Accountants doen het met steekproeven. **Paul Batenburg** legt uit hoe dat in z'n werk gaat. Aan de hand van een getallenvoorbeeld over een fictieve lease maatschappij die duizend auto's overneemt van een concurrent. Voor tien miljoen euro. De steekproef wordt niet uit de duizend auto's getrokken ...

Steekproeven in de accountantscontrole

In een vorig artikel[1] heb ik uitgelegd wat voor soort werk BWI-ers bij de afdeling DQI van Deloitte & Touche doen, en hoe hun opleiding aansluit bij de eisen die de dagelijkse praktijk aan hen stelt. In het onderstaande wil ik een van die werkzaamheden nader belichten en ingaan op de statistische methoden en technieken die wij gebruiken bij onze advisering aan accountants. Ik ben econometrist en sinds 1986 als statisticus verbonden aan Deloitte & Touche. Door jaren samen te werken met accountants begrijp ik redelijk wat zij doen en daarom adviseer ik hen over het opzetten, trekken en evalueren van steekproeven. De handleiding steekproefcontroles van de internationale firma waarvan D&T deel uitmaakt is mede door mij geschreven, en ik geef als Honorary Professor les in Statistical Auditing aan de Universiteit van Warwick (Coventry, UK). De tekst van dit artikel heb ik tegelijk gemaakt met mijn presentatie voor het jaarcongres van het Wiskundig Genootschap op 22 november 2002.

Accountantscontrole

Bedrijven en instellingen publiceren jaarrekeningen, waarin een financiële verantwoording wordt afgelegd over de geldstromen die de bedrijfshuishouding zijn binnengekomen dan wel hebben verlaten. Bij een bedrijf is dat een opsomming van inkomen en vermogen, bij een Ministerie een opsomming van uitgaven, verplichtingen en ontvangsten. Die bedragen komen tot stand als totalen uit administraties, en het is eigenlijk tegenwoordig vanzelfsprekend dat dergelijke administraties worden gevoerd met behulp van computersoftware en databestanden. Accountants controleren die bedragen door het vergelijken van in de computer opgeslagen boekingen met bewijsmateriaal dat meestal gewoon als papieren dossiers in archieven ligt opgeslagen. Zo ligt bijvoorbeeld aan een betaling aan een leverancier een ontvangstbewijs, een factuur en een order ten grondslag, aan een salarisbetaling een arbeidscontract en misschien een urenbriefje, en aan een binnengekomen betaling een afgiftebewijs van levering en een verzonden factuur.
Het is wat te naïef om te veronderstellen dat dankzij de computer dit controleren 'met een druk op de knop' kan gebeuren: natuurlijk bestaan er bestanden met informatie die kunnen helpen bij het vaststellen van de juistheid van een betaling, bijvoorbeeld een bestand met binnengekomen goederen en een bestand met prijsafspraken, maar:

- ook die bestanden moeten een keer door iemand vanaf papier in de computer zijn ingebracht
- controle dient te gebeuren door van de administratie onafhankelijk opgestelde bewijsstukken.

De verregaande automatisering lost dus het controlevraagstuk niet op, maar verplaatst het soms wel. Blijft daarom de vraag naar het toetsen van in computerbestanden opgenomen bedragen aan buiten die computer opgeslagen bewijsmateriaal. Uit efficiency kan dan worden overwogen om steekproeven te gebruiken. Om dit artikel overzichtelijk te houden beperk ik me tot een van de vele methoden die gehanteerd worden, namelijk de geldsteekproef op de juistheid van bedragen, waarbij wordt gecontroleerd of zij niet te hoog in een verantwoording zijn opgenomen.

Steekproefmodel
Dergelijke geldsteekproeven zijn gebaseerd op de veronderstelling dat de te controleren populatie het verantwoorde bedrag is, een verzameling euro's die het gevolg is van een verzameling geboekte transacties. De accountant trekt euro's, en controleert de transacties waarin die euro's zijn opgenomen. Het lijkt op een selectie van transacties met verschillende kansen (namelijk evenredig met hun geboekte waarde), maar dat is niet zo: een transactie kan namelijk meer dan één keer worden geselecteerd, omdat verschillende euro's in die transactie zijn getrokken. We trekken dus zonder terugleggen euro's, maar met terugleggen transacties.
Een uitspraak over de onbekende foutkans p in die populatie van euro's groot M wordt gedaan met behulp van een toets:

H_0: $p \geq p_m$ versus H_1: $p < p_m$, waarin p_m de gekozen kritieke foutkans is.

De keuze voor deze hypothesen (en niet andersom) zorgt ervoor dat de accountant alleen dan goedkeurt als er voldoende werk is gedaan om hoogstens een gekozen risico α_m te lopen om ten onrechte goed te keuren. Dit risico is:

$\alpha = P[k \leq k_m | M, n, p = p_m] \leq \alpha_m$

waarin k het aantal fouten (dat wil zeggen euro's die niet geboekt hadden mogen worden omdat er voor de bijbehorende transactie onvoldoende bewijs is (te vinden)) en k_m de daarvoor gekozen norm in de steekproef. De kansvariabele k is een hypergeometrisch verdeelde kansvariabele met als parameters de populatieomvang, de steekproefomvang en de onbekende foutkans. Deze kansverdeling is binomiaal en daarna Poisson te benaderen, en de in de praktijk relevante waarden van M, n en p laten dat vrijwel altijd toe.

Als de accountant heeft gekozen voor p_m, α_m en k_m kan de steekproefomvang n worden bepaald. Om dit artikel simpel te houden wordt afgesproken dat $k_m = 0$, dus dat de accountant alleen maar goedkeurt als de steekproef foutloos is. Voor de praktijk is dat geen restrictie want:

- dit soort steekproefcontroles wordt eigenlijk alleen maar toegepast om te bevestigen dat de accountant inderdaad kan goedkeuren, met andere woorden, dat de werkelijke fout p ruim onder de norm p_m ligt
- verder bouwt de accountant nog een buffertje in door p_m (circa 20%) lager te kiezen dan de norm waarmee de evaluatie van de controle uiteindelijk gaat gebeuren. Als er (kleine) fouten komen is p_m wellicht overschreden, maar de 'echte' norm nog niet.

Door uit te gaan van nul toegestane fouten wordt het rekenwerk aardig versimpeld:

- de Poissonverdeling geeft dat $\exp(-np_m) \leq \alpha_m$ als $n(\text{Poiss}) \geq -\ln(\alpha_m) / p_m$
- de binomiale verdeling levert $(1-p_m)^n \leq \alpha_m$ als $n(\text{binom}) \geq \ln(\alpha_m) / \ln(1-p_m)$ en de McLaurin expansie van $\ln(1-p)$ naar p laat zien dat $n(\text{Poiss}) > n(\text{binom})$
- de hypergeometrische verdeling moet worden geprogrammeerd (in een EXCEL-macro, bijvoorbeeld) om $n(\text{hyper})$ op te leveren. We weten dat die kleiner is dan $n(\text{binom})$, dus zo'n programmaatje start met $n(\text{binom})$ en bepaalt in de hypergeometrische verdeling de omvang van α. Vervolgens wordt de steekproef steeds één kleiner gemaakt, totdat α_m is overschreden. Dan is de steekproef één te klein geworden en weten we $n(\text{hyper})$.

Stel bijvoorbeeld dat een leasemaatschappij in één koop duizend auto's wil overnemen van een concurrent voor een bedrag van tien miljoen euro, maar wel steekproefsgewijs de waardering daarvan controleert. Door de populatie niet te beschrijven als de nummers 1 tot en met 1.000 maar als 1 tot en met 10.000.000 kan een steekproef worden getrokken waarin de waarde van een auto evenredig is met de kans op controle. Zo krijgt een te hoog gewaardeerde auto een grotere kans dan een correct gewaardeerde auto van dezelfde waarde. Dat maakt deze methode ideaal voor zogenaamde positieve controles op de juistheid van verantwoorde bedragen. Als de leasemaatschappij nu stelt dat zij 95% zeker willen merken dat zij 5% of meer te veel betalen, geldt:

$\exp(-n \times 0{,}05) \leq 0{,}05$ als $n \geq 60$.

In een lijst met duizend waardebedragen wordt een cumulatiekolom gemaakt, die wordt vergeleken met zestig random getallen tussen één en tien miljoen. Als de cumulatiekolom het randomgetal overschrijdt, wordt de bijbehorende auto als steekproefelement gecontroleerd. Dit leid ik af door met de Poissonverdeling te bepalen dat bij zestig waarnemingen de kans op nul fouten uit een oneindige massa met 5% fouten onder 5% ligt, maar 59 niet.

Had de maatschappij het bestaan van de duizend auto's willen controleren zonder daar een bedrag aan te willen koppelen, dan was een steekproef uit de nummers 1 tot en met 1.000 voldoende geweest. Nu is de massa klein genoeg voor de hypergeometrische verdeling, en die geeft als steekproef 57, om 95% zeker te merken dat 5% of meer auto's niet zouden bestaan. Dit leid ik af door hypergeometrisch te bepalen dat bij zestig waarnemingen de kans op nul fouten uit een massa van 1.000 met 5% fouten onder 5% ligt, bij 59 ook, bij 58 idem, 57 nog steeds, maar 56 niet meer.

Als je zoals ik een dergelijk voorbeeld aan wiskundestudenten of wiskundigen vertelt merk je vaak een tevreden gehoor. 'Met zo weinig werk een zo belangrijke beslissing ondersteunen, wat een mooi vak hebben we toch gekozen'. Accountancystudenten denken daar heel anders over. 'Zestig waarnemingen? Heeft die vent daar budget voor?' Dat komt doordat accountants dingen kunnen doen met hun normen p_m en α_m waar statistici wel eens van kunnen opkijken. In een volgende bijdrage wil ik daar wel eens over uitweiden.

Nu nog even terug naar het voorbeeld. Stel nu eens dat een geldsteekproef van zestig euro is getrokken en er één euro correspondeerde met een auto die helemaal niet te vinden is. Een beetje vreemd voorbeeld, maar dat is nodig voor de verdere uitleg van mijn theorie. Na lang zoeken wordt die auto op nul gewaardeerd. Hoeveel zou men nu dit bedrag van tien miljoen moeten afwaarderen omdat er nog meer auto's kwijt zouden kunnen zijn?

De beste schatting is natuurlijk $\frac{1}{60}$ van 10.000.000 (merk op dat de verantwoorde waarde van de kwijtgeraakte auto onbelangrijk is, er is immers maar één steekproefelement dat nul in plaats van één had moeten zijn), maar hoe onnauwkeurig is die schatting? Wat is met andere woorden de maximale fout? Hiervoor bepalen we de maximale fout p^u, met k_a het aangetroffen aantal fouten en p^u de maximale fout bij onbetrouwbaarheid α_m door:

$P[k \leq k_a \mid M, n, p \geq p^u] \leq \alpha_m$.

Ook hier is de Poissonverdeling het makkelijkst, al blijft het lastig. Immers, zelfs bij $k_a = 1$ moeten we al oplossen:

$\exp(-np) + np\exp(-np) = \alpha$

en dat is geen gesloten vergelijking. Ik heb dit ooit opgelost door te kijken naar het product np, dat 3 is bij 5% risico (α) en 0 fouten (want $\exp(-3) = 0{,}05$, of liever

$\ln(0{,}05) \approx 2{,}9957$.

Als ik één fout meer krijg (dan 0) gaat np zeker 1 en maximaal 2 omhoog, dus zoek ik naar de juiste np tussen 4 en 5 waarvoor de vergelijking voldoet. Dat zoeken kan lekker snel door eerst een keer 4 en een keer 5 in te vullen, na te gaan hoever die twee een uitkomst opleveren die

van 0,05 is verwijderd, en dan de slechtste van de twee te verruilen voor een gewogen gemiddelde ervan, gewogen met inversen van afstanden tot 0,05.
Ik heb dat in BASIC (jawel, 1986!) geprogrammeerd, voor de drie kansverdelingen en voor maximaal 500 fouten, inclusief een glijdende overgang naar de normale verdeling tussen de 65 en de 70 fouten die ik moest gebruiken omdat mijn computer weigerde 69 (zonder nadenken!) uit te rekenen. Alleen $\alpha = 50\%$ lukt(e) me niet goed. Natuurlijk is dat door betere programmeurs snel daarna in PASCAL overgedaan, en dat programma is ooit geijkt aan soortgelijke programma's van de Belastingdienst en van de Algemene Rekenkamer.
De uitkomst in het voorbeeld is overigens 4,75, zodat we kunnen stellen dat de maximale fout om en nabij de 6% is; in geld: $\frac{4,75}{60}$ van 10.000.000.
Een meer realistisch voorbeeld zou zijn dat van die zestig getrokken euro's er een zou zijn die een auto aanwees met een verantwoorde waarde van 30.000 die volgens de kopende partij maar 20.000 waard zou zijn. Een $\frac{1}{3}$ fout, dus.
De beste schatting zou nu logischerwijze op $\frac{1}{3}$ van $\frac{1}{60}$ van 10.000.000 komen, maar voor de maximale fout lopen we tegen de discreetheid van de Poissonverdeling aan. K.W. Stringer suggereerde in 1963 om een gewogen gemiddelde van (np)-waarden te nemen met de relatieve fout als gewicht, om zo de maximale fout te bepalen op:

$$\left\{\frac{\frac{2}{3} \times 3 + \frac{1}{3} \times 4,75}{60}\right\} \times 10.000.000.$$

Het intrigerende van deze methode (de Stringer bound) is dat deze vanaf 1963 is gebruikt door accountantskantoren, maar dat er pas in 1989 een bewijs voor is geleverd door de Hongaar G. Pap en de Nijmeegse hoogleraar M. van Zuijlen.
Nu het rekenwerk aan de orde is geweest, zal ik in een volgende bijdrage laten zien hoe dit allemaal past in de methodologie van de accountantscontrole. Daarbij komt aan de orde wat het nut is van maximale fouten en beste schattingen, en hoe accountants hun normen p_m en α_m bepalen.

Paul van Batenburg, Deloitte & Touche

Noot

[1] Batenburg, P. van (2002). BWI-ers in de praktijk. *Nieuwe Wiskrant, 22*(2), 18-20.

ICT bij wiskunde in de les en erbuiten

ICT2003 is de derde landelijke conferentie over ICT-gebruik in het wiskundeonderwijs, waarbij wiskundedocenten zich kunnen verdiepen in recente ontwikkelingen op dit gebied. Op deze conferentie, die wordt georganiseerd door het Freudenthal Instituut en APS-wiskunde, staat het gebruik van ICT binnen, maar ook buiten de wiskundeles centraal:

- voorbeelden uit de klassenpraktijk van basisvorming, VMBO en tweede fase
- ICT-gebruik van leerlingen thuis
- mogelijkheden voor de wiskundesectie
- overzicht van nieuwe ontwikkelingen.

In parallelpresentaties worden ervaringen met ICT-gebruik in de klas gepresenteerd. In hands-on workshops kunt u zelf ondervinden welke mogelijkheden software biedt. In de keuzewerkgroepen komen onder andere de volgende onderwerpen aan bod:

- excel benutten voor de wiskundeles
- wiskundesoftware voor taalzwakke leerlingen
- applets aanpassen en veranderen.

Daarnaast wordt de uitslag bekend gemaakt van de Nederlandse Wiskunde Webstrite, een wedstrijd waarin de beste website van een wiskundesectie wordt gekozen door het publiek en door een deskundige jury. Met mooie prijzen!

ICT2003 vindt plaats op donderdag 24 april 2003 van 9.30 tot 16.15 uur te Utrecht.
Meer informatie over deze conferentie kunt u vinden op www.fi.uu.nl/ict/2003.
Voor inlichtingen en inschrijving: Yolanda Velo, APS-wiskunde, tel. 030-2856722.