De steekproefomvang ontmaskerd - deel 3

Steekproefomvangen berekenen doen we meestal met rekenbladen in Excel of statistische software. Hoewel velen van ons statistiek hebben gehad, is het berekenen van de steekproefomvang in de audit toch lastig. In deze column gaan we in op hoe je handmatig een steekproefomvang kunt berekenen.

Niels van Leeuwen

In vorige columns hebben we verschillende manieren besproken om tot een steekproefomvang te kunnen komen. We stelden vast dat wanneer we waarnemingen classificeren in goed of fout, we spreken over discrete data. De verdeling die het beste past bij zulke data is de hypergeometrische. We hebben ook gezien dat de binomiale en Poisson-verdeling het rekenen vereenvoudigen. We hebben drie manieren besproken om tot een steekproefomvang te kunnen komen, namelijk het benaderen in Excel, werken met np-tabellen en gebruikmaken van de inverse gammafunctie in Excel. De inverse gammafunctie is een alternatief voor de inverse Poisson-verdeling.

In deze column wordt duidelijk gemaakt waarom het terecht is dat je deze formule kunt gebruiken. De inverse Poisson-verdeling komt namelijk tot stand door de Poisson-formule zo in te richten dat de uitkomst van de formule de np-waarde is, in plaats van de betrouwbaarheid. Dit herinrichten gebeurd op vergelijkbare wijze als bij 12 = 3 x 4, wat gelijk is aan 3 = 12 / 4. Als we in de bestaande Poisson-formule alle parameters invullen behalve de np-waarde, dan kunnen we deze gesloten formule oplossen. Net zoals wanneer we de waarde van x kunnen vinden in de formule 3 = 12 / x. Dit kan handmatig wanneer het verwachte aantal fouten 0 is. Dit kan ook met de binomiale formule.

Laten we beginnen met het ophalen van de parameters uit de vorige columns. We werkten met een populatieomvang N = € 1mln. We hebben een materialiteit van p = 5 procent wat neer komt op € 50.000. We willen 95 procent betrouwbaarheid behalen, dus 1 - b = 95 procent. Omdat er alleen een gesloten formule bestaat voor nul fouten, houden we het verwachte foutpercentage op 0 procent, dat betekent dat het aantal verwachte fouten in de steekproef gelijk is aan k = 0. Dit kunnen we samenvatten in de volgende parameters:

- 1-b = 95 procent
- n = 1.000.000
- p = 5 procent
- k = 0

We beginnen met de binomiale verdeling. Hoewel in Excel de inverse functie van de binomiale verdeling niet leidt tot een steekproefomvang, kun je wel handmatig de steekproefomvang berekenen met de binomiale formule. In onderstaande tabel staat stapsgewijs toegelicht hoe dit werkt.

Stap en toelichting	Formule
Stap 1: De kans op 0 fouten juist definiëren We willen graag met ten minste 95 procent zekerheid kunnen stellen dat het aantal fouten niet groter is dan 0. Dat is gelijk aan maximaal 5 procent kans dat het aantal fouten 0 is.	$P(k \leq 0) \geq 95\%$ = $P(k = 0) < 5\%$
Stap 2: de parameters in de formule invullen We definiëren de formule voor de onderstaande parameters als volgt: • β = 5% • p = 5% • k = 0 De binomiale formule ziet er als volgt uit:	$P(k) > \binom{n}{k}(p)^k(1-p)^{n-k}$
Stap 2: De inverse formule invullen Als we de genoemde parameters invullen in bovenstaande formule, ziet dit er als volgt uit: In deze formule zit een notering van een combinatie $\binom{n}{0}$. Voor deze notering gebruik je de nCr functie op je rekenmachine. Wanneer het onderste getal 0 is, zal de uitkomst echter altijd op 1 uitkomen. Dit geldt ook voor iets tot de 0e macht verheffen. Hierdoor kunnen we het bovenstaande ook presenteren als:	$5\% > \binom{n}{0}(0{,}05)^0(0{,}95)^{n-0}$ $5\% > 1 * 1(0{,}95)^n$
Stap 3: De macht wegwerken via de natuurlijke logaritme Om de bovenstaande formule op te lossen kun je de natuurlijke logaritme op beide zijde toepassen. Dat is de ln functie op je rekenmachine. Hierdoor komt de machtsverheffing als vermenigvuldiging uit op de rechterzijde, ofwel: Wordt dus: En dit is weer gelijk aan: Ofwel	$ln(0{,}05) > ln(0{,}95)^n$ $ln(0{,}05) > \ln(0{,}95) * n$ $n > \frac{\ln(0{,}05)}{\ln(0{,}95)}$ $n > \frac{\ln(\beta)}{\ln(1-p)}$

De uitkomst hieruit is 58,40, wat afgerond naar boven neer komt op 59. Dit is inderdaad gelijk aan de uitkomst van de benaderingsmanier van de binomiale verdeling uit de vorige column. Wanneer je met meer verwachte fouten wilt werken zal niet meer uitkomen op 1, waardoor het handmatig een stuk moeilijker zal worden om de formule te verwerken.

Deel dit artikel

GERELATEERD

STATISTICAL AUDITING (96) | *21 september 2022*

De steekproefomvang ontmaskerd - deel 2

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

STATISTICAL AUDITING (95) | *25 juli 2022*

De steekproefomvang ontmaskerd - een introductie

Steekproefomvangen 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... →