al te snel te concluderen dat (achteraf gezien) de beslissingen van gebruikers anders zouden hebben uitgepakt indien de (achteraf) gedetecteerde fouten zouden zijn verwerkt in de jaarrekening. De onderzoekers erkennen als belangrijke beperking in hun onderzoek dat ze alleen de absolute hoogte van fouten hebben onderzocht zonder daarbij onderscheid te maken naar de aard van de jaarrekeningfouten.

Tegenover dit onderzoeksresultaat staan onderzoeken naar de audit expectation gap (Humphrey, 1997; Lee e.a., 2009), die aangeven dat het maatschappelijk verkeer juist een lagere materialiteit verwacht dan door accountants gehanteerd, of zich zelfs nog moet realiseren dat het begrip materialiteit überhaupt bestaat en een rol speelt in de controleaanpak en daarmee de evaluatie van door de accountant geconstateerde controleverschillen.

Deze resultaten dienen onzes inziens echter wel in het perspectief te worden geplaatst van onderzoeken waarbij de focus vooral op de verwachtingskloof tussen accountant en maatschappelijk verkeer lag. Uit dergelijke onderzoeken volgt dat een deel van deze verwachtingskloof te maken heeft met onbegrip over door accountants gehanteerde begrippen als 'redelijke mate van zekerheid' en 'materialiteit'. Zo blijkt bijvoorbeeld uit een van deze onderzoeken (Church e.a., 2008) dat het maatschappelijk verkeer ermee gebaat is als de accountant expliciet rapporteert over het bij de controle gehanteerde materialiteitsniveau.

2.7 Materialiteit en tekortkomingen in interne beheersing

Het concept materialiteit heeft mede naar aanleiding van het Enron-debacle en andere boekhoudschandalen een nieuwe dimensie gekregen. Met de implementatie van Auditing Standard No. 2 in 2004 (in 2007 vervangen door AS No. 5) heeft de PCAOB in haar controlestandaarden een directe relatie gelegd tussen het begrip materialiteit en het plannen van testwerkzaamheden inzake het stelsel van interne beheersingsmaatregelen (planningsfase) alsmede het evalueren van tekortkomingen in de interne beheersing (evaluatiefase). Volgens deze standaard kunnen tekortkomingen worden gecategoriseerd in: tekortkomingen, significante tekortkomingen en materiële zwakten. Vanuit onderzoeksperspectief is echter nog relatief weinig onderzoek gedaan naar dit onderwerp. De enige studie die ons op dit terrein bekend is, betreft een (gedateerd)

onderzoek van Mayper (1982). In dit onderzoek werden drie onafhankelijke variabelen betrokken: het type ontbrekende interne beheersingsmaatregel (het ontbreken van functiescheiding of autorisaties), het type activa (kas, materiële vaste activa, voorraden) en het meest waarschijnlijke effect in geldswaarde. Uit het onderzoek komt de conclusie naar voren dat het niveau van consensus (mate waarin de oordeelsvorming van accountants onderling overeenstemt) van een gematigd niveau is (73%) als het gaat om het evalueren van tekortkomingen in interne beheersing. Als we hierbij bedenken dat in het onderzoek vooral senior accountants participeerden, komt het consensusniveau ons eerder hoog dan gematigd voor. Een interessant onderzoeksresultaat betreft voorts de suggestie dat een meerderheid van accountants in hun oordeelsvorming ‘configural processing’ laat zien. Dit impliceert dat het evalueren van tekortkomingen geen lineair verband laat zien waarbij individuele tekortkomingen geïsoleerd worden geëvalueerd, maar juist in onderlinge samenhang. Dit wordt in de praktijk juist als een van de moeilijkste evaluatieprocessen gezien: de aggregatie van tekortkomingen in interne beheersing, waarbij de tekortkomingen ook in hun onderlinge samenhang worden gezien. Dit onderzoeksresultaat wordt bij andere typen beslissingen, zoals bij risicoanalyse inschattingen, niet vaak gerapporteerd. Een van de weinige onderzoeken die dit wel rapporteert, is het onderzoek van Brown en Solomon (1991).

3 Het materialiteitsbegrip in de nieuwe controlestandaarden

In de vorige sectie is een uiteenzetting opgenomen van uitkomsten van uitgevoerde onderzoeken over materialiteit. Uit deze onderzoeken is onder meer gebleken dat het begrip materialiteit door accountants niet altijd eenduidig werd gehanteerd onder de toen geldende internationale controlestandaarden. In deze sectie gaan we nader in op de nieuwe controlestandaarden voor zover dit het onderwerp materialiteit raakt. De nieuwe controlestandaarden vloeien voort uit het zogenaamde Clarity project van het IFAC. In dit grootschalige project heeft het IFAC een aantal bestaande controlestandaarden tekstueel herschreven (‘redrafted’), een aantal andere controlestandaarden herzien (‘revised’) en ten slotte een beperkt aantal nieuwe standaarden ontwikkeld. In het kader van dit artikel zal de aandacht vooral liggen op ISA 320 (herziening van bestaande ISA 320 die het onderwerp materialiteit in de planningsfase behandelt) en ISA 450 (nieuwe standaard, die is gericht op het onderwerp materialiteit in de evaluatiefase). Wij ronden dit artikel af met een beschouwing over de vraag of de nieuwe controlestandaarden voor wat betreft het onderwerp materialiteit inderdaad de bedoelde helderheid (‘clarification’) verschaffen.

4 Materialiteit in de plannings- en uitvoeringsfase

Startpunt voor een verhandeling over de rol van materialiteit in de planning en evaluatie van de controle zou een explicitering moeten zijn van het doel van de controle, zoals bijvoorbeeld in ISA 200.3: Het doel van een controle is het versterken van de mate waarin de beoogde gebruikers in de financiële overzichten vertrouwen stellen. Dit wordt bewerkstelligd door het door de accountant tot uitdrukking brengen van een oordeel of de financiële overzichten in alle van materieel belang zijnde opzichten opgesteld zijn in overeenstemming met een van toepassing zijnd stelsel inzake financiële verslaggeving.

Het begrip materialiteit is voor de accountant een sleutelbegrip voor het plannen en uitvoeren van de controle. Uit ISA 320.6 volgt dat de materialiteit als input geldt voor het bepalen van aard, timing en omvang van uit te voeren controlewerkzaamheden. De materialiteit bepaalt dus om zo te zeggen de diepgang van de geplande controlewerkzaamheden. Concrete handvatten, zoals deze bijvoorbeeld zijn opgenomen in kantoor specifieke controlehandboeken, worden niet verstrekt.

4.1 Uitvoeringsmaterialiteit en niet-onbeduidend

ISA 320 introduceert een nieuw begrip: de uitvoeringsmaterialiteit. Onder uitvoeringsmaterialiteit wordt het volgende verstaan (ISA 320.9): Het bedrag of de bedragen die door de accountant op een lager materialiteitsniveau voor de financiële overzichten als geheel is (zijn) vastgesteld om de

waarschijnlijkheid dat het geheel van niet-gecorrigeerde en niet-ontdekte afwijkingen het materialiteitsniveau overstijgt tot een passend laag niveau te reduceren.

Hieruit volgt dat bij het plannen van de controle rekening dient te worden gehouden met een lager niveau dan de (verslaggevings)materialiteit in verband met onder andere niet-ontdekte afwijkingen. Naast het begrip uitvoeringsmaterialiteit verschijnt in ISA 530 ('audit sampling') het begrip 'toegestane afwijking' (tolerable misstatement) waarvan we eerder zagen dat dit begrip voorheen ook in wetenschappelijke onderzoeken werd gehanteerd. ISA 320 geeft geen concrete aanwijzingen hoe gekomen moet of kan worden tot het bedrag van de uitvoeringsmaterialiteit. In de praktijk tot vóór ISA Clarity werd de 'toegestane afwijking' meestal bepaald als percentage van de overall materialiteit, variërend van 50% (hoog risico cliënten) tot 90% (laag risico cliënten), waarbij het vooral kwalitatieve overwegingen waren die de omvang van het gehanteerde percentage beïnvloedden. ISA 320 geeft geen concrete aanwijzingen voor de bepaling van de uitvoeringsmaterialiteit noch geeft ISA 530 dit voor de bepaling van de toegestane afwijking. Daarom is het van belang de rol van deze begrippen nader te duiden zodat tot een afgewogen oordeelsvorming kan worden gekomen.

Een specifiek punt van aandacht in ISA 320 betreft de vraag in hoeverre het verplicht is om de materialiteit toe te rekenen naar jaarrekeningposten

en/of -toelichtingen. In expliciete bewoordingen lijkt dat niet het geval te zijn, maar door de introductie van de uitvoeringsmaterialiteit komt het daar wel impliciet op neer wanneer de accountant fouten in bepaalde posten wil kunnen isoleren⁴. Daarmee gaan de nieuwe standaarden verder dan de huidige waar toerekening in zijn geheel niet is voorgeschreven. De gebruikersbenadering die hier is gekozen, komt ons als redelijk argument hiertoe voor.

Los daarvan staat de bijzondere bepaling in ISA 320.10 die afzonderlijke vaststelling van de materialiteit voorschrijft ('dient') voor specifieke transactiestromen, rekeningsaldi of toelichtingen waarvan redelijkerwijs mag worden verwacht dat economische beslissingen door gebruikers daarop in het bijzonder worden gebaseerd. Er is derhalve sprake van een conditionele verplichting. Voor die posten wordt vervolgens ook een uitvoeringsmaterialiteit bepaald (ISA 320.A12).

Hoe kan nu in de praktijk de impact van materialiteit op de aard, omvang en timing van uit te voeren controlewerkzaamheden worden bepaald? Een eerste stap (uitgaande van de controle van een enkelvoudige jaarrekening, derhalve zonder rekening te houden met aspecten van groepscontroles) vormt het bepalen welke rekeningsaldi, transactiestromen en toelichtingen als materieel worden beschouwd waarvoor dientengevolge controlewerkzaamheden zullen worden gepland. Hiervoor kunnen de volgende benchmarks gelden:

- rekeningsaldi, transactiestromen en toelichtingen die het niveau van de overall uitvoeringsmaterialiteit in kwantitatieve zin te boven gaan;
- rekeningsaldi, transactiestromen en toelichtingen die het niveau van de gealloceerde uitvoeringsmaterialiteit in kwantitatieve zin te boven gaan;
- rekeningsaldi, transactiestromen en toelichtingen die in kwalitatieve zin als materieel worden beschouwd. Kwalitatieve factoren kunnen bijvoorbeeld betreffen: (1) de aanwezigheid van frauderisico's, (2) complexiteit van rekeningsaldi, transactiestromen en toelichtingen, (3) aanwezige schattingselementen, (4) mogelijke doorbreking van interne beheersing door de leiding ten aanzien van specifieke rekeningsaldi, transactiestromen en toelichtingen.

Zodra de accountant de selectie van rekeningsaldi, transactiestromen en toelichtingen heeft bepaald, is de vervolgstap het bepalen van verder uit te voeren werkzaamheden. Dit impliceert als tweede stap dat op het niveau van de jaarrekening als geheel, het niveau van rekeningsaldi, transactiestromen en toelichtingen, alsmede het niveau van getrouwheidsaspecten (controledoelstellingen) het risico op een afwijking van materieel belang wordt ingeschat. Voor concrete controlewerkzaamheden spelen significante en frauderisico's een essentiële rol. Dit onderwerp valt buiten het bestek van dit artikel. Een derde stap is het inzicht verkrijgen in de totstandkoming van de posten. Bij deze stap is het dienstig een tabel op te stellen waarbij in de eerste kolom de

rekeningen, transactiestromen en toelichtingen worden vermeld, en in de overige kolommen de onderkende gegevensverwerkende processen. Deze stap leidt tot een verder inzicht in de controleaanpak die bij de process audit (interimcontrole) wordt gehanteerd alsmede tot verdere inzichten in de totstandkoming van rekeningsaldi, transactiestromen en toelichtingen. Deze inzichten zijn mede richtinggevend voor de aard, omvang en timing van de uit te voeren gegevensgerichte werkzaamheden. Tenslotte is een belangrijk element van de controleaanpak dat raakvlakken heeft met de uitvoeringsmaterialiteit: het uitvoeren van een gegevensgerichte cijferanalyse. Uit ISA 520 volgt dat de accountant bij het uitvoeren van cijferanalyse zelf verwachtingen dient op te bouwen ten aanzien van de jaarrekening-gerelateerde cijfers. Afwijkingen tussen verwachte en werkelijk verantwoorde cijfers kunnen aanleiding zijn voor het noodzakelijkerwijs uitvoeren van aanvullende gegevensgerichte werkzaamheden. Welke grenswaarde (ISA 520.A16) hiervoor wordt gehanteerd, verschilt per accountantsorganisatie. Aspecten die hierbij aanvullend (niet limitatief) een rol spelen zijn: de absolute geldswaarde van de jaarrekeningpost en de aard van de jaarrekeningpost (routinematig, niet-routinematig, schattingspost). Deze grenswaarde mag echter niet verward worden met de praktijkregel die suggereert dat afwijkingen beneden 5%-10% van de materialiteit duidelijk onbeduidend ('clearly trivial') zouden zijn en daarom niet in de oordeelsvorming betrokken hoeven te worden (ISA450.5 en A2). Het gaat in ISA 520 namelijk

niet om werkelijke afwijkingen van materieel belang, maar om het verkrijgen van een indicatie dat mogelijk sprake is van een afwijking van materieel belang, reden om daarvoor aanvullende gegevensgerichte werkzaamheden uit te voeren.

4.2 Gegevensgerichte werkzaamheden

De materialiteit bepaalt in grote mate de hoeveelheid controlewerkzaamheden, en dat controlewerk zal altijd voor een deel bestaan uit gegevensgerichte werkzaamheden. Paragraaf A42 van ISA 330 vermeldt zelfs waarom risicoanalyse en toetsing van de interne beheersing alleen niet afdoende zijn:

ISA 330.A42: Op grond van paragraaf 20 is vereist dat de accountant gegevensgerichte controles opzet en uitvoert voor elk van de materiële transactiestromen, rekeningsaldi en in de financiële overzichten opgenomen toelichtingen, ongeacht de ingeschatte risico's van een afwijking van materieel belang. Deze vereiste geeft weer dat:

- (i). *de risico-inschatting door de accountant een kwestie van oordeelsvorming is en daardoor dus niet alle risico's van een afwijking van materieel belang zou kunnen onderkennen en*
- (ii). *er inherente beperkingen zijn met betrekking tot de interne beheersing, waaronder het doorbreken daarvan door het management.*

Gegevensgerichte werkzaamheden zijn daarom een onmisbaar bestanddeel van de controleaanpak, maar het is niet zo dat de controleaanpak tot alleen gegevensgerichte werkzaamheden beperkt kan worden:

ISA 315.29. Met betrekking tot sommige risico's kan de accountant van oordeel zijn dat het niet mogelijk is of vanuit praktisch oogpunt niet haalbaar is voldoende geschikte controle-informatie te verkrijgen door middel van gegevensgerichte werkzaamheden alleen. Dergelijke risico's kunnen betrekking hebben op de onjuiste of onvolledige vastlegging van routinematige en belangrijke transactiestromen en rekeningsaldi, waarvan de kenmerken dikwijls een hoge mate van geautomatiseerde gegevensverwerking met weinig of geen handmatige tussenkomst mogelijk maken. In dergelijke gevallen zijn de interne beheersingsmaatregelen van de entiteit relevant voor de controle en dient de accountant daarvan kennis te verkrijgen.

5 Materialiteit in de evaluatiefase

Het gebruiken van de materialiteit in de evaluatiefase wordt beschreven in ISA 450, zoals gezegd een geheel nieuwe standaard die overigens ook elementen bevat uit de voormalige ISA 320. Het doel van ISA 450 (450.3) is het evalueren van: (a) de invloed van onderkende afwijkingen, en (b) de invloed van niet-gecorrigeerde afwijkingen.

5.1 Begrippen

Het zal de lezer duidelijk zijn dat de nieuwe ISA's niet langer de term 'fouten' hanteren, maar 'afwijkingen'. Beide begrippen lijken sterke overeenkomsten met elkaar te hebben, hoewel ISA 450.4 suggereert dat het begrip fouten enger gedefinieerd lijkt te worden. Een afwijking wordt gedefinieerd als een verschil tussen een gerapporteerd

bedrag (respectievelijk rubricering, presentatie of toelichting) en het vereiste bedrag (respectievelijk rubricering, presentatie of toelichting) ingevolge het stelsel van financiële verslaggeving. Wat wel afwijkend is ten opzichte van de voormalige ISA 320 is het begrip '(niet) duidelijk onbeduidend'. In de evaluatiefase behoeft de accountant afwijkingen die duidelijk onbeduidend zijn niet mee te wegen in het eindoordeel over de verantwoording.

5.2 Communicatie en correctie van afwijkingen

Een ander nieuw element van ISA 450 is dat alle verzamelde afwijkingen gecommuniceerd dienen te worden met het passende management-niveau (dit zal veelal de directie zijn). Hieraan is gekoppeld het vereiste dat de accountant het management dient te verzoeken om alle afwijkingen (uitgezonderd die afwijkingen die duidelijk onbeduidend zijn) te corrigeren. Gelet op allerhande complexe uitvoeringsvraagstukken rond het turn-around/roll-over effect van in voorgaande jaren niet gecorrigeerde afwijkingen vinden wij de ISA-stellingname op dit punt begrijpelijk. Evenwel plaatst dit de accountant, zeker bij grote al dan niet beursgenoteerde ondernemingen, wel voor praktische problemen. Het overall beeld van gevonden afwijkingen wordt vaak zichtbaar vlak voor de eindbespreking met directie en governance organen en vlak voor afgifte van de controleverklaring. Eventuele aanpassingen kunnen dan heel erg bewerkelijk zijn en door tijdsdruk in deze fase gemakkelijk tot nieuwe (onbedoelde) afwijkingen leiden. Kortom, vaktechnisch gezien

begrijpelijk, maar praktisch gezien niet eenvoudig uitvoerbaar. Bovendien schept de ISA een boekhoudkundig probleem: voor bekende fouten is de tegenboeking bekend, voor geprojecteerde fouten niet altijd. Als de tegenboeking niet bekend is, brengt dit een vraagstuk met zich mee in de communicatie richting de cliënt.

5.3 Niveau van evaluatie en samenstelling overzicht van controleverschillen

Belangrijk is het voorschrift dat gevonden afwijkingen zowel individueel als op geaggregeerd niveau dienen te worden geëvalueerd. Dit vereiste houdt mede verband met het in ISA 320 paragraaf 10 opgenomen vereiste (in voorkomende gevallen) van het hanteren van een afzonderlijke materialiteit voor specifieke rekeningsaldi, transactiestromen en toelichtingen en is om die reden begrijpelijk.

5.4 Basisvragen in de evaluatiefase

Essentieel in ISA 450 is dat de accountant bij de evaluatie van de werkzaamheden zichzelf twee vragen stelt:

1. Heb ik voldoende werk gedaan? en
2. Wat is nu mijn conclusie over de verantwoording?

Het zal duidelijk zijn dat zonder een positief antwoord op de eerste vraag, elk antwoord op de tweede vraag nutteloos is:

ISA 450, 3: De doelstelling van de accountant is het evalueren van:

- (a). *De invloed van onderkende afwijkingen op de controle; en*

- (b). *De invloed van niet-gecorrigeerde afwijkingen, indien deze bestaan, op de financiële overzichten.*

Zo duidelijk als ISA 450 deze twee eisen verwoordt, zo onduidelijk wordt aangegeven hoe de accountant die twee vragen kan beantwoorden, zal straks blijken. Wel geeft ISA 450 goed aan wat allemaal onder afwijkingen wordt verstaan. Dat zijn niet alleen de gevonden fouten maar ook hun projectie:

450, A3: Om een hulpmiddel te vormen voor de accountant bij het evalueren van de invloed van tijdens de controle verzamelde afwijkingen en bij het communiceren van afwijkingen aan het management en aan degenen belast met governance, kan het zinvol zijn onderscheid te maken tussen feitelijke afwijkingen, inschattingsafwijkingen en geprojecteerde afwijkingen.

- *Feitelijke afwijkingen zijn afwijkingen waar geen twijfel over bestaat;*
- *Inschattingsafwijkingen zijn verschillen die ontstaan door standpunten van het management met betrekking tot schattingen die door de accountant als onredelijk worden beschouwd, dan wel door de keuze of toepassing van de grondslagen voor financiële verslaggeving die door de accountant als niet passend worden beschouwd;*
- *Geprojecteerde afwijkingen zijn de beste inschatting van de accountant van afwijkingen in populaties die de projectie inhouden van in steekproeven onderkende afwijkingen voor de populaties als geheel waaruit de steekproeven werden getrokken. Richtlijnen met betrekking tot het vaststellen van geprojecteerde afwijkingen en de evaluatie van de resultaten zijn gegeven in ISA 530.*

In de nieuwe ISA is dus expliciet gemaakt wat in de vorige versie alleen impliciet af te leiden was: de evaluatie van de controlewerkzaamheden omvat ook de extrapolatie van steekproeven.

Vraag 1: Heb ik voldoende werk gedaan?

Met zoveel woorden geeft ISA 450 in paragraaf 5 aan dat de verzamelde afwijkingen met uitzondering van die afwijkingen die duidelijk onbeduidend zijn, dienen te worden geaccumuleerd en te worden geconfronteerd met de materialiteit, en die zelfs niet mag benaderen:

450, 6: De accountant dient vast te stellen of de algehele controleaanpak en het controleprogramma moeten worden herzien indien:

- (a). De aard van de onderkende afwijkingen en de omstandigheden waarin deze plaats hebben er op wijzen dat andere afwijkingen kunnen bestaan die, wanneer geaggregeerd met tijdens de controle verzamelde afwijkingen, van materieel belang zouden kunnen zijn; of (zie: par. A4)*
- (b). Het totaal van tijdens de controle verzamelde afwijkingen de materialiteit nadert die is vastgesteld overeenkomstig ISA 320 (zie: par. A5)*

Het 'waarom' van het 'benaderen' wordt verduidelijkt in A5, maar het 'hoe' niet:

450, A5: Indien de aggregatie van de tijdens de controle verzamelde afwijkingen het niveau van materieel belang dat is vastgesteld overeenkomstig ISA 320 benadert, kan er sprake zijn van een groter dan aanvaardbaar laag risiconiveau dat mogelijk niet-ontdekte afwijkingen, indien opgeteld bij de aggregatie van de tijdens de controle verzamelde afwijkingen het niveau

van het materieel belang zou kunnen overschrijden. Niet-ontdekte afwijkingen kunnen voorkomen ten gevolge van het bestaan van een sampling risk en een non-sampling risk.

De gecumuleerde afwijking mag de materialiteit niet benaderen omdat het risico van ten onrechte goedkeuren wel eens te groot zou kunnen worden. Het subjectieve begrip 'benadert' kan echter objectief worden gemaakt door de introductie van het begrip geschatte maximale fout, voor statistici de foutbovendgrens. Het sampling risk wordt dan vanzelf ook objectief gemaakt door de kans dat de materialiteit wordt overschreden. Door de berekende foutbovendgrens bij een vooraf gekozen onbetrouwbaarheid te confronteren met de materialiteit is er hoogstens nog een kans gelijk aan die onbetrouwbaarheid dat de werkelijke fout de materialiteit overschrijdt. Als dus de uit de gegevensgerichte werkzaamheden berekende geschatte maximale fout bij een gekozen betrouwbaarheid onder de door de accountant gekozen materialiteit blijft kan de accountant stellen dat het totaal van ontdekte en niet-ontdekte fouten de materialiteit niet benadert.

De vraag 'heb ik voldoende werk gedaan?' impliceert dus dat de accountant een afweging maakt om de controleaanpak te herzien indien de gecumuleerde afwijkingen de gestelde materialiteit benaderen. Hierin is de impliciete veronderstelling begrepen dat de accountant een controleaanpak heeft ontworpen die, bij een vooraf gekozen betrouwbaarheidsniveau, aannemelijk maakt dat materiële afwijkingen worden gedetecteerd.

In de evaluatiefase van de controle is deze impliciete veronderstelling derhalve nog relevant voor zover de gevonden afwijkingen het materialiteitsniveau naderen. Het begrip maximale fout, zoals hier gebaseerd op statistische uitgangspunten heeft het voordeel dat de statistische theorie het begrip ‘redelijke mate van zekerheid’ objectief vormgeeft (dan wel kwantificeerbaar maakt). Dit laat natuurlijk onverlet dat niet elke accountant bij elke jaarrekeningcontrole alle jaarrekeningposten gegevensgericht controleert op basis van een statistische steekproef. Een niet-statistische steekproef zou evenwel niet tot een andere conclusie mogen leiden en de verantwoordelijkheid daarvoor berust bij de accountant.

Vraag 2: Wat is mijn de conclusie over de verantwoording?

Nadat de accountant heeft vastgesteld dat de aard, omvang en timing van de uitgevoerde werkzaamheden toereikend is en voordat de accountant een oordeel geeft over de verantwoording, wordt de huishouding in staat gesteld correcties aan te brengen. Belangrijk daarbij is het woord ‘alle’ in paragraaf 8 van ISA 450:

450, 8: De accountant dient alle afwijkingen die tijdens de controle zijn verzameld tijdig te communiceren aan het passende managementniveau, tenzij dit is verboden op grond van wet- of regelgeving. De accountant dient het management te verzoeken deze afwijkingen te corrigeren. (Zie: par. A7-A9).

Het woord ‘alle’ impliceert namelijk dat niet alleen gevonden afwijkingen, of verschillen van in-

zicht over inschattingen, maar ook geprojecteerde afwijkingen uit steekproeven aan de huishouding ter correctie worden voorgelegd. Voor velen is dat een novum in de ISA na Clarity, voor anderen is dit een voorbeeld van Clarity: ook foutprojecties dienen ter correctie worden voorgelegd. De huishouding zal zo’n geprojecteerde fout – alleen al om boekhoudkundige reden, want de tegenboeking is immers niet altijd bekend – niet vanzelfsprekend willen corrigeren. De accountant zal die fout dan bij zijn oordeel over de verantwoording als geheel moeten meewegen. Het verzoek tot correctie blijkt namelijk wat dwingender dan eerst aangegeven:

450, 9: Indien het management weigert om sommige of alle door de accountant gecommuniceerde afwijkingen te corrigeren, dient de accountant inzicht te verwerven in de redenen van het management voor het niet maken van de correcties en dient hij met dat inzicht rekening te houden bij het evalueren of de financiële overzichten als geheel vrij zijn van afwijkingen van materieel belang. (Zie: par. A10)

450, A13: Elke afzonderlijke afwijking wordt in overweging genomen om de invloed ervan op de relevante transactiestromen, rekeningsaldi of in de financiële overzichten opgenomen toelichtingen te evalueren, met inbegrip van de vraag of het materialiteitsniveau voor die bepaalde transactiestroom, dat bepaalde rekening-saldo of die bepaalde in de financiële overzichten opgenomen toelichting, indien aanwezig, is overschreden.

Model voor evaluatie, ter concretisering van ISA
Samenvattend geven we aan hoe ISA 450 –

ondanks Clarity – een extra slag zou kunnen maken wat betreft duidelijkheid, door enerzijds aan te geven dat foutevaluatie geschiedt op zowel het niveau van jaarrekeningposten als op niveau van de jaarrekening als geheel, en door het begrip ‘benaderen’ te vervangen door het begrip maximale fout: (zie figuur onder).

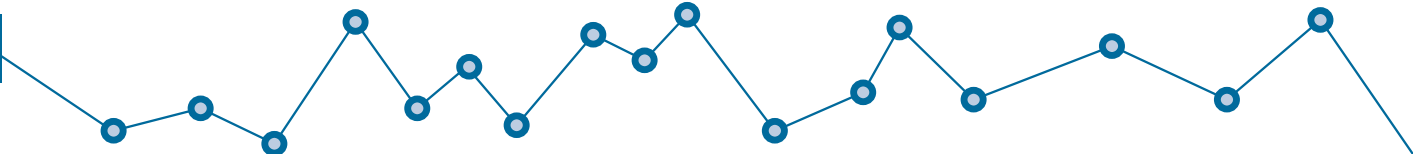
Nu is de statistiek een redelijk recht-toe-recht-aan oplossing voor een enkele post van de verantwoording, omdat daarvoor de maximale fout met eenvoudige hulpmiddelen kan worden berekend of benaderd. Voor de verantwoording als geheel is een dergelijke rekenkundige exercitie minder eenduidig. Bij gebrek aan overall maximale fout wordt daarom het risico dat de gecumuleerde fout de materialiteit overschrijdt, verkleind door de materialiteit per post lager te kiezen dan de overall materialiteit. Dat is de rol van de uitvoeringsmaterialiteit. In de hierna volgende paragraaf wordt echter betoogd dat het kiezen van een lagere materialiteit per post van de verantwoording alleen nut heeft als posten heterogeen zijn en

fouten daarom geïsoleerd kunnen worden naar de post.

6 Toekennen van materialiteit naar groepsonderdelen bij groepscontroles (ISA 600)

Een essentieel begrip in ISA 600 is de term ‘groepsonderdeel’. Paragraaf 21c vereist dat de accountant naast het bepalen van de groeps-materialiteit ook een materialiteit aan het groepsonderdeel toekent, die lager is dan de groepsmaterialiteit. Deze aanwijzing is een duidelijke aanscherping ten opzichte van de ‘pre-clarity’ versie: daarin stond dat de materialiteit per onderdeel niet groter mocht zijn dan de groepsmaterialiteit. Toekennen van materialiteit aan onderdelen moet nu. De reden waarom materialiteit moet worden toegekend staat in 21c: (...) *Teneinde het risico, dat het totaal van niet-gecorrigeerde en niet-ontdekte afwijkingen in de financiële overzichten van de groep de materialiteit voor de finan*

	Gevolgen voor verantwoording	Voldoende werk gedaan?
Per post	Bepaal verzamelde fout (bekende fout, inschattingsverschillen en geprojecteerde fout) en vraag huishouding om correctie	Confronteer maximale fout met materialiteit
Verantwoording als geheel	Accumuleer niet-gecorrigeerde verzamelde fout voor zover niet-duidelijk onbeduidend en confronteer met de materialiteit	Confronteer overall maximale fout met materialiteit voor verantwoording



ciële overzichten van de groep als geheel overschrijdt, tot een passend laag niveau terug te brengen, dient de materialiteit van het groepsonderdeel lager te zijn dan de materialiteit voor de financiële overzichten van de groep als geheel;

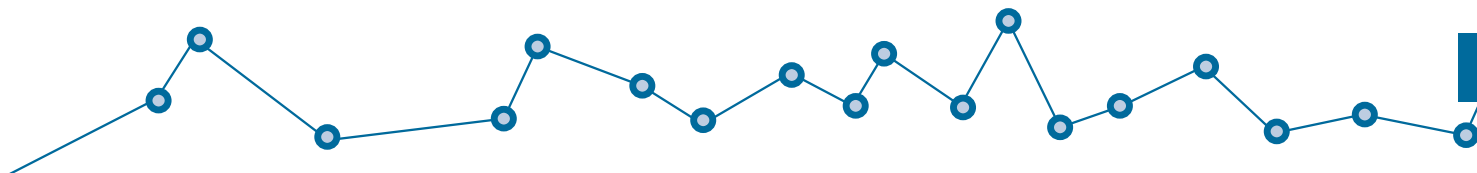
Wat hier wordt bedoeld is het zogenaamde aggregatierisico: het optellen van twee heterogene massa's geeft een foutverdeling die ongunstiger is dan de foutverdelingen van de onderdelen. Neem als voorbeeld een persoon die 100 loten van € 1 aan zijn vrienden verkoopt en 1 lot uitkiest om een prijs van € 100 uit te keren. Aan elke bezitter van een lot is de meest waarschijnlijke uitkering € 0 en de verwachte uitkering € 1, maar aan alle bezitters samen is de meest waarschijnlijke uitkering niet 100 keer € 0, maar € 100, en de verwachte uitkering 100 maal € 1. Meest waarschijnlijke fouten tellen dus niet, en verwachte fouten tellen dus wel op. Voor statistici is dit een gegeven: elk leerboek statistiek behandelt de centrale limietstelling (Rice, 1995) die zegt dat de som van onafhankelijke kans-variabelen tendeeft naar een normale verdeling. Die normale verdeling is ook bij niet-statistici bekend als een symmetrische kansverdeling met een meest waarschijnlijke fout gelijk aan zijn verwachtingswaarde.

Het optellen van onafhankelijke groepsonderdelen geeft dus een foutverdeling met een hogere maximale fout dan de uitkomsten van de kansverdelingen van de onderdelen apart. Om die hogere maximale fout toch aan de groepsmaterialiteit te laten voldoen, moet de materialiteit per onder-

deel lager zijn dan de groepsmaterialiteit. Hoeveel lager zegt ISA 600 niet; A43 zegt alleen dat de optelling van de materialiteiten per onderdeel hoger mag zijn dan de groepsmaterialiteit. Daarom hanteren wij in dit artikel niet de term alloceren, omdat in het Engels 'to allocate' betekent 'toekennen met als restrictie dat de toegekende bedragen samen optellen tot het materialiteitsbedrag'. Een materialiteit van € 100 kan dus naar 2 even grote en even risicovolle delen worden gealloceerd als 2 maal € 50, hetzelfde bedrag kan worden toegekend (of toegerekend als daar een wiskundig model achter zit) als 2 maal bijvoorbeeld € 63.

Overigens maakt het voor deze analyse van foutverdelingen niet uit wie de controle op groepsniveau en wie de controle op niveau van de onderdelen uitvoert. ISA 600 is niet alleen bedoeld voor het stellen van de materialiteit wanneer andere accountants(organisaties) de controle van onderdelen van een groep uitvoeren, maar ook wanneer het dezelfde accountant betreft. Het aggregatierisico uit 21c wordt daar namelijk niet door beïnvloed.

Toekennen van een lagere materialiteit per groepsonderdeel leidt tot meer werkzaamheden, maar heeft ook een groot voordeel: door een materialiteit per groepsonderdeel toe te kennen, bevestigt de accountant de onafhankelijkheid van het groepsonderdeel van de andere onderdelen. Geconstateerde afwijkingen in de controle van het ene onderdeel zijn dus niet relevant bij de evalua-



tie van andere groepsonderdelen. Toekennen van materialiteit aan groepsonderdelen schept daarmee ruimte voor het isoleren van fouten.

Als fouten daarentegen naar hun aard niet geïsoleerd kunnen worden is er geen sprake van een groepsonderdeel en hoeft dus ook de materialiteit niet toegekend te worden. Essentieel voor het toepassen en begrijpen van ISA 600 is de zorgvuldige afweging van het benoemen van groepsonderdelen. Alleen wanneer een deel van de activiteiten volkomen los staat van het geheel en dus ook volkomen afzonderlijk beoordeeld kan worden zonder gevolgen voor andere onderdelen, is er sprake van een groepsonderdeel waarvoor materialiteit moet worden toegekend maar ook waarvoor aangetroffen fouten mogen worden geïsoleerd. Andersom: wil de accountant de materialiteit niet toerekenen naar groepsonderdelen dan is de consequentie dat fouten gevonden bij de controle van het ene onderdeel moeten worden meegewogen bij de evaluatie van het andere onderdeel. Tegenover de efficiëntere controle staat een grotere exposure van fouten.

Methoden om de materialiteit toe te kennen zijn in de literatuur bekend. Van Schaik en Van Batenburg (2004) geven een overzicht van in de praktijk gehanteerde methoden zonder een daarvan te kiezen. Glover e.a. (2008) beschrijven een model voor de toerekening van de groepsmaterialiteit naar groepsonderdelen. Dit model wordt weliswaar in de praktijk als handvat omarmd, maar schiet wel op een belangrijk punt tekort omdat hun model

slechts de kans op overschrijding van de materialiteit meet en niet de mate van overschrijding. Toekenning volgens Glover e.a. laat dus geen compensatie toe van intensiever gecontroleerde onderdelen naar minder intensief gecontroleerde. Voor het model van Glover e.a. is inmiddels een beter model beschikbaar, het model van Stewart en Kinney. Dit model is daarom beter ten opzichte van het model van Glover, omdat daarin het foutbedrag in de onderdelen in plaats van de kans op een materiële fout per onderdeel wordt gemodelleerd. Het model zelf is nog niet gepubliceerd in artikelen maar op symposia en conferenties heeft Stewart (2010) al vele rekenvoorbeelden van hun toerekenmodel laten zien. Door een Gammaverdeling te veronderstellen voor de fout per onderdeel leiden zij een aggregatieformule af naar de groepsmaterialiteit.

7 Terug naar de jaarrekening: uitvoeringsmaterialiteit en toegestane afwijking: Clarity?

In een vorige paragraaf hebben wij uitvoerig stilgestaan bij het toekennen van de materialiteit voor een onderdeel van een groep, omdat wij denken dat de argumentatie voor een lagere materialiteit per onderdeel ook binnen een jaarrekeningcontrole een overweging zou moeten zijn. Het is daartoe interessant om de tekst van ISA 600, 21c hierboven eens te leggen naast ISA 320, A12.

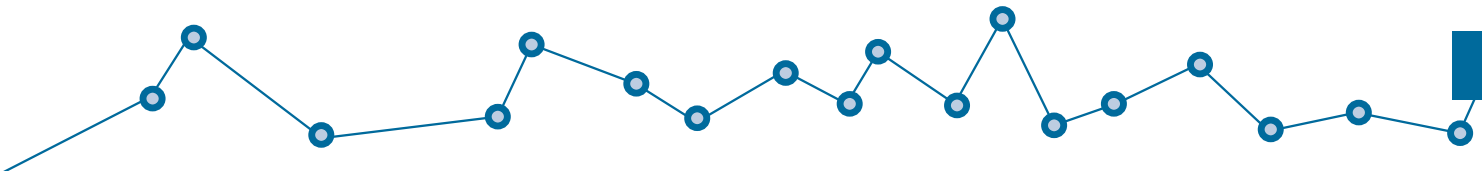
Het dusdanig plannen van de controle om alleen

individuele afwijkingen van materieel belang te ontdekken, houdt geen rekening met het feit dat het geheel van de individuele afwijkingen die niet van materieel belang zijn ertoe kunnen leiden dat de financiële overzichten een afwijking van materieel belang kunnen bevatten en laat geen ruimte over voor mogelijke niet-ontdekte afwijkingen. Uitvoeringsmaterialiteit (welke, zoals omschreven, bestaat uit een of meerdere bedragen) is vastgesteld om de kans dat het geheel van niet-gecorrigeerde en niet-ontdekte afwijkingen in de financiële overzichten de materialiteit voor de financiële overzichten als geheel overstijgt, tot een passend laag niveau terug te brengen. Op gelijksoortige wijze is uitvoeringsmaterialiteit met betrekking tot een materialiteitsniveau dat voor bijzondere transactiestromen, rekeningsaldi of in de financiële overzichten opgenomen toelichtingen bepaald is, vastgesteld teneinde de waarschijnlijkheid dat het geheel van niet-gecorrigeerde en niet-ontdekte afwijkingen in die bijzondere transactiestromen, rekeningsaldi of in de financiële overzichten opgenomen toelichtingen het materialiteitsniveau voor die specifieke transactiestromen, rekeningsaldi of in de financiële overzichten opgenomen toelichtingen overstijgt, tot een passend laag niveau terug te brengen. Het bepalen van de uitvoeringsmaterialiteit is geen eenvoudige mechanische berekening en impliceert het toepassen van professionele oordeelsvorming. Het wordt beïnvloed door het door de accountant verworven inzicht in de entiteit; die wordt geactualiseerd tijdens het uitvoeren van de risico-inschattingswerkzaamheden; alsmede door

de aard en omvang van de afwijkingen die zijn onderkend bij eerdere controles en via die weg door de verwachtingen van de accountant met betrekking tot afwijkingen in de huidige verslagperiode.

De argumentatie voor het toekennen van materialiteit per onderdeel in een groep kan dezelfde zijn als die voor het vaststellen van uitvoeringsmaterialiteit in een jaarrekening. Wij leiden hieruit af dat binnen een jaarrekening de materialiteit moet worden toegekend aan heterogene delen, dat wil zeggen jaarrekeningposten waarvoor de accountant bij de evaluatiefase fouten wil kunnen isoleren. De uitvoeringsmaterialiteit is daar het instrument voor. Wil de accountant de materialiteit niet toerekenen dan kunnen gevonden fouten niet worden geïsoleerd.

Het schema van de twee vragen en de manier om ze te beantwoorden wordt nu concreter:



	Gevolgen voor verantwoording	Voldoende werk gedaan?
Per post	Bepaal verzamelde fout (bekende fout, inschattingsverschillen en geprojecteerde fout) en vraag huishouding om correctie	Confronteer maximale fout met uitvoeringsmaterialiteit
Verantwoording als geheel	Accumuleer niet-gecorrigeerde verzamelde fout voor zover niet duidelijk onbeduidend en confronteer met de materialiteit	Confronteer overall maximale fout met materialiteit voor verantwoording

Zo bezien is ISA 450 niet echt duidelijker geworden. ISA 320 verplicht de accountant bij de opzet van de controle de uitvoeringsmaterialiteit toe te kennen, maar ISA 450 geeft niet aan dat op jaarrekeningpostniveau die uitvoeringsmaterialiteit de norm zou moeten zijn die niet mag worden benaderd door het totaal van geaccumuleerde fouten. ISA 450 zou helderder zijn geweest als er een aparte paragraaf was gewijd aan de evaluatie van de jaarrekeningpost en een aparte paragraaf voor de evaluatie van de jaarrekening als geheel.

En dan ten slotte nog het begrip toegestane fout uit ISA 530. Ervan uitgaande dat de uitvoeringsmaterialiteit de norm is voor de evaluatie van de jaarrekeningpost is het logisch om een iets scherpere norm te kiezen voor de opzet van de steekproefcontrole zodat ook bij een gering aantal fouten de post nog kan worden goedgekeurd door middel van een geschatte maximale fout die de uitvoeringsmaterialiteit nog niet overschrijdt. Die strengere norm is de toegestane afwijking. Het is een hulpmiddel om iets meer werk te doen

dan strikt noodzakelijk zodat de uitkomsten van dat werk een geringe fout mogen bevatten. De toegestane afwijking is dus een grootte die bepalend is voor de opzet van de steekproefcontrole maar geen rol speelt bij de evaluatie. Paragraaf A22 van ISA 530 bevat daarom ten onrechte het begrip toegestane afwijking in plaats van de uitvoeringsmaterialiteit: dat bij het vinden van fouten niet aan de toegestane afwijking is voldaan, is vanzelfsprekend. De steekproefomvang is namelijk zo opgezet dat alleen bij 0 fouten daaraan is voldaan. Het is de vraag of bij het gevonden aantal fouten de uitvoeringsmaterialiteit is overschreden.

8 Slotbeschouwing

Met Clarity zijn nieuwe begrippen geïntroduceerd die lijken op de oude begrippen, maar in feite zijn de begrippen voor en na Clarity niet vergelijkbaar en bovendien inconsistent. De introductie onzerzijds van het begrip geschatte maximale fout

zou kunnen helpen bij het inzichtelijk maken van deze begrippen en hun onderlinge samenhang. Wij merken hierbij op dat onze analyse vooral is ingestoken vanuit een kwantitatieve benadering omdat de statistische benadering voor meer eenduidigheid in interpretatie en toepassing van concepten leidt. De nieuwe ISA's zetten meer in vanuit de gedachte van professional judgment en terzake van het onderwerp 'materialiteit' een meer kwalitatieve benadering. Het risico van deze benadering is dat accountants juist onderling meer zullen gaan verschillen in het toepassen van materialiteitsconcepten. Dergelijke verschillen dragen niet bij aan het dichten van de verwachtingskloof. Een principle-based benadering zoals gehanteerd in de ISA's heeft zo ook haar nadelen. Wij zijn niet direct een voorstander van een rules-based benadering, maar maken wel het statement dat een principle-based benadering pas haar doelstelling realiseert als de principles helder (clarity!) zijn.

Tevens is sprake van onduidelijkheid over de vraag waarom ISA 600 toekennen van de materialiteit aan groepsonderdelen wel verplicht stelt, maar ISA 320 dit niet expliciet maar slechts impliciet doet door middel van het begrip uitvoeringsmaterialiteit. De introductie van de toegestane afwijking in ISA 530 is vervolgens nuttig bij de opzet van de (steekproef-)controle maar dit begrip zou bij de evaluatie geen rol mogen spelen. Aanvullende guidance vanuit het IFAC is daarom naar onze mening gewenst.

Wij constateren dat ISA Clarity in het algemeen, en ten aanzien van materialiteit in het bijzonder, nadruk legt op de gebruikers van de jaarrekening. Meer helderheid in begrippenkaders vanuit gebruikersoptiek is gewenst. Maar ook zien we vanuit de onderzoeksliteratuur dat het onderwerp materialiteit een aspect vormt dat de verwachtingskloof eerder vergroot dan verkleint. Meer verplichte communicatie door de accountant richting stakeholders over gehanteerde materialiteitsniveaus en een verplichte duiding in controledossiers wie de stakeholders zijn, zijn wat ons betreft belangrijke ingrediënten die helpen om de verwachtingskloof te overbruggen.

Evaluatie van tekortkomingen in de interne beheersing

Een specifiek vraagstuk rond materialiteit is het maken van een vertaalslag vanuit dit concept naar de evaluatie van tekortkomingen in interne beheersing te maken. De Amerikaanse Auditing Standard No. 5 (PCAOB, 2007) besteedt hieraan duidelijk meer aandacht, hetgeen begrijpelijk is gelet op het directe verband in Amerika tussen de jaarrekeningcontrole en in-control statements (Sarbanes-Oxley). De oude ISA's vermelden het evalueren van de impact van tekortkoming in de interne beheersing slechts sporadisch. De nieuwe ISA 265 geeft aan dit vraagstuk wel enige aandacht (verplichte rapportering van significante tekortkomingen aan het management), maar het begrip significant (wellicht vergelijkbaar met het Amerikaanse begrip material weakness) wordt niet geoperationaliseerd. Het is daarmee niet

duidelijk hoe het begrip 'significante tekortkoming' zich verhoudt tot het begrip 'materialiteit' of 'uitvoeringsmaterialiteit' als gehanteerd in ISA 320 en 450. Gegeven de maatschappelijke en toenemende actualiteit van in-control vraagstukken ten aanzien van financiële verslaggeving, was ISA Clarity onzes inziens de kans geweest om op dit punt meer helderheid te verschaffen. Een gemiste kans dus!

Literatuur

Arnold, V., S.G. Sutton, S.C. Hayne, C.A.P. Smith. (2000). Group Decision Making: The Impact of Opportunity-Cost Time Pressure and Group Support Systems. *Behavioral Research in Accounting*. Vol.12: p. 69-96.

Batenburg, P.C. van, F.D.J. van Schaik. (2004). Allocatie van materialiteit. *MAB*. September: p. 376-383.

Blokdijs, J.H., F. Drieënhuizen, D.A. Simunic, en M.T. Stein. (2003). Factors Affecting Auditors' Assessments of Planning Materiality. *Auditing: A Journal of Practice & Theory*. September: p. 297-307.

Braun, K.W. (2001). The disposition of audit-detected misstatements: An examination of risk and reward factors and aggregation effects. *Contemporary Accounting Research* 18 (1): p. 71-99.

Brown, C.E., I. Solomon. (1991). Configural information processing in auditing: the role of domain-specific knowledge. *The Accounting Review*. Vol. 66, Issue No. 1: p. 100-119.

Burgstahler, D., S.M. Glover, en J. Jiambalvo. (2000). Auditing: *A Journal of Practice & Theory*. Vol. 19, No. 1: p. 79-99.

Church, B.K., S.M. Davis en S.A. McCracken. (2008). The Auditor's Reporting Model: A Literature Synthesis and Overview. *Accounting Horizons*. Vol. 22, No. 1: p. 69-90. A Literature Overview and Research Synthesis. *Accounting Horizons*. Vol. 22, No. 1, Maart 2008, p. 69-90.

De Martinis, M.R., A.W. Burrowes. (1996). Materiality and Risk Judgments: A Review of Users' Expectations. *Managerial Finance*. Vol. 22, No. 9: p. 16.

Eilifsen, A., W.F. Messier Jr. (2000). The Incidence and Detection of Misstatements: A Review and Integration of Archival Research. *Journal of Accounting Literature*. Vol. 19: p. 1-43.

Estes, R., en D.D. Reames. (1988). Effects of Personal Characteristics on Materiality Decisions: A Multivariate Analysis. *Accounting and Business Research*, Vol. 18, No .72: p. 291-296.

Glover, S.M., D.F. Prawitt, J.T. Liljegen, W.F. Messier Jr. (2008). Component Materiality for Group Audits. *Journal of Accountancy*. December: p. 42-46.

Graham, L., W.F. Messier Jr. (2006). Audit Risk and Materiality in Conducting an Audit. *Journal of Accountancy*. May: p. 116-145.

Humphrey C. (1997). 'Debating Audit Expectations'. In: Sherer M and Turley S (Eds.), *Current Issues in Auditing*. 3rd Edition, Paul Chapman Publishing Ltd., London.

Kranacher, M-J. (2007). Determining Materiality: Relativity and Professional Judgment. *The CPA Journal*. August. Vol. 77, No. 8: p. 80.

Lee, T.H., A. Md. Ali, D. Bien. (2009). Towards an Understanding of the Audit Expectation Gap. *The Icfai University Journal of Audit Practice*, Vol. VI, No. 1: p. 7-35.

Mayper, A.G. (1982). Consensus of auditors' materiality judgments of internal accounting control weaknesses. *Journal of Accounting Research* 20 (2): p. 773-783.

Messier Jr., W.F., N. Martinov-Bennie, A. Eilifsen. (2005). A Review and Integration of Empirical Research on Materiality: Two Decades Later. *Auditing: A Journal of Practice & Theory*. Vol. 24, No. 2: p. 153-187.

Nelson, M.W., S.D. Smith, Z-V. Palmrose. (2005). The Effect of Quantitative Materiality Approach on Auditors' Adjustment Decisions. *The Accounting Review*. Vol. 80, No. 3: p. 897-920.

Nieuw Amerongen, C.M. van, P.C. van Batenburg. (2010). 'Materiality – Clarity'. Artikel 4050 in *Handboek Accountancy*, supplement 26. Kluwer, Deventer.

Pany, K., S. Wheeler. (1989). Materiality: An Inter-Industry Comparison of the Magnitudes and Stabilities of Various Quantitative Measures. *Accounting Horizons*. December: p. 71-78.

PCAOB. (2007). *Auditing Standard No. 5 – An Audit of Internal Control over Financial Reporting that is integrated with An Audit of Financial Statements*. Public Company Accounting Oversight Board. NY.

Pratt, M.J., K. van Peursum. (1996). Auditing Risk, Materiality & Judgment Standards: An International Comparison. *Managerial Finance*. Vol. 22, No. 9: p. 86-99.

Rice, J. (1995). *Mathematical Statistics and Data Analysis* (Second ed.), Duxbury Press.

Rosner, R.L., C.L. Comunale, Th.R. Sexton. (2006). Assessing Materiality. A New 'Fuzzy Logic' Approach. *The CPA Journal*. Juni. Vol. 76, No. 6: p. 26-28.

Securities and Exchange Commission. (1999). *SEC Staff Accounting Bulletin: No. 99 – Materiality*.

Stewart, T.R. (2010). *Audit Assurance Modeling That Makes a Difference*. Derde Symposium Statistical Auditing van het Limperg Instituut, Vrije Universiteit, 28 april 2010.

Tuttle, B., M. Collier, R.D. Plumlee. (2002). The Effect of Misstatements on Decisions of Financial Statement Users: An Experimental Investigation of Auditor Materiality Thresholds. *Auditing: A Journal of Practice & Theory*. Vol. 21, No. 1: p. 11-27.

Wallace, W.A, R.W. Kreutzfeldt. (1995). The Relation of Inherent and Control Risks to Audit Adjustments. *Journal of Accounting, Auditing & Finance*. Vol. 10, Issue No. 3: p. 459-483.

Wallin, J. (2001). Auditors' Selection of Tolerable Error and Risk Levels in the Context of Sample Size Decisions: A Cross-cultural Experiment. *LTA 2/01*: p. 260-273.

Wright, A., S. Wright. (1997). An examination of factors affecting the decision to waive audit adjustments. *Journal of Accounting, Auditing & Finance* 12 (1): p. 15-36.

Wright, W.F., N. Jindanuwat, J. Todd. (2004). Computational Models as a Knowledge Management Tool: A Process Model of the Critical Judgments Made during Audit Planning. *Journal of Information Systems*. Vol. 18, No. 1: p. 67-94.

Noot

- 1 Dit artikel is eerder verschenen in Handboek Accountancy, artikel nummer 4050.
- 2 Het is op basis van het artikel niet vast te stellen of hiermee mogelijk bedoeld wordt dat de kwantitatief (berekende) materialiteit wordt bijgesteld op basis van kwalitatieve factoren. In de praktijk zien we dat accountants kwalitatieve factoren op een verschillende wijze hanteren: sommige accountantsorganisaties staan slechts een neerwaartse bijstelling toe op basis van kwalitatieve factoren, andere accountantsorganisaties staan ruimte toe naar beneden en naar boven. Een andere interpretatie kan zijn dat de auteurs niet 'materialiteit' bedoelen, maar 'toegestane fout als fractie van de materialiteit'. In dit geval impliceert een toename van de kwaliteit van de interne beheersingsomgeving, dat met minder werk dezelfde conclusie getrokken kan worden.
- 3 Bij wetenschappelijke onderzoeken komt het geregeld voor dat het niet mogelijk is een bepaalde doelgroep representatief samen te stellen. Hiervoor wordt dan naar een aanvaardbaar alternatief (surrogaat) gezocht. De onderzoeksuitkomsten dienen echter wel in het licht geplaatst te worden van deze belangrijke veronderstelling.
- 4 Het begrip uitvoeringsmaterialiteit kan dus op meerdere manieren worden geïnterpreteerd. Wij prefereren in dit artikel de 'statistische' bril waarmee naar begrippenkaders wordt gekeken aangezien deze insteek naar onze mening beter objectieverbaar is. Deze benadering bevat ook het uitgangspunt dat het materialiteitsbegrip eenduidig gehanteerd kan worden voor homogene massa's (de achter jaarrekeningposten liggende dataset heeft dezelfde kenmerken) waardoor toerekening niet noodzakelijk is. Zijn de verschillende massa's niet homogeen (en daarvan is vaak sprake door bijvoorbeeld verschillen in schattingsposten en routinematige posten) dan is toerekening noodzakelijk. Een alternatieve bril om

naar materialiteitsbegrippen te kijken is dat vanwege het ontbreken van integrale controle een veiligheidsmarge wordt ingebouwd (verschil tussen materialiteit en uitvoeringsmaterialiteit). Het samenstel van controlewerkzaamheden dient vervolgens zodanig gepland te worden dat de werkelijk aanwezige afwijkingen geaggregeerd het niveau van de uitvoeringsmaterialiteit niet overschrijden.

Professional judgement, feilbaar denken¹

Ed Broeze

1 'Feilbaar denken'

Professional judgement speelt in de accountantscontrole bij onder meer de keuze van de materialiteit, het inschatten van inherent risico, de vraag 'steunen op de IB?', de keuze voor een (statistische?) steekproef en het bepalen van de omvang van gegevensgerichte controles. Kortom, professional judgement is cruciaal.

In een drieluik wil ik daar kanttekeningen bij plaatsen.

De lezer herkent in 'feilbaar denken' de titel van het boek van Daniel Kahneman (2012) 'Thinking, Fast and Slow', vertaald met 'Ons Feilbare Denken'. De vertalers hebben niet voor niets voor die niet-letterlijke vertaling gekozen.

Het boek laat zien hoe feilbaar ons denken is: hoe gevoelig voor een veelheid aan al dan niet bewuste, misleidende invloeden zoals: de volgorde waarin informatie tot ons komt, blijven vasthouden aan eerdere informatie, zelfs al is die waarneembaar nergens op gebaseerd, de wijze waarop informatie wordt aangeboden (het met de verkiesingscampagnes weer eens extra in de aandacht gebrachte *framing*), het onvermogen om kansen in

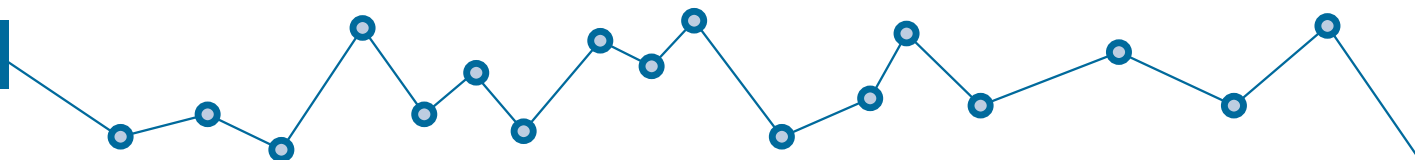
te schatten, de gevoeligheid voor de mate waarin informatie in het bewustzijn beschikbaar is (vliegen gevaarlijker dan autorijden?), et cetera.

Kahnemans boek roept de vraag op hoe betrouwbaar professional judgement kan zijn in het licht van en met de wetenschap van zoveel valkuilen. In deze aflevering van mijn drieluik over dat denken zal ik een – zeer onvolledige – schets geven van Kahnemans benadering van het denken. Die benadering heeft hij samen met Amos Tversky² ontwikkeld. Ik vind hun werk tot het boeiendste uit de literatuur behoren.

In Kahnemans (en Tversky's) (1974) theorie over probleemoplossend denken gaat het over een snel denken (systeem 1 genoemd) en een langzaam denken (systeem 2).

Systeem 1 komt snel tot conclusies, dit is het systeem van de intuïtie. En daarmee een zeer functioneel deel van ons denken, bij onder meer: een auto besturen, snel de juiste schaakzet zien (als je grootmeester bent).

Systeem 2, het trage denken, is het systeem van de rationaliteit. Systeem 1 reageert op een enquête



waarin 60% voorkeur uitspreekt voor een premier niet anders wanneer het aantal respondenten 100 is dan wanneer het 1000 is. Systeem 1 houdt niet van onzekerheid en zo werkt de wet van de kleine aantallen: mensen doen (vaak) alsof een kleine steekproef even goed is als een grote. En systeem 2 verbindt er vervolgens conclusies aan. Systeem 1 wordt vergaand buiten het bewustzijn om beïnvloed, maar heeft vervolgens invloed op systeem 2.

Tversky en Kahneman (1981) laten dat (ook) zien in een experiment, waarin een beschrijving is opgenomen van Linda. Een vrouw van 31, slim, openhartig, sociaal betrokken, fel op discriminatie, actief anti-kernenergie. Studenten moesten inschatten wat waarschijnlijker was: dat zij bankbediende was of bankbediende en actief in de vrouwenbeweging.

Rond de 85% koos voor de tweede optie, ook ondanks allerlei pogingen deze evident foute inschatting te voorkomen. Immers: er zijn meer bankbedienden dan feministische bankbedienden. De waardering van Linda als representatief voor feminisme (een systeem 1 inschatting) won het van de logica!

Als ons denken zo feilbaar is, hoe is het dan met professional judgement? Eén geruststelling: professionals blijken valkuilen te kunnen ontlopen (al kiest ook 85% van de statistiekstudenten bij Linda de verkeerde optie). Maar er valt veel meer over te zeggen: zie daarvoor de delen 2 en 3 van dit drieluik.

In paragraaf 2, zal ik ingaan op het vermogen van mensen om kansen in te schatten, in eenvoudige (en ook berekenbare) kanssituaties.

In paragraaf 3 zal ik ingaan op hoe dat ligt als er ook nog een niet zomaar in een kansmodel om te zetten situatie vraagt om de inschatting van een kans, zoals in risicoanalyse³. Zoals u weet tellen die inschattingen flink mee in de uiteindelijke kwantificering van het accountantscontrole risico.

2 Intuïtief schatten berekenbare kansen

Na het feilbare denken in paragraaf 1 gaat het nu over het intuïtief inschatten van berekenbare kansen.

Ik heb heel wat introducties in de statistiek verzorgd. Ter introductie van de introductie was een terugkerende vraag aan mijn gehoor: hoe groot schat je de kans dat in een gezelschap van 23 willekeurige personen er twee of meer op dezelfde dag jarig zijn? Het gros van mijn studenten schatte die kans op zo'n 5 procent, hoogstens 10 procent. Immers, de kans dat 2 personen op dezelfde dag jarig zijn is zó klein, dat die voor 23 ook nog steeds niet groot kan zijn.

Dat, althans, is de intuïtieve 'redenering', het systeem 1 denken van paragraaf 1 van dit essay. Verankering en onvoldoende aanpassing noemden Tversky en Kahneman dat. En het is ook *framing*, want als de vraag was geformuleerd als 'Schat de

kans in dat niet twee van de 23 verjaardagen op dezelfde datum vallen', dan was die inschatting zeer waarschijnlijk meer in de buurt gekomen van de 50 procent die hij ook in werkelijkheid is.

Zoals u ziet zijn de gebeurtenissen in de eerste formulering ('twee of meer op dezelfde dag') en die in de tweede formulering ('niet twee van de 23') elkaars complement en dus geldt volgens de kansrekening dat de kans die we met de eerste vraag zochten gelijk was aan (1 - de tweede kans), dus ook 50 procent. De intuïtie die op circa 5 procent kwam werkte dus niet.

In 1986 kwam ik bij de Algemene Rekenkamer te werken, als statistisch adviseur. In het begin werd mij weleens (vriendelijk) toegesproken: "Toen je er nog niet was vonden we een steekproef van 100 al fantastisch en als we daarin geen enkele fout vonden waren we volkomen zeker van een goedkeuring en bij één (of twee) eigenlijk ook nog wel. En nu zeg jij dat we er minstens 300 moeten trekken met 0 fout en dat ten onrechte goedkeuren dan nog altijd 5 procent kans heeft. We dachten dat je ons leven makkelijker kwam maken."

Mijn antwoord was natuurlijk: "Ja, als je met een betrouwbaarheid van 95 procent wilt kunnen goedkeuren en je materialiteit is 1 procent, DAN... En zoals jullie het deden had je maar een betrouwbaarheid van 63% en dan moest je nog aselekt hebben getrokken. Anders kende je je betrouwbaarheid niet eens."

Ziehier twee van de vele voorbeelden die ik ben tegengekomen over hoe mensen ertoe neigen kan-

sen verkeerd in te schatten (ik heb die 63 procent ook op twee manieren opnieuw berekend voor deze column: is hij echt zo laag?).

In de literatuur is er ook stelselmatig onderzoek naar gedaan: een incomplete greep, waaruit hetzelfde beeld naar voren komt (ja, er kan een bias in mijn keuze zitten, maar dan nog).

Bijvoorbeeld Messier et al. (2001) laten zien dat auditors op basis van de AICPA 'Audit Sampling Guide' stelselmatig kiezen voor een steekproefomvang die kleiner is dan volgens de statistiek vereist. Zij vrezen *under-auditing*⁴.

Burgstahler et al. (2000) laten zien dat auditors de neiging hebben de onzekerheid te onderschatten die hoort bij de projectie van een steekproefresultaat op de rekening, en daarbij tot te lage correcties te komen.

Ook de mens als 'aselector' doet het niet goed. Hall et al. (2001) laten zien dat bij 'lukraak steken' verdubbeling van de steekproefomvang nauwelijks invloed heeft op de vertekening die zo'n steekproef geeft: 'aantrekkelijke' objecten zijn in hun experiment met 30 procent oververtegenwoordigd – aantrekkelijk, door kleur, formaat en dergelijke.

In mijn proefschrift (Broeze, 2006), hoofdstuk 3, geef ik een uitgebreider overzicht. Sindsdien is er bij mijn weten slechts weinig op dit terrein verschenen.

We zijn dus slecht in het inschatten van kansen, onderschatten de onzekerheid die hoort bij uitspraken die op steekproeven zijn gebaseerd, onderschatten de omvang van benodigde correcties en kunnen niet 'aselecteren'. En toch...

En toch wordt alleen al in ISA 530 ruim baan gegeven aan professional judgement: bij het bepalen van de steekproefomvang (in A11) wordt van geen statistische methode gerept als het over het projecteren van fouten gaat (in A22), terwijl er wel vrolijk van sampling risk wordt gesproken en er in A22 nog andere bedenkelijke suggesties (*understatement*) worden gedaan over hoe met misstatements om te gaan.

Maar die horen bij fouten isoleren (zie columns 9 en 11) en bij het uitbreiden van de steekproef als de bestaande niet tot goedkeuring mag leiden (net als isoleren op z'n minst zeer tricky, maar onuitroeibaar; zie ook column 12).

Daarnaast geeft ISA530 in appendices 2 en 3 aanwijzingen voor het vaststellen van een steekproefomvang, analoog aan de AICPA-sampling guide waarvan Messier et al. hebben laten zien dat die niet werkt. Je vraagt je ook af waarom in die appendices niet gewoon een formule staat.

Voor alle duidelijkheid: ik wil en kan niemand dwingen om statistische methoden te gebruiken. Maar als je daarvan afziet, zie dan ook af van de suggestie van beheersing van de betrouwbaarheid van je controle.

3 Risicoanalyse

Een accountant maakt een procesmodel dat iets over kansen op mogelijke uitkomsten zegt, als hij een beoordeling ontwikkelt van de kwaliteit van de interne beheersingsmaatregelen. Wat hij doet is het vaststellen van de kwaliteit van een proces en dat vertalen in kansen voor het product.

Die vertaling kan op minstens twee fronten last hebben van het feilbare denken:

1. Wordt de kwaliteit van het proces wel op voldoende niveau ingeschat; en
2. Wordt die inschatting wel correct vertaald in een kans dat de jaarrekening geen materiële fout(en) bevat.

Bij het inschatten van het eerste, de *kwaliteit van het proces* (de AO, de IB) heeft de accountant last van alle vertekening veroorzakende valkuilen in het (professionele) oordelen. Die zijn:

- *Framing* (gaat men uit van 'een sterk groeiend' of juist een 'frauderend' bedrijf?).
- *Verankering* en (onvoldoende) *aanpassing* (*anchoring and adjustment*); de werkomgeving oogt zeer betrouwbaar, is er oog genoeg voor zwakke punten?
- Van bevindingen te gemakkelijk *representatief* achten (*representativeness*); één positieve lijncontrole slechts aanvullen met weinig systeemtests.
- Van beschikbaarheid (*availability*); onder meer oorzaak van het *recency-effect*; de laatst verwerkte informatie krijgt te veel gewicht.

Bij het tweede, *de vertaling van de inschatting van de kwaliteit* van de IB-processen naar de kans op fouten, kan men kijken of de accountant zijn eigen vertaling serieus neemt. Oftewel, dekt hij met zijn controleaanpak inderdaad het ingeschatte risico af?

Onderzoek laat zien dat dit vaak niet het geval is: het overgebleven risico is groter dan gewenst en ook wordt de controleaanpak niet aangepast aan veranderende risico's.

Daarnaast is de geplande omvang van tests soms wel weer positief gecorreleerd met risico-inschattingen op sommige controledoelstellingen. Er is in een enkel onderzoek ook gekeken of de risico-inschatting positief correleert met de omvang van de gevonden fouten. Daaruit komt een licht positief verband.

En we houden ook nog de tot weinig optimisme stemmende bevindingen uit deel 2 van dit essay dat we niet goed zijn in het inschatten van kansen.

Een veel uitgebreider overzicht van onderzoek naar de kwaliteit van risico-inschattingen staat in het derde hoofdstuk van mijn proefschrift (Broeze, 2006). De conclusie daarvan is onbestemd: er zijn heel veel valkuilen, maar accountants hoeven daar lang niet altijd in te trappen (door goede documentatie, ervaring met bedrijfstak, peer review et cetera).

Door dat onbestemde blijft het de vraag hoe goed uiteindelijk de risico-inschattingen zijn die uit de gebruikelijke risicoanalyse komen.

Die vraag heb ik in mijn proefschrift proberen te beantwoorden door uitkomsten van risicoanalyse te relateren aan de uiteindelijk gevonden fout. Hieruit kwam een positief verband, maar wel één dat sterk varieerde over de elf organisaties waar ik naar gekeken heb. Per organisatie betekende dat verband dus nog niet veel.

Ik wilde er niet omheen dat het bij de risico-inschatting om een kans gaat, namelijk de kans dat de werkelijke fout de materialiteit overschrijdt. Die kans kon ik berekenen bij gegeven steekproefomvang en de gevonden fout. En als ik die (*a posteriori*) kans als criterium gebruikte, was het verband tussen risico-inschatting en bevindingen in de gegevensgerichte detailcontrole helemaal weg; hetgeen betekent dat de risico-inschatting zowel te hoog, als te laag, als goed kan zijn, maar je weet dat niet.

Ik heb nog diverse andere grootheden als criterium voor de kwaliteit van de risico-inschatting gebruikt. En steeds bleek dit tot andere uitkomsten te leiden.

Per saldo is er bij het gebruik van risicoanalyse dus een groot risico dat het uiteindelijke controle-risico groter is dan het model laat verwachten.

Samengevat:

- Onderzoek is unaniem: in veel gevallen kunnen we geen kansen inschatten.
- Onderzoek naar risicoanalyse laat heel veel open over de kwaliteit.
- Mijn onderzoek laat op het meest in aanmer-

king komende criterium géén verband zien tussen de risico-inschatting en de bevindingen in de gegevensgerichte detailcontrole.

Men mag hoop houden, omdat met andere criteria soms wel positieve verbanden worden gevonden.

Het is opmerkelijk dat in de researchliteratuur nooit een probleem is gemaakt van het criterium voor de kwaliteit van de risico-inschatting.

Literatuur

Broeze, G.B. (2006); *Validation of Risk Assessment in Auditing*. Limperg Instituut, Amsterdam.

Burgstahler, D., S.M. Glover & J. Jiambalvo (2000); Error Projection and Uncertainty in the Evaluation of Aggregate Error; *AUDITING: A Journal of Practice and Theory* Vol. 19 No. 1, p.p. 79-99.

Hall, Thomas W., Terri L. Herron, Bethane Jo Pierce, Terry J. Witt (2001); The Effectiveness of Increasing Sample Size to Mitigate the Influence of Population Characteristics in Haphazard Sampling; *AUDITING: A Journal of Practice & Theory*; March 2001, Vol. 20, No. 1, pp. 169-185.

ISA 530 (2009); *International Standard on Auditing 530 Audit Sampling*.

Kahneman, Daniel (2012); *Thinking, Fast and Slow*; Penguin.

Kahneman, D. (2003); A Perspective on Judgment and Choice, Mapping Bounded Rationality. *American Psychologist* Vol 58.

Messier, William F., Steven J. Kachelmeier & Kevan L. Jensen (2001); *An Experimental Assessment of Recent Professional Developments in Nonstatistical Sampling Guidance*. Auditing, a Journal of Practice and Theory Vol. 20, No. 1, pp. 81-96.

Tversky, A. & D. Kahneman (1981); *The Framing of Decisions and the Psychology of Choice*; *Science*, Vol 211.

Tversky, A. & D. Kahneman (1974); Judgment under Uncertainty: Heuristics and Biases, *Science*, Vol 185, pp. 1124-1131.

Noot

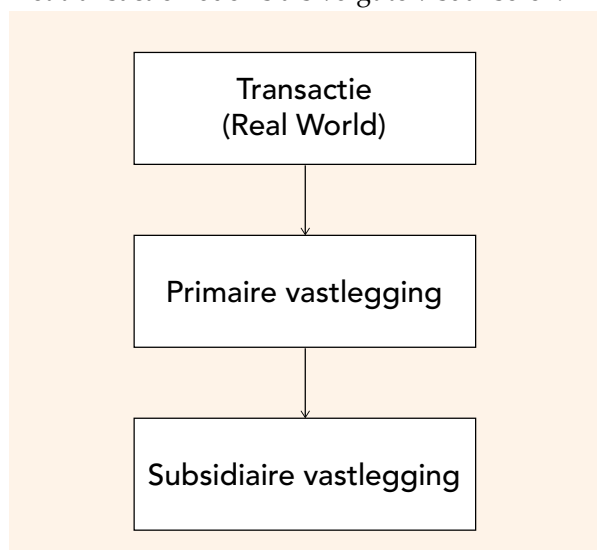
- 1 Dit essay is een samenvoeging van de columns Statistical Audit 16, 17 en 18.
- 2 Het artikel over 'Heuristics and Biases' (Tversky en Kahneman 1974) is in vertaling in het boek 'Ons Feilbare Denken' opgenomen; zie ook onder andere: Tversky en Kahneman (1981) en Kahneman (2003).
- 3 Bedoeld is het Audit Risk Model (ARM).
- 4 Met under-auditing wordt bedoeld dat er te weinig wordt gecontroleerd, de controle is dan niet effectief (genoeg).



Interne beheersing

Het transactiemodel is opgezet om de relatie tussen de 'echte wereld' en de financiële controle te beschrijven¹. De nadruk ligt daarbij op de rol van de onvervangbare interne controle. Het model maakt verder het 'indirecte controle'-paradigma van financial auditing duidelijk. Het transactiemodel maakt geen onderscheid tussen brongegevens op papier en andere, b.v. elektronische, brongegevens.

Het transactiemodel is als volgt te visualiseren.



Het model gaat ervan uit dat bedrijven transacties genereren in de 'echte wereld' (of: 'the real world').

De transacties die in de echte wereld plaatsvinden (tussen organisatiedelen onderling en vooral tussen de organisatie en de buitenwereld ervan) plegen door de organisatie te worden vastgelegd. De eerste vastlegging van een transactie noemen we de primaire vastlegging. De primaire vastlegging van een transactie dient alle data-elementen te bevatten die voor de bedrijfsvoering en de verantwoording daarover nodig kunnen zijn. De primaire vastlegging van de transactiegegevens leidt vervolgens tot een meer formele vastlegging: de subsidiaire vastlegging.

De verzameling subsidiaire vastleggingen is te omschrijven als het opnemen van (een deel van) de data van de primaire vastleggingen en het toevoegen van rubriceringsgegevens in informatiesystemen (niet per se geautomatiseerd).

Men kan ervan uitgaan dat of een van de eigenschappen of een combinatie van een of meer van de eigenschappen van een transactie de financiële consequenties ervan beschrijven. Die eigenschap (of combinatie van eigenschappen) plegen we de eigenschap: 'bedrag'² te noemen. Deze eigen-

schap is te zijner tijd nodig om de controle met behulp van steekproeven op geldeenheden te beschrijven.

Van real world naar primaire vastlegging

De creatie van de primaire vastlegging van een transactie, ofwel de verzameling brongegevens van die transactie, is idealiter aan maatregelen van interne beheersing onderhevig. Dit kan zijn directe controle en beheersing, indirecte controle en latere terugkoppeling of een combinatie van beide varianten. Ingeval van indirecte controle is er sprake van meerdere primaire vastleggingen vanuit verschillende bron over hetzelfde fenomeen. De veronderstelling is dan dat als de data overeenstemmen die uit verschillende bron afkomstig zijn, dan al die data juist zijn. Onder 'uit verschillende bron' wordt verstaan dat de data onafhankelijk van elkaar tot stand zijn gekomen, of nog beter, de bronnen ten opzichte van de data een belangentegenstelling hebben zodat de bronnen 'wel gek zouden zijn als ze deze gegevens onvolledig of onjuist zouden weergeven'. Dit fenomeen heet 'controle-technische functiescheiding'.

In het geval er IT wordt toegepast kan de controle-technische functiescheiding worden gerealiseerd als iedere bron of functie slechts die gegevens mag invoeren in een systeem die met zijn functie verband houden; bovendien moet de oorsprong van de gegevens, de bron of functie, navolgbaar blijven. In dat geval zijn er beheersmaatregelen nodig in de IT-omgeving. Deze moe-

ten de exclusiviteit van de behandeling van de gegevens garanderen. De randvoorwaarden voor die exclusiviteit bereikt men door maatregelen die als general IT-controls worden aangeduid.

Voor wat betreft de primaire vastleggingen wordt doorgaans, zij het impliciet, verondersteld dat deze gegevens alle voorwaarden voor indirecte controle in zich herbergen. Dat betekent dat integriteit zowel als herkomst als vaststaand worden verondersteld. Deze vooronderstelling moet expliciet worden gemaakt zodra er sprake is van elektronische primaire vastleggingen. Een andere vooronderstelling is dat van de primaire vastleggingen de Soll-positie kan worden afgeleid. Doorgaans is het mogelijk de Soll-positie af te leiden van de primaire vastleggingen als deze worden gevoegd bij (allerlei) documentatie omtrent de transacties.

De transacties die in de 'real world' plaatsvinden, moeten sporen nalaten in de wereld van de primaire vastleggingen, doen zij dat niet dan ontstaan er omissies. Die omissies kunnen gehele transacties betreffen (onvolledigheid van de primaire vastlegging van transacties) of gegevens(soorten) omtrent transacties (onvolledigheid vastlegging van attributen van transacties; ofwel onvolledigheid metadata). Verder kan het voorkomen dat gegevens over transacties onjuist zijn vastgelegd (onjuiste vastlegging van waarden van attributen van transacties).

De onderlinge relaties tussen (delen van) transacties moeten zien te voorkomen dat er sprake is

van onvolkomen registratie van transacties. De beheersmaatregelen hieromtrent behoren tot de categorie onvervangbare interne controle. De Soll-positie van de primaire vastleggingen bevindt zich in de 'real world'. De primaire vastleggingen zijn de Ist-posities.

Van primaire vastleggingen naar subsidiaire vastleggingen

De vertaling van primaire vastleggingen naar de subsidiaire vastleggingen impliceert dat de subsidiaire vastleggingen een deelverzameling aan gegevenselementen bevatten van de primaire vastleggingen en dat daar nog gegevenselementen die nodig zijn voor intern gebruik of voor het extern afleggen van verantwoording aan worden toegevoegd. De subsidiaire vastleggingen maken doorgaans deel uit van de verzameling formele informatiesystemen. De formele informatiesystemen moeten kunnen voldoen aan de informatiebehoefte van de gebruikers van die systemen. Willen ze dat kunnen dan moeten de primaire vastleggingen daarvoor toereikend zijn, dan wel: er moet achteraf op eenvoudige wijze voor aanvulling kunnen worden gezorgd³. Verder is het van belang dat alle subsidiaire vastleggingen van een transactie op de een of andere manier een verwijzing naar de brongegevens herbergen (management trail of audit trail).

De subsidiaire vastleggingen bevatten dus behalve een subset van de data-elementen van de primaire vastleggingen een of meer rubriceringsgegevens. Die rubriceringsgegevens komen tot

stand door interpretatie van de primaire vastlegging.

In de categorie subsidiaire vastleggingen worden doorgaans de gegevens van verschillende transacties samengevoegd, geaggregeerd. Veel vormen financiële verantwoording zijn uit geaggregeerde grootheden samengesteld. Dat betekent dat één te onderzoeken Ist-positie uit de financiële verantwoording meerdere individuele vastleggingen van meerdere transacties kan betreffen.

Transacties en de controle

De financiële verantwoording is een weergave van de financiële aspecten van alle transacties over een periode. De controle ervan behelst dus:

- de volledigheid van de vastlegging van de transacties;
- de volledigheid van de vastlegging van de gegevens(soorten) over de transacties;
- de (inhoudelijke) juistheid van de vastlegging van de waarden van de gegevens(soorten) over de transacties.

Omdat er veel soorten financiële verantwoording zijn die schattingen of prognoses bevatten omvat de controle van een willekeurige verantwoording dus ook de beoordeling van de schattingen of prognoses die erin voorkomen. Wij laten vooralsnog de beoordeling van de schattingen en prognoses buiten beschouwing.

De relatie tussen deze controledoelstellingen en de primaire en subsidiaire vastleggingen geven wij hierna weer.

Volledigheid transacties

Alle transacties vinden plaats in de 'real world'. Transacties die niet hebben geleid tot een primaire vastlegging hadden alleen kunnen worden vastgesteld door directe waarneming. Bij de behandeling van de primaire vastleggingen kwam al naar voren dat er controlemaatregelen bestaan om een redelijke zekerheid te bieden omtrent de volledige registratie van de transacties, te weten controle-technische functiescheiding. Deze verzameling maatregelen zorgt voor een uitgebreide vastlegging van gegevenssoorten waarvan de waarden (per transactie) gelijk moeten zijn. De relatie tussen de waarden is gebaseerd op de waardenkringloop in de onderneming. Deze controle is een verbandscontrole op transactie(cycle)niveau. De juistheid van degene die de vastlegging heeft verricht en de juistheid van de waarde van de overige gegevenselementen hebben het karakter van maatregelen van onvervangbare interne controle⁴.

Het is daarnaast mogelijk de onderlinge relatie tussen verschillende transactiesoorten te benutten om verband te leggen met de waardenkringloop. Ook deze vorm van controle wordt verbandscontrole genoemd. Bij deze controle komen alleen transacties aan het licht die wel in de real world hebben plaatsgevonden en niet leiden tot een primaire vastlegging, als de gerelateerde transacties wél hebben geleid tot primaire vastleggingen. Deze controles kunnen niet (zonder meer) met behulp van de theorie van steekproeven op geldeenheden worden beschreven. Philip Elsas (1996) heeft een uitgebreide beschrijving gegeven van

het volgen van de waardenkringloop op transactieniveau op basis van een combinatie van primaire en subsidiaire vastleggingen.

Volledigheid gegevenssoorten

Als achteraf blijkt dat er gegevenssoorten omtrent de transactie hadden moeten worden vastgelegd, wat niet gebeurd is, heeft de accountant het nakijken. In beginsel moet deze doelstelling tijdens de voorwaardelijke controles worden onderzocht. De controle is eigenlijk een informatie-analyse van de Soll-posities vanuit het gezichtspunt van de controle-instantie bezien. Iedere controle-instantie kan verschillende doelstellingen hebben. Die doelstellingen moeten al bij het inrichten van de administratie bekend zijn. De informatieanalyse moet al op dat moment mogelijk zijn. Bij een dergelijke informatieanalyse moet rekening gehouden worden met van elkaar onafhankelijke data-elementen en zogenoemde procesgegevens. Procesgegevens zijn data die uit een of andere combinatie van de onafhankelijke gegevenselementen af te leiden zijn. Het feitelijk duiden van de gegevenssoorten op het moment van de primaire vastlegging is doorgaans een activiteit van onvervangbare interne controle. Deze controle kan niet met behulp van de theorie van steekproeven op geldeenheden worden beschreven.

Juistheid waarden gegevenselementen per transactie

Bij deze controle is er de beschikking over een set transactiegegevens, zijnde de subsidiaire vast-

leggingen. Deze gegevens kunnen steeds met de primaire vastleggingen in verband worden gebracht. De Soll-positie van de waarden van de verschillende gegevenssoorten is dan van de primaire vastleggingen (de brongegevens) af te leiden. Deze afleiding van de Soll-positie is niet noodzakelijkerwijs eenvoudig. Deze controle is goed te beschrijven met behulp van de theorie van steekproeven op geldeenheden.

Samenvatting

Een financieel overzicht zoals een jaarrekening, is een weergave van vastleggingen van transacties. De primaire vastleggingen of brongegevens kunnen een papieren of elektronische verschijningsvorm hebben.

Als een accountant de kwaliteit van de primaire vastleggingen wil beoordelen zal hij dat moeten doen door middel van directe controle⁵.

In het geval van indirecte controle gaan we ervan uit dat de primaire vastleggingen in combinatie met aanvullende documentatie de Soll-positie(s) kunnen genereren.

De relatie tussen primaire en subsidiaire vastleggingen moet navolgbaar zijn.

De primaire vastlegging moet indirecte controle mogelijk maken. Als de primaire vastleggingen samenhangen met IT is de deskundigheid om de beheersing rond de onvervangbare interne controle te onderzoeken gelegen in het vakgebied IT-audit.

De primaire vastleggingen moeten de gegevenssoorten bevatten die kunnen behoren tot de verzameling gegevens die de informatiebehoefte van alle gebruikers kunnen dekken.

Aangevuld met interne interpretaties worden de gegevens opgenomen in de formele informatiesystemen en vormen zo databases met subsidiaire vastleggingen.

Het transactiemodel laat zien dat als een Ist-positie, die bevindt zich in de sfeer van de subsidiaire vastleggingen, moet worden gecontroleerd, de relatie tot de primaire vastlegging moet vaststaan. De primaire vastlegging geeft pas informatie over de Soll-positie als de onvervangbare interne controle heeft gewerkt.

De controle op de juistheid van de waarden van de vastleggingen van de transacties is met behulp van de theorie van steekproeven op geldeenheden te beschrijven.

Literatuur

Blokdijk, J.H. (1975); Een kernvraagstuk van de leer der accountantscontrole deel I en deel II; *Maandblad voor Bedrijfseconomie en Accountancy*, pp. 147-159 (deel I); pp. 190-207 (deel II).

Blokdijk, J.H. (1995); De betekenis van de onvervangbare interne controle voor de accountantscontrole; *Maandblad voor Bedrijfseconomie en Accountancy*, pp. 588-598.

Broeze, G.B., H.H.W. Kloosterman, (2003); *Overzichtsartikel onderzoek Statistical Audit*; Stuurgroep Statistical Audit.

Elsas, P.I. (1996); *Computational Auditing*, dissertatie, VU, Amsterdam.

Kloosterman, H.H.W, J.A. Lok, P.C. Waas (2005); Elektronisch factureren, Belastingcontrole en IT-audit, Artikel 9610, *Handboek EDP-audit*; NIVRA, Kluwer.

Kloosterman, H.H.W., R. Snoeker (2013); Informatietechnologie en Interne Beheersing, *De IT-auditor*; NOREA, Reed.

Popper, K.R. (1959); *The Discovery of Scientific Discovery*, Hutchinson Education.

Noot

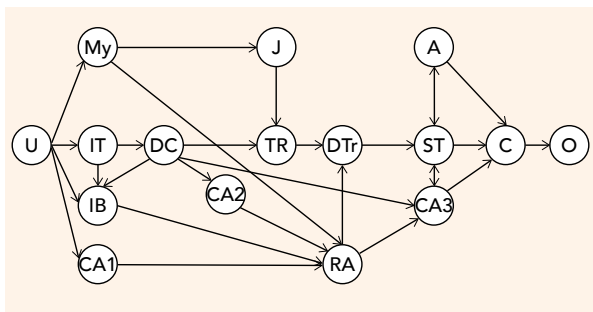
- 1 Het model is in de Stuurgroep Statistical Auditing naar voren gebracht door Kloosterman in een aantal brieven aan de Stuurgroep. Het is terug te vinden in Broeze en Kloosterman (2003), Kloosterman et al. (2005) en Kloosterman en Snoeker (2013).
- 2 Als bijvoorbeeld een verkoop van een artikel plaatsvindt kan het al voldoende zijn om bij de gegevens van de transactie de datum, het artikelnummer en het aantal ervan in te voeren. De prijs kan via het artikelnummer erbij gezocht worden. Het bedrag van de transactie kan daaruit worden afgeleid (afhankelijk gegeven). De verschuldigde omzetbelasting kan vervolgens uit een aantal andere hoedanigheden worden afgeleid.
- 3 Wat eventueel toegevoegd zou kunnen worden is een rubricering zoals een grootboekrekeningnummer.
- 4 Onder meer in Blokdijk (1975) en Blokdijk (1995) is de opvatting van Blokdijk over onvervangbaarheid te lezen. Markant zijn de twee belangrijkste kenmerken die hij geeft voor onvervangbaarheid. Hij noemt met nadruk de grenzen aan de deskundigheid van de

- accountant en die aan de beschikbaarheid van de accountant, dat laatste is het niet in continuïteit kunnen waarnemen van de transacties en gebeurtenissen die voor uiteindelijk de verslaglegging relevant zijn.
- 5 Als er directe controle plaatsvindt dan is die waarneming pas zinvol als hij de meting ervan niet beïnvloedt. Met andere woorden: er moet geen contaminatie plaatsvinden. Men kan bij zo'n waarneming zowel het al dan niet vastleggen, als de waarden van de vastlegging beoordelen. Als de waarnemingen aselekt in de tijd zijn gekozen en zij leggen geen feilen bloot, dan mag men stellen dat er voldoende directe waarnemingen hebben plaatsgevonden om te proberen te falsifiëren dat er sprake is van onvolledigheid. Zo'n bevinding kan gevoegd worden bij de overige indicaties omtrent het al dan niet volledig zijn van de verantwoording van de transacties als primaire vastlegging. Ter vergelijking een citaat uit Popper (1959): "...For I do not demand that every scientific statement must have been tested before it is accepted. I only demand that every such statement must be capable of being tested..." (p.23).

Interne beheersing en statistical audit; een schets

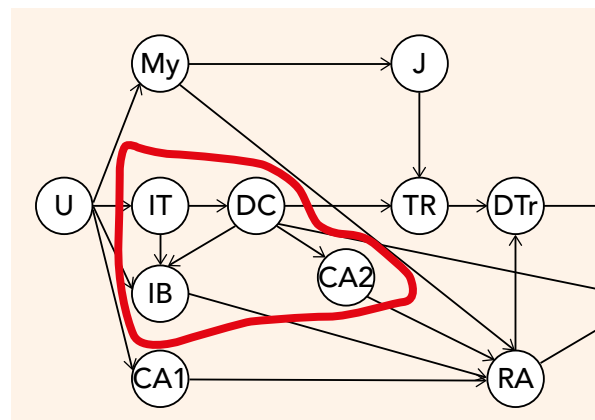
Hein Kloosterman

Het vakgebied statistical audit buigt zich over het toepassen van wiskundige methoden en technieken ter ondersteuning van de besluitvorming in de controle. Dat is in de inleiding bij deze essays al aangeduid door een netwerkmodel van de controle te tekenen waarin de kritieke besispunten in de controle zijn neergezet.



Figuur 1

Het beoordelen van de interne beheersing is daarin uitgelicht met de volgende figuur.



Figuur 2

Het mag duidelijk zijn dat deze figuren niet in detail weergeven welke wiskundige methoden en technieken allemaal mogelijk zijn in de interne beheersing. Dit artikel beperkt zich tot het opsommen van de bekendste technieken. Het gaat hier vooral om het verwijzen naar andere literatuur waarin die technieken meer worden toegelicht.

De wereld van de IT, als essentieel onderdeel van de interne beheersing, kent een groot aantal modellen en technieken die bedoeld zijn om bedrijfsprocessen te optimaliseren. Grofweg noemen wij die modellen en technieken operational research¹.

Tot operational research horen onder meer technieken als lineaire programmering, dynamische programmering, wachttijdentheorie, netwerktheorie en meerdere statistische methoden. Dergelijke technieken zijn vooral terug te zien in de sfeer van logistiek (bijvoorbeeld voorraadbeheersing², berekenen machine- en personeelscapaciteit, inroosteren van personeel en machines, plannen van complexe onderling afhankelijke activiteiten).

In de vakken bedrijfseconomie voorafgaande aan de accountancystudie worden dergelijke technieken mondjesmaat onderwezen. Tenzij er sprake is van een zekere mate van specialisatie. Veel IT-specialisten, bijvoorbeeld, hebben kennisgemaakt met die technieken omdat die te maken hebben met het inrichten van het primaire proces dat met IT moet worden ondersteund of uitgevoerd.

In de column van Paul van Batenburg en Jacques de Swart (2017) over de haat-liefdeverhouding van de accountantsopleidingen met statistiek is uiteengezet dat de meeste accountancystudenten in de propedeuse-, de bachelor- en de masterfase te maken krijgen met statistiek. Zowel beschrijvende als wiskundige statistiek, zoals steekproeven en regressieanalyse. Deze statistiek wordt vaak niet herhaald in de hoofdvakken interne beheersing en auditing. Soms wordt toegelicht welke betekenis die statistiek voor het accountantsberoep heeft en waar dat in het vakgebied terug te vinden is (zoals detailcontroles en cijferanalyses). In sommige specifieke cursussen wordt aandacht besteed aan statistiek in rela-

tie tot interne beheersing. De onderwerpen die dan aan bod komen zijn dan vooral compliance tests³, voorraadopnames, verbandscontroles en bestandsopschoning⁴. Deze drie benaderingen zullen wij hierna schetsen. Ten slotte staan wij nog even stil bij het fenomeen process mining.

Compliance tests (klassiek)

Compliance tests zijn doorgaans steekproeven om achteraf te constateren of de interne beheersing die sporen heeft achtergelaten die vooraf nodig werden geacht. De omvang van dergelijke steekproeven is niet met kwantiteiten aan te duiden. De veel aangetroffen waarden voor de steekproefomvang als 25, 40 en 60 hebben geen andere inhoud dan 2, 3 of 5 waarnemingen per maand (in vijfvouden).

Vaak worden compliance tests achteraf als enige soort compliance test beschreven. Daarmee wordt impliciet uitgesloten dat de controle ook tijdens processen kan plaatsvinden: directe controle. Om een en ander met elkaar in verband te zien is een essay opgenomen over het transactiemodel⁵.

Omdat aan het verrichten van compliance tests en dergelijke geen kwantitatief model ten grondslag ligt, kan een toereikende hoeveelheid van dergelijke tests voortkomen uit twee bronnen. Ten eerste moet de test uitvoerbaar zijn naar de geest van Popper (1959)⁶. Ten tweede moet men durven te uiten wat de subjectieve overtuigingskracht is van de hoeveelheid werk die verricht wordt. Die hoeveelheid kan variëren met de complexiteit van

de waar te nemen fenomenen. Hoe complexer hoe minder waarnemingen, met een minimum vergelijkbaar met een stel lijncontroles.

Wil men compliance tests ook nog uitvoeren door in de laag tussen 'real world' en primaire vastlegging te steken, dan loopt men tegen het probleem aan dat de actie niet herhaalbaar is. Mijns inziens is vooral van belang te onderkennen dat een paraaf op een bescheid niet de directe waarneming van de 'real world' kan vervangen. Evenzo van belang is het besef dat als er niet een integrale directe waarneming plaatsvindt er misschien wel vastleggingen zijn die of equivalent of ten naaste bij equivalent daaraan kunnen zijn.

Als er compliance tests worden uitgevoerd door middel van een steekproef zou het goed zijn zowel de benadering van attributensteekproeven toe te passen als die van geldsteekproeven. Ook Leslie, Teitlebaum en Andersen (1979) houden daar een pleidooi voor. Het effect van het niet goed toepassen van de regelgeving kan op die manier in geld worden uitgedrukt. Zo kan ook duidelijk worden gemaakt dat sommige slordigheden of verzuimen niet per se een financieel effect hoeven te hebben.

Compliance tests (IT-omgeving)

Als in een IT-omgeving gebruikers inloggen, wordt daar idealiter een vastlegging van gemaakt. Omdat er gebruikers zijn met veel bevoegdheden, zoals beheerders, kan er ook misbruik van die bevoegdheden worden gemaakt. Als de bedrijfs-

leiding redelijke zekerheid wil hebben dat zulk misbruik niet plaatsvond kan hij de betreffende logfiles (laten) onderzoeken. Voor een deel kan dergelijk onderzoek met data-analysetechnieken plaatsvinden. Zo kunnen de records van het personeel met administratorbevoegdheden geselecteerd worden. Ook kan vaak de aard van de actie in de log worden gezien. Bijvoorbeeld het aan- of uitzetten ervan. Omdat soms het beheer complex is zien we ook een complexe log. En soms met (heel) veel onverwachte acties. In dergelijke gevallen loont het de moeite per kritieke mutatie na te gaan wat er verder is gebeurd. Hoe complexer, arbeidsintensiever, dat is, hoe meer tijd het kost. Ook hier geldt: in de eerste plaats moet een accountant kijken of een dergelijke compliance test achteraf kan plaatsvinden (is de IT Popper-proof), in de tweede plaats moet er een schatting gemaakt worden hoeveel tijd er in het onderzoek mag of moet worden gestoken.

Voorraadopnames

Voorraadopnames en het valideren van uitstaande vorderingen met saldobiljetten zijn controles die te maken hebben met de omvang van balansposten en die ook met de interne beheersing rond goederen en vorderingen te maken hebben. Van beide typen controles is bekend dat zij een heleboel fouten plegen te laten zien. In die gevallen zijn toetsingsmethoden om zo snel mogelijk goed te keuren niet toepasbaar. Die zijn er immers op geënt dat in de steekproef geen of heel weinig fouten zullen voorkomen. In een apart essay, een toelichting eigenlijk, over het toepassen van de

familie van de regressieschatter laten we zien hoe er met gebruikmaking van de beschikbare informatie een zo klein mogelijke steekproefomvang kan worden gebruikt met een zo hoog mogelijke nauwkeurigheid. In dat essay wordt uitgelegd onder welke condities dat mogelijk is.

De genoemde methode is geschikt om iedere deelpopulatie waarin veel afwijkingen voorkomen op deze manier aan te pakken. Zo is de methode in de praktijk ook toegepast om de fouten in persoonsdeclaraties te schatten.

Verbandscontroles

Het beoordelen van verbanden in de waardenkringloop van organisaties gebeurt met onder meer inventarisaties (een doorsteek naar de real world) en met onderzoek naar de consistentie van de verbanden. Het werk van Elsas (1996) illustreert de mogelijke nauwgezetheid waarmee die verbandscontroles vorm kan worden gegeven. Elsas (1996) behandelt de afzonderlijke transaction cycles. De verbanden kunnen ook vormgegeven worden met behulp van geaggregeerde gegevens. Een belangrijke bron voor het toepassen van statistiek is de aanpak van Stringer en Stewart (1985). De accountant onderzoekt de kwaliteit van de verbanden tussen transactiestromen in periode A. Hij maakt daarvan een model en in periode B gaat hij met behulp van regressieanalyse na of de verbanden van periode A ook in periode B (significant) gelden. Indien dat niet zo is, verricht hij nader onderzoek.

Bestandsopschoning

Bij de essays over risicoanalyse in de controle is een artikel opgenomen van Veenstra en Heertje (2006). Dat artikel gaat in op de zogenoemde AOQL-methode⁷. Dit is een techniek waarbij het betreffende bestand, dan wel een populatie andere elementen, in een hoeveelheid compartimenten wordt opgedeeld. Vervolgens wordt in ieder compartiment een steekproef getrokken. Vooraf wordt bepaald in welke van de compartimenten hoeveel fouten mogen voorkomen. Het aantal fouten dat is toegestaan kan variëren van nihil tot meerdere. Hoe meer fouten er worden toegestaan, hoe groter de steekproef in dat compartiment moet worden.

Wordt er bij het controleren een aantal fouten aangetroffen dat kleiner is dan of gelijk aan het toegestane aantal fouten, dan worden die foute elementen gecorrigeerd. Als er meer fouten worden gevonden in een compartiment dan is toegestaan, dan wordt het betreffende compartiment in zijn geheel (integraal dus) gecontroleerd. Achteraf is er dus een hoeveelheid werk verricht die minimaal gelijk is aan de som van de steekproeven in de compartimenten. De maximale hoeveelheid werk is dat alle compartimenten integraal worden gecontroleerd. Na controle, en dus verbetering, is er sprake van een populatie (een bestand, bijvoorbeeld) dat ten minste voldoet aan de vooraf gestelde eisen.

Process Mining

Process mining is een serie technieken waarbij men gebruikmaakt van data die worden vast-

gelegd in de transactieverwerkende systemen zoals ERP-pakketten. Idealiter wordt iedere transactie die door middel van het ERP-systeem wordt uitgevoerd gelogd. Zo'n logfile wordt eventlog genoemd⁸. Als de eventlog compleet is kan die bijvoorbeeld worden gebruikt om de consistentie van verschillende transactiesoorten te onderzoeken. Dat komt neer op dat in detail de waardenkringloop wordt onderzocht die met het ERP-systeem is vastgelegd. Aangevuld met de methodiek van computational audit zoals uiteengezet door Elsas (1996) bieden de technieken een aanvulling op de klassieke verbandscontroles⁹. Zowel op de control cycles zoals die in de Engelstalige literatuur zijn beschreven als op de Nederlandse controleaanpak met totaalverbandscontroles.

Voor de aansluiting met de real world (zie Essay transactiemodel) blijft directe controle of een (echt) equivalent nodig. Voor een aantal soorten transacties zou zo'n eventlog als equivalent kunnen gelden. In de praktijk moet nog ervaring worden opgedaan met de uitzonderingen op die equivalentie. Zeker in die situaties waar een bedrijf met meerdere ERP-systemen tegelijkertijd werkt.

Literatuur

Aalst, W.M.P. van der, A.J.M. Koopman (2015);
Process mining: data analytics voor de accountant
die wil weten hoe het nu echt zit, *Maandblad voor
Accountancy en Bedrijfseconomie*, pp. 359-368.

Veenstra, R.H., A. Heertje (2006); Doeltreffende
administratieve controle door steekproeven,
Maandblad voor Accountancy en Bedrijfseconomie,
(april), pp. 611-619.

Batenburg, P.C. van, J.J.M. de Swart (2017);
Haat-liefdeverhouding tussen accountants en statistiek,
Column Statistical Audit 61, Accountant.nl.

Broeke, ten, A.M., W.P.A. van der Heijden,
R.W. Sierenberg, H. Terpstra (1977);
Operationele Research, Samsom.

Elsas, P.I. (1996); *Computational Auditing*,
dissertatie, VU, Amsterdam.

Kloosterman, H.H.W. (2012); *Mogen mijn bestanden
schoon?* Column Statistical Audit 15, Accountant.nl

Lee, P.H. van der (1992); *Automatisering in
Voorraadbeheer en Productiebesturing*; Kluwer
Bedrijfswetenschappen.

Leslie, D.A., A.D. Teitlebaum en R.J. Anderson
(1979); *Dollar Unit Sampling: A Practical Guide for
Auditors*, Copp Clark Pitman, Toronto.

Popper, K.R. (1959); *The Discovery of Scientific
Discovery*, Hutchinson Education.

Noot

- 1 Zie onder meer Ten Broeke et al. (1977).
- 2 Zie ook Van der Lee (1992).
- 3 Onder compliance tests pleegt te worden verstaan een controle waarbij er gekeken wordt of er bijvoorbeeld een correcte paraaf is opgenomen op een brondocument. Dat zou men een klassieke compliance test kunnen noemen. Een nieuwere soort compliance test kan zijn dat in een logfile wordt gekeken of er niet-geoorloofde transacties in voorkomen (bijvoorbeeld het uit- en weer aanzetten van log-acties door beheerders met administratorbevoegdheid).
- 4 Zie bijvoorbeeld de column van Kloosterman (2012) over dit onderwerp.
- 5 Zie daartoe het aparte essay van Kloosterman dat erover in de bundel is opgenomen.
- 6 In de Inleiding van de essaybundel is dit citaat van Popper (1959) opgenomen: "...For I do not demand that every scientific statement must have been tested before it is accepted. I only demand that every such statement must be capable of being tested..." (p. 23)
- 7 AOQL staat voor Average Outgoing Quality Limit.
- 8 Zie Van der Aalst en Koopman (2015).
- 9 Daar voeg ik aan toe dat de artikelen die ik over process mining zag, ervan uitgaan dat de complete database (dus ook de semipermanente gegevens) correct is en volledig. En daar hebben we dan weer AOQL voor. Ook een impliciete stelling in die artikelen is de (continue) goede werking van General IT-controls.



Detailcontroles

Steekproefselectie, de basis

Paul van Batenburg

Wie bij het gebruiken van een data-analysetool als IDEA of ACL om een (geld)steekproef vraagt, krijgt de keuze uit drie opties: een rechttoe rechtaan steekproef, een steekproef op basis van een vast interval, of een steekproef door middel van Cell-sampling¹. In dit artikel worden deze opties kort uit de doeken gedaan.

Laten we beginnen met de standaard steekproef. Niks mis mee: we laten de computer gewoon n waarnemingen uit M steken. In geval van een geldsteekproef zijn dat eurocenten, en de software vertelt welke transacties geraakt zijn. Het is mogelijk dat een transactie tweemaal of vaker geraakt wordt, en door even te zoeken in de menustructuur van de software kan ook duidelijk worden welke euro(cent) er precies gestoken is.

Nogmaals, niks mis met deze methode, ware het niet dat accountants graag hun steekproef gelijkmatig over de populatie verdelen. Niet zozeer vanwege een wat overdreven behoefte aan representativiteit, maar omdat het handig plannen is als men kan voorspellen hoeveel waarnemingen alvast gecontroleerd kunnen worden in een deel van de te controleren populatie.

Daarvoor is de methode van het vaste interval bedacht. We steken n euro's uit M door eerst $J = M/n$ te bepalen (naar beneden afronden a.u.b.!) en uit de eerste J euro's er random één aan te wijzen. Door daar vervolgens telkens J bij op te tellen krijgt men n waarnemingen die mooi gelijkmatig over de populatie zijn verdeeld en een post van omvang X heeft selectiekans $X \cdot n/M$.

Deze simpele methode was ook toepasbaar geweest in de niet meer voor te stellen situatie dat u met de hand 10 dossiers uit een stapel van 100 moet kiezen: door de 3e, de 13e, 23e, etc. te pakken zorgt u voor een eerlijke steekproef waarin elk dossier evenveel kans op selectie heeft en voorkomt u dat u na 60 dossiers er al 9 heeft aangewezen.

Is het echt zo simpel? Ja, en een beetje nee. Als u die eerste waarneming maar netjes random kiest is er wat dat betreft niets aan de hand. Maar, hoe kiest u 3 waarnemingen uit 10?

De oplossing is om de populatie eerst aan te vullen met de niet-bestaande 11e en 12e waarneming en dan uit 4 cellen steeds een waarneming te kiezen. In 1 op de 3 gevallen zult u dus 3 waarnemingen vragen en er 4 krijgen.

Verder is er voor statistici een interessante vraag of deze steekproef wel voldoende random is, omdat zodra de eerste waarneming gekozen is alle andere waarnemingen vastliggen. In plaats van M^n , zijn er nog maar J mogelijke steekproeven te trekken. Over dat probleem heeft de beroemde statisticus Hoeffding (1956) zich druk gemaakt.

Zijn conclusie is: wie zo de steekproef trekt mag niet meer gebruikmaken van formules voor trekken zonder terugleggen. Maar ja, dat waren accountants toch al niet van plan, want de steekproef is natuurlijk maar een fractie van de populatie.

De methode van het vaste interval heeft een interessant gevolg voor grote posten: Een post van omvang J moet minstens één keer aangewezen worden, en als een post meer dan eenmaal is aangewezen moet hij groter zijn dan J .

Gevolg is dat fouten in posten groter dan J (het zogenaamde top-stratum) niet geëxtrapoleerd worden; hun foutbedrag wordt opgeteld bij de geprojecteerde en de maximale fout, en na correctie mogen die bedragen weer afgetrokken worden.

Er is wel een ander probleem. Stel we controleren de salarissen met een steekproef met een vast interval. Het is mogelijk dat het repeterende karakter van het salarisbestand met bedragen per persoon per maand in de steekproef terugkomt. J zal maar gelijk zijn aan het totaal van de salarissen per maand...

De derde methode is Cell-sampling, beroemd geworden door Leslie, Teitlebaum en Anderson (1979), al is hun boek meer beroemd geworden door de voorgestelde evaluatiemethode dan door de selectiemethode. Leslie et al. (1979) wilden gehoor geven aan de twijfel of het vaste interval wel random genoeg was en stelden voor de populatie van M euro's wel in cellen van $J=M/n$ (omlaag afronden!) op te delen, maar dan uit iedere cel een nieuwe random selectie te maken.

Dat lijkt een stuk meer random, toch? Groot voordeel is dat er geen risico meer is van een patroon in de populatie dat in de steekproef doorkomt.

Maar nadelen zijn er ook:

- Het top-stratum bestaat niet meer uit posten groter dan J maar groter dan 2^*J , behalve in de eerste en de laatste cel, daar is de grens weer J ;
- Een post van twee euro kan (zei het met kans $J-2$) tweemaal gestoken worden (en dan telt de fout ook twee keer mee...);
- De kans dat een post van omvang X wordt geselecteerd is een derdegraadsfunctie van X .

Al met al lijkt het er op dat Cell-sampling een paardenmiddel is, en dat de nadelen van selectie met een vast interval gemakkelijk kunnen worden opgevangen door de populatie (random, of naar omvang van de post) te sorteren.

Literatuur

Gill, R.D. (1983); *The Sieve Method as an Alternative to Dollar-Unit Sampling: the Mathematical Background*; Mathematisch Centrum, Amsterdam.

Hoeffding, W. (1956); On the Distribution of the Number of Successes in Independent Trials. *Ann. Math. Statist.*; ed. 27, pp. 713-721.

Leslie, D.A., A.D. Teitlebaum, R.J. Anderson (1979), *Dollar Unit Sampling: A Practical Guide for Auditors*, Copp Clark Pitman, Toronto.

Noot

- 1 Gill (1983) heeft uitgebreid de trekkingsmethoden Systematic-sampling, Cell-sampling en de Zeefmethode (Sieve-sampling) behandeld. De zeefmethode voert te ver voor dit artikel. De toelichting op beide andere methoden is heel helder. Gill vindt Cell Sampling en de zeefmethode aanvaardbaar.

Fundamenteel onderzoek naar wiskundige steekproeven

Harrie Hendriks en Hein Kloosterman

Door een groep mensen aangevuurd door Martien van Zuijlen is een respectabele hoeveelheid onderzoek gedaan naar de wiskunde die gepaard gaat met wiskundige steekproeven in accountancy. Omdat het publiek van het symposium Statistical Audit voor het merendeel bestaat uit mensen die in de praktijk van de controle werken hebben wij gemeend de uitkomsten van het voornoemde onderzoek voornamelijk in de vorm van een samenvatting te vertolken. Daarnaast geven wij een uitgebreide literatuurverwijzing.

Er zijn heel veel researchmemoranda, dan wel artikelen in onze beschouwing begrepen. Het bleek moeilijk te zijn om een samenvatting te maken van de samenvattingen van de afzonderlijke onderzoeksrapporten. Dat heeft te maken met het verschil in formuleren in enerzijds de wiskunde en anderzijds de audittheorie.

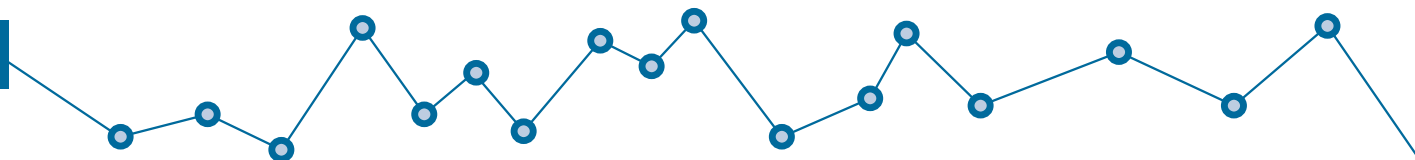
De eerste rapporten laten het gevolg zien van wat Popper noemt 'growth of scientific knowledge' (Popper, 1979). In die rapporten is er steeds iets minder van de praktijk geabstraheerd.

Uitgangspunt in de opgesomde literatuur is dat als de bovengrenzen van de fout in een popula-

tie worden berekend met behulp van de Stringer-bound, deze bovengrenzen erg conservatief worden bevonden. De meeste bevindingen omtrent dat conservatief zijn komen uit onderzoeken die met behulp van simulatie¹ zijn uitgevoerd.

Het eerste onderzoek in de reeks (Pap en Van Zuijlen, 1994a) onderwerpt de Stringer-bound aan een test. In deze (wiskundige) test is gekeken naar taintings met een uniforme verdeling op het traject van nul tot en met één, met een verdeling waarbij de meeste taintings nihil zijn en de overige uniform verdeeld tussen nul en één. Het onderzoek heeft naast de taintings ook nog het risico van ten onrechte goedkeuren gevarieerd. De uitkomst is, en dat blijkt uit een tegenvoorbeeld, dat de Stringer-bound niet altijd conservatief is. Namelijk als de kans op ten onrechte goedkeuren groter is dan 50% en de steekproefomvang groter is dan 16 is de uitkomst niet meer als conservatief te bestempelen.

Pap en Van Zuijlen (1994b) komt met een aangepaste Stringer-bound, die (asymptotisch) precies de correcte betrouwbaarheid heeft. De restrictie dat het conservatief zijn van de Stringer-bound alleen geldt als een kans op ten onrechte goedkeu-



ren van kleiner dan 50% mag worden gehanteerd, is in dit artikel algemeen gemaakt. Dat wil zeggen die restrictie gaat op voor alle eindige steekproeven. Op basis van de centrale limietstelling komt Pap en Van Zuijlen (1994b) daarnaast nog met een aanpassing van de Stringer-bound, waardoor de op die manier berekende bovengrens geldt voor alle waarden van de kans op ten onrechte goedkeuren.

In De Jager et al. (1996) is de aanpassing op de Stringer-bound uitgewerkt met een voorbeeld waarmee het rekenwerk in de praktijk met behulp van een spreadsheet kan plaatsvinden.

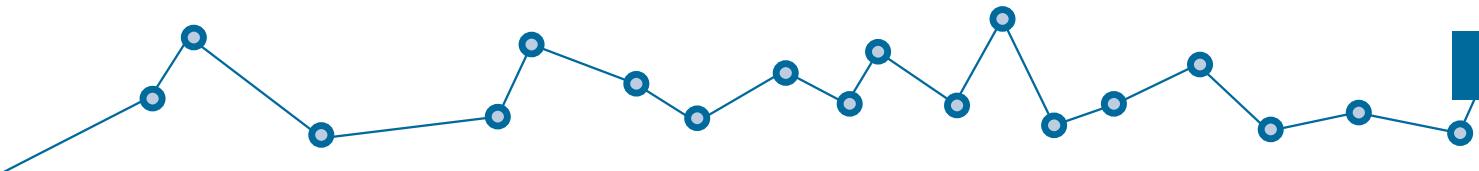
Pap en Van Zuijlen (1996) onderzoekt de bovengrens nogmaals. Nu echter met een andere benadering, te weten een Berry-Esseen Bound. Deze techniek voert te ver voor dit artikel waarin wij enkel de hoofdlijnen van het onderzoek aanstippen.

Bij het STW-project² van 2003 en verder waren mathematisch statistici en probabilisten betrokken die een breed scala aan wiskundige technieken beheersen. Het belangrijkste deel van het onderzoek in het project betrof het zoeken naar een alternatief voor de Stringer-bound. Het startpunt was de binnen de wiskunde zeer bekende Hoeffding-ongelijkheid (Hoeffding 1963), die een grote inspiratiebron geweest is in het wetenschappelijk onderzoek. In het STW-project is er veel inspanning geleverd om de Hoeffding-ongelijkheid te verbeteren, met succes, en er zijn verbeteringen gezocht door het importeren van voorinformatie

over de verdeling van taintings: over de variatie, de eventuele unimodaliteit, symmetrie in de verdeling, et cetera.

Het project heeft geleid tot duidelijke verbeteringen van de Hoeffding-ongelijkheid, en dus van tot dan toe in de statistiek bekende betrouwbaarheidsbovengrenzen. Deze nieuwe grenzen leverden in een uitgebreide simulatiestudie vrij algemeen hogere betrouwbaarheidsbovengrenzen op dan de Stringer-bound, tenzij extra prior-informatie gebruikt kon worden betreffende de taintings. Bovendien was de constructie technisch gecompliceerd. De resultaten hebben dientengevolge weinig of geen toepassing gevonden in de auditing.

Als slot en als kanttekening bij al dit onderzoek willen wij noemen dat het toepassen van een kans op ten onrechte goedkeuren die groter is dan 5% in beginsel valt onder de paraplu van het toepassen van een Audit Risk Model. Aan die toepassingen is een groot aantal artikelen in deze bundel gewijd. Dit artikel voegt in hoofdzaak toe dat behalve simulatiestudies ook een wiskundige studie laat zien dat de Stringer-bound conservatief is. Bij deze kanttekening hoort nog de opmerking dat de methode van Leslie et al. (1979), de Cell Evaluation Method wel een plausibele intuïtieve verklaring kent, maar een 'hard' bewijs – zowel mathematisch als door middel van simulatiestudies – tot nu toe ontbreekt.



Literatuur

Bentkus, V. (2004); On Hoeffding Inequalities. *Annals of Probability*. Vol. 32, No. 2, 1650-1673.

Bentkus, V., G.D.C. Geuze, M.C.A. van Zuijlen (2006); Optimal Hoeffding-like inequalities under a symmetry assumption, *Statistics*, Vol. 40, pp. 159-164.

Bentkus, V., M.C.A. van Zuijlen (2003); On conservative confidence intervals. *Lietuvos Matematikos Rinkinys (Lithuanian Mathematical Journal)*. Vol. 43, No. 2, 169-193.

Bentkus, V., M.I. Kalosha, M.C.A. van Zuijlen (2007); Confidence bounds for the mean in nonparametric multi sample problems, *Statistica Neerlandica*, Vol. 61, pp. 209-231.

Bentkus, V., M.I. Kalosha, M.C.A. van Zuijlen, On domination of tail probabilities of (super) martingales: Explicit bounds. *Lietuvos Matematikos Rinkinys (Lithuanian Mathematica Journal)*. Vol. 46, No. 1 (2006).

Bringle, G.S., (1986); *A study of upper error limits in accounting populations*, Airforce Institute for Technology US (AD-A174 172).

Hendriks, H., C. Kraaikamp, L. Meester, P. Mokveld, M. Nuyens (2005); *Isolating and correcting errors while auditing accounts* Proceedings of the 48th European Study groep Mathematics

with Industry, Delft, 15-19 March 2004. C. Kraaikamp, H. Lin, C. Oosterlee (Eds.), 9-21. Delft University Press, Delft.

Hoeffding, Wassily (1963); Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*; Jg 58 (301): pp. 13-30.

Jager, N.G. de, G. Pap, M.C.A. van Zuijlen (1996); *Facts, Phantasies and a New Proposal Concerning the Stringer-bound*; Research memorandum Limperg Instituut, Amsterdam.

Leslie, D.A., A.D. Teitlebaum, R.J. Anderson (1979); *Dollar Unit Sampling: A Practical Guide for Auditors*, Copp Clark Pitman, Toronto.

Lucassen, A., H. Moors en P.C. van Batenburg (1996); Modifications of the Stringer-bound: a simulation study on the performance of audit sampling evaluation methods, *Kwantitatieve methoden*, nr. 51, 1996, pp. 17-25.

Neter, J., Loebecke, J.K. (1975); *Behavior of Major Statistical Estimators in Sampling Accounting Populations: An Empirical Study*. Auditing Research Monograph no. 2, New York: AICPA.

Newiak, M., (2009); *Prüfungsurteile mit Dollar Unit Sampling*, Universität Postdam.

Panel on Nonstandard Mixtures of Distributions (1988); *Statistical Models and Analysis in Auditing*; National Academy Press, Washington, D.C.

Pap, G., M.C.A. van Zijlen (1994a); *On the Stringer-bound in case of uniform taintings*; Rapport 9433; Radboud Universiteit Nijmegen en Limperg Instituut, Amsterdam.

Pap, G., M.C.A. van Zijlen (1994b); *On the asymptotoc behaviour of the Stringer-bound*; Rapport 9434; Radboud Universiteit Nijmegen en Limperg Instituut, Amsterdam.

Pap, G., M.C.A. van Zijlen (1996); *A Berry-Esseen Bound for Linear Combinations of Order Statistics*; Rapport 9618; Radboud Universiteit Nijmegen en Limperg Instituut, Amsterdam.

Pijnenburg, M.G.F., N.I. Kalosha, M.C.A. van Zijlen (2006); *Optimaility of Sampling Plans in auditing*, Rapport 0601, Radboud Universiteit Nijmegen en Limperg Instituut, Amsterdam.

Popper, K.R. (1979); *The Growth of Scientific Knowledge*, Klostermann, Frankfurt am Main.

Roberts Donald M. (1978); *Statistical Auditing*, AICPA.

Schapendonk-Maas, H. (1998); *Betrouwbaarheidsbovengrenzen en steekproeftechnieken in de accountancy*, februari 1998; Research Memorandum, Limperg Instituut.

Noot

- 1 Het simulatieonderzoek zoals in Bringle (1986), Neter et al. (1975), Newiak (2009) en Schapendonk (1998) moet onzes inziens gezien worden als een motor om het theoretisch onderzoek zoals door Van Zuijlen en anderen is gedaan te stimuleren. Tussen deze benaderingen in staat de (intuïtieve) logica van Leslie et al. (1979) die leidt tot Cell Evaluation Method die zij als minder conservatief alternatief voor de Stringer-bound bieden.
- 2 Dit onderzoek is verricht in het kader van een STW-project. Doelstelling daarvan was zowel de theorie met betrekking tot het berekenen van bovengrenzen te verbeteren, de te berekenen bovengrenzen minder conservatief te maken én gereedschap te ontwikkelen voor die verbeterde versie(s) van een bovengrens. De rapporten van Van Zuijlen c.s. na 2003 behoren tot dat project.

Steekproeven bij volledigheidsc controles¹

Paul van Batenburg

Onder accountants heerst het hardnekkige misverstand dat steekproeven niet toepasbaar zijn bij volledigheidsc controles. In deze column worden enkele methoden besproken waarmee wel of niet bevindingen over volledigheid zijn af te leiden.

Een geldsteekproef is een selectie van posten uit een populatie waarbij elke geboekte euro gelijke kans heeft om geselecteerd te worden. Niet-geboekte euro's kunnen niet worden aangewezen. Met een geldsteekproef kan dus worden nagegaan of een geboekte euro terecht geboekt is. Of een niet-geboekte euro geboekt had moeten worden is niet vast te stellen.

Een gestileerd voorbeeld:

Uit drie posten met boekwaarden (Ist-positie, verantwoord) 100 euro trekken wij een geldsteekproef van 1 waarneming.

Elke post heeft kans 100/300 om geselecteerd te worden.

De auditwaarden (soll-positie, gecontroleerd) van de drie posten zijn 100, 80, respectievelijk 120. Als de huishouding geen fouten had gemaakt, waren de drie trekkingskansen respectievelijk 100/300, 80/300 en 120/300 geweest.

- De eerste post is correct en heeft de juiste trekkingskans.
- De tweede post is door de huishouding te hoog opgevoerd (*overstated*) en krijgt daardoor een grotere trekkingskans (100/300) dan wanneer die fout niet was gemaakt (80/300). Daarom zijn geldsteekproeven een uitstekend middel om te controleren op overstatements: hoe meer euro's de huishouding onterecht boekt, des te groter de kans op controle.
- De derde post is door de huishouding te laag opgevoerd (*understated*) en krijgt daardoor een lagere trekkingskans (100/300) dan wanneer die fout niet gemaakt was (120/300). Daarom zijn *geldsteekproeven een onjuist middel om te controleren op understatements: hoe meer euro's de huishouding onterecht niet boekt, des te kleiner de kans op controle.*

Wanneer is besloten gebruik te maken van een geldsteekproef is dus ook besloten om eenzijdig te controleren op de overstatements, en understatements niet te extrapoleren. Vaak is die keuze een vanzelfsprekend gevolg van de controlebenadering: de gegevensgerichte controle dekt de primaire controlerichting af, en de secundaire controlerichting wordt op een andere wijze afgedekt.

Als in een steekproef op salarissen een te laag geboekt salaris (*understatement*) wordt ontdekt zijn alle wel betaalde euro's correct geboekt. De accountant zal de bevinding melden en moeten beslissen of onderzoek nodig is naar het vaker voorkomen van understatements. Dat onderzoek kan niet worden gebaseerd op de getrokken steekproef, want de aangetroffen understatement is met een te kleine kans in de steekproef gevallen.

Alternatieve methoden voor evaluatie van understatements

Wanneer duidelijk is dat niet alleen overstatements maar ook understatements in de populatie kunnen voorkomen en beide foutsoorten in de evaluatie dienen te worden betrokken, zijn er voor de toepassing van steekproeven twee mogelijke alternatieven.

1. *Per post is een maximale potentiële auditwaarde bekend*

Een voorbeeld: de Belastingdienst controleert onder andere of de belastingplichtige zijn afnemers voldoende btw in rekening heeft gebracht. Er bestaan drie tarieven: nul procent, zes procent en eenentwintig procent. We gaan uit van drie facturen van 100 euro netto (exclusief btw) waaruit een steekproef van één waarneming wordt geselecteerd.

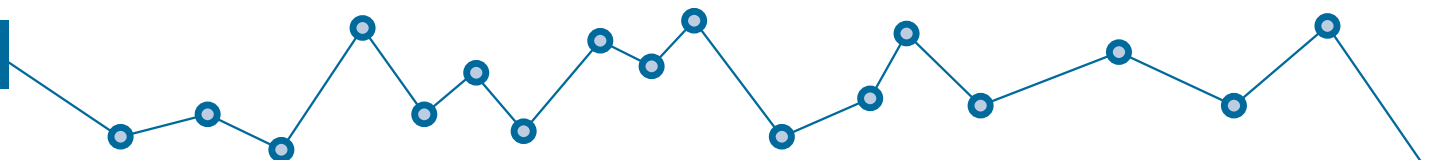
- *Factuur 1 van 100 euro netto is uitgegaan voor 121 euro.* De in rekening gebrachte btw kan alleen maar te hoog zijn, maar niet te laag. Voor de steekproefcontrole op understatement is dit irrelevant.

- *Factuur 2 van 100 euro netto is uitgegaan voor 106 euro.* De in rekening gebrachte btw kan te hoog zijn, maar ook te laag. Voor de steekproefcontrole op understatement is deze factuur wel relevant. De potentieel te laag in rekening gebrachte btw bedraagt 15 euro.
- *Factuur 3 van 100 euro netto is uitgegaan voor 100 euro.* De in rekening gebrachte btw kan niet te hoog zijn, maar wel te laag. Voor de steekproefcontrole op understatement is deze factuur relevant. De potentieel te laag in rekening gebrachte btw bedraagt 21 euro.

Door het bestand met factuurbedragen in te lezen in een softwarepakket zoals ACL (Audit Command Language) en aan te vullen met een kolom 'potentieel te laag in rekening gebrachte btw' waarin een bedrag wordt berekend:

Potentieel te laag in rekening gebrachte btw = 1,21 x nettobedrag – brutobedrag,
kan uit die kolom een geldsteekproef worden geselecteerd waarin factuur 1 kans 0, factuur 2 kans 15/36 en Factuur 3 kans 21/36 krijgt om gecontroleerd te worden.

Deze berekening geeft een *beter*e steekproef dan een selectie uit de nettobedragen, omdat de kans op selectie nu evenredig is met de potentiële fout. Het levert ook een *kleinere* steekproef op, omdat de te controleren populatie is teruggebracht van 300 naar 36 euro².



2. De maximale potentiële auditwaarde per post is onbekend

Als we hetzelfde getallenvoorbeeld gebruiken voor een controle op understatement van de gefactureerde omzet bestaat er eigenlijk geen duidelijke, bruikbare, bovengrens voor de potentieel onvolledige omzet. Bij gebrek aan beter vallen we terug op een postensteekproef waarin iedere post gelijke kans heeft op selectie, omdat iedere post eenzelfde, theoretisch oneindig groot, potentieel gemist bedrag met zich meedraagt.

Er is een methode beschikbaar om postensteekproeven in geld te evalueren. De techniek staat bekend als *verschil- of regressieschatter* (zie bijvoorbeeld Touw en Hoogduin, Statistiek voor Accounting en Controlling). Deze methode is vooral bedoeld voor steekproeven van posten uit populaties van posten waarin fouten naar twee kanten kunnen voorkomen. Typische voorbeelden zijn steekproefsgewijze voorraadopnames of controles op waarderingsprijzen van voorraden.

Groot voordeel van deze aanpak is dat hiermee door de selectie van posten *tegelijkertijd op overstatement en op understatement gecontroleerd* kan worden. In het eerste voorbeeld zouden daarvoor twee aparte geldsteekproeven nodig zijn.

Tegenover de ruime toepasbaarheid van de methode staan ook beperkingen.

- Er is geen formule die bepaalt hoe groot de postensteekproef moet zijn op grond van materialiteit en risico. Vuistregel is: evenveel posten als bij een geldsteekproef, maar niet

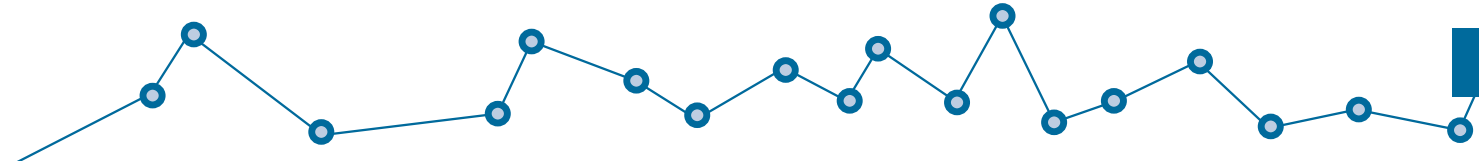
minder dan 100 als er veel verschillen worden verwacht en niet minder dan 200 als er weinig verschillen worden verwacht. Voordeel van een of twee geldsteekproeven zou zijn dat hun omvang uit de controlemix bepaald kan worden.

- Bij de selectie moet worden getracht de gemiddelde waarde van de te controleren steekproefposten niet te veel te laten afwijken van de gemiddelde waarde van alle posten. Dat kan door vooraf negatieve geboekte waarden af te zonderen (vaak zijn die niet plausibel), door waarden nihil bij de selectie te negeren mits deze vanzelfsprekend correct zijn (voorraadwaarden van niet meer gevoerde artikelen), en door een aantal hoogste waarden afzonderlijk te controleren.

Hoe het absoluut niet moet

Hierboven is aangegeven waarom in een geldsteekproef *understatements* niet kunnen worden geëxtrapoleerd en welke alternatieve steekproefmethoden wij ter beschikking hebben. Een alternatief is daarbij bewust niet behandeld omdat het onjuist is: het absoluut maken van fouten.

Terug naar het eerste voorbeeld: drie posten met geboekte waarden 100 en auditwaarden 100, 80, respectievelijk 120. De auditwaarde van deze populatie is 300. Post 1 is correct, post 2 bevat 100 euro's waarvan er 20 onjuist zijn en post 3 bevat 100 euro's die alle juist zijn. Post 3 is niet correct, maar de understatement van 20 zit niet in de 100 geboekte euro's maar had daar nog bijgeboekt



moeten worden. Het is dus onterecht om post 3 met een fout van -20 net zo te behandelen als post 2 met een fout van +20. Absoluut nemen van beide fouten zou immers suggereren dat de auditwaarde van de populatie 260 zou zijn!

De evaluatie van deze populatie zal moeten weergeven dat de 300 geboekte euro's een fout van 20 bevatten maar dat er ook nog 20 andere euro's ontbreken. Door in een geldsteekproef understatements absoluut te nemen lukt dat niet. Door een van de twee besproken alternatieve steekproefmethoden voor de opzet van de controle op understatements te gebruiken lukt dat wel.

Noot

- 1 Dit artikel is eerder geplaatst als column Statistical Auditing nummer 4 op Accountant.nl.
- 2 Dat is zo indien voor deze controle dezelfde toelaatbare afwijking wordt gebruikt om de steekproef op te zetten. Men kan ook redeneren dat nu niet de grondslag voor de btw wordt gecontroleerd maar de btw zelf. Dan kan men een interval in belastingeuro's gaan toepassen. Maar dat voert erg ver voor dit artikel.

Isoleren: atypische of triviale afwijkingen¹

Ed Broeze

1 Isoleren of corrigeren?

In de negentienjarige praktijk als statistisch adviseur bij de Algemene Rekenkamer ben ik vaak geconfronteerd met een controle waarin een 'incidentele', een 'niet-representatieve' of een 'heel uitzonderlijke' fout was gevonden. Zo uitzonderlijk, dat de controlerend accountant van mening was dat je die fout mocht 'isoleren'².

Of te wel: het ophogen (extrapoleren) van de gevonden fout naar de totale te controleren massa, was niet nodig. Immers, iets unieks heeft per definitie een ophogingsfactor van 1.

Analoog ging het met een 'systematische fout', zoals een tijdelijke fout in de programmatuur. Als je al die fouten kunt isoleren (en indien nodig/mogelijk) corrigeren, weet je zeker dat er niet meer van die fouten in de massa zitten, dus waarom zou je de zo gevonden fout extrapoleren?

Deze redenering is al jaren en tot op de dag van vandaag onuitroeibaar. Dit ondanks de vaststelling dat de elementen in de steekproef aselekt zijn getrokken en dat de weg van steekproef naar massa (naast het kansmodel waarmee wordt gerekend) is gebaseerd op die aselechte trekking. Het om wat voor reden dan ook skippen van elemen-

ten, doorbreekt de aselectheid en daarmee een geldige ophoging³.

Burghstahler en Jiambalvo (1986) werken deze vaststelling uit in een 'balletjes, blokjes en andere vormen in een vaasmodel'.

Zij zeggen: laten in een steekproef uit de vaas zwarte balletjes voor correcte items staan en rode balletjes en een rood blokje voor foute items. Van het rode blokje is vast te stellen (meent de accountant) dat het uniek is. Niet meetellen van het rode blokje voor het schatten van (ophogen naar) het aantal rode vormen in de vaas betekent echter dat je zeker moet weten dat er niet ook rode piramides, kegels, schijven, sterretjes etc. in de vaas zijn.

Niet-meetellen van het blokje laat dus al die andere mogelijke vormen buiten schot en leidt daarmee in het algemeen tot onderschatting van de werkelijke fout. Met als gevolg een te grote kans op onterechte acceptatie van een verantwoording met een materiële fout.

De auteurs memoreren: *"A fundamental assumption underlying audit sampling is that items are, for sampling purposes, homogeneous in the sense that observation of some subset of items is useful for drawing*

conclusions about the remainder of the whole population” (pagina 263).

Isoleren en niet-ophogen is strijdig met die homogeniteitsaanname, aldus Burghstahler en Jiambalvo (1986).

Wheeler, Dusenbury & Reimers (1997) trachten de isoleringsmisvatting te redden door te laten zien dat een vertekende schatter (zoals door isoleren ontstaat) wel preciezer kan zijn. En als dat zo is, dat je met die grotere precisie een meer wenselijke schatter hebt. Maar zij laten helaas buiten beschouwing dat die vertekening een onderschatting van de fout inhoudt.

Verontrustend is dat Allen & Elder (2005) melden dat in de periode 1994-1999 het ophogen van fouten minder gebruikelijk is geworden voor de big five-firma's in hun studie. De non-big five zijn wel vaker tot ophogen overgegaan. Zeer verontrustend is hun bevinding dat 'immaterialiteit' van een fout (ook) als argument voor niet-ophogen wordt gebruikt.

Verontrustend is ook dat ik nog zeer onlangs betrokken was bij een discussie of isoleren tot niet-ophogen mocht leiden. Die discussie is helaas verklaarbaar door het meest verontrustende, alleen al om zijn hardleersheid: ISA 530 in A19⁴ geeft ruim baan aan isoleren.

Het zal u, lezer, niet ontgaan zijn dat ik weinig over corrigeren heb gesproken. Wellicht is u ook

niet ontgaan dat de angel ook niet in het corrigeren zit, maar in het isoleren: de beslissing om een fout niet op te hogen.

En dan snapt u ook dat er in het koppel 'Isoleren en corrigeren' mogelijk zelfs een verdwyntruc kan zitten: als we die geïsoleerde fout corrigeren, hebben we toch ook niets meer op te hogen?!

Maar u snapt ook dat je zo'n statement alleen in een column kan maken.

In de volgende paragraaf wil ik graag ingaan op wat met isoleren en corrigeren verder fout kan gaan en op wat wél kan.

2 Isoleren en corrigeren, wat kan dan wel?

In column 9 op accountant.nl ben ik ingegaan op het 'waarom niet' van isoleren en corrigeren, vanuit het gezichtspunt van het doorbreken van de aseletheid van de steekproef en daarmee ongeldig worden van de foutschatting. Er is nog een ander gezichtspunt, wellicht nog meer doorslaggevend.

Vooraf: mijn eerste paragraaf over isoleren heeft een verheugend aantal, inhoudelijk doorwrochte reacties opgeleverd. Ik heb daar al twee keer bij die paragraaf op gereageerd. Een wezenlijk deel van de reacties (door Van Herk) sneed de invalshoek aan die ik in deze paragraaf voor het voetlicht breng. En er was vraag naar 'wat dan wel'.

De accountant overschrijdt, helaas, met zijn geschatte fout⁵ het materialiteitsniveau. Hij zou dus niet mogen goedkeuren. Maar hij ziet dat een of meer fouten 'geïsoleerd' kunnen worden: zij betreffen bijvoorbeeld een (zijns inziens) unieke, of een op een tijdelijke systeemfout terug te brengen fout. En dan 'hoeven die fouten tot niet meer dan het desbetreffende deel van de massa te worden opgehoogd' (zo is de door ISA 530 gesanctioneerde redenering). Wellicht maakt dat al dat de (beperkt) opgehoogde fout onder de materialiteit blijft. En anders brengt hij de vereiste correcties aan. En tot zijn vreugde (en die van de klant) blijft de geschatte en gecorrigeerde fout nu wel onder het materialiteitsniveau.

Met deze handelwijze wordt niet alleen de aselectie doorbroken (zie vorige paragraaf over isoleren), maar ook geeft hij een tweede kans op ten onrechte goedkeuren. De eerste kans: een van (meestal maximaal) vijf procent op ten onrechte concluderen dat de fout onder het materialiteitsniveau blijft. De tweede kans: als de steekproef indiceert dat het materialiteitsniveau wordt overschreden, een zodanige isoleringsstructuur voor de verantwoording en gevonden fout construeren dat, eventueel met behulp van correcties, de opgehoogde fout beneden het materialiteitsniveau blijft. Met opnieuw een kans dat deze procedure tot een onterechte goedkeuring leidt. Eerste en tweede kans tellen in principe op (moet je netjes doen, met voorwaardelijke kansen, maar dat nu even niet precies). Resultaat: een kans van boven de vijf procent op onterecht goedkeuren (van een foute populatie).

Hendriks et al. (2004) hebben deze 'kans te veel' uitgewerkt tijdens een symposium van een week van de TU Delft: Mathematics with Industry. In hun paper construeren zij gevallen waarin de kans op onterecht goedkeuren kan oplopen tot een veelvoud van de beoogde maximale vijf procent. Isoleren is dus beslist geen onschuldig 'een beetje gladstrijken van de bevindingen'.

Isoleren houdt feitelijk in dat men achteraf stratificeert en het ophogen van een in de steekproef gevonden fout tot de desbetreffende strata inperkt. Maar zo ontkent men het bestaan van niet ontdekte fouten, die op andere gronden isoleerbaar zijn. Er ontbreekt een statistische theorie om de gevolgen van isoleren voor de betrouwbaarheid van de audit te kwantificeren. Zou slim *vooraf* stratificeren een oplossing kunnen bieden? Hendriks et al. (2004) laten zien dat ook met vooraf stratificeren het 'tweedekanseffect' zorgt voor een te grote kans op ten onrechte goedkeuren. De verontrusting die in de vorige paragraaf werd geuit wordt er niet kleiner op.

Wat dan wel? Dat is eigenlijk vrij simpel: met een geldsteekproef is het steekproefgemiddelde van de gevonden taints (de fractionele fouten behorende bij de getrokken euro's) een zuivere schatter voor de gemiddelde taint in de massa. Hierbij kan op de bekende wijze⁶ de maximale fout (MAX) worden berekend. Als die boven het materialiteitsniveau uitkomt, kan bij de gewenste betrouwbaarheid niet worden goedgekeurd. Corrigeren (indien mogelijk) van fouten kan met

een bedrag COR; dit levert een nieuwe maximale fout op: $\text{MAX} -/\text{COR}$. Als deze uitkomst onder het materialiteitsniveau komt, kan worden goedgekeurd. De vraag wat de accountant kan doen als goedkeuring niet mogelijk is, komt (ook) aan de orde, via de invalshoek van het schatten, in een volgende artikel, van Paul van Batenburg.

Literatuur

Allen, R.D., R.J. Elder (2005); A Longitudinal Investigation of Auditor Error Projection Decisions; *AUDITING: a Journal of Practice and Theory* 24(2)

Burghstahler, D., J. Jiambalvo (1986); Sample Error Characteristics and Projection of Error to Audit Populations. *The Accounting Review*, Vol LXI no 2.

Hendriks H., C. Kraaikamp, L. Meester, P. Mokveld, M. Nuyens (2005); *Isolating and correcting errors while auditing accounts*. Proceedings of the 48th European Study group Mathematics with Industry, 15-19 March; (in Kraaikamp C., H. Lin, C. Oosterlee (Eds.); *Mathematics with Industry*; Delft University Press).

Wheeler, S.W., R.B. Dusenbury, J.L. Reimers (1997); Projecting Sample Misstatements to Audit Populations: Theoretical, Professional, and Empirical Considerations; *Decision Sciences* 28(2), pp. 261-278.

Noot

- 1 Dit essay is een samenvoeging van twee columns Statistical Audit van de hand van Ed Broeze op accountant.nl. Dit zijn de columns 9 en 11. In deze columns behandelt Broeze het weerbarstige thema 'isoleren en corrigeren'.
- 2 In de huidige regelgeving aangeduid als atypische fout (standaard 530).
- 3 Met ophoging wordt hier het berekenen van de projectie bedoeld.
- 4 Het noemen van atypische fouten komt ook nog in de huidige regelgeving voor.
- 5 De fout kan op twee manieren worden geschat. De geprojecteerde fout en de maximale fout. Beide zijn een schatting. Verderop in de tekst wordt duidelijk dat Broeze de berekening van de maximale fout bedoelt.
- 6 Broeze bedoelt hier met de 'bekende manier' het berekenen van de maximale fout door het toepassen van de zogenoemde Stringer-bound.

Uitbreiden van steekproeven om beter te schatten¹

Paul van Batenburg

COS 530 geeft aan dat een steekproef voldoende groot is geweest om een oordeel over de populatie te geven als de geprojecteerde fout niet te dicht bij de vooraf gekozen norm ligt. Statistici – niet behept met *professional judgement* – vinden een dergelijke beslissingsregel te vaag en suggereren het gebruik van het begrip maximale fout (bij een door de accountant te kiezen betrouwbaarheid). Door die maximale fout met de genoemde norm te confronteren wordt het begrip ‘dicht bij’ geobjectiveerd en voldoet de accountant ook meteen aan de eis uit COS 530 paragraaf 7 dat het steekproefrisico tot een aanvaardbaar laag niveau moet worden teruggebracht.

Kortom, veel accountants kennen en gebruiken het begrip maximale fout, maar in de COS komt het niet voor. Reden voor de International Auditing and Assurance Standards Board (IAASB) om het begrip maximale fout niet op te nemen (hoewel de draft versies er wel mee werkten) was de angst dat een betrouwbaarheidsbovendgrens als te complex zou worden ervaren. Echter, mede dankzij spreadsheets zoals Excel kan iedereen hem uitrekenen. De Excelformule voor de maximale fout is:

GAMMA.INV (betrouwbaarheid; 1+ aantal fouten; populatie in geld gedeeld door steekproefomvang).

Even spelen met de formule? Neem als voorbeeld een assistent die 100 waarnemingen uit 1 miljoen geldeenheden moet controleren en weet dat hij alleen bij 0 fouten mag goedkeuren.

Omdat:

GAMMA.INV (0,95; 1+0; 1.000.000/100) = 30.000

weten we dat de norm voor de maximale fout kennelijk 30.000 was.

Nu vindt hij één (hele) fout.

De geprojecteerde fout is € 10.000 (1% van de steekproef was fout, dus ook 1% van de massa) en de maximale fout is:

GAMMA.INV (0,95; 1+1; 1.000.000/100) = 47.500.

De maximale fout is te hoog voor goedkeuring van de populatie. Ook na correctie van de geprojecteerde fout van 10.000 is het maximum nog 37.500. Veel accountants denken nu aan uitbreiden van de steekproef (‘trek er 59 bij en als die geen fouten hebben zit je weer op 30.000’).

COS 450 (7-9) adviseert de accountant om de klant te vragen de populatie te corrigeren voor een bedrag gelijk aan de geprojecteerde fout. Over uitbreiden van de steekproef om alsnog goed te keuren als er niet nog meer fouten wor-

den gevonden is al voldoende geschreven en gediscussieerd: het is niet realistisch om opeens niet nog meer fouten te vinden, en het gaat ten koste van de betrouwbaarheid van de bovengrens. Immers, het steekproefrisico is al in de eerste ronde opgesoupeerd. Niet aan beginnen, dus.

Wat wel kan, is de schatting van de meest waarschijnlijke (= de geprojecteerde) fout aanscherpen. De COS wil immers de geprojecteerde fout laten corrigeren. Ook al suggereert de COS het, niet veel cliënten staan te trappelen om een geschat bedrag te corrigeren. Waar is de tegenboeking? Bovendien, als die 10.000 hier zou worden gecorrigeerd zou daarna de maximale fout nog steeds 37.500 zijn, en dus boven de norm van 30.000.

We zitten dus in een situatie waarin we niet kunnen goedkeuren, maar ook niet voldoende werk hebben gedaan om de noodzakelijke correctie te berekenen.

Maar er is COS 530 A18:

‘De accountant is ertoe gehouden om de afwijkingen te projecteren op de populatie om een globaal beeld te krijgen van de omvang van afwijkingen, maar deze projectie kan mogelijk niet voldoende zijn om het te boeken bedrag ter correctie van die afwijking te bepalen. De schatting moet dus verbeteren en daarvoor moet extra werk worden gedaan.’

De afstand tussen de maximale en de geprojecteerde fout overschrijdt de norm. Dit is het moment om een extra steekproef te trekken.

Stel, onze assistent doet 100 extra waarnemingen? We gaan ervan uit dat hij dus ook een extra fout vindt (in de eerste 100 vond hij er 1, dus nu waarschijnlijk ook) en we berekenen de maximale fout als volgt:

$$\text{GAMMA.INV}(0,95; 1+2; 1.000.000/200) = 31.500.$$

Nog steeds te veel om goed te keuren (dat willen we ook niet, we willen de fout niet wegpraten, maar pas na correctie goedkeuren), maar de afstand tot de geprojecteerde fout (2 van de 200 is nog steeds 1%) is nu 21.500, ruim beneden de norm! De projectie is nu zo nauwkeurig geschat dat de COS vindt dat deze als fout moet worden behandeld (COS 450 A3) en na correctie kan de populatie worden goedgekeurd.

Moet de steekproef echt worden verdubbeld? Zeker niet. Lucas Hoogduin en Harrie Hendriks (ook stuurgroepleden) hebben al eens laten zien dat men per fout ongeveer $1/3$ van de steekproef moet bijtrekken. U kunt het in Excel zelf narekenen: uit een populatie van M euro's trekken we er n en vinden k fouten.

De bovengrens is

$$\text{GAMMA.INV}(0,95; 1+k; M/n)$$

Als die waarde boven de norm ligt kunnen we (nog) niet goedkeuren.

De projectie is $k*M/n$ dus we kunnen bepalen of de afstand

$$\text{GAMMA.INV}(0,95; 1+k; M/n) - k*M/n$$

wel onder de norm ligt.

Als dat niet zo is moeten we e waarnemingen extra doen waarin we $K \cdot e/n$ extra fouten verwachten.

Na uitbreiding is de afstand:

$$\text{GAMMA.INV}(0,95; 1+k \cdot (1+e/n); M/(n+e)) - k \cdot M/n.$$

Merk op dat de foutprojectie gelijk blijft:

$$k(1+e/n) \cdot M/(n+e) = k \cdot M/n.$$

De zoekfunctie van Excel (goal seek) geeft die steekproefomvang die groot genoeg is om, ondanks extra fouten, na correctie goed te keuren door deze afstand te laten voldoen aan de vooraf gekozen norm.

Noot

- 1 Dit artikel is eerder opgenomen als column Statistical Audit 12 op Accountant.nl.

Als ik niet kan goedkeuren: zes tips¹

Paul van Batenburg

Steekproefcontrole wordt een stuk minder aantrekkelijk als de resultaten niet tot goedkeuring leiden. Het mogelijk vinden van fouten is waarschijnlijk de belangrijkste reden waarom steekproeven onder accountants minder populair zijn dan statistici hopen en verwachten.

Steekproefcontrole is een efficiënte manier om een populatie (jaarrekeningpost, transactiestroom, aangifte) die geen materiële fout bevat goed te keuren. Althans, als je je beperkt tot objectieve, reproduceerbare en valideerbare methoden.

Wanneer je van die eis afstapt, zijn er natuurlijk goedkopere manieren om goed te keuren: tegen 'professional judgement' kan wat dat betreft geen wetenschappelijk onderbouwde manier op.

Zodra er fouten worden gevonden vereist de COS dat die worden geëxtrapoleerd en dat die projectie ter correctie wordt voorgelegd aan de gecontroleerde. Deze eis geldt niet alleen voor statistische steekproeven maar voor alle vormen van detailwaarnemingen. De vraag die de accountant dan moet stellen en beantwoorden is tweeledig: is er voldoende werk gedaan om een oordeel over de verantwoording te geven, en wat zijn de gevolgen voor die verantwoording?

Wat nu als de accountant concludeert op grond van de gecontroleerde steekproef niet te kunnen goedkeuren. Wat dan?

1. *Keur niet 'af', maar keur 'niet goed'!*
Wie afkeurt, roept over zich af te moeten bewijzen dat er een materiële fout bestaat; wie niet goedkeurt, meldt slechts niet te kunnen aantonen dat de fout aanvaardbaar is. Laat het daarbij, dat is al erg genoeg!

2. *Ligt het aan te veel fouten of aan te weinig goede waarnemingen?*
Niet kunnen goedkeuren kan liggen aan of te veel fouten in de populatie – en dus in de steekproef – of aan het verrichten van te weinig werk: te kleine steekproef. Mijn advies is daarom: voeg een berekening toe van het effect van 0 fouten in de steekproef. Laat hiermee en met de evaluatie van de steekproefresultaten zien, dat duidelijk is dat het niet goedkeuren ligt aan het vinden van te veel foute, en niet aan het vinden van te weinig goede waarnemingen!

3. *Uitbreiden is geen wondermiddel maar paardenmiddel*

Bij het uitbreiden van steekproeven na het vinden van fouten vindt u meestal gewoon nog meer fouten, dus alsnog goedkeuren zit er niet in. En als die extra steekproef opeens foutloos is, dan gebeurt dat bij een te lage betrouwbaarheid om te kunnen goedkeuren, want de kans op ten onrechte goedkeuren was in de eerste steekproef al 5% en wordt door die tweede steekproef alleen maar groter.

4. *Corrigeren, maar met welke tegenboeking?*

Elke keer als ik accountants vertel dat de geprojecteerde fout door de gecontroleerde moet worden gecorrigeerd, is er minstens een toehoorder die verzucht: "Dat wil (of mag, of kan) mijn klant toch niet". Jammer dan, het staat in de COS. Wat de accountant kan (en dus moet) doen, is ervoor zorgen dat die geprojecteerde fout wordt berekend met de beste methoden en technieken en op basis van een steekproef die groot genoeg is om een zinnig oordeel te geven.

Volgende tegenwerping is dat de klant dan vraagt die fouten per transactie te overleggen. Maar de accountant is niet de boekhouder. Een lastiger tegenwerping is dat er geen grootboekrekening voor de tegenboeking bestaat. Dat is gedeeltelijk waar: bij balansposten is die er meestal wel (voorziening), bij resultaatposten zal een correctie inderdaad direct het resultaat in moeten. Maar wat van de COS moet, moet.

5. *Maak er een incident van*

Om een homogene massa alsnog heterogeen te verklaren en dus apart te evalueren, is veel nodig. En het leidt niet tot een lager correctiebedrag, maar geeft alleen aan waar dat bedrag waarschijnlijk te vinden is (en dus de tegenboeking vandaan moet komen).

6. *Beter dan genezen*

Een hoop van de hierboven beschreven problemen kunnen worden voorkomen door gewoon vooraf wat meer werk in te plannen. De COS heeft hiervoor de toelaatbare afwijking bedacht, die lager mag zijn dan de uitvoeringsmaterialiteit waartegen de populatie wordt geëvalueerd. Wat ook werkt is vooraf met de gecontroleerde afspreken dat de geprojecteerde fout onverkort zal worden gecorrigeerd. Daarmee kan de steekproef zo worden opgezet dat de maximale fout na correctie aan de materialiteit voldoet.

Een rekenvoorbeeld: het Directoraat-Generaal van de EU dat over landbouwsubsidies gaat, verwacht een steekproefcontrole door de lidstaten met 2% materialiteit en 95% betrouwbaarheid. De steekproefomvang is echter niet 150 (bij 0 fouten) want er moet een inschatting worden gemaakt van de vermoedelijke fout: is die 10% of 25% van de materialiteit? Bij 25% (van 2%, dus 0,5%) wordt de steekproefomvang 262 waarin 1,3 fouten mogen worden gevonden. Als we nu vooraf afspreken dat de geprojecteerde fout gecorrigeerd wordt, kunnen we de maximale

fout op 2,5% stellen en valt de steekproef terug naar 185.

Ten overvloede: ook bij lagere betrouwbaarheden werkt dit, al vraag ik me af hoe dat valt te rijmen met het vinden van fouten.

Noot

- 1 Dit artikel verscheen eerder als column Statistical Audit 38 op Accountant.nl.

Het evalueren van steekproeven

Conservatisme en evalueren¹ (deel 1)

Hein Kloosterman en Ferry Geertman

In dit deel van dit essay leggen wij uit wat in het vakgebied Statistical Audit wordt bedoeld met conservatisme en hoe ons dat kan parten spelen bij de controle. In deel 2 en deel 3 werken wij een voorbeeld uit hoe met dat conservatisme om te kunnen gaan.

Het voornaamste doel van het toepassen van steekproeven in de controle is het bepalen van een betrouwbaarheidsbovengrens voor de fractie (onbekende) fouten in de populatie. Trekken we een steekproef en vinden we geen fouten daarin, dan berekenen we deze bovengrens. We noemen deze bovengrens de maximale fout. De uitkomst van het onderzoek is nu gelijk aan de opzet ervan. De maximale fout is gelijk aan de toelaatbare afwijking (NV COS richtlijn 530). We weten dan dat we voldoende werk hebben gedaan om een uitspraak te doen over de gehele populatie.

Niet iedere onderzochte steekproef zal helemaal foutloos zijn. Zodra er fouten of foutjes in de steekproef voorkomen overschrijdt de berekende

maximale fout, die bovengrens, de toelaatbare afwijking. In accountantscontrole kennen we daarom nog twee grenzen: de materialiteit en de uitvoeringsmaterialiteit (NV COS 320). In grote lijnen staat het begrip materialiteit voor de 'worst case', de bovengrens van de fouten, met betrekking tot de jaarrekening als geheel. De uitvoeringsmaterialiteit staat dan voor de 'worst case' van de onderzochte deelpopulatie. In het algemeen, zo stelt de richtlijn, is de materialiteit groter dan of gelijk aan de uitvoeringsmaterialiteit en is de uitvoeringsmaterialiteit weer groter dan of gelijk aan de toelaatbare afwijking. De op die manier geplande speelruimte heeft met voorzichtigheid te maken.

Maar wat is eigenlijk de betrouwbaarheidsbovengrens?

De betrouwbaarheidsbovengrens bepalen we aan de hand van wiskundige/statistische functies op basis van de geplande betrouwbaarheid en het

verkregen steekproefresultaat. Het is hierbij goed ons te realiseren dat het verkregen steekproefresultaat afhankelijk is van het toeval, dat wil zeggen dat meerdere uitgevoerde steekproeven een verschillend resultaat zullen laten zien. Dit betekent ook dat een specifieke betrouwbaarheidsbovengrens afhankelijk is van het toeval. De uitkomst van de wiskundige/statistische functie losgelaten op het steekproefresultaat heet een 95%-betrouwbaarheidsbovengrens als de kans dat deze uitkomst groter is dan de werkelijke foutfractie 95% bedraagt.

Anders gezegd: als we heel vaak een steekproef zouden trekken, zeg 1.000 keer, en elke keer de betrouwbaarheidsbovengrens zouden berekenen, dan zal (ten minste) 950 keer de berekende bovengrens boven de echte foutfractie liggen en (hooguit) 50 keer beneden de echte foutfractie. Dit is de definitie van een 95%-betrouwbaarheidsbovengrens volgens de klassieke steekproeftheorie.

Nu de stap naar conservatief

Een 95%-betrouwbaarheidsbovengrens heet conservatief als de kans dat deze uitkomst groter is dan de werkelijke foutfractie meer dan 95% bedraagt. Anders gezegd: als we heel vaak een steekproef zouden trekken, zeg 1.000 keer en elke keer de betrouwbaarheidsbovengrens zouden berekenen, dan zal in meer dan 950 keer de berekende bovengrens boven de echte foutfractie liggen en minder dan 50 keer beneden de echte foutfractie.

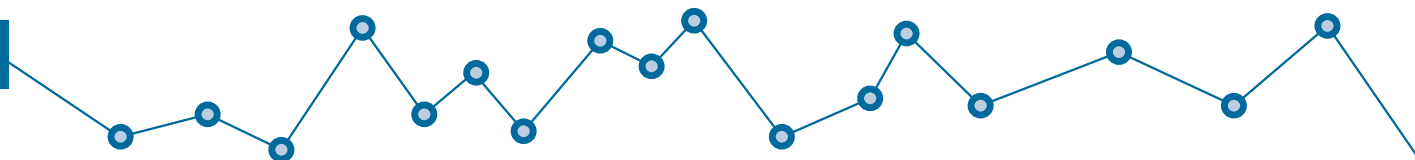
Het berekenen van bovengrenzen in geldsteekproeven gebeurt met de zogenoemde Stringer-bound, waarbij de gevonden relatieve fouten (taintings) in dalende volgorde aangeboden worden aan de functie om de bovengrens te berekenen. Een andere manier van berekenen van de foutbovengrens houdt in dat de gevonden fouten gesorteerd worden in afnemende volgorde van omvang in geld.

Van beide methoden is vastgesteld – op de manier zoals hierboven beschreven, dus door heel vaak een steekproef te trekken op een volledig bekende populatie – dat deze manier van rekenen conservatief is.

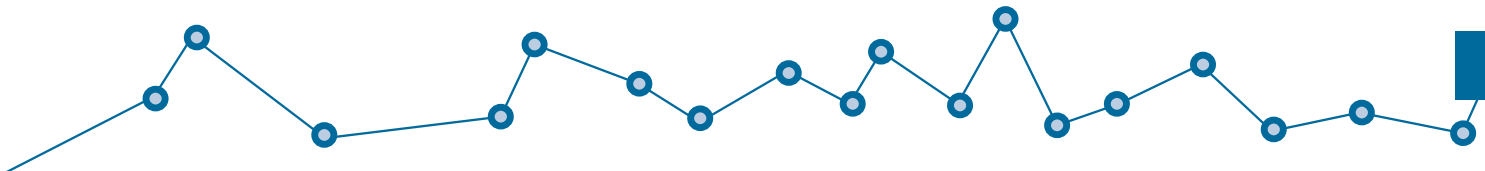
Vertaald naar ons vak betekent dit, dat we eigenlijk te veel werk hebben gedaan. Gegeven de uitgangspunten zijn we niet efficiënt. Anders gezegd: we doen er te lang over om tot goedkeuring over te gaan.

Het loont dan ook de moeite om in die gevallen waarbij de berekende maximale fout nèt boven de (uitvoerings-)materialiteit ligt, alternatieve berekeningswijzen te gebruiken. In volgende delen van dit essay geven we hiervan voorbeelden.

De veelgebruikte auditsoftware zoals ACL en IDEA maakt gebruik van respectievelijk de Stringer-bound en een iets lagere, meer precieze, bovengrens (die techniek voert voor dit deel te ver²).



Samengevat: het berekenen van een maximale fout in de populatie om te zien of er voldoende werk is gedaan om te kunnen goedkeuren, kan op meerdere manieren. Als de foutbovengrens net de (uitvoerings-)materialiteit overschrijdt, is het de moeite waard om na te gaan of er een alternatieve en geldige berekeningsmethode voorhanden is, waarmee de bovengrens minder hoog kan uitkomen en er toch nog kan worden goedgekeurd.



Conservatisme en evalueren (deel 2)

Hein Kloosterman en Ferry Geertman

Hier volgt een concreet voorbeeld om aan te geven hoe men kan omgaan met het conservatisme in de berekening van de bovengrenzen op basis van de uitkomsten van de steekproef.

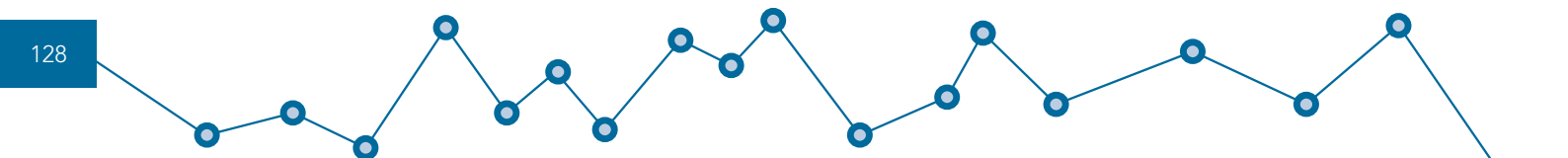
Steekproeven plegen zo te worden opgezet dat zo snel mogelijk kan worden goedgekeurd. Dat betekent dat er een steekproefomvang als uitgangspunt is genomen die zo is berekend dat die steekproef geen fouten mag bevatten. In controletermen: alleen bij nul aangetroffen fouten in de steekproef is de maximale fout van de gecontroleerde subpopulatie kleiner dan of gelijk aan

de goedkeuringsgrens (uitvoeringsmaterialiteit). Die nul-fouten-verwachting is een vrij algemeen uitgangspunt in de aanpak van accountants.

Stel dat de parameters in de controle als volgt waren:

- Audit Risk: 5%;
- Uitvoeringsmaterialiteit voor de onderzochte populatie: 135.000;
- Toegelaten afwijking = controletolerantie = MP = 105.000
- Er is geen risicoanalyse toegepast (dus geen reductie van de steekproefomvang).

Rek #	Geboekt bedrag	Conclusie na controle	Fout #	Fout totaal	Tainting
47500	1364,52	87,50 privé	1	87,50	0,0641
41400	32.971,81	Onderhoud; een deel te activeren	2	9.354,10	0,2837
40250	159,20	19,11 privé-uitgave	3	19,11	0,1200
15230	3250,00	Ten onrechte niet geactiveerd	4	3.250,00	1,0000
		Som relatieve fouten (taintings)			1,4678
Projectie	Interval = 35.000	Som Taintings * Interval			51.373
Maximaal	(ruwe methode)	$(4,75 + 1,55 \cdot 0,4678) \cdot \text{Interval}$			191.628
Grens					135.000



Uit de theorie (Poissonverdeling) volgt dat de factor R drie bedraagt bij een Audit Risk van 5% en nul fouten. Uit dezelfde theorie volgt dat het interval (de intensiteit waarmee de controle plaatsvindt) wordt berekend als $MP/R = 105.000/3 = 35.000$.

De tabel op pagina 128 laat de uitkomsten van de controle zien.
Voor de berekening van MP is relevant dat de uitkomst van de Poissonverdeling bij één fout een $R = 4,75$ en bij 2 fouten $R = 6,30$ laat zien. Die waarden verklaren de berekende bovengrens van 191.628. Immers 3 voor de nullast en 1,75 maal de som der taintings, wordt vermenigvuldigd met het interval. De aangroei van R bij 1 fout tot R bij twee fouten is 6,30 minus 4,75, ofwel 1,55. De bovengrens is dus (zie tabel op pagina 128) 4,75 maal het interval plus 0,4678 maal 1,55 maal het interval. Kortom, die 191.628.

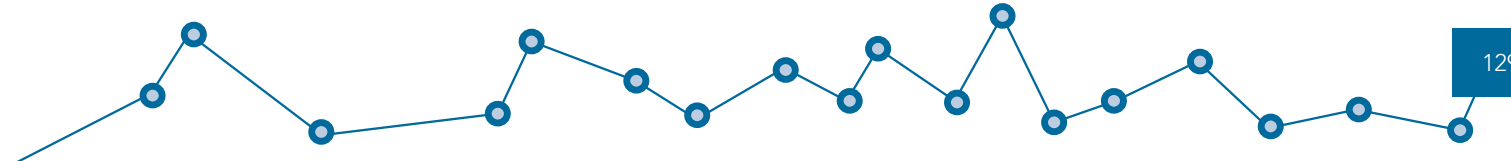
Omdat deze methode vrij ruw is en dus erg conservatief, kijken we naar de ‘next best’ optie en dat is de zogenoemde Stringer-bound. Die bovengrens is ook conservatief, maar houdt rekening met de dalende stijging van de Poissonverdeling.

Uit onderstaande tabel wordt duidelijk dat voor de berekening van de Stringer-bound de aflopende volgorde van de taintings bepalend is. Door de dalende afname van de ‘PGW’ (= ‘precision gap widener’) is de bovengrens lager dan bij de ruwe methode.

Bovenaan pagina 130 laten we nog een berekeningsmethode zien. Deze methode zet de fouten in afnemende volgorde van hun omvang.

Deze methode van evalueren (noemen we EVA-methode) is onderzocht en voldoende conservatief bevonden.

Fout #	Foutbedrag	Tainting	Stringer		PGW	MF
4	3.250,00	1,0000	1		1,75	0,875
2	9.354,10	0,2837	2		1,55	0,439735
3	19,11	0,1200	3		1,46	0,1752
1	87,50	0,0641	4		1,40	0,08974
		1,4678		Nullast:	3,00	3,00
				Totaal:		5,454675
					In geld	190.914



Fout #	Foutbedrag	Tainting	EVA		PGW	MF
2	9.354,10	0,2837	1		1,75	0,496475
4	3.250,00	1,000	2		1,55	0,775
1	87,50	0,0641	3		1,46	0,093586
3	19,11	0,1200	4		1,40	0,168
		1,4678		Nullast:	3,00	3,00
				Totaal		5,308061
					In geld	185.783,00

Recapitulerend zien we dus dat de bovengrens al lager wordt door de keuze van het rekenwerk, maar nog niet laag genoeg om direct te kunnen goedkeuren.

Maximale fout

Maximale fout ruwe methode:	191.628
Stringer-boundmethode:	190.914
EVA-methode:	185.783

Stel dat de cliënt desgevraagd bereid was de foutprojectie te corrigeren. Dan was de opstelling als volgt:

Maximale fout

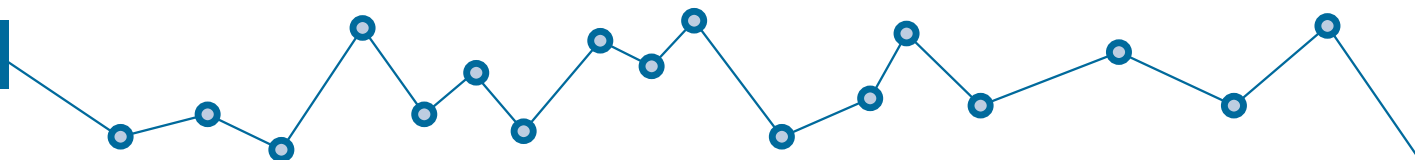
Ruwe methode +/- projectie:	$191.628 - 51.373 = 141.555$
Stringer-bound-methode:	$190.914 - 51.373 = 139.541$
EVA-methode:	$185.783 - 51.737 = 134.046$

Hieruit volgt dat als de EVA-methode wordt toegepast er precies kan worden goedgekeurd nadat de projectie is gecorrigeerd.

Wat is de les uit deze opstellingen?

Als de maximale fout groter is dan de uitvoeringsmaterialiteit kan de accountant (nog) niet goedkeuren. Hij moet dan aanvullend bewijs vergaren. Dat kan hij doen door het berekenen van een scherpere bovengrens van de fout. En dat kan hij doen door de klant de projectie te laten corrigeren.

In dit voorbeeld is er voldoende ruimte gecreëerd tussen de uitvoeringsmaterialiteit (van 135.000) en de controletolerantie (105.000) om de correctie effect te laten hebben: als de EVA-methode wordt toegepast en de foutprojectie wordt gecorrigeerd kan er goedgekeurd worden.



Conservatisme en evalueren (deel 3)

Hein Kloosterman en Ferry Geertman

Wat moet er gebeuren als de accountant zowel fouten vindt als het Audit Risk Model (ARM) heeft toegepast? Mag hij dan evalueren met het ontdekkingsrisico (steekproefrisico) dat met de formule van het ARM ($AR = IR * ICR * CAR * OR$) tot stand kwam? Of moet hij het Audit Riskpercentage hanteren? En wat is toegestaan? Wat is zinvol? En wat zeggen de regels hierover?

Het is een dilemma voor veel accountants. Men denkt vooraf dat de geldsteekproef (of andere substantive test) een foutloze uitkomst geeft maar dat blijkt anders uit te komen. Zijn een paar foutjes niet erg? Of gooien die de boel in de war?

Omdat het Audit Risk Model, of nu het rekenmodel wordt gebruikt of de Bayesiaanse variant, uitgaat van een verwachting dat er in de steekproef geen fouten zullen worden aangetroffen is het vinden van een fout op zijn zachtst gezegd lastig. Wij laten met een voorbeeld een aantal mogelijkheden de revue passeren.

Stel er is een deelpopulatie van € 1.000.000 en bij de controle ervan gaat men uit van een toelaatbare afwijking van € 10.000. In procenten is dus

de toelaatbare afwijking $p(kritisch) = 1\%$. Zonder reductie levert dat een geldsteekproef op van 300 euro's (want bij 95% betrouwbaarheid geldt dat $R = 3$; daaruit volgt de steekproefomvang: $R / p(kritisch) = 3 / 0,01 = 300$).

Risicoanalyse kan dan de volgende uitwerking krijgen.

Als er een laag inherent risico verwacht wordt, dan is dat een 'schone steekproef' van een aantal waarnemingen waard. Vaak past men $R = 2$ toe; dat betekent hier dat er nog 200 gegevensgerichte waarnemingen gedaan moeten worden en 100 schone waarnemingen worden meegeteld als gevolg van het laag zijn van dat inherent risico.

Als de conclusie luidt dat de kwaliteit van de interne beheersing goed is, is dat ook een 'schone steekproef' van een groot aantal waarnemingen waard; bijvoorbeeld 200. Dan moeten er dus nog 100 gegevensgerichte waarnemingen worden onderzocht (en goed bevonden).

En als de toepassing van cijferanalyse niet een negatieve uitkomst geeft, zou die ook nog eens een 'schone steekproef' waard kunnen zijn. Daarvan abstraheren wij hier.

Als de gegevensgerichte steekproef – die met dat berekende ontdekkingsrisico, zeg maar – geen fouten toont mag men als einduitkomst nemen, dat er een steekproef van in dit geval 300 elementen is getrokken die allemaal foutloos waren gebleken. De controle voldoet dan aan de eis dat als de populatie een fout ter grootte van de toelaatbare afwijking – in dit geval 10.000 – bevat, de kans om dat niet te ontdekken ten hoogste 5% bedraagt.

Stel in dit voorbeeld dat er wel een fout gevonden is in de beperkte gegevensgerichte steekproef van 100 gegevensgerichte elementen. Dat is strijdig met het idee van die verwachte schone steekproef. Hoe kunnen we zo'n uitkomst het hoofd bieden?

Het vinden van een fout levert een foutverwachting op. In het voorbeeld: stel dat er 100 gegevensgerichte waarnemingen zijn gedaan en daarin zat een fout. Dan is de foutprojectie 1 keer het interval (dat bij die gegevensgerichte steekproef van 100 is gehanteerd). Tot zover het simpele gedeelte.

Dan het berekenen van de bovengrens. Belangrijke literatuur, Leslie, Teitlebaum en Anderson (1979), geeft aan om met het gehanteerde betrouwbaarheidspercentage, het ontdekkingsrisico dat hoort bij $R = 1$ in dit voorbeeld, te gaan werken. Het ontdekkingsrisico dat bij $R = 1$ hoort is 36,8% (en dus de betrouwbaarheid 63,2%).

Allereerst de vooronderstelling, de premisse, dat het niet gecontroleerde deel foutloos zou zijn: is die stelling te handhaven? Waar moeten we nu

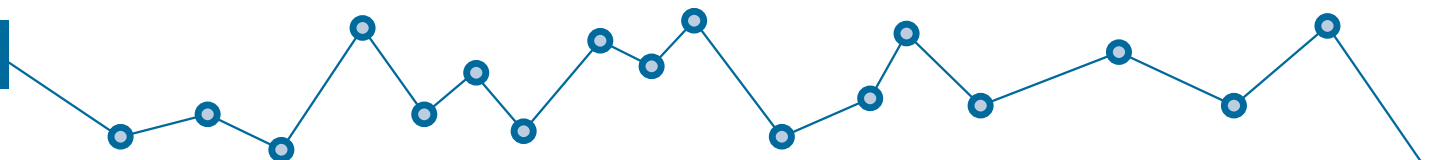
van uitgaan? Dat het aantal waarnemingen dat uit voorinformatie ontstond en de gegevensgerichte steekproef broertjes en zusjes van elkaar zijn? Mag ik dan – in het voorbeeld – verwachten dat die niet gegevensgericht gecontroleerde steekproef twee fouten bevat omdat dit steekproefequivalent twee keer zo groot is als de wel gegevensgericht gecontroleerde steekproef? En hoe bereken ik nu de bovengrens?

De bovengrens die wordt berekend met het ontdekkingsrisico is in het voorbeeld $R(1 \text{ fout})$ bij een risico van 36,8%; dat wordt $2,1458 * J = 2,1458 * 10.000 = 21.458$. We noemen dit optie 1. Verderop zetten we de verschillende opties in een tabel.

Om een vergelijkbare opstelling te maken en dan naar het ontdekkingsrisico van 5% te herrekenen nemen we het verwachte aantal fouten in een steekproef van 300, de niet-gereduceerde steekproef, dus. Dat levert $R(3 \text{ fouten})$ bij een risico van 5%; dat wordt $7,7537 * J = 7,7537 * 3.333 = 25.843$. Men kan ook dit zien als een redelijke schatting van de bovengrens, gegeven de verkregen informatie. We noemen dit optie 2.

Er zijn nog drie mogelijkheden, opties. Die omschrijven we als volgt.

Optie 3: Men controleert 100 elementen gegevensgericht, vindt daarin één fout en telt er 200 correcte elementen bij omdat men blijft steunen op de IB. De evaluatie gebeurt bij 95% betrouwbaarheid. Dit levert een bovengrens op van $4,7439 * 3.333 = 15.812$.



Optie 4: Deze optie is als optie 1, dus men controleert 100 elementen gegevensgericht en evalueert bij 63,2% betrouwbaarheid maar doet alsof dat 95% is omdat men blijft steunen op de IB.

Optie 5: Men controleert alsnog de 200 elementen die werden ‘verdiend’ door te steunen op de IB, omdat men vindt dat het vinden van een fout de nul-foutenpremissie frustreert.

Zie onderstaand schema.

De conclusie luidt in alle vijf de gevallen (bij optie 5 expliciet) dat nog niet goedgekeurd kan worden omdat de uitvoeringsmaterialiteit is overschreden. We zouden nog de projectie kunnen laten corrigeren.

Bij de vier opties in het voorbeeld waar een bovengrens is berekend biedt dat geen soelaas. Er zal dus extra werk moeten worden verricht om beter te kunnen schatten.

Opties	R	Bovengrens	Steunen IB	Projectie
Optie 1	2,1458	21.458	Ja, impliciet	10.000
Optie 2	7,7537	25.843	Nee	10.000
Optie 3	4,7439	15.812	Ja, expliciet	3.333
Optie 4	2,1458	21.458	Ja, expliciet	3.333
Optie 5	N.v.t.	Nog niet bekend	Nee	N.v.t.

Conservatisme en evalueren (deel 4)

Invalshoek literatuur

Hein Kloosterman en Ferry Geertman

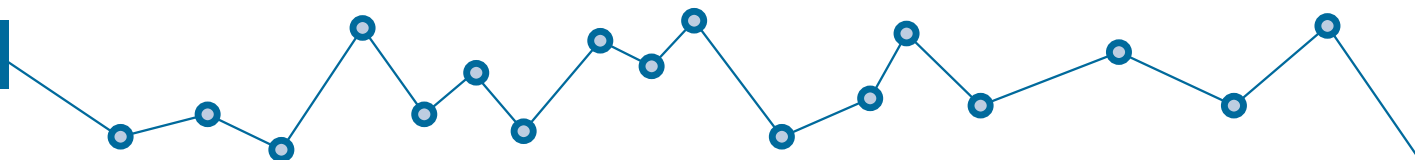
Dit is het vierde deel van een korte serie over het evalueren van steekproeven. Accountants willen wel met behulp van steekproeven controleren, maar de omvang ervan moet dan wel zo klein mogelijk zijn. Tegelijkertijd wil men nog kunnen goedkeuren als die steekproeven een paar foutjes bevatten.

Om te kunnen goedkeuren is het nodig dat de berekende bovengrens voor het foutbedrag in de populatie kleiner is dan de goedkeuringsgrens. Wij nemen even aan dat die goedkeuringsgrens gelijk aan de uitvoeringsmaterialiteit is. Natuurlijk duiken wij in de vakliteratuur om te zien welke middelen ons ter beschikking staan om die bovengrens op een verantwoorde manier zo klein mogelijk te krijgen.

In de literatuur die wij raadpleegden is geprobeerd om met een echt gecontroleerd bestand of met een nagemaakt gecontroleerd bestand te kijken hoe conservatief de verschillende evaluatiemethoden zijn. Met conservatief wordt dan bedoeld dat in minimaal 95 van de 100 getrokken steekproeven de berekende bovengrens met 95% betrouwbaarheid hoger is dan de werkelijk aanwezige fout.

Wat ons daarbij is opgevallen is dat er vooral twee soorten gecontroleerde gegevens (echt of namaak) worden gebruikt. De ene soort betreft gegevens van voorraadcontroles en de andere soort gegevens gaat over de resultaten van het versturen van saldobiljetten.

We noemen als literatuur hier Bringle (1986), Schapendonk-Maas (1998) en Newiak (2009). Bij hun onderzoek maken deze schrijvers gebruik van een populatie die (als het ware) integraal gecontroleerd is. Daaruit trekt men vervolgens een heleboel keren een steekproef. Men past verschillende evaluatiemethoden toe en komt tot de conclusie dat sommige methoden tot een conservatief resultaat leiden, sommige methoden een 'onvoldoende' scoren en dus (op deze populatie althans) niet bruikbaar zijn en nog andere methoden wel conservatief genoeg zijn maar nog leiden tot te grote steekproeven. Dergelijke onderzoeken zijn waardevol omdat de echte fout in de populatie bekend is. Nadeel van deze onderzoeken is, zoals gezegd, dat ze gaan over het onderzoeken van voorraden en uitstaande vorderingen en niet over transactiestromen.



In een ander artikel over steekproeven en accountantscontrole (Panel on Nonstandard Mixtures of Distributions (1988)) wordt de conclusie getrokken dat een boekhoudpopulatie – een transactiestroom, zeg maar – doorgaans nauwelijks fouten bevat. Daardoor zijn er voor het nauwkeurig schatten van de fout in de populatie heel veel waarnemingen nodig. Het artikel adviseert om niet te gaan schatten, maar te gaan toetsen. Met een steekproef die net groot genoeg is om te kunnen goedkeuren is het niet nodig om een schatting van onder- en bovengrenzen te geven, maar is het voldoende om een bovengrens (upper error limit; een worst case) uit te rekenen en die grens te vergelijken met de toelaatbare afwijking voor de betreffende populatie. Deze methode heeft als groot voordeel dat de steekproef zo klein mogelijk is, gegeven de kans op ten onrechte goedkeuren (van een populatie met een fout ter grootte van de toelaatbare afwijking). Bovendien doet op deze manier de kans op ten onrechte niet-goedkeuren er niet meer toe (strikt genomen is die nihil). Uit dit artikel hebben wij afgeleid dat bij veel onderzoek om het conservatisme van verschillende evaluatiemethoden te onderzoeken bestanden zouden zijn gebruikt waarin er sprake is van een geringe foutdichtheid.

In de inleiding op de essays hebben wij verwezen naar het overzichtsartikel van de Stuurgroep Statistical Auditing. Een van de wensen bij de aanvang van de Stuurgroep was het onderzoeken van het conservatisme van de meest gebruikte manier van evalueren, de zogenoemde ‘Stringer-bound’

(de door K.W. Stringer in 1960 geopperde manier om de bovengrens te berekenen). Er is inmiddels theoretisch bewijs geleverd dat de Stringer-bound conservatief is (als het gebruikte betrouwbaarheidspercentage groter is dan 50%). Daarnaast zijn er ook iets meer complexe methoden nog conservatief genoeg zo meldt een onderzoek van Schapendonk (1998) dat gebruikmaakt van een nagemaakt bestand met een geringe foutdichtheid. Dit meldt ook Lucassen et al. (1996). De datasets die Lucassen et al. hebben gebruikt zijn die van Neter en Loebecke (1975). Die bestanden zijn gebaseerd op een administratieve dataset waarin verschillende foutdichtheden zijn ingebracht (van 0,5% tot 70%). Ook nagemaakte bestanden dus. Lucassen et al. (1996) gebruikten voor hun onderzoek de bestanden met een geringe foutdichtheid. In deel 2 van dit essay hebben we de uitkomst van het onderzoek van Lucassen et al. (1996) toegepast; die noemden we de EVA-methode (zelf noemen zij die methode ATO of Amount Tainting Order).

In de inleiding op de essaybundel is een lans gebroken voor het toepassen van verschillende soorten steekproeven, afhankelijk van de aard van de deelpopulatie (de post van de jaarrekening of de verzameling posten). Voorgesteld wordt a priori geldsteekproeven te gebruiken om te toetsen. En wanneer er sprake is van veel fouten een schattingssteekproef te gebruiken van het type postensteekproef (uit de familie van de regressieschatter) om een uitspraak in geld te kunnen doen. Wij zijn dus op zoek naar integraal gecontroleerde

In Bringle (1986) en Newiak (2009) zijn datasets gebruikt met een hoge foutintensiteit. In gevallen waar veel fouten voorkomen zijn berekeningen met behulp van de Stringer-bound conservatief gebleken. Beide bronnen hebben geen regressieschatter gebruikt om een schatting van de fout in de populatie (voorraden) te maken.

Wat je snapt hoef je niet te leren: steekproefevaluatie op zijn allermakkelijkst! (deel 5)

Paul van Batenburg

Tijdens het symposium van de Stuurgroep Statistical Auditing van mei 2017 heb ik een methode gepresenteerd om de evaluatie van steekproeven te versimpelen. Nu weet ik dat er reacties kunnen komen dat de gemiddelde accountant dat toch niet snapt, of dat een bepaald boek met een softwaretool beter werkt dan zelf nadenken. Maar, ik schrijf ook niet voor de gemiddelde accountant, maar voor de accountants die dit onderwerp leuk vinden, en een “fool with a tool is still a fool”.

Mijn methode van evalueren is te onthouden als een rijmpje:
voor elk element van de steekproef is de bijdrage tot de foutprojectie,
gelijk aan het foutbedrag gedeeld door de kans op selectie.

Getallenvoorbeeld 1 (postensteekproef):
Stel dat 1.000 posten samen een boekwaarde van 1.000.000 euro hebben. We trokken een steekproef van 100 waarnemingen. Deze steekproef van 100 waarnemingen bevat 99 correcte uitkomsten en 1 post van € 500 die € 200 had moeten bedragen. Het foutbedrag is dus € 300.

De kans op selectie van deze post is bij een postensteekproef 100 op 1.000 dus 10%, dus de geprojecteerde fout op basis van deze postensteekproef is, omdat dit de enige fout is en alle andere uitkomsten dus een bijdrage 0 leveren, gelijk aan $€ 300/10\% = 3.000$ euro.

Getallenvoorbeeld 2 (geldsteekproef):
Stel dat ook hier de steekproefomvang 100 is (maar dan zijn het geldeenheden). De kans op selectie van een post van € 500 is bij een geldsteekproef van 100 geldeenheden 100 keer 500 op 1.000.000 dus 5%. Dus de geprojecteerde fout is nu $€ 300/5\% = € 6.000$.

Wat een verschil, hè? Maar ja, appels en peren...

Een postensteekproef geeft in deze voorbeelden 10% kans op een uitkomst € 3.000 (en 90% kans op uitkomst 0), een geldsteekproef geeft 5% kans op uitkomst € 6.000 (en 95% kans op 0) dus voor beide is de verwachte uitkomst: exact, € 300, de gevonden fout.

De postensteekproef in symbolen:
Nu in symbolen: T posten hebben samen een boekwaarde van M en de steekproef met omvang

n is bijna helemaal correct: hij bevat 1 post van x die y had moeten zijn. Het foutbedrag is $x - y$.

De kans op selectie van deze post is bij een postensteekproef n op T (n / T). Dus de geprojecteerde fout is, omdat dit de enige fout is en alle andere uitkomsten dus een bijdrage 0 leveren, gelijk aan $(x - y) / (n / T)$ euro. Omdat alle andere waarnemingen correct zijn, kun je dit schrijven als $T * (x - y) / n$ oftewel:

Aantal posten maal gemiddelde fout per steekproefpost bij een postensteekproef

In voorbeeld 1: 1.000 maal 300 gedeeld door 100.

Nu de geldsteekproef:

De totale boekwaarde van de T posten is M. De kans op selectie van deze post van x is bij een geldsteekproef $x * n / M$ (de kans hangt nu immers af van de boekwaarde van de post ten opzichte van de totale boekwaarde) dus de geprojecteerde fout is nu:
 $(x - y) / (x * n / M)$ euro.

Omdat alle andere waarnemingen correct zijn, kun je dit schrijven als:
 $(M / n) * (x - y) / x$.

We herkennen (M / n) als het selectie-interval voor de geldsteekproef, en $(x - y) / x$ als de foutfractie (ook wel taint genoemd). De formule wordt dus:
Interval maal totale foutfractie bij een geldsteekproef

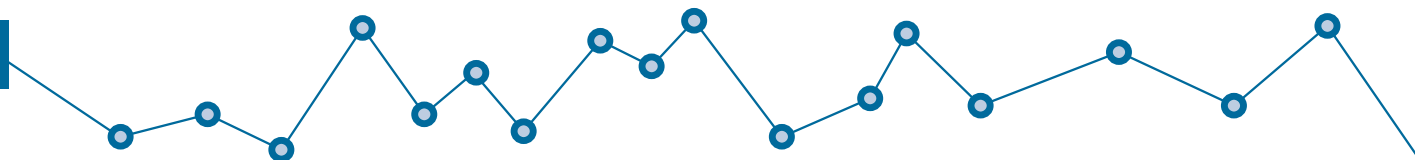
In voorbeeld 2: 10.000 maal € 300 gedeeld door 500 = € 6000.

Tot nu toe was het hopelijk simpel, maar nu komt de aap uit de mouw. Op het verkeerde been gezet door Tabel 5-1 van AICPA (2014), zijn er accountants die deze twee methoden door elkaar halen en de steekproef onjuist evalueren door het totale foutbedrag in de steekproef ($\text{som}(x - y)$) te delen door het totale gecontroleerde bedrag ($\text{som}(x)$) en dan te vermenigvuldigen met de omvang van de te controleren (deel)populatie (M).

In formule: $\text{som}(x - y) / \text{som}(x) * M$ en dat is alleen gelijk aan $T * \text{som}(x - y) / n$ als $\text{som}(x) / n = X / T$, dus als de gemiddelde omvang van een post in de steekproef gelijk is aan de gemiddelde omvang van een post in de populatie.

Bij een postensteekproef mag je daarop hopen, en als de steekproef netjes gestoken is zal dat gemiddeld ook zo zijn. Gemiddeld, maar eigenlijk dus nooit. Deze onjuiste methode geeft dus een onzuivere schatting van de foutprojectie in een postensteekproef.

Bij een geldsteekproef is het ondenkbaar dat de gemiddelde steekproefpost gelijk is aan de gemiddelde omvang van een post in de populatie. De steekproef is immers getrokken met kansen evenredig met de omvang van de posten en de gemiddelde steekproefpost is dus (veel) groter dan de gemiddelde populatiepost. Deze onjuiste methode deelt dus door een te groot gemiddelde



een geeft dus een onderschatting van de geprojecteerde fout!

Nog een versimpeld voorbeeld.

Ik onderstreep deze foute manier van evalueren met nog een getallenvoorbeeld: een populatie van 10.000 posten is samen 1.000.000 waard. Een steekproef van 3 geeft:

10.000 posten samen 1.000.000 euro			
ist	soll	fout	taint
50	42	8	16%
100	105	-5	0%
150	150	0	0%
300	297	3	16%
eva posten	$3/3 \cdot 10.000$		
eva geld	$0,16 \cdot 1.000.000/3$		
eva foutt	$3/300 \cdot 1.000.000$		

Merk op dat de foute evaluatie gelijk is aan de correcte evaluatie van een postensteekproef omdat de gemiddelde omvang van een steekproefpost 10.000 is!

Noot

- Deze artikelen zijn eerder geplaatst als columns Statistical Audit op accountant.nl. Het gaat om de columns 51, 52, 57, 60 en 64.
- Zie hiervoor het essay over Leslie et al. (1979).

Literatuur

AICPA (2014); *Audit Guide Audit Sampling*; AICPA.

Bringle, G.S. (1986); *A study of upper error limits in accounting populations*; Airforce Institute for Technology US (AD-A174 172).

Leslie, D.A., A.D. Teitlebaum, R.J. Anderson (1979); *Dollar unit sampling*; Copp, Clark & Pitman, Toronto.

Lucassen, A., J.J.A. Moors, P.C. van Batenburg (1996); Modifications of the Stringer-bound; A simulation study on the performance of audit sampling evaluation methods; *Kwantitatieve Methoden*, pp. 17-25.

Neter, J., J.K. Loebecke (1975); *Behavior of Major Statistical Estimators in Sampling Accounting Populations: An Empirical Study*. Auditing Research Monograph no. 2; New York: AICPA.

Newiak, M., (2009); *Prüfungsurteile mit Dollar Unit Sampling*; Universität Potsdam.

Panel on Nonstandard Mixtures of Distributions (1988); *Statistical Models and Analysis in Auditing*; National Academy Press; Washington, D.C.

Schapendonk - Maas, H. (1998); *Betrouwbaarheidsbovengrenzen en steekproeftechnieken in accountancy*; stageverslag Katholieke Universiteit Nijmegen.

Goedkeuren, ondanks fouten¹

Dankzij de COS een veel efficiëntere steekproef (deel 1)

Paul van Batenburg

Een vaak gehoord bezwaar tegen het gebruik van steekproeven is dat er te veel waarnemingen nodig zijn om tot goedkeuring te komen als de te controleren jaarrekeningpost niet brandschoon is. Op zichzelf al interessant: de accountant vindt dat hij te veel werk moet doen omdat de gecontroleerde zijn werk niet goed gedaan heeft. Met COS 450 in de hand is daar wat aan te doen.

Een voorbeeld: een jaarrekeningpost vertegenwoordigt een bedrag van 10 miljoen euro. De accountant heeft als materialiteit € 600.000. Bij 95% betrouwbaarheid en 0 toegestane fouten is de steekproef 50 waarnemingen. Daar kan geen data-analyse tegenop, denkt de accountant, maar de klant gooit roet in het eten door te melden dat er foutjes in de gegevens zitten waardoor in totaal een bedrag van € 320.000 fout kan zijn. Eerste reactie van de accountant moet zijn om de klant te vragen die foutjes te zoeken en te corrigeren. Belangrijk argument daarvoor is dat zo wordt voorkomen dat de steekproef fouten oplevert en er niet kan worden goedgekeurd. De klant werpt tegen dat € 320.000 niet materieel is. Moet de accountant dan maar slikken?

Een steekproef die ondanks een vermoedelijke fout van € 320.000 een maximale fout onder € 600.000 op kan leveren, is 221 waarnemingen groot. Daarin kunnen (ruim) 7 fouten nog steeds tot goedkeuring leiden. Op www.steekproeven.eu is een spreadsheet te vinden om dit te berekenen.

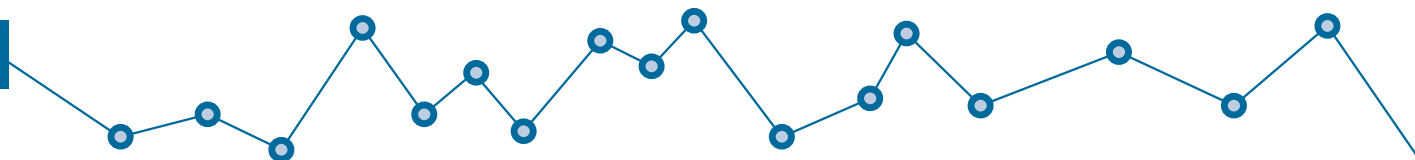
Een screenshot daarvanstaat op de volgende pagina.

De accountant gaat dus nu veel te veel werk doen omdat de klant niet genoeg werk gedaan heeft.

Gelukkig steekt COS 450 paragraaf 8 daar een stokje voor:

De accountant dient alle afwijkingen die tijdens de controle zijn geaccumuleerd, tijdig aan het management op het passende niveau mee te delen, tenzij dit op grond van wet- of regelgeving verboden is. De accountant dient het management te verzoeken deze afwijkingen te corrigeren.

Voor alle zekerheid: COS 450, paragraaf A3 meldt dat de geprojecteerde fout uit een steekproef tot die geaccumuleerde fout behoort. De klant zal dus nu toch moeten corrigeren!



Audit inputs			Smart Audit Sampling Helpfile SMASH			
Population amount		10.000.000	8.1 November 1, 2014			
Materiality for this population		600.000	AICPA Audit Sampling Guide (2014 ed.), appendix C			
Sample design not allowing for overstatements			R factor	required assurance	sample size	interval
Levels of substantive testing		Standard	3,00	95%	50	200.000
		Chosen	3,00	95%	50	200.000
Sample design allowing for overstatements in population						
Presumed overstatement <u>amount in</u>		Expected sample		required assurance		
population ("expected error")		overstatement proportion	R factor	from sample	sample size	interval
320.000		7,043	13,21	95%	221	45.415
Evaluation at chosen confidence level			Guidance			
Final population amount		10.000.000	excluding top stratum and isolated amounts			
Final sample size		50				
Interval		200.000	difference from design value needs to be explained			
Most Likely Error		320.000	MLE			3,20%
Upper Error Limit		1.137.200	UEL			11,37%
Lower Error Limit		41.617	LEL			0,42%
Work sufficient to accept population?		no	UEL compared to materiality			
Work sufficient to estimate overstatement?		no	UEL minus MLE compared to materiality			
Required entity adjustment		537.200	max(MLE, UEL-materiality) if UEL > materiality, else 0			
Precision		817.200	UEL-MLE			8,17%
Sample expansion		25	sample expansion cannot lead to acceptance without increased risk			
New Errors		0.800				
New Precision		598.342	if <materiality, sample is sufficient to estimate MLE			
Sample results for items below high value threshold containing overstatement fractions.						
Decide on their order: high to low is most conservative, random order or ordering by error amount is acceptable, sorting low to high is not.						
Rank	Recorded	Audited	Overstatement Fraction	Most Likely Error	Precision Gap	Upper Error Limit
0					3,00	600.000
1	100,00	0,00	1	200.000	1,75	350.000
2	100,00	40,00	0,6	120.000	0.94	187.200

Als de klant de correctie doorvoert kunnen we dat bedrag van zowel de geprojecteerde als maximale fout aftrekken. De maximale fout is dan slechts

€ 280.000. De accountant kan goedkeuren, maar hij heeft overbodig werk gedaan.

Hoe voorkomt de accountant dat overbodige werk? Door de klant er vooraf op te wijzen dat de geprojecteerde fout gecorrigeerd zal moeten worden. Er zijn dan drie manieren om met aanzienlijk minder werk goed te keuren:

- Als de klant vooraf de boel corrigeert, is een steekproef van 50 voldoende om bij 0 fouten goed te keuren.
- Als de klant dat niet vooraf doet, zal die steekproef van 50 om precies te zijn 1,6 fouten opleveren: de geprojecteerde fout is dan € 320.000 en de maximale fout € 1.137.200. Te hoog om goed te keuren en zelfs te hoog om de geprojecteerde fout te kunnen vaststellen. De onnauwkeurigheid van die schatting is immers € 817.200 terwijl de materialiteit € 600.000 is.
- De accountant kan de steekproef uitbreiden voor een voldoende nauwkeurige schatting: daar zijn slechts 25 extra waarnemingen voor nodig. Ondanks de 0,8 extra fouten (25/50 van 1,6) is de bovengrens bij 75 waarnemingen met 2,4 fouten € 920.000: te veel om goed te keuren, maar precies genoeg om na correctie van de geprojecteerde fout van € 320.000 goed te keuren.
- De accountant had ook, rekenkundig gezien, direct de materialiteit op € 920.000 kunnen zetten, wetend dat de maximale fout toch met € 320.000 verlaagd gaat worden. De steekproef zou dan weer op 75 uitkomen en daarin zouden 2,4 fouten toegestaan zijn:

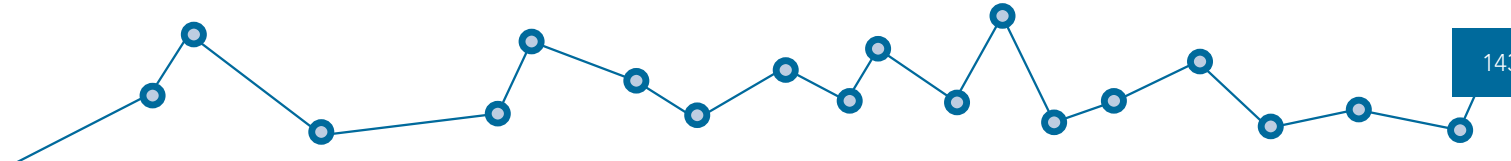
Dit is te zien in de volgende screenshots (pagina 142 en 143).

Evaluation at chosen confidence level			Guidance	
Final population amount	10.000.000		excluding top stratum and isolated amounts	
Final sample size	50			
Interval	200.000		difference from design value needs to be explained	
Most Likely Error	320.000	MLE		3,20%
Upper Error Limit	1.137.200	UEL		11,37%
Lower Error Limit	41.617	LEL		0,42%
Work sufficient to accept population?	no	UEL compared to materiality		
Work sufficient to estimate overstatement?	yes	UEL minus MLE compared to materiality		
Required entity adjustment	320.000	max(MLE, UEL-materiality) if UEL > materiality, else 0		3,20%
Precision	817.200	UEL-MLE		8,17%
Sample expansion	25	sample expansion cannot lead to acceptance without increased risk		
New Errors	0,800			
New Precision	598.342	if <materiality, sample is sufficient to estimate MLE		5,98%

Audit inputs		SMart Audit Sampling Helpfile SMASH			
Population amount		10.000.000	8.1 November 1, 2014		
Materiality for this population		920.000	AICPA Audit Sampling Guide (2014 ed.), appendix C		
Sample design not allowing for overstatements		R factor	required assurance	sample size	interval
Levels of substantive testing	Standard	3,00	95%	33	306.667
	Chosen	3,00	95%	33	306.667
Sample design allowing for overstatements in population					
Presumed overstatement amount in population ("expected error")	Expected sample overstatement proportion	R factor	required assurance from sample	sample size	interval
320.000	2,389	6,88	95%	75	133.778
Evaluation at chosen confidence level		Guidance			

Noot

- 1 De delen 1 en 2 van dit artikel zijn eerder geplaatst als columns Statistical Audit nummer 41 en 42 op accountant.nl.



Heterogeen misschien? (deel 2)

Wouter Gerards en Hein Kloosterman

Wij borduren voort op deel 1 van dit essay (van Paul van Batenburg). We gebruiken dezelfde casus. Nu voegen we als bijzonderheid toe dat er een deelpopulatie is waarin de *gecontroleerde* fouten verwacht.

De casus

Een financieel verslag heeft een omvang in geld van € 10.000.000. De materialiteit voor de verantwoording als geheel is vastgesteld op € 600.000. De gecontroleerde laat weten dat er een hoeveelheid afwijkingen, fouten, verwacht wordt. In geld is de vooraf verwachte afwijking € 320.000. In deel 1 zagen we dat als de projectie van de fout zal worden gecorrigeerd, de accountant in deze casus kan volstaan met hanteren van een steekproefomvang van 75.

De *gecontroleerde* meldt nu dat de fouten zich voordoen in maar een van twee deelpopulaties en dat daarom de populatie niet homogeen, maar heterogeen is.

De *gecontroleerde* beschrijft de deelpopulaties als volgt: er is een deel van € 4.000.000 waarin een fout van 8% wordt verwacht. In het complement,

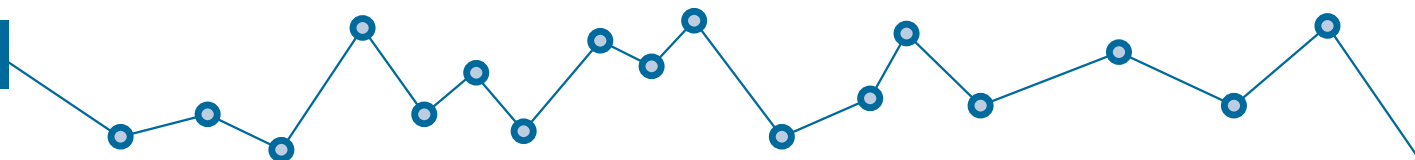
in € 6.000.000 dus, verwacht de *gecontroleerde* geen fouten.

De vraag is hoeveel werk er nu moet worden gedaan om beide deelpopulaties goedgekeurd te krijgen? Kan dat met diezelfde steekproefomvang van 75 of moet er meer worden gecontroleerd? Kortom: wat is het effect van de heterogeniteit ten aanzien van de foutdichtheid?

Kijken we naar de deelpopulaties dan zien we dat de deelpopulatie van € 6.000.000 gecontroleerd kan worden gebaseerd op een toelaatbare fout van € 600.000 en nul fouten in de steekproef. Dat leidt tot een interval van € 200.000 en dus een steekproefomvang van 30.

De steekproef op de deelpopulatie van € 4.000.000 kan, analoog aan wat in deel 1 is gesteld, worden berekend met een materialiteit ter grootte van € 600.000 plus de verwachte fout van € 320.000. Ook hier moet de gevonden foutprojectie worden gecorrigeerd. Dat levert een steekproefomvang op van 30.

Als de controle van de 30 elementen van de deelpopulatie van € 6.000.000 en de 30 elementen van de deelpopulatie van € 4.000.000 afgerond is, is de kous nog niet af. Ook al stemmen de uitkomsten



overeen met de verwachting. Immers er moet ook nog een oordeel komen op de populatie van € 10.000.000 als geheel!

Omdat hier sprake is van heterogeniteit moet berekend worden wat het effect is van het combineren van de uitkomsten van de controles op de twee deelpopulaties. En wel zodanig dat de maximale fout na correctie van de (geaggregeerde) projectie kleiner is dan de materialiteit.

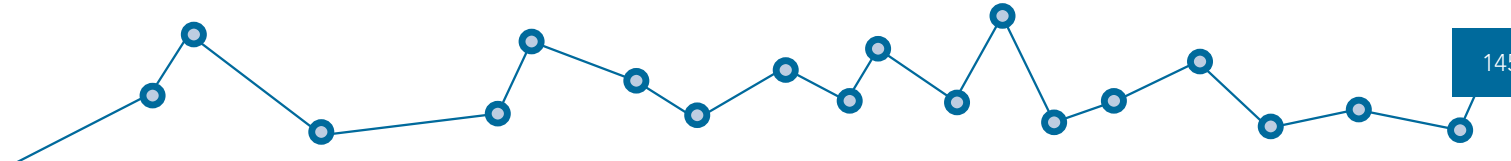
Na enig rekenwerk¹ is de uitkomst:

Deelpopulatie	Omvang steekproef	Taints	Projectie van de fout	Maximale fout
€ 4.000.000	56	4,48	€ 320.000	€ 700.741
€ 6.000.000	30	0	€ 0	€ 599.146

Aggregatie:

Totale populatie	Totaal steekproef	Geaggregeerde projectie	Geaggregeerde maximale fout	MF na correctie
€ 10.000.000	86	€ 476.522	€ 1.075.026	€ 598.504

Als de accountant volstaat met een oordeel over de verantwoording als geheel (de € 10.000.000) dan zagen we in column 41 dat hij kan volstaan met een steekproefomvang op het totaal van 75. Nu de *gecontroleerde* wil dat rekening wordt gehouden met de heterogeniteit van de beide deelpopulaties moet er extra worden gecontroleerd. En wel 11 elementen meer. Die elf elementen extra moeten voor de *gecontroleerde* informatie



toevoegen. Voor de beoordeling van de populatie als geheel voegen zij geen informatie toe.
Wie moet voor de extra kosten opdraaien?
Wij vinden: de *gecontroleerde*!

Zo zien we maar weer dat een verweer leidt tot meer werk. Mogen de lasten daarvan bij de juiste partij terechtkomen.

Noot

- 1 Het rekenwerk is verricht met behulp van de methode voor het aggregeren van uitkomsten zoals Trevor Stewart in 2013 die in zijn dissertatie 'A Bayesian Audit Assurance Model' heeft beschreven.

Data-analyse en statistical auditing¹

Maakt data-analyse steekproeven overbodig? (deel 1)

Paul van Batenburg

Wat is de plaats van steekproeven in een omgeving waar data-analyse wordt uitgevoerd? Die vraag staat centraal in dit deel.

Data-analyse omvat een verscheidenheid aan softwaretoepassingen waarmee we een beeld kunnen krijgen van de kwaliteit van een dataset. Bijvoorbeeld door te toetsen op vooraf gespecificeerde risico's.

Data-analyse kan veel soorten (mogelijke) fouten identificeren. Daarbij zijn verschillende gradaties van foutdefinities te benoemen (*zie onder dit deel*). Maar ook gegevens die voor deze 'tests' slagen, kunnen nog fout zijn. Een datum kan wel op correcte wijze zijn ingevuld, maar hoe weet de controleur of auditor of iemand werkelijk op 1 augustus 1986 in dienst is getreden? Ook bij geconstateerde strijdigheid met gegevens in een ander bestand rijst direct een vervolgvraag: is dat andere bestand correct?

Dit soort controles met behulp van data-analyse verschuiven het controleprobleem dus alleen maar.

Steekproeven doet men alleen als men niet integraal kan controleren. Omdat voor een controle

(onder meer) brongegevens nodig zijn, is dat vaak het geval. Tenzij men heel zeker weet dat die brongegevens honderd procent correct digitaal beschikbaar zijn.

Steekproeven zijn (nog) vaker nodig dan gedacht, omdat vaak ten onrechte wordt aangenomen dat de juiste gegevens zijn opgeslagen. Alles wat niet matcht is fout, maar niet alles wat matcht is per se correct! Een voorbeeld: betalingen toetsen aan betaalopdrachten, zonder dat men zich afvraagt waar die betaalopdrachten vandaan komen

Voordeel van steekproeven is dat men ook kan stuiten op foutsoorten waarvan men niet wist dat ze in een integrale data-analyse moesten worden meegeenomen (de *unknown unknowns* van Rumsfeld), omdat ze niet als risico waren benoemd. Data-analyse kijkt immers alleen naar die zaken waarvan vooraf is vastgesteld dat je ernaar moet kijken.

Andersom kan data-analyse heel efficiënt zijn om fouten op te sporen *voordat* ze in een steekproef tot foutmelding leiden. Niets is zo vervelend als bij een steekproefcontrole een fout constateren die al vooraf met data-analyse had kunnen worden opgespoord.

In de audit worden steekproeven gebruikt om assurance te geven dat een bestand goed (genoeg) is. Anders dan bij data-analyse is niet het zoeken van fouten het doel, maar het op zo'n manier vinden van voldoende correcte waarnemingen dat kan worden aangenomen dat deze ook representatief zijn voor de niet-gecontroleerde gegevens.

Steekproeven waren altijd alleen in bijzondere gevallen nodig. Data-analyse vervangt steekproeven niet en maakt ze niet overbodig, maar maakt ze effectiever en efficiënter. Andersom laten steekproeven zien of de data-analyse volledig is geweest.
Best of both worlds.

Foutdefinities data-analyse

Incorrect

- Velden die leeg zijn en gevuld zouden moeten zijn, kunnen niet correct zijn (een leeg veld waar een geboortedatum verwacht wordt).
- Velden die zijn gevuld met data die niet voldoen aan het vereiste format ('weet niet' bij een geboortedatum, maar ook 8-10-54 waar 08-10-1954 werd gevraagd), kunnen niet correct zijn (al is de juiste waarde soms wel te achterhalen).
- Velden die zijn gevuld met data die niet voldoen aan een vooraf gespecificeerde plausibiliteitseis, kunnen niet correct zijn (geboortedatum 31 februari).

Intern strijdig

- Velden die zijn gevuld met data die niet kunnen kloppen met andere gegevens in hetzelfde bestand (Diagnose Behandel Combinatie voor een man die een gynaecoloog bezoekt): we weten dat er iets fout is maar niet wat er fout is.

Extern strijdig

- Velden die zijn gevuld met data die niet kunnen kloppen met andere gegevens in een ander bestand (kilometerdeclaratie van zieke werknemer): we weten dat er iets fout is maar niet wat er fout is.

Misschien incorrect

- Velden die zijn gevuld met data die (wellicht onterecht) onwaarschijnlijk lijken (geboortedata van personeel vertoont een piek in maart).

Geauzomofo: u bent nog lang niet jarig! (deel 2)

Paul van Batenburg

Op het achtste symposium van de Stuurgroep Statistical Auditing, op 20 mei 2015, heb ik het begrip Geautomatiseerd Zoeken naar Mogelijke Fouten (Geauzomofo) geïntroduceerd.

Binnen de oceaan aan big data analytics zwemt 'GEAUZOMOFO' in het spartelvijvertje van toepassingen die specifiek interessant zijn voor de controle van verantwoordingen. Ik vroeg mij af of een accountant er gegevensgerichte zekerheid aan kan ontlennen.

Wat is Geauzomofo?

Op een bestand met geboortedata van personeelsleden kan men een query loslaten om (mogelijke) fouten af te zonderen (lege velden, 31 april, een datum voor 1900, maar ook 1 januari 1960 etc.) maar men kan niet zonder externe gegevens vaststellen dat een geboortedatum correct is ('Ik weet niet wanneer je wel jarig bent, maar wel wanneer je niet jarig bent'). In een bak met gekleurde ballen die groen zouden moeten zijn, kan ik dus niet die groene eruit vissen, maar wel rode of witte.

Het geautomatiseerd zoeken naar mogelijke fouten is dus een techniek die men kan toepassen bij de controle van een bestand zonder gebruik te maken van een externe bron. Door te definiëren hoe een mogelijke fout eruitziet, toetst men het bestand aan die norm. Het is dus nadrukkelijk geen controle, want bij controle toetst men aan de Soll-positie, de norm voor wat goed is (de 'goed-definitie'), en hier toetst men aan een 'foutdefinitie'.

Nu zal er vast wel een reageerder komen die zegt dat men ook op zoek kan gaan naar een bestand waar de Soll-positie in zit en dan integraal kunt matchen, maar dan hebben we het nu over de controle van dat bestand. Het bestand waarvoor geen Soll-bestand is. De moeder aller bestanden.

Helpt Geauzomofo?

Het grootste voordeel van het zoeken naar fouten is dat men bij controle van het bestand dat overblijft nadat alle uitzonderingsrapportages eruit gehaald zijn, een steekproef kan hanteren waarin men deze fouten niet meer zal vinden en niet ach-

teraf hoeft te isoleren en herstellen. Misschien kan de steekproef dus kleiner, omdat er minder (of geen?) fouten meer worden verwacht. Misschien.

Helpt Geauzomofu echt?

De meest relevante vraag voor deze tak van data-analyse is: als ik met hulp van automatisering fouten uit een bestand heb gehaald, is het bestand dan opgeschoond of zijn er nog meer soorten fouten? Gelooft u dat een vliegtuig veilig is als de fabrikant zegt dat er tijdens de testvlucht zo veel fouten zijn gevonden?

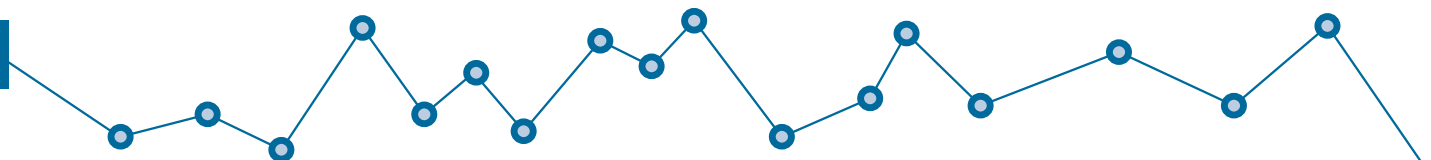
Hoe erachter te komen of Geauzomofu geholpen heeft?

De makkelijkste, snelste, meest transparante manier om te zien of nu ook alle foutoorzaken door een query zijn afgevangen, is een steekproef. Als de omvang goed is gekozen en de resultaten geen fouten bevatten, zijn er met redelijke zekerheid geen materiële fouten meer achtergebleven en is de Geauzomofu geslaagd (en 'volledig').

Wat heb ik nou aan Geauzomofu?

Iets, misschien. Men kan pas vertrouwen aan de Geauzomofu ontlene als die uitzonderingsrapportages geen fouten bevatten. Dat vertrouwen wordt pas assurance als de daarna getrokken

steekproef foutloos is. En die assurance was juist nodig om de steekproef te reduceren. Blijft staan dat het grote voordeel van het geautomatiseerd zoeken naar mogelijke fouten is dat al deze fouten opgespoord en gecorrigeerd kunnen worden. We hoeven dus niet achteraf steekproef-fouten te isoleren, met alle risico van het onderschatten van de mogelijke consequenties van de foutoorzaak. Voor wat betreft de ons bekende foutoorzaken: we blijven zonder die extra steekproef zitten met de vraag hoe erg de ons onbekende foutsoorten zijn.



Data-analyse als heilige graal: dacht het niet (deel 3)

Hein Kloosterman en Ferry Geertman

Kan data-analyse de steekproef vervangen? Wij deden een klein onderzoek op basis van een SVB-rapport. Het resultaat was verbluffend.

Op het Limperg symposium van 20 mei 2015 hebben we weer naar zeer interessante voordrachten geluisterd en gediscussieerd over steekproeven en data-analyse waarin iedere spreker duidelijk vanuit zijn eigen competentie, interesse of misschien wel commercieel belang sprak. Dat geeft niet zolang de discussie fair gevoerd wordt en daar waar mogelijk op feiten is gebaseerd.

Een van de discussies die ook in deze reeks columns gevoerd wordt is de vergelijking van data-analyse met steekproeven en de vraag of data-analyse een vervanger kan zijn van steekproeven.

We doen een klein feitenonderzoek op basis van het rapport van de SVB (2015) inzake de doorlooptijd van betalingen aan zorgverleners in relatie tot Persoonsgebonden Budgetten (PGB). De essentie van het rapport is dat de auditdienst van de SVB de betalingen aan zorgverleners heeft onderzocht op de tijdigheid ervan. De auditdienst heeft daarvoor een steekproef gehanteerd. Naast allerlei evaluaties van de steekproef heeft de

auditdienst ook gekeken naar foutoorzaken die vooraf als bekend werden aangemerkt.

Het rapport geeft op pagina 5 een lijst van mogelijke foutoorzaken. Het is een lange lijst: 39 mogelijke fouten maar liefst! Zo'n lange lijst biedt natuurlijk prachtige handvatten voor het definiëren van queries (en als die wat al te ingewikkeld worden maken we er scripts van) om de gehele populatie te onderzoeken op al die foutsoorten.

De auditdienst had dus met data-analyse de gehele populatie kunnen onderzoeken op deze foutsoorten en ze op een uitzonderingslijst voor nader onderzoek kunnen zetten. De rest van de populatie mag dan verondersteld vrij te zijn van fouten ... of niet?

Het antwoord op deze vraag kunnen we geven omdat de auditdienst ervoor heeft gekozen om een steekproef uit te voeren. In het rapport, om precies te zijn op pagina 14, heeft de auditdienst van 295 betalingen die een doorlooptijd van meer dan tien dagen hadden, beschreven wat de foutoorzaak is. Van zestig procent (179) van deze – te late – betalingen geeft de dienst aan dat de foutoorzaak niet op de lijst van (39) mogelijke foutoorzaken voorkomt.

Dat is nog niet alles: in bijlage V (pagina 22) staat dat van de te laat betaalde declaraties die na 1 april zijn ingediend tachtig procent (!) een andere foutoorzaak heeft dan gedefinieerd op de lijst van 39. En van de te laat betaalde declaraties die na 15 april zijn binnengekomen heeft zelfs negentig procent (!!) een andere foutoorzaak dan gedefinieerd op de lijst van 39.

Wij zijn benieuwd wat de gepresenteerde feiten zouden zijn geweest als alleen data-analyse was uitgevoerd op de mogelijke foutoorzaken. Had de ontvanger van het rapport dan wel een conclusie kunnen trekken?

Gelukkig dat de auditdienst zich niet heeft laten verleiden tot (alleen) geautomatiseerd zoeken naar mogelijke fouten ('geauzomofo' zoals Van Batenburg dat noemt)!

Zo'n steekproef biedt uitkomst(en).

Data-analyse maakt het onderscheid (deel 4)

Hein Kloosterman en Ferry Geertman

We schreven al eerder over het valideren van subjectieve risico-inschattingen van accountants. In dit deel geven we voorbeelden van modellen waarmee een deel van de risicoanalyses kan worden geobjectiveerd door gebruik te maken van data-analyse en statistiek.

Wij vinden steekproeven een uitermate krachtig hulpmiddel om heel snel (verhoogt de mate van efficiency, dus) een kwantificeerbare uitspraak (verhoogt de mate van effectiviteit, dus) te doen over een verzameling gegevens. Dat geldt voor de gegevensgerichte controles en het geldt ook in de fase van de analyse van specifieke risico's. Maar er is meer onder de zon dan steekproeven alleen.

Een brainstorm over ontwikkelingen in het accountantsberoep en de rol van statistiek, data en technologie leidde tot de stelling dat het met statistiek mogelijk is een voorspelling te doen over het al dan niet fout zijn van een jaarrekening. Daarvoor is dan data nodig en een aantal technologische voorzieningen.

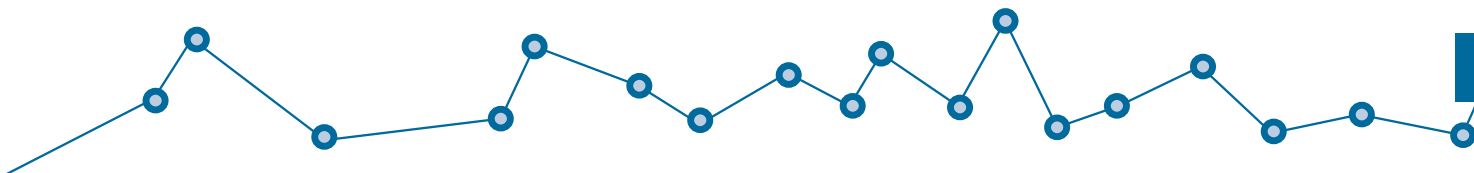
De gedachten gingen uit naar discriminantanalyse. Dit is een tak van sport binnen de statistiek waarbij een waarneming op basis van de waar-

genomen kenmerken wordt toegedeeld aan een bepaalde klasse of groep.

Een simpel voorbeeld om te illustreren hoe discriminantanalyse werkt, is de reclame voor de landmacht. Je ziet een persoon een aantal activiteiten uitvoeren en er verschijnt: 'Geschikt' of 'Ongeschikt'.

Wat hier vanuit de statistiek aan vooraf kan zijn gegaan, is het volgende: van een groep geschikte en ongeschikte kandidaten is een groot aantal gegevens verzameld. Deze gegevens zijn geanalyseerd om te bepalen welke soort gegevens echt (volgens de regels van de statistiek) onderscheidend zijn voor verschillende groepen. Deze gegevens worden discriminanten genoemd.

Een verdere uitwerking van het voorbeeld: van geschikte kandidaten is vastgesteld dat ze gemiddeld een 7,5 op het vak wiskunde hadden en de honderd meter binnen tachtig seconden zwemmend kunnen overbruggen. Ongeschikte kandidaten daarentegen hadden gemiddeld een 6,0 en deden over de honderd meter gemiddeld 95 seconden.



Van een nieuwe kandidaat, waarvan dus nog niet bekend is of hij geschikt dan wel ongeschikt is, worden de discriminanten – dat wiskundecijfer en de zwemtijd – gemeten en wordt op basis daarvan voorspeld of de kandidaat geschikt of ongeschikt is.

Omdat de discriminanten met behulp van statistiek worden bepaald, kennen ze een foutmarge. Die marge noemen we de misclassificatiekansen; men wijst een kandidaat af terwijl die eigenlijk wel geschikt is dan wel men classificeert een kandidaat als geschikt, terwijl die uiteindelijk ongeschikt blijkt. Het gaat te ver om dat hier in detail uit te leggen, maar hoe sterker (= hoe meer onderscheidend) de discriminanten, hoe kleiner deze misclassificatiekansen zijn en dus hoe sterker het voorspelmodel is.

Terug naar ons vak. Zouden we nu ook een model kunnen bouwen dat op basis van kenmerken van het object van onderzoek, de jaarrekening, deze classificeert als goed of fout? Dit zou dan een uitermate sterk stuk gereedschap zijn in de fase van analyse van de specifieke risico's van een audit.

Nu is 'goed of fout' misschien nog wel een stap te ver, maar een model dat objectief kan aangeven of er sprake is van een meer dan normaal risico, is al een hele stap voorwaarts.

De volgende voorbeelden van modellen geven al een aardig inzicht in de mogelijkheden. Schrikt u niet van de formules (die kunnen in Excel worden gezet). De inputgegevens bestaan simpelweg

uit waarden die in de jaarrekening voorkomen. Deze modellen maken gebruik van de hierboven omschreven discriminantanalyse.

1. Z-score van Altman

Een model dat bankroet voorspelt is de zogenoemde Z-score van Altman:

$Z = 1.2 \cdot X_1 + 1.4 \cdot X_2 + 3.3 \cdot X_3 + 0.6 \cdot X_4 + .999 \cdot X_5$,
waarbij geldt:

$X_1 = \text{Working Capital} / \text{Total Assets}$

$X_2 = \text{Retained Earnings} / \text{Total Assets}$

$X_3 = \text{Earnings Before Interest and Taxes} / \text{Total Assets}$

$X_4 = \text{Market Value of Equity} / \text{Total Liabilities}$

$X_5 = \text{Sales} / \text{Total Assets}$

De onderscheidende (discriminerende) waarden zijn:

$Z > 2.99$ - 'Safe' Zones

$1.81 < Z < 2.99$ - 'Gray' Zones

$Z < 1.81$ - 'Distress' Zones

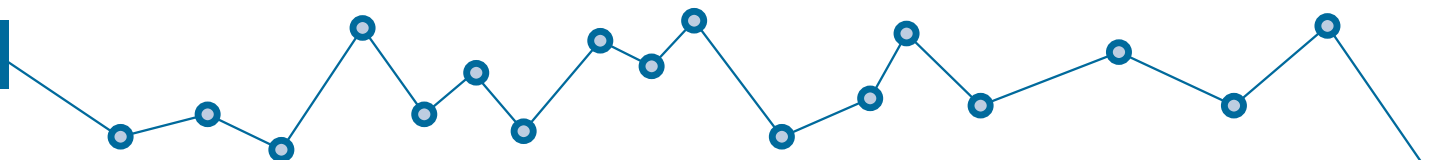
2. Model van Messod Beneish (1999)

Dit is een model uit 1999 waarbij gerommeld met de cijfers kan worden voorspeld, ontwikkeld door professor Messod Beneish. Het model gaat op zoek naar een M-score.

$M = -4.84 + 0.92 \cdot \text{DSRI} + 0.528 \cdot \text{GMI} + 0.404 \cdot \text{AQI} + 0.892 \cdot \text{SGI} + 0.115 \cdot \text{DEPI} - 0.172 \cdot \text{SGAI} + 4.679 \cdot \text{TATA} - 0.327 \cdot \text{LVGI}$

Waarbij geldt:

- Days Sales in Receivables Index (DSRI).
- Depreciation Index (DEPI)
- Sales Growth Index (SGI)
- Leverage Index (LVGI)



- Total Accruals to Total Assets (TATA)
- Gross Margin Index (GMI)
- Asset Quality Index (AQI)
- Sales, General and Administrative Expenses Index (SGAI)

Een M-score groter dan -2.22 is een aanwijzing (onderscheid) voor gerommel met de cijfers.

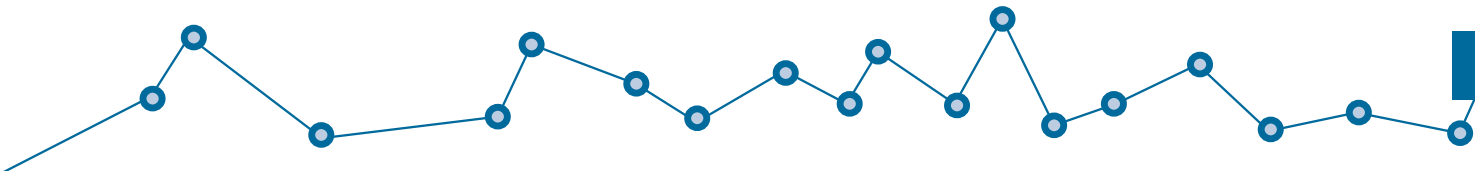
Ok-score Okkerse

Een ander voorbeeld is de OK-score van Okkerse. We hebben daarvan geen toelichting gevonden zoals van de twee voornoemde modellen.

Ongetwijfeld zijn er zijn nog meer van deze discriminantanalyses.

De elegantie van dergelijke modellen is dat ze een uitspraak doen over een jaarrekening als geheel. Men kan dan wellicht in een vroeg stadium eventueel een uitgebreid niveau van gegevensgerichte controle of misschien zelfs een zoeklichtcontrole met betrekking tot de alarmeringen aankaarten bij het auditcomité of het bestuur van de organisatie.

Er is vast ook een heleboel aan te merken op deze modellen. Wij zijn benieuwd of deze modellen en de achterliggende methodieken van toegevoegde waarde kunnen zijn in de discussie over risico-analyse in de accountantscontrole. Wat kan men met deze methoden en technieken objectiveren?



Integrale controle: als het kan, dan moet het? (deel 5)

Paul van Batenburg

Sommige vormen van data-analyse maken het mogelijk dat een accountant van nu integraal controleert waar vroeger een steekproef werd gebruikt. Voorwaarde is wel dat getoetst kan worden aan de normen die aangeven wat de juiste waarden van een te controleren element zijn.

Veel data-analyses zijn echter anders ingericht: zie hiervoor mijn artikel over het geautomatiseerd zoeken naar mogelijke fouten.

Bij gegevensgerichte controle kan een voorraadbestand bij de klant worden opgevraagd en integraal worden nagegaan of per artikel, per locatie, prijs maal hoeveelheid gelijk is aan voorraadwaarde. Toen dat voorraadbestand nog als voorraadjijst op papier werd afgestaan, werd die controle waarschijnlijk vaak steekproefsgewijs gedaan. Bij een proceduretest op inkopen werd vroeger een steekproef van betalingen vergeleken met ontvangsten en bestellingen. Nu kan dat integraal met de zogenoemde 3way-match.

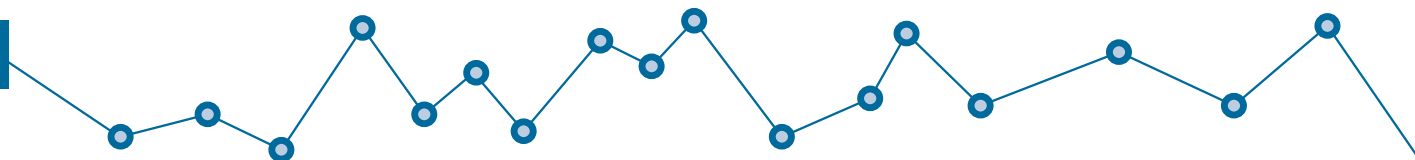
Voordeel van zo'n integrale controle is dat alle dergelijke fouten worden gevonden en kunnen worden gecorrigeerd. Een steekproef had altijd een kans (meestal vijf procent) dat een onaanvaardbaar foutenpercentage niet zou worden opgemerkt. Dat

betekent dat de accountant door het toepassen van data-analyse strenger is gaan controleren.

Die strengere controle kan tot een vreemd dilemma leiden: de accountant kan stellen dat er weliswaar niet kan worden goedgekeurd maar dat bij het controleren met behulp van een steekproef dat wel zou zijn gebeurd. Dit geldt vooral voor een proceduretest: de standaard omvang van 25 levert namelijk met vijf procent kans uit een populatie met twaalf procent (procedure-)fouten een foutloze steekproef op.

Eerste reactie op dit dilemma is dat data-analyse niet alleen is gepositioneerd om gegeven de auditkosten effectiever te controleren; wat heet: zelfs tegen lagere kosten! Dat is beter controleren. En beter controleren willen we toch allemaal? En, beter gecontroleerd worden toch ook?

Tweede reactie is dat de NV COS misschien kan worden aangescherpt om dit dilemma te voorkomen. NV COS 500 A53 geeft aan wanneer integrale controle mogelijk is en wellicht geschikt is. Maar waarom wordt er dan niet ook gemeld wanneer integrale controle moet? Zo voorkomt NV COS wellicht een ethisch dilemma!



Ik kan mij voorstellen dat aan NV COS 500 een zinsnede wordt toegevoegd als:
'Wanneer integrale controle technisch mogelijk is (doordat de accountant de beschikking heeft over zowel het te controleren bestand als het bestand met de normen voor die vastleggingen in elektronische vorm) dan verdient deze vorm van controle de voorkeur, waarbij de accountant speciale aandacht dient te geven aan de kwaliteit van het bestand met de te gebruiken normen.'

En dan zou ergens in COS 530 kunnen worden opgemerkt:
'Steekproefcontrole is een toelaatbare methode wanneer integrale controle technisch niet mogelijk of economisch niet haalbaar is. Dit is bijvoorbeeld als de normen voor de controle niet in elektronisch formaat beschikbaar zijn.'

Literatuur

Beneish, Messod D. (1999); *The Detection of Earnings Manipulation*; citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.3676.

NBA (2017); *Handreiking Regelgeving Accountants*; NV COS (nadere voorschriften controle en overige standaarden).

SVB (2015); *RAPPORT VAN FEITELIJKE BEVINDINGEN, Vervolgopdracht specifiek overeengekomen werkzaamheden inzake doorlooptijd uitbetalingen declaraties inzake Trekkingsrechten PGB*; www.rijksoverheid.nl/documenten/rapporten/2015/05/20/rapport-van-feitelijke-bevindingen-doorlooptijd-uitbetalingen-trekkingsrechten-pgb.pdf

Wikipedia (2017); *Z-Score (financieel)*; [nl.wikipedia.org/wiki/Z-score_\(financieel\)](http://nl.wikipedia.org/wiki/Z-score_(financieel)).

Noot

- 1 Dit essay is een samenvoeging van de columns 2, 45, 47, 49 en 55 Statistical Audit die zijn verschenen op Accountant.nl.

Fraude en statistical audit¹

Hein Kloosterman en Ferry Geertman

De NV COS geeft in richtlijn 240 een definitie van fraude die korthedshalve kan worden samengevat als ‘een opzettelijk handelen of nalaten van aan de organisatie verbonden personen om met misleiding een onrechtmatig of onwettig voordeel te verkrijgen’.

Vanuit het vakgebied Statistical Audit kunnen we vanuit twee verschillende controlemethodieken redeneren. De eerste is vanuit de gegevensgerichte detailcontrole, de tweede invalshoek is data-analyse. Beide invalshoeken kennen inmiddels een stevige modelbouw.

Gegevensgerichte detailcontrole

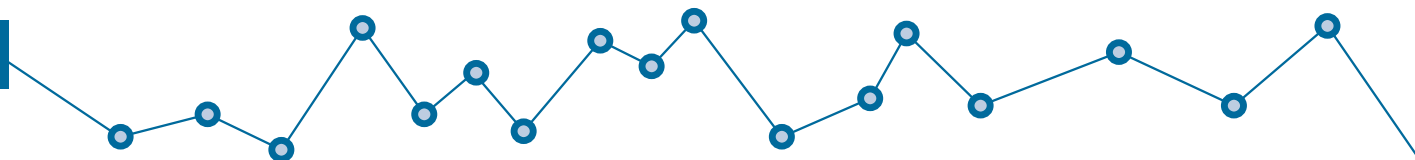
Gegevensgerichte detailcontrole ‘normaal’

Het model van de gegevensgerichte detailcontrole is vormgegeven met behulp van steekproeftheorie, en wel die van de geldsteekproeven (*monetary unit sampling*). De hoeveelheid bewijs die moet worden geleverd is gebaseerd op de parameters materialiteit (in geld) en het auditrisico (de kans) op ten onrechte goedkeuren.

Beide parameters hangen samen: als een verantwoording een hoeveelheid fouten ter grootte van het bedrag van de materialiteit bevat, mag het

risico van het goedkeuren van die verantwoording ten hoogste gelijk zijn aan het auditrisico. De aanpak van het onderzoek komt vervolgens neer op het verzamelen van zoveel bewijs – bewezen goede geldeenheden – dat de stelling dat de verantwoording **niet** kan worden goedgekeurd, kan worden verworpen.

Alle bewijs – dus dat betreffende alle steekproefelementen of hun als valide aangemerkte steekproefequivalenten – moet dus het goed zijn van de verantwoording onderschrijven. Anders gezegd: alle onderzochte gegevens (de steekproefelementen) van het grootboek (waarin immers de Ist-posities zich bevinden) moeten kloppen met hun brongegevens (de Soll-posities die in allerlei bronnen teruggevonden moeten worden, zoals dossiers, documenten en ook primaire vastleggingen ondersteund met bewijs omtrent de effectieve werking van de onvervangbare interne controle). Van iedere boeking wordt dus onderzocht of die goed is. Als hij goed is, draagt die boeking bij aan het mogelijk verwerpen van de stelling dat de verantwoording (nog) niet kan worden goedgekeurd.



Detailcontrole en fraudeonderzoek

Als bij de detailcontrole blijkt dat een boeking (Ist) niet overeenkomt met de brongegevens (Soll) hoeft er nog niet sprake te zijn van fraude. Er is sprake van een afwijking, een fout. Daarvan hoeft niet voorafgaand aan de controle bedacht te zijn aan welke kenmerken zo'n afwijking of fout voldoet. Wel moet worden bedacht wanneer en waarom een boeking 'goed' is. Als een boeking niet 'goed' is, kan het dus gaan om het ontdekken van zogenoemde *unknown errors*.

Als zich een of meer fouten of afwijkingen voordoen in een steekproef is het goed gebruik om aan foutanalyse te doen. Niet om de fouten weg te redeneren, daar doen we niet aan. Foutanalyse moet een indicatie leveren van de aard en de omvang van de verschillende soorten fouten in de populatie. De steekproef zoals die getrokken is om te gaan goedkeuren levert doorgaans te weinig informatie om zo'n indicatie te geven. Daarom is vaak een aanvullende steekproef wenselijk. Niet om alsnog te gaan goedkeuren, maar om een nauwkeuriger schatting van het totale bedrag aan fouten te krijgen. Een foutanalyse kan ook voor een deel worden uitgevoerd met data-analyse, daarover meer in een volgende paragraaf. Voorbeelden van boekingen die het stempel 'goed' niet kunnen krijgen zijn een teruggedraaide boeking terwijl de oorspronkelijke boeking 'goed' was, een boeking gedaan door een medewerker waarvan dit niet verwacht wordt, een boeking zonder omschrijving of juist met bepaalde woorden, een boeking die onverwacht op een tussenrekening terecht komt.

In het kader van fraudeonderzoek moeten dergelijke boekingen dan in detail worden onderzocht om de vraag te beantwoorden of er sprake is van benadeling en of er opzet in het spel is. Verder moet nog worden geanalyseerd of er meer van dergelijke boekingen zijn.

Voor de controle van de jaarrekening is het van belang om het maximale effect van de gevonden fouten te leren kennen; de bovengrens mag immers de materialiteit niet overschrijden. De relatie van de detailcontrole met fraudeonderzoek is dat gevonden fouten fraude kunnen betekenen. Ieder geval van fraude is een fout, maar niet iedere fout heeft te maken met fraude. Dat verdient nog enige toelichting.

'Normale' controle van de jaarrekening

Als bij de controle van de jaarrekening fouten aan het licht komen, en een zekere mate van fraudeonderzoek zou tot die jaarrekeningcontrole moeten behoren, dan ligt het voor de hand om de foute boekingen nader te onderzoeken.

Men moet zich ervan bewust zijn dat bij een goedkeuringsonderzoek zoals de jaarrekeningcontrole, de elementen (geldeenheden) die niet goedgekeurd kunnen worden het stempel 'goed' niet kunnen krijgen. Anders gezegd, er kan sprake zijn van twijfel óf van zeker weten dat de geldeenheden, en de boeking die erbij hoort, fout is.

Vervolgens moet van de niet-goede boeking worden vastgesteld dat die met opzet fout geboekt is. En nog wat zaken die met de benadeling van de gecontroleerde of de stakeholders te maken hebben. Men zou kunnen zeggen dat er een 180

graden andere blik moet worden geworpen op de relatie van de gecontroleerde niet-goede boeking met de bijbehorende elementen van de Soll-positie. Wat zich ook nog weleens voordoet bij het vinden van een fout die als fraude moet worden bestempeld, is dat de bovengrens van die fout opeens onbepaald groot wordt. Populair gezegd: zo'n fout kan een beerput opentrekken. Het statistisch berekenen van een bovengrens hoeft dan niet voldoende te zijn om te weten of de mogelijke (en dus maximale) fout de materialiteit overschrijdt. De beerput kan de fout dan tot ongeken- de hoogten brengen, zullen we maar zeggen.

Conclusie met betrekking tot fraude en gegevensgerichte controle

Onze conclusie is dat bij een 'normale' gegevensgerichte detailcontrole fouten aan het licht kunnen komen die te maken hebben met fraude. Hét voordeel van statistische steekproeven is dat zowel verwachte als niet-verwachte fouten aan het licht kunnen komen en zorgen voor een solide basis voor identificatie van risicogebieden, en dus leiden tot een betere prioriteitsstelling.

De accountant moet zich ervan bewust zijn dat als de kwalificatie 'goed' aan een onderzochte geldeenheid moet worden onthouden, er sprake kan zijn van fraude. Dat vereist een extra en anders gericht onderzoek. Dat op zijn beurt betekent extra werk ten opzichte van 'alleen maar' goedkeuren.

Data-analyse

Soorten data-analyse

Zoals al eerder in deze columns geschreven, heeft data-analyse meestal een ander onderzoeksdoel dan een gegevensgerichte detailcontrole. Er zijn grofweg drie soorten data-analyse:

- Data-analyse is bij uitstek geschikt om *consistentiecontroles* te verrichten op complexe verzamelingen aan elkaar gerelateerde gegevens. Zoals controles op de samenhang van de (registraties van de) waardenkringloop: heeft bijvoorbeeld iedere aflevering geleid tot een omzetboeking, en heeft die weer geleid tot de uitgifte van een factuur.
- Data-analyse wordt ook vaak gebruikt om *uitzonderingen op een verwachte samenhang* te toetsen. Het gaat dan vaak om zwakkere verbanden dan die van de waardenkringloop. De verwachte samenhang kan met bijvoorbeeld regressieanalyse tot uiting worden gebracht. Bijvoorbeeld uit oktober van het vorige jaar volgt een verwachting van de directe kosten van oktober van het huidige jaar, waarvan met behulp van statistiek een betrouwbaarheidsinterval is berekend.
- Een derde vorm van data-analyse is een methode waarbij het de bedoeling is om *vermoedens uit data 'los te peuteren'* (bijvoorbeeld het aantal en het tijdstip van wijzigingen van bankrekeningnummers en betaalmomenten). Deze laatste categorie wordt ook wel eufemistisch aangeduid als onderzoek naar bekende risico's.

Samengevat komen deze drie vormen van data-analyse neer op het volgende. *Data-analyse als consistentiecontrole* geeft bij een positieve uitkomst bewijs dat de waarden in de formele informatiesystemen onderling kloppen.

Data-analyse als toets of verzamelingen boekingen overeenstemmen met vergelijkbare en gecontroleerde verzamelingen boekingen is te zien als het vergelijken van een Ist-positie met een door die data-analyse gegenereerde Soll-positie.

Data-analyse om vermoedens uit data 'los te peuteren' probeert anomalieën boven water te krijgen, te onderzoeken of er ergens mogelijk iets fout zit. Onze ervaring is dat juist die analyses vaak leiden tot veel zogenoemde 'vals positieve' uitkomsten: het lijkt fout, maar is het niet.

Data-analyse en fraudeonderzoek

Als data-analyse om de consistentie vast te stellen (consistentie)fouten aan het licht brengt kan dat wijzen op fraude. Of en in welke mate daarvan sprake is vereist nader onderzoek. Voor zover er sprake is van positieve controles van de fouten kan de accountant detailcontroles met behulp van steekproeven toepassen. Als er sprake is van negatieve controles zijn er vaak andere technieken nodig zoals onderzoek bij derden. Bij zowel de positieve als de negatieve controle is het etiket 'fraude' niet zonder nader onderzoek op de verzamelde potentiële fouten te plakken.

Als data-analyse wordt toegepast om met regressieanalyse een Soll-positie te genereren voor een verzameling boekingen kan deze toets fouten aan

het licht brengen. Het wordt een beetje saai: deze techniek geeft weinig inzicht in de omvang en de oorzaken van de individuele fouten. Ook voor de conclusie of er sprake is van fraude is nader onderzoek nodig. Alweer is een beroep nodig op detailcontroles. Integraal of met behulp van steekproeven.

Bij data-analyse voor het vergaren van vermoedens is lastig aan te geven wat de vervolgactie moet zijn. Enerzijds doordat er veel vals positieve uitkomsten zijn en anderzijds doordat de verzameling potentiële fouten vaak vrij groot is. Alweer geldt dat voor de vervolgactie detailcontroles kunnen worden ingezet. Eventueel met behulp van steekproeven.

Altijd nader onderzoek

Net als bij de detailcontroles moet men naast de in de normale jaarrekeningcontrole toe te passen data-analyses altijd nog nader onderzoek verrichten om het eventuele fraudekarakter aan te tonen van de als fouten aangemerkte verzamelingen boekingen. Dat kan alleen door de individuele fouten intensief te beoordelen.

Data-analyse geeft weliswaar hints in de richting van fraude. Maar voor het vaststellen van fraude is detailonderzoek nodig.

Focus op fraudeonderzoek

Bij een onderzoek naar fraude spelen verschillende elementen een rol. Zoals hoe men een fraude kan ontdekken, hoe het effect van een fraude

gekwantificeerd moet worden en hoe men een schuldige aan kan wijzen. Wij behandelen die drie elementen. Naast het achteraf controleren, belichten we tot slot de mogelijkheid van controle in het heden.

Men kan om fraude proberen te ontdekken verschillende invalshoeken kiezen. De ene is het zoeken naar kenmerken die een indicatie van fraude kunnen geven, zogenoemde fraudesignalen of -kenmerken. De andere is 'gewoon controleren'², en daarbij de beoordeelde vastleggingen of gedragingen onderzoeken op potentiële fraude. Bij beide manieren moet bekend worden of een fout of omissie opzettelijk is geweest en niet door de beugel kan. Wij zoomen in op onderzoek op basis van fraudesignalen.

Ontdekken met behulp van fraudesignalen

Als de methode van controleren uitgaat van het zoeken op basis van bekende fraudesignalen lijkt zo'n methode sprekend op GEAZOMOFO (geautomatiseerd zoeken naar mogelijke fouten³). Zodra er sprake is van foutsignalen is het slim om die op te pakken en met behulp van programmatuur voor data-analyse te gaan kijken of die signalen ertoe leiden dat er fraude boven water komt. Je zou kunnen zeggen dat fraudesignalen de weg kunnen openen naar het ontdekken van fraude. Voor de volledigheid: zowel data-analyse als het controleren door middel van steekproeven kunnen zelfstandig leiden tot het ontdekken van fraudesignalen. En die kunnen dan weer leiden tot nader onderzoek en dan tot uitgewerkte conclusies.

Ontdekken door het beoordelen van organisatie

Een willekeurige controleaanpak kent naast substantive tests (gegevensgerichte detailcontroles) ook het beoordelen van de interne organisatie, ook wel de interne beheersing. Dit deel van de controle omvat het beoordelen van de volledigheid van de opbrengstverantwoording⁴. Dit is een voorwaardelijke controle om een uitspraak op basis van de substantive tests te kunnen doen. Het is al een hele klus om vast te stellen dat de beschikbare vastleggingen van het bedrijfsgebeuren onderling consistent zijn. Het is immers mogelijk dat de organisatie niet steeds heeft gewerkt zoals zou moeten. Bij een 'gewone controle' kan een deel van de controle-informatie omtrent de werking voortkomen uit data-analyse, process mining bijvoorbeeld. Die controle-informatie kan zijn dat de data laat zien dat de werking van de interne beheersing goed was. Het kan ook zijn dat blijkt dat de werking van de interne beheersing niet goed was. Die controle-informatie kan dan leiden tot de stelling dat er sprake kan zijn van fraude. De conclusie als de data geen fouten laat zien kan dan zijn dat die vastleggingen geen aanleiding geven tot een vermoeden van fraude.

Berekenen, schatten effect ontdekte fraude

Als er blijkt sprake te zijn van fraude dan moet er gekwantificeerd worden. Dat zou kunnen met behulp van statistische methoden en technieken. Bijvoorbeeld wanneer er een wiskundige steekproef is getrokken en daar blijken fouten in voor te komen die het etiket fraude verdienen. Om

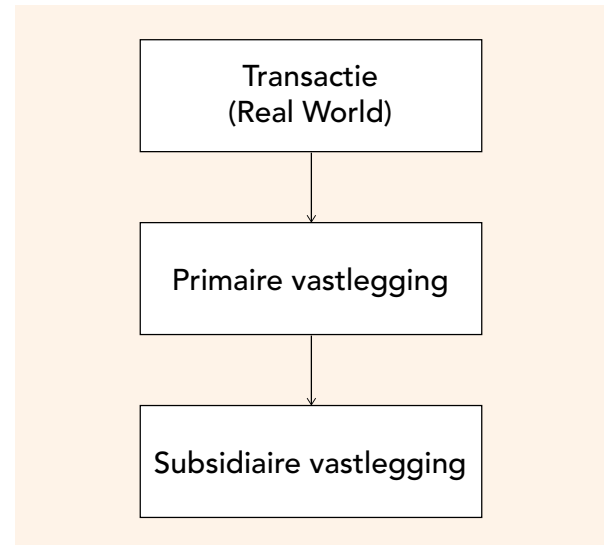
te kwantificeren moet het te gebruiken beslissingsmodel worden toegesneden op degene die beslist, zoals over eventuele vervolging. Stel dat de beslisser een Officier van Justitie is, we kunnen ons dan voorstellen dat hij zou willen weten wat het minimale nadeel is dat kan zijn berokkend. En ook wat het maximale nadeel is. Daarnaast, en dat geldt zeker als er ook nog civiele procedures kunnen volgen, zou hij willen weten wat de foutprojectie is. Om een minimale en een maximale fout te berekenen is het weer nodig de maximale onbetrouwbaarheid te kennen die een Officier van Justitie zou willen hanteren.

Aanwijzen schuldige(n)

Vaak volgt uit de oorzaak van een fraude de veroorzaker. Het ontdekken hoe de fraude is gepleegd, is dus een vingerwijzing naar wie de schuldige is of de schuldigen zijn. De meest lastige variant is als de schuldigen onderling hebben samengespannen. Als door dat samenspannen het net lijkt alsof alle interne beheersingsmaatregelen hebben gewerkt krijgt de onderzoeker het moeilijk om te bewijzen wie de schuldigen zijn. Fraude houdt in dat de veroorzakers er voordeel bij hadden. Het kunnen volgen van de geldstroom is dan idealiter de wegwijzer naar de fraudeur. Veel beschrijvingen van fraudegevallen laten zien dat als er sprake is van samenspanning, er tegelijkertijd is geprobeerd in een administratie geen sporen ervan na te laten. Het loont dan de moeite om te overwegen waarnemingen in de actualiteit te verrichten.

Er bovenop zitten: transactiemodel

Het transactiemodel (zie figuur) laat zien hoe de transacties in de echte wereld leiden tot gegevens, primaire vastleggingen, en die gegevens weer tot (beoogde) informatie, subsidiaire vastleggingen.



De tekening laat zien dat als er transacties in de werkelijke wereld zijn die niet leiden tot een primaire vastlegging, zij ook niet tot een subsidiaire vastlegging zullen leiden. De bestanden (of andere vastleggingen) van de klant zullen dan niet de mogelijkheid bieden een steekproef te trekken of een data-analyse uit te voeren. Er bovenop zitten betekent dan: realtime waarnemen, in de werkelijke wereld dus. Dat is over de schouder meekijken of woorden van gelijke strekking. Dus als er een vermoeden van fraude is en de vastleggingen

lijken geen houvast te bieden, dan zijn directe waarnemingen of een equivalent ervan (bijvoorbeeld videocamera's) in het gebied waar de fraude zou kunnen plaatsvinden een uitkomst.

All together now

Om het beste uit alle verschillende werelden te halen kunnen de hiervoor genoemde technieken worden gecombineerd. Dat betekent als er geen signalen van fraude zijn kan een gewoon onderzoek plaatsvinden, waarbij de beoordeling van de organisatie impliceert dat de consistentie van de vastleggingen is onderzocht.

De gegevensgerichte detailcontroles (substantive tests) worden allemaal kritisch uitgevoerd alsof de vastleggingen met fraude gepaard zouden kunnen hebben gaan.

Bij eventuele fouten, die fraude zouden kunnen herbergen, vindt nader onderzoek plaats. Daarbij gebruikt men zowel gereedschap voor data-analyse als statistische steekproeven⁵.

Als een meer nauwkeurig en meer betrouwbaar oordeel nodig is als er aanleiding is voor een vermoeden van fraude kan een grotere steekproef – geldsteekproef bijvoorbeeld – worden getrokken om de beslissers inzake de strafrechtelijke kant te ondersteunen.

Mocht het niet helder zijn hoe en door wie de fraude (nog steeds) wordt gepleegd zal men zijn toevlucht moeten zoeken tot methoden van directe waarneming.

Noot

- 1 Dit essay is een combinatie van de columns Statistical Audit 35, 36 en 59.
- 2 Die methode is toegelicht in de paragraaf over gegevensgerichte detailcontrole.
- 3 Bij de paragraaf over data-analyse hebben we het zoeken *data-analyse om vermoedens uit data 'los te peuteren'* genoemd. Dat is het op zoek gaan naar fraudevermoedens op basis van bekende signalen of bekende soorten signalen. 'GEAUZOMOFO' is een deelverzameling van dergelijke technieken. Het gaat alleen om zoeken naar bekende foutsoorten.
- 4 Eigenlijk gaat het om de volledigheid van de verantwoording van de transacties.
- 5 Dit past helemaal in de regelgeving (NV COS 530 paragrafen 12 en A17 en ook NV COS 450 paragrafen 9 en A10).

Postensteekproef om uitspraken in geld te doen¹

Ferry Geertman en Hein Kloosterman

Wat voor antwoord kunt u van ons verwachten als u vraagt om de opzet van een steekproef? Precies: “Dat hangt ervan af”. U moet zich eerst afvragen of een geldsteekproef mogelijk is. Pas daarna komen er andere mogelijkheden aan de beurt.

Als u een geldsteekproef wilt opzetten moet u van de (deel-)populatie die u wilt onderzoeken weten of de aanname van de zogenoemde ‘nul-fouten-verwachting’ wel klopt. Natuurlijk kan er in een onderzochte deelpopulatie een foutje of wat voorkomen. En, gelukkig, daar is ook best rekening mee te houden.

Stel er is sprake van een deelpopulatie van € 20.000.000. Stel dat de totale populatie € 100.000.000 is. En stel dat de materialiteit voor de jaarrekening als geheel is € 1.000.000. We stellen de uitvoeringsmaterialiteit en de toelaatbare afwijking op 90% daarvan. Als de accountant geen fouten verwacht wordt het interval € 300.000². De steekproefomvang voor de deelpopulatie van € 20.000.000 wordt dan (ongeveer) 67³.

De vraag is dan of het advies er beter van wordt als gegeven is dat er in voorgaande jaren in zo’n

vergelijkbare deelpopulatie fouten zijn gevonden. Ja, daar kan de accountant wat mee. Als er maar een paar afwijkingen werden gevonden kan de accountant besluiten een paar afwijkingen in de steekproef toe te staan.

Staat de accountant bijvoorbeeld twee fouten in de steekproef toe, dan wordt de steekproefomvang groter en wel $6,30 / 3,00 * 67 = 141$ (afgerond naar boven)⁴. En als de accountant wel 10 fouten wil toestaan komt zijn rekensom⁵, de steekproefomvang, op $16,962 / 3,00 * 67 = 379$.

Van zo’n grote steekproef schrikken accountants. ‘Mag het iets minder?’ vraagt men zich dan af.

In het verhaal hierboven zijn alleen de positieve fouten (overstatements) bij de kop gepakt. Wat doen we met understatements? Natuurlijk is er de mogelijkheid om de understatements in overstatements te vertalen. Mogelijkheid, want dat kan niet altijd. En als dat niet kan zijn we aangewezen op een postensteekproef. En als al bekend is dat er sprake is van zowel overstatements als understatements hadden we natuurlijk ook direct de koe bij de hoorns kunnen vatten: door dadelijk te kiezen voor een postensteekproef voor die deelpopulatie.

Wanneer kan men een postensteekproef gebruiken om de uitkomst in geld uit te drukken? In grote lijnen komen de voorwaarden neer op:

- er moeten 180 à 200 elementen worden getrokken;
- er komen veel afwijkingen voor in de steekproef (b.v. 20 in een steekproef van 200);
- deze afwijkingen zijn zowel positief als negatief (overstatements en understatementen);
- grofweg de helft is positief.

Het rekenwerk met de uitkomsten van de steekproef maakt gebruik van de zogenoemde normale verdeling. De methode geldt dus zeker als de afwijkingen normaal verdeeld zijn. Vandaar dat de literatuur bij de vereisten die grote trekking en het grote aantal afwijkingen heeft vermeld.

Als er gekozen wordt voor zo'n postensteekproef staat de omvang van de steekproef ongeveer vast. Die omvang, zagen we bij de voorwaarden, moet ongeveer 200 zijn⁶. In tegenstelling tot bij een geldsteekproef is de omvang niet uit te rekenen op basis van de vereiste nauwkeurigheid, de gewenste betrouwbaarheid en de omvang van de populatie⁷.

Er is een hoeveelheid wetenschappelijk onderzoek waaruit blijkt (er zijn simulatiestudies gedaan) dat optimale resultaten – in termen van nauwkeurigheid en betrouwbaarheid – worden gehaald bij een steekproefomvang van ongeveer 180 tot 200 (relevante) posten (Kriens et al., 1986, Kriens et al. 1989, en Lew 1998).

De techniek die is gebruikt in de studies is een algemene. Men gebruikte de regressieschatter (een schattingsmethode waarbij gebruik wordt gemaakt van hulpvariabelen).

Het aardige van de regressieschatter is dat deze techniek gebruik maakt van het voorhanden zijn van zowel de boekwaarde (de Ist-posities) als de auditwaarde (de Soll-posities), en dus niet alleen van de aangetroffen fouten. De verwachting is namelijk dat heel vaak de boekwaarde en auditwaarde zullen overeenkomen. En dat het toeval wil dat er af en toe (maar wel vaak) verschillen tussen beide waarden zijn. Met de regressieschatter kijkt men dus niet alleen naar de Soll-posities van de onderzochte elementen (posten), maar zowel naar de Ist- als de Soll-posities.

Zou men alleen de Soll-posities bekijken en daarmee schatten dan heet de uitkomst een directe schatter. Nu de hulpvariabele, de Ist-posities, wordt gebruikt kan men de regressieschatter gebruiken. Zoals gezegd: de regressieschatter is de meest algemene techniek. Er wordt ook gewag gemaakt⁸ van een verschilschatter en een ratioschatter. Bij de verschilschatter wordt verondersteld dat de Ist- en de Soll-posities steeds om en nabij dezelfde waarde verschillen. Bij de ratioschatter wordt verondersteld dat de lijn waaromheen de Ist- en Soll-posities liggen door de oorsprong zal gaan. Bij de regressieschatter hoeven die voorwaarden niet vooraf te worden bepaald. Laat maar komen die auditwaarden.

Steekproef	Getelde waarde in €
Artikelsoort 1	150
Artikelsoort 2	200
Artikelsoort 3	300
Artikelsoort 4	345
Artikelsoort 5	5
Artikelsoort 6	500
Gemiddelde	250 (=150+200+300+345+5+500/6)

Zie bovenstaand voorbeeld. Om het voorbeeld leesbaar te houden, gebruiken we een steekproef van maar 6 eenheden.

Het gaat om een voorraad. Er zijn 8.500 verschillende artikelsoorten. De totale boekwaarde is € 2.118.595.

De geschatte voorraadwaarde op basis van de Soll-posities (getelde waarden) is dan:

$8.500 * € 250 = € 2.125.500$.

Er is maar een klein verschil met de boekwaarde. Maar de vraag is hoe nauwkeurig deze schatting is.

De (on)nauwkeurigheid is afhankelijk van de spreiding in de steekproef. En die is groot. In dit geval loopt het 95% betrouwbaarheidsinterval (afgerond) van € 960.000 tot € 3.300.000. De maximale fout is dan het verschil van de boekwaarde met de verst verwijderde grens van het betrouwbaarheidsinterval. In casu is dat € 1.200.000.

Als we nu zowel de boekwaarden als de auditwaarden in aanmerking nemen krijgen we het volgende plaatje.

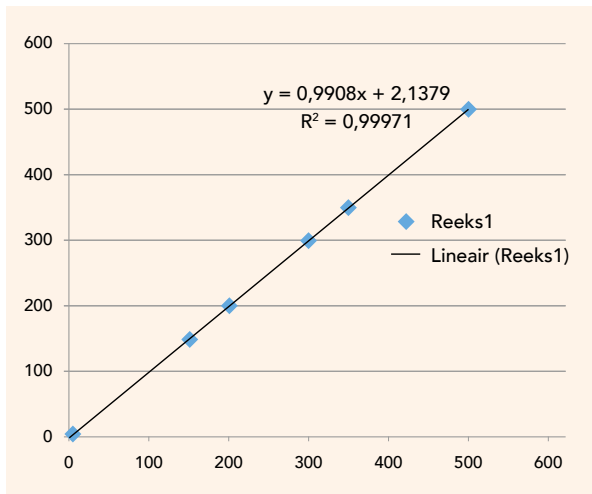
Steekproef	Administratie	Getelde waarde in €
Artikelsoort 1	145	150
Artikelsoort 2	200	200
Artikelsoort 3	302	300
Artikelsoort 4	350	345
Artikelsoort 5	4	5
Artikelsoort 6	500	500
Gemiddelde	250,17	250 (=150+200+300+345+5+500/6)

De geschatte voorraadwaarde op basis van de Soll-posities (getelde waarden) is weer:

$$8.500 * € 250 = € 2.125.500.$$

Uit het rekenwerk blijkt dat het 95%-betrouwbaarheidsinterval⁹ loopt van (afgerond) € 2.097.000 tot € 2.137.000. De maximale fout is nu € 21.595.

Het verschil met de directe schatter is al een beetje uit de tabel af te leiden. Als we de plot van de waarden bekijken wordt de techniek nog overtuigender.



Tot slot

Kortom: als men verwacht dat de fouten om de oren zullen vliegen, zoals vaak bij voorraden het geval is, is het slim om een postensteekproef te trekken en gebruik te maken van de relatie tussen boekwaarde en auditwaarde. Met behulp van een van de leden van de familie van de regressieschatter (de ratioschatter, de verschillenschatter of de regressieschatter) kan dan een betrekkelijk nauwkeurige schatting met een redelijke betrouwbaarheid worden gedaan.

Literatuur

Kriens, J., en J.J.M. Peterse (1986); Toepassing van de regressieschatter in de accountantscontrole, *Statistische Methoden*.

Kriens, J., H. Timmermans, H. van den Wildenberg, J.P.C. Kleijnen (1989); Regression sampling in statistical auditing, *Statistica Neerlandica*; pp. 193-207.

Lew, T.R., (1998); *Regressieschatter, Literatuurstudie naar de praktijk aannames voor toepassen van normaliteit*. Doctoraalscriptie VU Amsterdam.

Muilwijk, J., T.A.B. Snijders en J.J.A. Moors (1992); *Kans-steekproeven*; Stenfert Kroese.

Noot

- 1 Dit artikel is een bewerking van column 65 op Accountant.nl.
- 2 Als een steekproefrisico gelijk aan het auditrisico van 5% wordt toegepast is de R-factor (een klein beetje minder dan) drie. De toelaatbare afwijking van 900.000 gedeeld door 3 wordt dus 300.000.
- 3 20.000.000 gedeeld door 300.000 is 66,666. Afgerond naar boven is dat 67.
- 4 De Poissonverdeling laat een R-waarde van ongeveer 6,30 zien bij 95% betrouwbaarheid en 2 fouten in de steekproef.
- 5 De R-waarde bij 10 fouten is 16,962 bij 95% betrouwbaarheid.
- 6 Er is inmiddels een grote hoeveelheid literatuur over dit onderwerp. Zie Kriens et al. (1986) en Lew (1998).
- 7 Geldsteekproeven zijn meestal steekproeven om te toetsen. De methode die hier wordt besproken is een schattingssteekproef. Bij schattingssteekproeven pleegt een centrale tendentie te worden uitgerekend en een betrouwbaarheidsinterval. In accountantscontrole pleegt het te gaan om een eenzijdig 95%-betrouwbaarheidsinterval.
- 8 Bijvoorbeeld in Hoofdstuk 5 van Muilwijk et al. (1992).
- 9 Zoals eerder aangegeven had kunnen worden volstaan met het berekenen van een 90%-betrouwbaarheidsinterval omdat een eenzijdig interval relevant is, zodra de uitkomsten worden gebruikt om uiteindelijk toch te toetsen.



Gegevensgerichte cijferanalyses

Gegevensgerichte cijferanalyse met behulp van regressierekening

Paul van Batenburg

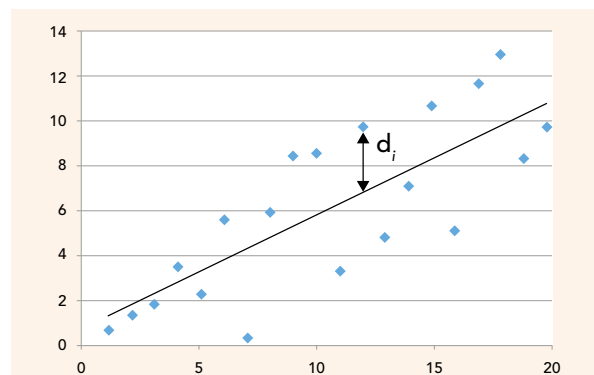
Essentiële voorwaarde voor het toepassen van gegevensgerichte cijferanalyse is dat de accountant voorafgaand aan het verzamelen van het te controleren gegeven een verwachtingswaarde opstelt en een grensbedrag aangeeft voor het aanvaardbare verschil met de realisatie (Standaard 520). Stringer en Stewart (1996) hebben aangegeven hoe de statistische techniek van regressie hiervoor gebruikt kan worden.

Regressie is een verzamelterm voor wiskundige methoden om een formule te schatten die de relatie tussen een aantal waarnemingen het beste representeert. In het geval van één verklarende en één te verklaren variabele zou het betekenen het trekken van een rechte lijn die zo goed mogelijk een puntenwolk vervangt (zie onderstaande grafiek).

Stringer en Stewart gebruiken deze techniek om aan de hand van gegevens uit het verleden (gecontroleerde jaren) een model te schatten en dat model vervolgens te gebruiken om gegevens die nog gecontroleerd moeten worden te voorspellen. Bijvoorbeeld maandbedragen die optellen tot een te controleren jaarrekeningpost omzet, voorspeld uit betrouwbare gegevens ten aanzien van de kost-

prijs van de omzet. De mate waarin de gegevens in het verleden aan de voorspelling voldeden wordt mede gebruikt om het grensbedrag te berekenen, naast de uitvoeringsmaterialiteit en het gewenste niveau van gegevensgerichte controle.

Bij de bepaling van het grensbedrag wordt een redenering gevolgd die ook terug te vinden is in de zogenoemde 'Load and Spread' berekening van Leslie, Teitlebaum en Anderson (1979)¹, en in de afleiding van materialiteit voor een onderdeel uit groepsmaterialiteit in Stewart (2013). Er wordt steeds gekeken naar de verdeling van een materiële fout over te controleren posten die het moeilijkst te ontdekken is.



Op deze manier kan van elk afzonderlijk maandbedrag van de te controleren omzet worden aangegeven of het binnen de bandbreedte van het grensbedrag valt of dat nader onderzoek nodig is. Voordeel van cijferanalyse als middel van gegevensgerichte controle boven steekproeven is dat niet naar een gestoken transactie wordt gekeken maar een deelbedrag aan transacties wordt voorspeld op basis van een (uit historische gegevens) gevalideerde bedrijfseconomische gedragsrelatie met andere gegevens. De accountant steunt dus meer dan bij steekproeven op de vergaarde kennis van de huishouding en geconstateerde afwijkingen kunnen meer waardevolle informatie bieden dan de juistheid of onjuistheid van een gestoken factuur.

Zoals bij elke vorm van data-analyse (deze statistische methode is bij uitstek een voorbeeld van data-analyse) is de toepasbaarheid afhankelijk van de kwaliteit van de aangeleverde gegevens. Als men data wil controleren door middel van andere data, dan moeten die andere data betrouwbaar, en dus (bijvoorbeeld in de juiste controlerichting) gecontroleerd, zijn. Deze vorm van cijferanalyse gedijt dus het best bij een controle waar de interne beheersing garant staat voor bruikbare data.

Literatuur

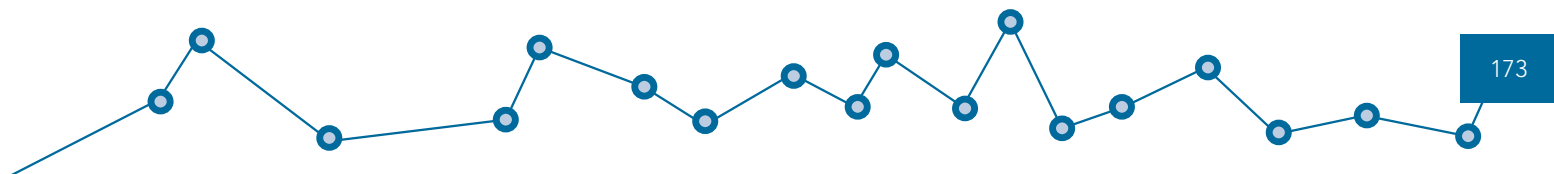
Leslie, D.A., A.D. Teitlebaum and R.J. Anderson (1979); *Dollar Unit Sampling: A Practical Guide for Auditors*, Copp Clark Pitman, Toronto.

Stringer, K.W. and T.R. Stewart (1996); *Statistical Techniques for Analytical Review*, Wiley, 2nd ed.

Stewart, T.R. (2013); *A Bayesian Audit Assurance Model*. Proefschrift Vrije Universiteit Amsterdam.

Noot

- 1 De methode wordt met voorbeelden toegelicht in Appendix B (pp. 278-288) en intuïtief behandeld in hoofdstuk 7 (pp. 135-145) van Leslie et al. (1979).



The background features a light orange color with several white abstract shapes. There is a large white circle on the left side, a white rectangle in the upper left, and a white trapezoidal shape below the rectangle. The text is positioned in the lower right area, overlapping the white circle and the orange background.

Risicoanalyse in de accountantscontrole

Het Bayesiaanse model van Deloitte & Touche ter ondersteuning van de controlemix¹

Paul van Batenburg en Roger Dassen

Inleiding

De discussie over het risicoanalysemodel heeft enige tijd gesmeuld, maar is weer opgelaaid dankzij bijdragen in De Accountant van Kinds (maart 1995) en van Kloosterman (september 1995, met reactie van Kinds). Het Bayesiaanse model komt daarbij aan de orde als mogelijk alternatief, maar niet op een manier die de lezer inzicht geeft in het hoe en waarom daarvan. In dit artikel wordt het Bayesiaanse model geschetst dat Deloitte Touche Tohmatsu International gebruikt ter ondersteuning van de controlemix. Hierbij wordt vooral ingegaan op de methodologische overwegingen en op de praktische toepassing, en minder op de wiskundige achtergronden.

Aanleiding

Twee ontwikkelingen lijken relevant om juist nu op het Bayesiaanse model in te gaan:

- In accountantsopleidingen, waaronder die van de Rijksuniversiteit Limburg en die van NIVRA/Nyenrode, wordt het Bayesiaanse model naast het risicoanalysemodel gedoceerd. Ook andere discussiepunten op het gebied van 'Sta-

tistical Auditing', zoals het al dan niet toedelen van materialiteit en controletolerantie, en de samenhang tussen trekkingstechniek en evaluatietechniek, komen daarbij aan de orde. Voor de NIVRA/Nyenrode opleiding wordt een keuzevak 'Quantitative Audit Support' ontwikkeld, waar studenten zich kunnen verdiepen in het gebruik van statistiek bij het ontwerpen van een beslissingsmodel voor de bepaling van de controlemix. In het curriculum Leer der Accountantscontrole van die opleiding wordt vanaf dit najaar ook aan 'Quantitative Audit Support' aandacht besteed.

Nu de Bayesiaanse aanpak in de opleiding een plaats heeft gekregen, is het goed om ook reeds afgestudeerden hiermee kennis te laten maken.

- Deloitte Touche Tohmatsu International heeft in haar nieuwe internationale audit manual, AuditSystem/2, de door de aangesloten Nederlandse firma aangedragen methodiek (BRAM, Bayesian Risk Assessment Model, gedoopt, als vertaling van Bayesiaans Risico Analyse Model) opgenomen naast het Audit Assurance Model als "another acceptable method of determining the scope of our audit work for each significant account balance and potential error".

De Bayesiaanse aanpak wordt dus daadwerkelijk in de praktijk toegepast, en is daarom relevant voor in de praktijk werkende accountants.

De plaats van gegevensgerichte controle in het controleproces

Accountantscontrole heeft als doel een uitspraak te doen over de getrouwheid van een jaarrekening. In de controleaanpak van de accountant wordt dit geoperationaliseerd door de mogelijke financiële fout in een jaarrekening te confronteren met een daarvoor gestelde norm, het materialiteitsbedrag. Een goedkeurende accountantsverklaring heeft dan ook de strekking dat een redelijke mate van zekerheid is verkregen dat de jaarrekening geen onjuistheden van materieel belang bevat.

In dit kader wordt niet ingegaan op de berekening van het materialiteitsbedrag. Ook wordt niet (direct) ingegaan op de kwantificering van het begrip ‘redelijke mate van zekerheid’, omdat niet het Audit Assurance Model (AAM) maar het Bayesiaanse model (BRAM) centraal staat. De doelstelling in de vorige alinea is zo geformuleerd omdat daaruit is af te leiden dat de accountant dient te kwantificeren. Argumenten daarvoor gaven Wennink en Van Batenburg in De Accountant (januari 1996) in hun reactie op het artikel van Diekman (juni 1995). Gekwantificeerde resultaten kunnen alleen maar worden afgeleid uit gegevensgerichte (detail-)controles.

De functie van de controlemix is daarbij:

- het bepalen van de toereikende inspanning aan gegevensgerichte controle, gegeven de reeds bekende of de verwachte conclusies uit andere fasen van de controle (de risicoanalyse en de organisatietests op de werking van de AO/IC), maar natuurlijk ook:
- het bepalen van de toereikende inspanning aan gegevensgerichte controle ingeval de verwachte conclusies uit die andere fasen van de controle minder gunstig uitpakken dan vooraf werd ingeschat.

De controlemix modelleert dus het constante (en volkomen terechte!) zoeken van de accountant naar de meest efficiënte wijze om tot het gewenste oordeel te komen. Efficiency zonder kwaliteitsverlies, dus.

Accountants worden dankzij de controlemix niet beloond voor het uitgevoerde werk maar voor de verantwoordelijkheid die zij op zich nemen door werkzaamheden te beperken op grond van andere, reeds verkregen, uitkomsten. Daarom is het essentieel dat de grondslagen voor die verantwoordelijkheid kunnen worden gevalideerd. Dat is het belang van het Bayesiaanse model.

Opzettelijk wordt de vage term ‘toereikende inspanning aan gegevensgerichte controle’ gebruikt in plaats van het simpeler ‘minimale steekproefomvang’, omdat ook gegevensgerichte cijferanalyses als manier van gegevensgerichte controle een plaats in een model voor de controlemix verdie-

nen. In dit artikel volgen nog enkele formules uit de steekproeftheorie, maar de conclusies uit die formules kunnen ook worden geïnterpreteerd bij het toepassen van cijferanalyses, net zoals accountants ervan uitgaan dat resultaten van cijferanalyses equivalent zijn met resultaten uit steekproefcontroles.

Oorzaak en gevolg in de controlemix

Het BRAM onderscheidt zich al bij de opzet van de controlemix van het AAM. Het AAM bepaalt immers slechts wat de noodzakelijk resterende hoeveelheid gegevensgerichte controle is. Het BRAM stelt vast wat deze hoeveelheid zou zijn als de accountant niet op voor het foutenbeeld gunstige voorkennis steunt (niet wil of kan steunen) en geeft vervolgens aan welke 'korting' daarop overeenkomt met welke gunstige voorkennis. Waar dus het AAM slechts de 'netto'-hoeveelheid gegevensgerichte controle berekent, geeft het BRAM zowel de 'bruto'-als de 'tarra'-hoeveelheden. Het BRAM werkt daarom twee kanten uit:

- De accountant heeft een bepaalde voorkennis over de te controleren populatie afgeleid en kan met het BRAM uitrekenen welke vermindering aan gegevensgerichte controle men zich kan permitteren door op die voorkennis te steunen;
- De accountant bepaalt met het BRAM wat de noodzakelijke voorkennis zou zijn voor een gewenst niveau van gegevensgerichte contro-

les, en richt de voorafgaande controlefasen zo in dat hij of zij denkt die voorkennis te kunnen vergaren.

De nadruk bij het BRAM ligt dus niet alleen bij de alsnog te controleren steekproefomvang, maar bij de wisselwerking tussen het vergaren en kwantificeren van valideerbare voorkennis en de daarmee samenhangende efficiency in de uitvoering van de controle.

Kwantificering van de verkregen dan wel te verkrijgen voorkennis

Welk model de accountant ook gebruikt, welke school deze ook aanhangt, er bestaan geen manieren om kwalitatieve informatie uit risicoanalyse of uit organisatietests objectief te kwantificeren. Wat dus het AAM en het BRAM bindt is dat beide manieren zijn om *subjectieve kwalitatieve* informatie *subjectief* om te zetten in *kwantitatieve* informatie. Wat AAM en BRAM scheidt is dat die subjectief verkregen informatie in het AAM niet en in het BRAM wel *objectief* toetsbaar is.

Kinds (in zijn reactie op Kloosterman, september 1995) slaat de spijker op de kop met zijn retorisch bedoelde vraag: "Maar als de accountant niet in staat is om de kans op een materiële fout (...) in te schatten, hoe kan hij dan wel de maximale onnauwkeurigheid (gegeven een vooraf bepaald risiconiveau) bepalen?" Toch kan die vraag beantwoord worden, met dien verstande

dat niet het ene moeilijk en het andere makkelijk is. Het is eerder zo dat het ene onmogelijk, en het andere knap lastig is. De kans op een materiële fout is een inschatting waar geen feedback over verkregen kan worden (u had 37% geschat, maar het had 47% moeten zijn). Achteraf is die kans namelijk gewoon 0 (er was geen materiële fout) of 1 (er was wel een). Het foutbedrag dat volgens de accountant zeker niet is overschreden kan achteraf wel worden getoetst aan de werkelijkheid (u had 50.000 geschat, maar er bleek 80.000 fout).

Een dergelijke toets achteraf is arbeidsintensief en dus duur, maar mogelijk. Integrale controle, of een steekproef van minstens de omvang die men eerst niet heeft hoeven uitvoeren als gevolg van de conclusies uit voorgaande controlefasen, zijn beide als validatiemethode denkbaar. Die valideerbaarheid achteraf maakt de inschatting vooraf op zichzelf niet eenvoudiger, maar daarom worden accountants ook zo goed opgeleid op het specificeren van hun voorkennis ('professional judgment').

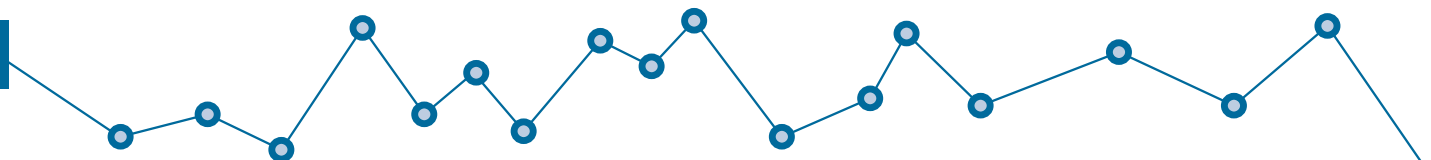
Uitgangspunt voor het BRAM is daarom dat voorkennis over een te controleren populatie moet worden gespecificeerd in termen van parameters van die populatie, en niet in termen van de wijze waarop die parameters onderzocht worden. Risicoanalyse en organisatietests worden uitgevoerd om de accountant voorkennis over de fout in de te controleren populatie te leveren, niet om het risico op een onjuist testresultaat te kunnen wijzigen. Het BRAM verwerkt die voorkennis in een for-

mule die de mogelijk aan te treffen foutbedragen weergeeft, en berekent hoeveel goede steekproefuitkomsten vervolgens nodig zijn om die formule om te vormen tot een voldoende nauwkeurige uitspraak over de onbekende foutkans in de populatie. Het AAM wijzigt het toelaatbare risico waarmee die foutkans empirisch wordt getoetst. Gevolg is dat die voorkennis in het AAM achteraf niet te valideren is, en het AAM daarom niet wetenschappelijk aanvaardbaar is.

Gevolg van dit uitgangspunt is dat de eertijds gebruikte formule rondom de factor f ('de stabiliteit van het administratieve proces') ook niet aan de eis van valideerbaarheid voldoet. Zelfs integrale controle van zowel de populatie uit dit jaar als die uit vorig jaar leert ons heel veel, maar niet hoe groot f was. Discussie over 'het' Bayesiaanse model als alternatief voor het AAM, zoals Kinds (1995) die entameert, zou dus niet (meer) over *dat* Bayesiaanse model moeten gaan.

De steekproefomvang in AAM en BRAM

Het model dat wèl ter discussie zou kunnen komen is het volgende. AAM en BRAM gaan er beide van uit dat de gegevensgerichte controle wordt uitgevoerd door middel van een aselechte geldsteekproef op een populatie groot M . De vooraf vereiste nauwkeurigheid van de controle, de controletolerantie, is een bedrag MP dat door de accountant (op enigerlei wijze afgeleid uit de materialiteit) is gekozen.



risico	1%	4%	5%	10%	14%	28%	37%	50%	77%
R	4,61	3,23	3,00	2,31	2,00	1,30	1,00	0,70	0,20

De noodzakelijke omvang van de steekproef (n) die bij het vinden van 0 fouten toereikend is voor de gestelde nauwkeurigheidseis MP is:

$$n = R \times M/MP$$

waarin R afhangt van de vereiste betrouwbaarheid van de uitspraak, of, analoog, van het toegelaten risico op een onjuiste uitspraak. Het getal R is te vinden in statistiekboeken onder ‘betrouwbaarheidsbovengrenzen voor de verwachting van een Poisson verdeelde variabele’. Een aantal voor dit artikel relevante waarden van R staat in de tabel, zie boven.

Het AAM is een manier om het toegelaten risico te kiezen aan de hand van inschattingen van het inherente risico en intern controle risico. Deze keuze bepaalt R in de hierboven gegeven formule. Dit artikel gaat over het Bayesiaanse model, waarin het toegelaten risico identiek blijft aan het toegelaten accountantsrisico.

Voorkennis in het BRAM

Een tweede formule in het Bayesiaanse model beschrijft de relatie tussen de gekwantificeerde subjectieve voorkennis over de fout in de populatie en het aantal waarnemingen (n*) dat op grond

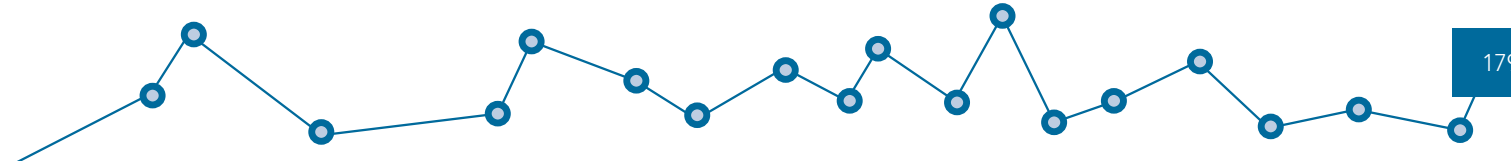
van die voorkennis niet meer nodig is. (Zoals hiervoor gemeld mag deze formule twee kanten uit gebruikt worden, en kan ook worden bepaald welke voorkennis verkregen moet worden om een gekozen aantal waarnemingen niet meer te hoeven uitvoeren). De moeilijkheid zit ‘m niet in de formule maar wel in de verantwoordelijkheid die op de schouders van de accountant rust bij het kwantificeren van die voorkennis.

De vraag die het BRAM de accountant stelt is: “Welk foutbedrag wordt in deze populatie zeker niet overschreden?” Het antwoord geven wij symbool MP* en wordt door ons voorzichtigheidshalve (BRAM kent nog een overschrijdingskans toe aan het bedrag dat de accountant zeker niet overschreden acht) geïnterpreteerd als het antwoord:

“Ik heb een steekproef van omvang n gecontroleerd en daarin geen fouten gevonden, zodat de geschatte maximale fout (bij een onbetrouwbaarheid overeenkomend met R) in deze populatie MP* bedraagt”.*

Daardoor kunnen wij stellen dat $n^* = R \times M/MP^*$, en de overblijvende te controleren steekproefomvang (die de afstand tussen MP* en MP overbrugt) bedraagt:

$$n - n^* = R \times M/MP - R \times M/MP^*.$$



Natuurlijk zijn deze formules benaderingen van een ingewikkelder wiskundig model. Daardoor kan de accountant die niet op voorkennis steunt stellen dat $MP^* = M$ (een massa groot M kan niet meer overgewaardeerd zijn dan het bedrag M zelf) om vervolgens op grond van deze formule de steekproefomvang een getal R te verkleinen ten opzichte van de eerste formule. Dat is een net niet toegestane concessie aan de eenvoud van deze benadering, die in onze interne handleiding en de ondersteunende software is verijdel.

Valideerbaar maken van het AAM

Door de steekproefreductie op grond van het AAM ten opzichte van het niveau zonder voorkennis gelijk te stellen aan n^* in het BRAM kan een relatie tussen het AAM en het BRAM worden afgeleid. Deze relatie kan worden gebruikt om het AAM valideerbaar, en dus wetenschappelijk aanvaardbaar, te maken. Die relatie beschrijft welke achteraf te valideren voorkennis MP^* nodig is om de verlaging te onderbouwen van de $R1$ die voor het accountantscontrole risico noodzakelijk is naar de $R2$ die in de steekproefcontrole gewenst is:

$$MP^* / MP = R1 / (R1 - R2) = \ln(ACR) / [\ln(IR) + \ln(CR)].$$

Oftewel: het gezamenlijk effect van de subjectieve inschattingen van inherent risico en intern controle risico, die niet achteraf valideerbaar zijn, kan valideerbaar gemaakt worden door te

berekenen dat deze keuzen equivalent zijn met het nemen van de verantwoordelijkheid voor de veronderstelling dat de fout in de te controleren populatie niet groter was dan een aantal malen de gestelde controletolerantie.

Een voorbeeld: de tabel van R -waarden liet zien dat bij 5% toegelaten risico $R1 = 3$ hoort. Stel nu dat een accountant het inherent risico inschat op $IR = 37\%$ en het intern controle risico op $CR = 28\%$. De AAM-formule: $(ACR = IR * CR * SR)$ levert nu als toegelaten steekproefrisico SR voor een accountantscontrole risico $ACR = 5\%$ op:

$$SR = ACR / [IR * CR] = 5\% / [37\% * 28\%] = 50\%, \text{ zodat } R2 = 0,7.$$

(Dat hier 50% uitkomt zal uw rekenmachine tegenspreken. Het komt er toch uit omdat de 37% en de 28% afrondingen zijn. In de praktijk rekent men met R , zodat (zie de tabel !) het sommetje wordt: $SR = R(ACR) - R(IR) - R(CR) = 3 - 1 - 1,3$)

De bovenaan de paragraaf geleverde formule geeft als te valideren voorkennis:

$$MP^* / MP = 3 / [3 - 0,7] = 3 / [1 + 1,3] = 1,3$$

Oftewel: het gezamenlijk effect van de subjectieve inschattingen van inherent en intern controle risico van 37% respectievelijk 28%, die niet achteraf valideerbaar zijn, kan valideerbaar gemaakt worden door te berekenen dat deze keuzen equivalent zijn met het nemen van de verantwoordelijkheid voor de veronderstelling dat de fout in de te controleren

populatie niet groter was dan 1,3 maal de gestelde controletolerantie.

Toepassing bij Deloitte & Touche

De AuditSystem/2 controleaanpak van Deloitte Touche Tohmatsu International gaat er vanuit dat de accountant aan de hand van een risicoanalyse de controlemix bepaalt tussen organisatiegerichte en gegevensgerichte controles. Deze laatste kunnen bestaan uit gegevensgerichte cijferanalyses of aselechte detailwaarnemingen, maar slechts in uitzonderingssituaties een combinatie van beide. AuditSystem/2 schrijft voor dat op grond van risicoanalyse en organisatietests het niveau van gegevensgerichte controle wordt gekozen uit 4 mogelijkheden:

- 'uitgebreid' als een specifiek risico (d.w.z. een meer dan normaal inherent risico) is onderkend en niet wordt gepoogd te steunen op IC-maatregelen die dat risico mitigeren;
- 'gematigd' als geen specifiek risico is onderkend en besloten is niet te steunen op de betrouwbaarheid van de AO/IC;
- 'beperkt' als een specifiek risico is onderkend én door organisatiegerichte controles de goede werking is vastgesteld van IC-maatregelen die dat risico mitigeren, of als geen specifiek risico is onderkend en door organisatie-

gerichte controles is vastgesteld dat gesteund kan worden op de betrouwbaarheid van de AO/IC;
als geen specifiek risico is onderkend en door een uitgebreide organisatiegerichte controle is vastgesteld dat gesteund kan worden op de betrouwbaarheid van de AO/IC.

'laag'

In de AAM-benadering van deze niveaus wordt aan elk niveau een R-waarde gekoppeld. De omvang van de steekproef voor het uitgebreide niveau dient te worden bepaald met behulp van $R = 3$, gematigd met $R = 2$ en beperkt met $R = 0,7$. Voor het 'lage' niveau worden wel zeer strenge eisen gesteld aan de omvang en diepgang van de organisatiegerichte controles, maar wordt geen R-waarde voorgeschreven. Dat is niet zo verwonderlijk, gegeven dat $R < 0,7$ overeenkomt met een toegelaten risico boven 50%. Voor een 'ontkenner' van het AAM behoeft dat geen verder betoog: het opgooien van een kwartje is dan zelfs een betere methode van steekproefcontrole. Er zijn echter toepassingen van steekproefcontroles door AAM-adepten met een toegelaten risico van 77%; het Nederlands record is 55%.

In de BRAM-benadering wordt nu gesteld dat de steekproefomvang behorende bij het uitgebreide niveau met een factor $2/3$ mag worden vermenigvuldigd ($R = 2$ in plaats van 3) mits de accountant zich zeker voelt dat de fout in de te controleren populatie niet groter is geweest dan 3 maal de controletolerantie:

$$2 M / MP = 3 M / MP - 3 M / MP^* \text{ als } MP^* = 3 MP.$$

Voor het beperkte niveau is het rekensommetje voor de bepaling van $MP^* = 1,3 MP$ in het voorbeeld hierboven uitgevoerd; de steekproefomvang is dan $1/4$ ($R = 0,7$ in plaats van 3 , en wat afronding) van die op het uitgebreide niveau.

Als illustratie van de eenvoud van het Bayesiaanse model in de praktijk wordt hier een bladzijde getoond uit het 'steekproefboekje' dat aan controleteams van Deloitte & Touche wordt uitgereikt, en waarin receptmatig alle stappen voor opzet, uitvoering en evaluatie van steekproefcontroles

zijn aangegeven. Voor de leesbaarheid zijn enkele symbolen uit het boekje (dat $1/4$ A4-formaat heeft en dus met ruimte moet woekeren) door lopende tekst vervangen.

Voor het lage niveau is, zoals gemeld, geen R -waarde voorgeschreven. Wel is in de AuditSystem/2 controleaanpak een groot aantal voor de hand liggende voorbeelden gegeven van invulling van een verhouding MP^*/MP in de formule:

$$3 M / MP - 3 M / MP^*.$$

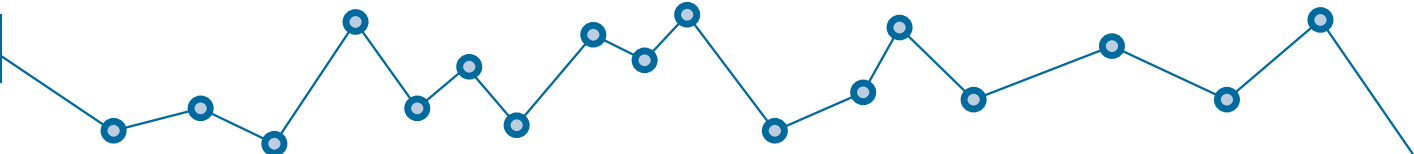
Gematigd / beperkt / laag niveau: opzet geldsteekproef

Bij **gematigd, beperkt of laag** niveau zal de vermoedelijke fout nihil zijn, dus worden geen fouten toegelaten in de steekproef en is de **R-waarde 3**. Dankzij het in de AuditSystem/2 controleaanpak verweven Bayesiaanse RisicoAnalyse Model is een **kleinere steekproef** dan **$n = 3 M/MP$** toereikend, mits uit risicoanalyse en/of proceduretests voldoende voorkennis is verkregen om te veronderstellen dat de fout in de populatie hoogstens MP^* is. In de formule $3 M/MP - 3 M/MP^*$ is voor het gematigd en beperkt niveau een te bereiken MP^* voorgesteld:

niveau	voorkennis MP^*	omvang
gematigd	$3 \times MP$	$2/3 n$
beperkt	$1,3 \times MP$	$1/4 n$
laag	MP^*	$3 M/MP - 3 M/MP^*$

De verkregen voorkennis is gekwantificeerd in een **valideerbare** uitspraak: de veronderstelling, dat de populatiefout hoogstens MP^* was, is **desgewenst** achteraf op juistheid te toetsen. Het initiatief daartoe kan het gevolg zijn van Maatschapsbeleid.

Vanwege de evaluatiemethode (zie aldaar) wordt altijd eerst n bepaald, waarna de steekproefomvang op het gehanteerde niveau berekend wordt.



Een voorbeeld: de steekproefomvang die overeenkomt met $R = 0,2$ (77% toegelaten steekproefrisico, het ons bekende 'wereldrecord AAM') bedraagt volgens de eerder gegeven AAM-formule:

$$n = 0,2 M / MP.$$

Deze steekproefomvang is d.m.v. BRAM ook toereikend voor de accountant die de veronderstelling kan onderbouwen dat $MP^*/MP = 15/14$, want dan volgt:

$$n = 3 M / MP - 3 M / (15/14 MP) = (45/15) M / MP - (42/15) M / MP = (3/15) M / MP.$$

De omvang van de nog te controleren steekproef is nu een gevolg geworden van de verkregen voorkennis, en dus ook van de inspanning die de accountant zich heeft getroost om die voorkennis te verkrijgen. Belangrijkste pre van deze aanpak blijft dat de accountant zich er achteraf (al dan niet op 'verzoek' van anderen) van kan vergewissen of zijn voorkennis terecht heeft geleid tot de veronderstelling dat niet meer fouten dan MP^* in de te controleren populatie aanwezig waren. Een efficiënte controle, maar tegelijkertijd met een duidelijk handvat voor kwaliteitsbeoordeling.

Conclusies?

Een artikel met beschrijvende probleemstelling ("In dit artikel wordt het Bayesiaanse model geschetst dat Deloitte Touche Tohmatsu International gebruikt ter ondersteuning van de controlemix") mag geen conclusies bevatten. Ter afsluiting, daarom, zeven stellingen:

1. Doel van de accountantscontrole is het vellen van gekwantificeerde oordelen over populaties van gegevens.
2. Daartoe zijn gegevensgerichte controles (steekproeven of cijferanalyses) nodig op elementen van die populaties.
3. De controlemix is een manier om de hoeveelheid gegevensgerichte controle zonder kwaliteitsverlies zo klein mogelijk te maken.
4. Efficiency zonder kwaliteitsverlies bestaat niet zonder kwaliteitstoetsing.
5. Daarom dienen in de controlemix gedane aannames op hun terechtheid getoetst te (kunnen) worden.
6. Het BRAM maakt een objectief toetsbare kwantificering van subjectieve kwalitatieve grootheden mogelijk.
7. Dat die kwantificering moeilijk is, mag niet het model verweten worden, maar wordt gerechtvaardigd door het belang van de uitkomst ervan.

Literatuur

Batenburg, P.C. van, P.T.F.M. Wennink (1996); Toleranties in de accountantscontrole: zonder kwantiteiten geen kwaliteit, *De Accountant*, nr. 5, januari, pp. 314-317.

Diekman, P.A.M. (1996); Toleranties in de accountantscontrole, *De Accountant*, nr. 10, juni, pp. 665-668.

Kinds, J.H.G. (1995); Controleren in onzekerheid, *De Accountant*, nr. 7, maart, pp. 458-461.

Kloosterman, H.H.W. (1995); Onzekerheid en risicoanalyse, een nieuw gevecht?, *De Accountant*, nr 1, september, pp. 26-29.

Noot

- 1 Dit artikel is geplaatst geweest als:
Batenburg, P.C. van, R.J.M. Dassen, (1996); Het Bayesiaanse model van Deloitte & Touche ter ondersteuning van de controlemix, in: *De Accountant*, jg. 103, no. 1, september, pp. 31-35.

Wat is eigenlijk risicoanalyse in de accountantscontrole¹

Hein Kloosterman

Samenvatting

Risicoanalyse in de accountantscontrole heeft een lange geschiedenis. In dit artikel is die geschiedenis geschetst. De huidige stand van zaken is een volwassen model voor het gegevensgerichte deel van een financiële controle. Dat model is gebouwd met kennis uit het vakgebied Statistical Audit. Het artikel beschrijft ook dat een deel van de werkzaamheden van een financiële controle niet met Statistical Audit is te verklaren. Dat deel noemt de auteur voorwaardelijke controles. Het audit risk dat daarmee gepaard gaat en met bijvoorbeeld IT-audit is tot nu toe nog ontoereikend beschreven. Er wordt een poging gedaan daarvoor een oplossingsrichting te bieden.

1 Inleiding

Risicoanalyse in de accountantscontrole is in de afgelopen tien jaar nauwelijks nog expliciet geworden². Dat blijkt bijvoorbeeld uit de literatuur op de lijst voor de Leer van de Accountantscontrole (Knechel 2001). De ratio van risicoanalyse in de controle zoals die onder meer door Anthony Steele (1992) is beschreven is na 1992 nog enkele

malen vrij krachtig naar voren gebracht, maar is daarna in de vakliteratuur en de regelgeving tot niet meer dan een aantal impliciete regels geworden. Volgens Steele heeft de risicoanalyse in de accountantscontrole een hoeveelheid wiskunde en statistiek nodig om die risicoanalyse beter te kunnen toepassen en verder te ontwikkelen. Daarvoor is ondersteuning met computers nodig. Inmiddels zijn de mogelijkheden om computers in de controle te gebruiken fors toegenomen, maar de ontwikkeling van risicoanalyse lijkt de afgelopen jaren te hebben stilgestaan. Dat is reden genoeg om te laten zien wat de ontwikkeling van risicoanalyse in de accountantscontrole is geweest en om te zien welke rol de Bayesiaanse statistiek³, zeg maar de wiskunde die Steele propageerde, speelt.

Een aantal Nederlandse schrijvers – bijvoorbeeld Mooijekind (1991), Schilder (1991) en Diekman (1995) – probeerden in hun schrijven wiskunde weg te houden van risicoanalyse. In 1996 hebben Van Batenburg en Dassen (1996) en Kloosterman (1996) in De Accountant een poging gedaan om een model van de gegevensgerichte controle, gebaseerd op Bayesiaanse statistiek weer te geven.

Om mijn ongerustheid over deze begripsvermenging tot uitdrukking te brengen heb ik geprobeerd het 'audit risk' probleem te schilderen, de modellen die in de literatuur gangbaar zijn tegen hun historische achtergrond te plaatsen en de verschillende soorten risicoanalyse in hun context te laten zien. Ik geef mijn visie op de modellering rond de toereikendheid (sufficiency) van de financiële controle. Daarmee hoop ik dat de discussie over andere vormen van risicoanalyse, zoals Mollema (2003, 2004) met zijn aanzet beoogt, tot vooruitgang leidt.

2 Risicoanalyse in de accountantscontrole: aggregatie van evidence

Er zijn verschillende bronnen voor controle-informatie. Verondersteld wordt dat die bronnen elkaar niet beïnvloeden. Elke bron is een hoeveelheid gegevensgerichte steekprofelementen⁶ waard. Als die steekprofelementen bij elkaar worden opgeteld en die hoeveelheid voldoet aan de eisen die worden gesteld aan een controle die zou worden uitgevoerd met alleen maar gegevensgerichte controles, dan is er toereikend gecontroleerd om te kunnen goedkeuren. Impliciet gaat zo'n model uit van steekproeven waarin geen fouten worden aangetroffen.

186

omstandigheden niet voorhanden is betekent dat, dat de hoeveelheid gegevensgerichte controles die door die evidence werd vertegenwoordigd alsnog uitgevoerd moet worden.

Op dezelfde manier vertegenwoordigt bijvoorbeeld de stelling van de accountant dat de interne controle allerlei verslaggevingsrisico's mitigeert en dat hij daardoor in belangrijke mate kan steunen op de interne controle, een omvang van een steekproef van gegevensgerichte controles. Vaak betekent het zelfs dat de totale gegevensgerichte steekproef wordt geacht uitgevoerd te zijn door de interne controle.

Ook hier geldt: als de accountant tot dat oordeel komt, moet hij het kunnen onderbouwen. Als hij dat oordeel: "het kunnen steunen op de interne controle" onderbouwt, moet iedere andere accountant met dezelfde informatie tot datzelfde oordeel kunnen komen.

Het beeld dat Steele schetste heeft maar beperkt invloed gehad op de discussie in de Nederlandstalige literatuur⁷. Die discussie leidde tot een vrij breedgedragen conclusie dat het gegoochel met risicopercentages hooguit bestempeld kan worden als denkmodel.

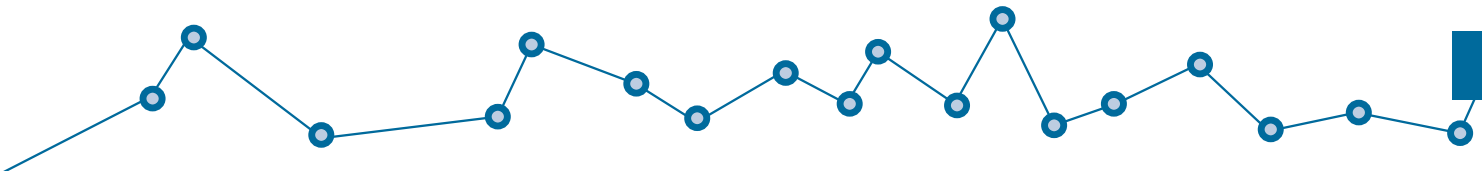
Dat is niet consistent met de waarneming dat veel kantoren vervolgens toch, en misschien wel stiekem, gaan rekenen. Uiteindelijk hebben Van Batenburg en Dassen (1996), in de geest van Steele, de zwakke punten van het ARM met onder meer Bayesiaanse statistiek geneutraliseerd. Een

deel van hun oplossing is vergelijkbaar met weersvoorspellingen. Die geven vaak de geschatte kans gegeven dat het morgen wel of niet gaat regenen⁸. Van Batenburg en Dassen schatten niet de kans maar schatten de maximale regenval van morgen.

Voor de volledigheid merk ik nog op dat zowel het model van Van Batenburg en Dassen als dat van Steele is vormgegeven met de theorie over geldsteekproeven⁹. Van Batenburg en Dassen hebben gebruikgemaakt van het niet hoeven toedelen van controletoleranties (Blokdijs 1996).

Ik wil benadrukken dat het evident zou moeten zijn dat het risicoanalysemodel alleen gaat over het beperken van de hoeveelheid gegevensgerichte controles. Met andere woorden over de positieve detailcontroles, de 'substantive tests'. Dat betekent dat de onvervangbare interne controle en de volledigheid van de transactieverantwoording al gecontroleerd moeten zijn voordat er gesproken kan worden over vermindering van de hoeveelheid gegevensgerichte controles (en pas als die ook goed genoeg zijn kan er goedgekeurd worden).

De controles met betrekking tot de onvervangbare interne controle en die met betrekking tot de ontvangstverantwoording zijn dus niet uitwisselbaar met de gegevensgerichte controles! Wel kan het zo zijn dat die voorwaardelijke controles informatie verschaffen over de kwaliteit van de gegevens. Met andere woorden: de accountant moet erop bedacht zijn dat de minimale controle-inspanning



organisatiegerichte controles omvat, die niet door welke gegevensgerichte controles dan ook, kunnen worden vervangen!

Uiteindelijk geeft de accountant met zijn verklaring aan, dat hij redelijke zekerheid biedt dat de verantwoording niet meer dan (het bedrag van) de materialiteit zal afwijken van de werkelijkheid. Deze waarborg maakt dus de gebruiker duidelijk dat zijn beslissingen met de mogelijke afwijking in deze financiële verantwoording niet zullen wijzigen.

3 Enige geschiedenis

Het ARM is ontstaan in de sfeer van steekproeven in de accountantscontrole¹⁰. In eerste instantie heeft het toevoegen van wiskunde aan de controleaanpak geleid tot het toepassen van steekproeven¹¹. In tweede instantie wordt de steekproefaanpak gebruikt als basis voor de ontwikkeling van een controleaanpak¹². In 1964 maakt Stringer (Bell en Wright, 1995, blz. 24) onderscheid in fouten die ontstaan in het accountingproces en het niet-ontdekken van fouten met (gezamenlijk) een materiële omvang. Voor de eerste soort risico moet de accountant vertrouwen op het systeem van interne beheersing, voor de tweede soort voert hij detailcontroles en dergelijke uit. Dit leidt tot de introductie van het begrip 'control risk'. Omstreeks 1975 voerde Stringer cijferanalyse als mathematische techniek ten tonele¹³. Dat leidt tot het toevoegen van een component 'analytical

review risk'. In het historisch overzicht van Bell en Wright (1995) is goed te zien hoe een heleboel combinaties van praktijkmensen en academici het gedachtegoed hebben uitgebreid.

Eind jaren zeventig en begin jaren tachtig (van de twintigste eeuw) heeft het model een extra duw in de rug gekregen door de ontwikkeling van Monetary Unit Sampling (Leslie, Teitlebaum en Anderson, 1979). Door het kunnen toepassen van geldsteekproeven kan een financiële verantwoording worden beschouwd als een bak met geld die gecontroleerd moet worden (bijvoorbeeld de jaarrekening als geheel). Die bak met geld kan zijn samengesteld uit een heleboel componenten. Voor zover over een of meer van de componenten een afzonderlijk oordeel nodig is kunnen die afzonderlijke componenten weer zelf als een bak geld worden beschouwd¹⁴.

Uiteindelijk is het met behulp van het ARM, als het Bayesiaans is bijgesteld, mogelijk om te beschrijven hoe een accountant in staat is om een garantie uit te spreken over de maximale fout die de onderzochte verantwoording bevat (Van Batenburg en Dassen 1996).

De geleidelijke ontwikkeling van het Audit Risk Model verbloemt wellicht dat het model niet de hele controle beschrijft. Eigenlijk, althans zo formuleer ik het, zegt het model niets over wat er voorafgaand aan de planning van de gegevensgerichte controles moet gebeuren. De preplanning (van de gegevensgerichte controles) is niet met het model beschreven. Die preplanning bestaat

uit de oriëntatie op het te controleren bedrijf en uit controleactiviteiten die uitsluitel geven of de gegevensgerichte controle überhaupt mogelijk is. Die laatste categorie maatregelen noem ik voorwaardelijke controles. De toevoeging: “om met de gegevensgerichte controles tot een goedkeurend oordeel te kunnen komen” pleeg ik weg te laten.

Samenvattend laat de geschiedenis van het Audit Risk Model zien dat geprobeerd wordt de informatie te gebruiken die verzameld is voordat met de gegevensgerichte controle, of substantive procedures, wordt begonnen. Die informatie die uit de preplanning van de gegevensgerichte controle komt, gaat in beginsel over alle aspecten van de onvervangbare interne controle. De ‘voorinformatie’ die leidt tot een reductie van de gegevensgerichte controle kan worden gezien als een nevenproduct.

4 Hoe werkt ARM

In deze paragraaf zet ik globaal de werking van het ARM uiteen. Daarna geef ik de toegevoegde waarde van een Bayesiaans model weer.

4.1 Het ARM zelf

Zoals in paragraaf 3 is aangegeven is het Audit Risk Model ingericht als een combinatie van (ten minste) steekproefrisico, interne-controlerisico en cijferanalyserisico. Later is daar nog het zogenoemde inherent risico bijgekomen. Een en ander leidde tot een formule:

$$\text{Audit risk} = \text{inherent risk} * \text{control risk} * \text{analytical review risk} * \text{sample risk}$$

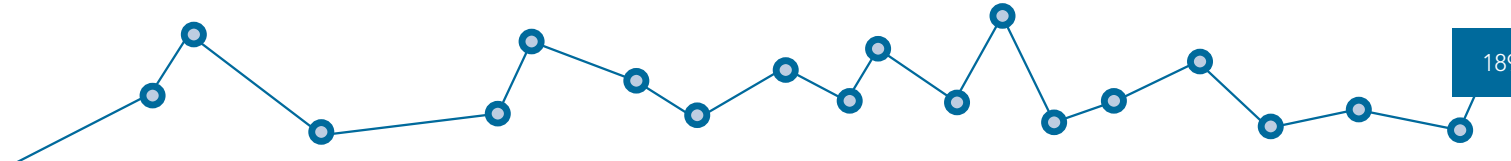
Met de formule wordt vervolgens bedoeld dat:

- het audit risico kan worden beperkt als gevolg van een
- verminderd inherent risico en/of een
- verminderd interne-controlerisico en/of een
- verminderd cijferanalyserisico en/of een
- verminderd steekproefrisico.

Meestal vindt er een herschikking van de formule plaats. Daarin is het audit risk gegeven (5%) en wordt het steekproefrisico uitgerekend. De tabellen van de verschillende grote accountantskantoren geven waarden uit de Poissonverdeling bij nul fouten in de steekproef weer in reeksen als bijvoorbeeld:

Kantoor	Reeks R-waarden
A:	3 2 1
B:	3 2 0,7
C:	3 2 0,5

Hoe groter het aandeel van de gegevensgerichte controles in de aanpak, hoe hoger de R-waarde die wordt gehanteerd. In de hierboven weergegeven reeksen is de waarde R=3 verbonden met de volledige steekproefomvang. Bij R=2 wordt de steekproefomvang met 1/3 gereduceerd als gevolg van een verminderd inherent risico. De laagste R-waarde wordt gehanteerd als de accountant steunt op de interne controle.



Het rekenwerk is dan eigenlijk al door het bureau vaktechniek uitgevoerd. De inschattingen van de verschillende risico's blijven dan voor de uitvoerende praktijk verborgen.

Deze reeksen lijken abacadabra, maar eigenlijk valt dat mee. Het is Amerikaanse pragmatiek om een beperking van de omvang van de gegevensgerichte controle aan te duiden. Die reductie is respectievelijk niets, een derde ten gevolge van het niet aanwezig zijn van een specifiek risico en respectievelijk ongeveer 67%, 77% en 83% als gevolg van het mogen steunen op de kwaliteit van de interne controle.

4.2 ARM en Bayes

Veenstra en Van Batenburg (1989) gaven aan dat het vermenigvuldigen van risico's zou kunnen worden voorkomen als gebruik zou worden gemaakt Bayesiaanse statistiek. Steele (1992) heeft eveneens Bayesiaanse statistiek als alternatief voor het vermenigvuldigen aangeboden. Hij presenteerde, zoals ik eerder al aangaf, onder meer een steekproef-equivalentenmethode. Van Batenburg en Dassen (1996) hebben een operationeel alternatief voor het dilemma van de vermenigvuldiging aangeboden. Dit alternatief vertaalt op elegante wijze het vermenigvuldigen in een variant die op dossierniveau uitspraken toelaat¹⁵.

Kort gezegd komt de methode neer op het schatten van een foutbovengrens. De accountant schat een foutbedrag in de populatie waarvan hij absoluut zeker is dat het niet overschreden zal wor-

den. Dit bedrag is input voor rekenwerk. Daaruit volgt dan tot welke reductie in de hoeveelheid gegevensgerichte werkzaamheden dat leidt. De kracht van het model is dat het eenzelfde reeks reductiefactoren kan opleveren (bijvoorbeeld 0, 33%, 77%) als het ARM (bij in casu de waarden $R=3$, $R=2$ en $R=0,7$).

Het Bayesiaanse model is ook transparant in die zin dat het er duidelijk over is dat alleen uitspraken over de omvang van de hoeveelheid gegevensgerichte controles worden gedaan. Niet over de controlewerkzaamheden in samenhang met bijvoorbeeld de onvervangbare interne controle.

Ik vind het jammer dat het huidige standaardwerk voor accountantscontrole, het boek van Knechel (2001), het vermenigvuldigen als vanzelfsprekend heeft gehandhaafd. Daarmee is Knechel voorbijgegaan aan de theorievorming¹⁶. Knechel negeert bovendien de historie van het wetenschappelijk onderzoek op het terrein van risicoanalyse in auditing (Bell 1995) door Statistical Audit en het Audit Risk Model niet in samenhang te behandelen.

5 Andere vormen van risicoanalyse

Ik heb gezien dat accountants het begrip 'risicoanalyse' niet eenduidig gebruiken. Zo gebruikt men vandaag de dag het begrip voor de analyse van bedrijfsrisico's, de analyse van IT-risico's, de analyse van de interne beheersing en ook

de analyse ten behoeve van de reductie van de gegevensgerichte werkzaamheden. Geen van die analyses is identiek aan een van de andere. Toch wordt dezelfde naam ervoor gebruikt. Ik ga daarom de verschillende soorten risicoanalyse langs.

5.1 Analyse van bedrijfsrisico's en ARM

Als een accountant de jaarrekening controleert beoordeelt hij het stelsel van maatregelen van interne beheersing. Het ligt daarom voor de hand dat hij kennis moet nemen van de analyse van de bedrijfsrisico's. Het ligt ook voor de hand dat hij de uitkomsten van dat onderzoek moet gebruiken voor de vaststelling of hij voldoende werk heeft verricht aan de voorwaardelijke controles. Verder zijn de mogelijke bedrijfsrisico's van belang voor de beoordeling van de schattingsposten in de jaarrekening zoals de waardering van duurzame productiemiddelen, voorzieningen en dergelijke.

De informatie die uit de analyse van de bedrijfsrisico's naar voren komt kan verder nog dienen als voorinformatie ten behoeve van de eventuele beperking van de omvang van de hoeveelheid gegevensgerichte controles. De uitkomsten kunnen dus voor het Audit Risk Model gezien worden als informatie over het Inherent Risico: als de informatie als 'gunstig' kan worden betiteld, is er een reductie van de hoeveelheid gegevensgerichte controles mogelijk. Dat betekent dus dat het ARM geen voeding biedt (en kan bieden) voor de weging van de invloed van bedrijfsrisico's op de jaarrekening, maar dat de analyse van bedrijfsrisico's

wel van invloed kan zijn op de manier waarop het accountantscontrole risico wordt gewogen.

5.2 Analyse IT-risico's en ARM

De analyse van IT-risico's speelt een belangrijke rol bij de beoordeling van de maatregelen van interne beheersing. In het bijzonder valt hier te noemen de rol die de IT kan vervullen bij de uitvoering van de primaire processen. In deze zin interfereren IT-risico's met bedrijfsrisico's. Een ander belangrijk aspect dat speelt in relatie tot de wereld van de IT is de controleerbaarheid. Zo gauw er sprake is van het gebruik van IT in relatie tot de bedrijfsvoering en de beheersing ervan is er ook sprake van de mogelijkheid tot het 'verdampen' van gegevens, en daardoor het niet meer achteraf controleerbaar zijn van een verantwoording. Er is dus sprake van een belangrijke rol voor IT om de onvervangbare interne controle te beheersen. De beoordeling van de IT, in samenhang met de beoordeling van de maatregelen van interne beheersing, kan voorinformatie leveren die kan leiden tot beperking van de hoeveelheid gegevensgerichte controles. Ook hier geldt dat de beoordeling van de IT wel het ARM kan voeden (met voorinformatie), maar het ARM biedt geen soelaas bij het beoordelen van de IT.

5.3 Analyse Interne Beheersing en ARM

Wat voor de beoordeling van de bedrijfs- en IT-risico's geldt, geldt ook voor de risico's met betrekking tot de interne beheersing. Met name de beoordeling van de onvervangbare interne controle, waarmee de controleerbaarheid (achteraf) van

de verantwoording moet worden gegarandeerd, kan voorinformatie ten behoeve van de gegevensgerichte controle bieden. Die beoordeling kan echter niet achterwege blijven of worden vermindert op basis van uit te voeren gegevensgerichte controles.

5.5 Voorwaardelijke controles en ARM

Het ligt volgens mij voor de hand dat de meeste niet-financiële audits gerelateerd zijn aan voorwaardelijke controles in het kader van de controle van een financiële verantwoording. Mollema (2003, 2004) stelt in zijn twee artikelen over risicoanalyse¹⁷ terecht dat als er geen sprake is van een financiële audit, het instrumentarium van geldsteekproeven (= monetary unit sampling) niet toepasbaar is.

Voor de analyse van bedrijfsrisico's, IT-risico's en risico's met betrekking tot de interne beheersing moeten bestaande modellen worden uitgebreid en eventueel nieuwe modellen worden ontwikkeld. Voor deze modellen heeft Mollema stof tot nadenken geboden. Een model dat is gebaseerd op geldsteekproeven is niet geldig, althans niet zonder meer.

6 Auditsufficiency en auditrisk

6.1 Inleiding

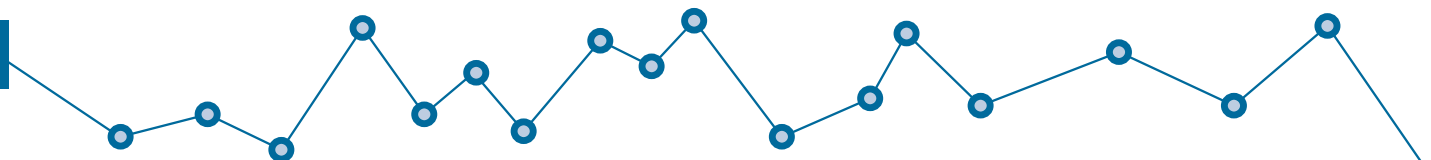
Auditors willen met zo weinig mogelijk werk tot een zo hoog mogelijke opbrengst komen. Een van de pijlers van die opbrengst is een oordeel over

een financiële verantwoording, dat op een deugdelijke grondslag is gebaseerd. Een auditor moet met andere woorden toereikend hebben gecontroleerd. Zo gauw het kwantitief wordt moet die toereikendheid dus in kwantiteiten kunnen worden uitgedrukt. Bij de financiële controle kunnen de hoeveelheden gegevensgericht werk worden uitgedrukt in te controleren geldeenheden. We zagen dat Bayesiaanse statistiek de mogelijkheid biedt om de voorinformatie uit andere werkzaamheden ook in geld, te controleren geldeenheden, uit te drukken. Mooi model, dus.

Maar een oordeel op deugdelijke grondslag omvat ook voorwaardelijke controles: dit zijn de volledigheidscntroles en de controles waarmee vastgesteld moet worden of de gegevensgerichte controles achteraf wel kunnen worden uitgevoerd. Dat betekent dat het geldsteekproefmodel het verdient te worden aangevuld met volledigheidscntroles, uitgedrukt in geld.

De vaststelling van de controleerbaarheid is helaas niet in financiële grootheden uit te drukken. Wellicht is er wel een andere manier te verzinnen om de toereikendheid van die controleactiviteiten te beschrijven.

Om dit hoofdstuk overzichtelijk te houden heb ik de controleactiviteiten in een workflow uitgebeeld. De flow volgt het beslissingsmodel van een financiële controle.

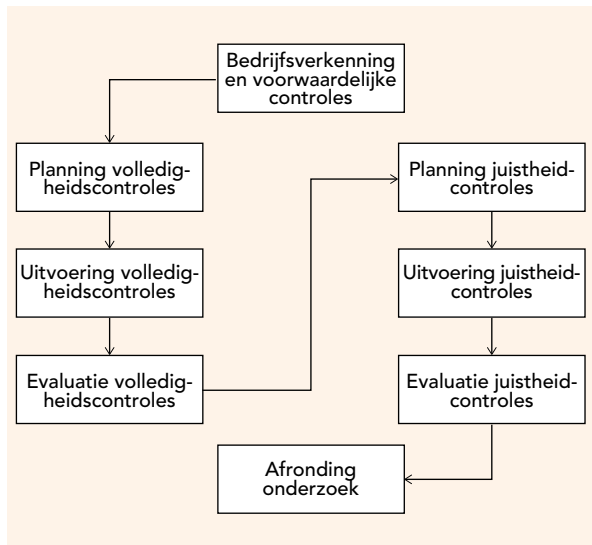


6.2 Tekening model financiële controle

Ik faseer een financiële controle als volgt:

- bedrijfsverkenning en voorwaardelijke controles¹⁸
- planning volledigheidscntroles
- uitvoering volledigheidscntroles
- evaluatie volledigheidscntroles
- planning juistheidscntroles
- uitvoering juistheidscntroles
- evaluatie juistheidscntroles
- totaal evaluatie en afronding.

Deze fasering is vooral besliskundig van aard. Dat is weer te geven met het volgende plaatje:



Met deze tekening probeer ik duidelijk te maken dat op het moment dat de planning van de juistheidscntroles plaatsvindt, er al vanuit verschillende invalshoeken naar de organisatie is gekeken. Een van de belangrijkste invalshoeken is het vaststellen dat de onvervangbare interne controle heeft gewerkt en tot een controleerbaar geheel heeft geleid. Een andere invalshoek is het uitvoeren van de volledigheidscntroles¹⁹. Als die activiteiten daadwerkelijk plaats hebben gevonden voordat de planning van de juistheidscntroles plaatsvindt kan de controle-informatie bij die planning worden meegewogen. Het wegen van die controle-informatie bij de planning van de omvang van de hoeveelheid gegevensgerichte controles is de risicoanalyse zoals die met behulp van het Audit Risk Model en de Bayesiaanse verbeteringen erop kan worden weergegeven.

6.3 Audit sufficiency en voorwaardelijke controles; volledigheid

De audit-riskmodellen waarin met Bayesiaanse statistiek (Steele 1992; Van Batenburg en Dassen 1996) is gewerkt, zijn gericht op de verklaring en de bepaling van de omvang van de gegevensgerichte werkzaamheden als onderdeel van een financiële controle. De modellen maken gebruik van de statistiek voor geldsteekproeven. De toereikendheid van de controle wordt gedefinieerd met behulp van materialiteit en steekproefrisico.

Materialiteit is dan een maatstaf voor de ergste populatie die nog op basis van de (gunstige) gegevens in de steekproef goedgekeurd zou kunnen

worden. Het steekproevenverhaal krijgt meestal vorm door in een verzameling (geldeenheden van) Ist-posities, van boekingen, te steken en daar de Soll-positie bij te zoeken. Meestal manifesteert die verzameling Soll-posities zich in de vorm van brongegevens. Laten we aannemen dat deze wijze van werken algemeen geldend is (of gemaakt kan worden).

Ik vind het vervolgens voor de hand liggen om de negatieve controles precies tegengesteld te definiëren. Dus de populatie waarin gestoken moet worden is de verzameling Soll-posities. Een Soll-positie wordt met de boeking, met de Ist-positie, vergeleken. Hier doen zich operationele problemen voor: de boekingen staan in beginsel de auditor ter beschikking in elektronische vorm. Daardoor zijn technieken voor het trekken van geldsteekproeven gemakkelijk toepasbaar te maken. Maar geldsteekproeven op Soll-posities, op brongegevens of erger nog: op directe waarnemingen: hoe komen we dan aan de bedragen? Zijn de gegevens wel beschikbaar voor trekkingen? Gelukkig staat de accountant een heel arsenaal aan wiskundige methoden en technieken ter beschikking waaruit hij kan kiezen. Als de geldsteekproef pur sang de beste oplossing is kan hij als next best oplossing een soort postensteekproef hanteren waarmee uitspraken in geld gedaan kunnen worden; bijvoorbeeld een uit de familie van de regressieschatter (Broeze en Kloosterman 2003). Een alternatief voor dergelijke steekproeven is het terugvallen op de handmatige methoden die indertijd de ontwikkeling van het

toepassen van geldsteekproeven nogal hebben geremd.

Concluderend:

De toereikendheid van de controle kan ook voor de financiële volledigheidscontroles worden gemodelleerd, analoog aan die voor de gegevensgerichte controles en inclusief de risicoanalyse volgens paragraaf 4 van dit artikel. Daarvoor kunnen accountants gebruik maken van de theoretische achtergrond die de theorie van geldsteekproeven biedt en die door het kennisgebied Statistical Audit wordt aangereikt om tot alternatieve oplossingen te komen.

6.4 Audit sufficiency en voorwaardelijke controles; controleerbaarheid

Hiervoor heb ik aangegeven dat er een hoeveelheid controlewerk is die, geredeneerd vanuit de gegevensgerichte controle, moet leiden tot de conclusie dat er gecontroleerd kan worden en die moet leiden tot de beslissing wat de minst kostenintensieve controleactiviteiten zijn om het doel van de goedkeuring te bereiken.

Dit werk is niet te beschrijven als het onderzoek van een zak geld, zoals hiervoor de gegevensgerichte controles en de financiële volledigheidscontroles zijn beschreven.

Mollema (2003, 2004) heeft weliswaar geprobeerd de verschillende soorten risicoanalyse op één noemer te brengen, maar is daar volgens mij (Kloosterman 2004) nog niet in geslaagd.

Om meerdere oplossingsrichtingen te vinden stel ik een paar vragen:

- Is het al dan niet achteraf controleerbaar zijn van een verzameling Ist-posities niet de hamvraag?
- Deze vraag wordt toch beantwoord met het vaststellen van het al dan niet beschikbaar hebben van data met betrekking tot de Soll-posities?
- Het antwoord op de vraag of alle Soll-posities binnen het bereik van de controle liggen kan (slechts) met een plausibiliteitsaanduiding worden gegeven?
- Is het mogelijk dergelijke plausibiliteitsantwoorden met een benadering in te vullen, die door middel van ervaringen te valideren is?

Simplistisch gesteld is controle het matchen van paren Ist- en Soll-posities. Daarom spreekt het voor zich dat de Ist-posities beschikbaar zijn. Voor de uitvoerbaarheid van de controle spreekt ook voor zich dat de Soll-posities beschikbaar zijn. Als de controle achteraf plaatsvindt moeten ook de paren Soll- en Ist-posities bewaard zijn gebleven. Laten we deze Soll-posities voor het gemak brongegevens noemen. Aan welke eigenschappen moeten dan die brongegevens voldoen? In de eerste plaats moet de echtheid (de herkomst is juist, en dat is vast te stellen) vaststaan. In de tweede plaats moet de betekenis van de gegevens juist geconserveerd zijn (de integriteit van de data is niet aangetast). Verder moet de interpretatie die eventueel op de brongegevens volgde navolgbaar zijn (bijvoorbeeld door middel van documentatie).

Belangrijke bottleneck worden nu die brongegevens. Waarom staat vast dat zij de 'wijsheid in pacht' hebben? Eigenlijk zijn die brongegevens de primaire vastleggingen van de transacties (en delen ervan) die binnen de gecontroleerde onderneming plaatsvonden. Zij zijn daarmee eigenlijk een afgeleide²⁰ van de werkelijkheid. Dit betekent dat indirecte controle kan plaatsvinden bij de gratie van kwalitatief goede brongegevens. Vaststellen of die brongegevens goed zijn kan plaatsvinden door de transacties in de werkelijkheid zelf waar te nemen en met de brongegevens te vergelijken. In dat geval is de werkelijkheid de Soll-positie en zijn de brongegevens de Ist-positie. Deze controleactiviteit, wie hem daadwerkelijk uitvoert laat ik hier even in het midden, zorgt er dus voor dat de brongegevens tot Soll-positie kunnen worden opgewaardeerd.

Hoeveel waarnemingen van deze klasse zouden er moeten plaatsvinden? Wat is het karakter van die waarnemingen? Mag ik hier ook zo'n soort Bayesiaanse vertaling voor toepassen?

Naar mijn overtuiging brengt een dergelijke invalshoek een soort risicoanalyse binnen bereik die betrekking heeft op werkzaamheden die niet in te controleren geldeenheden is uit te drukken. De vraag hoeveel voorwaardelijke controles er moeten plaatsvinden is daarmee teruggebracht tot een kostenvraagstuk. Dit formuleer ik als: hoeveel tijd wordt beschikbaar gesteld om de stelling te proberen te falsifiëren dat "de brongegevens zijn aan te merken als Soll-positie". Deze werkwijze

is analoog aan de benadering van Popper (1979). Deze benadering komt erop neer dat om het waarheidsgehalte van een stelling te toetsen we zoveel waarnemingen kunnen doen als wij nodig vinden, of waar wij op dat moment geld voor over hebben. Is de stelling dan (nog) niet gefalsificeerd, dan merken we hem als valide aan. Let wel, de hoeveelheid werk die wij nodig vinden of 'het maatschappelijk verkeer' nodig vindt (vaak: onze klant over heeft) voor die toetsingsactiviteit is subjectief. Eventueel is het mogelijk over die subjectiviteit collectieve afspraken te maken (b.v. in gedragslijnen voor de controle). En zo kunnen we komen tot de essentie van risicoanalyse in een audit. En dat is het geven van een antwoord op de vraag: "Wanneer heb ik voldoende werk gedaan om mijn goedkeurend oordeel te onderbouwen?"

7 Samenvatting

Dit artikel stelt dat het aggregeren van audit evidence audit risk met zich meebrengt. Het overzicht van de ontstaansgeschiedenis van het Audit Risk Model laat zien hoe audit risk in termen van Statistical Audit is beschreven. Daarmee brengt het artikel ook een dilemma van de audit risk modellen naar voren: de verschillende soorten controle-informatie zijn niet onderling uitwisselbaar. Om dit te onderbouwen is de financiële controle in stukken geknipt. De informatie uit bedrijfsverkenning, voorwaardelijke controles, volledigheidscntroles en juistheidscontroles (ge-

gevensgerichte controles) is allemaal afzonderlijk weergegeven.

Het artikel laat vervolgens het Audit Risk Model in de vermenigvuldigingsvariant en een Bayesiaanse variant zien. Het laat zien dat de modellen alleen maar spelen met de hoeveelheid gegevensgerichte controles. De andere vormen van risicoanalyse zoals de analyse van bedrijfsrisico's, IT-risico's en de risico's met betrekking tot de interne beheersing leveren wel informatie aan het ARM en zijn Bayesiaanse varianten, maar kunnen niet met die modellen worden beschreven. Tot slot is geprobeerd het model voor de sufficiency van de financiële audit zo algemeen mogelijk weer te geven, zodat daarvan in de toekomst een model voor sufficiency voor andere audits af te leiden valt.

Literatuur

Batenburg, P.C. van, R.J.M. Dassen (1996), Het Bayesiaanse model van Deloitte & Touche ter ondersteuning van de controlemix, *De Accountant*, jg. 103, no. 1, september, pp. 31-35.

Bell, T.B. A.M. Wright (eds) (1995), *Auditing, Practice, Research and Education*, American Institute of Certified Public Accountants, New York.

Blokdijk, J.H. (1996), Toedeling van de controletolerantie bij de accountantscontrole, *Maandblad voor Accountancy en Bedrijfseconomie*, jg. 70, no. 7/8, juli/augustus, pp. 350-360.

Broeze, G.B., H.H.W. Kloosterman (2003), *Statistical audit: overzicht van onderzoek over tien jaar*; Paper uitgereikt bij het onderbrengen van de Stuurgroep Statistical Auditing bij het Center voor Auditresearch Nivra-Nyenrode (CANN), januari.

Diekman, P.A.M. (1995), Toleranties in de accountantscontrole, *De Accountant*, jg. 101, no. 10, juni, pp. 665-668

Kloosterman, H.H.W. (1991), Schillenmodel, *De Accountant*, jg. 97, no. 7, maart, pp. 403-406.

Kloosterman, H.H.W. (1996), Intoleranties in de accountantscontrole, *De Accountant*, jg. 102, no. 8, april, pp. 534-537.

Kloosterman, H.H.W. (2004), Risicoanalyse in auditing, *Maandblad voor Accountancy en Bedrijfseconomie*, jg. 78, no. 9, september, pp. (412-416)

Knechel, W.R. (2001), *Auditing: Assurance & Risk*; 2nd edition, South-Western College Publishing, Cincinnati.

Koninklijk Nivra (2003), *Richtlijnen voor de Accountantscontrole*, Koninklijk Nivra, Amsterdam, NOVAA, Den Haag.

Leslie, D.A., A.D. Teitlebaum en R.A. Anderson (1980), *Monetary Unit Sampling, A practical guide for auditors*, Canadian Institute of Chartered Accountants, Toronto.

Mollema, K.Y. (2003), Auditrisico, meer dan ooit een issue (1), *Maandblad voor Accountancy en Bedrijfseconomie*, jg. 77, no. 12, december, pp. 551-556.

Mollema, K.Y. (2004), Auditrisico, meer dan ooit een issue (2), *Maandblad voor Accountancy en Bedrijfseconomie*, jg. 78, no. 1/2, januari/februari, pp. 5-15.

Mooijekind, M.J.Th. (1991), Het risico-analysemodel voor de accountant een (belangrijke) stap vooruit?, *De Accountant*, jg. 97, no. 11, juli/augustus, pp. 720-723.

Popper, K.R. (1979), *The Growth of Scientific Knowledge*; Klostermann Texte Philosophie, Frankfurt am Main.

Schilder, A. (1991), Risicoanalyse: rekenmodel, denkmodel of wedstrijdmodel?, *De Accountant*, jg. 97, no. 9, mei, pp. 573-576.

Steele, A. (1992), *Audit Risk and Audit Evidence*, Academic Press LTD, London.

Stringer, K.W. en T.R. Stewart (1996), *Statistical Techniques for Analytical Review in auditing*, 2nd edition, John Wiley and Sons Inc., New York.

Veenstra, R.H., P.C. van Batenburg (1989), Een doorbraak in steekproeftoepassingen door Bayesiaanse statistiek, *De Accountant*, jg. 95, no 11, pp. 561-564.

Veenstra, R.H., P.C. van Batenburg (1991),
Bayesiaanse steekproefcontrole op ernstige
fouten, *De Accountant*, jg. 97, no 10, pp. 641-645.

Wilschut, K.P.G. (1989), Het denkmodel achter de
risicoanalyse in de accountantscontrole,
De Accountant, jg. 96, no. 2, oktober 1989, pp. 86-91.

Wolde, J. ten (1990), Het controlemixmodel,
De Accountant, jg. 96, no 11, pp. 561-564.

Wonnacott, Thomas H., Ronald J. Wonnacott,
(1977), *Introductory Statistics for Business and
Economics*, 2nd Ed., John Wiley and Sons,
Inc., New York.

Noot

- 1 Dit artikel is eerder verschenen in het MAB 2004, pp. 570-578.
- 2 Ondanks de aanscherping van risicoanalyse in hoofdstuk 100 van de Richtlijnen voor de Accountantscontrole 100 (Nivra 2002) is er m.i. geen inhoudelijke vernieuwing opgetreden. Het opdelen van het opdrachtrisico (conform audit risk) in inherent risico, interne-beheersingsrisico en ontdekkingsrisico laat geen paradigmashift zien.
- 3 Wat Bayesiaanse statistiek is, is niet in een of twee zinnen te vatten. In paragraaf twee wordt het effect ervan getoond. Zo vertaalt Steele het “zeker weten dat de interne controle goed heeft gewerkt” in een steekprofequivalent. Het gebruik maken van dit “zeker weten” wordt met behulp van statistiek gemodelleerd. Deze statistiek maakt onder meer gebruik van de regel van Bayes (Wonnacott en Wonnacott, 1977; hoofdstukken 19 en 20).
- 4 Vaak weergegeven met behulp van de volgende formule:
$$AR = IR \times CR \times ARR \times DR$$

Waarbij:
AR = Audit Risk
IR = Inherent Risk
CR = Control Risk
ARR = Analytical Review Risk
DR = Detection Risk (wat gelijkgesteld wordt aan Sample Risk).
- 5 De term gegevensgerichte controle is een homoniem. Vaak wordt de term gebruikt om uitdrukking te geven aan de nadruk die in de controle wordt gelegd op detailcontroles, de controles die in de Angelsaksische literatuur worden aangeduid als ‘substantive tests’. Ook wordt de term gegevensgerichte controle gebruikt als de controleactiviteit zelf, als de substantieve test. In deze betekenis gebruik ik hier de term ‘gegevensgerichte controles’.
- 6 Het gaat dan om steekproeven op geldeenheden. In het Nederlands spreken we van geldsteekproeven, in het Engels van Monetary Unit Sampling.
- 7 Veenstra en Van Batenburg (1989) zwengelden een discussie aan. Wilschut (1989) gooide de knuppel in het hoenderhok. Veel schrijvers klommen daarop in de pen. Ondermeer Ten Wolde (1990), Mooijekind (1991), Kloosterman (1991) en Schilder (1991). Naar mijn indruk hebben Van Batenburg en Dassen (1996) de discussie netjes afgesloten.
- 8 Dat kan immers niet, althans niet voor die ene precieze, bijvoorbeeld volgende, dag. Men kan wel een opmerking maken over de groep dagen waartoe de dag van morgen ook hoort. Wat handiger is: als je dan toch iets zegt over het regenen van morgen: schat wat er morgen maximaal aan millimeters regen kan vallen, die hoeveelheid valt achteraf te meten. Dat is bij een controle niet anders. We willen iets weten over die ene en slechts die ene gecontroleerde. Of hij deel uitmaakt van een groep gecontroleerden waarbij het in 50% van de gevallen best wel goed gaat is voor de handtekening helemaal niet relevant. Het is van belang te realiseren dat als een vooraf geschatte kans niet kan worden waargenomen, er geen mogelijkheid bestaat om uit de ervaring te leren.
- 9 De uitdrukking geldsteekproeven is synoniem met Monetary Unit Sampling. Voor de komst van de euro werd de geldsteekproef aangeduid als guldensteekproef. In het geval van een wiskundige steekproef heeft ieder element uit de te beoordelen populatie dezelfde trekkingskans. In het geval van een geldsteekproef is de populatie een verzameling geldeenheden en heeft iedere geldeenheid dezelfde kans om in de steekproef te worden betrokken.
- 10 In Bell en Wright (1995) is in het hoofdstuk Risk Orientation een uitgebreid overzicht opgenomen van de historie van het Audit Risk Model.
- 11 Het toepassen van steekproeven komt in de plaats van het integraal uitvoeren van detailcontroles. De beschrijving van steekproeven in de accountantscontroles is daarmee doorgaans gericht op de

zogenoemde 'substantive tests'. De termen 'gegevensgerichte controles' en 'substantive tests' of 'substantive procedures' zijn synoniem.

- 12 In Bell en Wright (1995) op bladzijde 23 is een voorbeeld opgenomen van de invloed van het toepassen van steekproeven op de controleaanpak bij Haskins & Sells, onder invloed van de pionier K.W. Stringer. En dat al in 1963.
- 13 De cijferanalysemethode met behulp van statistiek is uitgewerkt in Stringer en Stewart (1996). Bell & Wright (1995) noemen de toevoeging van het cijferanalyse-risico aan het Audit Risk Model op bladzijde 37.
- 14 Deze gedachtegang is mede gebaseerd op het niet hoeven toedelen van controle toleranties (Blokdijk 1996).
- 15 Ik bedoel hier falsifieerbaar zoals Popper (1979) dat aanduidt. Het waarheidsgehalte van een stelling moet onderzocht kunnen worden. Falsifiëren is dan voldoende bewijs leveren dat de stelling klopt én één (en slechts één) tegenbewijs is voldoende om de stelling onderuit te halen. In dit geval: de verschillende soorten risico's in termen van kansen kunnen wel worden geschat, maar niet per onderzoek worden gefalsificeerd. De fout in de populatie kan daarentegen zowel geschat als gemeten worden; daardoor is zo'n methode wel falsifieerbaar. Het gebruik van bijvoorbeeld steekproeven is bedoeld om zo objectief mogelijk te laten zien dat er toereikend is geprobeerd de stelling onderuit te schoffelen. De methode van Van Batenburg en Dassen voldoet aan die eis.
- 16 Knechel gaat in hoofdstuk 3 in op 'risk, evidence and materiality'. In hoofdstuk 16 zegt hij pas iets over 'Statistical Evidence'. Hij brengt beide gebieden, risicoanalyse en steekproeftheorie, niet met elkaar in verband.
- 17 Mollema stelt dit impliciet. Hij zegt dat het Audit Risk Model niet toepasbaar is. Daarom stelt hij voor om de analyse van bedrijfsrisico's en IT-risico's met een model uit de verzekeringswiskunde aan te pakken. Mits een auditor zich dan houdt aan de vereiste dat de conclusies op 'multiklantniveau' zijn mag dat. Wil iemand conclusies trekken op uniklantniveau dan klopt het model niet meer.
- 18 Bedrijfsverkenning houdt onder meer in: te weten komen hoe het bedrijfsgebeuren is vormgegeven, welke rol de informatietechnologie in het bedrijf speelt en hoe de administratieve context is. Voorwaardelijke controles zijn controles waarbij vastgesteld wordt of en in hoeverre aan basale voorwaarden van controleerbaarheid is voldaan. De uitkomst ervan moet zijn vast te stellen of en zo ja, welke beperkingen het onderzoek met zich brengt. De controle op de volledigheid van de vastgelegde gegevenssoorten is een van de voorwaardelijke controles.
- 19 Een deel van de controles op de volledigheid van de transacties heeft te maken met de onvervangbare interne controle. Deze controles kunnen geacht worden deel uit te maken van de voorwaardelijke controles. Ik heb het onderscheid gemaakt om te laten zien dat er controleactiviteiten zijn die op andere objecten gericht zijn, maar misschien wel tegelijkertijd uitgevoerd zouden kunnen worden met de juistheidscontroles.
- 20 Eigenlijk de eerste afgeleide. Deze gegevens bevatten het meest puur een weerspiegeling van de echte wereld. Ze bieden de mogelijkheid verantwoording af te leggen.

Doeltreffende administratieve controle door steekproeven¹

Ruud Veenstra en Arnold Heertje

Samenvatting

Administratieve controle door middel van steekproeven kan de doeltreffendheid, doelmatigheid en beheersing van waardenstromen en informatiestromen aanzienlijk verbeteren en de onaantastbaarheid van de verslaglegging daarover garanderen. Dit artikel draagt de bouwstenen daarvoor aan. Met deze bouwstenen kunnen de essentiële voorwaarden worden vervuld voor herstel van vertrouwen in de integriteit van het financieel-economische verkeer. Het Limpergiaanse concept van de volkomen controle en daarop gebaseerde steekproefmethoden voor accountantscontrole bieden hiervoor een stevige grondslag. Daarnaast is een belangrijke doorbraak bereikt door administratieve controle gebaseerd op de voor de industriële kwaliteitscontrole ontworpen AOQL-steekproefmethode (AOQL = Average Outgoing Quality Limit).

In dit artikel blijven steekproefmethoden die op het Audit Risk Model zijn gebaseerd buiten beschouwing. Het Audit Risk Model plaatst de risico's van de accountantscontrole centraal en leidt daardoor niet tot een vastliggend oordeel over het object van onderzoek, maar tot een van de inge-

schatte 'audit risk' afgeleid oordeel. Op dat model gebaseerde steekproeven kunnen dan ook niet als bouwsteen worden beschouwd.

1 Inleiding

Om het vertrouwen van het maatschappelijk verkeer te herstellen na de grote boekhoudaffaires, moet de door Limperg geïntroduceerde vertrouwensnorm opnieuw worden gedefinieerd. Hierop wijst Wallage (2005, p. 130): *"Juist nu in een turbulent tijdsgewricht het vertrouwen in het accountantsberoep relatief laag is, moet elke externe accountant zich doorlopend realiseren dat de functie gericht is op het vervullen van de verwachtingen van het maatschappelijk verkeer. Zo niet, dan is het bestaansrecht in gevaar."*

Wallage (2005, p. 128) citeert de Chief Auditor van de Amerikaanse toezichthouder op het accountantsberoep (Public Company Accounting Oversight Board: PCAOB), Doug Carmichael: *"The common principle for the practice of auditing that Limperg postulated is now embedded in the mechanism for developing professional standards."* *"This proposed standard involves something much more than additio-*

nal information being covered by the auditor's report. It is Limperg's concept related to expansion of the audit function in action."

In Nederland zijn vanuit de Limpergiaanse vertrouwensnorm krachtige en goed toepasbare steekproefmethoden ontwikkeld in het kader van de accountantscontrole van grote controleobjecten. Deze kunnen dienen voor het herdefiniëren van de vertrouwensnorm. In ieder geval kan de interne controle aanzienlijk verbeterd worden door toepassing van steekproeven in de administratie door middel van Average Outgoing Quality Limit (hierna AOQL).

Ironisch genoeg speelde Ahold daarin een voor-
aanstaande rol in het verleden (Veenstra en Kriens 1982, p. 252; Kriens en Veenstra 1985, p. 386-389). Deze steekproefmethode kan in alle omvangrijke administratieve processen en geldstromen een aanzienlijke bijdrage leveren aan het herstel van de objectiviteit en de effectiviteit van het beheer en de beheersing van de administratieve processen (Veenstra 1988, p. 72-75). Juist de objectiviteit, inzichtelijkheid, valideerbaarheid en (bij goede documentatie zelfs) identieke herhaalbaarheid die wordt ingebouwd, kan een krachtige impuls geven aan het noodzakelijke herstel van het op objectieve gronden rustende vertrouwen in de integriteit van de financiële verslaglegging en de integriteit van de besteding en allocatie van de financiële middelen in zowel het bedrijfsleven als in de overheids- en semi-overheidssfeer. Het AOQL-steekproefstelsel en de uitkomsten van het steekproefsgewijze onderzoek verbinden op

objectieve en overzichtelijke wijze het inzicht in het proces op individueel en transactieniveau met de beleidsuitgangspunten én de uitkomsten van het beleid. Op grond van de uitkomsten kan op ieder gewenst moment tijdige en nauwkeurige bijsturing van het beleid plaatsvinden.

In paragraaf 4 wordt uitvoerig ingegaan op de mogelijkheden van AOQL als steekproefmethode voor administratieve kwaliteitscontrole. Voor een juiste plaatsbepaling ten opzichte van de accountantscontrole wordt in paragraaf 2 ingegaan op de (toekomstige) betekenis van het klassieke concept van de volkomen controle en in paragraaf 3 op het Audit Risk Model waarop de huidige richtlijnen voor de accountantscontrole in Nederland zijn gebaseerd.

In paragraaf 5 worden de steekproefmethoden voor de volkomen controle besproken. In paragraaf 6 volgt de conclusie.

2 Volkomen controle

In de vroege Nederlandse controlepraktijk en controletheorie wordt uitgegaan van een stelsel van omspannende verbandscontroles en totalencontroles, die door gerichte en slimme detailcontroles te verifiëren zijn (Kleerekoper 1933). Kleerekoper bouwt onder meer voort op eerdere pogingen van Schmalenbach (1911/1912) om de afweging tussen volledige controle en gebruik van steekproeven wetenschappelijk te baseren op de waarschijnlijkheidslere en op Van Klein (1911/1912) die deze

uitgangspunten van Schmalenbach uitwerkt in zijn artikel 'Über die Wahrscheinlichkeit der Entdeckung von Fehlern bei Revisionen' (Kleerekoper 1933, p. 76-79). Een revolutionaire verbetering van de efficiëntie van de detailcontroles wordt bereikt door de originele statistische analyse van De Wolf (1956 1959) en van Van Heerden (1961). Door de nadrukkelijk door Kleerekoper getrokken conclusie, dat *"er geen enkele methode [is] aan te wijzen die ons de eenvoudige steekproef leert bepalen"* en dat de steekproef een *"hap hazard method"* is *"die op intuïtie of op gevoel of iets dergelijks berust"* (t.a.p. p. 125), laat de officiële acceptatie van steekproeven in de accountantscontrole door het Nederlands Instituut van Registeraccountants (NIVRA) nog tot 1982 (!) op zich wachten (NIVRA 1974, 1982).

In hedendaagse termen kan 'volkomen controle' worden gedefinieerd als de (accountants)controle die de onderneming in al zijn facetten centraal stelt en zich fundamenteel richt op de vaststelling dóór de auditor dat alle bedrijfseconomische verbanden gedurende de gehele te controleren periode effectief zijn gehandhaafd. Deze bedrijfseconomische verbanden tussen inkomende en uitgaande goederen- en dienstenstromen en de daaraan gerelateerde inkomende en uitgaande kasstromen zijn naar hun aard volkomen (Veenstra 1972, 1994; Elsas 1996; Veenstra 1998, 2000a, 2000b, 2005; De Backer en Snoeck 2006). Dat de bedrijfseconomische verbanden zijn gehandhaafd stelt de auditor vast door omspannende verbandcontroles, totalencontroles en gerichte verificatie van de primaire bedrijfsgegevens op transactieniveau

gebruikmakend van wiskundige steekproeftechnieken en auditsoftware (Veenstra 1981).

Elsas (1996) heeft aangetoond dat er sprake is van een coherent model dat ook consistent en verifieerbaar is alsmede, in moderne termen, ook 'computational' te beschrijven en te ondersteunen is, indien in het Value Cycle System ook de waardensprong (brutowinst) mede in het controleconcept wordt opgenomen (Veenstra 1972, p. 48-49; 1994, p. 23-26). In zijn proefschrift heeft Elsas het controleconcept volledig wiskundig geformuleerd met gebruikmaking van Petrinetten (Petri 1962)². In overeenstemming met de bevindingen van Elsas en daarop aansluitend, concluderen De Backer en Snoeck (2006 p. 10-15) dat de door Petri ontwikkelde modelleringstechniek ook effectief en verifieerbaar gebruikt kan worden voor 'tool-onafhankelijke' overdraagbaarheid van bedrijfsproces-modellen over de bedrijfsgrenzen heen en dat daaraan in het kader van de huidige ontwikkelingen, zoals de opkomst van webservices, als veelbelovende technologie voor de implementatie van gedistribueerde bedrijfsprocessen³ grote toekomstige betekenis moet worden toegekend. Vast staat derhalve dat, ook nu de techniek voortschrijdt, er sprake is van een algemeen toepasbaar, coherent, consistent en verifieerbaar controleconcept met een veelbelovende toekomst.

3 Het Audit Risk Model

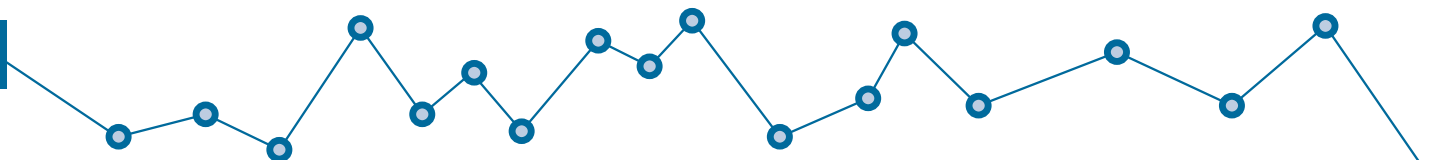
Ook in de VS wordt het belang van het Limpergi-aanse vertrouwensbeginsel onderschreven (Wallage 2005) en zijn deugdelijke steekproefmethoden ontwikkeld (Wilburn 1984). Toch zijn in de VS de subjectieve methoden niet helemaal uitgebannen. Een voorbeeld van een subjectieve methode is het uit de VS afkomstige Audit Risk Model uit 1981 (American Institute of Certified Public Accountants 1981, Appendix, U.S. Auditing Standards section 350 Audit Sampling). Uitsluitend omdat de huidige richtlijnen voor de accountantscontrole in Nederland zijn gebaseerd op dit Amerikaanse Audit Risk Model wordt hier wat dieper op het onderwerp ingegaan.

In de VS wordt het Audit Risk Model in bovengenoemde Appendix gedoogd (t.a.p. paragraaf 4): *“The model is not intended to be a mathematical formula...; however, some auditors find such a model to be useful...”*.

Dit model maakt het de accountant mogelijk om van detailcontroles af te zien op grond van a priori aannames ten aanzien van door de accountant veronderstelde lage risico's, te weten: Audit Risk (AR), Inherent Risk (IR), Control Risk (CR) en Risk of Analytical Procedures (AP). Bij een lage inschatting van deze risico's kan de accountant de jaarrekening goedkeuren zonder enige detailcontrole uit te voeren. De jaarrekening en – zelfs – mogelijk wel uitgevoerde verbandcontroles blijven daarmee in de lucht hangen. Het verband met de onderliggende werkelijkheid wordt door

de accountant slechts a priori verondersteld maar ten onrechte niet geverifieerd. Door het ontbreken van objectief geverifieerde (en ook niet te verifiëren) parameters (AR, IR, CR en AP) kan het model ook overigens geen objectieve uitkomst opleveren voor het accountantsoordeel over de jaarrekening of andere verantwoording. In het beste geval is het oordeel dat de accountant bij toepassing van dit model verkrijgt een oordeel over het audit risk en een daarvan afgeleid ongewis oordeel over de jaarrekening.

Toepassing van het model leidt niet tot een objectief en gekwantificeerd oordeel over de aan controle onderworpen informatiestromen en waardenstromen (Veenstra 1998 p. 14; 2000a, p. 74-77; 2000b, p. 77). Ook aan de op grond van dit model ontworpen steekproeven mankeert de objectieve grondslag. Dit model geeft in de praktijk dan ook, mede door de druk op de declaraties voor controlewerkzaamheden, ruim baan aan subjectief gewenste uitkomsten en legitimeert slechte praktijk, doordat de uitkomsten van het model niet valideerbaar zijn. Voor een bespreking van de praktijkproblemen die zich voordoen bij de Nederlandse toepassing van het Audit Risico Model wordt verwezen naar Eimers (2006). De support van het NIVRA voor het bedoelde model en de daarop gebaseerde tekst van de accountantsverklaring, belemmert het toepassen van objectieve controlemethoden, waaronder steekproeven in de accountantscontrolepraktijk en de administratieve praktijk. In de visie van de auteurs verdient het vanuit het vertrouwensbeginsel



de voorkeur om de focus van de praktijk uitdrukkelijk te richten op objectieve controlemethoden.

4 AOQL-steekproefmethode voor administratieve kwaliteitscontrole

De AOQL-steekproefmethode is ontwikkeld voor gebruik voor industriële kwaliteitscontrole bij Bell Telephone System en voor het eerst gepubliceerd in 1941, in e Bell System Technical Journal. AOQL is de afkorting van Average Outgoing Quality Limit. De methode stelt een grens aan de gemiddelde kwaliteit van de gecontroleerde subpopulaties. Dat de totale populatie ná controle, ongeacht de kwaliteit voorafgaande aan de controle, voldoet aan de voorafgaande aan de controle gestelde kwaliteitseis, is gewaarborgd doordat alle in de uitgevoerde substeekproeven aangetroffen fouten worden verbeterd. Veenstra en Kriens (1982, 1985) hebben deze steekproefmethode getransformeerd naar de administratieve kwaliteitscontrole. Voor uitgebreide tabellen wordt verwezen naar Dodge en Romig (1959). Zij hebben dertien tabellen samengesteld voor een AOQL variërend van 1% tot en met 10%. Afhankelijk van de eis die gesteld wordt, kan voor iedere situatie een optimale keuze worden gemaakt. Voor administratieve toepassingen is een AOQL van 1% veelal een acceptabele eis. In Tabel 1 (zie p. 207) zijn de waarden van Dodge en Romig dan ook overgenomen voor de eis dat de totale populatie ná controle maximaal 1% fouten bevat.

De werkwijze voor administratieve kwaliteitscontrole is als volgt.

1. Er wordt vastgesteld welk totaalbedrag (bijvoorbeeld: jaarbedrag) aan fouten de leiding van de organisatie ná de steekproefsgewijze kwaliteitscontrole maximaal aanvaardbaar acht.
2. Uit dit jaarbedrag en het verwachte jaartotaal volgt de fractie fouten (p) die na controle maximaal in de jaarpopulatie mag overblijven. In dit voorbeeld en in Tabel 1 is dit 0,01, dus 1%.
3. De totale populatie wordt verdeeld in subpopulaties. Bijvoorbeeld: een jaarpopulatie wordt verdeeld in 52 weekpopulaties.
4. Uit iedere weekpopulatie (met een omvang tot $N=100.000$) wordt aselekt bijvoorbeeld een steekproef van $n=85$ stuks getrokken, waarbij een goedkeurgrens hoort van $k_0=1$ (zie Tabel 1, laatste regel, eerste drie kolommen). Is de fractie fouten in de jaarpopulatie vóór controle naar verwachting groot of sterk wisselend, dan kan in de tabel meer naar rechts een meer optimale keuze worden gedaan voor $n=135$ en $k=2$, respectievelijk $n=255/k=4$, $n=445/k=7$, $n=790/k=12$ of $n=1520/k=22$.
5. Wordt in een weeksteekproef niet meer dan k fouten gevonden (k_0 =goedkeurgrens) dan wordt de weekpopulatie van die week goedgekeurd en worden de in de steekproef gevonden fouten verbeterd.
6. Wordt in een weeksteekproef meer dan k_0 fouten aangetroffen, dan wordt (uitsluitend) deze week afgekeurd en vervolgens volledig

gecontroleerd. De aangetroffen fouten worden verbeterd zowel van de weeksteekproef als van de volledig gecontroleerde weekpopulatie.

7. Door dit consequent te doen is zeker dat de jaarpopulatie voldoet aan de door de leiding van de organisatie gestelde eis. In dit voorbeeld bevat de jaarpopulatie na uitvoering van de administratieve kwaliteitscontrole maximaal 1% fouten. Het bereikte efficiencyvoordeel is afhankelijk van de hoogte van de fractie fouten vóór controle in de jaarpopulatie en de verdeling van de fouten vóór controle over de 52 weekpopulaties. Hoe minder weekpopulaties in eerste instantie moeten worden afgekeurd (en dus daarna volledig moeten worden gecontroleerd) des te groter is het efficiencyvoordeel dat met de methode wordt behaald.

Toelichting tabel 1:

p_m = fractie fouten die na controle maximaal in de jaarpopulatie mag overblijven

N = aantal elementen in de subpopulatie

n = steekproefomvang per subpopulatie

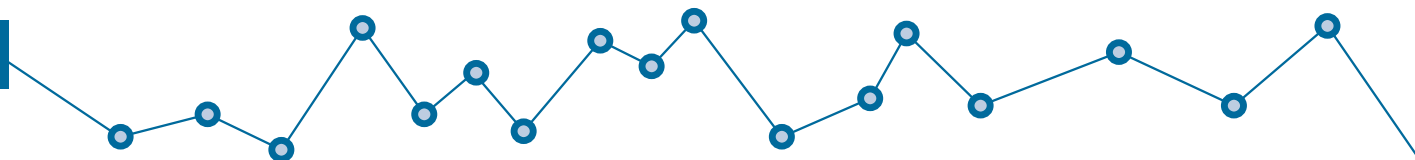
k_0 = goedkeurgrens per subpopulatie

p = fractie fouten in de subpopulatie vóór controle

Deze tabel is overgenomen uit Dodge en Romig (1959). Deze tabel geeft de waarden van n (steekproefomvang per subpopulatie) en k_0 (goedkeurgrens) aan indien de eis is dat de jaarpopulatie na controle maximaal 1% fouten mag bevatten.

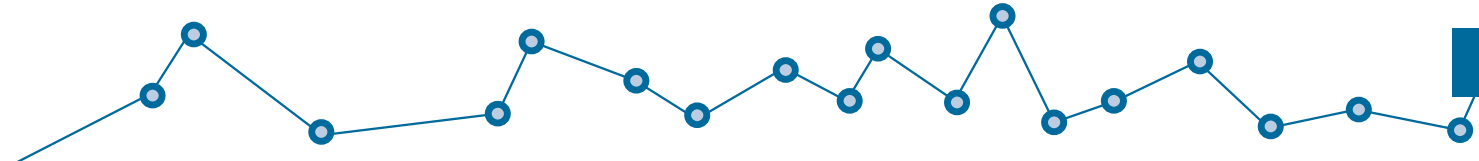
Inmiddels zijn enkele statistische verbeteringen aangebracht (Veenstra en Buijsse 1985, Van Batenburg et al. 1988; Van Batenburg en Kriens 1988, Talens 2005). Het blijkt dat de berekeningen van Dodge en Romig bij kleinere subpopulaties (N) tot iets te lage steekproeven (n) leiden. Voor toepassingen bij kleine subpopulaties (N) is het daarom wel raadzaam om genoemde literatuur daarover te raadplegen. Voor subpopulaties $N > 10.000$, zoals in bovenstaand voorbeeld waarin $N = 100.000$, heeft dit geen invloed op de opzet, uitvoering en uitkomsten van het AOQL-kwaliteitscontrolesysteem. Heeft men een 'exacte' verwachting over het percentage fouten vóór controle, hetgeen meestal niet het geval zal zijn, dan kan de toepassing preciezer geoptimaliseerd worden dan Tabel 1 en de overige tabellen van Dodge en Romig aangeven. De bij het artikel van Veenstra en Buijsse (1985) opgenomen tabellen zijn nauwkeuriger berekend en geven optimale combinaties van n , k_0 (steekproefomvang, goedkeurgrens) voor maximale percentages fouten ná controle van 0,1%, 0,5%, 1,0% en 2,0% voor N tot 100.000.

Kriens en Dekkers (1979 p. 117-127) beschrijven de AOQL-methode in het kader van accountantscontrole. Veenstra en Kriens (1982, 1984, 1985) kwamen bij toepassing bij Ahold tot de conclusie dat deze methodiek zich niet goed leent voor accountantscontrole omdat daarmee vermenging optreedt tussen de verantwoordelijkheden van de leiding en de controlerend accountant. De methodiek bleek zich echter bijzonder goed te lenen om op basis daarvan een kwaliteitsslag te maken ten



Tabel 1 Waarden n en k₀ voor p_m = 0,01

p in % tussen	0-0,02		0,03-0,20		0,21-0,40		0,41-0,60		0,61-0,80		0,81-1,00	
N	n	K ₀	n	K ₀	n	K ₀	n	K ₀	n	K ₀	n	K ₀
1-25	Alle	0	Alle	0	Alle	0	Alle	0	Alle	0	Alle	0
26-50	22	0	22	0	22	0	22	0	22	0	22	0
51-100	27	0	27	0	27	0	27	0	27	0	27	0
101-200	32	0	32	0	32	0	32	0	32	0	32	0
201-300	33	0	33	0	33	0	33	0	33	0	65	1
301-400	34	0	34	0	34	0	70	1	70	1	70	1
401-500	35	0	35	0	35	0	70	1	70	1	70	1
501-600	35	0	35	0	75	1	75	1	75	1	75	1
601-800	35	0	35	0	75	1	75	1	75	1	120	2
801-1.000	35	0	35	0	80	1	80	1	120	2	120	2
1.001-2.000	36	0	80	1	80	1	130	2	130	2	180	3
2.001-3.000	36	0	80	1	80	1	130	2	185	3	235	4
3.001-4.000	36	0	80	1	135	2	135	2	185	3	295	5
4.001-5.000	36	0	85	1	135	2	190	3	245	4	300	5
5.001-7.000	37	0	85	1	135	2	190	3	305	5	420	7
7.001-10.000	37	0	85	1	135	2	245	4	310	5	430	7
10.001-20.000	85	1	135	2	195	3	250	4	435	7	635	10
20.001-50.000	85	1	135	2	255	4	380	6	575	9	990	15
50.001-100.000	85	1	135	2	255	4	445	7	790	12	1520	22



aanzien van de administratieve processen en de kwaliteit van de administratieve gegevensbestanden. In samenspraak met de toenmalige CFO van Ahold (L. Coren) en de desbetreffende controller (B. Hagens) werd een specifiek AOQL-controlesysteem ontwikkeld dat vervolgens met succes in de interne organisatie is geïmplementeerd. Na een korte aanlooptijd waarin de kwaliteit van het administratieve proces aanzienlijk werd opgevoerd en de gegevensbestanden werden geschoond, kon in het kader van de accountantscontrole worden volstaan met een beperkte (foutloze) steekproef op ernstige fouten in het op de jaarrekening aansluitende gegevensbestand. Daarenboven werd door deze aanpak tevens een – expliciet in de desbetreffende jaarrekeningposten ingebed – robuust accountantsoordeel verkregen over de opzet en de werking van het desbetreffende systeem van interne controle. Ook deze aanpak past in het concept van de volkomen controle en maakt de beoordeling van de opzet en werking van de administratieve organisatie en de daarin vervatte interne controle robuust en nagenoeg volledig objectief en met de leiding communiceerbaar zonder een beroep te hoeven doen op het zogenaamde ‘professional judgement’. Deze verbeteringen in de effectiviteit van de totale aanpak en de scherpere beoordeling van zowel de administratie als de administratieve organisatie werden bovendien bereikt terwijl per saldo ook nog een aanzienlijke kostenbesparing op zowel de administratie als de accountantskosten werd verkregen. Met toestemming van het Aholdconcern werden deze overtuigende resultaten met het ten behoeve van

Ahold ontwikkelde AOQL Internal Control System, zowel in Nederland als internationaal in de Statistician gepubliceerd (Veenstra en Kriens 1982; Veenstra 1983; Kriens en Veenstra 1985).

Daarmee bleek het ontwikkelde AOQL-systeem als onderdeel van de beheersingsmaatregelen niet alleen te leiden tot kwaliteitsverbetering van het administratieve proces en de uitkomst daarvan, maar kon ook de vanuit de externe accountantscontrole uitgevoerde controle op ernstige en niet-ernstige fouten, alsmede van de beoordeling van de administratieve organisatie en de daarin vervatte interne controle, aanzienlijk effectiever worden uitgevoerd tegen in totaliteit lagere kosten (zie paragraaf 5.2 en 5.3).

De grote verdienste van een op basis van de AOQL-steekproefmethode voor een specifieke situatie uitgewerkt AOQL-controlesysteem als onderdeel van het totale beheersingssysteem van de onderneming, is dat alleen al de voorbereiding dwingt tot een duidelijke, rationele en transparante inrichting van de processen en een hoge kwaliteit van zowel het administratieve proces als de gegevensbestanden. De investering verdient zich vrijwel altijd in korte tijd terug in blijvende verbetering van kwaliteit van het administratieve proces. Nadat immers een nagenoeg foutloos proces is bereikt, kan de (interne) kwaliteitscontrole zich in het vervolg beperken tot de minimale steekproefomvang per subpopulatie, waarin dan geen of zelden nog een fout wordt aangetroffen. Het bij Ahold ontwikkelde AOQL-systeem is ook

elders geïmplementeerd, onder meer op enkele plekken bij de overheid. In zijn artikel over financieel beheer van non-profit- naar profit-organisaties wijst Veenstra (1988 p. 72-75) op de grote mogelijkheden voor kostenbesparing bij de overheid door goede scholing in controleaspecten bij automatisering en door het toepassen van de AOQL-steekproefmethode. Toepassingsgebieden waaraan moet worden gedacht zijn bijvoorbeeld:

- Controle op subsidies, vergunningen e.d.;
- Controle op loongegevens, pensioengegevens enz.;
- Controle op gegevens in computerbestanden (file maintenance);
- Controle op kostendeclaraties.

5 Steekproeven voor de volkomen controle

5.1 Steekproefcontrole op ernstige fouten

Feitelijk heeft De Wolff (1956) een eerste steekproefmethode ontwikkeld die als prototype van de steekproefcontrole op ernstige fouten kan worden beschouwd. Een ernstige fout is in deze steekproefopzet een fout die niet mag voorkomen. Als die wel wordt aangetroffen kan de te controleren populatie niet worden goedgekeurd en is een volledige controle noodzakelijk. Daarbij overweegt De Wolff dat men aan controle op administratieve bewerkingen zeer hoge eisen stelt. Men acht in de techniek een uitvalpercentage van 3% in vele gevallen een zeer behoorlijk resultaat. Een administratie waarbij in 1% van de transacties fouten

voorkomt, zal echter meestal al ontoelaatbaar zijn. Daarbij komt dat men, aldus De Wolff, als regel niets met zekerheid kan zeggen over de draagwijdte van de mogelijke fouten. In de administratie is de omvang van de schade, die door het maken van een fout ontstaat, in vele gevallen rechtstreeks afhankelijk van de grootte van de fout.

Impliciet gaat De Wolff dus uit van een postensteekproef. Toch is er volgens De Wolff ook vanuit dit uitgangspunt wel een vermindering van controlewerkzaamheden mogelijk ten opzichte van volledige controle. Men moet echter wel bereid zijn om een kleine concessie te doen en genoeg te nemen met een exact te berekenen waarschijnlijkheidsuitspraak over een (door de auditor) maximaal toelaatbaar geacht foutbedrag in de populatie, behoudens een geringe kans dat die wordt overschreden. Het aanvaardingscriterium van de steekproef houdt in dat geen enkele ernstige fout in de steekproef wordt ontdekt. Wanneer dit wél het geval is, wordt de gehele populatie volledig gecontroleerd. In 1956 gaat De Wolff in zijn methode uit van een stratificatie van de populatie in twee lagen. In 1959 verfijnt hij de methode door stratificatie van de populatie in een willekeurig aantal grootteklassen mogelijk te maken (De Wolff 1959).

Van Heerden (1961 p. 453) concludeert dat de benadering van De Wolff principieel juist is, namelijk afgestemd op de rechtstreekse eigen actie van de auditor ten aanzien van de te controleren totalen in de jaarrekening in het kader van de volkomen controle, maar voor praktische toepassing

in de accountantscontrole nog niet aanlokkelijk genoeg. Door de stratificatietechniek moet voor elke grootteklasse een afzonderlijke steekproef-fractie worden bepaald en het voorgeschreven aantal elementen aselekt worden aangewezen. Zeker in 1961 was de noodzakelijke handmatige voorbereiding van zo'n steekproef zeer arbeidsintensief en woog de te bereiken besparing vaak niet op tegen integrale verificatie van de hele populatie. De door Van Heerden geïntroduceerde 'guldenrangnummethode' (monetary unit sampling) lost dit probleem op door een te controleren postenreeks op te vatten als een reeks gulden (monetary units). Aan elke gulden wordt een eigen rangnummer gegeven. Daardoor heeft ieder element in de steekproef per definitie hetzelfde gewicht voor de uitkomst. Aldus wordt de postensteekproef getransformeerd in een 'monetaire eenheid-steekproef', dus Gulden-steekproef, Euro-steekproef of Dollar-steekproef, enz. Het grote voordeel is dat hiermee zowel een grote eenvoud van de steekproefmethode als, door de ideale stratificatie, de minimale omvang van de steekproef is bereikt. De statistische uitspraak is hetzelfde gebleven als bij De Wolff (Van Heerden 1961, p. 464). Hoe groot de prestaties van De Wolff en Van Heerden zijn, blijkt uit het feit dat de vergelijkbare maar internationaal meer bekend geworden 'dollar unit sampling' pas in 1973 werd gepubliceerd door Anderson en Teitlebaum (1973). Wilburn (1984 p. 161) zegt hierover: *"It appears, however, that Van Heerden of the Netherlands was the pioneer in the field of monetary-unit sampling. He developed the guilder unit method in the 1950's."*

Kriens en Dekkers (1979) publiceren het eerste Nederlandse handboek over steekproeven in de accountantscontrole. Hoofdstuk 4 is gewijd aan het toepassen van steekproeven bij controles op ernstige fouten (t.a.p. p. 19-21 en 53-62). Het aanvaardingscriterium van de steekproef is dat geen enkele ernstige fout in de steekproef wordt ontdekt. Anders dan De Wolff, die de gehele populatie volledig controleert als er een ernstige fout wordt aangetroffen, gaan Kriens en Dekkers niet verder dan constateren dat het aantreffen van een ernstige fout in de steekproef consequenties heeft voor het werkprogramma, maar niet welke dat zijn (t.a.p. p. 20/21).

5.2 Steekproefcontrole op ernstige en niet-ernstige fouten

Veenstra en Kriens (1984) passen de steekproefcontrole op ernstige fouten op ruime schaal toe in de auditpraktijk en ontwikkelen de methode 'op ernstige fouten' door tot 'steekproefcontrole op ernstige en niet-ernstige fouten'. Deze steekproefmethode wordt in 1984 gepubliceerd. De steekproefmethode op ernstige fouten werkt in zijn algemeenheid tot tevredenheid. Het consequent toepassen van de steekproefmethode dwingt onderneming en accountant tot het controleerbaar maken van de complexe onderneming, de informatiesystemen en de bedrijfsprocessen.

De gekwantificeerde uitspraken leiden tot objectieve onontkoombare bevindingen. Er blijven echter ook beperkingen. *“In de eerste plaats beantwoordt het systeem niet de vraag, hoe ernstig de aangetroffen situatie is wanneer er afgekeurd wordt. Een tweede bezwaar ligt in de keuze van de uitgangspunten, waardoor het systeem beperkt is tot controle op ernstige fouten.”* (Veenstra en Kriens 1984 p. 403). Om de ernst van de aangetroffen situatie beter te kunnen beoordelen, komen Veenstra en Kriens met een gecombineerde oplossing. Wordt er geen ernstige fout aangetroffen, dan worden de steekproef en de populatie goedgekeurd; wordt er wel een ernstige fout aangetroffen, dan wordt een schatting gemaakt van de fractie ernstige fouten in de populatie. Omdat het bijbehorende interval uitsluitend wordt berekend indien $k > 0$ is en niet indien $k = 0$, wordt een ‘voorwaardelijk betrouwbaarheidsinterval’ uitgerekend. Deze informatie wordt mede in de evaluatie van de steekproef en de besluitvorming over de te ondernemen aanvullende controlestappen betrokken.

Daarenboven breiden Veenstra en Kriens de steekproefcontrole uit met de mogelijkheid om het onderzoek mede te richten op fouten die niet leiden tot afkeuring van de populatie en andere aspecten van de populatie die voor het controleinzicht in de populatie relevant zijn. Aldus wordt een aangetroffen ‘ernstige fout’ die niet in de populatie behoort voor te komen ingebed in aanvullende, voor de beoordeling relevante, kennis over die populatie. In het artikel van Veenstra en Kriens zijn diverse concrete voorbeelden opgeno-

men van steekproefcontroles waarin een controle op ernstige fouten (bijvoorbeeld verificatie van de ontvangst van de tegenprestatie) wordt gecombineerd met niet-ernstige fouten (bijvoorbeeld boekingen op verkeerde rekeningen en onjuiste rubriceringen) en toetsing van de werking van de administratieve organisatie (waaronder naleving procedures, instructies en autorisatieregels). *‘Uitgaande van een bepaalde opzet van de administratieve organisatie, is het doelmatig en gebruikelijk – voor zover dit controletechnisch mogelijk is – in het kader van de verificatiewerkzaamheden tevens vast te stellen of de administratieve organisatie gedurende de te controleren periode heeft gewerkt zoals verwacht mocht worden’* (Veenstra en Kriens 1984 p. 405 e.v.). Om geen enkele ruimte voor misverstand te laten bestaan, wordt hierbij opgemerkt dat ook de toetsing van de werking van de administratieve organisatie en de daarin vervatte interne controle zonder uitzondering uitgaat van de integrale waarneming van de volledige populatie die door de te beoordelen administratieve organisatie is gegenereerd, en volledig is ingebed in de organische bedrijfseconomische verbanden en de desbetreffende posten van de jaarrekening (zie paragraaf 2 en 4 van dit artikel). Dus ook deze controlewerkzaamheid wordt vanuit de gegevens en met betrekking tot de volledige populatie uitgevoerd en leidt tot gekwantificeerde uitspraken over de kwaliteit van de opzet en de werking van het desbetreffende deel of aspect van de administratieve organisatie.

5.3 Bayesiaanse steekproefcontrole op ernstige en niet-ernstige fouten

Vanuit het inzicht dat de conventionele steekproefopzet uitgaat van 'alle' theoretisch mogelijke uitkomsten, dus zelfs van de theoretische mogelijkheid dat er tot 100% fouten in de populatie aanwezig is, is gezocht naar mogelijkheden om de steekproefcontrole op ernstige en niet-ernstige fouten efficiënter te maken. Deze mogelijkheid werd gevonden in de zogenaamde bestaande Bayesiaanse statistiek, die het mogelijk maakt om harde voorinformatie over een populatie te interpreteren als kansverschijnsel en op basis daarvan een priorkansverdeling (voorafgaande waarschijnlijkheidsverdeling van de te onderzoeken verschijnselen) vast te stellen en daardoor op een kwalitatief degelijke grondslag de steekproefomvang te verkleinen (Veenstra en Van Batenburg 1989 p. 562). Het concept van de volkomen controle biedt een duidelijke structuur waarop de priorverdeling gebaseerd kan worden. De ervaringen met AOQL werden als aanmoediging ervaren (zie paragraaf 4). Dat de in dit artikel beschreven controlebenadering ook in het licht van de huidige technische ontwikkelingen een valide benadering blijft, blijkt ook uit de studies van Elsas (1996) en De Backer en Snoeck (2006) (zie paragraaf 2).

In 1990 publiceren Veenstra en Van Batenburg een vanuit de auditpraktijk onderbouwde Bayesiaanse steekproefcontrole op ernstige (en niet-ernstige) fouten. Het gehanteerde vaktechnisch onderbouwde kernbegrip waarin de Bayesiaanse voorinformatie wordt uitgedrukt is de factor f ,

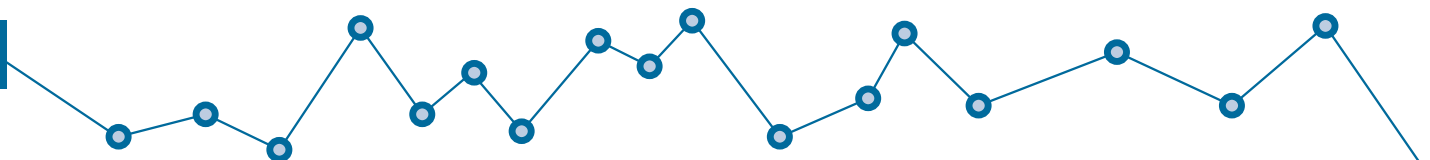
'de stabiliteit van het administratieve proces'.

"De stabiliteit van het desbetreffende administratieve proces dat de te onderzoeken populatie als uitkomst heeft, wordt dan uiteindelijk uitgedrukt in de waarde van f ." (Veenstra en Van Batenburg 1990 p. 19).

De stabiliteit van het administratieve proces ter zake van een specifieke te onderzoeken populatie hangt in sterke mate af van:

- De kwaliteit van de inrichting van de administratieve organisatie ter zake van de te onderzoeken populatie en de daarin bewust opgenomen maatregelen van interne controle;
- Specifieke wijzigingen in de kwaliteit van de inrichting van de administratieve organisatie ter zake van de te onderzoeken populatie en wijzigingen in de daarin bewust opgenomen maatregelen van interne controle;
- Voor de te onderzoeken populatie relevante specifieke wijzigingen in de controleomgeving en/of in de omgeving waarvan het administratieve proces onderdeel is, waaronder specifieke wijzigingen in de bedrijfsvoering en wijzigingen in het management.

Validatie van een toepassing van de Bayesiaanse steekproefcontrole op ernstige (en niet-ernstige) fouten is te allen tijde heel goed mogelijk, bijvoorbeeld door tijdens een stillere periode het (op grond van de f verminderde) aantal steekproefelementen conform de conventioneel berekende niet-Bayesiaanse steekproefomvang alsnog te controleren, waardoor de uitkomst kan worden vergeleken met de uitkomst van de conventio-



neel uitgevoerde steekproef. Zo kan ook ervaring worden opgedaan met de methode en de methode verder worden verbeterd. Desgewenst kan de methode zelfs ook gevalideerd worden door integraal onderzoek van de gehele populatie. De Bayesiaanse steekproefmethode op ernstige en niet-ernstige fouten is derhalve volledig valideerbaar en verifieerbaar en blijft ook overigens ingebed in de volkomen controle. De vaststelling van de factor f (de stabiliteit van het administratieve proces) mag dus niet verward worden met de niet-verifieerbare inschattingen van de risico's van de accountantscontrole in het Audit Risk Model, Audit Risk (AR), Inherent Risk (IR), Control Risk (CR) en Risk of Analytical Procedures (AP) (zie paragraaf 3).

Op de kritiek op de voorgestelde methode is door Veenstra en Van Batenburg (1991) uitvoerig gereageerd. In hoofdzaak is de kritiek beperkt gebleven tot de verdediging van het Audit Risk Model. Als belangrijk argument werd gehanteerd dat ook de organisatie waaraan Veenstra en Van Batenburg verbonden waren het audit risk model zou hanteren. De repliek daarop heeft bepaald nog niets aan actualiteit ingebracht: *“Naar onze mening zou het een stap op de goede weg zijn als de Nederlandse kantoren en beroepsbeoefenaren weer meer zelfstandig over de auditproblematiek gingen nadenken en zich niet zouden beperken tot ‘vertaling’ van Amerikaanse opvattingen, die niet noodzakelijkerwijze op een voor ons deugdelijke basis hoeven te berusten.”* (Veenstra en Van Batenburg 1991 p. 642). Internationaal en in de VS heeft de door Veenstra en Van Batenburg ontwikkelde steekproefmethode onder de naam

BDS (Bayesian Discovery Sampling) erkenning gekregen doordat het is opgenomen als expliciet toegestane steekproefmethode in het Audit Sampling Guide van Deloitte Touche Tohmatsu International (1991). Samen met O'Hagan is door Veenstra en Van Batenburg nog verdergaande research gedaan waarbij soortgelijke resultaten werden geboekt (Van Batenburg et al. 1992, 1994).

Naar de mening van de auteurs van het onderhavige artikel is het na alle boekhoudaffaires en het bereikte dieptepunt in het vertrouwen in het accountantsberoep tijd om in het voetspoor van pioniers, zoals Limperg, Kleerekoper, De Wolff en Van Heerden, de hoofdweg van de volkomen controle weer te volgen. Zoals in paragraaf 2 werd geconcludeerd is sprake van een concept met een veelbelovende toekomst.

6 Conclusie

De Wolff heeft met zijn publicaties in 1956 en 1959 een doorslaggevende rol gespeeld in een efficiëntere benadering van administratieve controle en accountantscontrole door steekproeven. Het is teleurstellend dat, hoewel goede steekproefmethoden beschikbaar zijn, deze nog niet op ruime schaal bij grote ondernemingen en overheden zijn geïmplementeerd. Na de recente affaires en gezien de moeizame relatie tussen overheid en burgers over besteding van overheids gelden, lijkt de tijd gekomen om deze impasse definitief en op grote schaal te doorbreken.

In het bijzonder kan door adequate toepassing van ICT in combinatie met de bovenbedoelde steekproefmethoden, handen en voeten worden gegeven aan het concept van de volkomen controle van de jaarrekening, zoals dat is ingezet door pioniers als Limperg, De Wolff en Van Heerden. Wellicht nog belangrijker, en los daarvan, heeft de oorspronkelijk voor de industriële kwaliteitscontrole ontworpen steekproefmethode AOQL een grote potentie voor het verbeteren van administratieve processen zowel in de overheidssfeer als in het bedrijfsleven. Hierdoor is niet alleen directe winst te behalen in de vorm van substantiële efficiëntievoordelen en een grotere effectiviteit, maar ook een indirecte winst die van onschatbare waarde is, namelijk herstel van vertrouwen in de integriteit van bedrijfsleven en overheden.

Literatuur

American Institute of Certified Public Accountants (1981), *Statement of Auditing Standards No. 39*, New York, (Appendix U.S. Auditing Standards section 350 Audit Sampling).

Anderson, R. en A. Teitlebaum (1973), Dollar-Unit Sampling: A solution to the audit sampling dilemma, *The Canadian Chartered Accountant/CA Magazine*, vol. 102, no. 4 (april), pp. 30-38.

Backer, M. de en M. Snoeck (2006), Bedrijfsprocessen en webservices, *Maandblad Informatie*, jaargang 48 (januari-februari), pp. 10-15.

Batenburg, P.C. van., J. Kriens en R.H. Veenstra (1988), Average outgoing quality limit: een herziene en verbeterde versie, in: J.G. de Gooijer, M.J.T.J. van Nieuwburg en J.A.M. Wesseling (redactie), *Economische statistiek: ontwikkelingen in het kwantitatief onderzoek*, Meppel, pp. 271-281.

Batenburg, P.C. van en J. Kriens (1988), E.O.Q.L. – a revised and improved version of A.O.Q.L., *FEW 348*, University of Tilburg.

Batenburg, P.C. van en R.H. Veenstra (1990), Enkele beschouwingen bij risicoanalyse in de accountantscontrole, *Maandblad voor Accountancy en Bedrijfseconomie*, jg. 64, no. 3 (maart), pp. 102-106.

Batenburg, P.C. van, A. O'Hagan, en R.H. Veenstra (1992), *Bayesian Discovery Sampling in financial auditing; practical applications and some remarks on the theoretical background*, Paper Congress over Practical Bayesian Statistics van The Institute of Statisticians, juli 1992.

Batenburg, P. C. van, A. O'Hagan en R.H. Veenstra (1994), Bayesian discovery sampling in financial auditing: a hierarchical prior model for substantive test sample sizes, in: *The Statistician*, vol. 43, no. 1, pp. 99-110.

Dodge, H.F. en H.G. Romig (1959), *Sampling Inspection Tables*, 2nd edition, New York, John Wiley.

DRTI International (1991), Bayesian Discovery Sampling, in: *Audit Sampling; A guide to applying The DRTI International Audit Approach*, Deloitte Touche Tohmatsu International, New York 1991, pp. 57-73.

Eimers, P.W.A. (2006), Het Audit Risico Model is springlevend! *Maandblad voor Accountancy en Bedrijfseconomie*, jg. 80 nr. 3 (maart), pp. 76-83.

Elsas, P.I. (1996), *Computational Auditing*, dissertatie Vrije Universiteit Amsterdam, 1996.

Heerden, A. van (1961), Steekproeven als middel van accountantscontrole, *Maandblad voor Accountancy en Bedrijfshuishoudkunde*, jg. 35, no. 12 (december), pp. 453-475.

Kleerekoper, S. (1933), De steekproeven als middel van accountantscontrole in de literatuur, *Maandblad voor Accountancy en Bedrijfshuishoudkunde*, jg. 7, februari 1933, pp. 26-32; maart 1933, pp. 50-57; april 1933, pp. 74-79; mei 1933, pp. 103-106; juni 1933, pp. 121-126.

Klein, A. (1911/1912), Über die Warscheinlichkeit der Entdeckung von Fehlern bei Revisionen, *Zeitschrift für Handelswissenschaftliche Forschung*, jg. 6.

Kriens, J. en A.C. Dekkers (1979), *Steekproeven in de accountantscontrole*, Stenfert Kroese, 1979.

Kriens, J. en R.H. Veenstra (1985), Statistical Sampling in Internal Control Systems by Using the A.O.Q.L.-System, *The Statistician*, vol. 34, pp. 383-390.

NIVRA interimrapport (1974), De statistische steekproef in de accountantscontrole, *De Accountant*, jg. 48.

NIVRA-geschrift (1982), *Accountantscontrole en steekproef*, Nivra-geschrift 25, Kluwer, januari.

NIVRA (1996-2005), *Richtlijnen voor de Accountantscontrole*, Amsterdam, vanaf editie 1996 tot en met editie 2005.

Petri, C.A. (1962), *Kommunikation mit Automaten*, Dissertation Universität Bonn.

Schmalenbach, E. (1911/1912), Über Einrichtungen gegen Unterschlagung und über Unterschlagungsrevision, in: *Zeitschrift für Handelswissenschaftliche Forschung*, jg. 6.

Talens, E. (2005), *Statistical Auditing and the AOQL-method*, Proefschrift Rijksuniversiteit Groningen.

Veenstra, R.H. (1972), *Handleiding Accountantscontrole*, intern document Van Dien + Co, 1972; uitgegeven als 'Handleiding Assistenten' en in ruime kring verspreid, januari.

Veenstra, R.H. (1981), *Gevolgen van de ontwikkelingen in de automatisering vanuit de algemene controlepraktijk*, Rapport uitgegeven en in ruime kring verspreid door de Nederlands Accountants Maatschap, januari.

Veenstra, R.H. en J. Kriens (1982), Toepassen van steekproeven in de administratie en interne controle door middel van het AOQL-systeem, *Bedrijfskunde*, jg. 54, no. 3, pp. 252-262.

Veenstra, R.H. (1983), Toepassen van steekproeven in de interne controle, *De Accountant*, jg. 90 (november), pp. 159-161.

Veenstra, R.H. en J. Kriens (1984), Steekproefcontrole op ernstige en niet-ernstige fouten, *Bedrijfskunde*, jg. 56, no. 4, pp. 403-408.

Veenstra, R.H. en J.C. Buysse (1985), Optimaliseren van steekproeftoepassingen in de administratie, *De Accountant*, jg. 91 (juni), pp. 561-563.

Veenstra, R.H. (1988), Financieel beheer van non-profit- naar profitorganisaties, in: K. D. Waagenaar en J. D. Voskamp (redactie), *Ervaringen met Privatiseren*, Kluwer Bedrijfswetenschappen, pp. 63-75.

Veenstra, R.H. en P.C. van Batenburg (1989), Een doorbraak in steekproeftoepassingen door Bayesiaanse statistiek, *De Accountant*, jg. 95 (juli/augustus), pp. 561-564.

Veenstra, R.H. en P.C. van Batenburg (1990), Een Bayesiaanse steekproefcontrole op ernstige fouten; een doorbraak in steekproeftoepassingen, *De Accountant*, jg. 97 (september), pp. 18-21.

Veenstra, R.H. en P.C. van Batenburg (1991), Bayesiaanse steekproefcontrole op ernstige fouten; een methode om uit de impasse te geraken, *De Accountant*, jg. 97 (juni), pp. 641-645.

Veenstra, R.H. (1994), *De accountant beoordeeld- Effectieve beheersing van de accountancy in uw organisatie*, Kluwer Bedrijfswetenschappen.

Veenstra, R.H. (1998), Het schemergebied tussen interne en externe controle belicht, in: *Handboek Automatisering van de Informatieverzorging*, Samsom Bedrijfsinformatie.

Veenstra, R.H. (2000a), Kwaliteit accountantscontrole verankeren in de wet, *Tijdschrift Financieel Management*, jg. 20 (januari/februari), pp. 71-77.

Veenstra, R. H. (2000b), High Quality Audit: controlegrondslagen nader gedefinieerd, *Tijdschrift Financieel Management*, jg. 20 (maart/april), pp. 77-85.

Veenstra, R.H. (2005), *De accountant beoordeeld – De onmisbare schakel voor herstel van vertrouwen*, Kluwer.

Wallage, Ph. (2005), De actuele waarde van Limpergiaans vertrouwen, *Maandblad voor Accountancy en Bedrijfseconomie*, jg. 79, no. 4 (april), pp. 125-131.

Wilburn, A. J. (1984), *Practical Statistical Sampling for Auditors*, Marcel Dekker, New York.

Wolff, P. de (1956), Steekproeven bij administratieve controle, *Statistica Neerlandica*, jg. 10, no. 1, pp. 35-44.

Wolff, P. de (1959), Productiviteitsverhoging bij accountantscontrole door toepassing van gelaagde steekproeven, *Statistica Neerlandica*, jg. 13, no. 2, pp. 215-232. (Engelstalige versie: Wolff, P. de (1959), *On the application of stratified sampling to an auditing problem*, Universita degli studi di Roma)

Noot

- 1 Dit artikel is oorspronkelijk gepubliceerd in MAB 2006, pp. 611-619.
- 2 Een Petri-net bestaat uit een verzameling plaatsen en een verzameling transities die met elkaar worden verbonden door gerichte pijlen (flowrelaties). Aan de hand van de plaatsen, de transities en de flowrelaties kan de structuur van een bedrijfsproces worden gemodelleerd.
- 3 Gedistribueerde bedrijfsprocessen zijn lokale bedrijfsprocessen van verschillende ondernemingen die met elkaar gekoppeld worden zodat additionele functionaliteit gecreëerd wordt.

Samenvatting dissertatie 'Valideren van risico-inschattingen in accountantscontrole'¹

Ed Broeze

In **hoofdstuk 1** geven we aan hoe in de gebruikelijke praktijk van auditing het toegestaan is de gegevensgerichte controle aanzienlijk te verminderen, als een van de uitkomsten van risicoanalyse, het bestaansrisico, op 'laag' wordt ingeschat. Het bestaansrisico is het risico (of de kans) dat in de jaarrekening, zoals deze wordt voorgelegd aan de auditor, zich een materiële fout bevindt (een fout die zo groot is, dat hij goedkeuren in de weg staat). Dit risico is het resultaat van het inschatten van:

- het inherente risico: het risico op een materiële fout ten gevolge van omgevingsfactoren;
- het interne controle risico (control risk): het risico dat de administratieve processen onvoldoende mogelijke fouten ten gevolge van omgevingsfactoren en als gevolg van eigen onvolkomenheden voorkomen, opmerken en/of corrigeren, zodat een materiële fout kan resulteren.

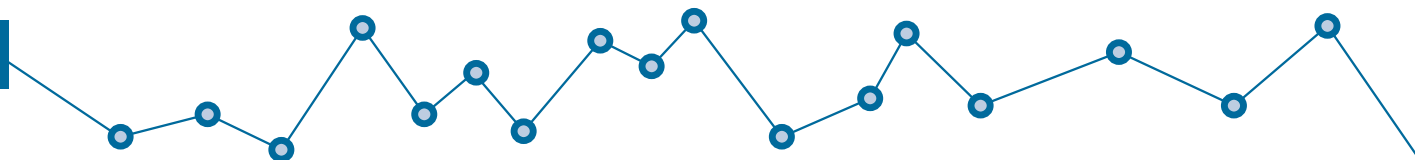
In deze praktijk worden standaardtabellen gebruikt, die een ingeschat bestaansrisico omzetten in een toegestane vermindering van de gegevensgerichte controle. Dit gebruik riep vragen bij ons op: hoe kunnen deze tabellen worden gerechtvaardigd? Is er een gezonde empirische basis voor

het omzetten van een ingeschat bestaansrisico in een vermindering van de gegevensgerichte controle? Om een antwoord te vinden op deze vragen, besloten we een onderzoek op te zetten naar de validiteit van het inschatten van het bestaansrisico.

In dit onderzoek hoopten we helderheid te krijgen over het verband tussen een ingeschat bestaansrisico en het 'werkelijke risico'. Hiertoe hebben we meer inzicht in het 'werkelijke risico' en valideringscriteria nodig.

In **hoofdstuk 2** ontwikkelen we een rechtvaardiging voor vier valideringscriteria:

1. Het foutpercentage (omvang van de fout als deel van de omvang van de rekening) en drie grootheden die op basis hiervan geconstrueerd kunnen worden:
2. De 'audit positie' van het foutpercentage (is deze groter dan de materialiteit?),
3. Het steekproefrisico (de kans dat een materiële fout niet wordt ontdekt, omdat slechts een steekproef is getrokken),
4. De empirische verdeling van de foutpercentages in de verzameling van alle gevallen die in ons onderzoek geanalyseerd zijn (voor valide inschattingen verwacht je hierin voor een laag



bestaansrisico dominantie van de lagere foutpercentages en voor een hoger bestaansrisico van de hogere foutpercentages).

De inschatting van het bestaansrisico richt zich op het vaststellen van de kwaliteit van de administratieve processen. Voor deze kwaliteit is het foutpercentage een directe indicator. In steekproeven zullen processen van dezelfde kwaliteit een variatie in de foutpercentages laten zien. In combinatie met de materialiteit (de goedkeurgrens) kan het foutpercentage een betere indicatie van het risico geven dat de administratieve processen een materiële fout zouden kunnen voortbrengen. 'Audit positie' en steekproefrisico zijn mede op de materialiteit gebaseerd. Daarom zullen zij in principe betere valideringscriteria zijn. We werden geholpen in het vinden van rechtvaardiging voor onze valideringscriteria door het 'Ensemble Prediction System' – de basis voor weervoorspellingen – te exploreren en dit, naast het risico dat water (de fout) een dijk (de materialiteit) overstroomt, als metafoor te gebruiken. Deze exploratie gaf ook grond aan het gebruik van de empirische verdeling van foutpercentages ter validering van het inschatten van risico's.

In **hoofdstuk 3** onderzoeken we de logische eigenschappen van het Audit Risk Model (ARM) en geven een overzicht van relevant onderzoek. We behandelen het 'heuristics and biases'-paradigma van Tversky en Kahneman (T&K). Dit paradigma vormt een basis voor het beschouwen van de risico-inschattingen en voor een meerderheid van

de beschikbare onderzoeken. We exploreren het ARM op zijn gebeurtenissenstructuur, zijn statistische validiteit en op het niveau waarop het risico wordt ingeschat. We concluderen dat oneigenlijke uitkomsten op de loer liggen, als gevolg van tekortkomingen m.b.t. de drie genoemde aspecten. Voor het ordenen van de beschikbare onderzoeken gebruiken we naast het T&K-paradigma ook 'consistency', 'complexiteit' en 'werking van het systeem' als ordenende begrippen. Onze overzicht laat zien dat:

- De 'biases' van het T&K-paradigma zich inderdaad voordoen bij het inschatten van risico's, maar dat zij met een geschikte training beperkt blijven;
- Het inschatten van risico's consistent is met relevante criteria; speciaal m.b.t. het foutpercentage. Maar ongelukkigerwijs is deze consistentie niet stabiel over de onderzoeken;
- Het inschatten van risico's op het niveau van 'controle beweringen' (ook wel 'controle criteria') tot verbeteren van de inschatting kan leiden;
- Complexiteit van het controle object een negatieve invloed heeft op de kwaliteit van de inschatting, maar dat ervaring het omgaan met complexiteit verbetert;
- De paar onderzoeken naar de effectiviteit van systeemtests en naar de invloed van kenmerken van het administratieve systeem op het foutpercentage, wijzen op een voorspellende kracht voor het foutpercentage, maar ook de hoogste kwaliteit is niet sterk genoeg om fouten te voorkomen.

Er is dus veel onderzoek op het terrein van risico-inschattingen, met positieve uitkomsten. Maar het grootste deel van dit onderzoek heeft slechts betrekking op aspecten van deze beoordeling. Voor een sluitend antwoord op de vraag of de huidige praktijk van kwantificering van de risico's is geoorloofd, zullen we het verband tussen het ingeschatte bestaansrisico en (de 4) valideringscriteria (onze indicatoren voor 'het werkelijke risico') moeten onderzoeken. Dat zal het doel van ons onderzoek zijn. Uit onderzoek dat wijst op een voorspellende kracht van specifieke aspecten van de administratieve processen kan worden geconcludeerd dat het decomponeren van risico-inschattingen de validiteit ervan kan verbeteren. Deze conclusie heeft ons ertoe gebracht om in de eerste studie niet alleen de klassieke risico-inschattingen te valideren, maar ook mogelijkheden te onderzoeken voor verbetering hiervan door de risico-inschatting op te delen over het inschatten van risico-indicatoren.

Voorafgaand aan het behandelen van ons eigenlijke onderzoek lossen we in **hoofdstuk 4** het probleem op, hoe het steekproefrisico te berekenen met behulp van de data die wij tot onze beschikking hadden: het geschatte foutpercentage en de bijbehorende steekproefomvang. In een simulatiestudie laten we zien dat het gebruik van de Beta-verdeling, waarin deze twee grootheden worden gebruikt, geschikt is voor het berekenen van het steekproefrisico. Als een extra vonden we sterke indicaties dat deze manier van het evalueren van steekproefresultaten een betrouwbaarheidsgrens

geeft voor de fout in de jaarrekening, die valide is en scherper dan de gebruikelijke Stringer-bound.

In **hoofdstuk 5** verantwoorden wij onze onderzoeksofzet. Acht organisaties deden mee in de eerste studie. Hen werd een vragenlijst voorgelegd naar gegevens over ongeveer 20 controle gevallen (uit de jaren 1996, 1997): het foutpercentage, het ingeschatte bestaansrisico en scores op de risico-indicatoren. De antwoorden waren zoveel mogelijk gebaseerd op de controledossiers.

De eerste studie werd gevolgd door een tweede: deze hield de replicatie in van het valideren van het klassieke ingeschatte bestaansrisico, zoals dat al in de eerste studie was gedaan. Ten slotte werd in een derde studie de voorspellende kracht van 'systeemtests' onderzocht. De tweede en derde studie werden gedeeltelijk bepaald door de beschikbaarheid van de noodzakelijke data. In hoofdstuk 5 lichten we ook onze keus voor een veldstudie toe. Hij is gebaseerd op de wens om risico-inschattingen te valideren in realistische zettingen. We kozen niet voor een experimentele aanpak, omdat het ons onmogelijk lijkt risico's dicht genoeg te benaderen, zoals die blijken of verstopt zijn in de werkelijkheid.

In **hoofdstuk 6** rapporteren wij de resultaten van de eerste studie naar de validiteit van de klassieke risico-inschatting. Zij varieerden per valideringscriterium, zoals de volgende samenvatting laat zien.

- Met betrekking tot de 'audit positie' werd slechts voor één organisatie een significante

negatieve correlatie gevonden. Precies de helft van de relevante correlaties had een '+'-teken; in die gevallen bestond dus een tendentie dat een hoog ingeschat risico samenging met een 'audit positie' die goedkeuren niet in de weg stond. Bij een valide risico-inschatting zou je juist het omgekeerde (en dus een negatieve correlatie) verwachten.

- De validiteit met het foutpercentage als criterium was bevredigend voor de gepoolde organisaties: de relevante correlatie was hier .43 (p -waarde $< .01$). Toch was er een probleem: dezelfde correlatiecoëfficiënt per organisatie varieerde sterk: van 0 tot .72. Hoewel dus gemiddeld de risico-inschatting valide is m.b.t. het foutpercentage, geeft dit aan een organisatie geen garantie m.b.t. zijn eigen risico-inschattingen.
- De validiteit m.b.t. het steekproefrisico kan niet worden aangetoond: voor de gepoolde organisaties werd een negatieve correlatie gevonden (terwijl validiteit om een positieve vraagt). In twintig van de geanalyseerde auditgevallen bleek er een serieus gevaar voor te weinig controle-inspanning te bestaan en in negentien gevallen bleek er een gevaar van een te hoge controle-inspanning (meer dan nodig) te bestaan. In het eerste geval dus ineffectiviteit, in het tweede inefficiëntie. Geanalyseerd per organisatie bleek dat bij twee organisaties de risico-inschatting valide was op dit criterium (een sterke positieve correlatie); bij twee andere was de relevante correlatie juist sterk negatief.

- We vonden een bevredigende mate van validiteit met betrekking tot de empirische verdeling van foutpercentages op het niveau van de gepoolde organisaties. We hadden niet genoeg data om deze analyse op het niveau van de afzonderlijke organisaties uit te voeren.

Deze uitkomsten zijn niet eenduidig: op twee valideringscriteria vonden wij validiteit, ook al was deze niet stabiel over de organisaties; op twee criteria vonden we geen validiteit. Analyses op een min of meer constant niveau van de materialiteit leidde niet tot verbetering van dit beeld, evenmin als analyses met potentiële moderatorvariabelen. Deze instabiliteit en het niet verbeteren voor potentiële moderatorvariabelen leidden tot de conclusie dat risico-inschatting door een organisatie niet vanzelfsprekend valide is. Tenzij hij deze validiteit heeft vastgesteld, moet een organisatie er dus van uitgaan dat zijn risico-inschattingen niet valide zijn in de zin van ons onderzoek. Onze bevindingen maken de conclusie aannemelijk dat een auditor eerder het foutpercentage dan het risico inschat. Met de bedoeling deze conclusie te testen zetten wij onze tweede studie op (zie hoofdstuk 8).

In **hoofdstuk 7** onderzoeken wij of het gebruik van risico-indicatoren de validiteit van risico-inschattingen met betrekking tot het foutpercentage verbetert. Voor dit deel van onze eerste studie ontwikkelden wij een verzameling van rond de 20 risico-indicatoren, met behulp van een voorstudie, door het analyseren van controledossiers, con-

trolehandboeken, tekstboeken en interviews met auditors. Wij analyseerden het verband tussen de indicatoren en het foutpercentage met behulp van (multiple) correlaties en regressie.

De indicatoren bleken consistent te zijn met het ingeschatte bestaansrisico: zowel de bivariate correlaties als de multiple correlaties, met een door de hele analyse gehanteerde standaard deelverzameling van indicatoren, waren bevredigend. De indicatoren waren minder consistent met het foutpercentage. Van de bivariate correlaties hadden er veel het 'verkeerde' (+) teken (6 van de 19) ook al liet de tekentoets zien dat het aantal 'juiste' (-) tekens significant was (op het 5% niveau). Vele spreidingsdiagrammen bevestigden dat de bivariate relatie van foutpercentage en indicatoren zwak is.

Slechts voor twee organisaties werd een regressiemodel gevonden dat het foutpercentage voorspelde en dat significant was op het 5% niveau en dat meer dan 20 procent verklaarde variantie (lees: 'verklarende kracht') had. Maar deze twee modellen bleken erg gevoelig voor het wegnemen van een indicator: dan verdween de verklarende kracht vrijwel geheel. Bovendien hadden vele indicatoren het 'verkeerde' teken, wat zou betekenen dat je voor zo'n indicator moet aannemen dat de verwachte fout afneemt als het op die indicator ingeschatte risico toeneemt. Die onlogische bevinding kan worden verklaard door het verschijnsel van 'suppressie'. Het regressiemodel is daarmee wel valide, maar het kan nauwelijks nog geïnterpreteerd worden. Al met al hebben wij met de indicatoren geen overtuigende predictoren voor

het foutpercentage weten te vinden. We hebben geprobeerd de voorspellende kracht te vergroten door regressie op factorscores en op door onszelf geconstrueerde schalen, beide gebaseerd op de indicatoren, te analyseren. Dit leidde niet tot een consistente verbetering. Onze conclusie kan alleen zijn dat wij geen verbetering van de voorspelling van het foutpercentage door het gebruik van risico-indicatoren konden aantonen. Omdat er rondom de risico-indicatoren veel onbeantwoorde vragen waren, kozen wij voor een voortzetting van ons onderzoek waarin de meer belovende relatie tussen bestaansrisico en foutpercentage verder zou worden geëxploreerd.

In **hoofdstuk 8** rapporteren we de resultaten van deze gedeeltelijke replicatie van de eerste studie betreffende de validiteit van het bestaansrisico. Hierbij kregen wij toegang tot data in de controle-dossiers over rekeningen uit 2001 van vier departementale auditdiensten. Helaas werden de relatief sterke correlaties van bestaansrisico en foutpercentage van de eerste studie, niet teruggevonden in deze replicatie.

Op organisatieniveau waren de correlaties van bestaansrisico met foutpercentage of steekproefrisico vrijwel gelijk aan nul. Ook de correlaties met 'auditpositie' waren niet significant. Voor analyses met de conditionele verdeling hadden we te weinig gevallen per organisatie.

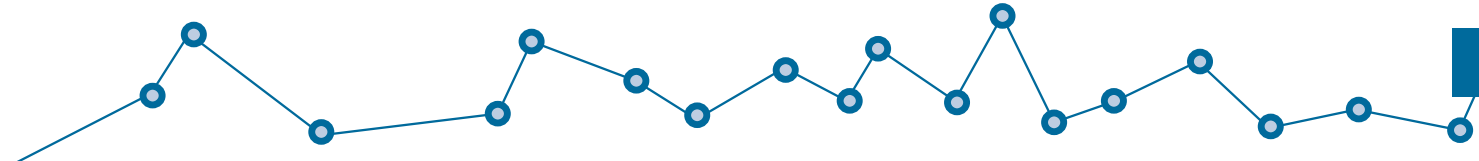
In deze replicatie breidden we onze onderzoeksvragen uit tot de vraag of de 'fout van het vorige jaar' de voorspellende kracht heeft voor de 'fout van dit jaar'. Voor twee organisaties konden we

de benodigde correlaties uitrekenen. In beide gevallen waren deze vrijwel gelijk aan nul, en was dus de voorspellende kracht niet aantoonbaar. Op organisatieniveau leidde dit tot de conclusie dat onze replicatie alleen aanwijzingen gaf voor het afwezig zijn van validiteit. Op het niveau van de gepoolde organisaties vonden we een positieve correlatie tussen foutpercentage en bestaansrisico. Dit werd ook teruggevonden met de conditionele verdeling; deze liet zien dat het bestaansrisico eerder het voorkomen dan de omvang van een fout voorspelt.

In **hoofdstuk 9** onderzoeken we of systeemtests gebruikt kunnen worden als onderbouwing van de inschatting van het bestaansrisico, zoals dit in de praktijk gebruikelijk is. Deze onderbouwing betreft meestal de gevallen waarin het bestaansrisico als 'laag' is ingeschat. In zo'n geval staat, zoals al gezegd, de auditmethodologie toe de omvang van de gegevensgerichte controle te verminderen, mits de risico-inschatting wordt onderbouwd door systeemtests. Dit betekent, naar ons inzicht, dat systeemtests voorspellend moeten zijn voor de omvang van de fout. Om dit te onderzoeken analyseerden wij de relatie tussen systeemtests en de fout zowel op het niveau van een transactie als op het niveau van een rekening, in een zeer grote verzameling van 'dual purpose tests', verdeeld over vijf jaar (1995 tot en met 1999). Op het niveau van een transactie was in alle jaren het vóórkomen van een systeemfout voorspellend voor het vóórkomen van een gegevensfout. Dat nam niet weg dat slechts in ca.

4% van de transacties waarin systeemfouten voorkwamen er ook een gegevensfout werd gevonden. Op het niveau van een rekening vonden wij in vier van de vijf jaren (1996 tot en met 1999) dat het voorkomen van een systeemfout voorspellend is voor het voorkomen van een gegevensfout. Voor de jaren '98 en '99 vonden we bovendien dat het voorkomen van een systeemfout voorspellend is voor de omvang van de gegevensfout. In deze derde studie analyseerden wij ook de voorspellende kracht op het rekeningniveau van de 'fout van vorig jaar' voor de 'fout van dit jaar'. Voor de foutomvang werd deze alleen gevonden voor het jaar 1999, voor het voorkomen van een gegevensfout vonden we drie significante correlaties; alleen de relatie tussen de jaren '95 en '96 was afwezig. Onze conclusie was dat systeemtests niet vanzelfsprekend zijn als onderbouwing, maar wel deze werking kunnen hebben.

In **hoofdstuk 10** vatten we onze conclusies samen; zij betreffen de validiteit van het inschatten van het bestaansrisico op de klassieke manier en de mogelijkheden dit te verbeteren via decompositie met behulp van risico-indicatoren. Daarnaast betreffen zij de kwaliteit van systeemtests als voorspeller van het foutpercentage en van de 'fout van vorig jaar' voor de 'fout van dit jaar'. Belangrijkste conclusie is dat een organisatie alleen mag aannemen dat zijn risico-inschattingen valide zijn, als hij dit ook empirisch heeft vastgesteld. Voorlopig mogen we niet aannemen dat deze inschattingen het 'werkelijke risico' geven. In dit hoofdstuk geven wij ook suggesties voor



voortzetting van dit onderzoek. Hierbij gaat onze voorkeur uit naar de mogelijkheid een ontwikkelingsproject op te zetten waarin de empirische verdeling van het foutpercentage, zoals gevonden in een 'geschikte verzameling' (van audit gevallen), als uitgangspunt wordt genomen. Deze verdeling kan als voorverdeling voor het foutpercentage dienen in een nieuw auditgeval. Ook het inschatten van het bestaansrisico komt in beeld, omdat deze voorverdeling afhankelijk gemaakt kan worden van het ingeschatte risico. Dit kan zowel voor de verdeling in de 'geschikte verzameling', als in de voorverdeling voor de nieuwe audit. Als risico-inschatting een voorspellende kracht heeft, zullen de conditionele verdelingen verschillen; als de voorspellende kracht er niet is, zullen de conditionele verdelingen samenvallen. In beide gevallen kan de empirische verdeling, gebruikt als voorverdeling, echter toch leiden tot een vermindering van de omvang van de gegevensgerichte controle. Dat zal met name gebeuren als deze verdeling veel gewicht geeft aan de lage foutpercentages. In enkele gevallen zal hij ook tot een vermeerdering van de controle-inspanningen kunnen leiden: als veel gewicht aan hogere foutpercentages wordt gegeven.

Op voorwaarde dat bij de afleiding van de voorverdeling uit de empirische verdeling marges worden ingebouwd voor veranderingen en nieuwe risico's en ook het preventieve karakter van een controle niet uit het oog wordt verloren, is deze benadering bruikbaar als een 'geschikte verzameling' is gedefinieerd en daarin de empirische

verdeling van de foutpercentages is vastgesteld. Het voortgaande onderzoek kan dan leiden tot het vaststellen van de voorspellende kracht van risico-inschattingen voor de empirische verdeling. Zo wordt deze risico-inschatting gevalideerd en zelfs gekalibreerd, waarmee deze het 'werkelijke risico' aangeeft. Naast deze aanpak met voorverdelingen uit 'geschikte verzamelingen', bevelen wij aan om een andere manier van modelleren van a priori informatie te gebruiken. Daarmee worden de problemen met statistische validiteit van het Audit Risk Model vermeden en de door ons voorgestelde modellering kan op een vanzelfsprekende manier geïntegreerd worden in de aanpak met de empirische verdeling van 'geschikte verzamelingen'.

Noot

- 1 Deze samenvatting is letterlijk en met toestemming van de auteur overgenomen uit het proefschrift. De auteur verontschuldigt zich voor de mogelijke stijlbloemen.

Rekenen met plaatjes: een proefschrift over een nieuwe kijk op steekproeven

Paul van Batenburg

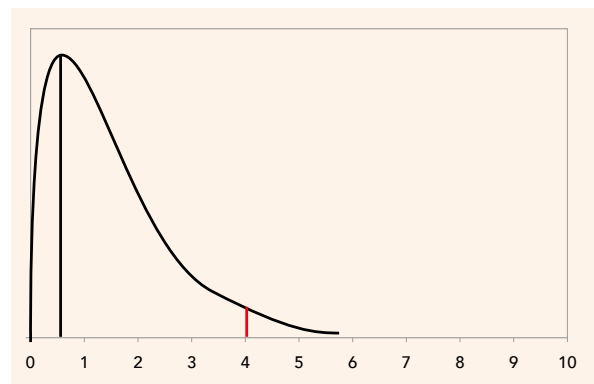
In 2013 promoveerde Trevor Stewart, gepensioneerd partner van Deloitte USA, aan de Vrije Universiteit te Amsterdam. De titel van zijn proefschrift luidt: 'A Bayesian Audit Assurance Model with Application to the Component Materiality Problem in Group Audits' (Een Bayesiaans risicoanalysemodel toegepast op het toekennen van materialiteit voor groepsonderdelen).

Het resultaat is te downloaden met deze link:

<http://dare.uvu.nl/bitstream/handle/1871/39813/dissertation.pdf>

Resultaten van dat project zijn ook terug te vinden in Stewart et al. (2012).

Samen met Philip Elsas – ook bekend op dit forum – was ik paranimf bij die promotie. Ik heb vanaf 2005 met Trevor samengewerkt aan een onderzoek naar de effectiviteit en efficiency van statistische steekproeven. Aanleiding voor dat onderzoek was de – toen nog – concept ISA 600 in het kader van het Clarity Project van de IAASB. Die nieuwe richtlijn gaf (in paragraaf 21c) aan dat de materialiteit voor een onderdeel in een groepscontrole kleiner dan, en niet meer zoals in vorige ISA's kleiner dan of gelijk aan, de overall materialiteit moet worden gesteld.



Figuur 1

Het vernieuwende van de manier waarop in dit project is gekeken naar steekproeven is dat we niet meer kijken naar schattingen zoals de geprojecteerde fout of de maximale fout, maar naar de gehele kansverdeling van de fout: welk bedrag aan fouten in de te controleren of in de gecontroleerde massa heeft welke kans? Die aanpak is essentieel voor Bayesiaanse statistiek: we nemen aan dat de onbekende fout in een populatie een kansverdeling heeft. In dit onderzoek is een Gammaverdeling verondersteld, mede omdat die verdeling mooi aansluit bij de Poissonverdeling voor het aantal fouten in een steekproef die uit die populatie kan worden getrokken, en die

Poissonverdeling is de basis voor de bij velen bekende tabel met R-waarden.

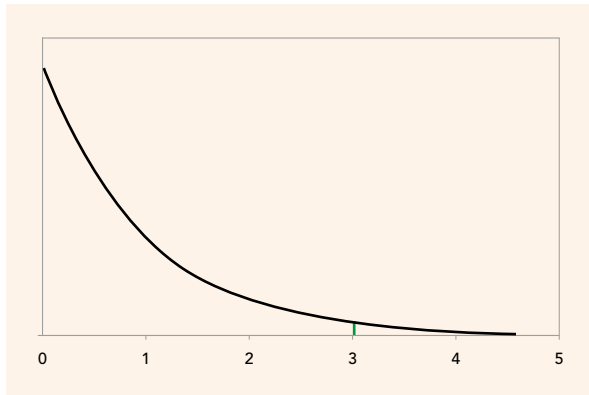
We zien in figuur 1 een populatie met een geprojecteerde (meest waarschijnlijke) fout van ongeveer 0,6 (miljoen euro) en een 95% betrouwbare maximale fout van bijna 4 (miljoen euro). Het eerste zien we aan de top van de curve, het tweede zien we doordat er 5% oppervlakte rechts van het rode streepje ligt. Zo'n plaatje kan een steekproefresultaat zijn, maar kan ook een inschatting van de accountant zijn op basis van risicoanalyse, de beoordeling van de AO/IB, of ervaringen uit het verleden.

Als voorbeeld: een steekproef van 100 uit een massa van 1 miljoen geeft bij 0 fouten een geprojecteerde fout 0 en een 95% betrouwbare maximale fout van 30.000 (figuur 2) en bij 1 fout een

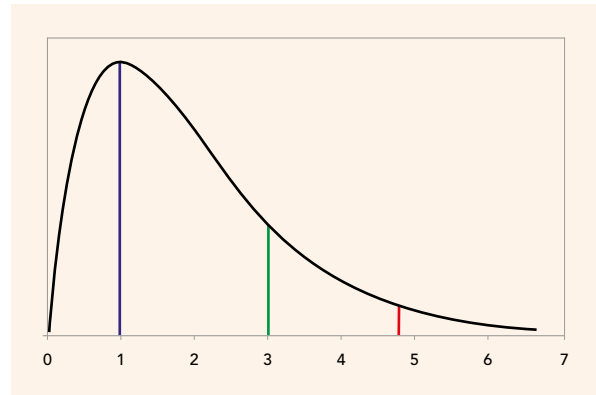
geprojecteerde fout 10.000 en een 95% betrouwbare maximale fout van 47.500 (figuur 3).

Het interessante van deze plaatjes is dat er als het ware mee gerekend kan worden. Voor het wiskundige bewijs zie het artikel of het proefschrift. Optellen en aftrekken, vermenigvuldigen, delen...

Natuurlijk zijn er bij dat rekenen voorwaarden te stellen waaronder de uitkomsten beter of juist minder goed passen. Hoe groter bijvoorbeeld het verschil tussen de omvang van twee populaties des te minder exact zullen de resultaten van deze berekeningen zijn. Maar, het gaat nu om het principe en niet om de details.



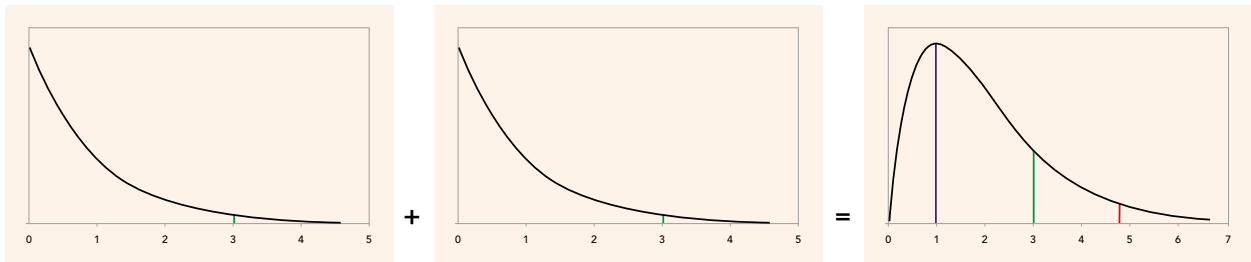
Figuur 2



Figuur 3

Bijvoorbeeld:

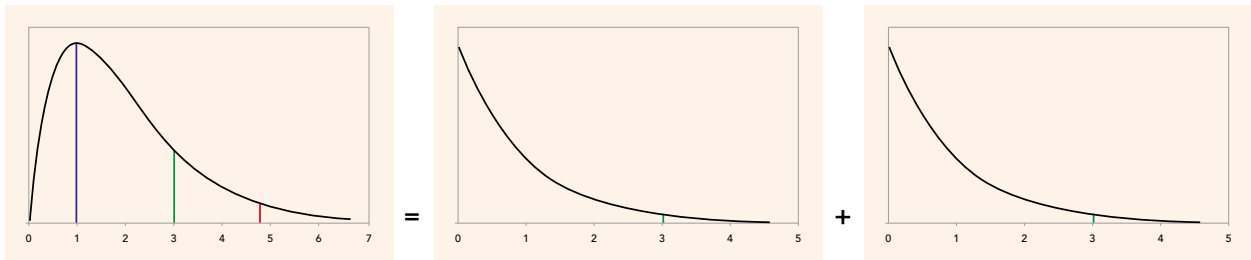
1. Ik heb een plaatje voor massa 1 en een plaatje voor massa 2. Hoe ziet het plaatje eruit als ik die massa's samenvoeg? Dit slaat op het geven van één oordeel over het totaal van 2 of meer heterogene (= onafhankelijke) massa's.



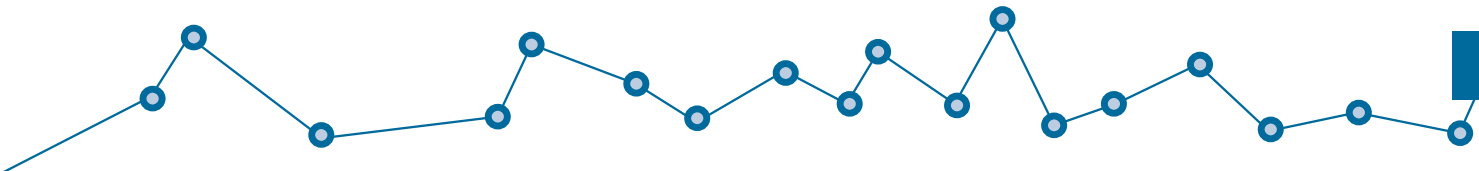
Figuur 4

Merk op dat de geprojecteerde fout per massa 0 is, maar voor het totaal niet; dat is het aggregatierisico uit COS 600, 21c. Voor statistici is dit vanzelfsprekend: de centrale limietstelling geeft aan dat het optellen van onafhankelijke kansverdelingen uiteindelijk leidt tot een (symmetrische) normale verdeling.

2. Ik weet aan welk plaatje de totale massa moet voldoen. Wat zijn de beste plaatjes voor delen 1 en 2 die optellen tot het gewenste plaatje?



Figuur 5

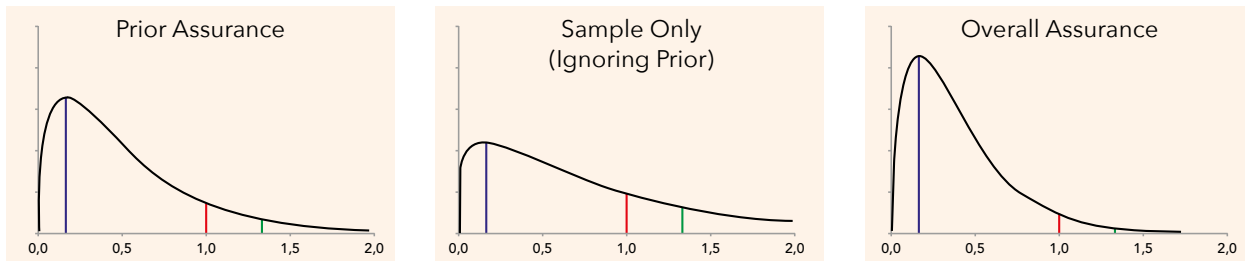


Dit is het toerekenen van de materialiteit naar heterogene deelpopulaties: het bepalen van de materialiteit voor een groepsonderdeel (COS 600, 21c en A43) of van de uitvoeringsmaterialiteit daar weer uit (COS 320, 9 en A12). Het nut van dit opdelen is dat fouten per deelpopulatie geëvalueerd kunnen worden.

Er bestaan oneindig veel antwoorden op de vraag welke twee plaatjes optellen tot een gegeven plaatje, maar het bijzondere van het proefschrift van Trevor Stewart is, dat hij heeft bepaald wat de beste keuze is gegeven een gekozen criterium zoals de totale steekproefomvang of de totale controlekosten.

3. Vervolgens kunnen we ook vermenigvuldigen:

Dit is het Nieuwe Generatie Audit Risk Model: de accountant veronderstelt het linkerplaatje en berekent hoeveel waarnemingen met hoeveel fouten leiden tot het gewenste rechterplaatje. Het klassieke model werkt alleen bij het veronderstellen en vinden van 0 fouten; nu kan ook voorkennis leiden tot efficiëntere steekproeven als een geprojecteerde fout vooraf wordt verondersteld. Deze convolutie van de Gammaverdeling is standaard in de Bayesiaanse statistiek.



Figuur 6

4. Ten slotte kunnen we ook delen: dat is van toepassing als een homogene massa op verschillende plaatsen wordt verantwoord. Dan bepalen we de voorkennis voor de deelmasa die volgt uit de controle op de gehele massa. Het eindplaatje van de drie plaatjes hiervoor voor de gehele populatie levert een beginplaatje voor eenzelfde analyse voor een deelpopulatie. Ook hier geldt natuurlijk dat wordt aangenomen dat de fout in de deelpopulatie weer een gammaverdeling volgt. Belangrijk daarbij is dat de deelpopulatie niet te klein mag zijn ten opzichte van de geheel populatie.
Voorbeeld van een toepassing van deze methode is een Shared Services Center waar transacties worden afgehandeld die later in verschillende verantwoordingen terechtkomen. Per verantwoording is een materialiteit gesteld, maar de controle zou efficiënter op de gehele populatie kunnen worden uitgevoerd.

Ik hoop dat deze vier voorbeelden hebben aangegeven hoe belangrijk dit proefschrift voor het vakgebied statistical auditing is en raad geïnteresseerden aan kennis te nemen van het artikel of het boek.

Literatuur

Stewart, Trevor R. and Jr. Kinney, R. William (2012); Group Audits, Group-Level Controls, and Component Materiality: How Much Auditing is Enough? (August 2012). *Accounting Review*, Forthcoming. Available at SSRN: <http://ssrn.com/abstract=2142891>

Het 'push left'-principe als motor van data-analytics in de accountantscontrole

Jacques de Swart, Jan Wille en Barbara Majoor

Samenvatting

Het businessmodel van accountantskantoren staat onder druk. In diverse gremia is de vraag gesteld of accountants in voldoende mate en tijdig hebben opgemerkt of de ondernemingen die zij controleren financieel gezond zijn. Tegelijkertijd bieden data-analytics veel mogelijkheden voor innovatie in de accountantscontrole, maar worden die mogelijkheden nog onvoldoende benut. Dit artikel presenteert het 'push left'-principe: een nieuwe manier om data-analytics te positioneren in de accountantscontrole.

De toepassing van data-analytics volgens het 'push left'-principe biedt drie voordelen: ten eerste heeft de accountant meer inzicht in de kans dat hij een goedkeurende verklaring afgeeft terwijl de jaarrekening toch een materiële fout bevat. Ten tweede is hij in staat om te garanderen dat zijn controlemix optimaal is: gegeven de verwachting die hij vooraf uitspreekt is het statistisch onmogelijk om tegen minder inspanning meer zekerheid te verkrijgen.

Ten derde vergroot het 'push left'-principe de relevantie van de accountant door meer dan alleen zekerheid te bieden. Het 'push left'-principe kan eenvoudig gekwantificeerd worden door een

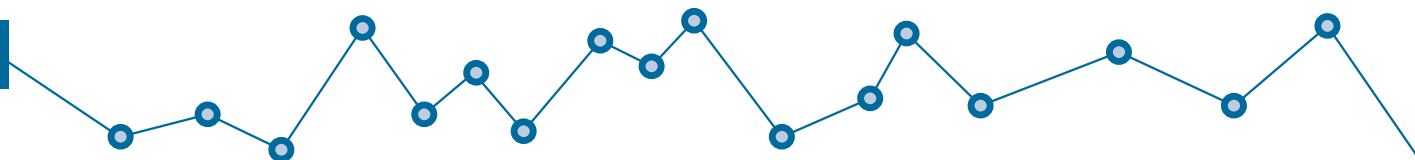
nieuwe Bayesiaanse methode genaamd Beta. Deze positionering past binnen de controlestandaarden, hoewel aanscherping van de standaarden soms wenselijk is.

Relevantie voor de praktijk

De toepassing van data-analytics is een steeds belangrijker wordend aspect van de accountantscontrole. Voor accountants werkzaam in de controlepraktijk maar ook voor de ondernemingen die gebruikmaken van accountantswerkzaamheden laten wij in dit artikel zien welke nieuwe mogelijkheden en inzichten data-analytics kunnen bieden om enerzijds de controle effectiever en efficiënter te maken en anderzijds hoe de controle door toepassing van het 'push left'-principe meer toegevoegde waarde kan hebben.

1 Inleiding

Wij zien in de huidige discussie over het accountantsberoep drie belangrijke uitdagingen voor de accountant die als een rode draad de toekomst van het accountantsberoep bepalen. De centrale



probleemstelling van dit artikel is welke rol data-analytics kunnen spelen om deze uitdagingen op korte termijn aan te gaan. Wij stellen daartoe het 'push left'-principe voor, een nieuwe manier om data-analytics te positioneren binnen de accountantscontrole. In deze context verstaan we onder data-analytics de kwantitatieve analyse van de betrouwbaarheid van gegevensverzamelingen. Zo'n analyse kan niet alleen bestaan uit steekproeven, maar ook uit de ontwikkeling of review van een kwantitatief model. Data-analytics ondersteunen diverse onderdelen van de controleaanpak, zoals het begrijpen van de business door middel van benchmarking, het testen van de effectiviteit van interne beheersingsmaatregelen, cijferanalyses en detailcontroles. Voor een uitgebreider overzicht van de verschijningsvormen van data-analytics in de accountantscontrole verwijzen we naar De Swart, Briene & Bakkum (2011). Om het 'push left'-principe te kwantificeren, maken wij gebruik van BETA, een op Bayesiaanse statistiek gebaseerde methode om de bij een controlemix behorende zekerheid te berekenen en deze zodoende op efficiënte wijze te optimaliseren. Ook plaatsen wij het 'push left'-principe in de context van de controlestandaarden.

De opbouw van dit artikel is als volgt. In deze inleiding schetsen we drie uitdagingen voor het accountantsberoep en hun link met data-analytics. In paragraaf 2 voeren we het 'push left'-principe in als manier om data-analytics in de accountantscontrole te positioneren. Dit principe is gebaseerd op het welbekende Audit Risk Model (ARM). We bespreken hoe onze interpretatie van ARM zich

verhoudt tot eerdere besprekingen van ARM in dit blad en andere literatuur en leggen een link tussen ARM en het Bayesiaanse Risico Analyse Model (BRAM). In paragraaf 3 bespreken we hoe aanpassing van bestaande statistische Bayesiaanse methoden nodig is voor toepassing binnen het 'push left'-principe. In paragraaf 4 wordt de link met de auditstandaarden besproken. Enerzijds tonen we hoe de voorgestelde methode al binnen de huidige standaarden werkt, anderzijds laten we zien waar aanscherping van de standaarden wenselijk is.

1.1 Drie uitdagingen voor het accountantsberoep

1.1.1 Onvermijdelijke vraag naar kwantificering van zekerheid

Van de accountant wordt primair verwacht dat hij zekerheid geeft bij financiële verantwoordingen. De kwaliteit waarmee hij dat doet staat ter discussie. Zo stelt de AFM in haar op 11 juli 2013 verschenen rapport dat "ingrijpende maatregelen nodig zijn om de controles door NBA-kantoren op niveau te krijgen" (AFM, 2013). Hoe goed die kwaliteit ook moge worden, de accountant loopt nog steeds het risico (ook wel het AccountantsControleRisico (ACR) genoemd) dat hij een goedkeurende verklaring afgeeft bij een materieel onjuiste jaarrekening. Volgens NV COS 200 dient hij het ACR tot een aanvaardbaar niveau terug te brengen (NV COS). Dit maakt de vraag aan accountants of ze ook kwantitatief kunnen duiden wat hun goedkeurende verklaring bij een jaarre-

keningcontrole inhoudt op termijn onvermijdelijk. Wij wijzen op de belangstelling die er bestaat voor ratings en het gebruik daarvan om de continuïteit van een onderneming te evalueren. In veel controleprotocollen is deze kwantitatieve zekerheid al ingevoerd.

We noemen enkele voorbeelden:

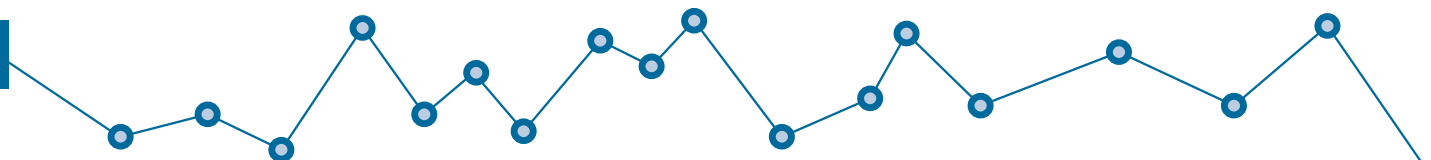
- Voor subsidies uit het Europees Fonds voor Regionale Ontwikkeling geldt dat alleen een goedkeurende controleverklaring verstrekt mag worden *“als met een betrouwbaarheid van minimaal 95% gesteld kan worden dat de fouten in het financieel verslag van de aanvraag tot subsidie vaststelling kleiner of gelijk dan 1% zijn”* (Europees Fonds voor Regionale Ontwikkeling, 2012).
- De Nederlandse Zorgautoriteit (NZa) eist van Geestelijke GezondheidsZorg (GGZ)-instellingen *“dat met 95% betrouwbaarheid gesteld moet kunnen worden dat niet meer dan 3% van het totaalbedrag van de gefactureerde Diagnose Behandeling Combinaties (DBC's) en de mutatie onderhanden DBC's in 2012 onjuist is”* (NZa, 2011).

Data-analytics hebben bij uitstek de potentie om de vraag naar gekwantificeerde zekerheid als resultaat van de jaarrekeningcontrole te beantwoorden en zo een antwoord te geven op deze uitdaging.

1.1.2 Toegenomen vraag naar meer dan alleen zekerheid

Roger Dassen (2012) schrijft in het rapport van de adviescommissie herziening eindtermen in

zijn voorwoord: “Van de accountant wordt meer verwacht dan het geven van zekerheid bij (financiële) verantwoordingen. Hij moet in staat zijn om risico's te signaleren, bespreekbaar te maken bij zijn cliënt en een rol te spelen bij de rapportage hierover aan het maatschappelijk verkeer.” Het rapport van de Commissie De Wit (Tweede Kamer der Staten Generaal, 2010) geeft aan dat het vooral gaat om toekomstgerichte informatie waar de accountant tekortschiet in zijn taakinvolving. “De financiële crisis heeft duidelijk gemaakt dat de financiële verantwoordingsinformatie van financiële instellingen voor aandeelhouders en andere gebruikers van de jaarrekening ontoereikend is geweest om de grote risico's binnen de instellingen of het financiële systeem te zien.” Dit betekent dat de accountant meer zal moeten doen dan een verklaring over cijfers afgeven die primair een verantwoording over het verleden zijn. Daartoe heeft hij al de ‘management letter’ als middel, maar omdat deze normaliter niet publiek beschikbaar is, neemt dit niet de behoefte aan vernieuwende inzichten over de instelling weg. Europees commissaris Barnier stelt vragen of de reikwijdte van de controle zich verder moet uitstrekken tot een oordeel over de financiële gezondheid van ondernemingen (European Commission, 2010). Ook Eumedion is van mening dat de accountant in zijn controleverklaring beleggers veel vaker moet wijzen op de risico's waarvoor de onderneming zich geplaatst ziet. Daarbij moet de accountant de toelichtende paragraaf niet alleen gebruiken voor het benadrukken van financiële risico's, maar juist ook van niet-financiële risico's,



zoals risico's van het gehanteerde governance model, strategische risico's en duurzaamheidsrisico's (Eumedion, 2011).

Uit recent onderzoek van Nyenrode Business Universiteit onder 1222 bedrijven met een accountantscontrole (Nyenrode, 2011) blijkt dat de waardering voor accountants niet zozeer afhangt van hoe goed zij hun primaire taak vervullen, maar vooral van de kwaliteit en hoeveelheid van deze vernieuwende inzichten. Het kunnen 'dichtvinken' van de jaarrekening wordt eerder gezien als een 'license-to-operate' of noodzakelijke voorwaarde, maar niet als voldoende voorwaarde om een gewaardeerde accountant te zijn.

Toch maakt toegenomen regelgeving het accountants steeds moeilijker om nog voldoende toe te komen aan de vorming van vernieuwende inzichten. De toegenomen mogelijkheden van data-analytics zullen hier een oplossing bieden, omdat een accountant die meerdere cliënten bedient, de mogelijkheid heeft om industriebrede inzichten te ontwikkelen, iets waar hun cliënten lastiger toe in staat zijn.

1.1.3 Druk op optimalisatie van de controlemix

De omzet en winstgevendheid van accountantskantoren staan sinds 2008 onder druk (ABN AMRO Bank, 2012). De branche wordt verweten dat zij met te weinig daadkracht heeft opgetreden bij het signaleren van de onjuiste implementatie en toepassing van accounting standaarden tijdens de financiële crisis (Kothari & Lester 2011). Dit verwijt knaagt aan de reputatie van accountants, waardoor opdrachtgevers zich kritischer opstellen

ten aanzien van de verhouding tussen de prijs en de geleverde diensten. Daarnaast neemt de regelgeving en het toezicht daarop sinds 2006, met de introductie van de Wet toezicht accountantsorganisaties (Wta) gestaag toe. Alle reden dus voor accountants om de controlemiddelen op elkaar af te stemmen ter optimalisatie van de controlemix. Nieuwe controlemogelijkheden op het gebied van data-analytics zullen dus niet bovenop de huidige controles geplaatst moeten worden, maar de huidige controles moeten vervangen. Aanpassing van de controlemix betekent wel het overwinnen van een drempel: wat als de nieuwe mix inzichten aan het licht brengt die in eerdere jaren ook valide waren maar toen niet werden opgemerkt? Meegaan in de neerwaartse spiraal van tariefdruk lijkt niet de oplossing voor het accountantsberoep vanwege de vicieuze cirkel minder controle – meer schandalen – nog meer tariefdruk, etc. Het antwoord zal juist moeten bestaan uit het bieden van meer toegevoegde waarde van accountants.

2 Data-analytics volgens 'push left'-principe als innovatie

Centraal in dit artikel staat de stelling dat data-analytics bij uitstek geschikt zijn om de genoemde uitdagingen voor de accountant aan te gaan. Dit vereist een positionering van data-analytics in de vorm van het 'push left'-principe, dat een verfijning van het Audit Risk Model is.

2.1 'Push left' als generalisatie van het Audit Risk Model

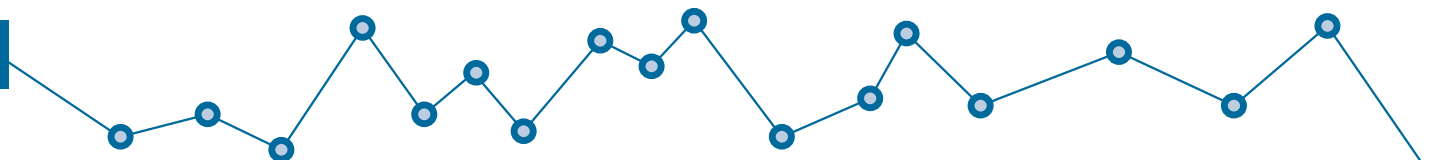
Het 'push left'-principe helpt om de verschillende vormen van data analytics en de afstemming met andere controlemiddelen in de accountantscontrole te duiden. Wij noemen dit het 'push left'-principe omdat het doel is om de zekerheid die verkregen wordt in het controleproces zoveel mogelijk voorin de verantwoordingsketen te verkrijgen. In die zin lijkt het 'push left'-principe op andere 'upstream'-methodieken als Lean (Womack & Jones 2003).

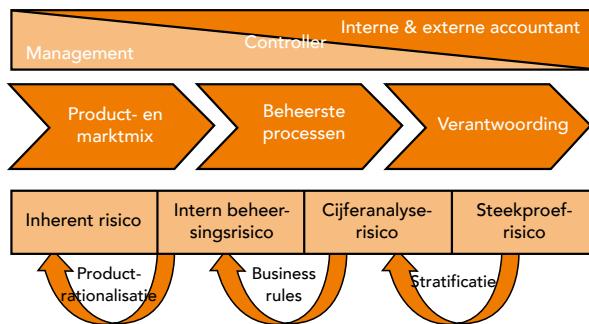
Het 'push left'-principe is gebaseerd op het aloude Audit Risk Model (Risicoanalysemodel), dat onder anderen door Knechel (1998) en Majoor en Van Kollenburg (2011a) uitgebreid beschreven is. Volgens dit model is het risico dat de externe verantwoording een materiële fout bevat (het AccountantsControleRisico (ACR)) gelijk aan het product van vier risico's:

- het Inherente Risico (IR), dit is het risico dat een fout een organisatie binnendringt;
- het Interne BeheersingsRisico (IBR), het risico dat de interne beheersing de fout niet opmerkt;
- het CijferAnalyseRisico (CAR), het risico dat cijferanalyses, zoals uitzonderingsrapportages en dwarsverbanden, de fout ook niet opmerken; en
- het SteekproefRisico (SR), het risico dat de fout door eventueel uitgevoerde steekproeven ook niet opgemerkt wordt (hetzij doordat de fout niet in de steekproef voorkwam, hetzij doordat de zorgvuldigheid van de steekproef te wensen overliet).

Dit model impliceert onder andere dat als IR of IBR gelijk aan 0 zijn, achteraf geen cijferanalyses of steekproeven meer nodig zijn. In de praktijk zal deze extreme situatie niet vaak voorkomen, en gebruikt men het model eerder om bijvoorbeeld de verkleining van de benodigde steekproef te kwantificeren als men deels kan steunen op interne beheersingsmaatregelen. Op deze wijze wordt het model ook in de nadere voorschriften controlestandaarden gebruikt (NV COS). In NV COS 315 worden de vier beschreven risico's van het ARM benoemd als onderdeel van de risico-inschattingswerkzaamheden. Op basis van het inzicht en de kennis die tijdens deze werkzaamheden is verkregen dient de accountant vervolgens de risico's van een afwijking van materieel belang te bepalen, waarbij de accountant expliciet moet vaststellen of een risico een significant risico betreft. Voor deze werkzaamheden dient de accountant vervolgens een controleplan met werkzaamheden op te stellen. Dit controleplan omvat veelal een mix van werkzaamheden gericht op toetsing van de interne beheersingsmaatregelen en gegevensgerichte controles. Steekproeven kunnen onderdeel zijn van de gegevensgerichte controles (NV COS 330).

Figuur 1 visualiseert de vier risico's. Boven de risico's staat de keten van besluiten die deze risico's beïnvloeden: de keuze van het management voor een product- en marktmix bepaalt het inherente risico. Besluiten van het management over de inrichting en beheersing van de processen die deze mix leveren, bepalen het IBR en – als cijferanalyses niet achteraf maar als onderdeel van





Figuur 1 Het 'push left' -principe

de interne beheersing plaatsvinden – het CAR. De wijze van verantwoorden door de accountant legt het steekproefrisico vast en – bij cijferanalyses achteraf – het CAR. De primair verantwoordelijken voor deze keuzes staan bovenaan in figuur 1. Van links naar rechts neemt de invloed van het management af, terwijl die van de accountants juist toeneemt.

In het verleden is in het MAB al vaker discussie gevoerd over de rol van risicoanalyse in de accountantscontrole. In meerdere artikelen is kritiek geleverd op het ARM. In Blokdijk (2001) wordt reeds de vraag gesteld of het ARM wel een deugdelijke grondslag levert voor de accountantscontrole. Er zijn veel verbeteringen gesuggereerd: Mollema (2003/2004) doet aanbevelingen om de directe vermenigvuldiging van risico's binnen het ARM te generaliseren, Kloosterman (2004) stelt dat hiertoe Bayesiaanse technieken zoals beschreven in Van Batenburg en Dassen (1996) meer op zijn plaats zijn. Veenstra en Heertje (2006) bepleiten het gebruik van de AOQL-methode. Eimers

(2006) geeft een goed historisch overzicht van deze discussie. Touw en Hoogduin (2011) tonen aan dat het Bayesiaanse Risico Analyse Model (BRAM) binnen de context van het ARM valt. Zij doen dit door te laten zien dat de Bayesiaanse priorverdeling omgerekend kan worden naar een equivalente hoeveelheid gegevensgerichte werkzaamheden, die op zijn beurt weer gelijk is een verlaging van het IR en het IBR. In dit artikel houden wij deze laatste lijn aan: het ARM is een denkmodel waarin Bayesiaanse technieken kunnen worden ondergebracht. Het gebruik van het ARM als denkmodel sluit ook aan bij de toepassing in de NV COS. De NV COS gebruikt het ARM om de risico-inschattingswerkzaamheden zoals beschreven in NV COS 315 te sturen en op basis daarvan de risico's van een afwijking van materieel belang te bepalen en de controlewerkzaamheden.

2.2 Antwoord op de uitdagingen; continuous improvement naast efficiënt zekerheid kwantificeren

Het kwantificeren van zekerheid is door ons als een van de uitdagingen voor het accountantsberoep benoemd. De toepassing van het 'push left'-principe zorgt voor concrete zekerheid omdat elk risico in figuur 1 gekwantificeerd wordt. Als een accountant de werking van een controle vaststelt middels een steekproef van 30 controles waarin hij 0 fouten vindt, dan weet hij met 95% zekerheid dat er niet meer dan 10% van de controles faalt. Waar zo'n kwantificering niet mogelijk is, schrijft het voorzichtigheidsbeginsel een risico van 100% voor. Maar er is meer: de accountant dwingt zich-

Naast deze twee toepassingen geeft het 'push left'-beginsel ook de handvatten aan de accountant en zijn cliënt om meer zekerheid te krijgen uit de controlewerkzaamheden, met name door

De externe accountant bevindt zich het meest rechts in de keten, terwijl hij via zijn controlemix wel de risico's in de hele keten in kaart brengt. Hij heeft dus optimaal inzicht om aan het management te duiden op welke wijze een beweging naar links te maken. Naast het afgeven van een goedkeurende verklaring als daar aanleiding toe is, initieert hij zo ook verbetering bij zijn cliënt en kan hij invulling geven aan de signalerende rol waar de maatschappij om vraagt, namelijk inzicht geven in de risico's die een onderneming loopt. Dit zijn de zogenaamde vernieuwende inzichten waarvoor de accountant meer waardering zal oogsten dan zijn goedkeurende verklaring, waarbij de trigger ligt in het continu verbeteren van zijn eigen controleproces. Hiermee wordt op een proactieve wijze invulling gegeven aan de communicatie met de cliënt die prima past in de onderwerpen die in NV COS 260 worden besproken als aangelegenheden die gecommuniceerd moeten worden aan degenen belast met gover-

nance. Zo weerstaat de accountant de neerwaartse druk op het budget waardoor hij een deel van het budget kan alloceren aan de focus op veranderingen ten opzichte van het vorige jaar.

Casus Uitkeringsinstantie

Een uitkeringsinstantie slaagde erin om voor een bepaald type uitkering een managementrapportage te maken met daarin niet alleen de klassieke kengetallen over aantallen en volumes over de tijd, maar ook een uitsplitsing naar het soort vaststellingen dat had geleid tot een uitkering. Op basis van beperkte steekproeven in met name de meest complexe vaststellingen werden de financiële foutpercentages per soort vaststelling ook gerapporteerd. Deze steekproeven vervingen de omvangrijke lijncontroles, die alleen op procedurele fouten waren gefocust. Deze rapportage diende zowel de externe verantwoording over de rechtmatigheid van de uitkeringen als de verbetering van de interne beheersing. De externe accountant, voor wie dit type uitkering van materieel belang was, voerde alleen een kwaliteitsreview uit op de door de uitkeringsinstantie verrichte controles, alvorens er op te kunnen steunen.

Casus Pensioenverzekeraar

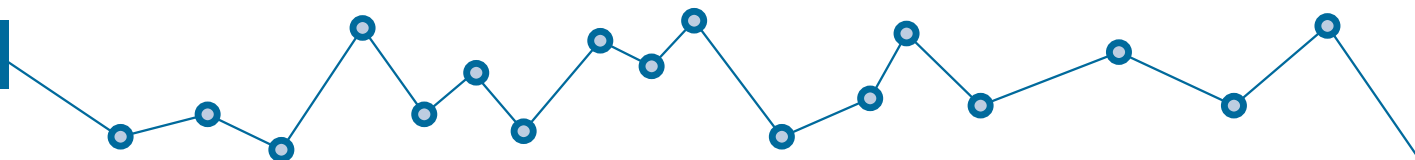
In het kader van haar toezichtsproject Quinto vroeg DNB een aantal pensioenverzekeraars om inzicht te geven in de juistheid van de aanspraken in de administratie. Een slimme data-analyse toonde welke mutaties in welke systemen het meest foutgevoelig waren en nader bekeken moesten worden. Met name voor complexe tailor made pensioenproducten bleek het herleiden van de mutaties in aanspraken naar pensioencontracten een ingewikkelde en arbeidsintensieve klus. Het inzicht in de controle-inspanning en foutgevoeligheid van complexe producten overtuigde het management van het nut en de noodzaak van productrationalisatie en vormde daarmee een waardevol hulpmiddel in de besluitvorming over de te voeren productportfolio. In termen van het 'push left'-principe, het risico dat aanspraken niet kloppen werd deels gemitigeerd door het inherente risico te verkleinen.

Die focus zal weer leiden tot vernieuwende inzichten en nieuwe bewegingen naar links. Kortom, de accountant maakt zichzelf relevant 'beyond-compliance'. Het 'push left'-principe is een verfijning van het ARM op drie vlakken. Ten eerste zien we de vier risico's IR, IBR, CAR en SR niet als onafhankelijke componenten die simpelweg vermenigvuldigd kunnen worden, maar als conditionele kansen. Dit maakt het mogelijk om steekproefbevindingen naar links te duwen,

bijvoorbeeld door interne beheersingsprocessen aan te passen. Daarnaast kan door middel van stratificatie het 'push left'-principe worden toegepast op verschillende deelpopulaties. Ten derde worden risico-inschattingen binnen het 'push left'-principe weergegeven door middel van kansverdelingen, in plaats van puntschattingen. Hierdoor kan de onzekerheid van inschattingen expliciet meegenomen worden. In principe is het ARM vrij eenvoudig te generaliseren op de eerste twee vlakken, maar voor de verfijning in het kwantificeren van risico's door middel van kansverdelingen zijn aanvullende, Bayesiaanse technieken vereist. In de volgende paragraaf gaan we hier dieper op in.

3 Kwantificering in het 'push left'-principe

Volgens het 'push left'-principe zou vanwege de toenemende efficiency van de controlemix steeds meer zekerheid links in de keten moeten ontstaan. Tegelijkertijd kwantificeren klassieke, gegevensgerichte steekproeven juist het ACR rechts in de keten. Er is dus een verdieping in de klassieke steekproeftechnieken nodig om de zekerheid over de hele keten heen te kunnen kwantificeren. Het gebruik van voorinformatie bij het plannen van steekproeven vormt hierin de sleutel. Deze voorinformatie dient om de te controleren populatie op te delen in deelpopulaties die elk een eigen risico op fouten hebben. Dit proces heet stratificatie en maakt steekproeven veel gericht-



ter en dus efficiënter door de steekproefomvang per deelpopulatie te variëren. Data-analytics zijn onmisbaar voor de extractie van deze deelpopulaties uit de totale populatie. Doordat steeds meer informatie waaraan transacties getoetst kunnen worden elektronisch beschikbaar komt, ontstaan er steeds meer deelpopulaties waarvoor helemaal geen steekproef benodigd is omdat op basis van data-analytics volledige zekerheid verkregen kan worden. In de praktijk zien we nog veel accountants worstelen met deze verdieping van klassieke steekproeven c.q. het toepassen van (gestratificeerde) steekproeven. Ten eerste komt dit doordat de technische ondersteuning bij de evaluatie van steekproeven vaak tekortschiet. Het doen van uitspraken per deelpopulatie op basis van steekproeven en deze consolideren tot één uitspraak, is een weerbarstig wiskundig probleem dat nog niet volledig is opgelost. Dankzij het werk van Stewart en Kinney (2013) is de oplossing van dit probleem wel een stuk dichterbij gekomen. Ten tweede bestaat er onder statistici onenigheid over de manier waarop het steekproefrisico bepaald moet worden. Bayesiaanse statistici houden er een andere mening op na dan klassieke statistici. Over dit onderwerp is onder accountants al diverse malen discussie gevoerd. In deze paragraaf stellen wij de Bayesian Evaluation Tool for Auditors (BETA) voor, een op Bayesiaanse statistiek gebaseerde methode om gestratificeerde steekproeven te evalueren.

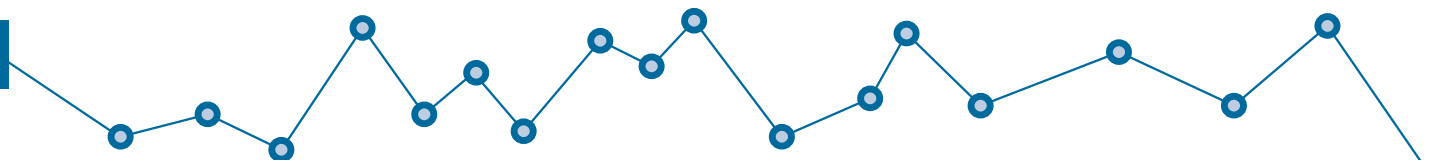
3.1 Klassiek versus Bayesiaans

De aanhangers van de klassieke statistiek gebruiken het al eerder besproken Audit Risk Model om het maximale steekproefrisico te bepalen. Vervolgens voeren ze eventueel een steekproef uit en berekenen de kans dat het steekproefresultaat uit een geldmassa komt met precies één materiële fout. Als deze kans kleiner is dan het maximale steekproefrisico, keuren ze de geldmassa goed. De audit guides van de meeste accountantskantoren en ook veelgebruikte softwarepakketten als IDEA en ACL hanteren deze aanpak. Bayesianen stellen dat het interessanter is om de kans op een foutfractie in de geldmassa te berekenen op basis van waargenomen informatie. Zij berekenen een kansverdeling voor de fout in de geldmassa (de zogenaamde posterior) door voorinformatie (de zogenaamde prior) te verrijken met de steekproefuitslag (de zogenaamde likelihood). Het voorbeeld in kader 2 geeft aan hoe deze begrippen in de praktijk werken. Deze aanpak is onder accountants minder gangbaar en wordt ook slechter ondersteund door software. Daarnaast slagen Bayesianen er niet altijd in om de onduidelijkheid over het formuleren van een prior weg te nemen. Dit is een nadeel, want de meeste accountants denken niet in termen van een prior maar in inschattingen op een driepuntschaal Hoog, Gemiddeld, Laag voor de risico's uit het ARM.

3.2 Bayesian Evaluation Tool for Auditors

Zoals de naam al doet vermoeden maakt BETA gebruik van betaverdelingen¹. Hoewel BETA uitgaat van Bayesiaanse statistiek, vraagt de tool alleen om de risico-inschattingen uit het ARM en vertaalt deze vervolgens onder de motorkap naar de priorverdeling. Deze vertaling maakt ook gebruik van de op voorhand meest waarschijnlijke foutfractie en de materialiteit. Omdat bestaande methoden deze inschatting ook al gebruiken om de steekproefomvang te bepalen, is dit geen beperking. Stel de prior is $Beta(a_prior, b_prior)$ verdeeld en er is een steekproef getrokken van omvang n waarin k fouten zijn aangetroffen². Dan ligt de verdeling van de foutfractie in de geldmassa vast: deze is $Beta(a_prior + k, b_prior + n - k)$ verdeeld. Dit was al bekend. Het nieuwe van BETA is dat de twee parameters van de priorverdeling, a_prior en b_prior , berekend worden uit de inschattingen voor IR, IBR en CAR. Dit werkt als volgt. Eerst berekenen we welke virtuele steekproef als binomiale likelihood van een niet-informatieve prior een posteriorverdeling maakt waarin de materialiteit een overschrijdingskans heeft die gelijk is aan het steekproefrisico zonder voorkennis³. Dan vertalen we de drie risico-inschattingen uit het ARM naar een steekproefrisico met voorkennis. In het voorbeeld in kader 2 (zie pag. 241) hebben wij hiervoor de waarden genomen zoals gehanteerd door het IODAD (2007), maar deze parameters kan de gebruiker zelf wijzigen. Vervolgens berekenen we de omvang van een tweede steekproef die als binomiale likelihood van de niet-informatieve prior een posterior-

verdeling maakt waarin de materialiteit een overschrijdingskans heeft die gelijk is aan het steekproefrisico met voorkennis. Het verschil tussen de twee steekproefomvangs vatten we op als de omvang van de steekproef die equivalent is met de voorkennis. Die rekenen we terug naar een niet-triviale priorverdeling. Deze prior geeft nu samen met de tweede steekproef als binomiale likelihood per definitie de gewenste posteriorverdeling. De Casus DBC geeft een voorbeeld van deze berekening (zie kader 2 op pag. 241). De geschetste BETA-methode kan ook in meer complexe situaties gebruikt worden. Zo kunnen dankzij de mooie eigenschappen van convoluties van betaverdelingen gestratificeerde steekproeven geëvalueerd worden. De afzonderlijke betaverdelingen per deelpopulatie worden dan samengevoegd tot één betaverdeling die de kansverdeling van fouten in de gehele populatie weergeeft.



Casus DBC

Als voorbeeld beschouwen we de controle van een DBC-geldmassa. Volgens het protocol van NZa (2011) moet een 95%-betrouwbare foutbovengrens van 5% worden aangetoond. Stel dat de meest waarschijnlijke fout op basis van het verleden 2,5% bedraagt. We beginnen met de statistisch eenvoudige situatie waarin we de benodigde steekproefomvang plannen zonder voorkennis over IR, IBR en CAR. We schatten deze drie risico's dus Hoog in. Volgens de klassieke methode bedraagt de benodigde steekproefomvang dan 240. Dit is 9,1% hoger dan de steekproefomvang van 220 waarop BETA uitkomt. Tabel 1 geeft de uitgangspunten en berekeningen aan voor de twee verschillende methoden.

Tabel 1 Steekproefplanning met klassieke methode en BETA voor DBC-casus zonder voorkennis

Methode	Uitgangspunt	Berekening	Steekproefomvang
Klassiek	De kans dat een steekproef met 2,5% fout uit een DBC geld-massa komt met 5% fout moet kleiner zijn dan 5%.	De binomiale verdeling met $n = 240$ en $p = 5\%$ geeft een kans van 5% voor $k \leq 2,5\% * 240 = 6$.	240
BETA	De kansverdeling voor de fout in de DBC geldmassa op basis van een steekproef met 2,5% fout moet een 95%-betrouwbare bovengrens hebben van 5%.	De Beta(1;1) verdeling als prior met een binomiale likelihood met $n = 220$ en $p = 2,5\%$ geeft $k = 5,5$ en een Beta(6,5; 215,5) als posterior waarvan het 95%-percentiel 5% is.	220

Stel dat we nu over de voorkennis beschikken dat het risico dat een materiële fout niet door de interne beheersing wordt opgemerkt Gemiddeld is. Dit komt volgens het IODAD-handboek overeen met IBR = 60%. Tabel 2 (zie pag. 242) laat zien hoe BETA leidt tot een steekproefomvang van 169 en de klassieke methode tot een steekproefomvang van 190, wat 12,5% minder efficiënt is dan BETA.

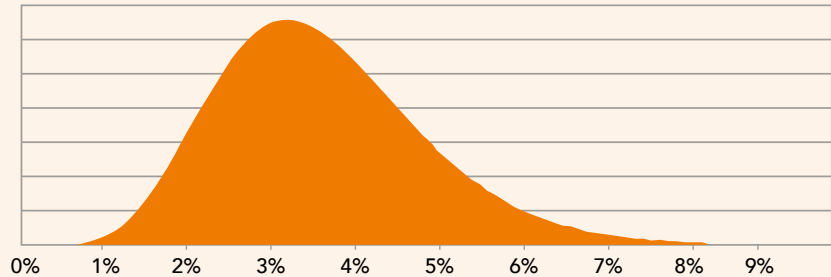
Vervolg casus DBC

Tabel 2 Steekproefplanning met klassieke methode en BETA voor DBC-casus met voorkennis

Methode	Uitgangspunt	Berekening	Steekproefomvang
Klassiek	Steekproefrisico mag 5%/60% = 81/3% zijn. De kans dat een steekproef met 2,5% fout uit een DBC geldmassa komt met 5% fout moet dus kleiner zijn dan 81/3%.	De binomiale verdeling met $n = 190$ en $p = 5\%$ geeft een kans van 81/3% voor $k \leq 2,5\% * 190 = 5$.	190
BETA	De posteriorkansverdeling voor de fout in de DBC geldmassa op basis van een niet-triviale prior, waarin de voorkennis over IBR verwerkt is, en een steekproef met 2,5% fout moet een 95%-betrouwbare bovengrens hebben van 5%.	De Beta(2,275; 50,725) verdeling komt als prior overeen met de voorkennis IBR = 60%. De binomiale likelihood met $n = 169$ en $p = 2,5\%$ geeft $k = 4,225$. Dit leidt tot een Beta (6,5; 215,5) als posterior waarvan het 95% percentiel 5% is.	169

Noot: De afleiding is als volgt: De Beta(1;1) verdeling als prior met een binomiale likelihood verdeling met $n = 169$ en $p = 2,5\%$ geeft $k = 4,225$ en dus een Beta (5,225;165,775) verdeling als posterior. Deze verdeling heeft het 912/3% percentiel bij 5%. Dit betekent dat de voorkennis equivalent is met een steekproef van $220 - 169 = 51$ met daarin 2,5% fouten. Dit levert $k = 2,5\% * 51 = 1,275$. Dus moet de prior die de voorkennis vertegenwoordigt de Beta (2,275; 50,725) verdeling zijn.

Kansverdeling van het foutpercentage in de DBC geldmassa bij IBR = Medium, steekproefomvang 169 en 6 gevonden fouten



Figuur 2 Informatie over DBC geldmassa op basis van Bayesiaanse steekproef

Vervolg casus DBC

Om te illustreren dat de Bayesiaanse aanpak in BETA beter aansluit bij de werkwijze van accountants, veronderstellen we nu dat er 6 fouten worden aangetroffen in de steekproef in plaats van de verwachte 4,225. In de klassieke benadering levert dit doorgaans discussie op vanuit de redenering dat $6/169 = 3,56\%$ kleiner is dan 5%. De Bayesiaanse aanpak levert de Beta(8,275; 213,725) verdeling op als kansverdeling voor de foutfractie in de DBC geldmassa. Deze kansverdeling staat in figuur 2 (zie pag. 242) en kan de discussie over het vervolg faciliteren, bijvoorbeeld door te constateren dat de 95%-betrouwbare foutondergrens 1,9% bedraagt en de 95%-betrouwbare foutbovengrens 6,1%. Meer specifiek, deze kansverdeling kan – door toevoeging van verliesfuncties voor het ten onrechte goed- of afkeuren van populaties – worden gebruikt voor het berekenen van de maatschappelijk optimale controlemix.

3.3 Overige toepassingsmogelijkheden

Het gebruik van BETA maakt ook andere toepassingen mogelijk. In het geval van niet-moneitair postensteekproeven kan BETA een discrete predictieve kansverdeling leveren, die vooral bij kleine populaties efficiënter is. Ook hebben we een generalisatie naar gebroken fouten gemaakt, zodat – net als in de Stringer-bound-methode – onderscheid gemaakt kan worden tussen bijvoorbeeld twee halve fouten en één hele fout. Tot slot is het mogelijk om bij de beslissing of de geldmassa goed- of afgekeurd wordt, het maatschappelijk nut te maximaliseren. Volgens Wille (2003) hangt dit nut niet alleen af van de kosten voor het uitvoeren van de steekproef en het risico op onterecht goedkeuren, maar ook van het risico op onterecht afkeuren. Met name dit laatste risico blijft vaak onderbelicht (zie De Swart, 2004). Dit maakt BETA tot een tool die niet alleen praktischer en efficiënter is dan de klassieke aanpak, maar ook robuuster onder verschillende scenario's.

4 Projectie op controlestandaarden

Data-analytics spelen, zoals vermeld in de inleiding, een rol in diverse onderdelen van de controleaanpak. De vraag is nu of de controlestandaarden op die onderdelen het gebruik van data-analytics in voldoende mate stimuleren. Gegevensgerichte werkzaamheden, waaronder steekproeven, maken een belangrijk onderdeel uit van een controlemix omdat deze verplicht zijn voor elke stroom transacties, jaarrekening-post en toelichting van materieel belang, voor elk risico van materieel belang dat als significant is gekwalificeerd (NV COS 330). Steekproeven zijn een heel effectieve en efficiënte keuze ingeval een voorraad als een materiële jaarrekeningpost is gekwalificeerd of bijvoorbeeld als onderdeel van de controle van de opbrengsten als die uit veel losse transacties bestaan. Ook als onderdeel van de gegevensgerichte maatregelen, ter toetsing van de effectieve werking van een interne beheersingsmaatregel kan een steekproef worden gebruikt (Majoor en Van Kollenburg, 2011b, p. 132). Bij het

gebruik van steekproeven kan een accountant niet volstaan met het louter berekenen van de fout in de steekproef, want volgens NV COS 530-7 is het noodzakelijk om het steekproefrisico te beschouwen. Dit geldt ook voor steekproeven in het kader van rechtmatigheidscontroles zoals beschreven in de eerste twee voorbeelden in paragraaf 1.1.1. Hier moet namelijk met een zekere betrouwbaarheid vastgesteld worden dat de onrechtmatige betalingen niet meer dan een van tevoren vastgesteld percentage van het totaal vormen. De controlestandaarden adresseren data-analytics dus wel, maar niet op een manier die leidt tot een optimale afstemming tussen diverse soorten data-analytics, zoals wij voorstellen in het 'push left'-principe. Wij achten aanscherping van de controlestandaarden op dit punt dan ook wenselijk.

5 Samenvatting en conclusies

Met data-analytics toegepast in het 'push left'-beginsel krijgt het Audit Risk Model een nieuwe dynamiek. Dit geeft de accountant een handvat om een adequaat antwoord te geven op een aantal uitdagingen waarmee het accountantsberoep worstelt. Zo geeft het de mogelijkheid om zekerheid te kwantificeren, de controle effectiever en efficiënter te maken.

Daarnaast geeft het ook een handvat voor de belangrijkste uitdaging die het accountantsberoep op dit moment heeft, namelijk het antwoord geven op de maatschappelijke vraag naar een accountant die kritisch is en de ondernemingen

wijst op risico's die de financiële continuïteit van een onderneming bedreigen. Op basis van de risico-indeling van het ARM kan de accountant met behulp van data-analytics daadwerkelijk bijdragen aan inzicht in de risico's die van invloed zijn op de toekomstige financiële continuïteit en kan hij ook daadwerkelijk zijn signalerende rol vormgeven. Om dit te bewerkstelligen is aanscherping van de controlestandaarden wel wenselijk.

Transparantie en coherentie in de kwantificering van zekerheid en voorkennis is onontbeerlijk voor het implementeren van het 'push left'-principe. De in paragraaf 3 geïntroduceerde Bayesian Evaluation Tool for Auditors combineert de intuïtie en flexibiliteit van Bayesiaanse statistiek met de gangbare inschattingen van risico's binnen het ARM. De geschetste BETA-methode kan echter ook in meer complexe situaties gebruikt worden.

Literatuur

ABN AMRO Bank (2012). *Visie op zakelijke dienstverlening - Accountantskantoren*. Geraadpleegd op www.abnamro.nl/nl/zakelijk/visie.

Adviescommissie Herziening Eindtermen (2012). *Een accountantsopleiding op maat voor het maatschappelijk verkeer*. Geraadpleegd op www.nba.nl.

Batenburg, P.C. van, & R.J.M. Dassen (1996). Het Bayesiaanse model van Deloitte & Touche ter ondersteuning van de controlemix. *De Accountant*, jg. 103, nr. 1, september, p.p.31-35.

Blokdijk, J.H. (2001). De effectiviteit van de systeemgerichte aanpak in de accountantscontrole. *Maandblad voor Accountancy en Bedrijfseconomie*, 75(3), 71-80.

Eimers, P.W.A. (2006). Het Audit Risico Model is springlevend!, *Maandblad voor Accountancy en Bedrijfseconomie*, 80(3), 76-83.

Eumediton (2011). *Eumediton response to the consultation paper on Enhancing the value of auditor reporting: Exploring options for change*. Amsterdam. Geraadpleegd op <http://www.ifac.org>.

European Commission (2010). *Audit Policy: Lessons from the Crisis*. Green paper, Brussel. Geraadpleegd op <http://eur-lex.europa.eu/>.

Europees Fonds voor Regionale Ontwikkeling (EFRO) (2012). *Controleprotocol voor de huisaccountant van de begunstigde voor de controle van het financiële verslag van de eindrapportage voor definitieve vaststelling van subsidies uit het Europees fonds voor regionale ontwikkeling*. Geraadpleegd op <http://www.nba.nl>.

Interdepartementaal Overlegorgaan Departementale Accountantsdiensten (2007). *Handboek Auditing Rijksoverheid 2007*, vastgesteld door het Interdepartementaal Overlegorgaan Departementale Accountantsdiensten (IODAD) op 28 maart 2006 en 29 mei 2007.

Kloosterman, H.H.W. (2004). Wat is eigenlijk risicoanalyse in de accountantscontrole? *Maandblad voor Accountancy en Bedrijfseconomie*, 78(12), 570-578.

Knechel, R. (1998). *Auditing: Text & cases*. First edition. London: International Thomson Publishing.

Kothari, S.P., & R. Lester (2011). *The role of accounting in the financial crisis: Lessons for the future*. Report 1972354. MIT Sloan School of Management. Geraadpleegd op <http://ssrn.com/abstract=1972354>.

Majoor, B., & J. van Kollenburg (2011a) (red.). *Elementaire theorie accountantscontrole. Grondslagen van Auditing & Assurance*, zesde druk. Groningen/Houten: Noordhoff Uitgevers.

Majoor, B., & J. van Kollenburg (red.) (2011b). *Elementaire theorie accountantscontrole. De praktijk van auditing & assurance*, zesde druk. Groningen/Houten: Noordhoff Uitgevers.

Mollema, K.Y. (2003). Auditrisico, meer dan ooit een issue (1). *Maandblad voor Accountancy en Bedrijfseconomie*, 77(12), 551-556.

Mollema, K.Y. (2004). Auditrisico, meer dan ooit een issue (2). *Maandblad voor Accountancy en Bedrijfseconomie*, 78(1/2), 5-15.

NV COS (2013). *HRA deel 1A - NV COS* (geldend per 15 januari 2011). Geraadpleegd op <http://www.nba.nl/Wet-en-regelgeving/Beroepsregels/HRA/>.

Nyenrode (2011). *MKB master scriptie NU traject*. Geraadpleegd op www.nyenrode.nl.

Nederlandse Zorg Autoriteit (2011), *Protocol vereveningsonderzoek Zvw met oplevering in 2012*. Geraadpleegd op www.nza.nl.

Stewart, T.R., & Jr., W.R. Kinney Jr. (2013). Group audits, group-level controls, and component materiality: How much auditing is enough? *The Accounting Review*, 88(2), 707- 737.

Swart, J.J.B. de (2004). Hoe controleert de Belastingdienst met steekproeven verstrekkingen aan werknemers? *Praktijkblad Arbeidsvoorwaarden*, 2004(9), 18-19.

Swart, J.J.B. de, Briene, A, & A. Bakkum (2011). Data-analyse essentieel onderdeel van de controle. *Spotlight*, 18(1), 18-21. Geraadpleegd op <http://www.pwc.nl/nl/spotlight/assets/documents/spotlight2011-1-3-data-analyse-essentieel-onderdeel-van-de-controle.pdf>.

Touw, P., & L. Hoogduin (2011). *Statistiek voor audit en controlling*. Schoonhoven: Academic Service.

Tuyl, F., R. Gerlach & K. Mengersen (2009). Posterior predictive arguments in favor of the

Bayes-Laplace prior as the consensus prior for binomial and multinomial parameters. *Bayesian Analysis*, 4(1), 151-158.

Tweede Kamer der Staten-Generaal (Commissie De Wit) (2010). *Parlementair onderzoek financieel stelsel, Deelonderzoek 1 en 2*. Den Haag. Geraadpleegd op www.tweedekamer.nl.

Veenstra, R., & A. Heertje (2006). Doeltreffende administratieve controle door steekproeven. *Maandblad voor Accountancy en Bedrijfseconomie*, 80(12), 611-619.

Wille, F.J. (2003). *Auditing using Bayesian decision analysis*. Amsterdam: Tinbergen Institute.

Womack, J.P. & D.T. Jones (2003). *Lean thinking: Banish waste and create wealth in your corporation*. London, Simon & Schuster.

Noot

- 1 Het voordeel van beta-verdelingen is dat in de Bayesiaanse interpretatie een beta-verdeling een zogenaamde geconjugeerde prior is, dat wil zeggen dat een beta prior-verdeling gecombineerd met een binomiale likelihood-verdeling leidt tot een posterior die weer beta verdeeld is.
- 2 Deze parameter k mag ook een gebroken getal zijn.
- 3 Overigens is er enige discussie over wat de minst informatieve prior is, maar uit Tuyl et al. (2009) blijkt dat Beta (1,1), die notabene door dominee Bayes zelf werd voorgesteld, in ieder geval een goede kandidaat is.

Het Bayesiaanse model en Tax Assurance¹

Paul van Batenburg en Robbert Hoyng

In dit artikel over het gebruik van steekproeven in fiscale controles gaan wij in op het wiskundige model waarmee de reductie in de steekproefomvang kan worden berekend.

Zoals eerder vermeld, is het Bayesiaanse model niet bedoeld om, en niet in staat om subjectieve beslissingen over te nemen². Er bestaat geen formule die aangeeft welke reductie past bij welke mate van volgroeiheid van een Tax Control Framework. Er bestaat wel een formule die ervoor zorgt dat dergelijke subjectieve beslissingen transparant en consistent zijn.

In eerdere artikelen is laten zien op basis van welk model de reductie tot stand komt in de Controle-aanpak Belastingdienst. Dat model is gebaseerd op het vinden van nihil fouten omdat ervan wordt uitgegaan dat het Tax Control Framework geen fouten meer doorlaat. Zodra er in de gereduceerde steekproef toch fouten worden gevonden is de basis onder de genomen reductie ter discussie.

Zo'n aanname helpt goed bij het bepalen van een zo klein mogelijke steekproef, maar de vraag is of die aanname realistisch is. Met name in de begintijd van het leerproces van het TCF is het niet

ondenkbaar dat er fouten, foutjes, zullen optreden. En als daar geen ruimte voor is, dan moet er heel veel extra werk gedaan worden. De oorzaak van de fout moet in kaart worden gebracht en alle soortgelijke transacties waar die oorzaak ook mogelijk aanwezig was, moeten nader worden onderzocht. Pas als dat een bedrag aan correcties oplevert dat de meest waarschijnlijke fout uit de steekproef benadert, is de fout naar onze mening afdoende geïsoleerd en kan de aangifte alsnog goedgekeurd worden.

Om dat extra werk te voorkomen is het in veel gevallen aan te raden om niet uit te gaan van een minimumberekening, maar bij de opzet van de steekproef wat meer te doen dan strikt noodzakelijk is: liever vooraf gepland wat extra werk, dan achteraf ongepland veel extra werk. Het in dit artikel getoonde model maakt van het steunen op de werking van het TCF geen zwart/wit beslissing: er zijn ook grijs tinten in de zin van een te controleren bestand waar vast nog wel wat mis is, maar waar de interne beheersing er wel voor zorgt dat er geen grote fouten doorglijpen.

De geldsteekproef

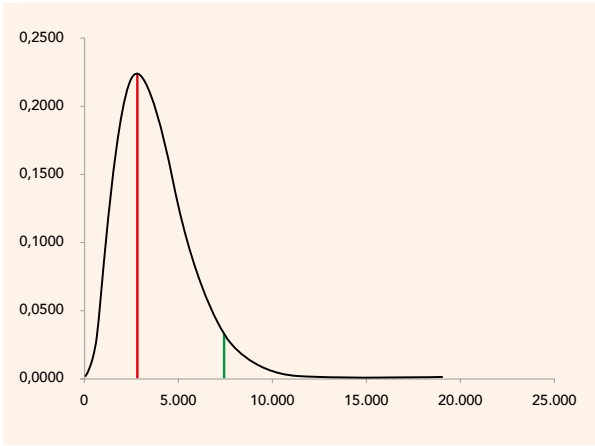
Hieronder geven wij een tabel voor de opzet en evaluatie van steekproeven. Bij de opzet wordt eerst de materialiteit bepaald, en wordt een interval berekend door die materialiteit te delen door het getal R uit die tabel. Dat getal wordt bepaald door het aantal toegestane fouten in de steekproef, en de tabel gaat uit van de keuze om een 95%-betrouwbare uitspraak te doen. Met dat interval wordt de steekproef gestoken: elk interval levert één waarneming op. Bij de evaluatie gebruiken we deze tabel weer, maar nu met het aantal gevonden fouten, om de R-waarde te bepalen die vermenigvuldigd met het gebruikte interval de maximale fout oplevert. Als die maximale fout beneden de materialiteit ligt kan de gecontroleerde populatie worden goedgekeurd.

Aantal fouten	R
0	3,00
1	4,75
2	6,30
3	7,76
4	9,16
5	10,52

Tabel 1 Tabelwaarden voor de opzet en evaluatie van steekproeven.

De meest waarschijnlijke fout in de populatie ontstaat door hetzelfde interval te vermenigvuldigen met het aantal gevonden fouten in de steekproef. Die meest waarschijnlijke fout is het werkelijke kwaliteitsoordeel over de populatie; de maximale fout geeft alleen aan of de controle toereikend is geweest om dat oordeel te geven.

Hieronder tonen wij een grafiek van de mogelijke fout in de populatie op grond van de bevindingen uit de steekproef. Daar zien wij de meest waarschijnlijke fout (op de horizontale as ter hoogte van de verticale lijn tussen 0 en 5.000) en de maximale fout (op de horizontale as ter hoogte van de verticale lijn tussen 5.000 en 10.000).



Figuur 1 De kansverdeling van de onbekende fout in de te controleren massa gegeven de steekproef-uitkomsten

Bayesiaanse statistiek

Het bijzondere aan Bayesiaanse (spreek uit als Beez, zoals de zangeres Joan Baez, en niet als bajes zoals gevangenis) statistiek ten opzichte van de 'klassieke' statistiek is dat er aannames gedaan kunnen worden over de te onderzoeken populatie die wiskundig gecombineerd worden met steekproefresultaten uit die populatie tot een oordeel over die populatie.

Google op 'de Vos wiegendood' en u wordt doorgeleid naar in onze ogen de allerbeste uitleg van Bayesiaanse statistiek die ooit een krant gehaald heeft (NRC, oktober 1987).

In de fiscale controle toegepast kunnen we vooraf (prior) een aanname doen over de vorm van figuur 1, vervolgens een steekproef trekken en controleren, en dan achteraf (posterior) figuur 1 opnieuw bepalen op basis van zowel de aannames als de resultaten. Dat betekent ook dat we kunnen afspreken hoe figuur 1 er posterior uit moet zien, en als we denken te weten hoe die figuur er prioruit ziet, kunnen bepalen welke steekproefomvang er nodig is en hoeveel fouten daarin mogen worden gevonden.

Het wiskundige model dat wij hier bespreken is afkomstig uit het proefschrift van Trevor Stewart, een in 2009 gepensioneerd partner van Deloitte USA (Stewart 2013).

En dan nu de wiskunde!

Wij nemen aan dat figuur 1 een voorbeeld is van een Gammaverdeelde variabele. De formule om figuur 1 te maken is tamelijk indrukwekkend en zullen wij daarom niet opnemen. Belangrijk is dat die formule bepaald wordt door twee te kiezen getallen: a en b . Als je die twee kiest, ligt het figuur vast en andersom. Het getal a beschrijft de kwaliteit van de populatie: hoe lager a des te minder fouten men verwacht of des te minder fouten men heeft ontdekt. Het getal b beschrijft de kwaliteit van de controle: hoe lager b des te meer werk men moet gaan doen of heeft gedaan.

De meest waarschijnlijke fout is nu te berekenen als $(a-1)*b$ en voor de maximale fout is een Excel-functie beschikbaar: `GAMMA.INV(0,95;a;b)`.

We gaan nu uit van een situatie waarin men op grond van subjectieve voorkennis de parameters a en b kiest: a zegt hoeveel fouten men verwacht, b zegt hoeveel werk er in zit om die verwachting te onderbouwen. Als we nu uit een populatie van omvang M een steekproef trekken van n waarnemingen en daarin k fouten vinden, verandert die aangenomen figuur in een nieuwe figuur met nieuwe getallen:

a verandert in $a+k$

b verandert in $1/(1/b+n/M)$

De meest waarschijnlijke fout is nu $(a+k-1)/(1/b+n/M)$ en de maximale fout is te berekenen als `GAMMA.INV(0,95;a+k;1/(1/b+n/M))`.

Een getallenvoorbeeld, dan maar? We controleren een populatie van $M = 1.000.000$ met een materialiteit van 10.000. Bij 95% betrouwbaarheid en 0

toegestane fouten zou daar een steekproef van 300 waarnemingen voor nodig zijn: $R = 3$ dus $n = 3 \times 1.000.000/10.000 = 300$.

Nu denken wij op grond van de beoordeling van de Interne Beheersing dat de fout in deze populatie niet groter kan zijn dan 30.000. Het BRAM-model (zie aflevering 3) geeft als formule:

$$n = 3M/PM - 3M/PM^* = 3 \times 1.000.000/10.000 - 3 \times 1.000.000/30.000 = 200$$

Het model dat wij nu tonen zegt:

- de prior heeft een a van 1 en een b van 10.000 zodat de meest waarschijnlijke fout is $(a-1)b = (1-1)10.000 = 0$ en de maximale fout is $\text{GAMMAINV}(0,95;1;10.000) = 30.000$ (afgerond)
- de posterior die wij willen afleiden heeft een a van 1 en een b van 3.333, zodat de meest waarschijnlijke fout weer 0 is (die kan natuurlijk niet van de waarde van de prior afwijken) en de maximale fout gelijk is aan de materialiteit: $\text{GAMMAINV}(0,95;1;3.333) = 10.000$ (afgerond)
- dat betekent dat $k = 0$ (nieuwe a minus

oude a) moet worden en dat

$$1/(1/10.000 + n/1.000.000) = 3.333 \text{ dus } n = 200.$$

De uitkomst van dit model is dus dezelfde als van BRAM, omdat we geen fouten verwachten. Is dit model nu omslachtiger dan BRAM? Zeker. Maar de kracht van het model is dat het ook werkt als het aantal toegestane fouten ongelijk 0 is. Alleen, dan is een getallenvoorbeeld niet meer zo simpel door te rekenen en is er op zijn minst een Excel toepassing nodig.

Op www.steekproeven.eu is zo'n Excelsheet te vinden. Als u daar kijkt zal het opvallen dat wij de uitkomsten voor de leesbaarheid in dit artikel een beetje hebben afgerond.

Dit was het eerste voorbeeld, zie onder:

Als we ermee rekening houden dat de populatie niet brandschoon is maar bijvoorbeeld 1.000 (10% van de materialiteit) aan fouten bevat, wordt de benodigde steekproef groter, maar mogen we ook foutjes (0,26 totale foutfractie) vinden, zie pag. 251 boven:

population	1.000.000						1 define population and materiality
materiality	1.000						
Prior			Sample		Posterior		
a	1		n	200	a	1	
b	10.000		errors	0	b	3.333	
Most Likely	0				Most Likely	0	
Maximum	30.000				Maximum	10.000	
2 choose a and b to make Most Likely and Maximum consistent with prior knowledge				3 find sample size to yield Maximum below materiality			

population	1.000.000					1 define population and materiality
materiality	10.000					
Prior		Sample		Posterior		
a	1,11	n	263	a	1,37	
b	9.346	errors	0,26	b	2.703	
Most Likely	1.000			Most Likely	1.000	
Maximum	30.000			Maximum	10.000	
2 choose a and b to make Most Likely and Maximum consistent with prior knowledge			3 find sample size to yield Maximum below materiality			

Dit modelletje vereist nogal wat trial-and-error zoeken naar parameters die passen bij de toe te passen voorkennis. Natuurlijk zijn er meer geavanceerde tools, maar die vragen weer meer rekenkracht en zijn daarom niet geschikt om vrij uit te delen.

Als afsluiting stellen wij u enkele vragen die relevant zijn bij de keuze voor een steekproefaanpak.

1. Is het doel van de steekproef om een beeld over de gehele interne beheersing (monitoring van de IB) te geven, of alleen de fiscale aspecten (de monitoring van het TCF), of bent u alleen geïnteresseerd in een enkele aangifte?
2. Houdt er in het laatste geval rekening mee dat u de materialiteit nu bepaalt aan de hand van de omvang van de post en niet aan de hand van de omzet, waardoor u naar verhouding meer werk doet!

3. Denkt u dat de processen die ten grondslag liggen aan de steekproefsgewijs te controleren gegevens zo goed werken dat u foutloze steekproeven gaat krijgen, of kunt u beter in een grotere steekproef ondanks het vinden van wat foutjes nog steeds een goedkeurend oordeel afleiden?
4. Hoe past de omvang van de geïntegreerde steekproef volgens de CAB bij de capaciteit die u kunt vrijmaken voor de controle?
5. U kunt die omvang altijd verkleinen door (zwaarder) te steunen op de procedures die tezamen het TCF vormen. Maar is dat terecht? (Hoe) kunt u uw aannames waarmaken?

Literatuur

Steele, A. (1992), *Audit Risk and Audit Evidence: the Bayesian Approach to Statistical auditing*; Academic Press & the Chartered Association of Certified Accountants (ACCA).

Stewart, T.R. (2013); *A Bayesian Audit Assurance Model with Application to the Component Materiality Problem in Group Audits*, Proefschrift, Vrije Universiteit Amsterdam.

Noot

- 1 Dit artikel is eerder geplaatst in het tijdschrift Tax Assurance, editie april 2015. Het is het vierde en laatste van een reeks.
- 2 Zie ook Steele (1992) om te lezen dat Bayesiaanse statistiek 'het geloof' rekenkundig kan doen benaderen, maar niet verklaren. Wel kan met Bayesiaanse statistiek een berekening worden verklaard. Bijvoorbeeld in de zin dat de hoeveelheid werk en de uitspraak welke zekerheid men vereist leiden tot de verklaring: "Dus u vertrouwt erop dat de maximale fout niet MP^* overschrijdt?"

Auteurs



eindredactie

Hein Kloosterman

H.H.W. Kloosterman RE RA werkt als zelfstandig adviseur op het gebied van IT-audit en Statistical Audit; hij adviseert daarbij ook aan en namens de KEY Group. Hij is betrokken bij de accountantsopleiding van de Erasmus Universiteit

(ESAA) en bij de IT-auditopleiding van de VU. Hein is lid van de Stuurgroep Statistical Auditing van het Limperg Instituut.



eindredactie

Paul van Batenburg

Drs. Paul C. van Batenburg is werkzaam bij Deloitte en lid van de Stuurgroep Statistical Auditing van het Limperg Instituut.



eindredactie

Kees Voorhoeve

Dr. J.C. Voorhoeve is werkzaam in het toezicht op de banken bij De Nederlandsche Bank en werkt op persoonlijke titel mee aan deze bundel. Voorheen werkte hij tien jaar bij de adviesbureaus Bakkenist Management Consultants en

Deloitte. Hij is IBIS UvA gecertificeerd Lean Six Sigma Master Black Belt en studeerde en promoveerde in de experimentele natuurkunde aan de RU Leiden. Hij is lid van de Stuurgroep Statistical Auditing.



Ed Broeze

Dr. G.B. Broeze is gepensioneerd statistisch adviseur van de Algemene Rekenkamer. Hij was lid van de Stuurgroep Statistical Auditing van het Limperg Instituut.



Roger Dassen

Prof. dr. R.J.M. Dassen RA is CFO bij ASML en hoogleraar Auditing bij de VU te Amsterdam.



Ferry Geertman

Ir. F. Geertman RE CISA is managing director bij de KEY Group. Vanuit zijn achtergrond als toegepast wiskundige heeft Geertman ruime ervaring met data-analyse waaronder het opzetten en evalueren van steekproeven als onderdeel van de interne en externe controle.

Daarnaast heeft hij uitgebreide ervaring bij adviseren rondom en beoordelen van (IT-)risicomanagement en interne beheersing bij grote, multinationale, beursgenoteerde bedrijven. Ferry is lid van de Stuurgroep Statistical Auditing van het Limperg Instituut.



Wouter Gerards

Drs. W. Gerards doceerde na zijn afstuderen in de financiële economie in 2009 econometrie aan de Universiteit van Maastricht. Sinds 2010 is hij werkzaam bij het Agentschap Sociale Zaken en Werkgelegenheid. Zijn hoofdtaak is het

ontwerpen van steekproeven en de daarbij behorende controleonderzoeken. Hierbij heeft hij in samenwerking met Ed Broeze de huidige controlemethodiek ontwikkeld. Hij lid van de Stuurgroep Statistical Auditing van het Limperg Instituut.



Arnold Heertje

Prof. dr. A. Heertje is emeritus hoogleraar economie aan de Universiteit van Amsterdam.



Harrie Hendriks

Dr. H.W.M. Hendriks is gepensioneerd wiskundige aan de Radboud Universiteit Nijmegen. Hij heeft als specialiteit mathematische statistiek. Harrie is lid van de Stuurgroep Statistical Auditing van het Limperg Instituut.



Robbert Hoyng

Drs. R. Hoyng is Group Tax Director bij Naspers Group Company



Barbara Majoor

Prof. dr. G.C.M. Majoor RA werkt bij de AFM. Zij is als hoogleraar Accountancy verbonden aan Nyenrode Business Universiteit.



Niels van Nieuw Amerongen

Dr. C.M. van Nieuw Amerongen RA is eigenaar van advies- en trainingsbureau V&A en is als associate professor Auditing & Assurance verbonden aan Nyenrode Business Universiteit.



Jacques de Swart

Prof. dr. J.J.B. de Swart MBA is hoogleraar toegepaste wiskunde bij Nyenrode Business Universiteit en werkzaam bij de adviespraktijk van PwC. Jacques is als voorzitter lid van de Stuurgroep Statistical Auditing van het Limperg Instituut.



Ruud Veenstra

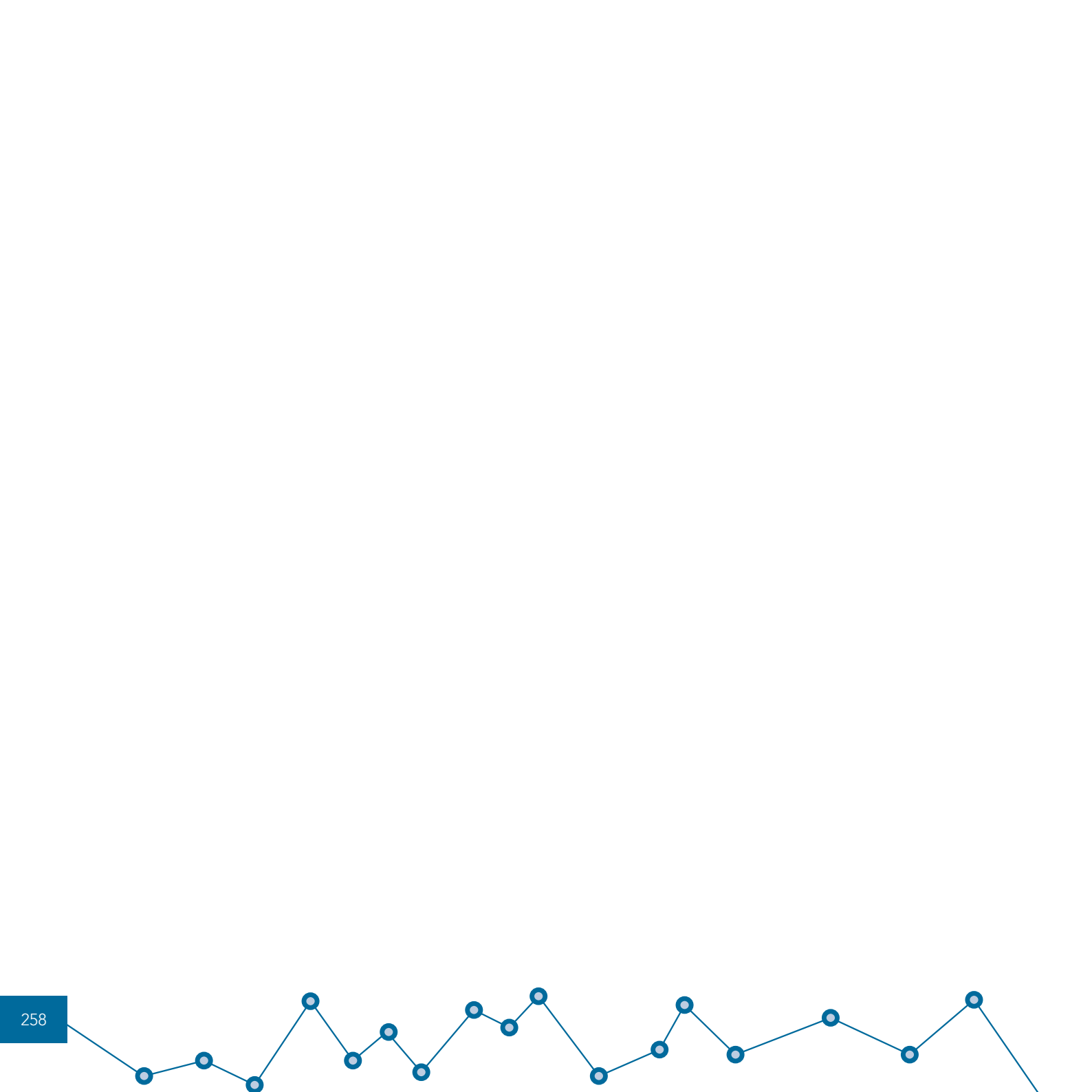
Drs. R.H. Veenstra RA was vanaf 1970 openbaar accountant bij grote en internationale ondernemingen, waaronder tot 1991 bij Ahold, en is vanaf 1992 gevestigd als onafhankelijk accountancy-consultant. Ruud is lid van de Stuurgroep

Statistical Auditing van het Limperg Instituut.



Jan Wille

Dr. J.F. Wille FRM is werkzaam als wiskundige bij PwC Consulting.

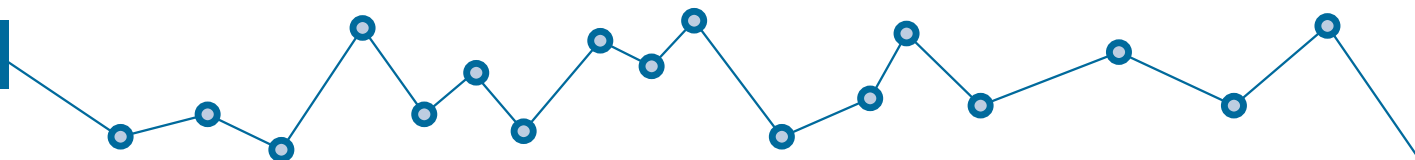


Verklarende woordenlijst

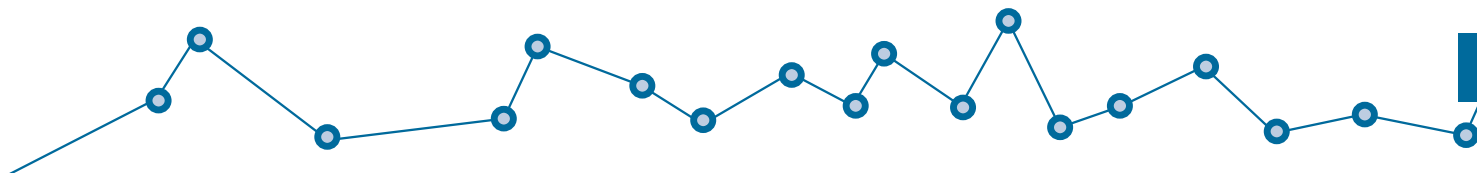
Deze verklarende woordenlijst bevat een aantal in deze essaybundel gebruikte begrippen die toegepast worden binnen het vakgebied Statistical Auditing. Algemene begrippen uit de accountantscontrole en de statistiek worden bekend verondersteld.

AOQL-methode	AOQL is een afkorting van 'Average Outgoing Quality Limit'. Deze methode is afkomstig uit de industriële kwaliteitsbeheersing op basis van een steekproef per batch met als doel het verwachte percentage fouten beneden een vooraf gestelde limiet te houden. Een van de kenmerken is dat de als fout aangemerkte eenheden worden verbeterd of vervangen door goede exemplaren. De noodzakelijke hoeveelheid werk kan variëren tussen de geplande steekproefomvang en integrale controle. Dit is er van afhankelijk of er meer dan het toegestane aantal fouten in een of meer van de steekproeven worden gevonden.
Atypische fouten	Fouten waarvoor geen verklaring kan worden gegeven; deze lijken door het toeval te zijn ontstaan. Deze soort fouten worden wel aangemerkt als te isoleren fout. Ook wel: fouten met een oorzaak die niet elders is voorgekomen
Audit Assurance	Toegevoegde zekerheid door de accountantscontrole. Audit assurance is een kansbegrip en het complement van audit risk. Bijvoorbeeld: een audit assurance van 95% komt overeen met een audit risk van 5%.
Audit Risk	Audit Risk is de kans dat een auditor een populatie met een fout ter grootte van tenminste de toelaatbare afwijking ten onrechte goedkeurt.

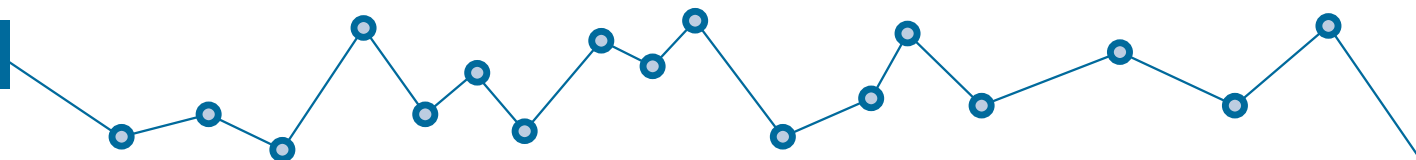
Bayesiaanse statistiek	Een vorm van statistiek waarin geen hypothesen worden getoetst, maar kansverdelingen worden berekend op basis van waarnemingen. Als startpunt wordt een priorkansverdeling gebruikt. Deze kan bijvoorbeeld gebaseerd zijn op een inschatting van de kwaliteit van de interne beheersing. Bayesiaanse statistiek is de tegenhanger van frequentistische statistiek.
Bovengrens van de fout	De bovengrens geeft het foutbedrag in de populatie aan dat met een vooraf gestelde betrouwbaarheid niet overschreden wordt. Een veel gekozen betrouwbaarheidspercentage is 95%. Een 95% betrouwbare bovengrens van € 100.000 betekent dat de kans dat het werkelijke foutbedrag groter is dan € 100.000, kleiner is dan 5%.
Cell sampling	Bij Cell Sampling wordt een populatie in gelijke stukken verdeeld (intervallen of Cells). Het aantal gelijke stukken is de steekproefomvang. Per cell wordt er aselekt een element van die cell aangewezen. In het geval van een geldsteekproef wordt de post behorende bij deze geldeenheid gecontroleerd.
Conservatisme	Algemeen principe van voorzichtigheid. In deze bundel betekent conservatisme (meestal) het hoog (voorzichtig) inschatten van betrouwbaarheidsbovengrenzen.
Data-analyse	Een verzameling technieken voor het extraheren en compileren van ruwe gegevens met als doel concrete controle-informatie te genereren, dan wel daaraan een bijdrage te leveren en met ook als doel het ondersteunen van de te nemen controlebeslissingen.
Directe schatter	Het berekenen van een geschatte waarde (bv het gemiddelde of het totaal) uit een verzameling enkelvoudige waarnemingen, d.w.z. waarnemingen van een enkele onafhankelijke variabele. De boekwaarden van de waarnemingen worden niet gebruikt.
Discriminantanalyse	Techniek om op basis van een beperkt aantal kenmerken van vastleggingen in een populatie tot de conclusie te komen of die populatie wel of niet aan de vereiste norm voldoet.
Estimation sampling	Steekproeven om te schatten. Zowel centrale tendenties als betrouwbaarheidsintervallen. Engelse term voor de schattingssteekproef.



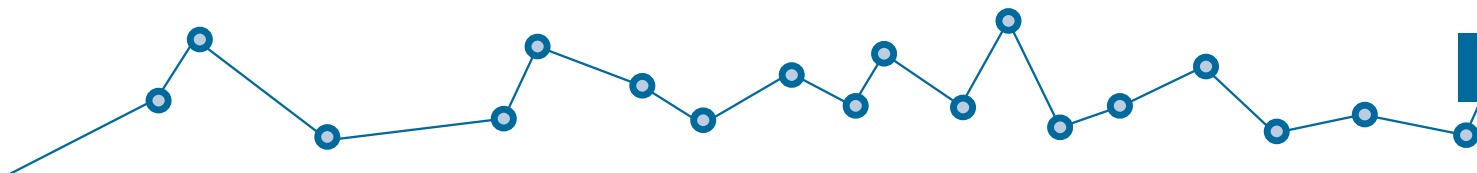
Frequentistische statistiek	In de frequentistische statistiek wordt vooraf een nulhypothese geformuleerd over de populatie. Op basis van data, bijvoorbeeld uit een steekproef, en een vooraf gekozen significantieniveau wordt deze nulhypothese wel of niet verworpen. Frequentistische statistiek is de tegenhanger van Bayesiaanse statistiek.
Geldsteekproef	Een geldsteekproef beschouwt de geldeenheden van een populatie met bedragen, als de eenheden waaruit getrokken gaat worden. Hoe groter de afzonderlijke post, hoe groter de trekkingskans ervan.
Te isoleren fout	Een fout die als systematisch wordt gezien. Onder bepaalde - strikte - voorwaarden kan men zo'n type fout isoleren, waarna men de omvang van dat type fouten integraal kan bepalen. Zo ontstaat een subpopulatie, die als integraal gecontroleerd kan worden beschouwd.
Ist-positie	De waarde van een eenheid zoals die in de boekhouding aangetroffen wordt. Synoniem van het begrip 'boekwaarde'.
Kritische deelwaarneming	Manier van selecteren waarbij de accountant posten om te controleren kiest op basis van criteria in plaats van 'at random'. De criteria zijn bedoeld om fouten te vinden en kunnen bijvoorbeeld hoge of afwijkende bedragen zijn of posten afkomstig van een specifieke afdeling of kritisch land. Bij deze manier van selecteren zijn de resultaten van de controle niet op statistisch verantwoorde wijze naar het niveau van de hele populatie te projecteren. Dit een specifieke soort 'niet-statistische steekproef'. Zo'n selectie wordt ook wel 'judgemental sampling' genoemd.
Materialiteit	Die omvang van een onjuistheid of omissie die de gebruiker van de betreffende financiële verantwoording een andere beslissing doet nemen (of kan doen nemen). De materialiteit wordt uitgedrukt als een absoluut bedrag of als een percentage van een grootte van de financiële verantwoording.
Materiële controle	Controleactiviteit om te beoordelen of het onderzochte fenomeen of event inhoudelijk goed genoeg is of zal worden vastgelegd. Meestal: zijn de bedragen juist gegeven de overige kenmerken van de vastlegging van de transactie.



Negatieve controle	Controle om te onderzoeken of er sprake is van ontbrekende boekingen of elementen van boekingen zoals bedragen. Is alles geboekt dat geboekt had moeten worden?
Overstatement	Een overstatement is een te hoog geboekt bedrag.
Poissonverdeling	Discrete statistische verdeling waarbij het aantal fouten in de te beoordelen populatie heel klein is. De zogenoemde foutintensiteit nadert naar nul. Veelgenoemde voorbeelden hiervan zijn fouten in een administratieve populatie en weeffouten in een lap stof.
Positieve controle	Controle om te onderzoeken of er sprake is van onjuistheden. Is alles wat geboekt is, geboekt zoals het geboekt had moeten worden?
Postensteekproef	Een postensteekproef beschouwt de afzonderlijke posten van een populatie als de elementen waaruit getrokken gaat worden (dit in tegenstelling tot een geldsteekproef).
Primaire vastlegging	De eerste registratie van een transactie of event. Soms moeten de gegevenselementen nog aangevuld worden, soms is er sprake van overtollige gegevenselementen. De duiding van het event vindt plaats bij de vertaling naar de subsidiaire vastlegging(en) ervan.
Process mining	Een serie technieken (van data-analyse) om uit de vastleggingen van een proces de actoren en de deelprocessen te herleiden. Daaruit kan bijvoorbeeld worden afgeleid of en in hoeverre de bedoelde functiescheiding volgens de vastleggingen heeft gewerkt.
Regressieschatter	Wanneer er sprake is van een al of niet statistisch verband tussen twee (een onafhankelijke en een afhankelijke) variabelen kan met de familie van de regressieschatters uit de waarnemingen de waarde (b.v. het gemiddelde of het totaal) nauwkeuriger geschat worden dan met de directe schatter.
Risicoanalyse	Het begrip risicoanalyse is niet eenduidig. De betekenissen zijn onder meer het beoordelen van het inherent risico, het beoordelen van bedrijfsrisico's (bedreiging van de ondernemingsdoelen, dan wel de kansen op het beter halen van die doelen), het beoordelen van IT- risico's en tenslotte het beoordelen van het risico op het ten onrechte goedkeuren van een financiële verantwoording die een geaggregeerde hoeveelheid fouten ter grootte van de materialiteit bevat.



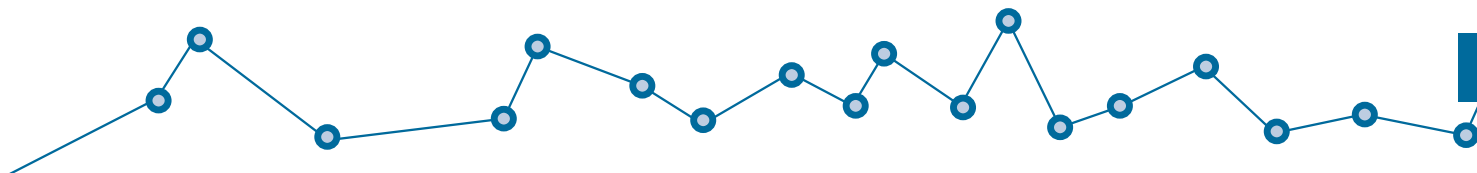
Schattingssteekproef	Een steekproef om de parameters van een populatie te schatten. Deze betreffen zowel de centrale tendentie als de betrouwbaarheidsgrenzen. Vertaling van Estimation Sampling.
Significantieniveau	In de frequentistische statistiek betekent dit begrip het vooraf gestelde maximale risico dat de nulhypothese ten onrechte verworpen wordt. Na het vergaren van de data wordt het risico op onterecht verwerpen van de nulhypothese, ook wel aangeduid met de p-waarde, berekend. Als de p-waarde onder het significantieniveau blijft, wordt de nulhypothese verworpen.
Soll-positie	De waarde die geboekt had moeten worden. Waarbij onder waarde de inhoud van alle 'velden van het record' moet worden verstaan. Synoniem van auditwaarde.
Statistische steekproeven	Steekproeven die aselect en dus representatief uit een populatie worden getrokken. De projectie en de berekende grenzen hebben betekenis voor de besluitvorming over de populatie als geheel.
Steekproefequivalent	Met behulp van Bayesiaanse statistiek wordt de beschikbare voorinformatie omgezet in een steekproefomvang en een betrouwbaarheidsinterval. B.v. de beoordeling van de kwaliteit van de Interne Beheersing (IB) wordt omgezet in een steekproefomvang. Deze 'steekproef' hoeft niet meer te worden gecontroleerd omdat die, door de beoordeling van die IB al als '100% goed' wordt aangemerkt.
Stringer Bound	De Stringer Bound is de maximale fout berekend voor een geldsteekproef waar de taintings in volgorde van afnemende omvang zijn gesorteerd en per besmette geldeenheid de bijdrage aan de maximale fout is berekend.
Subsidiaire vastlegging	Een subsidiaire vastlegging is een interpretatie van de primaire vastlegging. De gegevens van de primaire vastlegging worden aangevuld met gegevens voor rubricering en dergelijke. De als overtoollig aangemerkte gegevenssoorten van de primaire vastlegging worden niet meegenomen.
Substantive test	Een substantive test is de inhoudelijke beoordeling van een boeking (meestal subsidiaire vastlegging). Engelstalige term voor een gegevensgerichte detailcontrole. Gelijk aan materiële controle.

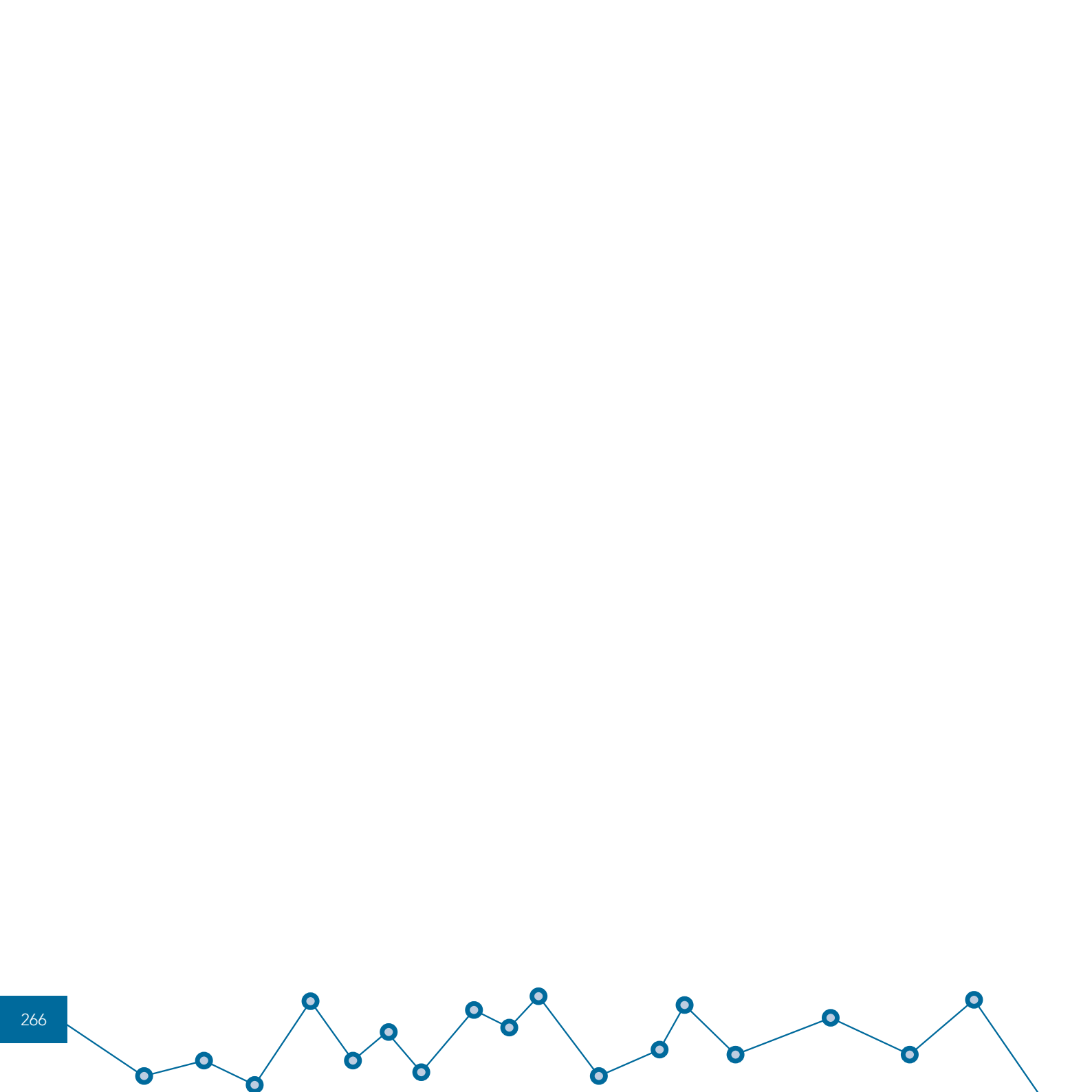


Systematische trekking	Een vorm van aselechte steekproeftrekking. In het eerste interval wordt aselekt een element aangewezen. Verdere selectie vindt dan plaats door steeds een volgend element aan te wijzen dat precies een interval verder ligt. Een geldsteekproef kan op deze wijze worden getrokken. Wordt ook wel trekking met een vast interval genoemd.
Tainting	Een tainting is een relatieve fout, dus een getal tussen nul en één. De omvang van de relatieve fout ontstaat door het verschil tussen boekwaarde en de auditwaarde te bepalen en vervolgens te delen door de boekwaarde.
Toelaatbare afwijking	De 'materialiteit' in geld voor de verantwoording als geheel wordt vertaald in een 'uitvoeringsmaterialiteit' voor de subpopulatie waarop de steekproef wordt ontworpen. Om ook bij het vinden van fouten te kunnen goedkeuren kan een 'toelaatbare afwijking' daar weer onder worden gekozen. Dat is dan de grootte die leidt tot het steekproefinterval (en uiteindelijk de steekproefomvang).
Transactiemodel	Model dat de vertaling van de transacties van een organisatie in de echte wereld via de primaire vastleggingen naar de subsidiaire vastleggingen weergeeft. Met subsidiaire vastleggingen wordt zowel de administratie als de jaarrekening of andere financiële verantwoording bedoeld.
Trekking met een vast interval	Synoniem van systematische trekking.
Triviale afwijkingen	De NV COS gebruikt deze uitdrukking voor geprojecteerde afwijkingen die geen betekenis voor de verantwoording hebben.
Uitvoeringsmaterialiteit	De uitvoeringsmaterialiteit is een beoordelingsnorm waaraan de uitkomsten van een steekproef op een subpopulatie worden getoetst. Die steekproef is ontworpen met een toelaatbare afwijking. De toelaatbare afwijking mag kleiner zijn dan de uitvoeringsmaterialiteit. Het extra werk dat zo wordt gedaan voorkomt dat kleine afwijkingen tot 'niet goedkeuren' leiden. De uitvoeringsmaterialiteit dient zo te worden gekozen dat als elk onderdeel aan zijn uitvoeringsmaterialiteit voldoet, zeker is dat de verantwoording als geheel aan de materialiteit voldoet.

Understatements	Understatements zijn te laag weergegeven bedragen van boekingen. Understatements kunnen ook niet geboekte bedragen (missende records) betreffen.
Voorinformatie	Voorinformatie is een term uit de Bayesiaanse statistiek die de waarde van bijvoorbeeld de kwaliteit van de IB weergeeft in de vorm van een kans of als een waarde. Met Bayesiaanse statistiek wordt deze voorinformatie omgezet in een priorverdeling.
Waarschijnlijkheid	De NV COS gebruikt deze uitdrukking in het geval van het opzetten en evalueren van cijferanalyses. Deze term is identiek aan 'statistische betrouwbaarheid'.

De redactie.





Statistical Auditing for dummies

Kees Voorhoeve

Inleiding

Limperg sprak in zijn tijd over de verstandige leek. Tegenwoordig is er een uitgebreide boekenreeks voor dummies, leken, over uiteenlopende, vaak ingewikkelde, onderwerpen gericht op mensen die er weinig of niets van afweten maar er wel benieuwd naar zijn. Voor deze doelgroep beogen wij hier de belangrijkste begrippen uit vakgebied “Statistical Auditing” in hun samenhang te presenteren.

Doel van de accountant is op effectieve en efficiënte manier een verantwoorde uitspraak te doen over de ‘getrouwheid’ van een financiële verantwoording die hem ter controle is aangeboden.

Effectief wil zeggen dat dit op een methodologisch verantwoorde manier gebeurt. Efficiënt wil zeggen dat de accountant hiervoor niet meer werk doet dan noodzakelijk, opdat de kosten voor de gecontroleerde beperkt blijven. De accountant kan daarom niet de hele administratie controleren. Als gevolg hiervan zal de verklaring altijd enige onzekerheid omtrent de gestelde ‘getrouwheid’ inhouden. Het controleproces uitgevoerd door de accountant bestaat naast de oriëntatie op de business van de gecontroleerde uit vier delen.

Die zijn de beoordeling van de interne beheersing, risicoanalyse ten behoeve van het verrichten van (minder) detailcontroles, gegevensgerichte detailcontroles en gegevensgerichte cijferanalyses. De financiële verantwoording kan bijvoorbeeld een jaarrekening of een belastingaangifte betreffen.

De accountantsverklaring houdt altijd enige onzekerheid in omtrent de getrouwheid van de financiële verantwoording.

De accountant stemt daartoe met de gecontroleerde de materialiteit af, dat is de fout die de financiële verantwoording kan bevatten zonder dat de accountant die, gegeven genoemde onzekerheid, heeft kunnen ontdekken. Deze onzekerheid, de onbetrouwbaarheid in de uitspraak van de accountant, wordt het audit risk of audit risico genoemd. De materialiteit wordt als absoluut bedrag uitgedrukt of als een percentage. Naarmate de materialiteit kleiner is wordt dus meer nauwkeurigheid gevraagd en moet de accountant meer werk doen.

Het audit risico is het risico dat de accountant een onjuiste verklaring afgeeft

Op grond van de bevindingen van de controle komt de accountant, in het geval van goedkeu-

ring, tot de uitspraak dat met een betrouwbaarheid groter of gelijk aan 95% de maximale fout in de financiële verantwoording de materialiteit niet overtreft. Wij gebruiken hier het veelgebruikte betrouwbaarheidsniveau van 95%, maar andere percentages komen in de praktijk ook voor. Dit begrip heet Audit Assurance, de toegevoegde zekerheid door de accountantscontrole. Het complement van die betrouwbaarheid is de Audit Risk en is standaard - bij 95% betrouwbaarheid - 5%. Dit is dus de kans dat de accountant een verantwoording goedkeurt met daarin fouten die opgeteld gelijk zijn aan de materialiteit.

De methoden en technieken van het vakgebied "Statistical Auditing" helpen om op efficiënte wijze tot een statistisch verantwoorde uitspraak te komen over de getrouwheid van de financiële verantwoording.

Met de beoordeling van de interne beheersing gaat de accountant na of de kwaliteit van de administratie van de gecontroleerde zodanig is dat deze achteraf controleerbaar is. Doel van de risicoanalyse is het audit risico te beperken en een uitspraak te doen over de hoeveelheid werk die de accountant moet doen aan gegevensgerichte detailcontroles. Met andere woorden, hoe groot moeten de steekproeven zijn die de accountant moet controleren om tot een verantwoorde uitspraak te komen.

Wij behandelen posten- en geldsteekproeven.

Als meest arbeidsintensieve deel van de controle is hier naast effectiviteit, efficiëntie van belang.

Als antwoord op de vraag hoe in situaties met een geringe foutenkans met een beperkte steekproef

toch tot een statistisch verantwoorde uitspraak te komen, behandelen wij de technieken van geldsteekproeven en het opdelen van de populatie boekingen in subpopulaties. Geëigende methoden om de hoeveelheid steekproeven te beperken zijn de toepassing van Bayesiaanse statistiek en gegevensgerichte cijferanalyses. Een vorm van interne beheersing waarin het gebruik van statistiek centraal staat is de AOQL-methode. Aan het einde van dit artikel komen wij kort op het idee hierachter terug.

Met de beoordeling van de interne beheersing kwalificeert de accountant primair de controleerbaarheid van de administratie.

Een organisatie kent een bedrijfsgebeuren met beheersmaatregelen die ervoor moeten zorgen dat de organisatie haar doelen behaalt en daarbij aan wet- en regelgeving voldoet. Een van de regels is dat de jaarrekening een getrouw beeld geeft van wat er in het bedrijfsgebeuren in werkelijkheid is gebeurd. Daartoe moet de administratie waarheidsgetrouw en deugdelijk zijn gevoerd.

De interne beheersing moet met beleid en procedures hiervoor zorgdragen. Het COSO-raamwerk definieert de elementen van interne beheersing en schrijft voor dat de accountant opzet, bestaan en werking van de procedures toetst naast de betrouwbaarheid van de financiële informatievoorziening en de naleving van relevante wet- en regelgeving. De aanname is dat naarmate de interne beheersing beter is, er minder fouten in de administratie en dus in de jaarrekening zullen voorkomen.

Het transactiemodel is hulpmiddel bij het beoordelen van het waarheidsgehalte van de administratie.

De kans bestaat dat de werkelijke transacties niet of onjuist in de administratie zijn vastgelegd. Het transactiemodel beschrijft het proces van fysieke transacties naar vastleggingen in de administratie. De accountant moet dus rekening houden met de IST-positie, datgene dat geboekt is, en de SOLL-positie, datgene wat geboekt had moeten worden. Met positieve controles controleert de accountant of transacties juist geboekt zijn, met negatieve controles of er transacties onterecht niet geboekt zijn.

De risicoanalyse dient om het audit risico te beperken.

De richtlijnen voor de accountantscontrole schrijven een op risicoanalyse gebaseerde benadering voor. Hulpmiddel daarbij is het zogenoemde Audit Risk Model (ARM), dat stelt dat het audit risico een functie is van het inherente risico, het interne-beheersingsrisico en het detectierisico. In deze bundel maken wij binnen het detectierisico nog onderscheid in het steekproefrisico en het cijferanalyserisico. In veel publicaties wordt het audit-risk als het product van de verschillende risico's beschreven, maar dit staat mathematisch ter discussie, onder meer omdat het interne-beheersingsrisico en het inherente risico niet onafhankelijk van elkaar zijn. Aan de andere kant is zeker het interne beheersingsrisico niet absoluut vast te stellen, wat de precieze vorm van de functie minder relevant maakt. Het ARM is dus eigenlijk meer een denkmodel dan een wiskundige formule en laat in ieder geval zien dat er sprake is van het stapelen van onzekerheden.

Buiten het bestek van deze bundel staat het ARM ook maatschappelijk ter discussie omdat het primair gebruikt wordt om de hoeveelheid detailcontroles te beperken in een tijd waarin de maatschappij juist om meer controles vraagt.

Definitie van de verschillende begrippen binnen het Audit Risk Model.

Het inherente risico is de kans dat de financiële verantwoording een materiële fout bevat als gevolg van factoren anders dan door het gebrek aan interne beheersing. Dit risico wordt hoger beoordeeld wanneer het transacties betreft waar een grote mate van 'professional judgement' in het geding is, bijvoorbeeld bij vastgoedwaarderingen, of in complexe organisaties, bijvoorbeeld met veel werkmaatschappijen. Het interne beheersingsrisico is het risico dat de interne beheersing een materiële fout niet opmerkt. Het cijferanalyserisico is het risico dat cijferanalyses, zoals uitzonderingsrapportages en dwarsverbanden, de fout ook niet opmerken. Het detectierisico is het risico dat de accountant een materiële fout in de financiële verantwoording met behulp van gegevensgerichte detailcontroles niet ontdekt. In deze bundel is het detectierisico synoniem met het steekproefrisico, het risico dat een materiële fout door uitgevoerde steekproeven ook niet opgemerkt wordt.

Toepassing van het Audit Risk Model levert hoeveel werk de accountant gedurende de gegevensgerichte detailcontrole moet verrichten.

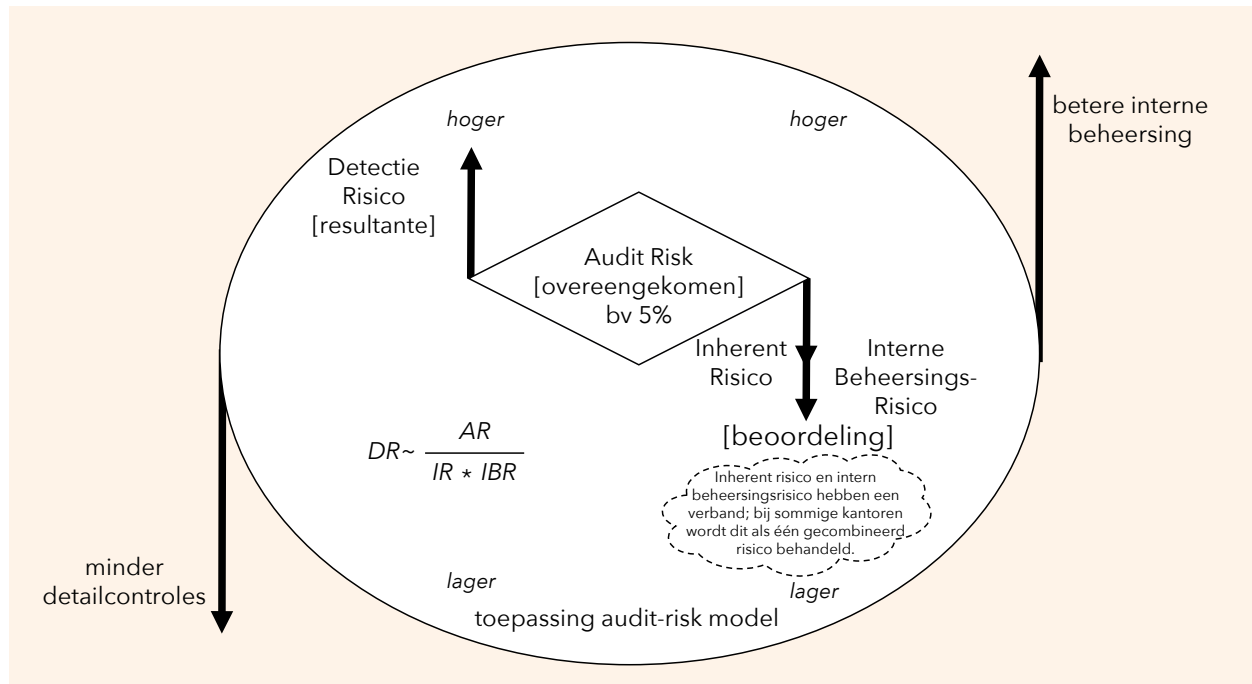
De accountant komt een aanvaardbaar audit risico met de opdrachtgever overeen. Vervolgens beoor-

deelt de accountant het inherente risico op fouten en de kwaliteit van de interne beheersing uitgedrukt in het interne-beheersingsrisico. Toepassing van het Audit Risk Model leidt vervolgens tot een detectierisico. Het detectierisico is voor de accountant leidraad meer werkzaamheden te doen zoals uitgebreidere gegevensgerichte cijferanalyses of uitgebreidere steekproeven. Bij het toestaan van een hoog detectierisico is dit niet nodig. Een betere interne beheersing betekent dat de accountant minder detailcontroles behoeft te doen. Dit is grafisch in onderstaande figuur uitgebeeld.

Met behulp van Bayesiaanse statistiek is dit verband tussen de kwaliteit van de interne beheersing en de hoeveelheid detailcontroles te objectiveren. Hierop komen wij later terug.

Binnen de gegevensgerichte detailcontroles staat het trekken van steekproeven centraal.

Het vakgebied 'statistical auditing' gaat ervan uit dat de accountant aselechte steekproeven trekt, ook wel statistische steekproeven genoemd. In dat geval zijn de resultaten van de steekproefcontrole te projecteren naar het niveau van de hele populatie. De accountant kan ook de meest foutgevoe-



lige posten uitkiezen, 'judgemental sampling' of kritische deelwaarneming genoemd. Dat kan de accountant doen voor individuele posten, maar ook voor hele subpopulaties. Dat doet de accountant wanneer deze specifiek op zoek is naar fouten of fraude. De resultaten van zo'n steekproef zijn dan niet naar het niveau van de hele populatie te projecteren en de accountant kan niet tot een uitspraak voor de hele populatie komen.

De uitkomst van een aselekt getrokken steekproef leidt tot een uitspraak voor de hele onderzochte populatie met een zekere betrouwbaarheid binnen grenzen, de maximale fout.

Vanwege efficiencyredenen neemt de accountant steekproeven. De omvang van de steekproef hangt af van de beoogde mate van zekerheid. Die zekerheid wordt weergegeven in betrouwbaarheidsgrenzen. De accountant verklaart uiteindelijk met een bepaalde betrouwbaarheid, in het algemeen 95%, dat de financiële verantwoording een getrouw beeld geeft. Meestal is voor de verklaring alleen een betrouwbaarheidsbovendgrens van belang. Deze betrouwbaarheidsbovendgrens wordt ook wel de maximale fout genoemd. Voor een goedkeurende verklaring moet de maximale fout kleiner zijn dan de materialiteit.

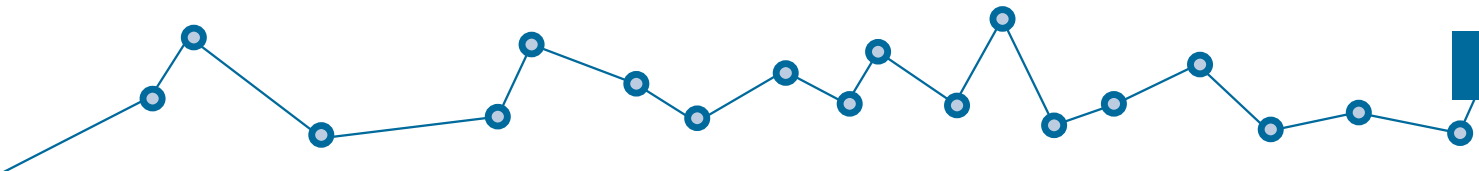
De accountant kan aselekt posten- en geldsteekproeven trekken.

Afhankelijk van de kenmerken van de posten in de administratie, statistisch de populatie genoemd, trekt de accountant aselekt een postensteekproef of een geldsteekproef. Bij een pos-

tensteekproef zijn de posten uit de administratie de eenheden waaruit getrokken wordt. Bij een geldsteekproef trekt de accountant uit de geldeenheden in de administratie. Daartoe wordt het totale bedrag van de administratie in zoveel gelijke cellen verdeeld als de omvang van de steekproef. Hierop selecteert de accountant willekeurig de steekproefelementen door systematische trekking of aselekt trekking binnen elke cel. Bij een geldsteekproef controleert de accountant de posten die bij willekeurig getrokken geldeenheden horen. Hoe groter de afzonderlijke post, hoe groter de trekkingskans ervan. Als in een populatie weinig fouten verwacht worden is een geldsteekproef de eerste keus. Bij een niet-financiële populatie is alleen de postensteekproef beschikbaar. Als de accountant in een populatie veel fouten verwacht, zoals bijvoorbeeld bij voorraadinventarisaties het geval kan zijn, ligt ook een postensteekproef voor de hand.

De uitdaging is om bij een geringe foutdichtheid met een beperkte steekproef toch een goede schatting van de totale fout te maken.

Een administratie bevat doorgaans weinig fouten, met een statistische term een geringe foutdichtheid genoemd. Bij een geringe foutdichtheid is het bijna niet mogelijk om met een steekproef van beperkte omvang een goede schatting te maken van de fout in de populatie boekingen. Extra complicatie is dat een administratie in het algemeen grote en kleine bedragen zal bevatten, wat statistisch wil zeggen dat de bedragen in de administratie niet homogeen, maar scheef verdeeld zijn. Wan-



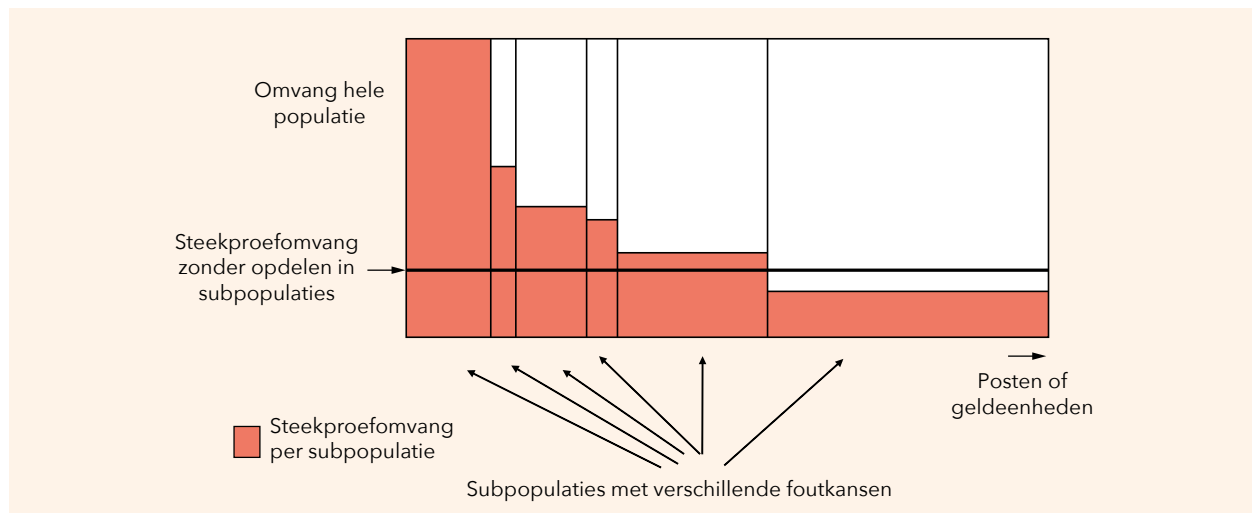
neer een fout in een klein bedrag wordt gemist is de impact bij projectie naar de hele populatie gering, maar bij het missen van een fout in een groot bedrag is de impact veel groter. Dan komt de effectiviteitsvraag, het komen tot een statistisch verantwoorde uitspraak, in het geding.

De accountant kan op twee manieren, en in de combinatie, met deze statistische onzekerheid omgaan:

1. Door het trekken van geldsteekproeven in plaats van postensteekproeven. De kans is dan groter dat de grote posten met hun eventuele fouten in de steekproef vallen.
2. Door de populatie van boekingen te verdelen in subpopulaties met verschillende foutkansen en per subpopulatie een posten- of geldsteekproef trekken. Verschillende fout-

kansen per subpopulatie betekent ook dat de uitvoeringsmaterialiteit en als gevolg hiervan de steekproefomvang per subpopulatie kan verschillen. Dit is grafisch weergegeven in onderstaande figuur.

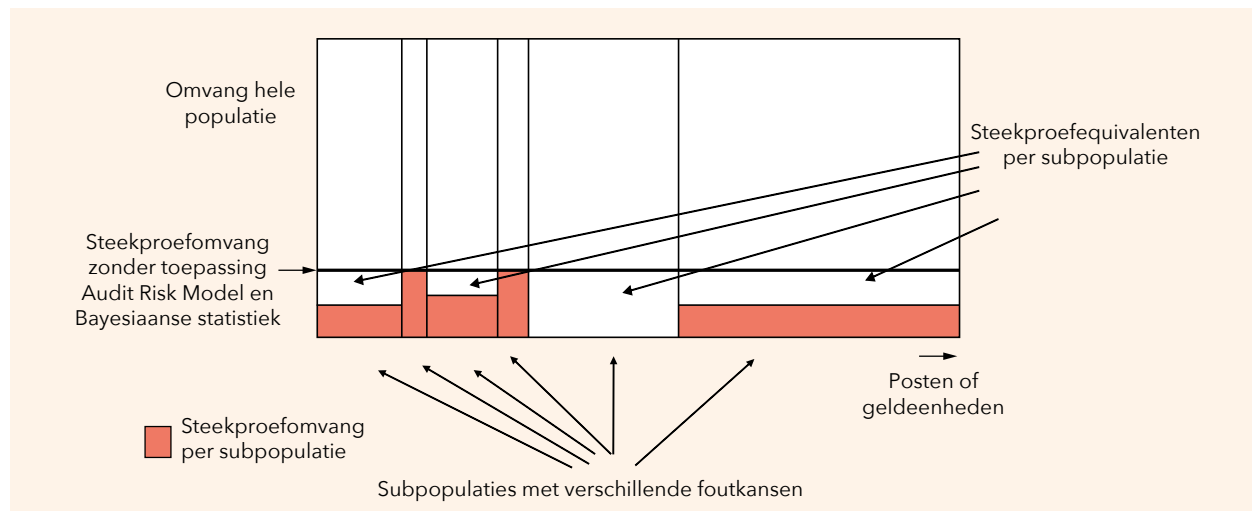
Het bovenstaande is statistisch te onderbouwen met frequentistische statistiek. Met behulp van de Poissonverdeling is, gegeven het auditrisico en de uitvoeringsmaterialiteit, een steekproefomvang te bepalen.



Toepassen van Bayesiaanse statistiek helpt om het aantal detailcontroles te beperken.

In de praktijk zal de accountant, op grond van de opgedane kennis van de organisatie, 'understanding the business', de beoordeling van de interne beheersing en de initiële cijferanalyse weten welke delen van de administratie een grotere foutkans hebben dan andere. De Bayesiaanse statistiek, tegenhanger van de frequentistische statistiek, biedt de mogelijkheid om hiermee om te gaan bij de bepaling van de steekproef. In deze vorm van statistiek wordt voorinformatie beschreven als een a priori kansverdeling waarmee gerekend kan worden. De beoordeling van de interne beheersing door de accountant wordt dan als voorinformatie beschouwd. Vervolgens kan de accountant met Bayesiaanse statistiek uitrekenen welke vermindering van de gegevensgerichte con-

trole, de steekproef, aanvaardbaar is door van die voorkennis gebruik te maken. Deze vermindering van de steekproefomvang, de steekproefequivalent genoemd, hoeft niet meer apart gecontroleerd te worden omdat die al als '100% goed' is aangemerkt. In een extreem geval hoeft de hele subpopulatie niet meer gecontroleerd te worden. Dit is grafisch weergegeven in onderstaande figuur. Bayesiaanse statistiek in de audit helpt dus om de steekproefomvang, en dus de hoeveelheid werk, te beperken en toch op verantwoorde wijze tot een verklaring te komen.



Met behulp van Bayesiaanse statistiek is het verband tussen de kwaliteit van de interne beheersing en de hoeveelheid detailcontroles te objectiveren

Algemeen probleem bij het toepassen van Bayesiaanse statistiek is hoe de voorinformatie als een kansverdeling te beschrijven. Dit geldt ook voor de voorinformatie naar aanleiding van de beoordeling van de interne beheersing. Een praktische oplossing, beschreven in deze bundel, is door de interne beheersing te beoordelen op een vierpunt-schaal en aan elke score een maximale foutkans te koppelen. Voor deze foutkansen, behorende bij de vier scores variërend van “laag” tot “hoog”, steekt de accountant zijn hand in het vuur dat deze niet overschreden worden. Door naderhand het aantal fouten te evalueren is op deze manier de kwantitatieve score wel objectief toetsbaar. Een aantal grote accountantskantoren werkt op deze manier.

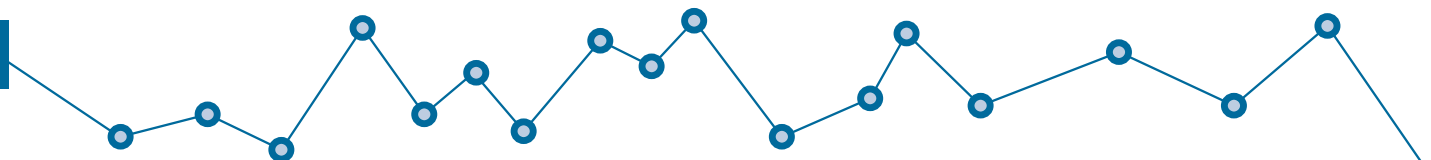
Andere kantoren werken wel met een ‘3-2-1’ schaal waarmee zij het inherent risico en de kwaliteit van de interne beheersing beoordelen. De hoogste waarde $R=3$ betekent dan dat de accountant geen steekproefreductie toepast. Een lagere R -waarde betekent dat de accountant als gevolg van een verminderd inherent risico een steekproefreductie toepast ($R = 2$). De laagste R -waarde, $R=1$, betekent dat de accountant steunt op de interne beheersing. De bijbehorende steekproefreducties zijn dan door het bureau vaktechniek voorgeschreven.

Ook gegevensgerichte cijferanalyse kan helpen het aantal detailcontroles te beperken.

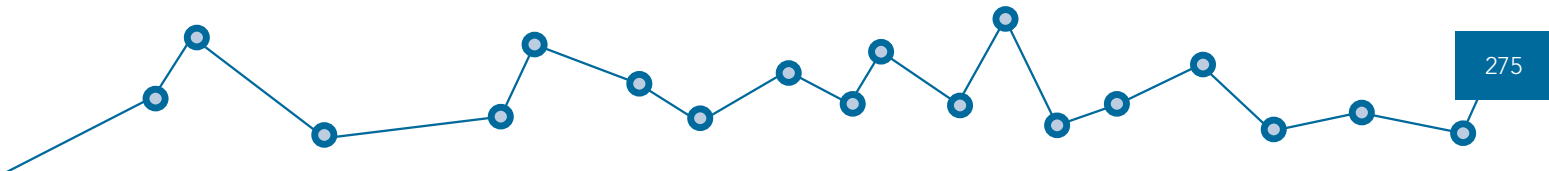
Voorwaarde voor het toepassen van gegevensgerichte cijferanalyse is dat er verbanden bestaan tussen verschillende groepen gegevens in de administratie. In dat geval kan die cijferanalyse de gegevensgerichte detailcontroles of overbodig maken of reduceren. Met een statistische techniek, regressieanalyse, leidt men voor een verzameling al gecontroleerde gegevens een verband af met een andere te controleren gegevens. Het verband betreft bijvoorbeeld het verband tussen verkopen, inkopen en geregistreerde directe uren. Als het verband tussen gecontroleerde en te controleren gegevens in de gecontroleerde periode sterk was kan men dit verband ook onderzoeken in de te controleren periode. Wanneer het verband nog steeds sterk is, dan vervalt de noodzaak de te controleren gegevens aan detailcontroles te onderwerpen. Als het ‘oude verband’ niet in dezelfde mate meer blijkt te bestaan in de nieuwe periode kan men alsnog overstappen op gegevensgerichte detailcontrole voor die nieuwe periode.

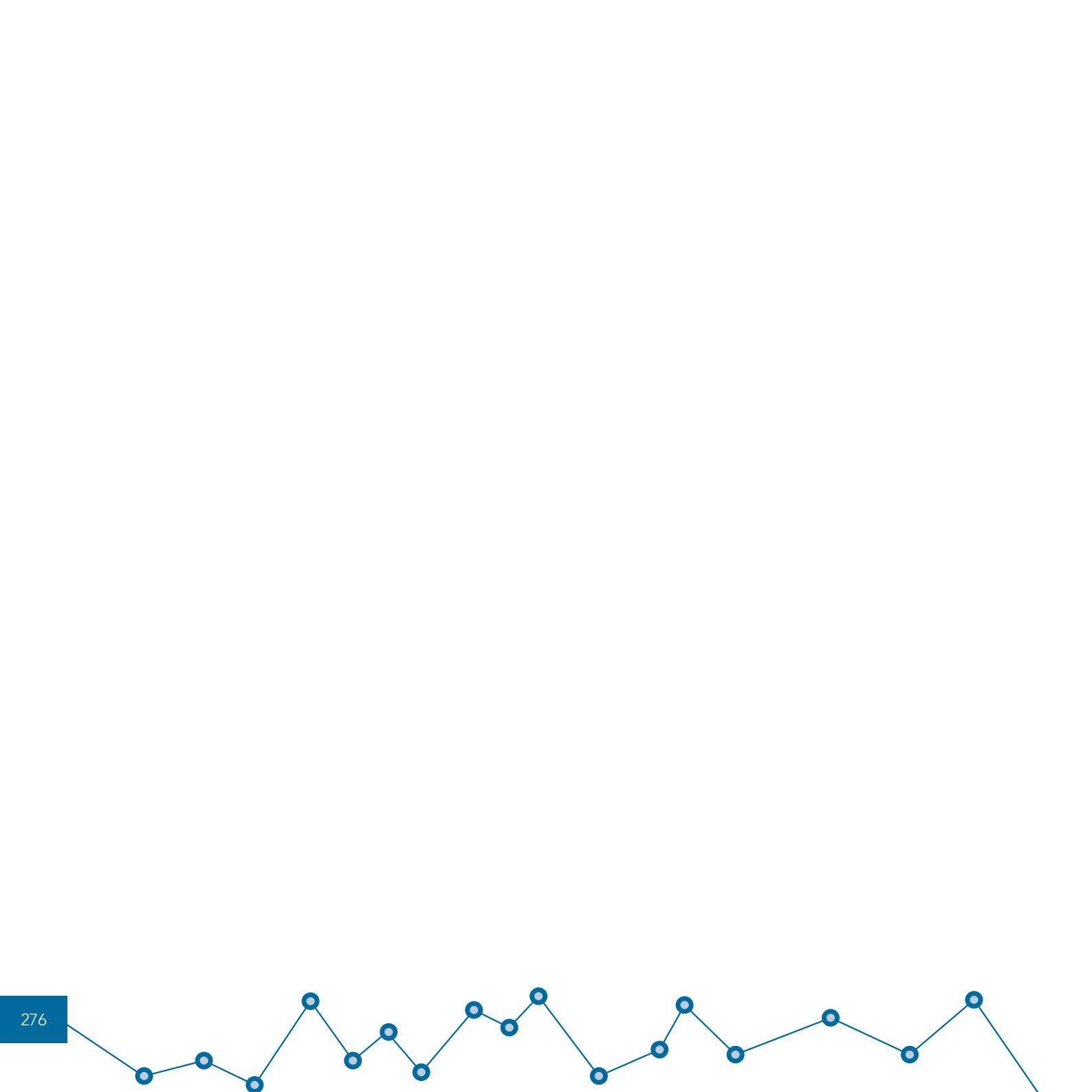
AOQL-methode is een vorm van interne beheersing met behulp van statistiek.

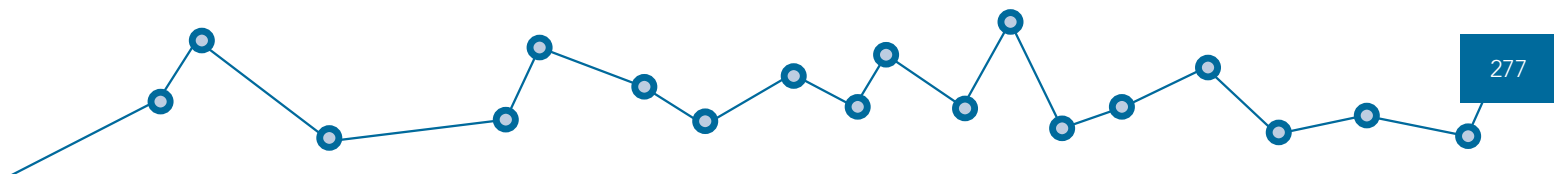
AOQL (Average Outgoing Quality Limit) is een methode afkomstig uit de procesindustrie. De AOQL-methode is strikt genomen geen controlemiddel achteraf, maar een manier om ‘going concern’ de kwaliteit van de administratie te beheersen. Periodiek worden steekproeven genomen waarvan de uitkomst een tevoren afgesproken foutlimiet niet mag overschrijden. Een van



de kenmerken van de methode is dat de als fout aangemerkte eenheden worden vervangen door goede exemplaren. Voorwaarde voor toepassing van AOQL is dat de productie aan boekingen in onderling vergelijkbare batches te verdelen is. Bij overschrijding van de foutlimiet moet de batch integraal worden gecontroleerd. De hoeveelheid werk kan dus variëren tussen de geplande steekproefomvang per batch en integrale controle van alle batches, afhankelijk van het aantal fouten dat in de steekproeven wordt gevonden. Met behulp van statistiek wordt de steekproefomvang zo klein mogelijk gehouden om verantwoord te kunnen beweren dat het gemiddelde percentage fouten voor de hele populatie boekingen binnen de limiet blijft. Naderhand kan de accountant zich baseren op deze manier van interne beheersing.







Essaybundel Statistical Auditing

De Stuurgroep Statistical Auditing is in 1991 opgericht als een informeel gremium van statistici en accountants en anderen die geïnteresseerd zijn in de combinatie van auditing en statistiek, kortweg Statistical Auditing.

Doel van de Stuurgroep en haar leden is het bevorderen van het correcte, effectieve en efficiënte gebruik van op data gerichte methoden en technieken bij de controle van verantwoordingen. De Stuurgroep tracht dit doel onder andere te bereiken door presentaties, waaronder een jaarlijks symposium onder auspiciën van het Limperg Instituut sinds 2008, en publicaties, waaronder columns op Accountant.nl sinds 2011.

Deze bundel bevat een groot aantal van de inmiddels 70 columns die door leden van de Stuurgroep zijn gepubliceerd, aangevuld met artikelen door leden van de Stuurgroep geschreven samen met anderen. Vanzelfsprekend is van deze medeauteurs daarvoor toestemming gekregen.

De redactie van deze bundel bestond uit Hein Kloosterman, Paul van Batenburg en Kees Voorhoeve. Paul en Hein hebben de meeste tekstbijdragen geleverd, Kees heeft zich toegelegd op het duidelijk en helder (her-)formuleren van de tekst en de transformatie van een stapel columns met een nietje erdoor naar een echt boek. Hein heeft van het begin af aan het project gedragen.

