

## Een statistische steekproef die echt statistisch is

Bij gegevensgerichte werkzaamheden voeren drie vragen de boventoon.

- Hoeveel posten moeten we controleren?
- Welke pakken we dan?
- Hebben we dan wel genoeg gedaan?

Goed dat deze vragen gesteld worden, want voldoende en geschikte controle-informatie is nu eenmaal het belangrijkste element van onze taak.

Vanwege de *gepercipieerde* objectiviteit, de wetenschappelijke onderbouwing en de concreetheid bestaat bij sommige mensen de voorkeur voor het uitvoeren van een steekproef. Maar zoals ieder krachtig gereedschap een gebruiksaanwijzing heeft, geldt dit ook voor steekproeven. Door deze te lezen, wordt optimaal gebruik gewaarborgd.

Gebruiksaanwijzingen voor statistiek zijn waardevolle NIVRA geschriften, een Engelstalige uitleg van de AICPA en er kan altijd nog een expert worden geraadpleegd. De belangrijkste leerpunten uit het lezen van die gebruiksaanwijzing en het inschakelen van een expert zijn opgenomen in dit artikel.

- Het type verdeling.
- Het belang van stratificatie.
- De inschatting van de verwachte fout.

### Het type verdeling

Het type verdeling is voor mij al leek het minste grijpbare begrip, dus laten we met het moeilijkste beginnen. In de statistiek worden bijvoorbeeld hypergeometrische (met terugleggen), binominale (zonder terugleggen) en Poisson verdeling (gemaakt voor het schatten van het tijdsverloop van gebeurtenissen) toepasbaar geacht voor de accountantscontrole.

Belangrijkste wat uit de gebruikshandleiding naar voren komt is dat de bijbehorende berekening in volgorde van complexiteit aflopend zijn en in volgorde van hoeveelheid werk oplopend staan. Aangezien de complexiteit van berekeningen in goede ondersteunende spreadsheets kan worden opgelost, bepaalt dan de keuze voor het type verdeling alleen nog de hoeveelheid werk. Dus door het kiezen van een andere verdeling als basis, kan zonder concessies te doen aan de kwaliteit, werk worden bespaard. Dat is een belangrijk leerpunt.

### Het belang van stratificatie

Een belangrijke randvoorwaarde voor het uitvoeren van een steekproef is dat de massa waaruit elementen worden getrokken 'homogeen' zijn. Daar komt subjectiviteit om de hoek kijken. Subjectiviteit, omdat we bezig zijn met het verzamelen van controlebewijs ten aanzien van het risico op een afwijking van materieel belang. Wordt de massa dan niet opgeknipt in onderdelen die *qua risico* gelijk zijn, zorgt het toepassen van een selectie-interval er automatisch voor dat grote of vaak voorkomende posten een grotere kans krijgen om te worden getrokken. Dit hoeft niet altijd overeen te komen met de elementen waarin de grootste kans op een fout aanwezig is.

- Artikelen waarvan grote voorraden aanwezig zijn, kunnen wel de meest courante zijn.
- Kostensoorten die 12 keer per jaar voorkomen zoals lease-auto's, huur en hostingkosten zijn vaak niet de veroorzaker van mogelijke materiele fouten.

- Bij 'de grootste' transacties zijn de mensen die bij die transactie betrokken zijn (inclusief degene die moet autoriseren) het meest alert.

Door op dit punt niet de gebruiksaanwijzing van de statistiek te volgen, neemt de kans toe dat de gegevensgerichte werkzaamheden zullen bestaan uit het controleren van de items waar het risico dat wordt gecontroleerd helemaal niet op van toepassing is. Dit leidt mogelijk tot een andere uitkomst dan wanneer de massa was opgedeeld in meerdere strata en er per stratum een steekproef was getrokken. Ieder met een eigen verwachte fout, het volgende en meest belangrijke leerpunt.

### **De inschatting van de verwachte fout**

De meest subjectieve parameter bij het plannen van een steekproef is zonder meer de verwachte fout. Bij een steekproef op basis van een binominale verdeling neemt de omvang van de uit te voeren werkzaamheden toe met 65% als de verwachte fout toeneemt van 0 naar 25% van de uitvoeringsmaterialiteit.

Bij een steekproef op basis van een poisson verdeling neemt de steekproeffactor toe van 3 naar 6,3 bij de inschatting van 2 defecten in plaats van nul. Hier steken 2 belangrijke vraagstukken de kop op:

- Waar moeten we de inschatting van die verwachte fout op baseren?
- Welke mate van onderbouwing in het dossier is hiervoor genoeg.

Gezien het sterk bepalende effect en subjectieve karakter had ik verwacht hier al meer over te hebben gehoord. Vooral aangezien het gebruik van steekproeven de laatste tijd weer aardig in opkomst is.

### **Suggestie**

Uit overleg met een deskundige is naar voren gekomen dat het voor controleteams minder subjectief en waarschijnlijk effectiever is om dezelfde wiskunde andersom te gebruiken.

- Begin met het selecteren van de meest risicovolle items. Dat kunnen grotere items zijn, maar ook soms juist items in de middencategorie (omdat mensen dan minder alert zijn tegen fraudes en fouten). Door het meest risicovolle stratum te onderzoeken en met die uitkomsten te evalueren, vermijd je dat ieder stratum apart moet worden onderzocht. En omdat bij de evaluatie ervanuit wordt gegaan dat de hele massa bestaat uit zulke risicovolle items, wordt automatisch voorzichtigheid ingebouwd. Ervaring wijst uit dat dit deel van de werkzaamheden het meest interessant is, vooral indien goed uitgevoerd met behulp van data-analyse.
- Bereken op basis van de uitgevoerde werkzaamheden, de objectief waarneembare gevonden fouten en de omvang van de totale massa het niveau van zekerheid dat ondertussen is bereikt en stel vast of dat genoeg is. Belangrijkste voordeel is dat zo de inschatting vooraf van de verwachte fout niet hoeft te worden gemaakt.

Zo is men toch bezig met professionele oordeelsvorming zonder dat men het risico loopt door een onjuiste inschatting van de verwachte fout veel te weinig te doen.