

Toepassing van data-analyse om de steekproef te rationaliseren

Koen Derks, Lotte Mensink, Jacques de Swart, Ruud Wetzels

Received 28 March 2024 | Accepted 29 August 2024 | Published 18 September 2024

Samenvatting

Data-analyse wordt steeds belangrijker in de audit. Hoewel het potentieel om de audit efficiënter te maken groot is, worstelen veel auditors met het integreren van de resultaten van deze analyses om daarmee de vervolgwerkzaamheden te verlichten. Dit artikel bespreekt de integratie van de resultaten van data-analyse in de steekproeffase middels Bayesiaanse statistiek, en beargumenteert dat dit kan leiden tot een verbetering van de efficiëntie en transparantie. Drie voorbeelden van regressie, classificatie en clustering illustreren deze aanpak. Het artikel sluit af met aanbevelingen voor de praktische toepassing van deze methodiek.

Relevantie voor de praktijk

Data-analyse kan de audit efficiënter maken. Desondanks is het voor veel auditors een uitdaging om data-analyse te benutten voor efficiëntiewinst in de audit. In dit artikel worden aan de hand van drie voorbeelden de mogelijkheden van Bayesiaanse statistiek belicht, waarmee voortgebouwd kan worden op data-analyse om de steekproeffase te verkleinen.

Trefwoorden

Steekproef, Bayesiaanse statistiek, data-analyse, machine learning, prior-verdeling

1. Inleiding

Data-analyse wordt steeds vaker toegepast in de audit-praktijk (NBA 2019), een trend die mede mogelijk wordt gemaakt door het toenemende gebruik van (big) data door bedrijven (Appelbaum et al. 2017). Deze trend leidt tot meer beschikbare data voor de auditor en een breed scala aan mogelijkheden om deze data te analyseren. Hoewel data-analyse de potentie heeft om de audit efficiënter te maken (Gepp et al. 2018), is het voor veel auditors een uitdaging om de resultaten van deze analyses te integreren om latere werkzaamheden te rationaliseren. Dit artikel bespreekt een statistisch raamwerk om de resultaten van data-analyse te integreren in de audit en illustreert hoe dit kan leiden tot een transparantie- en efficiëntiewinst aan de hand van een drietal voorbeelden. Deze voorbeelden zijn omwille van de leesbaarheid enigszins versimpelde, doch representatieve versies van casussen die men aantreft in de praktijk.

Data-analyse kan in elke fase van de audit worden toegepast. Auditors kunnen bijvoorbeeld data-analyse gebruiken om inzicht te verkrijgen in de auditee en de controleomgeving. Een voorbeeld van een techniek die hierbij veel wordt gehanteerd, is trendanalyse. Hierbij worden data van eerdere perioden vergeleken met de huidige periode om inzicht te krijgen in seizoenspatronen, marge- en prijsontwikkelingen, nieuwe of vervallen geldstromen of pieken in omzet of kosten (Snoei and Van Nieuw Amerongen 2015). Daarnaast kunnen auditors ook data-analyse gebruiken om risicobeoordelingsactiviteiten uit te voeren. Voorbeelden hiervan zijn het real-time monitoren van nieuwsberichten over de auditee om daar vervolgens met gebruik van natural language processing technieken een sentimentanalyse op uit te voeren (ING 2018), of met gebruik van deze technieken te evalueren hoe klanten de auditee of

andere marktspelers beschrijven. Verder kunnen auditors data-analyse gebruiken om de opzet en werking van interne beheersingsmaatregelen te beoordelen. Een voorbeeld hiervan is het gebruik van process mining om knelpunten in transactiestromen te vinden (Van der Aalst and Koopman 2015; Jans and Hosseinpour 2017). Als laatste kan data-analyse worden toegepast in de steekproeffase van de audit (De Swart et al. 2013). Zo kunnen gestratificeerde steekproeven bijvoorbeeld efficiënter of nauwkeuriger worden geëvalueerd door gebruik te maken van hiërarchische statistische modellen (Derks et al. 2022b). Hoewel data-analyse in de audit vaak afzonderlijk wordt ingezet – zie bijvoorbeeld de voorbeelden in NBA (2019) – kan de integratie van data-analyse in de audit leiden tot meer efficiëntiewinst.

Steekproeffase van de audit

Het streven naar efficiëntiewinst is bijna altijd gericht op de steekproeffase van de audit. De gegevensgerichte controles die in deze fase worden uitgevoerd blijven namelijk een veelgebruikt onderdeel van het auditproces (Christensen et al. 2015), met name als de norm voor de data niet integraal digitaal beschikbaar is. Daarnaast vereisen de standaarden dat de auditor een bepaald niveau van gegevensgerichte controles uitvoert (Standaard 330, paragraaf 18), zelfs als de processen effectief lijken, bijvoorbeeld omdat de interne controle door het management kan worden gepasseerd.¹ Echter, gegevensgerichte controles in de vorm van steekproeven zijn tijdrovend en daardoor kostbaar. Door data-analyse toe te passen voordat een steekproef wordt genomen, kunnen auditors de efficiëntie van de steekproeffase aanzienlijk verbeteren. Een voorbeeld hiervan is een gegevensgerichte analyse waarbij de totale huuropbrengsten van een appartementencomplex worden voorspeld op basis van de huurprijzen, het aantal appartementen en de leegstand in het gebouw (Standaard 520, paragraaf A8). In het geval dat de totale huuropbrengst niet te veel afwijkt van de voorspelde opbrengst, kan de auditor besluiten om een minder grote steekproef uit de populatie huurbetalingen te trekken. Vanzelfsprekend is het hierbij belangrijk dat de auditor een afweging maakt; als het meer tijd of geld kost om de relevante data te verzamelen en de analyse uit te voeren en te interpreteren dan de reductie in de steekproef waard is, dan is het verstandiger om de volledige steekproef te nemen.

Auditrisicomodel

Het idee dat data-analyse op deze manier kan bijdragen aan auditcomfort en daarmee het werk in de steekproeffase kan verminderen lijkt intuïtief gezien een no-brainer. Echter, het kwantificeren van de impact van data-analyse op het werk in de steekproeffase is al lang een punt van discussie. In de praktijk worden verminderingen in de omvang van de steekproef vaak gerechtvaardigd

door het gebruik van het auditrisicomodel. Dit model suggereert dat het risico op een onjuiste verklaring van de auditor het resultaat is van het inherente risico, het interne beheersingsrisico, het cijferanalyserisico, en het steekproefrisico (Kloosterman 2004b). Hoewel het auditrisicomodel geen verantwoord statistisch model is (Van Batenburg and Dassen 1996; Kloosterman 2004a), biedt het een kader voor het verminderen van de steekproefgrootte op basis van de beoordeling van de kans op een materiële fout in de populatie. In het eerdergenoemde voorbeeld kan de uitgevoerde analyse bijvoorbeeld leiden tot een verlaging van het cijferanalyserisico van ‘hoog’ naar ‘laag’, waardoor het steekproefrisico kan toenemen en de steekproefgrootte kan afnemen. Mits de relevante data gemakkelijk beschikbaar en te analyseren is, resulteert het gebruik van het auditrisicomodel op deze manier in een praktisch te rechtvaardigen efficiëntiewinst voor de auditor. Overigens is een andere tekortkoming van het auditrisicomodel dat het ook kan worden toegepast zonder enige data om de risico-inschattingen te onderbouwen, waarmee het vatbaar is voor ongefundeerde beweringen.

Hoewel het auditrisicomodel in de praktijk veelvuldig wordt toegepast, werkt het vooral beperkend met betrekking tot de integratie van de resultaten van data-analyse in de steekproeffase. Dit komt omdat het raamwerk de auditor verplicht om de resultaten om te zetten naar de kans op een materiële fout in de populatie. Echter, de resultaten van data-analyse zijn vaak niet direct te koppelen aan de kans op een materiële fout. In het eerdergenoemde voorbeeld is het bijvoorbeeld onduidelijk hoe een afwijking van de voorspelde opbrengst gerelateerd is aan de kans op een materiële fout in de populatie huurbetalingen. Praktisch gezien leidt het gebrek aan verband tussen deze twee grootheden tot twee mogelijke vervolgstappen. Ten eerste kan het zijn dat de auditor subjectief te werk moet gaan om de link tussen de twee grootheden te leggen, wat de verantwoording van de risico-inschatting afzwakt. Ten tweede kan het ertoe leiden dat de resultaten van de data-analyse helemaal niet worden geïntegreerd, wat het reeds uitgevoerde auditwerk tenietdoet. Beide opties hebben negatieve gevolgen voor de kwaliteit of de efficiëntie van de audit.

Bayesiaanse statistiek

Het is niet eenvoudig om informatie uit data-analyse die niet direct relateert aan de kans op een materiële fout in de populatie te integreren met traditionele (i.e., frequentistische) statistiek in de steekproeffase. Echter, het gebruik van Bayesiaanse statistiek biedt hiervoor een oplossing. In dit artikel laten we zien hoe de auditor middels deze vorm van statistiek kan voortbouwen op de uitkomsten van verschillende soorten data-analyse om de steekproeffase te verkleinen. Hiermee draagt het artikel bij aan de vraag naar de inzet van data-analyse om de efficiëntie van de audit te bevorderen (NBA 2019).

Dit artikel is als volgt gestructureerd. In hoofdstuk 2 wordt de theorie achter de Bayesiaanse statistiek uitgelegd en wordt besproken hoe bestaande informatie over de foutfractie in de populatie kan worden geïntegreerd in de statistische analyse via de prior-verdeling. In hoofdstuk 3 worden drie voorbeelden besproken om een prior-verdeling te vormen op basis van verschillende soorten analyses, zoals regressie, classificatie en clusteranalyse. In dit hoofdstuk wordt tevens getoond hoe dit leidt tot een efficiëntiewinst die direct kan worden verantwoord door de resultaten van de analyse. In het laatste hoofdstuk worden de conclusies en aanbevelingen voor de praktijk gepresenteerd.

2. De Bayesiaanse benadering van steekproeven

Binnen het vakgebied statistiek zijn er twee gangbare benaderingen: de frequentistische statistiek en de Bayesiaanse statistiek (Bayarri and Berger 2004; Wagenmakers et al. 2008). In de auditpraktijk wordt voornamelijk de frequentistische statistiek toegepast. Bij deze benadering baseert de auditor zich uitsluitend op de beschikbare data, zoals waarnemingen uit een steekproef, om uitspraken te doen over kenmerken van een populatie, bijvoorbeeld de foutfractie. De Bayesiaanse statistiek daarentegen hanteert een andere aanpak, waarbij de auditor een zogenaamde prior-verdeling opstelt die de bestaande informatie over de foutfractie in de populatie representeert. Deze prior-verdeling wordt vervolgens bijgewerkt met de data uit de steekproef om een kansuitspraak te doen over de foutfractie in de populatie, rekening houdend met alle informatie. Het voordeel van het gebruik van de prior-verdeling is dat deze kan worden opgesteld op basis van allerlei soorten bestaande informatie. Zo kan niet alleen informatie uit risico-inschattingen – die direct relateren aan de kans op een materiële fout – worden opgenomen in de prior-verdeling, maar bijvoorbeeld ook informatie uit meer complexe cijferanalyses die op een andere manier relateren aan de fout(fractie) in de populatie (Derks et al. 2021a), of expertkennis van de auditor (Crosby 1981). Op deze manier geeft Bayesiaanse statistiek auditors de mogelijkheid om in de steekproeffase voort te bouwen op diverse soorten data-analyses. Dit bevordert de integratie van data-analyse en de efficiëntie van de audit. Echter, het formaliseren en verantwoorden van de prior-verdeling wordt lastig gevonden (Chesley 1975), en daarom wordt er in de praktijk weinig gebruik gemaakt van Bayesiaanse statistiek.

Voorbeeld

Om de Bayesiaanse manier van steekproefevaluatie nader uit te leggen, schetsen we een voorbeeld. Stel dat een auditor de taak heeft om de inkoopfacturen van een

onderneming te controleren op overwaarderingen. De onderneming in kwestie heeft in enig jaar in totaal één miljoen euro aan boekingen op de betreffende grootboekrekening, waarvoor de auditor een uitvoeringsmaterialeiteit van drie procent hanteert. Dat betekent dat van de één miljoen euro aan boekingen maximaal drie procent een fout mag bevatten. In plaats van alle boekingen te controleren op fouten, kan de auditor ervoor kiezen om een monetaire, statistische steekproef uit de boekingen te trekken en de daaruit verkregen informatie te extrapoleren naar de gehele populatie.

Bij het evalueren van een statistische steekproef streeft de auditor ernaar om een kansuitspraak te doen over een bepaalde eigenschap, θ , van de populatie. In dit voorbeeld vertegenwoordigt θ bijvoorbeeld de foutfractie in de geldwaarde van de boekingen op de grootboekrekening. Merk op dat θ soms de monetaire foutfractie aangeeft en soms de foutfractie in het aantal boekingen, afhankelijk van het feit of de auditor een monetaire steekproef neemt of niet. In dit artikel laten we beide de revue passeren. Omdat de auditor niet de hele populatie inspecteert, maar alleen de data, y , afkomstig van een steekproef van deze populatie, moet de informatie uit de steekproef worden geëxtrapoleerd naar de hele populatie met inachtneming van onzekerheid. Dat betekent dat de auditor een kansuitspraak moet doen over θ . De Bayesiaanse manier om deze kansuitspraak te maken, gegeven de data y , is via de posterior-verdeling $p(\theta | y)$. Deze verdeling wordt door de stelling van Bayes gedefinieerd als proportioneel aan het product van de prior-verdeling $p(\theta)$ en de aannemelijkheidsfunctie $l(y | \theta)$ (Vergelijking 1). Het symbool \propto in Vergelijking 1 geeft aan dat de twee grootheden gelijk zijn aan elkaar, vermenigvuldigd met een schalingsfactor.

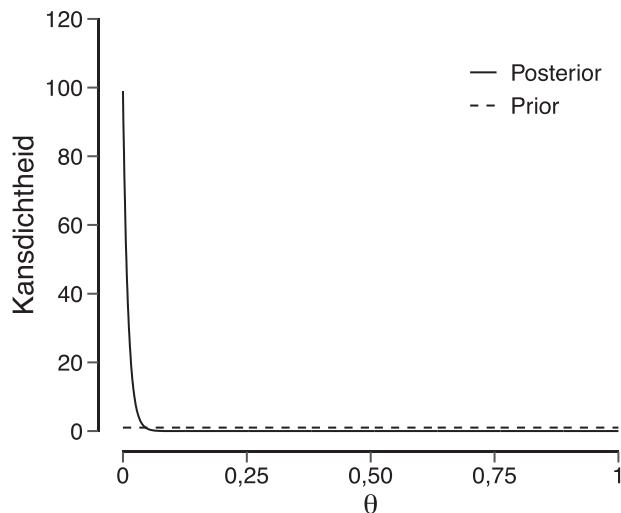
$$\frac{p(\theta | y)}{\text{Posterior}} \propto \frac{l(y | \theta)}{\text{Aannemelijkheid}} \times \frac{p(\theta)}{\text{Prior}} \quad (1)$$

Zoals Vergelijking 1 aangeeft, bestaat een Bayesiaans model uit drie componenten: de prior-verdeling, de aannemelijkheid, en de posterior-verdeling. In de volgende secties worden deze drie componenten van een Bayesiaans model kort uitgelegd.

2.1. De prior-verdeling

De prior-verdeling $p(\theta)$ representeert de bestaande auditinformatie over de foutfractie θ voordat er data uit een steekproef zijn gezien. De prior-verdeling is een kansverdeling, waarin aan elke mogelijke waarde van θ een relatieve plausibiliteit wordt toegekend, zodat deze integreert tot één. Een veelgebruikte prior-verdeling voor θ is een uniforme bèta ($\alpha = 1, \beta = 1$) verdeling. Deze prior-verdeling, weergegeven in Figuur 1, impliceert dat elke waarde van de foutfractie even waarschijnlijk is. Er zijn twee redenen waarom de uniforme prior-verdeling in de praktijk vaak wordt gebruikt. Ten eerste is deze

Figuur 1. De uniforme bèta(1, 1) prior-verdeling en de bèta(1, 99) posterior-verdeling na het zien van 98 foutloze waarnemingen. In het geval van de uniforme prior-verdeling is de posterior-verdeling gelijk aan de aannemelijkheidsfunctie. Het 95^e percentiel van de posterior-verdeling ligt onder de uitvoeringsmaterialiteit van drie procent, wat betekent dat minder dan vijf procent van de kansmassa boven de uitvoeringsmaterialiteit ligt.



gemakkelijk te begrijpen; ten tweede levert het evalueren van een steekproef met deze prior-verdeling resultaten op die bijna identiek zijn aan die van een frequentistische steekproefevaluatie (Derks et al. 2022a).

2.2. De aannemelijkheid

De aannemelijkheid representeert de informatie die de steekproefdata y over de foutfractie θ bevatten. De aannemelijkheidsfunctie $l(y | \theta)$ geeft de waarschijnlijkheid dat de steekproefdata zich voordoen onder specifieke waarden van θ (Edwards 1992; Etz 2018). In een auditsteekproef wordt doorgaans aangenomen dat de kans om k fouten te vinden in een steekproef van n boekingen met een foutfractie θ binomiaal verdeeld is, aangeduid als binomiaal($k | n, \theta$) (Stewart 2012).²

In het voorbeeld besluit de auditor om een steekproef van $n = 98$ waarnemingen te trekken en ontdekt dat in deze steekproef $k = 0$ waarnemingen een fout bevatten. Het aannemen van de binomiale aannemelijkheid houdt in dat de data uit deze steekproef binomiaal($k = 0 | n = 98, \theta$) zijn verdeeld.

2.3. De posterior-verdeling

Door de stelling van Bayes toe te passen, kan de auditor de informatie in de prior-verdeling combineren met de informatie in de data om tot de posterior-verdeling $p(\theta | y)$ te komen. De stelling van Bayes impliceert dat waarden van de foutfractie θ die de steekproefdata relatief goed voorspellen, waarschijnlijker worden dan

ze waren volgens de prior-verdeling, terwijl waarden van θ die de steekproefdata relatief slecht voorspellen, minder waarschijnlijk worden. De posterior-verdeling bevat dus de bijgewerkte kennis van de auditor over θ na het zien van de steekproefdata. Deze verdeling is dé bron van informatie voor de auditor, omdat het de informatie die de auditor had voor het nemen van de steekproef combineert met alle verkregen informatie uit de steekproef.

Net als de prior-verdeling is de posterior-verdeling een kansverdeling, waardoor de bijgewerkte kennis over θ kan worden samengevat met statistieken zoals het gemiddelde, de mediaan en de modus. Zo kan de auditor bijvoorbeeld een uitspraak doen over de meest waarschijnlijke fout door de modus van de posterior-verdeling te bepalen (oftewel, de waarde van θ met de hoogste waarschijnlijkheid). Evenzo kunnen percentielen van de posterior-verdeling worden geïnterpreteerd in termen van geloofwaardigheid (Kruschke and Liddell 2017). Het 95^e percentiel van de posterior-verdeling geeft bijvoorbeeld aan welke waarden van θ met 95 procent zekerheid onder deze waarde liggen. Het 95^e percentiel van de posterior-verdeling is daardoor een 95 procent zekere bovengrens voor de foutfractie in de populatie.

Na het zien van de binomiaal verdeelde steekproef van $n = 98$ items met $k = 0$ fouten is de posterior-verdeling een bèta($\alpha = 1 + k = 1, \beta = 1 + n - k = 99$) verdeling, die is weergegeven in Figuur 1. De modus van deze posterior-verdeling is nul, wat betekent dat de meest waarschijnlijke fout in de populatie wordt geschat op nul procent. Het 95^e percentiel van de posterior-verdeling is 2,9 procent. Dit betekent dat, met een waarschijnlijkheid van 95 procent, de foutfractie in de populatie lager is dan 2,9 procent. Deze 95 procent zekere bovengrens is lager dan de uitvoeringsmaterialiteit van drie procent. Dit betekent dat de auditor kan concluderen dat er minstens 95 procent kans is dat populatie geen materiële fout bevat, wat tevens impliceert dat er minder dan vijf procent kans is dat de populatie een materiële fout bevat. Deze laatste kans is klein genoeg om de populatie goed te keuren bij een auditrisico van vijf procent.

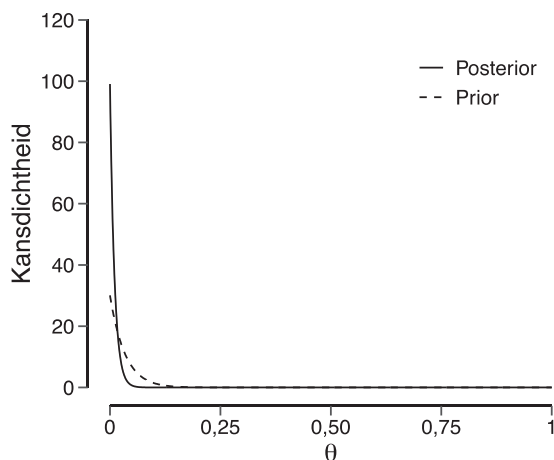
2.4. Incorporeren van een inschatting van de kans op een materiële fout

De auditor dient de prior-verdeling te onderbouwen, aangezien deze de beschikbare informatie over de foutfractie in de populatie bevat. Een voorbeeld van hoe de prior-verdeling onderbouwd kan worden is het gebruik maken van de inschatting van de kans op een materiële fout in de populatie. Ter illustratie, stel dat de auditor voordat de steekproef wordt genomen het risico op een materiële fout heeft vastgesteld met behulp van het auditrisicomodel (Vergelijking 2).

$$\text{Steekproefrisico} = \frac{\text{Auditrisico}}{\text{Inherent risico} \times \text{Intern beheersingsrisico} \times \text{Cijferanalyserisico}} = \frac{\text{Auditrisico}}{\text{RMM}} \quad (2)$$

In het voorbeeld heeft de auditor de effectiviteit van de interne beheersingsmaatregelen van de auditee beoordeeld, zoals de mate van functiescheiding en de kwaliteit van de computersystemen. De auditor heeft vastgesteld dat de interne beheersing voldoende effectief is om materiële fouten te voorkomen of te ontdekken. Omdat de interne beheersing effectief lijkt, wordt het interne beheersingsrisico als ‘laag’ beoordeeld. Volgens de auditgids van de auditor komt dit overeen met een kans van 40 procent op een materiële fout in de populatie die niet wordt opgemerkt als gevolg van de interne beheersing. De overige risico's beoordeelt de auditor als ‘hoog’. Dit betekent dat de kans op een materiële fout in de populatie (Risk of Material Misstatement, RMM in Vergelijking 2) kan worden vastgesteld op 0,4, ofwel 40 procent. In de frequentistische aanpak zou de auditor vervolgens het steekproefrisico verhogen van vijf procent naar $0,05 / 0,4 = 0,125$, ofwel 12,5 procent, en daarmee de steekproef uitvoeren. Dit resulteert in een minimale steekproefgrootte van 69 waarnemingen. Echter, in de Bayesiaanse benadering wordt de inschatting van het risico op een materiële fout direct in de prior-verdeling opgenomen.

Figuur 2. De $\text{b\eta}(1; 30,083)$ prior-verdeling en de $\text{b\eta}(1; 99,083)$ posterior-verdeling weer na het zien van 69 foutloze waarnemingen. De prior-verdeling heeft 40 procent van de kansmassa boven de uitvoeringsmaterialiteit van drie procent. Het 95^e percentiel van de posterior-verdeling ligt onder de uitvoeringsmaterialiteit van drie procent.



Bij een verwachting van nul fouten kan de informatie over de kans op een materiële fout in de populatie worden opgenomen in de prior-verdeling door de β -parameter van de prior-verdeling in te stellen op $\beta = \frac{\ln(0,4)}{\ln(1-0,03)} = 30,083$ (Derks et al. 2021a). Dit resulteert in een $\text{b\eta}(\alpha = 1, \beta = 30,083)$ prior-verdeling, die is weergegeven in Figuur 2. Onder deze prior-verdeling is de kansmassa boven de uitvoeringsmaterialiteit van drie procent gelijk aan 40 procent, wat direct te verantwoorden is door middel van de ingeschatte kans op een materiële fout van 40 procent. Met deze prior-verdeling zijn er eveneens maar 69 foutloze waarnemingen nodig om tot een posterior-verdeling te komen waarvan het 95^e percentiel onder de drie procent ligt. De inschatting van de kans op een materiële fout via

het auditrisicomodel staat dus gelijk aan een reductie van de steekproef met $98 - 69 = 29$ waarnemingen.

De integratie van deze informatie middels de prior-verdeling zorgt niet alleen voor efficiëntie, maar ook voor transparantie. Omdat zowel de risico-inschatting als het steekproefwerk relateren aan de foutfractie θ , wordt door het gebruik van de prior-verdeling namelijk duidelijk hoe de kans op een materiële fout het toekomstige werk beïnvloedt. Dat maakt de Bayesiaanse benadering bijzonder intuïtief en in lijn met de denkwijze van de auditor (Stewart 2012, p. 24). Echter, naast het onderbouwen van de prior-verdeling met informatie over de kans op een materiële fout, kan deze ook gebaseerd zijn op andere informatie die iets zegt over θ . Dit biedt vele mogelijkheden voor de auditpraktijk, omdat de auditor zo bijvoorbeeld duidelijk kan maken aan de toezichhouder hoe een uitgevoerde data-analyse die iets zegt over θ het auditwerk heeft beïnvloed. In de volgende sectie wordt besproken hoe kan worden voortgebouwd op de resultaten van verschillende vormen van data-analyse middels de Bayesiaanse statistiek.

3. Incorporeren van de resultaten van data-analyse

In dit hoofdstuk worden drie voorbeelden gepresenteerd waarin middels de Bayesiaanse statistiek wordt voortgebouwd op data-analyse. Ondanks dat het voorgaande hoofdstuk concludeerde met een methode om een prior-verdeling te formuleren aan de hand van het auditrisicomodel, is er geen standaardprocedure voor het integreren van informatie afkomstig uit data-analyse. Dit is te wijten aan het feit dat deze procedures sterk kunnen verschillen, afhankelijk van het soort informatie dat in de prior-verdeling wordt meegenomen. Het is daarom belangrijk om de data en veronderstellingen die in deze aanpak worden gebruikt, grondig te rechtvaardigen en zorgvuldig na te denken over hoe de resultaten in de prior-verdeling worden opgenomen. De volgende subsecties leggen uit hoe dit kan worden bereikt binnen de context van een regressieanalyse, een classificatieanalyse en een clusteranalyse.

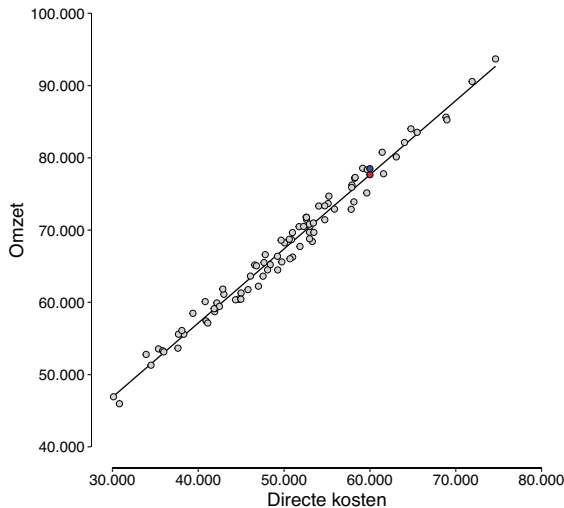
3.1. Voorbeeld 1: Regressieanalyse

De prior-verdeling kan worden opgebouwd aan de hand van de uitkomsten van een regressieanalyse (Kinney 1979; Stringer and Stewart 1986). Een concreet voorbeeld hiervan is een externe auditor die frauduleuze activiteiten met betrekking tot de omzetverantwoording wil onderzoeken. In dit geval wenst de auditor de omzetrekening van de auditee te controleren op overwaardingen. Daartoe wil de auditor eerst de relatie tussen de omzet en de directe kosten binnen de specifieke sector van de auditee analyseren, om vervolgens een uitspraak te doen over de afwijking van de verwachte omzet van de auditee.

De beschikbare informatie bestaat uit de gecontroleerde cijfers over de directe kosten C en de omzet R van $n = 85$ soortgelijke bedrijven in de sector waar de auditee

in opereert. Deze geaudite en dus betrouwbare benchmarkdata zijn weergegeven in Figuur 3. De omzet van de auditee R^* is respectievelijk € 78.500 en de som van de directe kosten C^* is € 60.000. Dit wordt aangegeven met een blauwe stip in Figuur 3. De uitvoeringsmaterialiteit voor de omzetrekening is € 2.355, oftewel drie procent van de totale waarde. Dat betekent dat de omzet met maximaal drie procent te hoog opgevoerd mag zijn.

Figuur 3. De verhouding tussen de directe kosten en de omzet van 85 bedrijven uit dezelfde sector als de auditee. De rode stip staat voor de voorspelde omzet op basis van de benchmarkdata, terwijl de blauwe stip de daadwerkelijke omzet van de auditee weergeeft.



De relatie tussen de omzet R en de directe kosten C kan worden gemodelleerd met een lineaire vergelijking (Vergelijking 3). In Vergelijking 3 stelt β_0 de gemiddelde omzet van de bedrijven bij geen directe kosten voor en representeert β_1 de stijging in de omzet als de directe kosten met één stijgen. De term ϵ stelt de residuen voor, waarvan de auditor de standaardveronderstelling maakt dat deze normaal verdeeld zijn met gemiddelde nul en standaarddeviatie σ_ϵ . Merk op dat er in dit voorbeeld een versimpeld model wordt gebruikt. In de praktijk is deze relatie vaak complexer dan hieronder is weergegeven en moet de auditor het regressiemodel zorgvuldig opstellen en evalueren.

$$R = \beta_0 + \beta_1 \times C + \epsilon \quad (3)$$

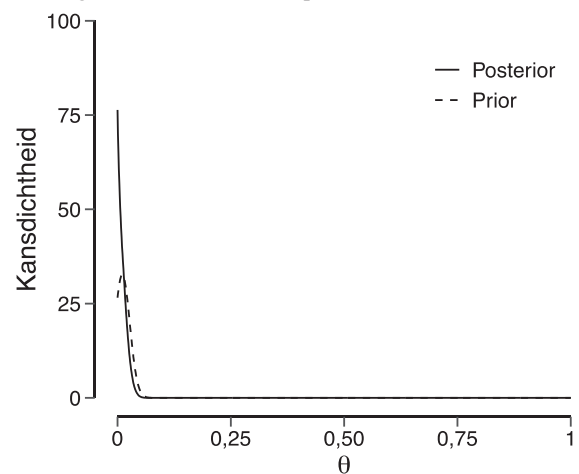
De auditor schat de parameters in Vergelijking 3 met behulp van de klassieke methode voor Bayesiaanse regressie middels oneigenlijke³ uniforme prior-verdelingen voor β_0 , β_1 en $\log(\sigma_\epsilon)$ (Gelman et al. 2023, pp. 354–357), welke dezelfde uitkomsten geeft als frequentistische regressie middels de kleinste-kwadratenmethode. Dit leidt tot de meest waarschijnlijke schattingen $\hat{\beta}_0 = € 16.089,27$, $\hat{\beta}_1 = 1,026$ en $\hat{\sigma}_\epsilon = € 1.314,37$. Invullen van de meest waarschijnlijke schattingen voor de coëfficiënten β_0 en β_1 in Vergelijking 3 geeft de meest waarschijnlijke schatting \hat{R} van de verwachte omzet van de auditee, weergegeven in Vergelijking 4.⁴ Deze meest waarschijnlijke schatting van de verwachte omzet is weergegeven met een rode stip in Figuur 3.

$$\hat{R} = € 16.089,27 + 1,026 \times € 60.000 = € 77.649,27 \quad (4)$$

Uitgaande van de veronderstelling dat de data representatief zijn voor de auditee, kan de auditor middels het regressiemodel in Vergelijking 3 de geaudite waarde van de omzet voorspellen en van hieruit een prior-verdeling opstellen voor de fractie overwaardering θ . Bij gebruik van uniforme prior-verdelingen voor de parameters is de verdeling die de onzekerheid van de verwachte omzet weergeeft een geschaalde t -verdeling met gemiddelde $\mu^* = \hat{R}$, standaardafwijking $\sigma^* = \sqrt{\sigma_\epsilon^2 \left(1 + \frac{1}{n} + \frac{(C^* - \bar{C})^2}{\sum(C - \bar{C})^2}\right)}$ en $n - 2$ vrijheidsgraden. De prior-verdeling voor de fractie overwaardering θ kan worden gespecificeerd als deze verdeling, die de onzekerheid van de voorspelling van de omzet weergeeft, uitgedrukt als een percentage van de omzet. Voor de duidelijkheid kiezen we hier voor een benadering van deze prior-verdeling, hoewel er meer specifieke methoden bestaan om deze prior-verdeling te definiëren (e.g., Kinney 1979).

Het gemiddelde van de prior-verdeling is de relatieve afwijking van de omzet in vergelijking met de meest waarschijnlijke verwachte omzet volgens de benchmarkdata, uitgedrukt als $\mu_\theta = \frac{R - \hat{R}}{R} = 0,011$, oftewel 1,1 procent. De standaardafwijking van de prior-verdeling kan worden berekend als $\sigma_\theta = \frac{\sigma^*}{R} = 0,017$, oftewel 1,7 procent. Aangezien de auditor zich uitsluitend richt op overwaarderingen, wordt de prior-verdeling afgekapt op het interval $[0; 1]$. Dit resulteert in de getrunceerde geschaalde t -verdeling met gemiddelde $\mu_\theta = 0,011$, standaardafwijking $\sigma_\theta = 0,017$, en 83 vrijheidsgraden als prior-verdeling voor θ , zoals weergegeven in Figuur 4.

Figuur 4. De getrunceerde geschaalde t prior-verdeling met parameters $\mu = 0,011$, $\sigma = 0,017$ en 83 vrijheidsgraden en de posterior-verdeling na het zien van 54 foutloze waarnemingen. Het 95^e percentiel van de posterior-verdeling ligt onder de uitvoeringsmaterialiteit van drie procent.



Intuïtief gezien heeft de auditor meer auditcomfort wanneer de omzet dicht bij de verwachte omzet ligt. Door de relatieve afwijking van de verwachte omzet op deze manier mee te nemen in de prior-verdeling, vertaalt

de auditor op een statistisch verantwoorde manier dit extra auditcomfort naar een kleinere steekproef. Ter illustratie, door gebruik te maken van deze prior-verdeling zijn er nog slechts 54 foutloze waarnemingen nodig in plaats van 98 om tot een posterior-verdeling te komen waarvan het 95^e percentiel onder de uitvoeringsmaterialiteit van drie procent ligt. Echter, de uitkomsten van deze regressieanalyse kunnen ook gecombineerd worden met de inschatting van een materiële fout middels het auditrisicomodel zoals beschreven in hoofdstuk 2.4. Doordat de auditor kan steunen op de interne beheersing kunnen de 29 waarnemingen die impliciet als correct zijn aangenomen van de vereiste steekproefgrootte worden afgetrokken. Dat betekent dat er nog slechts $54 - 29 = 25$ foutloze waarnemingen nodig zijn om voldoende zekerheid te krijgen over de populatie.

Voor dit voorbeeld, evenals de volgende twee voorbeelden, kan het gebeuren dat de auditor fouten ontdekt in de gereduceerde steekproef, waardoor de vereiste zekerheid niet wordt behaald. De auditor kan hier op drie manieren mee omgaan:

- 1) De auditor kan de steekproef uitbreiden tot de vereiste zekerheid van 95 procent om de populatie goed te keuren is bereikt, of totdat er genoeg zekerheid is behaald om de populatie af te keuren.
- 2) De auditor kan de populatie noch goedkeuren noch afkeuren (dat wil zeggen, oordeelonthouding) met vermelding van de resterende kans op een (niet-) materiële fout als de kansmassa onder de posterior-verdeling aan de linker- en rechterkant van de uitvoeringsmaterialiteit.
- 3) De auditor kan de posterior-verdeling gebruiken als de nieuwe prior-verdeling voor het uitvoeren van aanvullende auditactiviteiten. Denk hierbij bijvoorbeeld aan het implementeren van een extra analytische procedure of het testen van andere interne beheersingsmaatregelen. De nieuwe prior-verdeling is in dat geval een zo volledig mogelijke samenvatting van de kennis van de auditor met betrekking tot het te onderzoeken populatiekenmerk voorafgaand aan deze aanvullende

activiteiten. Door deze nieuwe prior-verdeling bij te werken met de informatie verkregen uit deze activiteiten, kan de auditor de behaalde extra zekerheid kwantificeren en zo mogelijk tot goedkeuring komen.

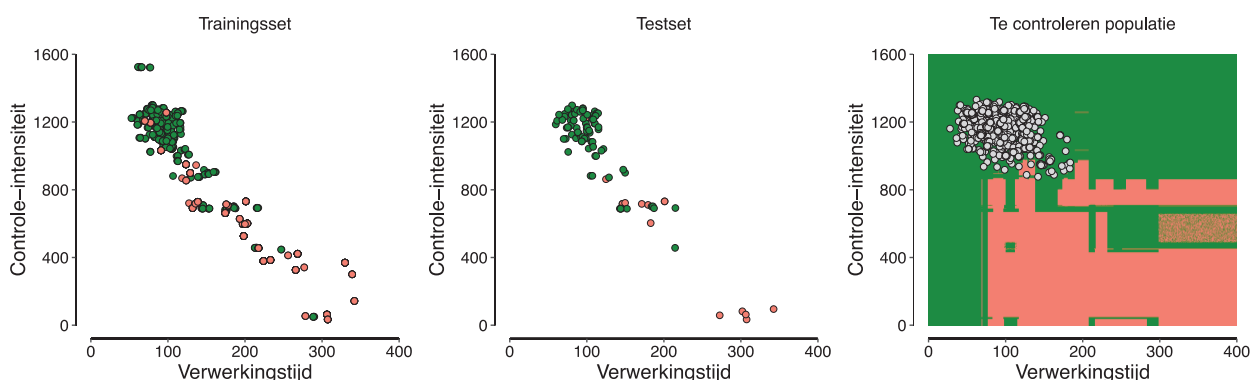
Het is aan de auditor om te beoordelen welke manier de voorkeur heeft.

3.2. Voorbeeld 2: Classificatieanalyse

De prior-verdeling kan eveneens worden opgesteld aan de hand van de uitkomsten van een classificatieanalyse. In dit voorbeeld wordt een situatie besproken waarin een externe auditor de taak heeft om een populatie facturen te controleren op fouten. Gedurende het jaar heeft een interne auditor wekelijks een steekproef van tien facturen gecontroleerd. De interne auditor heeft hierbij expliciet geprobeerd fouten te ontdekken door relatief veel facturen met een hoog risico te controleren. Deze facturen representeren echter maar een klein deel van de populatie. De externe auditor wil voortbouwen op het werk van de interne auditor om zo de kosten voor de audit te minimaliseren. Er zijn verschillende manieren om het werk gedaan door de interne auditor te verwerken in een prior-verdeling (e.g., Raats and Moors 2003). In dit voorbeeld kiezen we ervoor om de prior-verdeling op te stellen middels het trainen en toepassen van een machine learning-algoritme. De toelaatbare fout voor de populatie facturen is vastgesteld op drie procent, wat betekent dat maximaal drie procent van de facturen een fout mag bevatten.

In dit scenario heeft de externe auditor toegang tot data over de controle-intensiteit en verwerkingstijd van de facturen die door de interne auditor gecontroleerd zijn. Met behulp van een classificatieanalyse kan de auditor voor het huidige jaar een voorspelling doen over de correctheid van elke factuur in de nog te controleren populatie, gebaseerd op de controle-intensiteit en de verwerkingstijd. Deze twee kenmerken zijn voor de facturen in de te controleren populatie ($n = 976$) weergegeven in het rechter paneel in Figuur 5.

Figuur 5. De verwerkingstijd van de facturen tegenover de controle-intensiteit voor de ongeziene populatie (rechter paneel), de gebalanceerde trainingsset (linker paneel) en gebalanceerde testset (middelste paneel). In de trainings- en testset geven kleuren aan of de facturen fout (roze) of correct (groen) waren. Dezelfde kleuren worden in het rechter paneel gebruikt om de voorspellingen van het algoritme te visualiseren.



De belangrijkste aanname van een classificatieanalyse is het bestaan van een relatie, dat wil zeggen een functie f , tussen de verwerkingstijd V , de controle-intensiteit I en de uitkomst Y (een factuur die “Fout” of “Niet fout” is), zoals aangegeven in Vergelijking 5. In dit voorbeeld wordt opnieuw een vereenvoudigd model gebruikt, maar merk op dat deze relatie in de praktijk vaak complexer is dan hier wordt voorgesteld.

$$Y = f(V, I) + \epsilon \quad (5)$$

Indien de functie f bekend is, kan een voorspelling over de juistheid van een ongeziene factuur worden gemaakt op basis van de controle-intensiteit en de verwerkingstijd. Het is echter niet bekend hoe de kenmerken gerelateerd zijn aan de correctheid van een factuur, dus f is onbekend en deze functie moet uit de data geleerd worden. Hier kunnen verschillende soorten algoritmes voor worden gebruikt. In dit geval wordt het *random forest*-algoritme (James et al. 2023, p. 343–345) gehanteerd, omdat dit uiteindelijk een kansuitspraak geeft op een classificatie van een factuur als fout of niet fout.

Om de functie f te leren, wordt het algoritme een deel van de beschikbare data getoond – de zogenoemde trainingsset ($n = 420$) – die zowel de kenmerken als de doelvariabele bevat, dat wil zeggen het door de interne auditor waargenomen resultaat “Fout” en “Niet fout”. Een mogelijk probleem bij het toepassen van classificatie in deze context is dat er waarschijnlijk meer correcte facturen in de beschikbare data voorkomen dan foute facturen. Als deze ongelijke verdeling in de trainingsset aanwezig is, kan het ertoe leiden dat het algoritme de voorkeur geeft aan de correcte facturen, simpelweg omdat deze vaker voorkomen in de trainingsset. Dit probleem kan worden aangepakt door een trainingsset te creëren waarin de verhouding tussen correcte facturen en foute facturen gelijk is. In dit specifieke geval heeft de externe auditor een aantal foute facturen uit de beschikbare data meerdere malen willekeurig geselecteerd, waardoor de verhouding in de trainingsset evenwichtig is. Dit heet *oversampling* en is een robuuste methode om de informatie in de trainingsset over de foute facturen te vergroten, terwijl het aantal foute en correcte facturen gelijk blijft (He and Garcia 2009). De data uit deze gebalanceerde trainingsset worden weergegeven in het linker paneel van Figuur 5, waarbij de kleur aangeeft of de factuur een fout bevat (roze) of correct is (groen).

Om te testen of de geleerde functie f goed generaliseert naar data die het algoritme nog niet eerder heeft gezien, krijgt het algoritme in de leerfase niet de zogenoemde testset te zien. Deze testset bestaat wederom uit een deel ($n = 100$) van de beschikbare data (middelste paneel in Figuur 5). De kwaliteit van het algoritme wordt op de testset geëvalueerd door de daadwerkelijke classificatie van de facturen te vergelijken met de voorspelde classificatie. De tabel die hieruit volgt (Tabel 1) wordt ook wel een verwarringsmatrix genoemd.

Tabel 1. De verwarringsmatrix voor de 100 facturen in de testset. De werkelijke foutclassificatie wordt vergeleken met de voorspelde foutclassificatie volgens het algoritme. De vier cellen tonen de echt positieven (TP), de foutpositieven (FP), de foutnegatieven (FN) en de echt negatieven (TN).

		Werkelijk	
		Fout	Niet fout
Voorspeld	Fout	11 (TP)	1 (FP)
	Niet fout	4 (FN)	84 (TN)

Het is gebruikelijk om de minst voorkomende uitkomst te voorzien van het label “positief”. De meest voorkomende uitkomst heet dan “negatief”. In dit voorbeeld zijn de correcte facturen dus de negatieven, en de foute facturen de positieven. De verwarringsmatrix omvat de echt positieven, de foutpositieven, de echt negatieven en de foutnegatieven. De echt positieven (TP) zijn de 11 facturen in de testset die daadwerkelijk fout waren en door het algoritme correct als zodanig zijn voorspeld. De foutpositief (FP) is de factuur in de testset die in werkelijkheid correct was, maar door het algoritme onjuist als fout is voorspeld. De echt negatieven (TN) zijn de 84 facturen in de testset die daadwerkelijk correct waren en door het algoritme correct als zodanig zijn voorspeld. Ten slotte zijn de foutnegatieven (FN) de 4 facturen in de testset die daadwerkelijk fout waren, maar door het algoritme onjuist als correct zijn voorspeld.

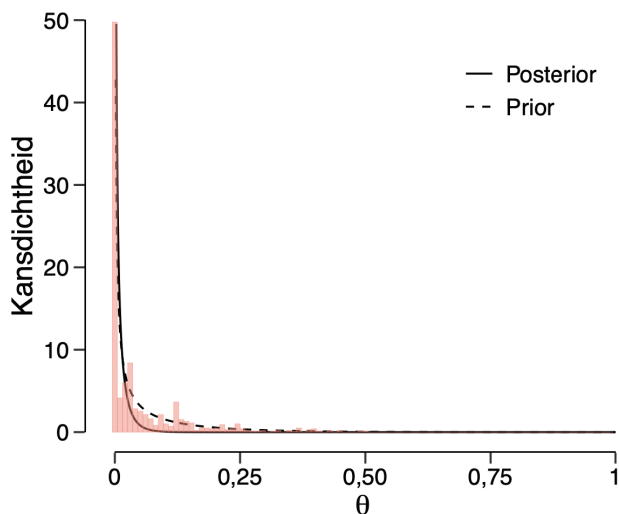
De verwarringsmatrix maakt het mogelijk om verschillende maatstaven te berekenen voor de kwaliteit van het algoritme. Allereerst is de nauwkeurigheid (*accuracy*) van het algoritme gelijk aan $\frac{TP+TN}{TP+FN+TN+FP} = 0,95$. Daarnaast is de precisie (*precision*) van het algoritme gelijk aan $\frac{TP}{TP+FP} = 0,917$. Bovendien is de gevoeligheid (*recall*) van het algoritme gelijk aan $\frac{TP}{TP+FN} = 0,733$. Ten slotte kan de *F1-score* worden berekend als $\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = 0,815$.

Uit deze maatstaven concludeert de auditor dat het algoritme voldoende presteert om te worden ingezet voor het voorspellen van fouten in de nog te controleren populatie.

Na de evaluatie van de kwaliteit van het algoritme kan er voor de 976 facturen in de te controleren populatie een voorspelling worden gemaakt voor de foutkans van elke factuur. De 7 facturen die door het algoritme een voorspelde foutkans krijgen toegewezen die groter is dan $1/2$ worden als fout bestempeld en worden vervolgens door de auditor integraal gecontroleerd. Op de 969 overgebleven facturen met een voorspelde foutkans lager dan $1/2$ wil de auditor een steekproef doen. De frequentieverdeling van de voorspelde foutkansen van deze facturen wordt getoond in Figuur 6. Door aan te nemen dat deze voorspelde foutkansen indicatief zijn voor de ware foutkansen van de facturen, kan de auditor de prior-verdeling voor de foutfractie θ in de populatie opstellen. Door het bepalen van het gemiddelde $\bar{x} = 0,047$ en de variantie $s^2 = 0,007$ van deze voorspelde foutkansen kan er via de momentmethode een bèta-verdeling worden

toegepast met parameters $\alpha = \left(\frac{\bar{x}(1-\bar{x})}{s^2} - 1\right) \bar{x} = 0,262$ en $\beta = \left(\frac{\bar{x}(1-\bar{x})}{s^2} - 1\right) (1 - \bar{x}) = 5,267$.⁵ Dit resulteert in de bèta(0,262; 5,267) verdeling als prior-verdeling voor θ , zoals weergegeven in Figuur 6.

Figuur 6. De bèta(0,262; 5,267) prior-verdeling op basis van de voorspelde foutkansen per factuur in de populatie (histogram) en de bèta(0,262; 42,267) posterior-verdeling na het zien van 37 foutloze waarnemingen. Het 95^e percentiel van de posterior-verdeling ligt onder het toelaatbare foutpercentage van drie procent.



Met deze prior-verdeling is het slechts nodig om een steekproef van 37 foutloze waarnemingen te zien in plaats van 98 om een posterior-verdeling te bereiken waarvan het 95^e percentiel onder het toelaatbare foutpercentage van drie procent ligt. Echter, bij het combineren van de uitkomsten van deze classificatieanalyse met de inschatting van een materiële fout middels het auditrisicomodel zoals beschreven in hoofdstuk 2.4, zijn er nog slechts $37 - 29 = 8$ foutloze waarnemingen nodig om voldoende zekerheid te krijgen over de overgebleven populatie facturen met een voorspelde foutkans lager dan $1/2$.

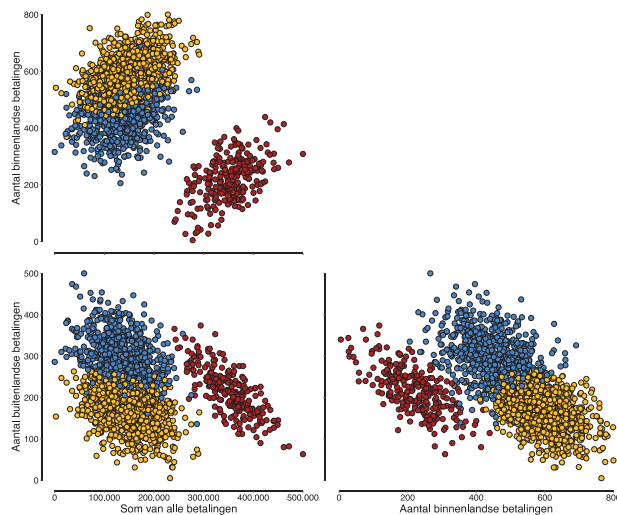
3.3. Voorbeeld 3: Clusteranalyse

In het laatste voorbeeld wordt clusteranalyse toegepast om een populatie te stratificeren en deze vervolgens efficiënt te analyseren met gebruik van de Bayesiaanse statistiek. Dit voorbeeld betreft een auditor die bij een toezichthoudende instantie werkt en de taak heeft om de risicoclassificatie van klanten door een financiële instelling te onderzoeken. Financiële instellingen geven al hun klanten een integriteits-risicoscore, en het doel van de auditor is om te controleren of de financiële instelling niet een te lage risicoscore aan de rekeninghouder heeft toegekend. De toelaatbare fout is in dit voorbeeld drie procent, wat betekent dat aan maximaal drie procent van de rekeninghouders een te lage risicoscore toegekend mag zijn.

De beschikbare data omvatten het aantal binnenlandse betalingen per rekeninghouder, het aantal buitenlandse betalingen, en de som van alle betalingen. In dit scenario

past de auditor een clusteranalyse toe om de populatie te verdelen in verschillende clusters op basis van deze drie kenmerken, en zodoende het auditwerk te focussen op rekeninghouders die een hoger risico op fraude vertonen (Thiprungsri and Vasarhelyi 2011). De auditor maakt voor het clusteren gebruik van een Gaussisch mixture-model (Fraley and Raftery 2002; Fraley and Raftery 2007; Scrucca et al. 2016). Met behulp van het Bayesian Information Criterion (BIC) bepaalt de auditor dat een model met drie clusters het beste op de data past. Figuur 7 toont de data en de drie clusters.

Figuur 7. De som van de betalingen van de rekeningen in de te controleren populatie tegenover het aantal binnenlandse en buitenlandse betalingen gemaakt met deze rekeningen. De kleuren van de rekeningen zijn bepaald door een clusteranalyse met drie clusters. De eerste en grootste cluster (geel) bestaat uit 1053 rekeningen, de tweede cluster (blauw) bestaat uit 697 rekeningen en de derde en kleinste cluster (rood) bestaat uit 250 rekeningen.

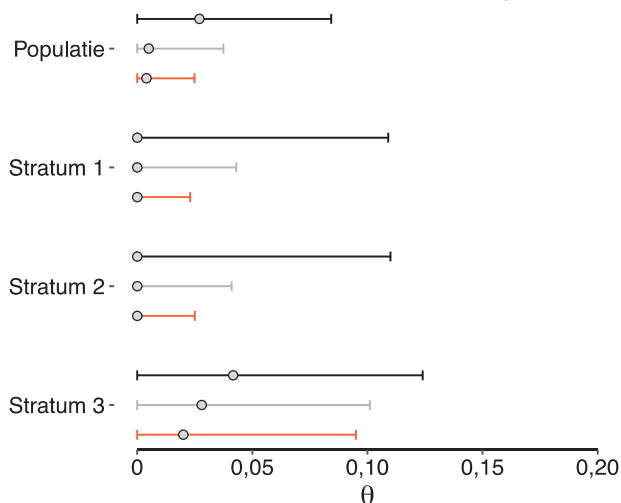


Het eerste en grootste cluster (geel) bestaat uit 1053 rekeningen met relatief veel binnenlandse betalingen ten opzichte van buitenlandse betalingen en een totaalbedrag lager dan € 250.000. Het tweede cluster (blauw) bestaat uit 697 rekeningen met relatief veel buitenlandse betalingen ten opzichte van binnenlandse betalingen en een totaalbedrag lager dan € 250.000. Het derde en kleinste cluster (rood) bestaat uit 250 rekeningen met relatief weinig binnenlandse en buitenlandse betalingen ten opzichte van de andere twee clusters en een totaalbedrag hoger dan € 250.000. De auditor besluit de minimale steekproef van 98 waarnemingen die nodig is bij een toelaatbare fout van drie procent over de drie clusters te verdelen. Daarbij wenst de auditor het auditwerk te focussen op het derde cluster, omdat deze rekeningen volgens de auditor een relatief hoog frauderisico hebben. Daarom neemt de auditor in zowel het eerste als tweede cluster een steekproef van 25 waarnemingen, waarbij er in beide steekproeven geen te laag toegekende risico-integriteitsscores worden gevonden. In het derde cluster neemt de auditor een steekproef van 48 waarnemingen. Hieruit blijkt dat bij twee

rekeningen de risico-integriteitsscore door de financiële instelling te laag is toegekend. De auditor wenst op basis van de resultaten van deze gestratificeerde steekproef een 95 procent zekere bovengrens voor de foutfractie (d.w.z. de fractie rekeninghouders met een te laag toegekende risico-integriteitsscore) in de populatie te berekenen.

De gebruikelijke manier om een gestratificeerde steekproef statistisch te evalueren is om de foutfracties in de drie clusters (i.e., strata) als onafhankelijk te beschouwen. Vanuit een Bayesiaans perspectief houdt dit in dat de foutfractie in elk cluster, θ_s , een onafhankelijke prior-verdeling toegewezen krijgt, in dit voorbeeld de $\text{b\^eta}(1, 1)$ prior-verdeling. Om tot een representatieve schatting van de foutfractie θ op populatieniveau te komen, weegt de auditor de posterior-verdelingen naar het relatieve aandeel van de clusters in de populatie. Het beschouwen van de foutfracties in de clusters als onafhankelijk is echter niet optimaal, omdat het onwaarschijnlijk is dat de fracties van te lage risico-integriteitsscore in de drie clusters geheel onafhankelijk van elkaar zijn (Efron and Morris 1977), bijvoorbeeld omdat alle risico-integriteitsscores door hetzelfde KYC-team zijn toegekend. Een aantrekkelijkere manier van evalueren is om deze data te analyseren onder de veronderstelling dat de foutfracties in de clusters in zekere mate op elkaar lijken. Dit kan worden gedaan door gebruik te maken van een Bayesiaans hiërarchisch model, met name Model 3 uit Derks et al. (2022b). Statistisch gezien betekent dit dat elke θ_s een $\text{b\^eta}(v\phi, v(1 - \phi))$ prior-verdeling toegewezen krijgt, waarbij de parameters van deze prior-verdeling zelf een prior-verdeling toegewezen krijgen. In dit voorbeeld maakt de auditor gebruik van een $\text{b\^eta}(1, 1)$ prior-verdeling voor ϕ en een $\text{Pareto}(1, 3/2)$ prior-verdeling voor v (Carpenter 2016). Figuur 8 laat voor beide modellen de schattingen van de foutfracties zien in de populatie en in de drie clusters.

Figuur 8. Eenzijdige schattingsintervallen en meest waarschijnlijke schattingen van de foutfracties θ voor het onafhankelijke model (zwart), het hiërarchische model (grijs) en het hiërarchische model met informatie uit het auditrisicomodel (oranje). Onder het hiërarchische model zijn de schattingen van de foutfracties efficiënter dan onder het onafhankelijke model.



Onder het onafhankelijke model is het 95^e percentiel van de posterior-verdeling van θ gelijk aan 0,0842, ofwel 8,42 procent.⁶ Daarentegen is het 95^e percentiel van de posterior-verdeling van θ in het hiërarchische model gelijk aan 0,0378, ofwel 3,78 procent. Met het hiërarchische model heeft de auditor dus meer zekerheid verkregen uit dezelfde data. De schatting van de fractie rekeninghouders met een te lage risico-integriteitsscore in de populatie bij gebruik van het hiërarchische model is daarom efficiënter dan de schatting bij gebruik van het onafhankelijke model. Echter, bij het samenvoegen van de resultaten van de clusteranalyse met de schatting van een materiële fout via het auditrisicomodel zoals beschreven in hoofdstuk 2.4, kan de prior voor ϕ worden ingesteld op de eerder opgestelde $\text{b\^eta}(1; 30,083)$ prior-verdeling. Door deze prior-verdeling voor ϕ in het hiërarchische model te gebruiken is het 95^e percentiel van de posterior-verdeling van θ gelijk aan 0,0251, ofwel 2,51 procent. Deze 95 procent zekere bovengrens is laag genoeg om de populatie goed te keuren bij een auditrisico van vijf procent.

4. Conclusies en aanbevelingen

In dit artikel is beargumenteerd dat Bayesiaanse statistiek auditors helpt om de resultaten van data-analyse in de audit te integreren. Aan de hand van drie voorbeelden – regressie, classificatie en clustering – is geïllustreerd hoe de integratie van data-analyse op deze manier de efficiëntie van de audit kan verbeteren.⁷ Hoewel de voorbeelden in het artikel vereenvoudigd zijn en praktijkcasussen vaak complexer, blijft het basisprincipe overeind: door Bayesiaanse statistiek te gebruiken, kunnen auditors de impact van data-analyse op vervolgvactiteiten en eindbeoordelingen duidelijk inzichtelijk maken. Dit vergemakkelijkt het volgen van de stroom van informatie voor toezichthouders en andere belanghebbenden. We adviseren auditors daarom aan de slag te gaan met het integreren van data-analyse middels de Bayesiaanse statistiek, om de praktische toepasbaarheid te evalueren en eventuele uitdagingen te identificeren.

Om het voor auditors zo makkelijk mogelijk te maken om de in dit artikel behandelde technieken in de praktijk toe te passen, bieden we een aantal praktische handvatten om hiermee aan de slag te gaan. Zo zijn de benodigde datasets en de R code voor het reproduceren van de drie voorbeelden in dit artikel beschikbaar in de online bijlage op <https://osf.io/2qawp/>. Daarnaast hoeven de Bayesiaanse berekeningen niet door de auditor zelf te worden gedaan. Zodra de prior-verdeling is vastgesteld, kunnen vervolgberekeningen, zoals het bepalen van de posterior-verdeling, worden uitgevoerd met gebruiksvriendelijke, open-source software zoals JASP for Audit (Derks et al. 2021b). Door deze technieken gemakkelijk voor iedereen beschikbaar te maken, hopen we de integratie van deze methodes in de audit te bevorderen.

In de praktijk is het belangrijk om de voor- en nadelen van de Bayesiaanse aanpak goed af te wegen. De potentiële reductie van de steekproefgrootte gaat

gepaard met de verantwoordelijkheid om aan te tonen dat de data-analyse geldig en relevant is voor de steekproef en dat de vertaling van de resultaten naar de prior-verdeling passend is. Of de voordelen van het integreren van data-analyse opwegen tegen deze nadelen hangt af van de tijd en moeite die het kost om een grotere steekproef te selecteren en te controleren. Als deze kosten de tijd en moeite om de data-analyse uit te voeren en daarmee een prior-verdeling op te zetten overschrijden, is het goedkoper om de data-analyse te integreren middels de prior-verdeling. Natuurlijk is het opstellen van een prior-verdeling op basis van data-analyse geen triviale taak en kan het aanzienlijke tijd en moeite kosten. Echter, omdat het controleren van steekproefwaarnemingen vaak veel tijd en geld kost, is het waarschijnlijk dat de mogelijke vermindering van de steekproefgrootte die wordt bereikt door de informatie uit de data-analyse te integreren opweegt tegen de tijd en moeite die gaat naar het specificeren van de prior-verdeling. Zelfs als de kosten voor een grotere steekproef klein zijn en de auditor daarom besluit dat het opstellen van de prior-verdeling de tijd en moeite niet waard is, kan deze terugvallen op een

prior-verdeling die geen bestaande informatie bevat. De Bayesiaanse benadering biedt de auditor de flexibiliteit om één van deze opties te kiezen.

Al met al biedt het Bayesiaanse raamwerk een robuuste basis voor de toekomst. Sinds de opkomst van de risicogebaseerde audit is er een verschuiving naar het verkrijgen van zekerheid uit andere activiteiten dan detailcontroles. Het toenemende gebruik van data-analyse markeert slechts een volgende stap in deze ontwikkeling. In combinatie met de groeiende complexiteit van data betekent dit dat auditors een intuïtief raamwerk zullen moeten hebben om de informatie uit deze procedures te integreren, kwantificeren en interpreteren. Dit zal vooral het geval zijn als zij willen voldoen aan de constante vraag naar een efficiëntere audit. Aangezien het Bayesiaanse raamwerk de flexibiliteit biedt om verschillende soorten data-analyse in de statistische analyse te integreren, beweren wij dat dit raamwerk op de lange termijn nuttiger zal zijn voor de auditor dan de huidige frequentistische methoden. Het omarmen van dergelijke innovatieve technieken vereist lef van auditors, maar wordt door allerlei partijen aangemoedigd (NBA 2019; Kwartiermakers toekomst accountancy 2023, p. 71).

-
- **Dr. K. P. Derks – Koen** is universitair docent bij Nyenrode Business Universiteit en lid van de stuurgroep statistical auditing van het Limperg Instituut. Daarnaast is hij softwareontwikkelaar bij JASP, een gratis en open-source statistiekprogramma, waarin hij de module voor auditing ontwikkelt.
 - **L. Mensink MSc – Lotte** is promovenda bij Nyenrode Business Universiteit waar zij onderzoek doet naar de toepassing van Bayesiaanse statistiek in de audit.
 - **Prof. dr. J. J. B. de Swart MBA – Jacques** is partner bij PwC Consulting, verantwoordelijk voor de data analytics groep. Daarnaast is hij hoogleraar Toegepaste Wiskunde aan Nyenrode Business Universiteit en voorzitter van de stuurgroep statistical auditing van het Limperg Instituut.
 - **Prof. dr. R. Wetzels – Ruud** is director Data Analytics bij PwC Consulting. Daarnaast is hij hoogleraar Data Science bij Nyenrode Business Universiteit en bestuurslid van stichting JASP.
-

Noten

1. Andersom heeft Blokdijs (1995) aangetoond dat een gegevensgerichte controle ook niet zonder het beschouwen van de interne controle kan.
2. Hoewel dit een monetaire steekproef is, gaan we er in dit voorbeeld van uit dat een boeking óf volledig fout is óf volledig correct. Er bestaan echter ook methoden om gedeeltelijke fouten te evalueren, zoals de Stringer bound (Stringer 1963).
3. Een oneigenlijke prior-verdeling integreert niet tot 1 en is daarom geen geldige kansverdeling. In dit voorbeeld verlenen de uniforme prior-verdelingen dezelfde plausibiliteit aan alle waarden van β_0 , β_1 en $\log(\sigma)$ tussen min oneindig en oneindig. Deze verdelingen zijn oneigenlijk omdat ze integreren tot oneindig. Echter, na het zien van de data zijn de posterior-verdelingen in dit geval niet meer oneigenlijk en zijn het daarom geldige kansverdelingen.
4. De coëfficiënten in Vergelijking 4 zijn afgerond ter illustratie, maar de verwachte omzet is berekend op basis van de niet-afgeronde coëfficiënten.
5. De α -en β -parameters van de bèta-verdeling zijn gebaseerd op de niet-afgeronde gemiddelde en variantie.
6. Het is belangrijk om te benadrukken dat, anders dan in eerdere voorbeelden, de bovengrens in dit geval is vastgesteld op basis van een steekproef waarin twee fouten zijn gevonden, niet nul. Als de auditor deze steekproef zou analyseren zonder stratificatie, zou de bovengrens 6,2 procent zijn; dit is lager dan de 8,42 procent, maar dat is te wijten aan het feit dat de aanname van onafhankelijkheid tussen de strata nogal conservatief is.
7. Belangrijk om te vermelden hierbij is dat het meenemen van informatie uit data-analyse geen garantie biedt voor een efficiëntere steekproeffase. Als de informatie uit data-analyse wijst op een hoog risico op materiële fouten, is het vanzelfsprekend dat de auditor meer werk moet doen in de steekproeffase om nog goed te kunnen keuren. De data-analyses in de voorbeelden in dit artikel bevatten steeds informatie die wijst op een relatief laag risico op materiële fouten.

Literatuur

- Appelbaum D, Kogan A, Vasarhelyi M A (2017) Big Data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory* 36(4): 1–27. <https://doi.org/10.2308/ajpt-51684>
- Bayarri MJ, Berger JO (2004) The interplay of Bayesian and frequentist analysis. *Statistical Science* 19(1): 59–80. <https://doi.org/10.1214/08834230400000116>
- Blokdijk H (1995) De betekenis van onvervangbare interne controle voor de accountantscontrole. *Maandblad Voor Accountancy en Bedrijfseconomie* 69(10): 588–595. <https://doi.org/10.5117/mab.69.11660>
- Carpenter B (2016) Hierarchical partial pooling for repeated binary trials. [online] Mc-stan.org. <https://mc-stan.org/users/documentation/case-studies/pool-binary-trials.html> [Accessed 28 February 2024]
- Chesley GR (1975) Elicitation of subjective probabilities: A review. *The Accounting Review* 50(2): 325–337. <https://www.jstor.org/stable/244714>
- Christensen B, Elder R, Glover S (2015) Behind the numbers: Insights into large audit firm sampling policies. *Accounting Horizons* 29(1): 61–82. <https://doi.org/10.2308/acch-50921>
- Crosby MA (1981) Bayesian statistics in auditing: A comparison of probability elicitation techniques. *The Accounting Review* 56(2): 355–365. <https://www.jstor.org/stable/245818>
- Derks K, De Swart J, Van Batenburg P, Wagenmakers E-J, Wetzels R (2021a) Priors in a Bayesian audit: How integration of existing information into the prior distribution can improve audit transparency and efficiency. *International Journal of Auditing* 25(3): 621–636. <https://doi.org/10.1111/ijau.12240>
- Derks K, De Swart J, Wagenmakers E-J, Wille J, Wetzels R (2021b) JASP for Audit: Bayesian tools for the auditing practice. *Journal of Open Source Software* 6(68): 2733. <https://doi.org/10.21105/joss.02733>
- Derks K, De Swart J, Wagenmakers E-J, Wetzels R (2022a) An impartial Bayesian hypothesis test for audit sampling. *PsyArXiv*. <https://doi.org/10.31234/osf.io/8nf3e>
- Derks K, De Swart J, Wetzels R (2022b) Een Bayesiaanse blik op gestratificeerde steekproeven heeft voordelen voor de auditor. *Maandblad voor Accountancy en Bedrijfseconomie* 96(1/2): 37–46. <https://doi.org/10.5117/mab.96.78836>
- Edwards AWF (1992) *Likelihood*. Johns Hopkins University Press (Baltimore, USA).
- Efron B, Morris C (1977) Stein’s paradox in statistics. *Scientific American* 236(5): 119–127. <https://doi.org/10.1038/scientificamerican0577-119>
- Etz A (2018) Introduction to the concept of likelihood and its applications. *Advances in Methods and Practices in Psychological Science* 1(1): 60–69. <https://doi.org/10.1177/2515245917744314>
- Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association* 97(458): 611–631. <https://doi.org/10.1198/016214502760047131>
- Fraley C, Raftery AE (2007) Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classification* 24: 155–181. <https://doi.org/10.1007/s00357-007-0004-5>
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2021) *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.
- Gepp A, Linnenluecke MK, O’Neill TJ, Smith T (2018) Big data techniques in auditing research and practice: Current trends and future opportunities. *Journal of Accounting Literature* 40(1): 102–115. <https://doi.org/10.1016/j.acclit.2017.05.003>
- He H, Garcia EA (2009) Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering* 21(9): 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- ING (2018) Using AI to assess credit risk. [online] <https://www.ing.com/Newsroom/News/Using-AI-to-assess-credit-risk.htm> [Accessed 26 March 2024]
- James G, Witten D, Hastie T, Tibshirani R, Taylor J (2023) *An introduction to statistical learning with applications in R*. Springer (New York, USA). <https://doi.org/10.1007/978-3-031-38747-0>
- Jans M, Hosseinpour M (2017) Process mining-technieken voor internecontroletoetsen - mogelijkheden nu en in de toekomst. *Maandblad Voor Accountancy en Bedrijfseconomie* 91: 90–95. <https://doi.org/10.5117/mab.91.24025>
- Kinney WR (1979) Integrating audit tests: Regression analysis and partitioned dollar-unit sampling. *Journal of Accounting Research* 17(2): 456–475. <https://doi.org/10.2307/2490513>
- Kloosterman HHW (2004a) Risicoanalyse in auditing. *Maandblad Voor Accountancy en Bedrijfseconomie* 78(9): 403–411. <https://doi.org/10.5117/mab.78.21824>
- Kloosterman HHW (2004b) Wat is eigenlijk risicoanalyse in de accountantscontrole? *Maandblad Voor Accountancy en Bedrijfseconomie* 78(12): 570–578. <https://doi.org/10.5117/mab.78.13818>
- Kruschke JK, Liddell TM (2017) Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review* 25(1): 155–177. <https://doi.org/10.3758/s13423-017-1272-1>
- Kwartiermakers toekomst accountancy (2023) Druk en tegendruk: Slotrapportage Kwartiermakers toekomst accountancy. [online] <https://open.overheid.nl/documenten/92348e83-becb-4988-bd-fc-4312f5a288c0/file> [Accessed 26 March 2024]
- NBA [Nederlandse Beroepsorganisatie van Accountants] (2019) NBA Handreiking 1141: Data-analyse bij controle: uitdagingen en vooral kansen. [online] <https://www.nba.nl/siteassets/wet--en-regelgeving/nba-handreikingen/1141/nba-handreiking-1141-data-analyse---18-juni-2019.pdf> [Accessed 26 March 2024]
- Raats VM, Moors JJA (2003) Double-checking auditors: a Bayesian approach. *Journal of the Royal Statistical Society Series D: The Statistician* 52(3): 351–365. <https://doi.org/10.1111/1467-9884.00364>
- Scrucca L, Fop M, Murphy TB, Raftery AE (2016) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal* 8(1): 289–317. <https://doi.org/10.32614/RJ-2016-021>
- Snoei W, Van Nieuw Amerongen N (2015) Toepassing van (big) data-analyse in de MKB-jaarrekeningcontrole in een relatief eenvoudige omgeving. *Maandblad Voor Accountancy en Bedrijfseconomie* 89(10): 377–385. <https://doi.org/10.5117/mab.89.31178>
- Stewart T (2012) Technical notes on the AICPA audit guide audit sampling. American Institute of Certified Public Accountants (New York, USA).

- Stringer KW (1963) Practical aspects of statistical sampling in auditing. In Proceedings of the business and economic statistics section. American Statistical Association (Washington, D.C., USA), 405–411.
- Stringer KW, Stewart TR (1986) Statistical techniques for analytical review in auditing. John Wiley & Sons.
- De Swart J, Wille J, Majoor B (2013) Het ‘push left’-principe als motor van data analytics in de accountantscontrole. *Maandblad Voor Accountancy en Bedrijfseconomie* 87(10): 425–433. <https://doi.org/10.5117/mab.87.11905>
- Thiprungsri S, Vasarhelyi MA (2011) Cluster analysis for anomaly detection in accounting data: An audit approach. *The International Journal of Digital Accounting Research* 11: 69–84. https://doi.org/10.4192/1577-8517-v11_4
- Van Batenburg PC, Dassen RJM (1996) Het Bayesiaanse model van Deloitte & Touche ter ondersteuning van de controlemix. *De Accountant* 103(1): 31–35.
- Van der Aalst W, Koopman A (2015) Process mining: Data analytics voor de accountant die wil weten hoe het nu echt zit. *Maandblad Voor Accountancy en Bedrijfseconomie* 89(10): 359–368. <https://doi.org/10.5117/mab.89.31176>
- Wagenmakers E-J, Lee M, Lodewyckx T, Iverson G J (2008) Bayesian versus frequentist inference. In: Hoijtink H, Klugkist I, Boelen PA (Eds) *Bayesian evaluation of informative hypotheses*. Springer (New York, USA): 181–207. https://doi.org/10.1007/978-0-387-09612-4_9