STAtOR

periodiek van de VVS jaargang 10 nummer 1, maart 2009

STAtOR
Jaargang 10, nummer 1, maart 2009

STAtOR is een uitgave van de Vereniging voor Statistiek en Operationele Research (VVS). *STAtOR* wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operationele research. Verschijnt 4 keer per jaar.

Redactie
Goos Kant (hoofdredacteur), Ana Isabel Barros, Mirjam Moerbeek, Gerrit Stemerdink (eindredacteur), Fred Steutel, Hilde Tobi, Marnix Zoutenbier.

Kopij en reacties richten aan
Prof. dr. G. Kant (hoofdredacteur), Faculteit der Economische Wetenschappen van de Universiteit van Tilburg, Postbus 90153, 5000 LE Tilburg, telefoon 013 - 4668234, mobiel 06-11045089, <G.Kant@uvt.nl>.

Bestuur van de VVS
Voorzitter: prof. dr. R. Gill <voorzitter@vvs-or.nl>
Secretaris: dr. C.G.H. Diks <c.g.h.diks@uva.nl>
Penningmeester: prof. dr. ir. C.A.G.M. van Montfort <kvmontfort@feweb.vu.nl>
Statistische dag: prof. dr. A.W. van der Vaart <aad@cs.vu.nl>
Namens de Bedrijfssectie (BDS):
prof. dr. R.J.M.M. Does <R.J.M.M.Does@uva.nl>
Namens de Biometrische Sectie (BMS):
prof. dr. A.H. Zwinderman <a.h.zwinderman@amc.uva.nl>
Namens de Economische Sectie (ECS):
dr. P.H.F.M. van Casteren <casteren@fee.uva.nl>
Namens het Ned. Genootschap voor Besliskunde (NGB):
prof. dr. J.J. van de Klundert <j.vandeklundert@math.unimaas.nl>
Namens de Sectie Mathematische Statistiek (SMS):
dr. P.J.C. Spreij <spreij@science. uva.nl>
Namens de Sociaal Wetenschappelijke Sectie (SWS):
prof. dr. J.K. Vermunt <j.k.vermunt@uvt.nl>

Leden- en abonnementenadministratie van de VVS
VVS, Postbus 244, 6700 AE Wageningen, telefoon 0317 - 419572, fax 0317 - 421364, <admin@vvs-or.nl>. Raadpleeg onze website over hoe u lid kunt worden van de VVS of een abonnement kunt nemen op *STAtOR* of op een van de andere periodieken.

VVS-website
http://www.vvs-or.nl

Advertentieacquisitie
Marieke Klein, p/a Vrije Universiteit, afdeling Econometrie & Operationele Research, De Boelelaan 1105, 1085 HV Amsterdam, <adverteren.stator@vvs-or.nl>. *STAtOR* verschijnt in maart, juni, september en december.

Ontwerp en opmaak
Pharos / M. van Hootegem, Nijmegen

Uitgever

ISSN 1567-3383

Inhoud

Met statistiek en OR de crisis te lijf!

Reeds enkele maanden is de economische crisis niet uit het nieuws te verdrijven. Alle media berichten er uitvoerig over, menig econoom doet zijn best om Bekende Nederlander te worden, en het kabinet heeft grote moeite een gezamenlijke koers te bepalen. Alle voorstellen van politici en economen zijn er met name op gericht om op macro-niveau bestedingen te stimuleren op de korte termijn terwijl de overheidsfinanciën op lange termijn gezond blijven.

Statistiek en Operations Research zijn onzichtbaar in deze discussie. Als het gaat om de overheidsfinanciën is dat mogelijk terecht. Echter, ons vak biedt wel andere mogelijkheden in crisistijd en de artikelen in deze Stator zijn daarvan mooie voorbeelden die we zullen categoriseren in twee klassen: 'het goede' en 'het schone'.

Onder 'het goede' verstaan we toepassingen van ons vak waarin primair kwaliteit verbeterd of geld bespaard wordt. Dit stimuleert weliswaar niet de economie maar kan voor bedrijven en instellingen wel essentieel zijn om de continuïteit te garanderen. Zo gaat het artikel van Janneke Meesters over kostenbesparingen in de petrochemische industrie via de bevoorrading, beschrijft Mirjam Moerbeek hoe testen zuiniger kunnen worden opgezet terwijl de kwaliteit goed blijft, en beschrijft Jelke Bethlehem hoe de opzet van steekproeven de conclusies beïnvloedt.

Onder 'het schone' verstaan we de bijdragen waarin ons vak het gevoel van crisis kan verminderen zonder dat geld hierbij een rol speelt. Luisteren naar mooie muziek is daar een voorbeeld van. Menno van Zaanen schrijft over de ontwikkelingen op dat gebied. Johan van Leeuwaarden geeft een leerzaam inkijkje in zijn gevoelsleven, en Fred Steutel laat ons genieten van de schoonheid van de kwantitatieve methoden. We kunnen minder geld te besteden hebben, maar zolang we kunnen blijven genieten van mooie muziek en wiskunde hebben we daar wellicht niet eens zoveel last van. En Goos Kant laat zien dat als we schonere routes rijden dit wat meer geld kost. De belangstelling hiervoor in deze tijd laat zien dat geld niet het enige is dat van belang is.

De gebruikte klasse-indeling in 'het goede' en 'het schone' is volstrekt onbruikbaar voor kwantitatieve analyse-doeleinden, bijvoorbeeld doordat de klassen elkaar overlappen. Voor een uitgebreide bespreking van de overwegingen die leiden tot klasse-indelingen verwijzen we naar het artikel van Gerrit Stemerdink, een voorbeeld van zowel 'het goede' als 'het schone'.

Veel leesplezier!
De redactie.

Naar welke MUZIEK zullen we nu luisteren?

De afgelopen jaren is de manier waarop naar muziek geluisterd wordt drastisch veranderd. Luisteraars hebben zelf controle over naar wat er geluisterd wordt en sinds de introductie van het MP3 formaat, dat gebruikt wordt om muziek te comprimeren, kunnen grote muziekverzamelingen makkelijk overal naartoe meegenomen worden. Het kiezen van specifieke muziek uit deze grote verzamelingen is moeilijk. Het zou fijn zijn als het mogelijk is om automatisch muziek te selecteren op basis van eigenschappen van muziek, zoals bijvoorbeeld stemmingen (vrolijk, droevig) of vergelijkbare componisten. Het classificeren van muziek vanuit veel verschillende invalshoeken kan hierbij helpen.

Menno M. van Zaanen

Muziek stamt al uit de oude steentijd (Kunej en Turk, 2000). In het begin bestond muziek vooral uit schreeuwen gecombineerd met simpele muziekinstrumenten. Tot in de 19e eeuw, toen Edison de fonograaf patenteerde, kon een lied alleen opnieuw beluisterd worden door het opnieuw te spelen of zingen.

Terwijl de uitvinding van de fonograaf de vrijheid gaf om precies dezelfde muziek meerdere malen en op verschillende locaties af te spelen, was een volgende stap in muziekproductie de uitvinding van de bandrecorder (en daarna cassetterecorder) waarbij het voor gebruikers ook mogelijk was zelf muziek op te nemen.

Rond 1990 volgde de introductie van MPEG-1 Layer 3. Dit bestandsformaat, dat nu MP3 genoemd wordt, maakt het mogelijk muziek in digitale vorm te comprimeren, waarbij muziek efficiënt

opgeslagen kan worden op bijvoorbeeld harde schrijven van computers.

Playlist

Met behulp van de zeer compacte MP3 spelers is het makkelijk om veel muziek mee te nemen en af te luisteren. Om niet na elk muziekstuk opnieuw te moeten kiezen, worden vaak playlists gemaakt. Een playlist is een lijst van muziekstukken, die van te voren worden gekozen op basis van bijvoorbeeld eenzelfde thema, componist of stemming.

Het maken van playlists is niet makkelijk en tevens een tijdrovende zaak. Hoe bepaal je welke muziek (en in welke volgorde) in een playlist hoort? Dit probleem wordt groter naar mate de muziekverzameling groter is, zeker wanneer de luisteraar niet alle muziek kent.

Uit onderzoek blijkt dat gebruikers tegenwoordig toegang hebben tot duizenden muziekstukken. Dat maakt het kiezen bij het maken van een playlist lastig. De praktijk wijst uit dat gedurende 80% van de tijd naar slechts 20% van de muziekverzameling wordt geluisterd en er is onderzoek waaruit blijkt dat 63% van de toegankelijke muziekstukken zelfs nooit beluisterd wordt.

Doelstelling van het automatisch maken van playlists is om afhankelijk van de wensen van de luisteraar muziekstukken uit de muziekverzameling te kiezen en op die manier de luisteraar te verrassen met muziekstukken die voldoen aan de gestelde eisen en de luisteraar waarschijnlijk de moeite waard vindt, maar wellicht nooit gevonden zou hebben.

Het zoeken naar vergelijkbare muziek kan gebeuren op basis van verschillende eigenschappen. Bijvoorbeeld, muziek uit eenzelfde periode, genre, stijl, stemming, componist of artiest zouden in playlists opgenomen kunnen worden. Sommige van deze eigenschappen worden expliciet in MP3 bestanden opgenomen. Helaas is dat niet voor alle denkbare eigenschappen het geval.

Het toekennen van eigenschappen van muziek die nog niet expliciet beschikbaar zijn, kan op verschillende manieren. Een mogelijkheid is om dit handmatig te doen en daarna gegevens te delen met anderen. Dit is een *social tagging* taak, een van de onderliggende gedachten van Web 2.0, waarbij handmatig toegekende informatie van objecten (muziek in dit geval) wordt gedeeld met andere gebruikers.

Op basis van de gegevens van anderen is het ook mogelijk om met behulp van machinaal leren nog niet geannoteerde gedeelten van de muziekcollectie automatisch te classificeren. Hierbij worden alleen voor de training van zulke systemen gegevens gebruikt die handmatig zijn toegekend. Het machinaal leersysteem kan na het trainen toegepast worden op ongeclassificeerde data die daarbij automatisch geclassificeerd wordt.

Er zijn veel eigenschappen die mogelijk automatisch toegekend kunnen worden. Twee voorbeelden hiervan zijn het automatisch herkennen van componisten en het toekennen van stemmingen op basis van liedteksten. Hier zullen slechts twee systemen worden geschetst; er zijn veel meer implementaties mogelijk.

Herkennen van componisten

Als we aannemen dat bepaalde componisten een eigen stijl hebben die in de structuur van de muziek te vinden is dan zouden, op basis van de typische structuur van de componisten, nieuwe muziekstukken aan componisten gekoppeld kunnen worden. Met behulp van Alignment-Based Learning (ABL) (van Zaanen, 2002) is het mogelijk zulke typische structuren te ontdekken (Geertzen en van Zaanen, 2008). Door het vergelijken van bladmuziek van muziekstukken van een componist zoekt ABL abstracte patronen. Als deze patronen ook in nieuwe muziekstukken voorkomen is

dat een indicatie dat ze door dezelfde componist geschreven zijn.

ABL krijgt als invoer een verzameling muziekstukken van een bepaalde componist. In paren worden deze muziekstukken met elkaar vergeleken. Hierbij worden delen die in de beide muziekstukken voorkomen aangemerkt. Deze delen worden gezien als typerend voor de componist en worden bewaard in de vorm van patronen. De patronen zijn globaal en kunnen relaties tussen (reeksen van) noten voorkomend op elke plek in het muziekstuk beschrijven.

Het classificeren van een nieuw muziekstuk is nu mogelijk door voor elk van de patronen te controleren of het te vinden is in het nieuwe muziekstuk. Omdat elk patroon typisch is voor een bepaalde componist is het mogelijk bij te houden hoe waarschijnlijk het is dat een muziekstuk door een componist geschreven is door te kijken naar het aantal passende patronen van die componist. Het systeem geeft dan als uitvoer voor elke componist de kans dat het muziekstuk door deze componist geschreven is.

Tot nu toe is er nog niks gezegd over de beschrijving van de muziek. Representatie van bladmuziek is inherent anders dan bijvoorbeeld opnames van uitvoeringen. In plaats van opnames, waarbij geluidsgolven beschreven worden, vormt bladmuziek een symbolische representatie van de noten.

Voor het vinden van de patronen wordt er gewerkt met bladmuziek in het zogenaamde humdrum **kern formaat. Dit elektronisch formaat is ontwikkeld om gemakkelijk met een computer te analyseren. Toonhoogte van een noot wordt bijvoorbeeld beschreven met een letter en toonduur met een cijfer.

In dit formaat kunnen ook andere symbolen die voorkomen in bladmuziek worden gerepresenteerd, maar niet elk muziekstuk is met dezelfde details beschikbaar. In dit onderzoek worden dan ook alleen toonhoogte en lengte van noten en rusten gebruikt.

Gegeven de weergave van muziek in het humdrum **kern formaat moet worden besloten hoe deze informatie aan het systeem gepresenteerd wordt. Qua representatie is uit veel mogelijkheden te kiezen, terwijl deze keuze een grote invloed heeft op de uiteindelijke kwaliteit van de classificatie. Bladmuziek beschrijft bijvoorbeeld absolute toonhoogte en toonduur, maar het is ook mogelijk relatieve representaties (oftewel toonhoogtes en duur ten opzichte van de vorige noot) te gebruiken. Ter illustratie, mensen kunnen een melodie herkennen, zelfs als die op een andere noot begint. In zulke situaties herkennen wij relatieve toonhoogtes.

De methode op basis van relatieve toonhoogte en duur zoals hier beschreven levert een systeem op dat vier op de vijf muziekstukken correct kan classificeren. Ter vergelijking, een versimpeld Markov model kent aan ongeveer twee op de drie muziekstukken de correcte componist toe. Dit model beslist op basis van n-grammen van noten welke de meest waarschijnlijke componist is voor een muziekstuk.

Toekennen van stemmingen op basis van liedteksten

Een ander onderzoek concentreert zich op de invloed van liedteksten op de stemming die mensen aan een muziekstuk toekennen. Veel playlists worden handmatig gemaakt op basis van stemmingen (Voong and Beale, 2007). Liedjes kunnen bijvoorbeeld vrolijk of droevig, of rustig of wild zijn. In de liedteksten komen vaak woorden voor die typisch zijn voor zulke stemmingen. Woorden als zwart, dood, of huilen associëren we bijvoorbeeld typisch met droevig.

De taak is nu om de woorden die typisch een stemming beschrijven automatisch te kunnen herkennen. Het herkennen van dit soort woorden gebeurt op basis van een verzameling liedteksten

waarvan de stemming vooraf bekend is. Voor elk woord kan per stemming een waarde berekend worden dat aangeeft hoe belangrijk dat woord is voor die stemming.

TF*IDF, een maat die vaak gebruikt wordt in de context van *information retrieval*, geeft aan hoe relevant documenten zijn gegeven zoektermen.

Dezelfde maat kunnen we gebruiken om te meten hoe relevant woorden, gebruikt in liedteksten, zijn met betrekking tot stemmingen. Eerst voegen we alle liedteksten die dezelfde stemming hebben samen, zodat we evenveel 'documenten' hebben als stemmingen. We berekenen nu de TF*IDF voor elk woord binnen de liedteksten van de stemming. Dit geeft voor elk woord een score die beschrijft hoe goed dat woord de stemming beschrijft.

Het berekenen van TF*IDF gaat, zoals de naam suggereert, door de *term frequency* (TF) te vermenigvuldigen met de *inverse document frequency* (IDF). TF is het aantal maal dat het woord voorkomt in het document genormaliseerd over het totaal aantal keer dat het woord in alle documenten te vinden is. IDF is de logaritme van het totaal aantal documenten gedeeld door het aantal documenten dat het woord bevat. Dit geeft de belangrijkheid van het woord binnen de stemming aan.

TF zorgt ervoor dat als woorden vaak in een document voorkomen, ze belangrijker worden. IDF daarentegen zorgt er voor dat woorden als bijvoorbeeld 'de', die in alle documenten vaak voorkomen en daarom niet representatief voor een stemming is, weer een lagere waarde krijgen. TF*IDF geeft dus woorden die veel voorkomen in slechts een beperkt aantal documenten een hoge waarde. Het classificeren van een nieuw lied kan nu door de TF*IDF van elk woord in de tekst voor elke stemming te berekenen en op te tellen. Op basis van deze waarden kan een stemming aan de liedtekst toegekend worden.

De methode van stemmingclassificatie van muziek op basis van alleen liedteksten is uiteraard beperkt. In dit onderzoek wordt geen informatie van de muziek zelf gebruikt. Dit onderzoek kan dan ook worden uitgebreid door ook eigenschappen van de muziek, zoals toonsoort of ritme, te gebruiken. Nu wordt dit nog niet gedaan, omdat het automatisch extraheren van deze informatie uit muziek niet triviaal is.

Het automatisch analyseren van muziek en op basis van deze resultaten aan bieden van nieuwe muziekstukken zal in de toekomst nog veel verder uitgebreid worden. Omdat dit soort analyses het mogelijk maakt muziek vanuit verschillende invalshoeken te organiseren zal deze trend het mogelijk maken ons meer vrijheid te geven om muziek selecteren.

Muziek waar we nog nooit van gehoord hadden of zouden hebben, wordt nu beschikbaar. Biedt hierdoor de toekomst nog meer luisterplezier?

Literatuur

Geertzen, J. en van Zaanen, M. (2008) *Composer classification using grammatical inference, Proceedings of the MML 2008 International Workshop on Machine Learning and Music held in conjunction with ICML/COLT/UAI 2008*, Helsinki, Finland, 17-18.

Huron, D. (1997) Humdrum and kern: selective feature encoding. In Selfridge-Field, E. (Ed.), *Beyond MIDI: The handbook of musical codes*, Cambridge: MIT Press, 375-401.

Kunej, D. en Turk, I. (2000) New Perspectives on the Beginnings of Music: Archeological and Musicological Analysis of a Middle Paleolithic Bone Flute. *The Origins of Music*, Cambridge: MIT Press, 235-268.

Voong, M. and Beale, R. (2007) *Music Organisation Using Colour Synaesthesia. Proceedings of The ACM Conference in Computer-Human Interaction (CHI)*, San Jose, CA.

Van Zaanen, M. (2002) *Bootstrapping Structure into Language: Alignment-Based Learning. PhD thesis*, University of Leeds, Leeds, UK.

Menno van Zaanen werkt als onderzoeker, docent en coördinator Human Aspects of Information Technology (HAIT) bij de faculteit Geesteswetenschappen van de Universiteit van Tilburg.
E-mail: <mvzaanen@uvt.nl>

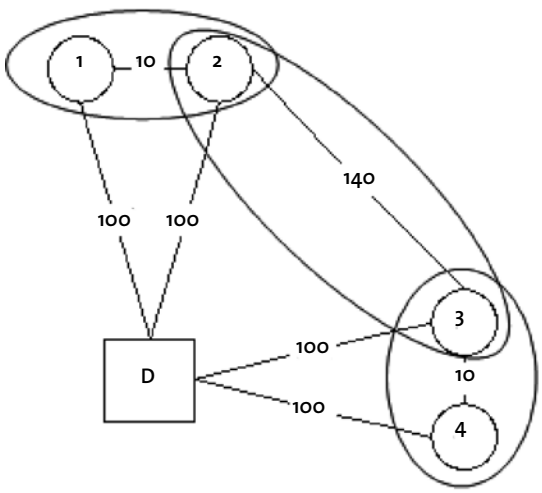
KOSTEN-EFFECTIEVE BEVOORRADING IN DE PETROCHEMISCHE INDUSTRIE

Janneke Meesters

Traditioneel beheren bedrijven zelf hun voorraad en beslissen ze wanneer ze een order bij hun leverancier plaatsen en om levering vragen. De grote uitdaging is dat er hierdoor vaak een behoorlijke volatiliteit ontstaat in de te leveren volumes, maar ook dat er geen rekening wordt gehouden met andere bedrijven die op hetzelfde moment bij dezelfde leverancier bestellen.
Nu is het in de petrochemische industrie al enige tijd zo dat de verantwoordelijkheid aan het verschuiven is naar de oliemaatschappij; gedeeltelijk doordat een deel van de te leveren adressen tot hetzelfde bedrijf behoren, zoals de tankstations bij een oliemaatschappij. Deze strategie wordt ook wel Vendor Managed Inventory (VMI) genoemd. De ervaring hiermee is dat de vrijheid die dit oplevert in het distributieproces gebruikt kan worden om een efficiëntere supply chain te creëren. Tevens biedt dit de mogelijkheid om piekperiodes uit te strijken en geografisch betere combinaties van leveringen te maken.

Vanuit een OR-oogpunt verandert dit het probleem echter aanzienlijk. In plaats van het standaard VRP (Vehicle Routing Problem) waarin de beslissing is welke wagen welke levering doet, wordt dit een IRP (Inventory Routing Problem). Behalve het feit dat er twee beslissingsvragen toegevoegd worden, namelijk *wanneer* lever ik deze klant en *hoeveel* lever ik deze klant, betekent dit ook dat de planningsperiode minder goed is af te bakenen. De beslissingen voor de korte termijn hebben namelijk invloed op de toekomstige kosten.

Een klein voorbeeld om het probleem te illustreren is het klassieke voorbeeld van Bell [1] (Figuur 1). De meest de hand liggende oplossing is om klant 1 en 2 te combineren in één rit, en klant 3 en 4. Dit betekent dagelijks twee ritten met een geleverd volume van 7.500 liter en 420 km per dag. Een betere strategie is echter om op de eerste dag enkel klant 2 en 3 te combineren in één rit, en op de tweede dag alle vier de klanten te beleveren, door klant 1 en 2 te combineren en klant 3 en 4. Dit betekent ook een totaal volume van 7.500 liter per dag, maar ditmaal tegen 380 km per dag, een besparing van 40 km dus. In de realiteit spreken we echter niet over 4 klanten en 1 wagen op 1 depot, maar over vele duizenden klanten, tientallen wagens en depots en moet de planning over een heel jaar geoptimaliseerd worden.



Figuur 1. Het voorbeeld van Bell.

Vanwege deze complexiteit wordt het probleem meestal in twee fases opgelost. Allereerst worden verschillende voorspellingsmethodieken gebruikt om het dagelijkse verbruik te voorspellen. Op basis van deze voorspelling en de laatste levering en daarmee het laatst bekende voorraadpeil, worden orders gegenereerd op het moment dat iedere klant voor het eerst voor een product de veiligheidsvoorraad bereikt. Pas in de tweede fase worden deze orders toegekend aan een wagen en op een specifieke rit gepland.

Verschillende industrieën hebben hun eigen kenmerken en daarmee verschilt ook de problematiek die opgelost dient te worden. In de afgelopen jaren is, onder andere door middel van afstudeerstages (zie 2 en 4) een beter begrip van deze problematiek ontstaan bij ORTEC, met als resultaat de ontwikkeling van algoritmen specifiek voor deze petroleum- en gasindustrie.

Uitdagingen van de petroleumindustrie

De petroleumindustrie draait voornamelijk om de bevoorrading van tankstations en automaten en kenmerkt zich door een beperkt aantal klanten, met een hoge omloopsnelheid in producten en daardoor meestal ook meerdere leveringen per

	CUSTOMER			
	1	2	3	4
tank capacity	5000	3000	2000	4000
daily usage	1000	3000	2000	1500

truck capacity 5000

week, of soms zelfs per dag. Vaak gaat het om meerdere producten die in gecompartimenteerde wagens worden vervoerd. Ook zijn er meestal meerdere leverpunten waar beperkingen zijn op de beschikbaarheid van de producten en bestaan er prijsverschillen per product tussen de verschillende leverpunten.

Doordat er in de meeste gevallen gewerkt wordt met een heterogene vloot waarbij iedere wagen een eigen compartimentenconfiguratie heeft, liggen ordergeneratie en routeplanning bijzonder dicht bij elkaar. Immers, doordat iedere klant verschillende producten ontvangt, moet de precieze compartimentering van de wagen bekend zijn bij het genereren van orders. Deze is echter pas bekend op moment dat de routeplanning al gemaakt is, maar die kan niet gemaakt worden als er geen orders bekend zijn.

Binnen onze oplossingssuite is dit opgelost door orders met ranges te maken voor de producten en door een set optionele orders te creëren die als complementen kunnen dienen voor klanten die geen volle wagen kunnen ontvangen[2]. Voor het 'kritieke' product is maar één leverhoeveelheid mogelijk, namelijk de maximale hoeveelheid. Voor de overige producten is er echter een keuze; er moet minimaal voldoende geleverd worden om evenzoveel dagen vooruit te kunnen. De maximale leverhoeveelheid hangt af van de grootte van de tank. Door deze ranges mee te geven aan elk product van de order in plaats van een vaste hoeveelheid is er wel een basis voor het maken van een routeplanning, zonder dat het noodzakelijk is vooraf al te weten in welke wagen een order ingepland zal gaan worden.

Doordat het probleem praktisch gezien vrij eenvoudig is, lukt het over het algemeen een planner met softwarematige ondersteuning vrij goed een eigen oplossing te vinden. Deze oplossing is zodanig goed dat het hen lukt vrij dicht te komen bij de ondergrens: lever alle klanten met een volle wagen. Een volledig automatische oplossing verslaat de planner dus ook niet op het gebied van gereden kilometers per geleverde liter, maar wel op het moment dat vaste kosten en kosten van overwerk meegenomen worden. Om VMI en VRP nog dichter bij elkaar te krijgen en daarmee binnen redelijke rekentijd de IRP op te lossen, is verder onderzoek nodig.

Uitdagingen van de LPG-industrie

In de gasindustrie zijn compleet andere kwesties aan de orde; hier gaat het om leveringen naar vele duizenden klanten, uiteenlopend van LPG-tankstations die meerdere malen per week beleverd worden tot particuliere klanten die slechts één of tweemaal per jaar beleverd worden. De tankgrootte bij klanten is dusdanig klein, dat het gemiddelde aantal stops per rit op kan lopen tot ongeveer 20 maar gemiddeld tussen de 8 en 10 uitkomt. Bovendien is het gasverbruik enorm cyclisch: bij gebruik voor verwarming geldt een laag verbruik in de zomer, een hoog verbruik in de winter. Bij gebruik in de landbouw of op campings kan het juist precies andersom zijn. Omdat voor het afleveren van gas specifieke wagens nodig zijn, kunnen er niet flexibel veel of weinig voertuigen ingezet worden en dus ligt de uitdaging voornamelijk in het afvlakken van seizoenspieken.

Per individuele klant is het niet verstandig om te vroeg te leveren, daarmee verlies je immers de efficiëntie doordat je op lange termijn vaker zult moeten leveren. Maar als een klant in een afgelegen gebied ligt met lage leveringsdichtheid en er toch al een rit gepland is voor dit gebied, kan het verstandig zijn deze en andere klanten in dit gebied vervroegd te leveren. Daarmee wordt een vervolgbezoek aan dat afgelegen gebied uitgesteld. Bovendien kan het zijn dat in deze periode er transportcapaciteit over is, terwijl in een latere periode transportcapaciteit ontbreekt.

Voor de retailindustrie hebben we voor deze

problematiek een concept ontwikkeld om een periodiek beleveringsschema te maken[3]. Als input voor deze zogenoemde *Tactical Route Planner* dient een lijst met potentiële leveringsschema's afhankelijk van de te leveren frequentie. Door een integer lineair programmeringprobleem op te lossen worden de klanten over de planningsperiode gebalanceerd door voor iedere klant een specifiek schema te kiezen. Dit concept is verder uitgewerkt en toegepast in de gasindustrie. In plaats van algemene schema's te gebruiken, wordt op basis van het voorspelde verbruik, per klant een individuele levering bepaald. (Figuur 2)

Om dit te berekenen wordt voor een week vooruit berekend welke klanten op welke dagen bezocht zouden kunnen worden met welke hoeveelheid. De onderliggende methode probeert vervolgens zoveel mogelijk opdrachten in hetzelfde gebied te clusteren op dezelfde dag, en daarnaast zoveel mogelijk de transportcapaciteit te balanceren over de dagen. Tevens wordt gekeken naar de middellange termijn (13 weken) trends in gasverbruik om het doelvolume te bepalen dat in deze week geleverd dient te worden. Dit gebeurt op basis van een kosten minimalisatie waarbij o.a. de afstand naar de dichtstbijzijnde klant die zeker geleverd moet worden in de planningsperiode en het voorraadpeil wordt meegenomen.

Wat de praktijk verschillend maakt van hetgeen in de literatuur is beschreven, is:

- de grootte van de probleeminstanties; veelal gaat het om vele tienduizenden klanten en meerdere depots;
- de heterogene vloot, tijdsvensters bij klanten en toegangsrestricties voor verschillende wagens op leverlocaties;
- praktische toepasbaarheid en rekentijd.

Toekomstig onderzoek zal zich voornamelijk richten op een multi-depot toepassing van dit concept en intelligent preprocessing om het clusteren en daarmee de rekentijd te verbeteren[4].

Dit concept is getest in een theoretische omgeving bij een grote Franse gasleverancier en leverde bijzonder goede resultaten. Voor een periode van 4 weken, met 13.000 klanten en even zoveel leveringen werd een potentiële stijging in volume per gereden kilometer van meer dan 40% gemeten ten opzichte vaste routes die momenteel in de praktijk gereden worden. Voldoende reden om een volgende stap in te gaan en dit concept in de realiteit te gaan gebruiken voor de dagelijkse planning.

Noten

1 Bell, W.J., Dalberto, L.M., Fisher, M.L., Greenfield, A.J., Jaikumer, R., Kedia, P., Mack, R.G., and Prutzman, P.J. 1983. *Improving the Distribution of Industrial Gases with an On-line Computerized Routing and Scheduling Optimizer.* Interfaces 13, 4-23.

2 Golbach, R. 2008. *Efficient Fuel Replenishment*. Master Thesis, University of Twente.

3 Hoendervoogt, A. 2006. *Period Scheduler: an algorithm for strategic planning*. Master Thesis, Tilburg University – dept. of Econometrics & Operations Research

4 Hulshof, P.J.H. 2008. *How VMI can be successful in Gas Distribution – A solution methodology for the inventory routing problem in gas distribution.* Master Thesis, University of Twente.

Janneke Meesters is senior consultant en Global Industrie Expert voor de Olie & Gas Industrie bij ORTEC. E-mail: <JMeesters@ortec.nl>

Fo33	1	1	%	26	0	0	0	0	0
Fo33	2	1	%	0	34	0	0	0	0
Fo33	3	1	%	0	0	41	0	0	0
Fo33	4	1	%	0	0	0	49	0	0
Fo33	5	1	%	0	0	0	0	57	0
Fo33	6	1	%	0	0	0	0	0	84

Figuur 2.

column

HET IS ALTIJD LENTE IN DE OR

Als beginnend columnist voor STAtOR had ik me geen zachtere landing kunnen wensen. Alles ligt open. Economische wetten gaan op de schop, oude filosofen komen van stal en herzieners grijpen hun kans: de wereld is een chaos, men weet het even niet meer. Maar als adolescente onderzoeker in de OR ben ik ook het spoor wat bijster. Wat doet dit wispelturig universum met mijn zo dierbare vakgebied?

Johan van Leeuwaarden

Een kink in de kabel

Het moet ergens in de lente van 2003 geweest zijn. Mijn netwerkverbinding werd weer eens om obscure redenen verbroken en de verloren tijd die het herstarten in beslag zou nemen besloot ik te gebruiken om even te genieten van de namiddagzon op mijn balkon. Mijn oog viel meteen op de zwarte geul die mijn straat over de volle lengte doorkliefde. Twee loopgravers waren pal onder me driftig in de weer met een kleine graafmachine. 'Wat staat hier te gebeuren?', riep ik uit nieuwsgierigheid, waarop één van hen opkeek en antwoordde: 'Wij leggen hier een glasvezelnetwerk aan meneer, met een haast onbegrensde bandbreedte, ontzaglijk veel sneller dan het huidige kabelnetwerk en dus ligt voor u digitale televisie en ultrasnel internet in het verschiet!' Zonder nog een woord uit te kunnen brengen ging ik naar binnen, sloot mijn balkondeur en besefte dat een deel van mij zojuist levend was begraven.

Nu moet u weten dat ik op dat moment al dertien maanden bezig was met mijn promotieonderzoek over hoe kabelnetwerken, met hun beperkte capaciteit, zo goed mogelijk in te zetten voor het internet. Driftig tikte ik mijn nieuwste wiskundige resultaten in terwijl drie verdiepingen lager het hele wezen van mijn onderzoek als

een dichtgeslibde ader uit de grond werd getrokken. Obsoleet geraken is de Nederlandse vertaling van een term die ik toentertijd opdeed in een anoniem beoordelingsrapport. Instantaan verouderd, of bij leven reeds achterhaald, is wellicht nog treffender.

Ik was een uilskuiken, los in het dons, en ongewapend tegen de harde wetten van het onderzoek. Dat een promovendus begint als uilskuiken is overigens niets vreemds, en eigenlijk zeer toepasselijk, want een uilskuiken verwijst naar het jong van een vogel wiens superieure inzicht vaak symbool staat voor wijsheid (de promotor dus). Wel vreemd dat het kroost van het toch notabel uilengeslacht zo laag wordt ingeschaald, maar dat terzijde.

Het besef

Als licht getraumatiseerde jongeling zwoer ik nimmer meer mijn lot te verbinden aan passerende hightech. In de jaren die volgden stripte ik de inleidingen van mijn artikelen tot wankele geraamtes met slechts wat algemene opmerkingen over hoe breed inzetbaar mijn wiskundige resultaten wel niet waren. Ik verzweeg mijn doelen en weigerde mijn formules te vervuilen met chaperonnerende teksten over vergankelijke toepassingen. Maar naarmate ik een beter beeld kreeg van mijn vakgebied groeide het besef dat mijn credo om slechts doelloze artikelen van onbeperkte houdbaarheid te schrijven niet langer te handhaven viel. Waren als onderzoeker in de toegepaste wiskunde mijn formules niet per definitie onrein, voor galg en rad geboren, en zat daar nou niet juist de uitdaging in? Bob Dylan zong ooit: 'Some are mathematicians, some are carpenter's wives, don't know how it all got started, I don't know what they're doin' with their lives.' Ik was op zoek naar nut en duiding.

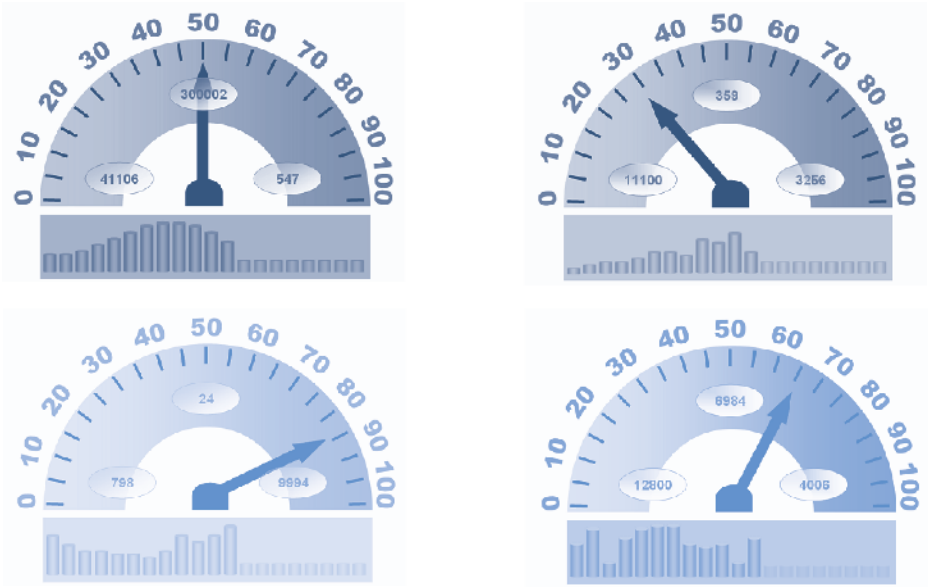
Gelukkig was daar de OR, mijn vakgebied dat floreerde in de met optimisme en groei doorspekte naoorlogse jaren, naadloos aansluitend bij de technologische ontwikkelingen en de toenemende verwevenheid van universiteiten en bedrijven. Wiskundige technieken werden geavanceerder, de complexiteit van de vraagstukken groter, en belangen gewichtiger. Zo groeide OR uit tot een machtig bolwerk, volop kansen biedend aan onderzoekers uit vele windstreken, onder één voorwaarde: de toepassing staat centraal!

De ommekeer

De kritieke zeven jaren van mijn huwelijk met de wetenschap zijn doorlopen, en zoals dat gaat bij langdurige verbintenissen: ik heb mijn mening drastisch bijgesteld. Ik omarm nieuwe ontwikkelingen. Ik ben dankbaar voor iedere ommekeer, iedere aanleiding om jezelf opnieuw uit te vinden, op zoek naar de fraaie wiskunde die voor het oprapen ligt tussen de brokstukken van de vorige werkelijkheid.

Een veranderend universum laat modellen in hun hemd staan, bij de waan van de dag ontmanteld. De omvallende banken trokken veel wiskundige modellen mee, net als de glasvezel mijn kabels de das omdeed. Daar waar het fout gaat, worden de modellen geslachtofferd. Het zij zo. Sterker nog: het is prachtig! Want bij zwaar weer wordt een dringend beroep gedaan op de OR en hebben we nog meer dan anders de wind in de zeilen. Verrassende wendingen brengen nieuw leven, nieuwe problemen die schreeuwen om oplossingen. In deze donkere tijden is het goed te weten dat het altijd lente is in de OR.

Johan van Leeuwaarden is werkzaam in de groep Stochastische Besliskunde bij de faculteit Wiskunde en Informatica van de TU Eindhoven. Tevens is hij research fellow bij EURANDOM.
E-mail: <j.s.h.v.leeuwaarden@tue.nl>



Wees milieubewust: neem een schonere route

Goos Kant

Veel auto's en vrachtwagens beschikken tegenwoordig over een navigatiesysteem dat in staat is om op basis van een gedetailleerde wegenkaart de kortste of de snelste route te berekenen van de huidige locatie naar de plaats van bestemming. Deze criteria minimaliseren niet noodzakelijk de CO_2 uitstoot. Tussen het brandstofverbruik en CO_2 uitstoot bestaat een lineair verband. Dit betekent dat als het verbruik vrij nauwkeurig kan geschat worden op een bepaalde route, dat daarbij ook de totale CO_2 uitstoot vrij eenvoudig achtergehaald kan worden. Het verbruik wordt bepaald door onder andere: motortype, verkeersdrukte, gemiddelde snelheid, variabiliteit in snelheid, rijgedrag chauffeur, beladingsgraad, weersomstandigheden en de aërodynamica van de auto. Sommige factoren zijn eenvoudig om op voorhand in te schatten, andere factoren zoals weersomstandigheden weer niet. Toch blijft het interessant en nuttig om een idee te krijgen welke invloed CO_2 emissies als minimalisatie criterium uitoefent op de optimale route en in welke mate de schoonste route verschilt met de kortste of snelste route. Bijvoorbeeld hoeveel tijd moet er meer ingecalculeerd worden bij gebruik van de schoonste route ten opzichte van de snelste route? Of: hoeveel bedraagt het verschil in kosten tussen de kortste, de snelste en de schoonste routes?

Opzet

Samen met Van Duren heb ik een test uitgevoerd (Van Duren, 2008) Alle routes werden berekend in dit project aan de hand van een Benelux-kaart ter beschikking gesteld door kaartleverancier NavTeQ. Bij een wegsegment behoort de volgende data: begin- en eindpunt, afstand in meter en wegtype. Voor elke wegtype kan een gemiddelde snelheid per voertuigtype gedefinieerd worden. Een 17-tal wegtypes zijn beschikbaar gaande van snelweg, over regionale weg in stadsgebied tot een ferry. Voor elk voertuigtype kan de gemiddelde snelheid op een bepaald wegtype aangepast worden. Vervolgens moeten we per wegsegment een CO_2 uitstoot schatten. Als we de veronderstelling van een constante snelheid aannemen, dan kan de CO_2 uitstoot bepaald worden door Formule 1 (Palmer, 2007).

De constante waarden in deze formule hangen vooral af van onder andere het wagentype, het type brandstof en het type motor. Zodoende beschikken we voor elk wegsegment niet alleen over een aantal meters, rijtijd, maar ook over de geschatte CO_2 uitstoot als het wegsegment gebruikt wordt door een route.

Vanuit een willekeurig gekozen locatie, wordt voor elk minimalisatie criterium (afstand, rijtijd, CO_2 uitstoot) de *optimale* route berekend naar 100 willekeurig gekozen locaties en hierover het gemiddelde genomen. Voor het bepalen van de optimale route voor elk criterium werd dezelfde methode gebruikt, namelijk het vrij bekende Dijkstra algoritme. Deze routes werden met elkaar vergeleken om op de vragen gesteld in de inleiding een antwoord te bieden.

Resultaten

Uit de hierboven beschreven test komen de resultaten zoals weergegeven in Tabel 1.

De *schoonste* route is dus een mooie route, tussen de snelle én de kortste route in. Hij is iets langzamer dan de snelste, iets langer dan de kortste, en slechts weinig duurder. De *schoonste* route volgt voor ongeveer 65 – 70% de *snelste* route. Dit volgen betreft vooral op de snel- en hoofdwegen. Als we de snelheden 10% verlagen, dan is de *schoonste* route nog maar 0,3% duurder en 2,7% langzamer dan de snelste route. Bij een lagere snelheid komt de *schoonste* route dus meer in de buurt van de snelste. De totale kosten van de route nemen overigens wel toe met 6% en de uitstoot met 0,5%. De ideale snelheid is (afhankelijk van diverse parameters) ongeveer 90 km/h, dus te langzaam rijden is slechter voor het milieu.

In deze test werd een constante snelheid per

$$\mathbf{CO_{2,segment} = (a + b \times v + c \times v^2 + d \times v^e + f \times \ln(v) + g \times v^3 + h \times v^{-1} + i \times v^{-2} + j \times v^{-3}) \times k}$$

$CO_{2,segment}$: CO_2 uitstoot in gram per kilometee (gram CO2/km).
v: constante snelheid op een wegsegment (km/u).
a, b, c, d, e, f, g, h, i, j en k zijn constante coëfficiënten.

Formule 1. Formules voor berekening CO2 op wegsegementen.

De schoonste route op het gebied van	Ten opzichte van de snelste route	Ten opzichte van de kortste route
Duur	4.8 % langzamer	29.3 % sneller
Lengte	3.5 % korter	2.0 % langer
CO2 – uitstoot	2.3 % minder uitstoot	6.0 % minder uitstoot
Kosten[1]	1.1 % duurder	16.1 % goedkoper

Tabel 1. Resultaten

wegtype verondersteld. Dit betekent dat op elk moment van de dag gedurende de hele week dezelfde snelheid over het stukje weg gereden wordt. Dit strookt natuurlijk niet met de werkelijkheid, maar het bovenstaande geeft toch al een idee hoe een *schone* route er kan uitzien en in welke mate de route verschilt met de *snelste* en de *kortste*. Dit idee is zeer recentlijk ook opgepakt door commerciële bedrijven. Garmin heeft recent de EcoRoute geïntroduceerd (zie www.garmin.nl) waarbij de optie 'minder brandstof' gekozen kan worden voor de route. Daarbij wordt tevens door het tonen van informatie de berijder geholpen om door aanpassing van het rijgedrag zo efficient mogelijk met brandstof om te gaan.

Conclusies

'Schone routes' zijn interessant. Hij stoot 2 % minder CO_2 uit, en is maar 5 % langzamer en 1 % duurder dan de snelste route. Een mooi alternatief als je iets meer de tijd hebt, al is het zonder hulpmiddelen nog niet eenvoudig deze route te bepalen. Dit alternatief wordt nog interessanter als de brandstofprijs stijgt of de kosten voor het uitstoten van CO_2. Zo blijkt bij CO2-emissiekosten van boven de 300 euro per ton de schoonste route zelfs goedkoper te zijn dan de snelste route. Merk daarbij ook op dat het brandstofverbruik in deze modellering een aantal aannames bevat, en dat deze testresultaten over de Benelux gaan. Dat er allerlei groene kansen liggen moge duidelijk zijn. Van de totale Europese CO_2 uitstoot wordt 27% veroorzaakt door transport, waarvan een groot gedeelte door het autoverkeer. Ter vergelijking voor uzelf: een gemiddeld gezin in Nederland verbruikt jaarlijks 9.000 kg CO_2 (zie www.milieucentraal.nl), waarvan onder andere 3.000 kg door de auto, 2.900 kg door aardgas en 2.100 kg door electriciteit. Als u toch al bezig bent met spaarlampen, de verwarming lager zetten en slimmere douchekoppen, denk dan ook eens aan een schone route. Of nog beter: pak de fiets.

Noot

1. Voor de kosten nemen we de vrachtwagenkosten (€ 35 per uur en € 0,35 per km) en gecorrigeerd voor CO2 uitstoot (sociale kosten van ongeveer € 10 per ton, zoals bepaald door DEFRA 2005)

Literatuur

DEFRA, *Experimental Statistics on Carbon Dioxide emissions at Local Authority and Regional Level*. <www.defra.gov.uk/environment/statistics/globatmos/download/regionalrpt/localregionalco2statssumm.pdf>, 2005.

Duren, D. Van, *Not the Shortest, the Fastest or the Cheapest Path, but the Cleanest Path,* Bachelor Thesis, Faculty of Economics and Business Administration, Tilburg University, 2008.

Palmer, A. *The development of an integrated routing and carbon dioxide emissions model for goods vehicles*, PhD thesis, Cranfield University, 2007.

Goos Kant is hoogleraar Operations Management & Informatietechnologie bij de Universiteit van Tilburg, Direkteur Logistiek bij ORTEC en hoofdredacteur van STAtOR.

E-mail: <GKant@ortec.nl>.

HET WAAROM EN HOE VAN STATISTISCHE POWER ANALYSES

Wetenschappelijk onderzoek kost geld. Om kosten te besparen dienen onderzoekers van te voren goed na te denken over de opzet van hun onderzoek. Eén van de onderdelen van een onderzoeksvoorstel is een berekening van de benodigde steekproefgrootte. Een te kleine steekproef resulteert in een te kleine kans om effecten aan te kunnen tonen; een te grote steekproef betekent dat er onnodig veel proefpersonen of -dieren moeten worden gebruikt. Een statistische power analyse kan gebruikt worden voor het berekenen van de benodigde steekproefgrootte. In deze bijdrage ga ik in op het waarom en hoe van statistische power analyses, en de problemen die ik als statistisch adviseur in de praktijk tegenkom en mogelijke oplossingen daarvoor.

Mirjam Moerbeek

Wetenschappelijk onderzoekers stellen zich dikwijls de vraag of er verschillen bestaan tussen groepen van mensen of proefdieren. Pedagogen stellen zich de vraag of kinderen van ouders die een speciale opvoedingstraining hebben gevolgd minder probleemgedrag vertonen dan ouders die

zo'n training niet hebben gevolgd, fertiliteitartsen vragen zich af of rokers een lagere kans van slagen hebben bij een IVF behandeling dan niet-rokers, en sociologen stellen zich de vraag of stadsbewoners een andere attitude ten opzichte van immigranten hebben dan plattelandsbewoners.

Om dit soort vragen te kunnen beantwoorden dient de onderzoeker een steekproef te trekken uit de onderliggende populatie. Een belangrijke vraag is hoe groot deze steekproef moet zijn. De steekproefgrootte wordt in de praktijk nog te vaak op basis van niet-statistische redenen bepaald. Men kiest vaak een steekproefgrootte die praktisch of financieel haalbaar is of laat zich leiden door wat al jaren gangbaar is op de afdeling. Deze methode wordt echter steeds minder geaccepteerd door instellingen die onderzoek financieren en een gedegen powerberekening wordt steeds vaker geëist in onderzoeksvoorstellen.

Het toetsen van hypothesen, statistische power en steekproefgroottes

Als voorbeeld gebruik ik het eerste voorbeeld uit de vorige paragraaf. Het probleemgedrag van kinderen kan worden gemeten met behulp van een subschaal van de *child behaviour checklist*; hoe hoger de score hoe hoger de mate van probleemgedrag. De verwachtingen van de scores in de trainingsconditie en controleconditie noteren we als μ_E en μ_C en de standaarddeviatie is gelijk aan σ. Onder de nulhypothese is er geen verschil in verwachte waarden ($H_0 : \delta = \mu_C - \mu_E = 0$). Onder de alternatieve hypothese verwachten we een lagere score in de trainingsconditie ($H_A : \delta = \mu_C - \mu_E > 0$). De parameter δ staat bekend als de effect grootte.

Uiteraard weten we niet welke hypothese waar is. Wisten we dit wel, dan was het immers niet nodig om het onderzoek uit te voeren. Op basis van een steekproef kunnen we een statistische toets uitvoeren en besluiten de nulhypothese al dan niet te verwerpen. Zoals in Figuur 1 te zien is kunnen er in deze procedure twee soorten fouten gemaakt worden. Een type I fout wordt gemaakt als de nulhypothese ten onrechte verworpen wordt; een type II fout wordt gemaakt als de nulhypothese ten onrechte *niet* verworpen wordt. De kansen op deze twee fouten worden doorgaans aangeduid met de Griekse letters α en β. De waarden van deze kansen moeten door de toegepast onderzoeker gekozen worden op basis van de consequenties van het komen tot een verkeerde conclusie als gevolg van een Type I of II fout. Het complement van β wordt de statistische

Onbekende werkelijkheid / Conclusie op basis van statistische toets	Geen verschil tussen condities (verwerp Ho niet)	Wel een verschil tussen condities (verwerp Ho)
Geen verschil tussen condities (Ho)		Type I fout
Wel een verschil tussen condities (HA)	Type II fout	

Figuur 1. Fouten die gemaakt kunnen worden bij het uitvoeren van een statistische toets.

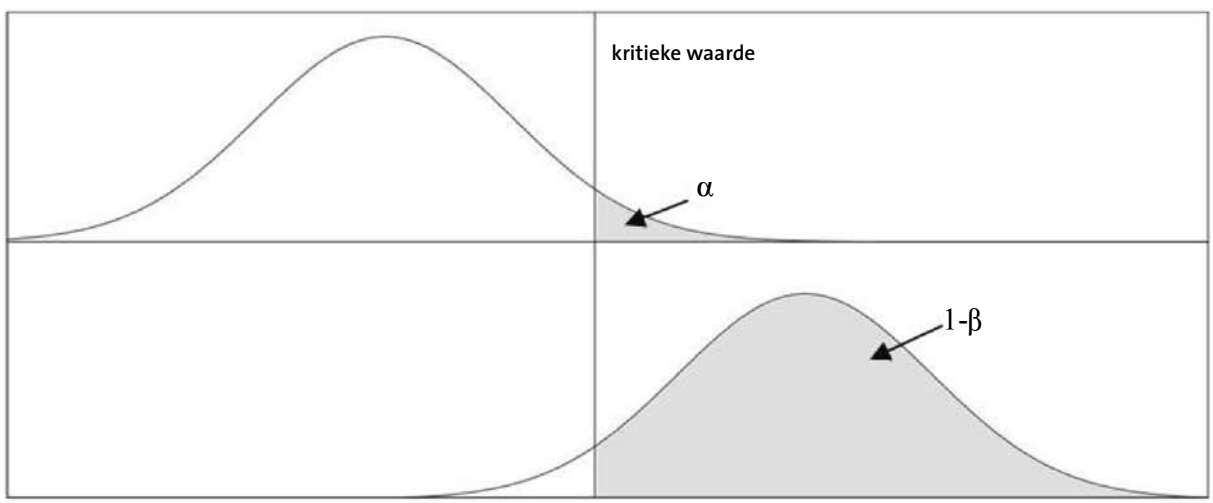
power genoemd en is de kans op het aantonen van een verschil tussen de trainingsconditie en controleconditie indien dit verschil ook bestaat in de populatie.

Voor het voorbeeld kan een onafhankelijke steekproeven t-toets gebruikt worden. Onder de nulhypothese heeft de bijbehorende toetsingsgrootheid een centrale t-verdeling met n_C+n_E-2 vrijheidsgraden, waarbij n_C en n_E de steekproefgroottes in beide condities zijn. De kansdichtheid van de verdeling onder de nulhypothese is weergegeven in het bovenste deel van Figuur 2. De verticale lijn in Figuur 2 noemen we de kritieke waarde. Wanneer de uitkomst van de toetsingsgrootheid groter is dan de kritieke waarde dan besluiten we de nulhypothese te verwerpen. De kans hierop is gelijk aan α en is weergegeven met het grijze oppervlak.

Onder de alternatieve hypothese volgt de toetsingsgrootheid een niet centrale t-verdeling met niet-centraliteitsparameter $\gamma = \delta / \sqrt{\alpha^2/n_E/+\alpha^2/n_C}$.

De kansdichtheid onder de alternatieve hypothese is naar rechts verschoven, zie onderste deel Figuur 2. De kans op het verwerpen van de nulhypothese wordt weergegeven door het grijze gebied in het onderste deel van Figuur 2. Hoe groter de waarde van de niet-centraliteits parameter γ, hoe groter de verschuiving naar rechts en dus hoe groter de power 1- β. Zoals te zien is hangt γ af van de grootte van het effect δ, de standaarddeviatie van de metingen σ, en de steekproefgroottes n_C en n_E. De eerste twee parameters liggen vast en zijn veelal onbekend in de planningsfase van een onderzoek. δ wordt vaak gekozen als het minimale klinische relevante effect grootte; een waarde van σ kan gekozen worden op basis van kennis van een expert of gerapporteerde waarden in de literatuur.

Om tot een gewenst power niveau te komen dienen de steekproefgroottes n_E en n_C zodanig gekozen te worden dat de kansdichtheid van de toetsingsgrootheid onder de alternatieve hypothese voldoende ver naar rechts verschuift. Uiteraard is hier software voor beschikbaar, zoals de functie normal.sample.size in SPLUS of het programma GPower. Figuur 3 geeft powerniveaus als functie van de totale steekproefgrootte waarbij is gekozen voor μ_C = 60 en drie verschillende waarden



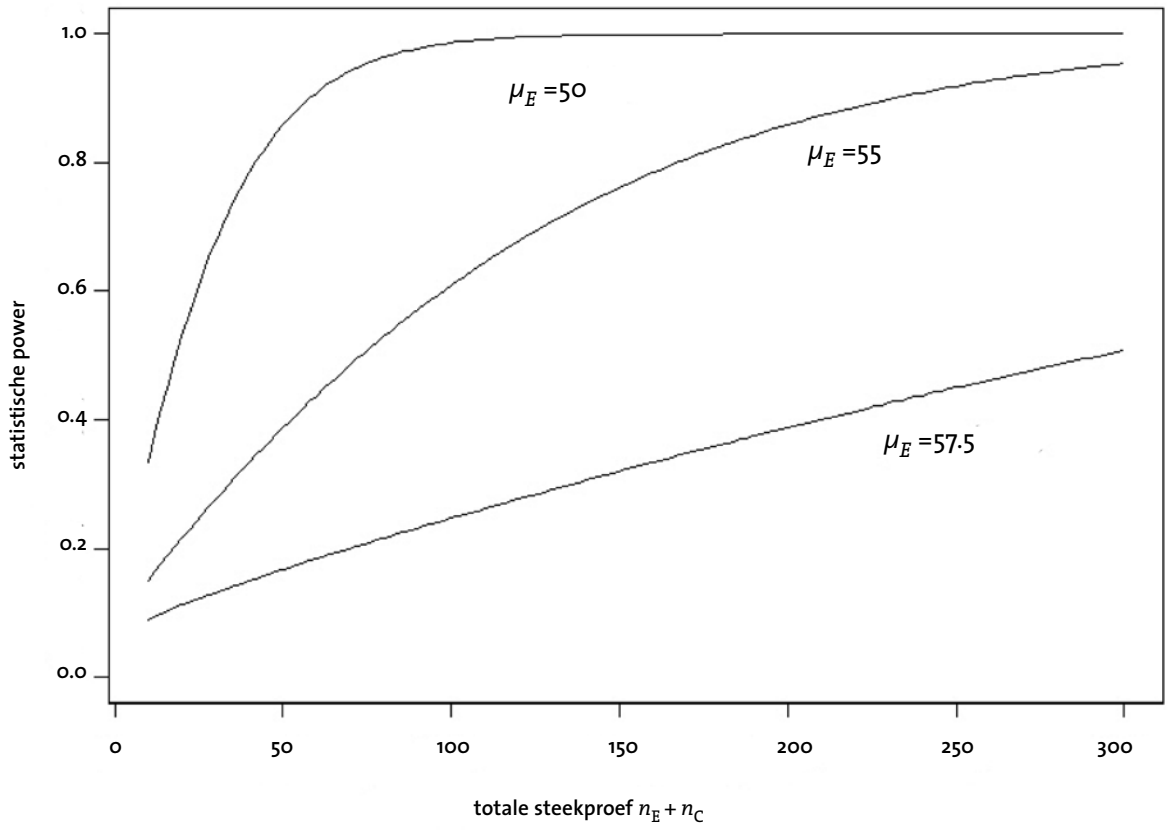
Figuur 2. Grafische presentatie van de kans op een type I fout α en de statistische power 1- β.

voor μ_E (μ_E = 57.5, μ_E = 55, en μ_E =50). De waarde van de standaarddeviatie σ is 13. Figuur 3 laat zien dat de power groter is naarmate het verschil in verwachtingen tussen beide experimentele condities groter is.

Figuur 3 illustreert waarom een berekening van de steekproefgrootte zo belangrijk is. Als men een verschil in verwachtingen van ten minste tien punten relevant vindt en men dit verschil met een kans van 90% wil kunnen aantonen dan zijn in totaal 58 personen nodig, 29 per conditie. Een te kleine steekproef betekent dat de power aanzienlijk lager is dan de gewenste 90%, terwijl vergroten van de steekproef nauwelijks tot een toename van de power leidt. Met andere woorden: in te kleine steekproeven is er een kans dat bestaande effecten niet aangetoond kunnen worden terwijl een te grote steekproef betekent dat er onnodig veel proefpersonen of –dieren moeten worden gebruikt.

Toepassing van power analyses in de praktijk

Bovenstaand voorbeeld is uiteraard een vereenvoudiging van hetgeen ik tijdens mijn adviesgesprekken aan toegepast onderzoekers tegenkom. Zoals in het voorbeeld hierboven is uitgelegd moeten de populatiewaarden van de effectgrootte δ en standaarddeviatie σ bekend zijn om de steekproefgrootte te kunnen bepalen voor



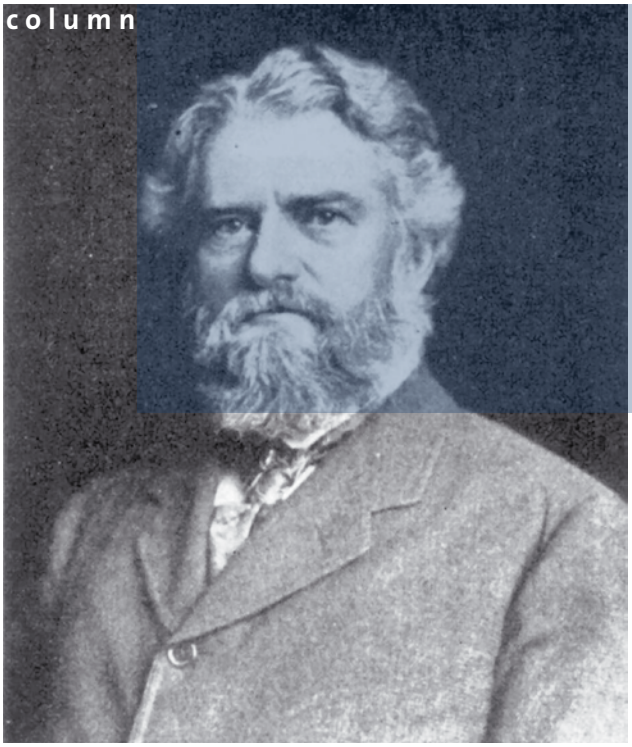
Figuur 3. Statistische power als functie van de steekproefgrootte.

een onderzoek waarin men twee onafhankelijke steekproeven met continue responsvariabelen. Voor ingewikkelder statistische modellen zijn veelal de waarden van meer dan twee parameters nodig. Dit impliceert een vicieuze cirkel: men wil een onderzoek uitvoeren om inzicht te kunnen krijgen in de waarden van bepaalde modelparameters, maar om het onderzoek goed op te kunnen zetten dienen deze waarden reeds in de planningsfase bekend te zijn. Dit probleem kan soms opgelost worden door schattingen van de modelparameters van reeds uitgevoerd soortgelijk onderzoek te gebruiken. Probleem hierbij is echter dat de waarden van deze parameters niet altijd netjes gerapporteerd worden in tijdschriftartikelen en dat niet altijd hetzelfde meetinstrument gebruikt wordt. Als alternatief kan een pilot studie gebruikt worden maar dit is praktisch gezien alleen haalbaar als deze niet al te veel tijd in beslag neemt. Wanneer de primaire uitkomstvariabele in het gebruikte voorbeeld het probleemgedrag twee jaar na het begin van het experiment is dan is het voor bijvoorbeeld een promovendus met een aanstelling van vier jaar nauwelijks haalbaar om een pilot studie uit te voeren. Een andere mogelijkheid is het toepassen van Bayesiaanse methoden waarbij een verdeling van de modelparameters wordt gebruikt in de power berekening. Momenteel is er in het veld van powerberekeningen en optimale proefopzetten veel aandacht voor het ontwikkelen van optimale designs die robuust zijn ten opzichte van incorrecte schattingen van modelparameters in de planningsfase van een onderzoek.

Een tweede probleem dat ik geregeld tegenkom is dat toegepast onderzoekers van te voren, vanuit bijvoorbeeld financiële of praktische overwegingen, reeds een steekproefgrootte hebben gekozen. Vervolgens willen zij een steekproefberekening uit te voeren zodanig dat de berekende steekproefgrootte de reeds door hen gekozen steekproefgrootte is. Uiteraard kunnen in zo'n berekening de waarden van α en β en de modelparameters zodanig gekozen worden dat het gewenste antwoord verkregen wordt, maar de gekozen waarden moeten altijd vanuit inhoudelijke gronden verdedigd kunnen worden. Een kleine steekproef is immers altijd te verantwoorden als men bereid is genoegen te nemen met slechts een kleine kans om een bepaald effect aan te kunnen tonen. Het ondersteunen van toegepast onderzoekers in hun power berekeningen betekent dus niet alleen het uitvoeren van complexe berekeningen maar ook het geven van een gedegen voorlichting.

Een laatste probleem dat ik regelmatig tegenkom is dat de problemen waarmee toegepast onderzoekers bij mij op adviesgesprek komen meestal veel complexer zijn dan het gebruikte voorbeeld en dat een kant-en-klare oplossingen niet altijd bestaan. Bij longitudinaal onderzoek gaat het niet alleen om het totaal aantal personen maar ook om het aantal metingen per persoon en de optimale plaatsing in de tijd. Bij toxicologische experimenten gaat het niet alleen om het aantal proefdieren maar ook om het aantal dosisniveaus. Poweranalyses voor dit soort problemen worden nog extra gecompliceerd wanneer de uitkomstvariabelen een niet-continue verdeling hebben. Daarnaast is het noodzakelijk om proefopzetten te ontwikkelen voor nieuw soort onderzoek, zoals bijvoorbeeld micro-array studies. Dit betekent dus dat er nog genoeg werk te verrichten is voor statistici die onderzoek doen naar optimale proefopzetten en power berekeningen.

Mirjam Moerbeek is universitair hoofddocent en onderzoeker aan de Universiteit Utrecht, departement Methoden en Technieken. Zij ontving in 2003 een VENI subsidie en in 2008 een VIDI subsidie van NWO. Doel van beide projecten was het ontwikkelen van power analyses en optimale proefopzetten met toepassingen in de sociale en gedragswetenschappen.
E-mail <M.Moerbeek@uu.nl>



Simon Newcomb, 1835 - 1909

Frank Benford, 1883 - 1948

DE WET VAN BENFORD

Fred Steutel

Getallen gedragen zich soms heel anders dan je zou verwachten. Een beroemd, maar toch niet zo bekend voorbeeld.

In 1881 merkte Simon Newcomb op dat in veel numerieke gegevens, zoals wiskundige tabellen en statistische waarnemingen, de voorste cijfers niet even vaak voorkwamen; vooral de 1 kwam veel vaker voor dan je zou verwachten. Alleen ingewijden wisten dit, en konden daar gebruik (misbruik) van maken. Feller (Vol. 2) begint zijn stukje over Benford met (voor Nederlanders verdietst): 'Een bekende toegepast wiskundige had veel succes met de weddenschap dat een willekeurig getal in de Enkhuizer Almanak met een cijfer kleiner dan 5 zou beginnen.' Maar, ter zake.

Voor de goede orde beginnen we met de definitie van 'voorste cijfer'. Ieder positief getal x kan worden geschreven in de vorm $x = a, b\ c\ \ldots \times 10^n$ met a een positief getal en n geheel. Het cijfer a heet wel het 'eerste significante cijfer' – verder aan te duiden met 'voorste cijfer' Het is duidelijk dat a de waarden 1, 2, ..., 9 kan aannemen. Newcomb kwam tot de conclusie dat voor de kansverdeling

van het voorste cijfer, beschouwd als een toevalsgrootheid D, geldt

$$P(D = d) = \text{Log}\,(d+1) - \text{Log}\,(d) \qquad (1)$$

voor d = 1, 2, ... , 9, waarbij Log de logaritme met grondtal 10 voorstelt. Dit betekent dat de kans op een 1 als voorste cijfer gelijk is aan Log 2 = 0, 3010, veel groter dan 1/ 9. P(D = 2) = Log 3 – Log 2 = 0,1761. De kans op 'kleiner dan 5' is 0,6990.

Formule (1) geldt natuurlijk niet voor alle getallenreeksen; zo beginnen alle telefoonnummers in Eindhoven met een 2. De vraag is dan natuurlijk voor welke, of wat voor, getallenreeksen deze verdeling wel geldt.

Het ligt voor de hand om te kijken naar stochastische grootheden, zeg X. Het voorste cijfer van X is d, als voor een natuurlijk getal n geldt dat

$$d.\,10^n \leq X < (d+1).10^n \qquad (2)$$

Hierbij kunnen we zonder bezwaar n=0 nemen, dus X tussen 1 en 10. Immers, het voorste cijfer verandert niet door vermenigvuldiging met een macht van 10. We blijven deze gereduceerde grootheid als X noteren. We kunnen nu formule (2) schrijven als

$$\text{Log}\, d \leq \text{Log}\, X < \text{Log}\,(d+1)$$

De kans dat deze ongelijkheid geldt, is precies het rechterlid van (1), als Log X (modulo 1) homogeen verdeeld is op (0,1), en dat is bij goede benadering het geval, als X het (gereduceerde) product is van een groot aantal toevalsgrootheden. In deze gevallen geldt dus inderdaad de wet van Newcomb.

Een heel andere vraag luidt: 'Wat is de kans dat een willekeurige rij eigenschap (1) heeft'. Je stuit dan op de moeilijkheid dat er geen kansverdeling bestaat op alle rijen, dat wil zeggen: een 'willekeurige rij' bestaat niet, is niet gedefinieerd. Dit probleem is met veel vernuft opgelost door Ted Hill, die regelmatig naar Nederland (VU) kwam. Zijn theorie zou ons te ver voeren, maar ook daar blijkt dat veel rijen aan de 'wet van Benford' gehoorzamen. Hij geeft ook wetten voor verdere cijfers, laat zien dat deze niet onafhankelijk zijn en dat de verre cijfers bij benadering homogeen verdeeld zijn op {0,1, ..., 9}. Er is zoveel vertrouwen in de wet van Benford dat wel gedacht is dat afwijkingen ervan bij getallen in belastingaangiften zouden wijzen op fraude.

Interessant is nog dat de Benford-eigenschap invariant is voor schaaltransformaties: als je getallen in een Benford-verzameling allemaal vermenigvuldigt met hetzelfde positieve getal, dan blijft de eigenschap behouden. Je kunt dit in eenvoudige gevallen controleren. Bij vermenigvuldiging met 2 gaan de voorste cijfers 1; 2; 3; 4; 5; 6; 7; 8; 9 over in, respectievelijk, 2 of 3; 4 of 5; 6 of 7; 8 of 9; 1; 1; 1; 1; 1, in precies de goede verhoudingen. Zo is de kans op een 1 na vermenigvuldiging met 2 weer Log 2.

De verdeling van voorste cijfers hangt natuurlijk af van het gebruikte talstelsel. Als we in plaats van het tientallig stelsel het n-tallig stelsel gebruiken (met n-1 positieve 'cijfers'), dan wordt de kans op een 1 gelijk aan $^n\log 2$. Voor n = 16 wordt P(D=1) = 0, 25; voor n = 4 is P(D=1) = 0,5. Voor n =2 is P(D=1) =1; immers in het tweetallig stelsel beginnen alle getallen met een 1.

Rest de vraag waarom het 'first digit problem' naar Benford is genoemd en niet naar Newcomb. Antwoord: de elektrotechnicus Frank Benford herontdekte de wet in 1938; toen de naam van Newcomb, na meer vijftig jaar te zijn vergeten, weer opdook, was de verwijzing naar Benford standaard geworden.

Fred Steutel is emeritus hoogleraar kansrekening aan de TU Eindhoven.
E-mail: <f.w.steutel@tue.nl>.

KLASSEGRENZEN

Gerrit Stemerdink

Wat zijn klassegrenzen?

Klassegrenzen zijn onlosmakelijk verbonden aan klassen. De eerste vraag die we daarom moeten stellen is: wat zijn klassen en waarom gebruiken we ze?

Klassen worden vaak gebruikt zonder er bij na te denken. We zeggen dat iemand van middelbare leeftijd is en bedoelen daarbij dat de leeftijd valt in de klasse van 40 tot en met 60 jaar. Die 40 en die 60 zijn dan klassegrenzen, daaronder en daarboven behoort een waarneming tot een andere klasse. Tenminste, dat nemen we maar aan: anderen hanteren een leeftijd vanaf 41 als middelbaar. Klassen lijden in de dagelijkse praktijk dikwijls aan een slordige definitie.

In de wetenschap moeten we echter streng zijn. Al kom je ook maar één dag tekort, als je nog geen 40 bent (of welk ander criterium ook) val je niet in die betreffende klasse.

Waarom delen we waarnemingen in klassen in?

Meestal wordt een klasse-indeling gemaakt om een beter overzicht te krijgen. Zo is een tabel met alle lengtes in cm tussen 120 en 210 door zijn lengte niet erg duidelijk. Ook voor sommige analysetechnieken kan een inperking noodzakelijk zijn. Een variantie-analyse met leeftijd als onafhankelijke is onwerkbaar als we alle leeftijden tussen 18 en 65 als afzonderlijke niveau's gebruiken. Een indeling in 3 of 4 klassen ligt hier meer voor de hand.

Hoe delen we waarnemingen in klassen in?

Bij het vaststellen van klassen moeten we veel aspecten in onze beslissing betrekken. Zo moet worden gekeken hoe de variabele tot stand is gekomen, wat we er mee willen gaan doen en wat er in andere onderzoeken mee is gebeurd.

Socialistisch-realistisch reliëf Boedapest. Foto Pieter Bosch

Dat lijkt triviaal, maar dat is het niet altijd. Bij een leeftijdsindeling moet goed gekeken worden naar het doel: voor gebruik bij werkloosheidscijfers kan een andere klasseindeling nodig zijn dan bij een demografische vergelijking of bij een voorspelling van kiesgedrag. Zo hanteert het Statistisch Jaarboek van het CBS bij de bevolkingsopbouw een leeftijdsindeling tot en met 19 jaar. Maar zo'n indeling is onbruikbaar voor onderwijsdoeleinden, waar men slechts tot en met 16 jaar leerplichtig is, of voor verkiezingsanalyses waarbij men vanaf 18 jaar stemgerechtigd is. Alleen al voor zoiets heel eenvoudigs als een indeling in leeftijdscategorieën zien we hier al direct drie verschillende mogelijkheden, afhankelijk van het gebruiksdoel. En er zullen er nog wel veel meer zijn.

Ook bij het tot stand komen van een variabele is trivialiteit ver te zoeken. Hebt u zich ooit afgevraagd waarom in allerlei medische publikaties een polsslag van 72 als normaal wordt genoemd? Waarom zou dat niet 71 of 73 zijn, of zelfs 72.8? Het antwoord is verrassend eenvoudig: vroeger beschikten artsen niet altijd over betrouwbare horloges, laat staan dat die secondenwijzers hadden. De polsslag werd daarom gemeten met een klein zandlopertje dat 15 seconden mat. Vervolgens werd dat getal met 4 vermenigvuldigd om het aantal slagen per minuut te krijgen waardoor de polsslag altijd een veelvoud van 4 werd. Daar komt die 72 als *normale* polsslag vandaan. Als we in een onderzoek de polsslag in klassen gaan indelen zonder daarbij te weten dat de waarden slechts veelvouden van 4 zullen zijn, krijgen we problemen. Bij een indeling als 66 - 70, 71 - 75, 76 - 80, 81 - 85 etc. zullen we in sommige klassen 2 waarnemingspunten hebben en in andere slechts één! Dit verschijnsel komt meer voor: berucht is een secundaire analyse op lengtegegevens die oorspronkelijk in inches waren gemeten. De omrekening naar centimeters had desastreuze gevolgen voor de klasse-indeling in centimeters!

Gelijke breedte of gelijke aantallen?

Bij het indelen in klassen kunnen twee hoofdprincipe's worden onderscheiden. Men kan klassen van gelijke breedte maken, of men kan trachten in iedere klasse evenveel waarnemingen te krijgen. Dat lijkt simpel, maar toch...

Neem de bekende schoolcijfers van 1 tot 10. We laten voor het gemak even cijfers als 5.7 of 8+ buiten beschouwing. Hoe gaat u indelen: 1-2, 3-4 etc? Realiseert u zich dan dat de onvoldoende 5 met de voldoende 6 in één klasse terechtkomt? Logischer lijkt het een indeling te maken waarbij de cesuur voldoende/onvoldoende wordt gehandhaafd. Trouwens: weet u zeker dat de nul niet is gebruikt?

Ook als we klassen willen maken die gelijke aantallen bevatten moeten we oppassen. Loopt een klassegrens niet net door een top in de verdeling? En zijn de klassegrenzen wel 'sporend' met dat wat met de variabele is gemeten? Ik verwijs maar weer naar de schoolcijfers: we willen waarschijnlijk toch de 5 en de 6 in verschillende klassen laten terechtkomen.

Klassegenzen moeten eenduidig gedefinieerd worden

Ook bij het vaststellen van een klassegrens moet goed naar de variabele worden gekeken. Men moet rekening houden met de waarden die de variabele kan aannemen. Het maakt nogal verschil of we alleen de gehele getallen 1-10 als schoolcijfer geven, of dat ook met 1 of zelfs 2 decimalen rekening moet worden gehouden. Zijn het gehele getallen, dan kan een klasse als 6 - 7 voorkomen, maar met 2 decimalen zal diezelfde klasse 5.50 - 7.49 zijn.

Financiën zijn helemaal een ramp!

Als de variabele die we willen indelen een financiële achtergrond heeft is extreme voorzichtigheid geboden. Is er de variabele inkomen en is dat gemeten aan de hand van een loonstrookje dan

valt het wel mee. De waarden zullen liggen tussen € 1000 en € 20.000 per maand en we kunnen rustig indelen. Uiteraard wel even opletten of er niet enkele toppen in de oorspronkelijke cijfers zitten. Daar dienen we dan rekening mee te houden.

Anders wordt het als we vragen naar een bedrag dat gegeven is voor een goed doel, bijvoorbeeld de jaarlijkse bijdrage aan een kerkgenootschap. Daar kunnen we vooral 'ronde' bedragen verwachten, de kans is klein dat er bedragen als € 23 of € 505 voorkomen. Maar kijk hier ook eerst goed naar de toppen in de oorspronkelijke bedragen alvorens klassegrenzen vast te stellen. En houd daarbij rekening met de neiging, tenminste tot 1 januari 2002, om in veelvouden van 25 te denken. Nederland had tenslotte kwartjes, rijksdaalders en bankbiljetten van fl. 25 en fl. 250.

Een tegengestelde conclusie geldt voor prijzen: daar kunnen we juist wél bedragen als € 4,95 verwachten. Waar bij giften een klasse van € 5,00 tot € 9,99 een goed idee kan zijn is het dat voor prijzen zeker niet!

Let op met andere tijden en andere landen of culturen!

Vergelijken van gegevens uit verschillende onderzoeken is vragen om moeilijkheden.

Neem een onderzoek met schoolcijfers uit Nederland, Duitsland en de USA. Nederlandse cijfers lopen van 1-10, van zeer slecht tot uitmuntend. Maar Duitse cijfers lopen van 1-5, waarbij de slecht/goed volgorde juist omgekeerd is! En de amerikanen geven een A voor de beste prestatie en een E voor de slechtste. U begrijpt het al, problemen te over, nog afgezien van het probleem van de feitelijke waardering: is een Nederlandse 7 of 8 te vergelijken met een Duitse 2 of een Amerikaanse B?

Ook vergelijkingen met historische gegevens leveren verrassingen. Vóór de mammoet-wet hadden we technische- en huishoudscholen, ulo/mulo en hbs/gymnasium als de drie belangrijkste vormen van voortgezet onderwijs. Hoe plaatsen we nu nà de invoering van de mammoet-wet de havo in dit rijtje? Toevoegen aan het vwo, óf is het een aparte klasse tussen mavo en vwo? Het wordt nog erger als we naar de perceptie kijken: de mulo had in de jaren twintig van de 20e eeuw zo ongeveer dezelfde status als de hbs kort voor het afschaffen van dat schooltype.

Pas op voor automatisme!

Er is al gewaarschuwd voor klassegrenzen die dwars door toppen lopen, of die geen recht doen aan de achterliggende aard van de gegevens. Automatisme is daarom altijd uit den boze.

Vooral bij het gebruik van computerprogramma's komen problemen om de hoek kijken. Er zijn programma's die de mogelijkheid bieden een variabele automatisch in een aantal klassen in te delen. Gemakkelijk als men bijvoorbeeld 5 klassen met gelijke breedte wil hebben. Maar let op: programma's nemen daarvoor soms simpelweg de hoogste en de laagste waarneming en delen dat interval in 5 gelijke stukken. Dat komt vaker voor dan u denkt. Bij een eenvoudige frequentietabel zoals bijvoorbeeld SAS of SPSS die maken gebeurt dat soms vanzelf als de variabele een zeer groot bereik heeft. Ziet u het al voor zich bij een lengtemeting: klassen als 172-183 cm? En weet u overigens zeker dat die uiterste waarden waarop deze indeling berust reëel zijn? Waren dat geen uitbijters?

Hoe eenvoudig en zelfs banaal het probleem van klassegrenzen ook moge lijken: het is gecompliceerd. Men dient met zeer veel aspecten rekening te houden en daarbij constant op zijn hoede te zijn.

Gerrit Stemerdink is redacteur van STAtOR.
E-mail: <gjstemerdink@hotmail.com>

PEILINGEN

Jelke Bethlehem

De Amerikaanse verkiezingen

De campagne voor de Amerikaanse verkiezingen werd door vele opiniepeilers op de voet gevolgd. Vrijwel dagelijks kwamen er nieuwe peilingen uit. Na eerdere debacles waren de Amerikaanse onderzoeksbureaus wel wat voorzichtiger geworden met hun prognoses. Toch ging nog steeds niet alles goed. Zo meldt de NOS op 8 november dat de Amerikaanse opiniepeilers in de aanloop van de presidentsverkiezingen een belangrijke groep hebben overgeslagen. Ze peilden alleen onder mensen met een vaste telefoon. Inmiddels heeft echter 20 procent van de Amerikaanse kiezers alleen een mobieltje. Zij stemden opvallend vaak op Obama. Dat blijkt uit een nationaal verkiezingsonderzoek. Vooral dertigers met alleen gsm stemden vaak op Obama: 63 procent tegen 51 procent van de dertigers met een vaste telefoon. De peilers wisten niet dat inmiddels zoveel mensen geen vaste lijn meer hebben en dachten ook dat het voor het stemgedrag geen verschil maakte.

Toch wel een merkwaardig bericht. Kennelijk is een belangrijke groep met een ander stemgedrag overgeslagen. Maar de vraag die opkomt is dan: als dit het geval is, hoe komt het dan dat de opiniepeilers de uitslag zo goed konden voorspellen? Zo voorspelde CNN dat 52% op Obama zou stemmen en 44% op McCain. Uiteindelijk stemde 53% op Obama en 46% op McCain. Als het bericht cor-

rect is dan zou de voorspelling dus lager moeten zijn uitgevallen. De peilers misten immers de op Obama stemmende jongeren met een mobieltje. De conclusie lijkt voor de hand te liggen dat minstens nog iets mis moet zijn gegaan in de peilingen, maar dan met een tegengesteld effect. Helaas gaat het bericht van de NOS hier niet op in.

Over slecht onderzoek

Nebahat Albayrak moet de nieuwe burgemeester van Rotterdam worden. Dat was de verrassende uitkomst van het Rotterdamse burgemeestersreferendum, georganiseerd door *AD Rotterdams Dagblad* in september 2008. Ruim 12.500 inwoners van de Maasstad brachten hun stem uit. Tientallen verslaggevers van de krant en ruim honderd vrijwilligers waren woensdag 24 september de hele dag in touw om Rotterdammers te laten stemmen. De inwoners konden hun favoriet kiezen uit acht kandidaten, die in politieke kringen werden genoemd als kanshebber om Ivo Opstelten op te volgen als burgervader. De kandidaten in het referendum waren behalve de genoemde drie: Geert Dales, Steven van Eijck, Hans de Boer, Robin Linschoten en Lodewijk de Waal. *AD Rotterdams Dagblad* besloot het referendum te organiseren, nadat de gemeenteraad had besloten dat de verkiezing van de nieuwe burgemeester niet via een officieel referendum zou plaatsvinden.

Het *AD* suggereerde in zijn berichtgeving dat het om een referendum ging. Dat was het natuurlijk niet als niet elke inwoner van Rotterdam formeel een oproep had gehad om zijn stem uit te brengen. De kop van het artikel ('Rotterdammers willen Albayrak') suggereerde dat dit onderzoek representatief was. Ook de hoofdredacteur zei op de radio (NOS Journaal) dat het een representatief onderzoek was omdat mensen uit alle wijken hadden meegedaan. Maar helaas was het geen representatief onderzoek. De opzet en uitvoering rammelen, methodologisch gezien, aan alle kanten.

Als we de berichten van de krant mogen geloven, zijn tientallen verslaggevers en honderd vrijwilligers de straat op gegaan om de mensen te laten stemmen. Dat betekent dat alleen mensen 'op straat' aan het onderzoek konden meedoen. Het is ook nog maar de vraag of niet-lezers van het *AD* voldoende op de hoogte waren van deze actie. Ook niet-Rotterdammers konden hun stem uitbrengen. En er was niets dat mensen weerhield om meerdere malen hun stem te brengen.

Dit onderzoek was een typisch voorbeeld van zelfselectie. De onderzoeker heeft dan geen enkele invloed op het selectiemechanisme van de steekproef. De steekproeftheorie leert dat je bij dit soort onderzoek een groot risico loopt een verkeerde conclusie te trekken.

Op filmbeelden van het *AD* blijkt dat de interviewers mee keken bij het invullen van de formulieren. Er is dus geen enkele sprake van privacy geweest. In dit soort situaties aarzelen mensen vaak om hun mening te geven. In plaats daarvan geven ze sociaalwenselijke antwoorden.

Samenvattend kan worden gesteld dat het geen representatief was. De kop 'Rotterdammers willen Albayrak' dekte de lading niet. En helaas namen vele andere media de conclusies zonder meer over zonder zich af te vragen of het nu wel of niet om een goed onderzoek ging.

Jelke Bethlehem is senior-methodoloog bij het Centraal Bureau voor de Statistiek in Den Haag en hoogleraar aan de de Faculteit der economische Wetenschappen en Econometrie van de Universiteit van Amsterdam. E-mail: <j.bethlehem@cbs.nl>.

EEN SOMMETJE

Fred Steutel

Emeritus hoogleraar Roel Doornbos, oud-voorzitter van de VVS en oud-hoofdredacteur van Statistica Neerlandica stelde mij de volgende vraag, die ik hier in mijn eigen woorden weergeef. Doornbos laat mij nog weten dat het probleem voorkomt in de afscheidsrede van prof. dr. I.S. Herschberg,

De burgemeester van Amsterdam wil kennismaken met de langste bewoner van zijn stad. Daartoe laat hij alle Amsterdammers, zeg één miljoen, opdraven en een lange rij vormen, genummerd 1 tot en met 1000.000. De burgemeester loopt de rij langs van 1 naar 1.000.000. Nummer 1 loopt met hem mee tot er een grotere Amsterdammer wordt bereikt;nummer 1 krijgt een biljet van 100 euro en gaat zijns weegs. De grotere loopt mee tot een nog grotere wordt gevonden en vertrekt ook met 100 euro. Dit gaat zo door tot de grotere de allergrootste blijkt te zijn. Ook die wordt beloond met een biljet van 100 euro, gaat naar huis met een grootste Amsterdammercertificaat.

De vraag tenslotte: hoeveel 100-eurobiljetten moet de burgemeester (gemiddeld) meenemen? Hoewel het gebruik van wiskunde in STAtOR niet wordt aanbevolen, is het probleempje zo aansprekend en de wiskunde zo eenvoudig, dat ik het er op waag.

EEN ANTWOORD

We vervangen het miljoen door n en proberen door inductie het antwoord te vinden. Laat K(n) het aantal benodigde biljetten voorstellen en E(n) de verwachting van K(n). Het is duidelijk dat K(1) = 1 met kans 1. Door toevoeging van een (n+1)-ste Amsterdammer aan de rij verandert er meestal niets; alleen als de toegevoegde Amsterdammer de grootste is van alle n+1, neemt het benodigde aantal hoeden met één toe. De kans hierop is 1/(n+1). Voor de verwachting E(n) geldt dan

$$E(n+1) = n / (n+1)\, E(n) + 1/ (n+1)\, \{E(n) + 1\} \qquad (1)$$

Dit betekent dat $E(n+1) = E(n) + 1/(n+1)$, en dus

$$E(n) = 1 + 1/2 + 1/3 + \ldots + 1/n \qquad (2)$$

Voor het gemiddelde aantal biljetten dat de burgemeester van Amsterdam moet betalen geldt dus (hier zijn tabellen of formules voor):

$$E(1000.000) = 1 + 1/2 + \ldots + 1/\, 1.000.000 = 14{,}39.$$

Een beetje verrassend: kleiner dan verwacht, maar de burgemeester zal toch een paar biljetten meer moeten meenemen dan 14 of 15. Maar, hoeveel meer? Om die vraag te beantwoorden hebben we ook de variantie van K(n) nodig. Op een soortgelijke manier als hier boven vinden we

$$\text{var } K(n) = 1+ 1/2 + 1/3 + \ldots +1/n - (1+ 1/4 + 1/9 + \ldots + 1/n^2). \qquad (3)$$

Dit levert voor de Amsterdamse burgemeester var

K(1.000.000) = 12,75. Als we de strenge eis stellen van 'zes sigma', dan moeten 36 biljetten genoeg zijn. Maar we kunnen nog meer uitrekenen. Met dezelfde methode die tot vergelijking (1) leidt, kunnen we de kansengenererende functie Pn van K(n) vinden:

$$P_n(z) = z/n! + P(K(n)=2)\,z^2 + P(K(n)=3)\,z^3 + \ldots + P(K(n)=n)\,z^n.$$

Het blijkt dat

$$P_n(z) = z/1 \cdot (1+z)/2 \cdot (2+z)/3 \cdot \ldots \cdot (n-1+z)/n.$$

Dit wijst erop dat K(n) verdeeld is als de som is van n onafhankelijke grootheden X(k), die nul of één zijn met kansen 1 - 1/ k, resp. 1/ k:

$$K(n) = X(1) + X(2) + \ldots + X(n), \qquad (4)$$

met $P(X(k) = 1) = 1 - P(X(k) = 0) = 1/k$.

Uit (4) volgt nu eenvoudig de verwachting E(n), en als we geloven dat de X(k) onafhankelijk zijn, ook var K(n) = var X(1) + ... + var X(n). Immers, $E\,X(k) = 1/k$ en $\mathrm{var}\,X(k) = 1/k - 1/k^2$.

Het blijft lastig om voor de burgemeester P(K(n) > k) uit te rekenen voor bijvoorbeeld k= 20, k= 25 en k= 30. Met wat moeite en wat hulp vinden we dat 20, 25, resp. 30 biljetten genoeg zijn met kansen, respectievelijk, 0, 950; 0, 9977 en 0, 99996. Voor deze rekenpartij kreeg ik de onmisbare hulp van collega Jos Jansen. De standaard normale benadering leidt voor k = 25 tot 0, 9987. De kans dat de burgemeester meer dan 30 honderd-eurobiljetten nodig heeft, is buitensporig klein.

Fred Steutel is emeritus hoogleraar kansrekening aan de TU Eindhoven.
E-mail: <f.w.steutel@tue.nl>

Hoe kijk je naar een woning? Iedereen ziet een woning anders. Een architect kijkt naar het ontwerp, een moeder naar de afstand tot een basisschool. Ook zijn er mensen die beroepshalve willen weten hoeveel woningen er eigenlijk zijn in Nederland, hoe groot de gezinnen zijn die er wonen en wanneer een woning aan vervanging toe is. Het CBS levert deze cijfers en vele andere feiten. Uit onze onderzoeken blijkt dat een gezin 100 jaar geleden gemiddeld meer dan 4 kinderen kreeg en nu minder dan 2. Onze cijfers laten zien dat gezinnen met kinderen minder vaak verhuizen dan paren zonder kinderen en dat huurders de overstap naar een andere woning eerder maken dan huiseigenaren. Door het werk van onze medewerkers weten we bijvoorbeeld dat in Nederland 83 procent van alle huishoudens toegang heeft tot internet en er elke dag ongeveer 500 kinderen worden geboren, dat bijna 30 procent van de bevallingen thuis plaatsvindt en 40 procent van de baby's een ongetrouwde moeder heeft. Ook weten we dat huis- en tuingereedschap de laatste twee jaar zo'n 3 procent duurder is geworden en melk, kaas en eieren meer dan 18 procent. Een enorme hoeveelheid cijfers is onze grondstof, waardevolle en betrouwbare informatie ons eindproduct. Betrouwbare informatie, die bijna dagelijks de krant haalt en waarop de overheid en het bedrijfsleven belangrijke beslissingen funderen. Bij het CBS registreren en onderzoeken we alles wat voor onze samenleving van belang is. We rapporteren niet alleen over de Nederlandse Economie, de volksgezondheid, het onderwijs en de werkgelegenheid, maar ook over demografische ontwikkelingen en verkeersveiligheid. Meer weten over onze vacatures? Kijk op www.werkenbijhetcbs.nl.

Centraal Bureau voor de Statistiek

www.werkenbijhetcbs.nl

ORTEC

Dat wil niet zeggen dat je van Mars moet komen

Als afgestudeerde wil je graag direct aan de slag. Bij ORTEC hoef je hier niet lang op te wachten. Je wordt direct op projecten ingezet en krijgt veel eigen verantwoordelijkheid. Bij ORTEC werken veel studenten. Sommigen schrijven bij ons een afstudeerscriptie, anderen werken enkele dagen per week als studentassistent.

Maar je staat er nooit alleen voor. Je kunt rekenen op de expertise van je collega's: stuk voor stuk experts op het gebied van complexe optimalisatievraagstukken in diverse logistieke en financiële sectoren. Hoogopgeleide, veelal jonge mensen die weten wat ze doen en jou naar een hoger niveau zullen brengen. Samen met je collega's help je klanten gefundeerde beslissingen te nemen. Dit doe je met gebruik van wiskundige modellen en het toepassen van simulatie- en optimalisatietechnieken.

Vanwege onze constante groei is ORTEC altijd op zoek naar enthousiaste studenten en afgestudeerden die de ruimte zoeken om zich te ontwikkelen en willen bijdragen aan de volgende generatie optimalisatietechnologie.

Hiervoor denken we aan bèta's in de studierichtingen:

- Econometrie
- Operationele Research
- Informatica
- Wiskunde

Voor vacatures en afstudeerplaatsen kun je kijken op www.ortec.com. Zit jouw ideale functie of afstudeerplek er niet bij, stuur dan een open sollicitatie of scriptievoorstel naar recruitment@ortec.com.