

De steekproefomvang ontmaskerd - deel 2

Deze tweede column in een serie van vijf bespreekt hoe je zelfstandig de juiste steekproefomvang kunt berekenen.

Niels van Leeuwen

In de vorige column zagen we dat het classificeren van waarnemingen als goed of fout het werken met een discrete verdeling vraagt. Wij behandelden drie discrete verdelingen, namelijk de binomiale, hypergeometrische en Poissonverdeling.

Bij een audit willen we voldoende zekerheid hebben dat er geen materiële hoeveelheid aan fouten of onzekerheden bestaan in onze populatie. De vraag is nu hoe we door middel van statistiek kunnen stellen dat we genoeg waarnemingen hebben gedaan om te spreken van voldoende zekerheid. Om hierachter te komen gaan we in op de vraag: hoe bereken je de steekproefomvang? We beantwoorden deze vraag door eerst het begrip 'voldoende betrouwbaarheid' te duiden in een formule. Vervolgens lossen we deze formule op drie verschillende manieren op:

- Benaderen in Excel voor meerdere verdelingen
- Werken met np-tabellen voor de Poissonverdeling
- Toepassen van de inverse gamma verdeling in Excel als benadering voor de Poissonverdeling

Voldoende betrouwbaarheid

Zoals we in de vorige column vaststelden is het doel van statistiek in de audit om met een beperkt aantal waarnemingen voldoende betrouwbaarheid te verkrijgen dat een populatie vrij is van materiële fouten en onzekerheden. Nu is het woord 'voldoende' natuurlijk onderhevig aan oordeelsvorming. Het hangt dus af van de context wanneer je spreekt over voldoende betrouwbaarheid. Om aan te duiden wat voldoende is gebruiken wij het audit-risk model.

Het audit-riskmodel en statistiek

Hoeveel je moet doen om van voldoende betrouwbaarheid te kunnen spreken hangt af van het geaccepteerde risico op de opdracht. Het audit-riskmodel zegt dat het accountantsrisico (AR) het product is van het inherente risico (IR), het controle risico (CR) en het detectierisico (DR). Het inherente risico is wat er wezenlijk fout kan gaan bij de controle, zonder dat de gecontroleerde organisatie hier iets aan doet. Het controlerisico is het risico dat een materiële fout niet wordt opgemerkt door het stelsel van interne beheersing. Het detectierisico is het risico dat de auditor fouten niet ziet. Hoe meer werk de auditor doet, hoe lager het detectierisico. De drie risico's moeten gezamenlijk leiden tot een acceptabel laag accountantsrisico. In formulevorm als volgt:

$$AR = IR \times CR \times DR$$

Statistiek kan worden toegepast in de werkzaamheden van de auditor om het detectierisico te verlagen. Hiervoor moet dit detectierisico nog worden vertaald naar een statistische maatstaf. Het detectierisico is in feite gelijk aan het beta-risico. Het beta-risico is het risico dat een hypothese wordt aangenomen terwijl deze feitelijk moet worden verworpen. Met andere woorden, het beta-risico is in deze situatie het risico dat de auditor een populatie goedkeurt, terwijl er feitelijk sprake is van een materiële fout.

Dit risico moet dus acceptabel laag zijn. Wat acceptabel laag is hangt samen met de hoogte van het inherente en controlerisico, maar als deze buiten beschouwing worden gehouden is het gebruikelijk om hooguit een beta-risico van 5% aan te houden. Dit is hetzelfde als een betrouwbaarheid van 95 procent. We kunnen deze 95 procent betrouwbaarheid invullen in de in de eerste column gepresenteerde formules die de kans op aantal fouten aangeeft voor de drie verdelingen. We willen met 95 procent betrouwbaarheid kunnen stellen dat het maximaal aantal fouten de materialiteit niet overschrijdt. In formulevorm ziet dat er als volgt uit:

Verdeling	Hypergeometrisch	Binomiaal	Poisson
Formule	$95\% \leq 1 - \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$	$95\% \leq 1 - \binom{n}{k} p^k (1-p)^{n-k}$	$95\% \leq 1 - \frac{e^{-\lambda} \lambda^k}{k!}$
Waarhi:	N = de populatie omvang K = aantal fouten in de populatie n = steekproefomvang k = aantal waargenomen fouten	Waarhi: n = het aantal waarnemingen k = het aantal fouten p = de foutfactor = K/N	Waarhi: λ = gemiddeld aantal fouten per control = n * p n = het aantal waarnemingen p = de foutfactor = K/N k = het aantal fouten e = Getal van Euler

De populatieomvang N is een vast gegeven en is de totale omvang waar je een steekproef op berekent. De grote K duidt het aantal fouten in de populatie aan. Deze parameter is onbekend, maar we weten wel dat deze maximaal zo groot mag zijn als de materialiteit. Daarom wordt voor K de materialiteit aangehouden. De kleine k is het aantal fouten dat je tegenkomt in de steekproef. Deze parameter is vooraf ook onbekend, dus bij de planning van een steekproef wordt deze geschat. De foutfactor uitgedrukt in letter p is gelijk aan het aantal fouten in de populatie gedeeld door de populatieomvang. De Griekse letter λ staat voor een gemiddeld aantal fouten per interval en dit is gelijk aan de steekproefomvang vermenigvuldigd met de foutfactor.

In feite is het berekenen van de steekproefomvang een kwestie van het invullen van de parameters tot we die 95 procent betrouwbaarheid hebben bereikt. De parameters N , K , k , p en λ zijn van tevoren bepaald, de enige variabele is n . Deze kun je steeds ophogen tot de uitkomst van de formule uiteindelijk 95 procent aangeeft. We kunnen nu met enkele manieren deze vergelijking oplossen.

Verschillende manieren om de steekproefomvang uit te rekenen

Wij illustreren de manieren van werken aan de hand van twee casussen, die we gedurende deze gehele serie van columns gebruiken. Eén casus, waarbij we geen fouten verwachten, en één waarbij we wel fouten verwachten.

Wanneer we namelijk geen fouten verwachten kunnen we de steekproefomvang ook handmatig uitrekenen, wat we in de volgende column laten zien. Wanneer we wel fouten verwachten kan dat niet handmatig en moeten we een van de andere manieren gebruiken. In beide casussen is de populatieomvang € 1 mln. We hebben een materialiteit van 5 procent, wat neer komt op € 50.000. We willen 95 procent betrouwbaarheid behalen, wat gelijk is aan een maximaal beta-risico van 5 procent. In de eerste casus verwachten we 0 fouten (k_1) en in de tweede casus verwachten we 1 fout (k_2). Dit kunnen we samenvatten in de volgende variabelen:

- $\beta = 5\%$
- $N = 1.000.000$
- $p = 5\%$
- $k_1 = 0$
- $k_2 = 1$

Benaderen met Excelfuncties

We beginnen met de manier die het meest aansluit bij de bovenstaande formules. Met deze manier vullen we de parameters in een Excel-formule in, waarbij we de steekproef n steeds verhogen tot we onze gewenste betrouwbaarheid hebben behaald. We moeten hiervoor de onderstaande formules in Excel invullen.

Verdeling	Functie
Binomiaal	$95\% \leq 1 - \text{BINOM.VERD}(k; n; P; 1)$
Hypergeometrisch	$95\% \leq 1 - \text{HYPGEOM.VERD}(k; n; K; N, 1)$
Poisson	$95\% \leq 1 - \text{POISSON.VERD}(k; n \cdot p, 1)$

Wanneer we een benadering doen voor beide scenario's komen we op de volgende steekproefomvang uit:

Verdeling	$k_1 = 0$	$k_2 = 1$
Binomiaal	59	93
Hypergeometrisch	60	95
Poisson	59	93

In het bijgevoegde Excel-rekenblad kun je dit zelf toepassen in Excel.

Werken met np-tabellen voor de Poissonverdeling

Een andere relatief eenvoudige manier is het werken met np-tabellen. Deze tabellen geven voor de Poissonverdeling een vaste verhouding in tabel vorm weer. De parameters waar je mee werkt zijn k en $1-b$. Door deze te gebruiken kun je de corresponderende np-waarde vinden. De n is eenvoudig te verkrijgen door deze np-waarde te delen door p . Als we deze waardes invullen komen we uit op 3 en 4,74. Als we deze waardes delen door een p van 5 procent komen we uit op

een steekproefomvang van 60 en 95, afgerond naar boven.

$k 1-\beta n$	99% (1- β)	(n)	95% (1- β)	(n)	90% (1- β)	(n)
0	4,61	93	3,00	60	2,30	47
1	6,64	133	4,74	95	3,89	78
2	8,41	169	6,30	126	5,32	107
3	10,05	201	7,75	156	6,68	134
4	11,60	233	9,15	184	7,99	160
5	13,11	263	10,51	211	9,27	186

Toepassen van de inverse gamma verdeling in Excel voor de Poissonverdeling

Wanneer je van een kansberekening weet dat de uitkomst 95 procent moet zijn en alle parameters behalve één bekend zijn, dan kun je de onbekende oplossen. Dit is precies wat de inverse gamma verdeling doet voor de Poissonverdeling. Excel kent geen inverse functie van de Poisson verdeling, maar wel van de gammafunctie. De gamma- en Poissonverdeling zijn gelijk aan elkaar. Het verschil is dat de gammafunctie ook geschikt is voor continue data. Met de inverse gammafunctie kun je dus de np-tabel overslaan en in Excel direct de steekproef terugrekenen. Je gebruikt hiervoor de volgende formule:

Verdeling	Functie
Inverse gamma	=GAMMA.INV(1- β ;k+1;1)/p

Ook met deze manier kom je bij k1 en k2 uit op 60 en 95 steken naar boven afgerond. Hoe je deze formule toepast is ook opgenomen in het bijgevoegde Excel-rekenblad.

Kleine verschillen

We hebben nu drie manieren toegepast om tot dezelfde steekproefomvang te komen. Hierbij zien we wel kleine verschillen in omvang. In de eerdere column hebben we de voor- en nadelen van de verschillende manieren besproken. Nu je deze verschillen kent kun je zelf de steekproefomvang op verschillende manieren berekenen en is het hopelijk geen mysterie meer hoe je tot de juiste steekproefomvang komt. Hopelijk geeft dat je meer vertrouwen bij het gebruik van statistiek in de audit.

In de komende column gaan we nog iets dieper in op de gepresenteerde manieren. We bekijken waarom het logisch is dat de inverse functie op een goede steekproefomvang uitkomt. In de vierde column automatiseren we de benaderingswijze in een VBA-script met als doel beter te begrijpen hoe je met behulp van software tot een steekproefomvang komt. In de laatste column laten wij zien hoe je de open software-omgeving JASP kunt gebruiken bij het berekenen van de software van de steekproefomvang.

- [Excel-rekenblad](#)

Bron: Stewart, T. (2012). *Technical notes on the AICPA audit guide Audit Sampling*. American Institute of Certified Public Accountants, New York, 5-8.

Deel dit artikel



GERELATEERD



STATISTICAL AUDITING (95) | 25 juli 2022

De steekproefomvang ontmaskerd - een introductie

Steekproefomvang berekenen doen we meestal met rekenbladen in Excel of statistische software. Hoewel velen van ons statistiek hebben gehad, is het berekenen van... →



STATISTICAL AUDITING (94) | 17 september 2021

Fraudeonderzoek en steekproeven

Wat kenmerkt een fraudeonderzoek? Hoe moet men zo'n onderzoek, methodologisch gezien, aanpakken? Kortom: wat wil men bewijzen? Dat de desbetreffende organisatie... →



STATISTICAL AUDITING (93) | 30 juli 2021

De pilotsteekproef: een gids die je ook de verkeerde weg kan wijzen!

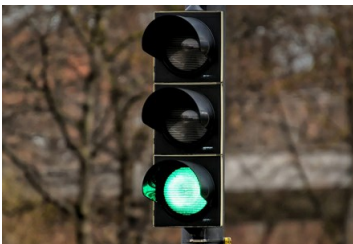
Deze column gaat over pilotsteekproeven. Hiervoor wordt vaak uit de (statistisch) losse pols een omvang van 25 voorgesteld. Deze column illustreert dat een kleine... →



STATISTICAL AUDITING (92) | 17 juni 2021

Van materialiteit naar uitvoeringsmaterialiteit

De vertaling van (overall) materialiteit naar uitvoeringsmaterialiteit komt niet zo maar uit de lucht vallen. Daar moet de accountant wel wat voor doen. Moet hij... →



STATISTICAL AUDITING (91) | 30 april 2021

Van uitvoeringsmaterialiteit naar toelaatbare afwijking

Deze column is een pleidooi voor het hanteren van een buffer tussen uitvoeringsmaterialiteit en toelaatbare afwijking. →
