

Maakt data-analyse steekproeven overbodig?

Wat is de plaats van steekproeven in een omgeving waar data-analyse wordt uitgevoerd? Die vraag staat centraal in de tweede column in de serie Statistical Auditing.

Paul van Batenburg

Data-analyse omvat een verscheidenheid aan softwaretoepassingen waarmee we een beeld kunnen krijgen van de kwaliteit van een dataset. Dit 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 deze column*). 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 doe je alleen als je niet integraal kunt controleren. Wanneer voor een controle brongegevens nodig zijn, is dat vaak het geval. Tenzij je 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 je af te vragen waar die betaalopdrachten vandaan komen

Voordeel van steekproeven is dat je ook kunt stuiten op foutsoorten waarvan je niet wist dat je ze in een integrale data-analyse moest meenemen (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 er naar moet kijken.

Andersom kan data-analyse heel effectief en 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 is 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 plausibiliteitsgrens, 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 Behandelings 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).

Stuurgroep Statistical Auditing

De Stuurgroep Statistical Auditing is verbonden met het Limperg Instituut en heeft als doel 'het bevorderen van het correcte (effectief en efficiënt) gebruik van statistische methoden en technieken bij accountantscontroles en daarmee verwante controles op financiële verantwoordingen en overzichten'.

Deel dit artikel



Drs. Paul C. van Batenburg is lid van de stuurgroep Statistical Auditing van het Limperg Instituut.

GERELATEERD



STATISTICAL AUDITING (88) | 15 december 2020

Inventarisatie ten behoeve van de controle op de NOW regeling: plan B in actie!

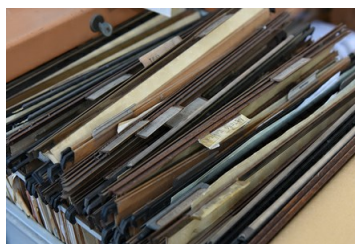
Tot nu toe heb ik altijd accountants aangeraden om bij de controle op de hoeveelheden in de voorraadadministratie geen geldsteekproef maar een postensteekproef te... →



STATISTICAL AUDITING (87) | 08 oktober 2020

Het gebruik van cijferanalyse ten tijde van Corona: zorg voor een plan B!

Gegevensgerichte cijferanalyse is te zien als een vorm van machine learning. Er wordt een relatie verondersteld tussen te controleren gegevens en andere gegevens.... →



STATISTICAL AUDITING (86) | 21 augustus 2020

Controle van een bestand: de steekproefomvang is niet het probleem, maar de populatie!

Het gebeurt nog wel eens dat een accountant wordt gevraagd om, los van de jaarrekeningcontrole, een uitspraak te doen over de kwaliteit van een bestand. →



STATISTICAL AUDITING (85) | 17 juli 2020

Boete? Waarom en waarover?

Onlangs kreeg ik een vraag van een fiscalist over een boete die werd opgelegd naar aanleiding van een correctie die was gebaseerd op een statistische steekproef.... →



STATISTICAL AUDITING (84) | 22 juni 2020

IPE testing: 'wat' is belangrijker dan 'hoeveel'

Onlangs had ik een discussie met accountants over de steekproefomvang bij IPE-testing (IPE=Information provided by entity). Conclusie: gebruik je eigen professionele... →
