

Data-analyse: de speld in de hooiberg

Drs. E.J. Wolters RE RA, drs. T.G. Kloosterman RA en drs. E.M. Peeters RE

Als gevolg van de economische stagnatie en de afnemende mogelijkheden voor autonome groei in veel markten worden ondernemingen gedwongen het inzicht in kosten en opbrengsten te vergroten. Met behulp van specifieke data-analysesoftware wordt het aloude adagium 'meten is weten' nieuw leven ingeblazen.

Inleiding

Voor veel ondernemingen is als gevolg van de economische stagnatie en volwassenheid van de markt het moeilijk om autonome groei te realiseren. 'Overall cost leadership' is dan een logische strategie. Hiervoor is gedetailleerd inzicht in de diverse kosten- en opbrengstencategorieën noodzakelijk.

Data-analyse actueel?

Uiteraard doen ondernemingen al jaren van alles om een zo goed mogelijk inzicht te verkrijgen in kosten en opbrengsten. Hierbij worden allerhande geautomatiseerde hulpmiddelen ingezet, zoals spreadsheets en rapportagesoftware. Voor het maken van statistische analyses en simulaties zijn data-analysetools duidelijk in opkomst.

Data-analysesoftware is erop gebouwd om met massale databestanden om te kunnen gaan. Spreadsheets kunnen maar met beperkte hoeveelheden data omgaan en ook rapportagesoftware wordt gauw traag wanneer data uit verschillende databasetabellen moeten worden gecombineerd en bewerkt. Maar de echte kracht van data-analysesoftware ligt in de functionaliteit. Met behulp van simulatietools als Matlab en Crystal Ball kunnen verschillende scenario's worden doorgerekend. Statistische pakketten als SPSS of SAS kunnen verbanden tussen variabelen aantonen en de relatieve belangrijkheid ervan berekenen.

In dit artikel lichten wij de toepassingsmogelijkheden van data-analyse toe en wordt een aantal praktijkcasussen gegeven.



Drs. E.J. Wolters RE RA is werkzaam als manager bij KPMG Information Risk Management. Hij is gespecialiseerd in de leasebranche en onder andere betrokken bij data-analyses. Tevens treedt hij op als opdrachtmanager voor diverse jaarrekeningcontroles.

wolters.eelco@kpmg.nl



Drs. E.M. Peeters RE is senior manager bij KPMG Information Risk Management. Hij is gespecialiseerd in de leasebranche en betrokken bij de selectie van brancheapplicaties, de implementatie hiervan, conversietrajecten, veranderingsmanagement en bedrijfsprocesanalyses.

peeters.marc@kpmg.nl



Drs. T.G. Kloosterman RA is manager bij KPMG Integrity & Investigation Services. Hij geeft sturing aan Forensic Technology, een unit binnen KPMG Integrity & Investigation die zich onder meer bezighoudt met data-analyse en het herleiden en interpreteren van sporen op digitale media. Verder is hij betrokken bij bijzondere onderzoeken, waaronder fraudeonderzoeken.

kloosterman.theunis@kpmg.nl

De auteurs danken hun collega's drs. L.A. Hoogduin RA, G. Schreurs en drs. ir. G.J. van Wiggen voor hun input en commentaar.

Toepassingsmogelijkheden

De toepassingsmogelijkheden van data-analyse zijn zeer gevarieerd. In het algemeen kan data-analyse worden toegepast voor:

- het zoeken van verbanden tussen variabelen. Hierdoor kunnen de ‘drivers’ van kosten of opbrengsten worden geïdentificeerd.
- het doelgericht zoeken van bepaalde gegevens in grote gegevensverzamelingen. Hierbij kunnen onregelmatigheden, onjuistheden of uitschieters worden geïdentificeerd.

Dit lijkt op het eerste gezicht niet zoveel nieuws, maar bedacht moet worden dat door het inzetten van data-analysesoftware een aantal problemen kan worden overwonnen die met oorspronkelijke geautomatiseerde hulpmiddelen zoals spreadsheets en rapportagesoftware niet (of niet efficiënt) op te lossen waren.

Ten eerste kan data-analysesoftware met massale gegevensbestanden omgaan – het analyseren van honderdduizenden databaserecords is geen enkel probleem. Spreadsheets zijn daarentegen begrensd tot een maximumhoeveelheid data en de performance van rapportagesoftware laat bij grote gegevensverzamelingen al gauw te wensen over.

Ten tweede blijkt uit de praktijk dat met behulp van traditionele rapportagesoftware gemaakte management-rapportages te veel gericht zijn op het verleden en management geen instrumenten in handen geeft bij te sturen of in te grijpen. Data-analysesoftware is juist in staat om gerichte analyses uit te voeren die een bijdrage kunnen leveren aan gefundeerde ingrepen door het management.

Statistische software

Het bovenstaande kan worden geïllustreerd aan de hand van statistische software. Deze is namelijk in staat de correlatie tussen variabelen *automatisch* te berekenen. Dit houdt in dat het pakket *alle* mogelijke combinaties van variabelen probeert, net zo lang tot de verklarende variabelen zijn geïdentificeerd en gekwantificeerd. Omdat doorgaans meerdere variabelen kosten of opbrengsten beïnvloeden, is het ondoenlijk dergelijke analyses met traditionele rapportagesoftware of spreadsheets uit te voeren. Dit omdat er met deze laatste twee typen tools een reeks van verschillende rapportages nodig is om de ontwikkeling van allerlei variabelen te monitoren. Daarnaast kunnen ‘drivers’ elkaar beïnvloeden en kunnen ze in de loop van de tijd veranderen. Dit geeft een extra complexiteit voor rapportagetools, omdat deze in beginsel een statisch karakter hebben.

Simulatiesoftware

Simulatiesoftware kan het management helpen bij antwoorden op ‘what if’-vragen. Met behulp van simulatie-

Leasemaatschappij: analyse van resultaten op reparatie, onderhoud en banden

Een leasemaatschappij zag sinds een aantal jaar haar resultaat op reparatie, onderhoud en banden (ROB) teruglopen. Het management had weliswaar een aantal indicaties die deze ontwikkeling verklaarden, maar kon de vinger niet op de zere plek leggen. Door middel van statistische analyses is aangetoond welke variabelen wel – en in welke mate – en niet het ROB-resultaat verklaarden. Op basis van de verklarende variabelen is de totale contractpopulatie ingedeeld in subgroepen met homogene eigenschappen. Hierbij is niet alleen aangetoond wat de gemiddelde ROB-kosten zijn per merk en model, maar ook welke exacte invloed de cilinderinhoud, het brandstoftype, de kleur van de auto of het (jaar)kilometragede, de leeftijd van de bestuurder, zijn/haar geslacht, de branche van de lessee, de plaats van vestiging van de lessee, de introductie van nieuwe modellen en de jaarovergang op het ROB-cijfer hebben.

Deze analyse heeft het management in staat gesteld het samenstel aan oorzaken van de dalende ROB-resultaten te identificeren en hierop passende maatregelen te nemen. Het wekte geen verbazing dat het automerk een belangrijke verklarende factor van de ROB-kosten bleek te zijn. Ook het bestaan van een sterke correlatie tussen kosten voor banden en onderhoudskosten werd door de analyse bevestigd. Wel was opvallend dat cilinderinhoud nauwelijks een rol speelde. Ook bleek er een sterk verband met de branche waarin de lessee actief is. Het belang hiervan was tot dat moment sterk onderschat. Ten slotte was interessant dat kostbare auto's per saldo net zoveel rendement leveren als bepaalde standaardgezinsauto's. De perceptie was tot dan toe dat luxeauto's naar verhouding juist een hogere bijdrage aan het resultaat leverden.

Casus 1.

Financiële instelling: analyse verkoopgegevens

Een financiële instelling die een scala van financieringsproducten aan particulieren aanbiedt, had al geruime tijd te kampen met stagnerende verkopen – met name aan nieuwe cliënten. Over de demografische en persoonlijke kenmerken van haar klantenbestand was een schat aan informatie verzameld. Door middel van statistische analyse konden de gemeenschappelijke eigenschappen van de nieuwe klanten worden vastgesteld.

Het bleek dat de meeste groei werd gerealiseerd in decentrale, minder ontwikkelde regio's. Ook bleek dat de leeftijd van de meeste nieuwe klanten gemiddeld een stuk hoger lag dan verwacht. Op basis van de statistisch aangetoonde correlaties heeft het management de marketingstrategie bijgesteld door zich intensiever te focussen op bepaalde groepen cliënten.

Casus 2.

tools zoals Matlab of Crystal Ball kunnen historische resultaten naar de toekomst worden geëxtrapoleerd of kan een aantal scenario's worden doorgerekend – bijvoorbeeld bij het doorrekenen van orderniveaus en bestelmomenten in distributiecentra of bij het bepalen van mogelijke rendementen van aandelenfondsen.

Voor wie is data-analyse interessant?

De resultaten die met data-analyse worden verkregen, kan het management gebruiken als sturingsinformatie. Wanneer bijvoorbeeld de 'drivers' van een bepaalde kostensoort zijn geïdentificeerd, kan het management maatregelen nemen om de kosten te beperken. Ook kan het identificeren van 'drivers' leiden tot een aanpassing of verfijning van de prijsstelling van producten of diensten. Data-analyse kan dan ook een bijdrage leveren in een Activity Based Costing (ABC)-traject.

Niet alleen het management van ondernemingen maakt gebruik van data-analyse. Data-analyse wint aan populariteit onder de auditor. Dit omdat het integraal controleren van gegevensverzamelingen en het statistisch aantonen van veronderstelde verbanden met behulp van data-analysetools tot onomstotelijk controlebewijs kan leiden. Dit vormt een welkome aanvulling op de ver doorgevoerde systeemgerichte controleaanpak die tegenwoordig door auditors wordt gepraktiseerd ([Scho04]).

Auditors maken gebruik van software die een betrouwbare methode van data-extractie en behandeling waarborgt. Auditsoftware – zoals ACL of WinIDEA – voorziet in tal van waarborgen ten behoeve van een betrouwbaar data-analyseproces. Auditsoftware zorgt hierbij voor de nodige waarborgen voor een betrouwbare audit trail. Ingelezen databestanden kunnen door auditsoftware niet worden gemanipuleerd. De bronbestanden blijven onaangetast waardoor de foutkans van de analyses verkleint, hetgeen ten goede komt aan de betrouwbaarheid van de uitkomsten van de uitgevoerde analyse.

ses. In bewijsvoering ten behoeve van bijvoorbeeld juridische procedures kan dit van groot belang zijn.

Voor wie is data-analyse geschikt?

Het werd al even genoemd: één van de voordelen van data-analysesoftware is dat het met massale gegevensverzamelingen om kan gaan. De ervaring leert echter dat data-analyse niet alleen geschikt is voor grote multinationals maar ook effectief kan worden ingezet bij kleinere organisaties.

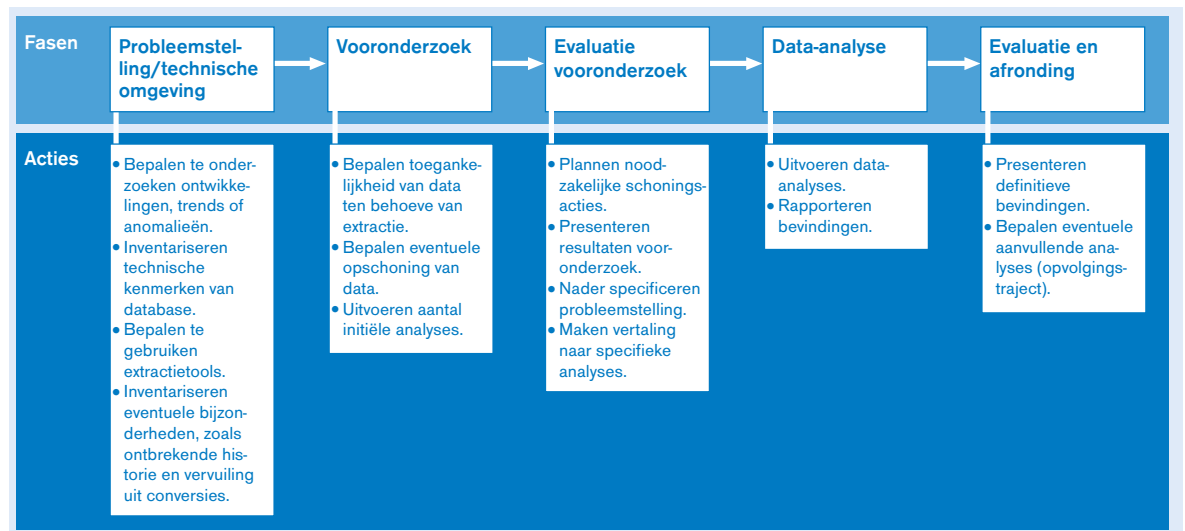
Data-analyse kan laagdrempelig blijven doordat er geen omvangrijke investeringen voor noodzakelijk zijn. Ook hebben data-analysetrajecten veelal een korte doorlooptijd. Eén of enkele workshops kunnen al tot bruikbare resultaten leiden. Waar multinationals waarschijnlijk zullen kiezen voor een meer structurele aanpak van data-analyse en geautomatiseerde tools zullen aanschaffen en mensen opleiden, kunnen kleinere organisaties opteren voor een ad-hoc aanpak – waarbij de geautomatiseerde hulpmiddelen van externe adviseurs worden gebruikt.

Haarlemmer olie

De ervaring leert dat data-analyse over het algemeen interessante uitkomsten genereert, onafhankelijk van het doel waarvoor zij wordt toegepast. Toch is data-analyse geen panacee.

Ten eerste is voor het adequaat bedienen van dergelijke tools een aanzienlijke hoeveelheid specialistische kennis noodzakelijk. Om bijvoorbeeld de door het al eerder genoemde statistische pakket SPSS gegenereerde resultaten juist te kunnen interpreteren, is gedetailleerde kennis van statistiek en modellering onontbeerlijk.

Ten tweede is het van belang data-analyses uit te voeren met een stevige dosis branchekennis en een goede afbakening van de probleemstelling. Anders dreigt het



Figuur 1. Fasering van een complex data-analysetraject.

gevaar dat data-analyse verzandt in steeds verder gedetailleerde analyses, met alle kosten van dien. Branche-deskundigen kunnen de probleemstelling of informatiebehoefte van het management vertalen naar voor de technische en statistische specialisten uitvoerbare data-analyses. Ook hebben zij een cruciale rol bij de evaluatie en interpretatie van de uitkomsten. Deze uitkomsten moeten immers weer worden vertaald naar een informatiebehoefte van het management.

Ten derde moet worden bedacht dat analyses die worden uitgevoerd op onbetrouwbare data tot verkeerde conclusies kunnen leiden. Het is daarom van groot belang om eerst te evalueren of de data toereikend zijn voor het analysedoel. Hierbij moet voordat tot data-analyse wordt overgegaan, worden onderzocht of en zo ja, in hoeverre het noodzakelijk is data te schonen, te corrigeren of buiten beschouwing te laten. Hiertoe is inzicht nodig in het datamodel en de databasestructuur van de onderneming. Datatabellen kunnen immers wat naamgeving betreft een bepaalde inhoud suggereren die in de praktijk anders wordt uitgelegd. Het is daarom een vereiste dat inzichtelijk is wat de betekenis is van de diverse variabelen en hoe dit vertaald wordt in de bestanden. Veelal wordt hierbij snel inzicht verkregen in de data-integriteit van de bestanden. Dit op zijn beurt geeft weer inzicht in de werking van procedures en geprogrammeerde controles.

Data-analysetools versus traditionele rapportagesoftware en spreadsheets

Data-analysetools zijn geen substituuft voor traditionele rapportagesoftware en spreadsheets. Ze moeten worden gezien als een aanvulling op de algemeen toegepaste hulpmiddelen zoals MS Excel, Business Objects, Crystal Reports en Cognos. Spreadsheets hebben dan wel een aantal beperkingen, als gevolg van hun gebruiksgemak en de lage aanschafkosten blijven ze voor iedere organisatie nuttig. Ook rapportagesoftware blijft een geschikt hulpmiddel. Eenmaal met statistische software aangevoerde 'drivers' kunnen bijvoorbeeld worden gevolgd met periodieke vastomlijnde maandrapportages. Kortom, het gecombineerde gebruik van de verschillende soorten geautomatiseerde hulpmiddelen zal in de praktijk waarschijnlijk tot de beste resultaten leiden.

Aanpak

De aanpak van een data-analysetraject zal afhangen van de aard, complexiteit en omvang van de onderneming. Zoals eerder aangegeven is het in ieder geval van belang een heldere, afgebakende probleemstelling te formuleren – meestal in de vorm van een vraag van het management. Daarnaast moet een initieel onderzoek worden uitgevoerd om vast te stellen of en zo ja, in hoeverre het schonen van data noodzakelijk is.

Verzekeringsmaatschappij: beheersing schaderesultaat

Het technisch resultaat van een verzekeringsmaatschappij bestaat uit schadelast, interne en externe kosten, premie- en provisiestelling en het resultaat op herverzekering. Met behulp van data-analyse is het mogelijk deze componenten van het technisch resultaat inzichtelijk te maken en kan de onderlinge samenhang van de verschillende factoren die invloed hebben op het technisch resultaat in kaart worden gebracht. Dit biedt het management handvatten voor het verbeteren van het beheersen van het technisch resultaat. Voorbeelden van analyses:

- Door het analyseren van onderlinge verbanden tussen premie en schadebetalingen kan naar voren komen dat een hogere concentratie bestaat van schadebetalingen op jonge polissen dan gewenst. Het doorrekenen van verschillende polisvoorwaarden en premiehoogten vanuit de schadehistorie naar de toekomst geeft informatie ten behoeve van het optimaliseren van het technisch resultaat.
- Door het analyseren van onderlinge verbanden tussen de externe kosten en de schade per shadedossier wordt een indruk verkregen van de efficiency van schadebehandeling door externe partijen.
- Met behulp van data-analyse kan de ideale verhouding worden berekend tussen de herverzekerde claim per branche en verschillende excess of loss-drempels. Op deze wijze kan de herverzekeringsstructuur worden geoptimaliseerd.
- Door het analyseren van onderlinge verbanden tussen premie en schadebetalingen kan naar voren komen dat bonus-malustreden ten onrechte niet naar beneden zijn bijgesteld in verband met uitbetaalde schade. Zo zijn er nog talrijke voorbeelden van zinvolle analyses op te noemen die het management inzicht geven in de wijze waarop verschillende variabelen invloed hebben op het technisch resultaat.

Casus 3.

Luchtvaartmaatschappij: analyse frequent-flyerprogramma

Een luchtvaartmaatschappij wilde inzicht hebben in het gebruik van verdiende punten in haar frequent-flyerprogramma om de kosten van de punten vast te stellen. Zij maakte gebruik van een systeem dat de transacties – het verdienen en uitgeven van punten – uitsluitend registreerde, maar geen mogelijkheden bood voor nadere analyse. Met een statistisch analysepakket is voor de gehele levensduur van het frequent-flyerprogramma inzicht verkregen in het aantal punten dat per periode is verdiend, het gedeelte dat hiervan openstond, het gedeelte dat verlopen is en het gedeelte dat opgenomen is voor incentives. Bij deze analyse is tevens onderscheid gemaakt naar de verschillende membership levels van

Casus 4.

De fasering voor grotere en complexe data-analyses wordt weergegeven in figuur 1. Hierbij wordt opgemerkt dat met name de laatste twee fasen iteratief van aard kunnen zijn. Uitkomsten en bevindingen kunnen nu eenmaal leiden tot nog niet eerder vermoede verbanden of nieuwe inzichten die nader onderzoek behoeven.

Conclusie

Data-analyse maakt gebruik van krachtige geautomatiseerde tools waarmee het mogelijk is om bruikbare managementinformatie uit massale gegevensbestanden te destilleren. Bovendien kan dit in relatief korte tijd en zonder grote investeringen. Uiteraard is data-analyse geen panacee en kent zij haar eigen uitdagingen. Niettemin is uit de praktijkcasussen af te leiden dat data-analyse tot interessante resultaten kan leiden en dus terecht volop in de belangstelling staat. Data-analysesoftware lijkt een welkome aanvulling op de al bestaande geautomatiseerde hulpmiddelen.

Literatuur

[Scho04] Drs. R.P. Schouten RE RA, *De revival van auditsoftware*, Compact 2004/1.

het programma. Op basis van deze analyses en het implementeren van deze analyses in de bedrijfsvoering heeft de luchtvaartmaatschappij de kosten van het frequent-flyerprogramma beter onder controle.

Leasemaatschappij: effectiviteit en efficiency onderhoudsafdeling

Een leasemaatschappij heeft een lease-applicatie met uitgebreide workflowmanagementfunctionaliteit. Als gevolg hiervan bevat de database een groot aantal records met daarin de exacte gebruikershandelingen van de onderhoudsafdeling: gegevens als de aard en het tijdstip van de handeling liggen allemaal vast.

Door het analyseren van deze data is het mogelijk inzicht te verkrijgen in de effectiviteit en efficiency van de onderhoudsafdeling. Zo kan worden geanalyseerd hoe lang het gemiddeld duurt voordat bijzondere reparaties zijn geautoriseerd of hoe lang het duurt voordat een aanvraag voor vervangend vervoer is verwerkt. Deze informatie kan niet alleen worden gebruikt voor het stroomlijnen van de bedrijfsprocessen, maar ook voor het identificeren van 'bottlenecks' of de beoordeling van de onderhoudsafdeling of zelfs van individuele werknemers.

Casus 5.