PRAKTIJKREEKS CREDIT MANAGEMENT

Kansrekening en statistiek voor credit managers

Alles geordend naar Maat & Getal

André J.M. Koch

VVCM
VERENIGING VOOR CREDIT MANAGEMENT

Kansrekening en statistiek voor credit managers

Alles geordend naar Maat & Getal

André J.M. Koch

VVCM
VERENIGING VOOR CREDIT MANAGEMENT

De Praktijkreeks Credit Management is een uitgave van de Vereniging Voor Credit Management (VVCM). De praktijkreeks biedt vakinhoudelijke ondersteuning aan credit managers en debiteurenbeheerders. In elk deel staat een onderwerp centraal uit het vakgebied credit management en debiteurenbeheer. De thema's worden op een praktische en toegankelijke manier behandeld zodat de informatie direct toepasbaar is bij de dagelijkse werkzaamheden.

Eerste druk, november 2014

CIP-gegevens
ISBN: 9789082305104

Vormgeving H. Tijbosch

Adres Computerweg 11
3542 DP Utrecht
Telefoon (0346) 55 80 50
www.vvcm.nl

Voorwoord

Al sinds haar oprichting in 1990 is de VVCM, de Vereniging Voor Credit Management, in Nederland actief als dé brancheorganisatie voor credit managers. Door het geven van opleidingen, het organiseren van kennissessies en het uitgeven van het vakblad 'De Credit Manager' wordt een toenemend aantal leden in staat gesteld om hun expertise te vergroten en daarmee nog meer bij te dragen aan de continuïteit en de winstgevendheid van de bedrijven waar zij voor werken.

De VVCM heeft de ambitie om uit te groeien tot hét kennisinstituut voor credit management en opent met het boekje dat u nu leest, de VVCM Praktijkreeks voor Credit Management. Met dit boekje en de volgende edities in deze reeks wil de VVCM credit managers praktijkgerichte informatie en tools aanreiken, waarmee zij hun vak nog professioneler kunnen uitoefenen.

André Koch (1961) is de schrijver van deze eerste uitgave in de VVCM Praktijkreeks voor Credit Management. Hij is oprichter en partner van Stachanov Solutions & Services BV, een bedrijf in Amsterdam dat is gespecialiseerd in data-analyse en modellenbouw. Hij doceert risicomanagement aan de Nyenrode. Business Universiteit. Hij geeft pakkende voorbeelden van de succesvolle inzet van statistiek en kansberekening in de dagelijkse praktijk van de credit manager.
Ik wens u met dit boekje veel leesplezier en vooral veel succes in uw werk.

Martin van der Hoek,
Voorzitter Vereniging Voor Credit Management

Inhoud

Inleiding

Hoewel de wereld zich in chaos en onzekerheid lijkt te presenteren, zijn wij er als mensen traditioneel toch van overtuigd dat er onderliggend een diepere structuur is. De Bijbel legt de wijze koning Salomon de uitspraak 'Maar Gij hebt alles naar maat en getal en gewicht geordend' in de mond, waarmee verwezen wordt naar een plan en regelmaat die aan onze warrige wereld ten grondslag liggen. In vele filosofieën en geloven komt deze gedachte naar voren: patronen, natuurwetten en getallen zijn de vingerafdrukken van het goddelijke en laten de volmaaktheid van de schepping zien.

Nu is de alledaagse werkelijkheid van de credit manager ver verwijderd van abstracte filosofieën over de schepping. Hij is vooral bezig met concrete zaken als debiteurenbeleid en het verbeteren van werkkapitaal en kasstromen. Risico, onzekerheid, onduidelijkheid en beperkte informatie zijn de beren waartegen moet worden gevochten. De uitdaging is nu om uit deze chaos en onzekerheid structuren en informatie te destilleren die leiden tot betere bedrijfsbeslissingen. Dit is nu het terrein van de statistiek en kansrekening!

Met statistiek en kansrekening zoeken we naar structuren en patronen in data en proberen we op een rationele wijze om te gaan met onzekerheid en risico. We scheppen daarmee orde in de chaos en brengen onze bedrijfsdoelstellingen dichterbij. Kansrekening en statistiek zijn onderdeel van de toegepaste wiskunde, een vak dat niet bij iedereen goede herinneringen oproept. Sommige bedrijfsprofessionals zijn er zelfs trots op dat ze geen jota van statistiek begrijpen...

In dit boekje leggen we belangrijkste beginselen van statistiek en kansrekening uit. Nu zijn er al veel van dit soort publicaties beschikbaar, maar de Vereniging voor Credit Management (VVCM) hecht eraan deze vakgebieden te belichten vanuit het perspectief van het credit management. De credit managers zijn eerst en vooral mensen van de praktijk, die vraagstukken waarmee ze in hun werkomgeving worden geconfronteerd willen oplossen.

Dit boekje onderscheidt zich daarom door een praktische insteek en veel praktijkvoorbeelden gericht op credit management. Dit betekent ook dat ervoor gekozen is om niet in detail uit te leggen hoe bepaalde begrippen uit de statistiek worden berekend. MS Excel is het standaardplatform voor bedrijven om dit soort berekeningen uit te voeren. Bevrijd van deze rekenlast kunnen we daarom de aandacht richten op het begrijpen en interpreteren van de gegevens en uitkomsten. Dat is winst.

Daarnaast leggen we kort uit hoe de modellen en begrippen in Excel kunnen worden uitgewerkt. Om niet te verzanden in de vele versies, taaluitvoeringen, lokale configuraties van Excel is ervoor gekozen dit summier te doen door aan te geven welke formules te gebruiken. De ervaren Excel-gebruiker zal met de gegeven aanwijzingen veelal zijn eigen weg weten te vinden. De praktische insteek komt ook terug in de afsluiting van de relevante hoofdstukken en paragrafen met de terugkerende vragen: 'Wat heeft de credit manager hieraan?' en 'Hoe pak ik het aan in Excel?'

André Koch
Stachanov Solutions & Services BV
andre@stachanov.com

Kansrekening

Kans of toeval speelt een grote rol in onze samenleving en in ons leven. Aan de ene kant roept kans een passief gevoel van machteloosheid op, waarbij wij, als eenvoudige mensen, ons overgeleverd voelen aan 'het lot' dat wordt beschikt door God of een hogere macht die ver boven ons staat. Van de andere kant heeft kans ook de betekenis van nieuwe mogelijkheden op succes die ons zomaar in de schoot kunnen worden geworpen. Dit is het lot uit de loterij, het 'gelukje' waar we niet op hadden gerekend. Of ons nu het geluk toelacht, of dat we gewoon domme pech hebben, in beide gevallen heeft kans te maken met onzekerheid. Het niet weten wat de toekomst gaat brengen, is een van de charmantere aspecten van het leven, maar kan ook een bron zijn van nervositeit en twijfel. Mensen proberen grip te krijgen op deze onzekerheid op allerlei manieren, via hun levensovertuiging, rituelen zoals het spugen op de dobbelsteen voor het rollen of door analyse en praktisch handelen.

Dit boekje heeft geen filosofische pretenties en richt zich uitsluitend op het begrijpen van kans en toeval in alledaagse praktische problemen waarbij we keuzes maken of beslissingen moeten nemen. Je kunt immers proberen om na bestudering van het probleem je kansen in te schatten en op grond van deze inschatting een beslissing te nemen. De kansrekening of waarschijnlijkheidsleer is een onderdeel van de wiskunde. De kansrekening helpt ons beslissingen te nemen in praktische problemen. Hoeveel budget moet ik aanvragen voor de bouw van een nieuwe brug? Hoeveel kassapersoneel moet ik inroosteren voor koopavond? Hoeveel broodjes moeten worden ingekocht voor de kantine? Kansrekening is een praktisch onderdeel van de wiskunde. Het is dan ook toegepaste wiskunde, omdat we de waarschijnlijkheidsleer gebruiken bij alledaagse beslissingen. Dit boekje gaat over het beter nemen van praktische, zakelijke beslissingen waarbij kans een rol speelt.

Hoe te rekenen met kansen?

Dit is een simpele vraag, maar het antwoord is niet zo eenvoudig. De grondslagen van de wiskunde zijn voor een groot deel in het Oude Griekenland gelegd. Merkwaardig genoeg hebben de Grieken kansrekening links laten liggen. De geschiedenis door bleven risico en onzekerheid ongrijpbaar voor de mens die in zijn aardse bestaan overgeleverd was aan een almachtige maar vaak ook ondoorgrondelijke en soms wispelturige God. Het antwoord was bidden en leren het lot te aanvaarden. In deze visie, die tot aan de moderne tijd gemeengoed bleef, is er geen ruimte voor een rationele bepaling en meting van risico.

Na de Middeleeuwen kwam hier verandering in. Anders dan in onze tijd, waarbij we wiskundigen vaak afserveren als wereldvreemde nerds, stonden statistiek en kansberekening toen midden in de samenleving. De wereld van kaartspelers en gokkers die professioneel met kansberekening te maken hadden, verbond zich met de hogere maatschappelijke lagen waar het beoefenen van wiskunde een gepast en gewaardeerd tijdverdrijf was. Net zoals er literaire of muzieksalons waren, had men ook wiskundige salons. Kansrekening was een geliefd onderwerp omdat het aansloot bij raadsels en problemen die men kende uit het kaart- en dobbelspel: hoe groot is de kans om tweemaal zes te gooien, of om een harten heer te trekken? Het antwoord lag in breuken en procenten. De kans om zes te gooien met één dobbelsteen, is één gedeeld door zes (1/6) enzovoorts. De kans om tweemaal zes te gooien is:

$$\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

Voor ons is dit gesneden koek, maar in de Vroegmoderne Tijd waren dit spectaculaire inzichten: kansen konden opeens worden uitgerekend en bepaald.

Kansrekening is een van de meer toegankelijke terreinen van de wiskunde, juist omdat je de problematiek vaak kent uit het dagelijkse leven. Voor het lezen en bestuderen van dit boekje is geen sterke achtergrond in wiskunde vereist. En hoewel wiskundige formules handig zijn om zaken duidelijk op te schrijven, schrikken ze vaak af. Ook zonder formeel genoteerde formules kan men een heel eind komen. Wel wordt verondersteld dat de lezer basisvaardigheden heeft in het gebruik van het Microsoft Excel spreadsheetprogramma De aanpak is steeds dat een praktisch probleem wordt beschreven, er een analyse wordt uitgevoerd met gebruikmaking van een snufje wiskundige theorie en er vervolgens een model wordt gebouwd in MS Excel om zo een gefundeerde beslissing te kunnen nemen in zakelijke vraagstukken. Aan de slag!

Voorbeelden uit de praktijk

Eén van de meest eenvoudige kansexperimenten die we ons kunnen voorstellen, is het tossen met een muntstuk. Kop of munt is een staande uitdrukking in onze taal en de toss is bekend uit de sportwereld voor het bepalen van de aftrap en speelrichting. Er zijn maar twee mogelijke uitkomsten: kop of munt.

In de kansrekening wordt echter niet volstaan met het benoemen van de mogelijkheden, in dit geval kop of munt, maar wil men ook uitrekenen hoe groot een kans is. Deze kans drukt men meestal uit in procenten. De kans om met een zuivere munt kop te gooien is 50% of 1/2. De kans op munt is ook 50%, omdat

we weten dat de kans op kop gelijk is aan de kans op munt. De optelsom van de kansen van alle mogelijke uitkomsten is altijd 100% of 1. In het geval van de munt zijn er twee mogelijke uitkomsten, kop en munt, met ieder een kans van 50%. De optelsom 50% kop plus 50% munt levert inderdaad 100% op.
Men kan de vraag stellen: hoe weet men dat de kans op kop 50% is? Eén manier om deze vraag te beantwoorden is door te wijzen op de omstandigheden van het kansexperiment: het gaat om een zuivere munt en de fysieke kenmerken van deze munt maken dat de kans op kop even groot is als de kans op munt. Men kan dus beredeneren dat de kans op kop 50% moet zijn.

Deze kans kan echter ook proefondervindelijk worden vastgesteld door bijvoorbeeld honderdmaal een munt op te gooien en bij te houden hoe vaak de muntzijde en hoe vaak de beeldenaar boven ligt. De uitkomst van dit experiment is waarschijnlijk dat men ongeveer vijftig maal munt gooit en vijftig maal kop. Helemaal zeker is dit niet te zeggen. Immers, het zou zomaar kunnen dat er honderdmaal achter elkaar kop wordt gegooid en munt in het geheel niet voorkomt. Deze kans is zeer klein, maar toch... Echter, als dit experiment een aantal malen wordt herhaald, zal men zien dat de gemiddelde uitkomst tendeert naar vijftig kop en vijftig maal munt.

Kop of Munt	Aantallen	%
Kop	49	49%
Munt	51	51%
Totaal	100	100%

Tabel 1: Aantal malen kop en munt.

Belangrijk is te zien dat kansen enerzijds kunnen worden beredeneerd en uitgerekend, maar anderzijds ook proefondervindelijk kunnen worden vastgesteld. Bij het tossen van het muntstuk zijn beide methodes toegepast. Het proefondervindelijk vaststellen van kansen kan ook worden gedaan door een beroep te doen op historische gegevens.

Indien een bedrijf bijhoudt hoeveel oninbare vorderingen voorkomen per duizend debiteuren, kan men iets zeggen over de kans op kredietuitval. Vaak wordt een gemiddelde uit het verleden gebruikt als verwachtingswaarde voor de toekomst. Door het totaal aantal oninbare vorderingen te tellen en te delen door het aantal leveringen, kan men de kans op een debiteurenrisico berekenen. Deze aanpak veronderstelt dat het verleden de toekomst voorspelt. Dit is een sterke aanname, want zoals we allemaal weten: in het verleden behaalde resultaten bieden geen garantie voor de toekomst!

Jaar	Aantallen	%
1996	12	1,2%
1997	9	0,9%
1998	23	2,3%
1999	12	1,2%
2000	11	1,1%
2001	9	0,9%
2002	15	1,5%
2003	15	1,5%
2004	16	1,6%
2005	18	1,8%
Gemiddelde	14	1,4%

Tabel 2: Aantal malen kop en munt.

In feite levert de databank van historische gegevens van dit bedrijf evenveel informatie op als een experiment waarbij tien groepen van duizend vorderingen worden geobserveerd en wordt bijgehouden hoeveel debiteurenverliezen zich voordoen. De uitkomst van dit 'experiment' is dat de kans op een oninbare vordering 1,4% is. Omdat de kansen van alle mogelijkheden moeten optellen tot 1 of 100%, kan ook iets worden gezegd over de kans dat een vordering wel inbaar is: 100% - 1,4% = 98,6%.

Kansrekening: Het nut voor de credit manager

Kansrekening speelt een centrale rol in credit management. Eén van de belangrijkste begrippen is de uitvalkans of in het Engels de probability of default. Deze uitvalkans wordt weergeven als een percentage en geeft de waarschijnlijkheid aan dat een debiteur niet volledig aan zijn verplichtingen voldoet gedurende een bepaalde tijdsperiode. We kunnen dergelijke uitvalkansen vaststellen voor individuele bedrijven of groepen van debiteuren maar ook voor financiële instrumenten zoals obligaties. Vaak wordt de afkorting PD (probability of default) gebruikt om aan te geven wat het risicoprofiel is.

Stel dat een bedrijf in haarverzorgingsproducten onder andere levert aan kleine zelfstandige kapperszaken. Deze leverancier weet op basis van zijn historische cijfers dat de PD-rate in dit marktsegment ongeveer 3% is. Dat betekent dat het haarverzorgingsbedrijf verwacht dat gedurende een jaar 3% van de leveringen op krediet als oninbaar moet worden geboekt. Dit is pure kansrekening.

Uitvalkans of PD is een gevleugeld begrip geworden doordat het een van de centrale begrippen is in de internationale Bazel-verdragen die het toezicht op banken regelen. In de Zwitserse stad Bazel is het hoofdkwartier gevestigd van de Bank for International Settlements (BIS), de centrale bank van de centrale banken. De BIS schrijft onder andere voor hoeveel kapitaal banken moeten aanhouden om zich in te dekken tegen bancaire risico's, waarbij het kredietrisico bovenaan staat. Het Bazel-verdrag vraagt banken hun kredietnemers te classificeren naar risicocategorieën en aansluitend PDs. Onder invloed van de Bazel-verdragen heeft credit scoring en credit rating een opmars gemaakt.

Een aanpalend kansbegrip uit de Bazel-akkoorden is de Loss Given Default (LGD), het te verwachten verlies bij wanbetaling. Stel we zijn een bedrijf dat financiële leaseproducten verkoopt op auto's waarbij de voertuigen ook het onderpand zijn voor de kredieten. Bij wanbetaling kan de auto worden ingevorderd en worden verkocht. Laten we aannemen dat dit na aftrek van kosten kan tegen gemiddeld 70% van de kredietwaarde. De LGD is dan 30%. Ook dit is een kansbegrip dat centraal staat in credit management.

De verwachtingswaarde

Het gooien met een dobbelsteen is een ander klassiek kansexperiment. De uitkomst van het kansexperiment is een variabele die verschillende waarden kan aannemen. In dit geval is de variabele het aantal ogen dat men gooit. Men zou kunnen stellen dat je bij het dobbelen niet van tevoren kan zeggen wat je gaat gooien. Dit is maar ten dele waar. Weliswaar kan men het aantal ogen niet van tevoren bepalen, maar er is meer vooraf over de uitkomst te zeggen dan men geneigd is te denken. Zo weet men dat de minimumuitkomst van het experiment 1 is, want minder dan 1 kan men niet gooien. Zo is het maximum 6 en zijn er ook maar zes mogelijke uitkomsten, namelijk: 1, 2, 3, 4, 5 of 6 ogen. Het is ook duidelijk dat de kans om 1 te gooien even groot is als de kans op 2, op 3 enzovoort. De uitkomsten moeten hele getallen zijn: het is onmogelijk om 1,27 te gooien met een dobbelsteen. De kans op 1 is één gedeeld door zes of 1/6 (één zesde) wat weer gelijk is aan 16,67% en zo is de kans op 2 ook 1/6 of 16,67% en zo verder. Alle kansen voor de verschillende mogelijkheden opgeteld, is weer 1 of 100%. Vooraf weet men al heel wat over dit kansexperiment. Schematisch kan men de verschillende uitkomsten en de daarbij behorende kansen weergeven in de volgende tabel:

Ogen	Kans	%
1	1/6	16,67%
2	1/6	16,67%
3	1/6	16,67%
4	1/6	16,67%
5	1/6	16,67%
6	1/6	16,67%
Totaal	1	100%

Tabel 3: Mogelijke uitkomsten bij het gooien met één dobbelsteen.

De hierboven beschreven kansexperimenten met de munt en de dobbelsteen worden in de wiskunde stochastische experimenten genoemd en de uitkomst van een stochastisch experiment heet een stochast. Een stochast is een variabele die ten minste deels door het toeval bepaald is.

Ogen	Kans	Berekening
1	16,67%	0,16667
2	16,67%	0,33333
3	16,67%	0,50000
4	16,67%	0,66667
5	16,67%	0,83333
6	16,67%	1,00000
	100,00%	3,50000

Tabel 4: Verwachtingswaarde aantal ogen bij het gooien met één dobbelsteen.

We zien dat alle kansen bij elkaar opgeteld 1 of 100% opleveren. Dat is geen toeval, want uiteindelijk zal bij het gooien met één dobbelsteen één uitkomst werkelijkheid worden. Vooraf is echter niet te zeggen of de uitkomst 1, 2, 3, 4, 5, of 6 zal worden. Vooraf is de kans op bijvoorbeeld een 2 gelijk aan 1/6 of 16,67%, achteraf, als blijkt dat er inderdaad 2 is gegooid, dan wordt deze kans 1 of 100% omdat alle mogelijke andere uitkomsten nu eenmaal geen rol meer spelen.

Door de kansen te vermenigvuldigen met de individuele uitkomsten en op te tellen, krijgen we de verwachtingswaarde. De verwachtingswaarde van het aantal gegooide ogen is 3,5. Wat betekent dit nu? Bij het rollen met één dobbelsteen kan men verwachten dat de uitkomst van het kansexperiment 3,5 is. Het gaat hier om een gemiddelde dat naar voren komt als het experiment maar vaak genoeg wordt herhaald. Natuurlijk is het zo dat ik met één worp van een dobbelsteen geen 3,5 kan gooien. Echter, als er honderdmaal met één dobbelsteen wordt geworpen, zal het gemiddelde van de uitkomsten dicht bij de 3,5 liggen. Ook bij het tossen van een muntstuk kan men de verwachtingswaarde uitrekenen. Dit kan eenvoudig

door de kans te vermenigvuldigen met de waarde van het muntstuk. Stel het gaat om een twee euro munt waarbij er wordt afgesproken dat bij het gooien van munt gooier de munt mag houden en hij bij het gooien van kop niet, dan is de verwachtingswaarde:

Verwachtingswaarde: € 2,00 * 50% + € 0,00 * 50% = € 1,00

Rekenkundig kan men zo bepalen dat de verwachtingswaarde van de winst bij het spelen van dit spel één euro bedraagt. Je zou hetzelfde kunnen doen door alle prijzen die de Staatsloterij uitlooft te vermenigvuldigen met de kans dat de prijs wordt gewonnen. Zou je dit doen voor alle prijzen en de resultaten bij elkaar optellen, dan krijg je de verwachtingswaarde van de inkomsten uit een staatslot. Dit komt waarschijnlijk uit op een bedrag dat veel lager ligt dan de prijs van een staatslot. In de praktijk ligt het uitkeringspercentage op ongeveer zeventig procent, waardoor de verwachtingswaarde van een staatslot uitkomt op minder dan € 10.50. Iedere keer als je een staatslot van vijftien euro koopt, verlies je al € 4,50 bij de kassa. Dit is echter een gemiddelde: je kunt helemaal niets winnen, maar ook de jackpot. De verwachtingswaarde is en blijft echter € 10,50. Je begrijpt nu waarom wordt gezegd dat de staatsloterij een extra belastingheffing is voor mensen die weinig van wiskunde begrijpen.

Deterministisch en probabilistisch

De verwachtingswaarde kan worden uitgerekend door voor alle mogelijke uitkomsten de kans te vermenigvuldigen met de waarde en deze uitkomsten bij elkaar op te tellen. Dit rekenkundig bepalen van de verwachtingswaarde wordt de deterministische methode genoemd. De uitkomst is bepaald door het toepassen van rekenkundige regeltjes die per definitie altijd dezelfde uitkomst opleveren. Er is dus een vast verband tussen de invoer en het resultaat. Dus als de mogelijke uitkomsten van de dobbelsteen één tot en met zes zijn en de kans in alle gevallen 1/6 of 16,67 %, dan is de verwachtingswaarde altijd 3,5. Er is een causaal verband tussen de invoer en de uitvoer, tussen de gegevens waarop onze som is gebaseerd en de uitkomst van 3,5. Het causale verband ligt vast in de rekenregels die zijn toegepast:

Verwachtingswaarde: 1 x 16,67% + 2 x 16,67%... + 6 x 16,67% + 3,5

De meeste zaken die we in ons dagelijks leven en zakenpraktijk uitrekenen, zijn deterministisch. Ook zijn bijna alle Excel-berekeningen deterministisch. Het causale verband ligt bij Excel besloten in de formule die je toepast op de invoerdata; deze zal altijd dezelfde uitkomst opleveren. De uitkomst wordt bepaald door de invoer en het toepassen van de rekenregels. In het latijn betekent *determinare* dan ook *bepalen* of *beslissen*; vergelijk met het Engelse to *determine*.

Het tegenovergestelde van deterministische berekeningen is dat je probalistisch een antwoord op een probleem krijgt. Het woord probabilistisch is afgeleid van het Latijnse *probare* dat we in het Nederlands kennen als *proberen*. De verwachtingswaarde van bijvoorbeeld de dobbelsteen kunnen we ook probabilistisch bepalen. Dit kan simpel door duizendmaal met de dobbelsteen te gooien, de resultaten bij elkaar op te tellen en dit weer te delen door duizend. Doe je dit, dan krijg je een uitkomst die dicht bij de eerder berekende verwachtingswaarde van 3,5 uitkomt. Beide methodes leveren dus hetzelfde resultaat op, met het verschil dat de deterministische berekening precies 3,5 oplevert en de probabilistische berekening ongeveer 3,5. Het is bij de probabilistische methode in theorie mogelijk dat je duizendmaal met de dobbelsteen gooit en duizendmaal één gooit. De kans hierop is verwaarloosbaar klein. Hoe vaker de dobbelsteen wordt gegooid, hoe preciezer de uitkomst wordt. Duizend worpen zijn voldoende om met een redelijke mate van zekerheid vast te stellen dat de verwachtingswaarde 3,5 is. We zullen verderop bespreken wat 'een redelijke mate van zekerheid' betekent.

Statistiek

De meeste mensen kennen de statistiek als onderdeel van de wiskunde van de middelbare school. Zag men als tiener wellicht niet altijd het nut van dit vak in, blijkt later in de praktijk van het beroepsleven vaak dat de statistiek wel degelijk nut heeft. De statistiek is een van de meer toegankelijke terreinen van de wiskunde omdat het zeer praktisch is gericht. De statistiek heeft te maken met ons alledaagse leven en is minder abstract dan andere terreinen van de wiskunde.

Wat is statistiek?

Statistiek is de wetenschap van het verzamelen en vergelijken van massale verschijnselen en van de weergave hiervan in tabellen of grafische voorstellingen, aldus Van Dale. Het gaat dus om gegevens of data van verschijnselen die we om ons heen kunnen waarnemen. We noemen deze gegevens uit de praktijk empirische data. In de statistiek worden gegevens kwantitatief dus in getalvorm weergegeven. Doel van de statistiek is om ons te helpen goede beslissingen te nemen. Statistiek heeft dus duidelijk een praktische dimensie. Statistische informatie kan op georganiseerde wijze worden weergegeven, bijvoorbeeld in tabellen, grafieken en diagrammen.

Een stukje geschiedenis

Een belangrijke ontwikkeling is het idee van de steekproef en de opkomst van de statistiek. De moderne staten van de zeventiende en achttiende eeuw zoals Frankrijk, het Verenigd Koninkrijk en onze eigen Republiek der Zeven Verenigde Nederlanden hadden behoefte aan het meten en schatten van de bevolking om belastingopbrengsten en inzet voor het leger te kunnen bepalen. Volkstellingen waren moeilijk en daarom ging men ertoe over om op basis van steekproeven schattingen te doen voor het geheel. Niet voor niets is de oorsprong van het woord statistiek terug te voeren op 'staat' wat verwijst naar de staathuishouding. Wilde men iets zeggen over de leeftijdsopbouw van de bevolking, dan kon men één stad helemaal in kaart brengen en deze steekproef uitvergroten naar het hele land. De statistiek heeft dus van het begin af aan een praktische insteek gehad.

Dit idee van het nemen van steekproeven liep parallel aan de vondst van de wet van de grote aantallen. Men ontdekte dat bijvoorbeeld de verhouding kop-munt bij het herhaaldelijk gooien van een muntstuk 50% / 50% was. Experimenten die vaak worden herhaald, tenderen naar een bepaald gemiddelde en dit gemiddelde is beter te bepalen naarmate men een grotere steekproef neemt. Dit is de grondslag van de wet van de grote aantallen.

Kwantitatieve en kwalitatieve data

In de statistiek gaat het om gegevens ofwel data die men heeft verzameld. Dit zijn feiten en informatie afkomstig uit waarnemingen, metingen en experimenten. Hierbij wordt een onderscheid gemaakt tussen aan de ene kant kwantitatieve data die te meten zijn en in getallen weer te geven zijn, en aan de andere kant kwalitatieve of categorische data. Kwantitatieve data zijn het resultaat van een waarneming of meting met behulp van een instrument zoals een thermometer, een liniaal of een kilometerteller. Bij categorische data gaat het om gegevens die zijn gegroepeerd aan de hand van een kenmerk, waarbij we de frequentie, waarmee deze eigenschap voorkomt, tellen. Zo kunnen we het aantal fietsen tellen van het merk Batavus dat iedere dag door de straat rijdt, of het aantal producten met een defect, het aantal scholieren van acht jaar oud enzovoorts.

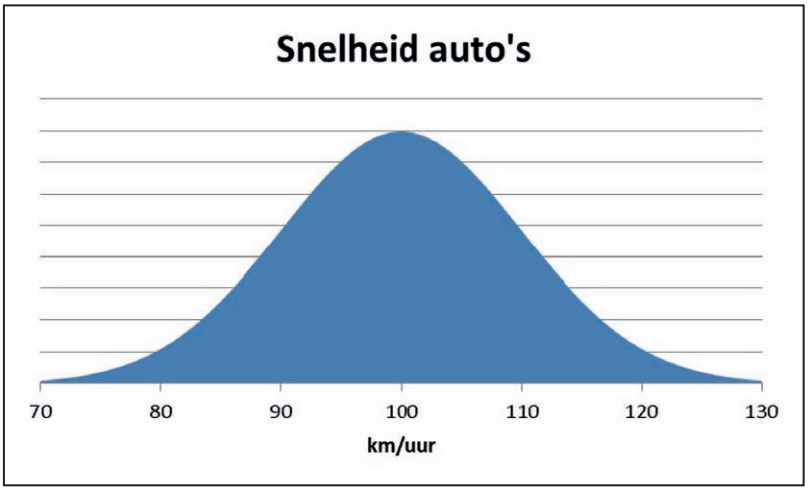
Variabelen

Een ander belangrijk concept in de statistiek is de 'variabele'. Een variabele is een grootheid die in waarde kan veranderen. De variabele temperatuur kan waarden aannemen zoals 10 °C of 12 °C, de variabele snelheid 110 km/uur, en de variabele 'studierichting' kan waarden hebben als wiskunde, psychologie, Frans, geschiedenis enzovoorts. Het tegenovergestelde van variabele is een 'constante'. De constante is een grootheid die niet van waarde verandert en dus vast is.

Continue en discrete variabelen

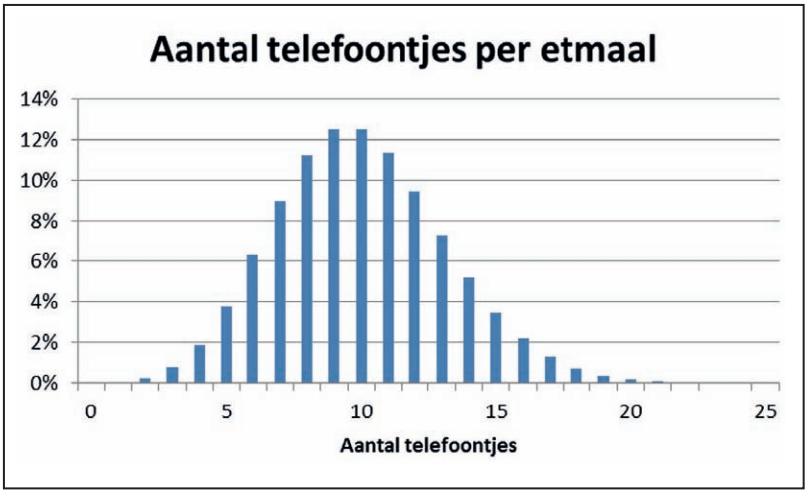
Variabelen kunnen continu of discreet zijn. Continue variabelen kunnen alle mogelijke waarden aannemen tot op een oneindig detailniveau, terwijl discrete variabelen maar een beperkt aantal mogelijke waarden aan kunnen nemen binnen een bepaalde bandbreedte.

De snelheid van een auto is een voorbeeld van een continue variabele. We kunnen de snelheid van auto's meten langs een snelweg en tot de conclusie komen dat de snelheid van een voorbijrazende auto 102 km/uur is. Echter, als we een nauwkeurigere snelheidscamera hebben kunnen we wellicht constateren dat de snelheid 102,3 is. Hebben we de beschikking over een nog nauwkeuriger meetinstrument dan komen we misschien uit op 102,34 of misschien wel 102,341 km/uur enzovoorts. Kortom, we kunnen de nauwkeurigheid steeds verder opvoeren. Dit geldt voor de snelheid van alle auto's die langsrijden. Op die manier zijn er in principe oneindig veel mogelijke snelheden te meten in de range 0 km/uur voor een stilstaande auto tot 325 km/uur van een langs flitsende Ferrari.



Figuur 1: Voorbeeld van een continue variabele.

Een discrete variabele heeft maar een beperkt aantal waarden in een bepaalde bandbreedte. Zo is het aantal auto's dat op maandagochtend tussen 9:00 en 9:15 u komt tanken bij een tankstation altijd een geheel getal. Het kunnen nul, één, twee, drie, vier, vijf enzovoorts auto's zijn, maar nooit 3,67 auto's. Hetzelfde geldt voor het aantal drukfouten dat men vindt op een pagina van een krant, of het aantal telefoontjes dat je op je mobieltje ontvangt gedurende een etmaal.



Figuur 2: Voorbeeld van een discrete variabele.

Overigens worden continue variabelen vaak in een computer verwerkt als discrete variabelen met een hele kleine stapgrootte. Andersom kunnen discrete variabelen met een voldoende kleine stapgrootte vaak als continue variabelen behandeld

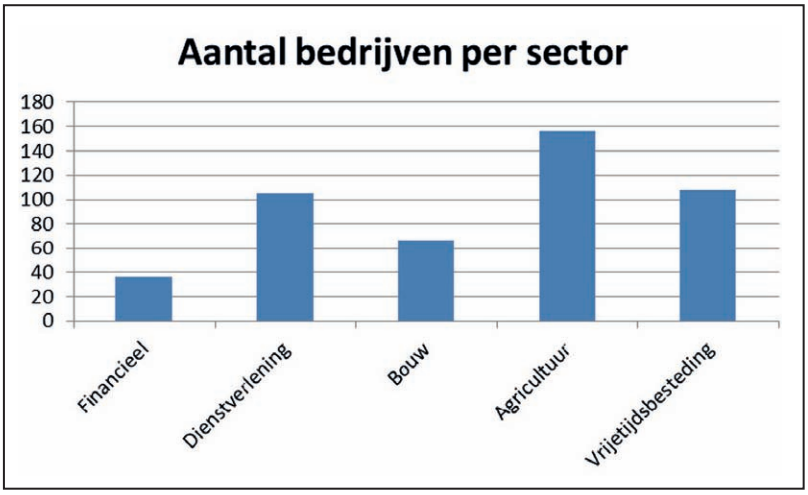
worden. Geldbedragen, bijvoorbeeld, zijn in principe discrete variabelen omdat er geen bedragen kleiner dan één cent zijn. Maar wanneer de bedragen groot genoeg zijn, speelt de afronding op één cent geen noemenswaardige rol meer.

Beschrijvende statistiek I: verdelingen en maten

Beschrijvende statistiek gaat over het weergeven en beschrijven van data. De data is vaak afkomstig van een steekproef, en het nut van de beschrijvende statistiek is om de resultaten van de steekproef duidelijk te maken zonder dat hiervoor de complete data set nodig is.

Kolommendiagram of histogram

De resultaten van een steekproef of telling kunnen we in kaart brengen door deze in een kolommendiagram af te beelden. Hierbij worden de gegevens gegroepeerd in groepen of klassen en geeft de hoogte van de kolom aan hoe vaak een waarneming is gedaan in die klasse. Het kolommendiagram geeft daarmee een beeld van de relatieve frequentie van de verschillende categorieën.



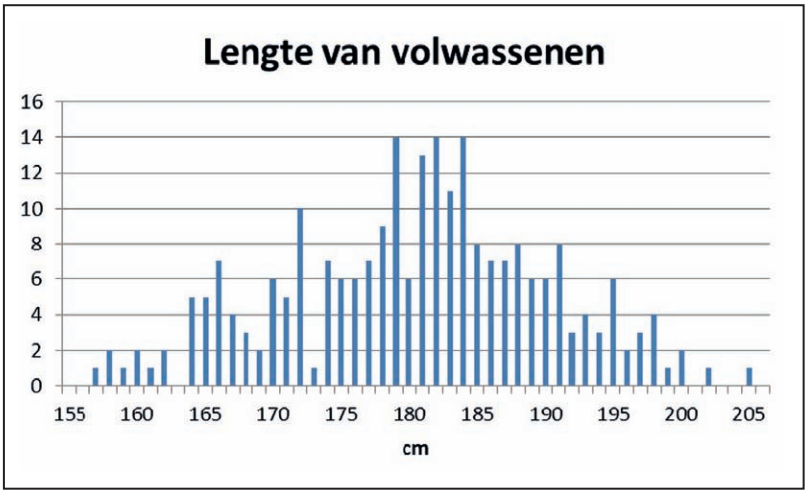
Figuur 3: Voorbeeld van een histogram van categorische data.

Zo kunnen we een overzicht maken van de bedrijven in een grote stad, en deze bedrijven in een aantal sectoren indelen. Vervolgens tellen we hoeveel bedrijven er in elk van de sectoren zit, en geven dit grafisch weer in een staafdiagram. Als label gebruiken we de namen van de verschillende sectoren. Dit is een voorbeeld van een histogram of kolommendiagram.

In het histogram kan eenvoudig afgelezen worden in welke sector de bedrijven in deze stad werkzaam zijn. Zo volgt uit bovenstaande figuur dat er tweemaal zoveel bedrijven in de bouw zitten als in de financiële sector, of dat de meeste bedrijven

tot de agrarische sector behoren. Wanneer we een zelfde soort histogram voor een andere stad maken, dan kunnen we in één oogopslag zien of de bedrijvigheid op een vergelijkbare manier over de verschillende sectoren verdeeld is of niet.

In het voorbeeld gaat het om categorische data: de data zijn al ingedeeld in een aantal groepen. We beperken ons tot het tellen van het aantal waarnemingen voor een bepaalde categorie of groep. In principe is de rangschikking van de groepen op de horizontale as willekeurig. Doen we hetzelfde met kwantitatieve data, dus gegevens die het resultaat zijn van een meting, dan staan de waarden op de horizontale as in een bepaalde volgorde. Gaat het om continue variabelen als gewicht, lengte of temperatuur dan zal men de variabelen willen samenvoegen in groepjes. Stel we meten de lengte van 250 volwassen personen die in de Kalverstraat voorbij komen. We ronden de lengte naar beneden af op hele centimeters. Effectief hebben we hiermee klassen gecreëerd van steeds 1 cm. Dit betekent dat iemand van 175,1 cm en een andere persoon met een lengte van 175,9 cm beiden in de categorie lengte 175-176 cm vallen.



Figuur 4: Voorbeeld van een histogram van kwantitatieve data.

In bovenstaand histogram staat de lengte op de horizontale as weergegeven en de frequentie op de verticale as. Er zijn blijkbaar vijf personen met een lengte van 163-164 cm en er is niemand met een lengte van 162-163 cm.

Histogram: Hoe doe je het in Excel?
Om een histogram of kolommendiagram te kunnen maken is het eerst nodig om je data in een aantal categorieën op te delen, en te tellen wat de frequentie van elke categorie is. Dit kan je in Excel doen met behulp van (NL)[=aantallen.als()] of (EN)[=countifs()]. Als je je thuis voelt in de matrix-functies in Excel dan kan je ook (NL)[=interval()] (EN)[=frequency()] gebruiken.

Beschrijvende statistiek

Het valt op dat er in het kolommendiagram van de lengtemetingen een bepaald patroon te ontdekken is. De meeste waarnemingen liggen rond de 180 cm en naarmate men verder naar links of naar rechts gaat wordt de frequentie minder. Het afbeelden van de data in een frequentieverdeling of histogram geeft ons een indruk van de algemene vorm van de verdeling ofwel distributie. We krijgen met andere woorden een grof idee van hoe de metingen en getallen zijn verdeeld. Verscheidene statistische gereedschappen kunnen worden gebruikt om de aard van de distributie of verdeling te beschrijven: gemiddelde, mediaan, minimum waarde, maximum waarde, enzovoorts. Hiermee zijn we aangeland op het terrein van de beschrijvende statistiek. In de beschrijvende statistiek worden grote verzamelingen metingen en gegevens gestructureerd en ingedikt tot overzichtelijke kerngetallen. In het voorbeeld van de lengtemetingen weet men al heel wat indien bekend is dat de gemiddelde lengte 181 cm is en de laagste meting 119 cm was en de langste persoon 240 cm was.

Het is gebruikelijk om deze kengetallen te verdelen in drie categorieën: centrummaten, spreidingsmaten en vormmaten.

Centrummaten

Centrummaten geven aan rond welke centrale waarde de gegevens zijn gegroepeerd. Met behulp van centrummaten probeert men antwoord te geven op de vraag: 'waar ligt het midden van de data?' Achtereenvolgens komen aan de orde: gemiddelde, mediaan en modus.

Gemiddelde

Het gemiddelde is het meest gebruikte kengetal uit de beschrijvende statistiek. De berekening is simpel: tel de waarden van alle data(punten) bij elkaar op en deel dit door het aantal data(punten). In het Engels worden de termen 'mean' of 'average' naast elkaar gebruikt.

Een mogelijke valkuil bij het gebruik van het gemiddelde als centrummaat, is de manier waarop met extreme waarden wordt omgegaan. Die hebben een vrij grote invloed op het uitgerekende gemiddelde. Voorbeeld: als Bill Gates met je in de

schoolbanken zou hebben gezeten, heeft het geen zin het gemiddelde salaris te berekenen van de alumni van je school. Iedereen, met uitzondering van Bill, zou dan een lager dan gemiddeld salaris hebben. In dergelijke situaties moet je jezelf afvragen of je echt de gemiddelde waarde wilt weten, of dat je een waarde wilt hebben die het grootste deel van de groep beschrijft.

Gemiddelde: Hoe reken je het uit in Excel?
In Excel kan dit eenvoudig door gebruik te maken van (NL)[=gemiddelde()] of (EN) [=average()]. Wanneer je hier een reeks van getallen ingeeft dan zoekt Excel naar de waarde die het vaakste in de reeks voorkomt.

Mediaan

De mediaan is niets anders dan de middelste waarneming. Hiervoor is noodzakelijk om de waarnemingen te sorteren in een rij van klein naar groot. Zijn er een oneven aantal waarnemingen dan is het simpel om de middelste te nemen. Is er een even aantal dan neem je de middelste twee, telt die op en deelt door twee. In dat geval is de mediaan het gemiddelde van de middelste twee waarnemingen. De mediaan is een robuuste centrummaat waarop extreme waarden geen effect hebben. De Engelse term is 'median'. De mediaan is handig als een extreme waarde of waarneming het gemiddelde sterk doet verschuiven. Als je uitzoekt wat voor een salaris de alumni van je oude school verdienen, terwijl Bill Gates vroeger met je in de schoolbanken zat, dan zegt de mediaan waarschijnlijk meer over de klas als geheel dan het gemiddelde.

Mediaan: Hoe reken je het uit in Excel?
In Excel kan dit eenvoudig door gebruik te maken van (NL)[=mediaan()] of (EN) [=median()]. Wanneer je hier een reeks van getallen ingeeft dan zoekt Excel naar de mediaan.

Andersom zijn er ook situaties waarin het gemiddelde betere informatie geeft dan de mediaan. Neem bijvoorbeeld een bank met een miljoen uitstaande hypotheken. Meer dan de helft van de huiseigenaren zal zijn hypotheek netjes terugbetalen, dus de mediaan van de winst die de bank maakt op een hypotheek zal altijd positief uitvallen. Maar doordat een enkeling zijn hypotheek niet af zal betalen is het wel mogelijk dat het gemiddelde, of de verwachtingswaarde, negatief wordt. In een dergelijk geval zou de bank zijn rente moeten verhogen om geen verlies te draaien.

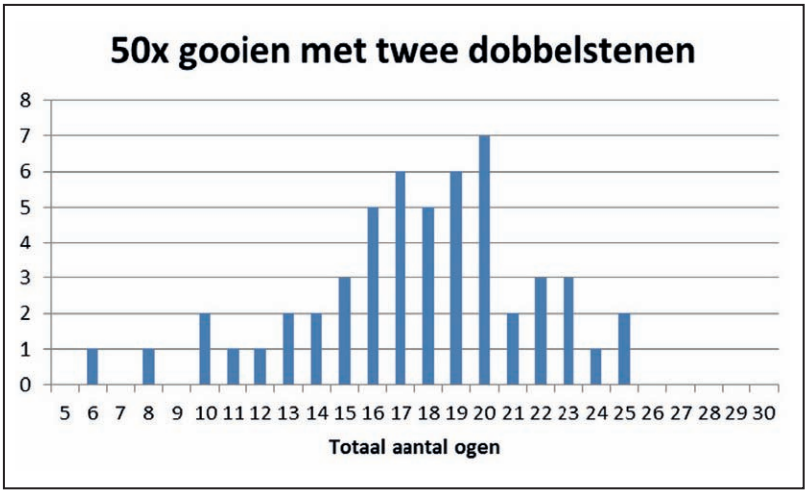
Modus

De modus is simpelweg de waarde of meeting die het meest voorkomt. Het kan zijn dat er meerdere *modi* zijn indien er meerdere meest voorkomende waarden zijn, het kan ook zijn dat er geen modus is in een databestand.

De term *modus* komt ook terug in Jan Modaal, in de politiek zo vaak aangehaald. Deze Jan Modaal verdient een modaal inkomen. Dit wil zeggen, de inkomenscategorie waar de meeste Nederlanders inzitten of het inkomen dat het vaakst voorkomt.

Voorts geeft de modus het antwoord op de vraag 'wat denk je dat de meest waarschijnlijke uitkomst is?'. Immers, de modus is de meest voorkomend waarde en als je op deze uitkomst gokt heb je de meeste kans om goed te zitten! In de kansrekening speelt de modus een belangrijke rol in de zogenaamde driehoeksverdeling. De Engelse term is 'mode'.

De modus is niet gevoelig voor extreme waarden, wat een voordeel is in vergelijking met het gemiddelde. De modus is wel gevoelig voor ruis, vooral in het geval van brede toppen in de verdeling, zoals in onderstaand histogram. Bovendien kan de modus alleen gebruikt worden bij discrete variabelen.



Figuur 5: Histogram van het totaal aantal ogen van twee dobbelstenen, vijftig worpen.

Modus: Hoe reken je het uit in Excel?
In Excel kan dit eenvoudig door gebruik te maken van (NL)[=modus()] of (EN)[=mode()], of in nieuwere versies: (NL)[=modus.meerv()] of (EN)[=mode.mult()]. Wanneer je hier een reeks van getallen ingeeft dan zoekt Excel naar de waarde die het vaakst in de reeks voorkomt.

Spreidingmaten

Soms kunnen centrummaten ronduit misleidend zijn: als je een voetbad neemt met één voet in een teiltje met kokend water en de andere voet in een bak met ijsklontjes dan is het gemiddeld lekker warm en comfortabel…

Spreidingsmaten geven de mate van verstrooiing of spreiding van data aan. We behandelen hier: range, variantie en standaarddeviatie.

Range of variantiebreedte

De spreidingsmaten laten zien wat de mate van verstrooiing is van de data. Zitten ze op kluitje, of zijn ze juist sterk verspreid? Een eenvoudige manier om iets te zeggen over de spreiding is door te kijken naar de variantiebreedte of range. Dit is niets meer en niets minder dan het verschil tussen maximum, de hoogste waarde, en minimum, de laagste waarde. De range wordt wel sterk beïnvloed door extreme waarden.

Range: Hoe reken je het uit in Excel?
In Excel kan dit eenvoudig door gebruik te maken van de formule [=max()-min()]. Wanneer je hier een reeks van getallen ingeeft dan zoekt Excel naar de maximale waarde en minimale waarde, en rekent het verschil uit.

Variantie en Standaarddeviatie

Variantie en standaarddeviatie zijn spreidingsmaten die het beste samen kunnen worden behandeld; ze zijn als het ware broer en zus van elkaar. De standaarddeviatie is de wortel van de variantie en omgedraaid is de variantie de standaarddeviatie in het kwadraat. Weten we de één, dan weten we ook de ander! Het blijkt dat velen moeite hebben met deze concepten. Ze zijn echter cruciaal in de statistiek en modellenbouw en het is daarom van belang langer bij dit onderwerp stil te staan. Kort door de bocht kan met zeggen dat de variantie het gemiddelde is van het kwadraat van de afwijking van het gemiddelde.

Een rekenvoorbeeld helpt het beste om uit te leggen wat variantie is. In nevenstaande afbeelding zien we een reeks van tien data. Het gemiddelde bedraagt 6. We kunnen nu voor ieder van de datapunten uitrekenen hoe ver het punt van het gemiddelde afligt. Dit kan eenvoudig door het gemiddelde af te trekken van het datapunt. Voor het eerste datapunt 2: dit ligt 2 – 6 = -4 van gemiddelde 6 af. Deze operatie herhalen voor alle datapunten. Deze afwijking gaan we vervolgens kwadrateren. De afwijking -4 levert gekwadrateerd 16 op. Deze kwadraten worden opgeteld. De som is 66. De som van de kwadraten wordt vervolgens gedeeld door het aantal waarnemingen min één. In ons geval moet de som van 66 worden gedeeld door 10 – 1 = 9. Het resultaat is een variantie van 7,33.

	Data	Gemiddelde	Afwijking Deviatie	Kwadraat
	2	6	-4	16
	4	6	-2	4
	4	6	-2	4
	5	6	-1	1
	6	6	0	0
	6	6	0	0
	6	6	0	0
	7	6	1	1
	8	6	2	4
	12	6	6	36
Som	60			66
Aantal data	10			
Gemiddelde	6			
Variantie				7,333333
Standaarddeviatie				2,708013

Tabel 5: Berekening van de standaarddeviatie.

Het uitrekenen van de standaardafwijking is daarna eenvoudig: we nemen de wortel uit 7,33 en komen dan uit op $\sigma = 2{,}71$. De kleine Griekse letter sigma (σ) wordt vaak gebruikt om de standaarddeviatie aan te duiden.

Variantie en standaarddeviatie: Hoe reken je het uit in Excel?
In Excel kan dit eenvoudig door gebruik te maken van [=var()], of in de nieuwere versies [=var.p()] en [=var.s()]. Hierbij wordt een reeks van getallen ingegeven en Excel antwoordt met de variantie. Het verschil tussen de 'p' (populatie of population) en de 's' (steekproef of sample) is dat men zich moet afvragen of de data betrekking hebben op een deel van het geheel of juist de hele populatie omvatten. Zodra het om wat grotere dataseries gaat maakt dit onderscheid overigens vrijwel niets uit.

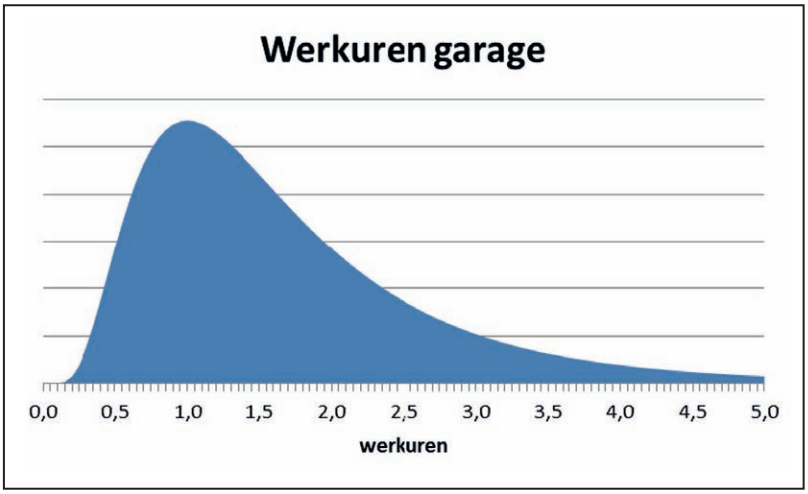
Vormmaten

Vormmaten zeggen iets over de vorm van de verdeling. Is deze symmetrisch of asymmetrisch? Spits of plat? We zullen de scheefheid en de platheid hier behandelen.

Scheefheid

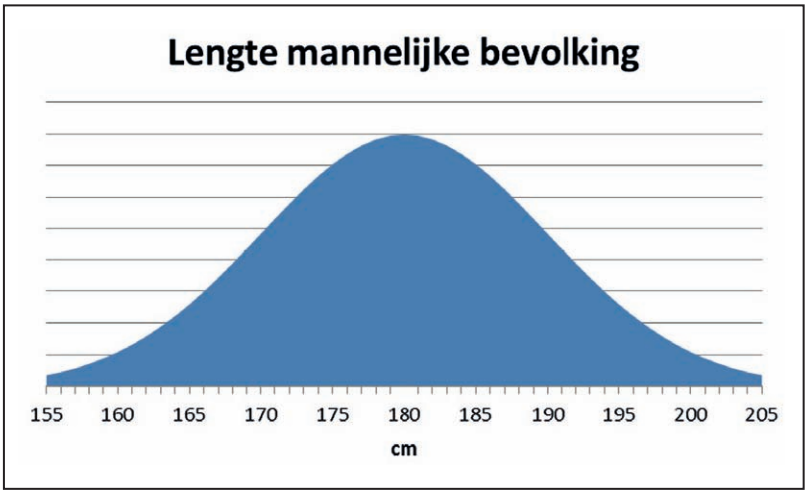
Scheefheid, in het Engels *skewness* genoemd, geeft de mate van asymmetrie aan van een dataset. De verdeling van de data kan symmetrisch zijn, waarbij de linker- en rechterstaart van de grafiek even lang zijn of asymmetrisch, waarbij één van de staarten langer is dan de andere.

Scheefheid zegt iets over de vorm van de grafiek en is om die reden dan ook een vormmaat. In termen van risico zijn deze staarten van belang. De lengte van de staarten zegt iets over mogelijke uitschietende waarden die ver afliggen van het midden van de grafiek. Deze uitschieters representeren vaak extreme situaties met veel risico. Dikwijls zijn uitschieters asymmetrisch verdeeld. Een voorbeeld legt dit het beste uit. Stel dat een onderhoudsbeurt aan een auto standaard een uur werk is voor de garage. In de grote meerderheid van de gevallen zullen klanten die hun auto brengen voor een beurt een rekening krijgen voor één uur werk. Het zal niet of nauwelijks voorkomen dat de garagist de klant na vijf minuten belt met de mededeling dat de auto al klaar is, maar het gebeurt vaker dat men een probleem heeft gevonden waardoor er veel langer dan één uur aan de auto wordt gewerkt. Het vervangen van de versnellingsbak kan de factuur opdrijven naar acht uur werk. De asymmetrie komt naar voren in het feit dat deze tegenvallers vaker voorkomen, maar de meevallers veel minder en bovendien ook minder extreme waarden kunnen aannemen. Vijf minuten voor een onderhoudsbeurt is niet erg waarschijnlijk en bovendien is er een harde ondergrens van nul minuten omdat er geen negatieve reparatietijd bestaat. Aan de andere kant van het spectrum is er in principe geen bovengrens. Immers, wat is de maximale tijd dat men aan een auto kan sleutelen?



Figuur 6: Voorbeeld van een asymmetrische verdeling.

De scheefheid kan groter dan nul zijn: dan is de scheefheid positief met vooral uitschieters naar de positieve zijde van de x-as, of negatief met uitschieters naar links, naar de negatieve zijde van de x-as. De scheefheid kan ook nul zijn of bijna nul waarbij de grafiek symmetrisch is.

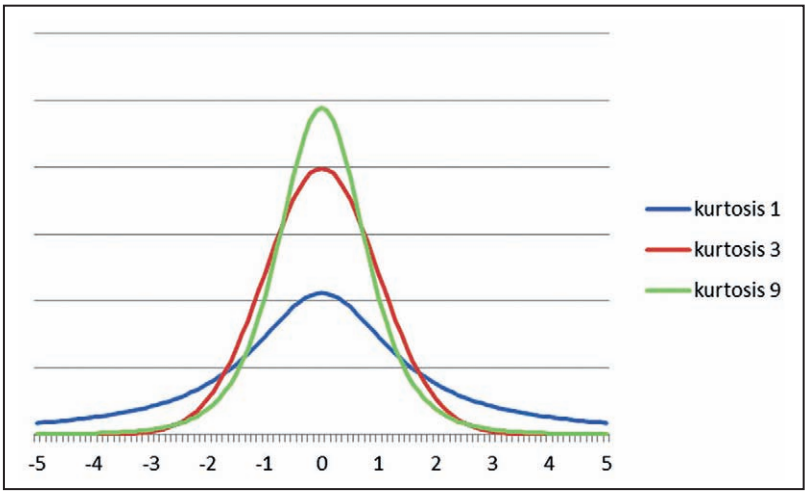


Figuur 7: Voorbeeld van een symmetrische verdeling.

Scheefheid: Hoe reken je het uit in Excel?
In Excel kan dit door gebruik te maken van (NL)[=scheefheid()] of (EN) [=skew()].
Hierbij wordt een reeks van getallen ingegeven en geeft Excel de scheefheid terug.

Kurtosis

De kurtosis van de grafiek is de laatste vormmaat die we hier bespreken. De kurtosis geeft een indicatie van hoe plat of hoe gepiekt een grafiek is. Kurtosis is dan ook Grieks voor welving. Bij een kurtosis waarde groter dan drie is de grafiek meer gepiekt dan een standaard normaalverdeling, dit wil zeggen een normaalverdeling met een standaarddeviatie van één. Bij een kurtosis kleiner dan drie is de grafiek platter dan bij een standaard normaalverdeling.



Figuur 8: Drie verdelingen met verschillende kurtosis.

In bovenstaand plaatje is een drietal grafieken te zien. De rode distributie is een normaalverdeling met een standaarddeviatie van één en een kurtosis van drie. De blauwe verdeling is duidelijk platter en heeft een kurtosis kleiner dan drie, terwijl de groene grafiek duidelijk gepiekt is met een kurtosis groter dan drie.

Kurtosis: Hoe reken je het uit in Excel?
In Excel kan dit simpel door gebruik te maken van (NL)[=kurtosis()] of (EN) [=kurt()]. Hierbij wordt een reeks van getallen ingegeven en geeft Excel de kurtosis terug.

Wilde momenten in de statistiek

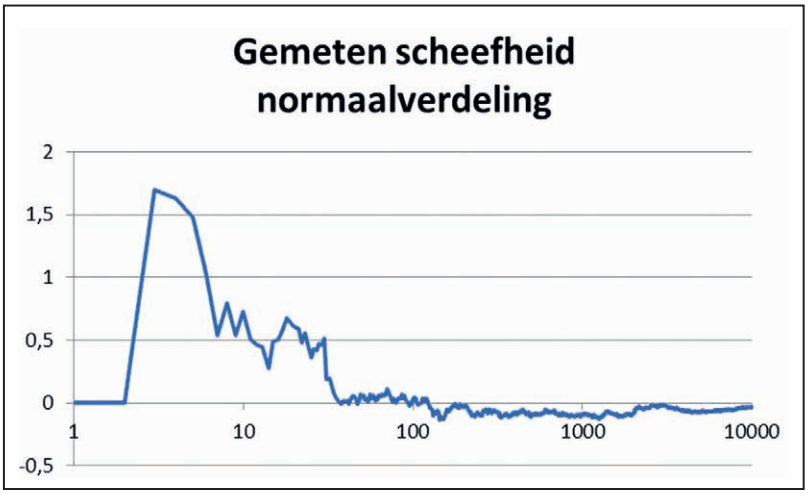
Het gemiddelde, de variantie, de scheefheid en de platheid worden ook wel momenten genoemd. De momenten beschrijven de data en geven ons een beeld van het vlees dat we in de kuip hebben. Distributies met dikke staarten geven aan dat er relatief veel risico is. Dat wil zeggen dat er grote kans is op extreme waarden. Nu is er met deze verdelingen met een dikke staart iets opmerkelijks aan de hand: de momenten zijn niet altijd gedefinieerd. Dit betekent dat we niet in staat zijn om met de gebruikelijke statistische gereedschappen iets te zeggen over de data. Kort door de bocht: deze verdelingen kunnen zo wild zijn, dat ze zich niet laten vangen in de structuren en kerndefinities die we gewend zijn te gebruiken in de statistiek. Bij verdelingen met een dikke staart zijn één of meerdere momenten niet gedefinieerd. Dit houdt in dat naarmate je meer data verzamelt, je niet meer zekerheid krijgt over sommige van deze momenten. Een toelichting is hier op zijn plaats.

In de wiskunde kennen we de centrale limietstelling die stelt dat het sommeren van een groot aantal toevalsvariabelen met eindige variantie resulteert in een normaal verdeelde stochast of toevalsvariabele. Dit leidt ertoe dat de waarden gevonden in een representatieve steekproef normaal verdeeld zijn rond het gemiddelde van de totale populatie. Stel ik wil weten hoe groot de aanhang is van partij X als er verkiezingen worden gehouden. Het is hiervoor niet nodig alle kiesgerechtigde Nederlanders naar hun politieke voorkeur te vragen: een steekproef volstaat hier. Zonder in details te treden zal het duidelijk zijn dat naarmate we een grotere steekproef nemen, het steekproefgemiddelde dichter bij het echte gemiddelde van alle kiesgerechtigden komt. Naarmate de steekproef groter wordt, convergeren de gevonden gemiddelden naar het echte gemiddelde en ook de foutmarge wordt steeds kleiner. Zo niet bij de dikke staartverdelingen. Hier blijkt dat één of meer momenten niet te definiëren zijn en steeds meer verspringen naarmate men meer data heeft.

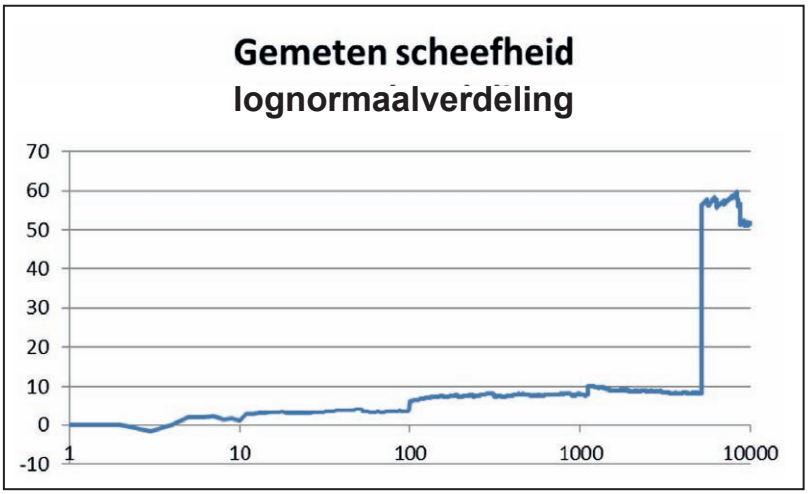
Lognormaal verdeling met minimum 0, gemiddelde 1 en σ 5				
Trekkingen	Gemiddelde	Variantie	Scheefheid	Platheid
1.000	0,77	4,99	8,08	82,62
10.000	1,05	59,98	51,86	3.222,73
100.000	1,01	22,72	40,48	3.058,44
1.000.000	1,00	28,43	125,59	41.334,22

Tabel 6: Dikke staarten zorgen voor wilde momenten.

In bovenstaand voorbeeld zien we dat de scheefheid en vooral de kurtosis sterk oplopen naarmate er meer trekkingen worden gedaan en er meer data beschikbaar komen. Deze tendens is tegenovergesteld aan de convergentie die we zien bij de wet van grote aantallen. In bovenstaand voorbeeld is duidelijk te zien dat bij dikke staarten de scheefheid en kurtosis sterk toenemen en op hol slaan naarmate er meer data worden verzameld.



Figuur 9: De gemeten scheefheid bij gesimuleerde normaal verdeelde data. De scheefheid convergeert wanneer het aantal datapunten vergroot wordt.



Figuur 10: De gemeten scheefheid bij gesimuleerde lognormaal verdeelde data. De scheefheid convergeert niet maar maakt onverwacht grote sprongen: er is sprake van onvoorspelbaar gedrag.

Dit is teleurstellend: in situaties waar zich grote en extreme risico's voordoen laat onze statistische gereedschapskist ons in de steek. Maar we kunnen dit fenomeen gebruiken als kanarie in de kolenmijn die stopt met zingen zodra er mijngas aanwezig is. Indien één of meerdere momenten niet gedefinieerd zijn is het oppassen geblazen want er zijn dan mogelijk grote, moeilijk voorspelbare, risico's in het spel.

Wilde momenten: Probeer het uit in Excel!

Om dit effect in Excel aan te tonen heb je de random number generator nodig.
Voer eerst de volgende formule in Excel in: (NL)[=stand.norm.inv(aselect())] of (EN)[=normsinv(rand())], of in de nieuwere versies (NL)[=norm.s.inv(aselect())] en (EN)[=norm.s.inv(rand())]. Met deze formule genereer je standaard normaal verdeelde getallen. Kopieer en plak deze formule naar een groep cellen, en bereken de momenten over deze groep. Als het goed is, dan zie je dat de momenten stabiliseren wanneer je de groep groter (honderden tot duizenden) maakt. Dit ligt in lijn met de wet van de grote aantallen.
Herhaal dit nu met de formule (NL)[=1/(1-aselect())] of (EN)[=1/(1-rand())], waarmee je getallen genereert uit een dikke staartverdeling. Als het goed is zie je nu dat de uitgerekende momenten niet stabiliseren wanneer je de groep groter maakt. Dit is een voorbeeld van een verdeling waarvan de momenten niet goed vastgelegd kunnen worden.

Beschrijvende statistiek: Het nut voor de credit manager
Het verstrekken van een lening betekent voor een credit manager dat hij risico op zich neemt. Om dat risico af te dekken, zal hij een financiële buffer moeten aanhouden. Deze buffer mag niet te klein zijn, omdat het risico dan onvoldoende afgedekt is. Maar aan de andere kant mag de buffer zeker ook niet te groot zijn, omdat geld dat stil staat niets opbrengt.

De credit manager wil dus graag weten hoeveel risico hij zal moeten afdekken. De eerste stap hierin is het in kaart brengen van de risico's die bij de afzonderlijke leningen horen.

Door historische gegevens te verzamelen en de verdeling van deze gegevens te bestuderen, kan de credit manager zich een verwachtingsbeeld van de toekomst vormen. Zo geven centrummaten een verwachtingswaarde voor de toekomst, tonen spreidingsmaten aan wat de onzekerheid (of risico) is, en kan uit een histogram worden afgelezen hoe vaak een waarde onder een zekere grenswaarde ligt.

De te verwachten verliezen op debiteurenvorderingen zijn ook van belang voor een goede berekening van de kostprijs en prijsstelling van een product. Immers, wanbetaling leidt tot extra kosten die gedekt moeten worden door de productprijs.

Hierbij moet wel goed bedacht worden dat voorspellingen op basis van een steekproef alleen een voorspellende waarde hebben wanneer de steekproef voldoende groot in omvang is, en representatief is. Dat laatste wil zeggen dat alle mogelijke situaties en uitkomsten in de steekproef aan bod komen.

De relevantie van de wilde momenten ligt in het feit dat de spreiding, en dus het risico, niet betrouwbaar gemeten kan worden voor een verdeling met een dikke staart. Een credit manager die hier niet op bedacht is zou een grove misrekening kunnen maken!

Beschrijvende statistiek II: samenhang in data

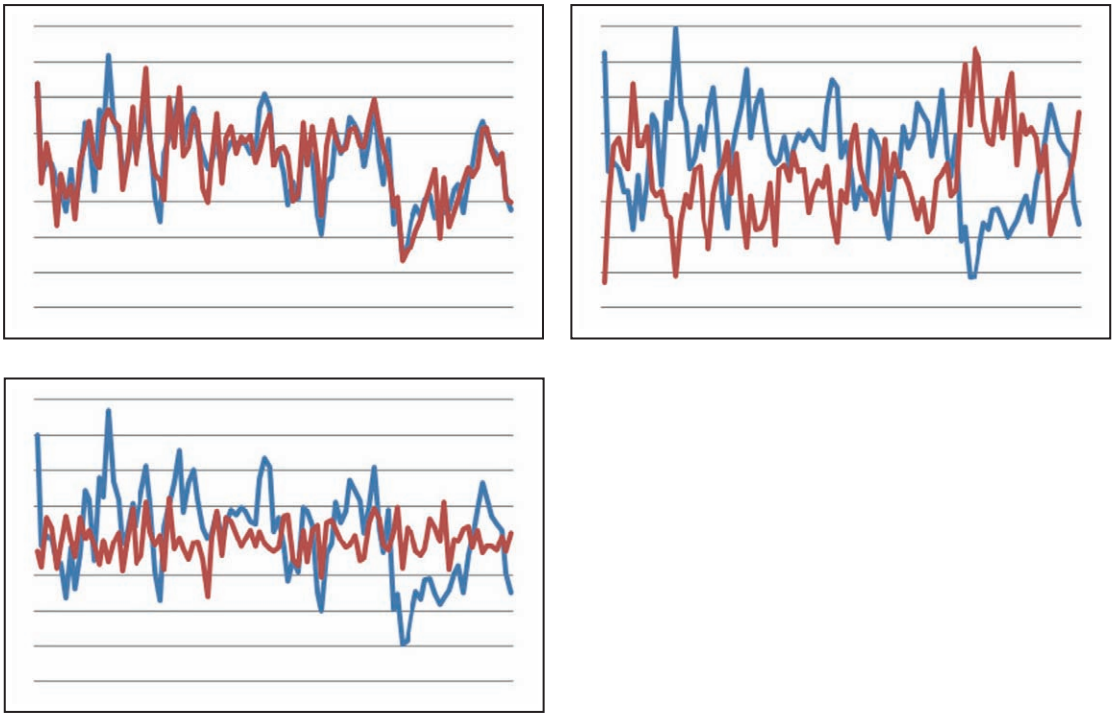
De histogrammen, centrummaten, spreidingsmaten en vormmaten kunnen gebruikt worden om een beschrijving van waargenomen data te geven, maar doen dit door iedere toevalsvariabele apart te behandelen. In sommige situaties voldoet dit, maar in veel andere situaties heb je met meerdere toevalsvariabelen tegelijk te maken.

Dit vraagt om enige extra aandacht bij het verwerken van de gegevens. Iedere keer dat we met meerdere toevalsvariabelen tegelijk te maken krijgen, zouden we de vraag moeten stellen: zijn deze variabelen onderling onafhankelijk, of afhankelijk?

Correlatie en covariantie

Correlatie en covariantie zijn maten voor de samenhang, of afhankelijkheid, van data. Correlatie wordt weergegeven in een correlatiecoëfficiënt die tussen -100% en +100% kan liggen. Bij een correlatie in de buurt van 100% bewegen de variabelen in tandem. Voorbeeld van een dergelijke positieve correlatie is de verkoop van zonnebrillen en zonnebrandolie. Bij een negatieve correlatie in de buurt van -100% zijn de bewegingen tegenovergesteld. Denk ter illustratie aan de verkoop van badpakken en sneeuwpakken: in de zomer lopen de badpakken goed, in de winter de sneeuwpakken. Bij een correlatie in de buurt van nul is er geen samenhang.

Correlatie is meestal een ervaringsgegeven dat we op het spoor komen in historische data. Een correlatie hoeft niet noodzakelijkerwijs te duiden op een oorzaak-en-gevolg relatie. Dit kan wel, maar hoeft niet. Het is helder dat de verkoop van zonnebrillen zijn oorzaak vindt in de temperatuur en de uren zonneschijn. Maar er is ook een positieve correlatie tussen de afname van het aantal ooievaarsbroedparen sinds de jaren zestig en het teruglopen van het geboortecijfer in diezelfde periode. Ondanks deze correlatie is er waarschijnlijk geen causaal verband!



Figuur 11: Data met 80% (links), 0% (beneden) en -80% (rechts) correlatie.

Correlatie: Hoe reken je het uit in Excel?

In Excel kan correlatie worden berekend door gebruik te maken van (NL) [=correlatie()] of (EN)[=correl()]. Hierbij kunnen twee reeksen van getallen worden ingegeven en Excel geeft dan de correlatie terug. Zoals gezegd bevinden correlaties zich altijd in het domein tussen min één en plus één. Het teken, positief of negatief, geeft aan met welke correlatie men te maken heeft. Een correlatie van nul of dicht bij nul laat zien dat de variabelen vrijwel onafhankelijk zijn. Omgedraaid, een correlatie van -1 of +1 geeft aan dat de variabelen volstrekt afhankelijk zijn. In dit laatste geval kun je je afvragen of deze variabelen apart moeten worden gemodelleerd: ze gedragen zich immers als één geheel.

Covariantie is een andere maat voor hetzelfde verschijnsel. Het verschil tussen deze twee maten is als volgt: waar de correlatie de mate van samenhang weergeeft als percentage van de totale variantie in de data, geeft de covariantie een absolute maat voor de hoeveelheid samenhang tussen data. Wanneer we de standaard-deviaties van A en B en de correlatie tussen A en B weten, dan kunnen we de covariantie uitrekenen:

$$\mathrm{Cov}(A,B) = \sigma(A) \times \sigma(B) \times \rho(A,B)$$

Andersom kunnen we de correlatie ook uitrekenen vanuit de covariantie:

$$P(A,B) = \frac{Cov(A,B)}{\sigma(A) \times \sigma(B)}$$

Bij credit management zijn we meer geïnteresseerd in een relatieve maat voor de samenhang tussen data dan in een absolute maat. Daarom wordt binnen dit vakgebied de correlatie vaker gebruikt dan de covariantie. Er zijn echter vakgebieden waarin dit andersom geldt.

Covariantie: Hoe reken je het uit in Excel?
In Excel kan dit eenvoudig door gebruik te maken van (NL)[=covariantie()] of (EN)[=covar()] of in de nieuwere versies (NL)[=covariantie.p()] en [=covariantie.s()] of (EN)[=covariance.p()] en [=covariance.s()]. Hierbij kunnen twee reeksen van getallen worden ingegeven en Excel antwoordt met de covariantie. Verschil tussen de 'p' (populatie of population) en de 's' (steekproef of sample) is wederom dat men zich moet afvragen of de data betrekking hebben op een deel van het geheel of juist de hele populatie omvatten. Zodra het om wat grotere dataseries gaat maakt dit onderscheid overigens vrijwel niets uit.

De mate van samenhang tussen variabelen heeft grote invloed op de frequentie waarmee een 'worst case scenario' zich voordoet. Correlatie en covariantie kunnen het totale risico versterken maar ook dempen.

Laten we weer naar een voorbeeld kijken. Stel een reddingssloep is zojuist van de zinkende Titanic weggevaren met aan boord een dertigtal passagiers. Deze zijn willekeurig over de sloep verspreid en het bootje vaart redelijk stabiel over de Atlantische Oceaan. Plotseling duikt er een orka aan stuurboord op. Alle passagiers zijn bang voor dit machtige zeedier en haasten zich naar bakboord. De boot raakt hier door instabiel en kapseist. Het systeem wordt instabiel door de grote correlatie in het gedrag van de passagiers die allen bang zijn voor orka's. Een andere reddingssloep heeft onder de dertig passagiers vijftien zeebiologen aan boord. Ook hier duikt een orka op aan stuurboord. Vijftien van de passagiers zijn bang en haasten zich naar bakboord, terwijl de vijftien zeebiologen zich inspannen om de orka te zien aan stuurboord. Wederom is er sprake van correlatie, maar ditmaal negatief. Dankzij de interesse van de zeebiologen blijft het systeem stabiel. In het kort kan men stellen dat positieve correlaties de kans op een onwaarschijnlijke gebeurtenis vergroten, terwijl negatieve correlaties deze kans juist verkleinen.

Risicospreiding

Ook in de portefeuilletheorie spelen correlaties een belangrijke rol. Het gaat hier om verzamelingen financiële instrumenten zoals aandelen, obligaties, hypotheken of beleggingspanden. Het centrale idee achter zo'n portefeuille is dat de risico's zoals weergegeven in de standaarddeviaties van de opbrengsten van deze instrumenten elkaar gedeeltelijk opheffen.

Een voorbeeld: een aandeel A heeft een jaarlijks rendement van 10%. Dit rendement is onzeker en heeft een standaarddeviatie van 5%. Een tweede aandeel, aandeel B, heeft precies hetzelfde rendement en standaarddeviatie, namelijk 10% en een σ van 5%. Gemiddeld gezien heeft een portefeuille van € 1.000 aandeel A en € 1.000 aandeel B hetzelfde rendement als een portefeuille van alleen € 2.000 aandeel A. Beide beleggingsportefeuilles leveren immers eenzelfde gemiddeld rendement van 10% maal € 2.000 is € 200.

Toch zal een verstandige belegger een voorkeur hebben voor de portefeuille met de combinatie van aandeel A en aandeel B omdat het risicoprofiel stukken lager is. Dit risico kan worden afgelezen aan de standaarddeviaties van de portefeuilles. De portefeuille met alleen € 2.000 aandeel A blijkt een standaarddeviatie te hebben van € 100, terwijl de gespreide portefeuille van € 1.000 aandeel A en € 1.000 aandeel B een sigma heeft van € 70.71. Het loont blijkbaar om niet alle eieren in één mandje te leggen. Beleggers proberen het rendement te maximaliseren terwijl ze het risico minimaliseren. Portefeuilles van aandelen of andere financiële instrumenten maken gebruik van het verschijnsel dat die risico's niet allemaal in hetzelfde tempo dezelfde kant oplopen.

Portefeuille A&B	**Waarde**	**Rendement**	**σ**	**Opbrengst**
Aandeel A	€1.000,00	10%	5%	€ 100,00
Aandeel B	€1.000,00	10%	5%	€ 100,00
Portefeuille A&B				€ 200,00
σ Portefeuille A&B				€ 70,71

Portefeuille A	Waarde	Rendement	σ	Opbrengst
Aandeel A	€2.000,00	10%	5%	€ 200,00

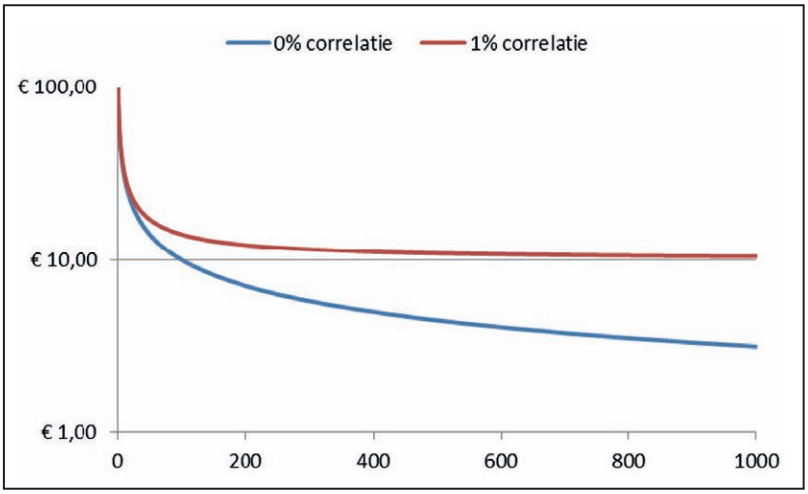
Portefeuille A	€ 200,00
σ Portefeuille A	€ 100,00

Tabel 7: Verwachte opbrengst en risico, onder aanname dat A en B niet onderling gecorreleerd zijn.

In bovenstaand voorbeeld is er impliciet vanuit gegaan dat er geen correlatie is tussen het rendement van aandeel A en dat van aandeel B. In de praktijk zal er wel sprake zijn van correlaties tussen de verschillende instrumenten in een portefeuille. Ideaal is vanzelfsprekend de negatieve correlatie waarbij de risico's elkaar min of meer opheffen in de portefeuille. Het is echter moeilijk op de beurs aandelen te vinden die een negatieve correlatie vertonen. De meeste aandelen zijn vrij sterk positief gecorreleerd omdat ze allemaal onderhevig zijn aan hetzelfde beurssentiment. Voor andere instrumenten zoals obligaties of onroerend goed geldt dit evenzo.

Portefeuilletheorie: Het belang voor de credit manager
Hoewel men de portefeuilletheorie vooral ziet als onderdeel van de beleggingsleer is de theorie ook voor credit managers van belang. Uitstaande vorderingen zijn immers ook activa, net als aandelen en obligaties dit kunnen zijn. Risicospreiding en waardering zijn evengoed op debiteurenvorderingen van toepassing.

Wanneer een verzameling aandelen onderling een positieve correlatie heeft lager dan 100%, dan is het nog wel zinvol om risico te spreiden, maar zal het effect altijd minder groot zijn dan wanneer er 0% correlatie zou zijn.



Figuur 12: Standaarddeviatie van een portfolio afgezet tegen het aantal portfolio onderdelen met 0% correlatie(blauw) of 1% correlatie(rood).

Correlaties van portefeuilles of van variabelen in een model worden vaak weergegeven in een correlatiematrix of correlatietabel. Een dergelijke matrix lijkt op een afstandstabel tussen verschillende steden. In het geval van de correlatiematrix worden alle variabelen op de horizontale as en op de verticale as van de tabel weergeven.

Correlatie-Matrix	Variabele A	Variabele B	Variabele C
Variabele A	**1,000**	0,700	0,700
Variabele B		**1,000**	0,700
Variabele C			**1,000**

Tabel 8: Een voorbeeld van een correlatietabel of correlatiematrix.

Op de kruispunten van de assen staan de correlaties weergegeven. De diagonale as bevat steeds de waarde één of 100% omdat iedere variabele uiteraard perfect met zichzelf is gecorreleerd. Verder kan de helft van de matrix leeg blijven omdat als eenmaal de relatie tussen variabele A en variabele B is gedefinieerd, het niet noodzakelijk is om ook de correlatie tussen B en A weer te geven. Deze zijn gelijk aan elkaar.

Correlaties: Het belang voor de credit manager
Credit managers zijn op zoek naar manieren om risico te verlagen. Een veelgebruikte methode is het spreiden van risico: wanneer niet alle krediet aan één cliënt uitgegeven wordt, maar verspreid wordt over meerdere cliënten, wordt aangenomen dat niet alle schuldenaren tegelijk betaling zullen weigeren.

Het probleem is dat het effect van risicospreiding gemakkelijk overschat wordt doordat vaak gerekend wordt met een correlatie van 0%. De praktijk blijkt vaak anders: als de economie slecht draait dan zit iedereen krapper bij kas, en als het weer in de zomer tegenvalt, draait de hele toeristische sector op een laag pitje. Zelfs kleine correlaties kunnen het effect van verregaande risicospreiding al inperken.

Voorts kunnen debiteurenrisico's veel groter zijn dan men zich bewust is. Positieve correlaties tussen debiteurenvorderingen leiden ertoe dat de economische golfbeweging wordt versterkt. In goede tijden zullen de meeste bedrijven ook lage afschrijvingen hebben op wanbetaling. Met de meeste debiteuren gaat het goed en de onderlinge correlaties versterken dit effect waardoor er over de gehele linie weinig problemen zijn. Zodra de wind draait is de kans groot dat niet een enkel een verdwaald schaap in de problemen komt maar de hele kudde kredietverliezen vertoont. We zien dat correlaties er in dit voorbeeld toe leiden dat de economische golfbeweging wordt versterkt: in goede tijden surfen we allemaal op een mooie draaggolf, in slechte tijden gaan we collectief ten onder in de tsunami van debiteurenproblemen.

De les voor de credit manager is hier om geen overoptimistische aannamen te doen, en om bij het analyseren van historische data te blijven onderzoeken of verdere risicospreiding nog wel voldoende effectief is.

Staartafhankelijkheid

In veel situaties is er weinig tot geen samenhang tussen toevalsvariabelen onder normale omstandigheden, maar des te meer onder extreme omstandigheden. Dit wordt staartafhankelijkheid genoemd omdat de afhankelijkheid in de staart van de kansverdeling zit.

Een voorbeeld van staartafhankelijkheid is de samenhang tussen claims van vloedverzekeringen en claims van windschadeverzekeringen in Nederland: Onder normale omstandigheden zullen deze onafhankelijk van elkaar bewegen, maar wanneer, bij een extreem zware storm, de deltawerken doorbreken, schieten beide omhoog.

Het risico wordt voor een groot deel bepaald door de kans op het 'worst case' scenario, of het slechtste scenario dat we kunnen bedenken. Staartafhankelijkheid kan er daarom voor zorgen dat het spreiden van risico minder effectief wordt.

Staartafhankelijkheid: Het belang voor de credit manager
Correlaties kunnen soms moeilijk te ontdekken zijn. Dit is ook het geval bij staart-afafhankelijkheid waarbij de correlatie alleen maar optreedt in extreme gevallen, wanneer zaken mis gaan. Omdat dergelijke extreme situaties meer uitzondering dan regel zijn is het moeilijk deze correlaties op te pikken in een statistische analyse. Variabelen lijken onder normale omstandigheden niet gecorreleerd te zijn maar blijken dit uiteindelijk wel te zijn onder druk. Als credit managers krijgen we als het ware zand in de ogen gestrooid.

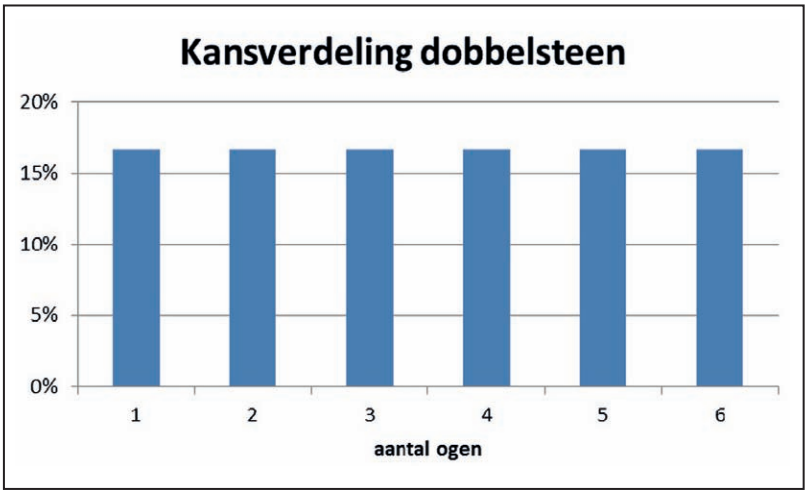
De grootte van het risico voor de credit manager wordt voor een groot deel bepaald door extreme situaties. Belangrijke vragen in het risicomanagement zijn dan ook: wat is het ergste dat kan gebeuren? (het zogenaamde worst case scenario) en hoe groot is de kans dat het gebeurt?

De credit manager doet er dan ook goed aan om bij het analyseren van historische data juist ook data te bekijken van perioden met veel tegenwind. Alleen op die manier kan de statistiek een betrouwbare schatting van het risico geven.

Kansverdelingen

Wat is een kansverdeling?

Een kansverdeling of distributie is een formule of grafiek die het gedrag van een kansvariabele beschrijft. De kansverdeling geeft op de horizontale x-as aan welke waarden deze toevalsvariabele kan aannemen en op de verticale y-as hoe groot de kans is dat deze variabele voorkomt.



Figuur 13: Voorbeeld van een kansverdeling.

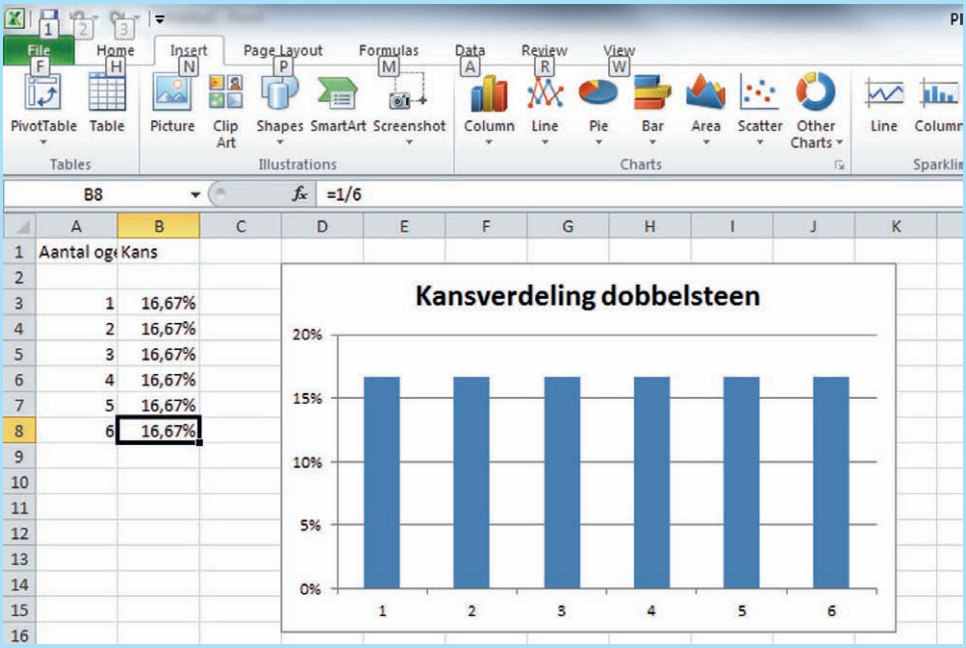
Aan de hand van het eerder gegeven voorbeeld van de kansverdeling van een dobbelsteen kan dit worden gedemonstreerd. De mogelijke waarden die de kansvariabele, ook wel *stochast* genoemd, kan aannemen zijn één, twee enzovoorts tot en met zes. Deze waarden zien we op de horizontale as. De kans is af te lezen op de verticale as. In alle gevallen is deze 1/6 of 16,67 %.

De blauwe kolommen in de grafiek geven alle mogelijke uitkomsten, dat wil zeggen de gewichten, en hun relatieve kansen. Bij iedere worp met de dobbelsteen, bij iedere verpakking die wordt afgevuld wordt deze distributie voor de toevalsvariabele weer gevolgd. De kansen op alle mogelijke uitkomsten vooraf, zijn opgeteld steeds één of 100%: de zes vlakken van de dobbelsteen hebben ieder een kans van 1/6 en tellen op naar één. Na het rollen van de dobbelsteen met bijvoorbeeld als resultaat zes, is de kans op een, twee, drie, vier en vijf in alle gevallen 0% en de kans op zes 100% omdat er nu zekerheid is over het resultaat. Het is belangrijk te zien dat het groene gebied onder de curve van de distributie alle mogelijke kansen vooraf vertegenwoordigt, dus voordat ik de dobbelsteen rol en voordat ik de zak vul. Doordat de oppervlakte in feite de optelsom weergeeft

van alle mogelijkheden en kansen kunnen deze distributies worden gebruikt voor allerhande interessante analyses.

Kansverdelingen: Hoe doe je het in Excel?
Om met Excel een plaatje van een kansverdeling te maken, is het eerst nodig om een reeks te maken met alle mogelijke uitkomsten. We zullen deze reeks in het vervolg X noemen. Als het om een uniforme verdeling gaat, deel de verdeling dan op in een eindig aantal intervallen met gelijke breedte.
Vervolgens maak je een reeks waarin je voor iedere mogelijke uitkomst de kans zet. Deze reeks zullen we in het vervolg Y noemen.

Vervolgens maak je een grafiek van de Y-reeks, en gebruik je de X-reeks als label. Het is gebruikelijk om een kolommendiagram te gebruiken voor discrete verdelingen, en een lijndiagram of oppervlaktediagram voor continue verdelingen.



Figuur 14: Een verdeling maken in Excel.

Kansverdeling of histogram?

De oplettende lezer zal gemerkt hebben dat kansverdelingen en histogrammen een aantal gelijkenissen vertonen. Het verschil is echter dat een histogram een weergave is van een verdeling zoals die bij een steekproef gevonden is, terwijl een kansverdeling het totaalplaatje geeft van de achterliggende verdeling.

Uit de wet van de grote aantallen volgt dat de vorm van een histogram steeds meer op de onderliggende kansverdeling gaat lijken wanneer de omvang van de steekproef wordt vergroot, maar er zal altijd een kleine onzekerheid overblijven.

In de praktijk worden vaak steekproeven en histogrammen gebruikt om de onderliggende kansverdeling te schatten. Exacte kansverdelingen bestaan dan ook alleen bij gedachtenexperimenten en als stochastische modellen.

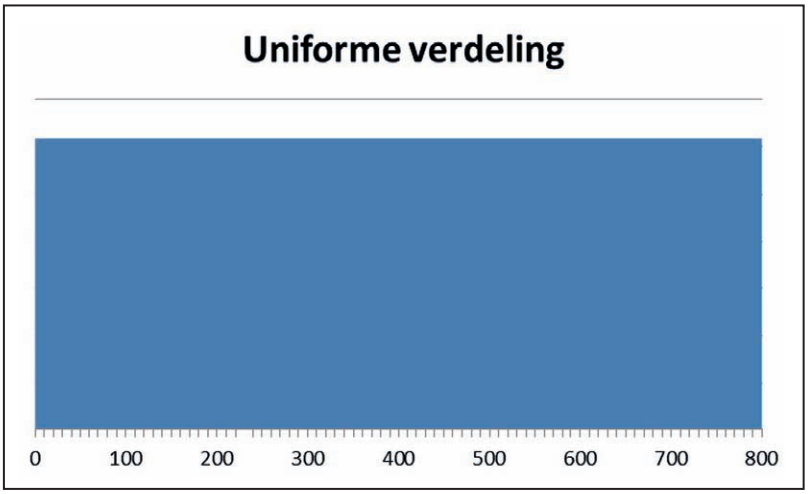
Welke kansverdeling?

Bij het modelleren van variabelen heeft men de keuze uit vele kansverdelingen. Welke distributie het best kan worden verbonden met de desbetreffende variabele is niet altijd eenduidig te zeggen. Hier komt enig vakmanschap van de modellenmaker bij kijken. Van de andere kant, er zijn wel enige vuistregels te geven en met nuchter verstand komt men ook ver. Het is goed te beseffen dat een variabele vaak met meerdere distributies kan worden gemodelleerd en er niet altijd één goede oplossing is. Voorts zijn er vele kansverdelingen in de gereedschapskist te vinden maar kan men met een handvol distributies al uit de voeten. Met een hamer, zaag, winkelhaak en schroevendraaier kan men een tafel maken. In dit hoofdstuk zullen de meest gebruikte kansverdelingen kort en bondig worden behandeld. Bij de keuze van de juiste distributie helpt het om de vraag te stellen of het hier om een continue of discrete variabele gaat en of de variabele symmetrisch of asymmetrisch is. Door deze vragen te beantwoorden kan men het aantal mogelijke verdelingen waaruit kan worden gekozen snel terugbrengen.

Uniforme verdeling

De eenvoudigste verdeling is de uniforme verdeling. Deze distributie wordt gedefinieerd door de minimum- en maximumwaarde vast te leggen. De kans op iedere waarde in dit gebied is even groot, vandaar de naam uniform. Men kan de uniforme verdeling kiezen omdat de onderliggende kansvariabele het beste op deze wijze kan worden gemodelleerd, maar ook omdat verdere gegevens ontbreken. Vaak geven mensen een schatting in de vorm van range: ik ben met dit karwei zes à tien uur bezig, of ik heb tien tot twaalf kilo materiaal nodig. Als men vraagt dit verder te preciseren kan men dit niet. Je zou ervoor kunnen kiezen het model te baseren op de gemiddelde waarden, dus acht uur en tien kilo maar dit doet geen recht aan de onzekerheid die werd gecommuniceerd. Hier dwingt

het gebrek aan precieze gegevens ons de uniforme verdeling te gebruiken. Een voorbeeld waar de uniforme verdeling principieel de beste keuze is bijvoorbeeld de simulatie van de plek waar zich een lekkage kan voordoen in een 800 km lange oliepijplijn. Het lek kan overal de kop opsteken: na 10 meter, 200 kilometer, 600 kilometer enzovoorts. Hier is op voorhand niets over te zeggen.

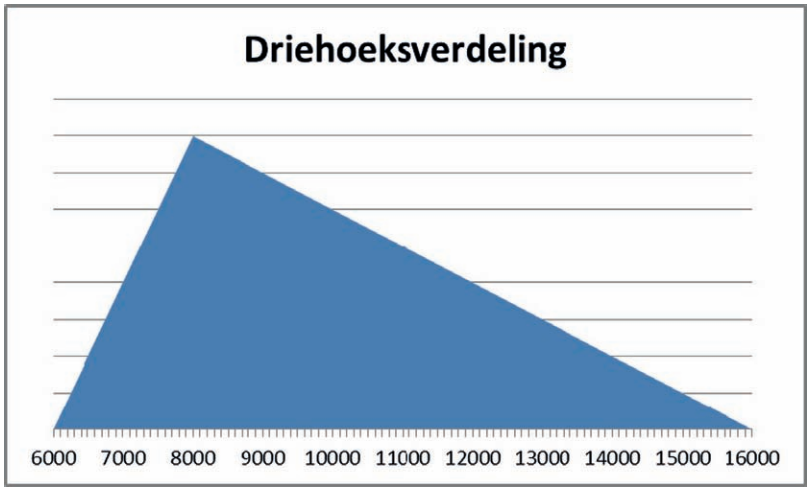


Figuur 15: Voorbeeld van een uniforme kansverdeling.

De uniforme verdeling is symmetrisch omdat de lengte van de grafiek links van het midden even hoog is als rechts van het midden. Het gaat hier om een continue verdeling omdat er geen discrete stappen zijn in de mogelijke uitkomsten, zoals bij het gooien van een dobbelsteen.

Driehoeksverdeling

Een eenvoudig te begrijpen en robuuste kansverdeling die vaak wordt toegepast is de driehoeksverdeling. Men definieert deze verdeling aan de hand van de meest waarschijnlijke waarde of de top van de driehoek, en een minimum en maximum als linker- en rechterhoek van de driehoek. Stel je moet een schildersbedrijf laten komen om regulier onderhoud te laten doen aan het houtwerk van het bedrijfspand. De vorige keren betaalde men ongeveer € 8.000,-. Je verwacht dat je nu weer een vergelijkbaar bedrag krijgt geoffreerd. De kans dat het veel goedkoper wordt is aanwezig, maar niet erg groot en lager dan € 6.000,- zal het nooit uitpakken. Van de andere kant is het wel mogelijk dat er extra kosten zijn en er meer dan de verwachte € 8.000,- moet worden afgerekend. Echter, meer dan het dubbele, dus € 16.000,-, lijkt onwaarschijnlijk. Dit is een goed voorbeeld van een driehoeksverdeling omdat we een meest waarschijnlijke waarde hebben, namelijk € 8.000,- en een minimum en maximum.



Figuur 16: Voorbeeld van een driehoeksverdeling.

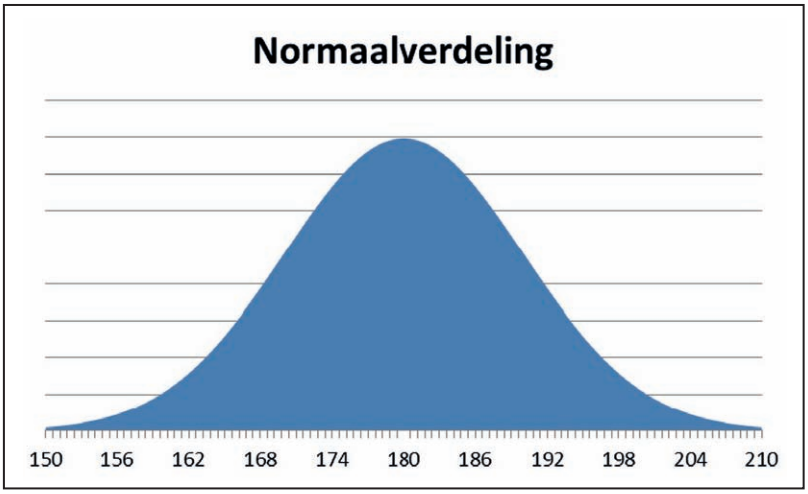
De top van de driehoek die de meeste waarschijnlijke of meest voorkomende waarde aangeeft, is in statistische termen de modus. De driehoeksverdeling is continu. In ons voorbeeld is de driehoeksverdeling ook asymmetrisch: de staart van de grafiek naar rechts is veel langer dan die naar links. Ook valt op dat het gemiddelde niet samenvalt met de top en modus van de driehoek. Dit vindt zijn oorzaak in de asymmetrische vorm waarbij budgetoverschrijdingen veel waarschijnlijker zijn dan onderbestedingen. Driehoeksverdelingen kunnen overigens symmetrisch zijn. Hiervoor is het nodig dat de topwaarde in het midden ligt tussen het opgegeven minimum en maximum. Bij een symmetrische driehoeksverdeling vallen de modus en het gemiddelde samen, bij asymmetrische driehoeksverdelingen niet. In de civiele techniek wordt de driehoeksverdeling vaak LTU genoemd, wat staat voor Laagste, Top en Uiterste waarde.

Normaalverdeling

De meest bekende kansverdeling is zonder twijfel de normaalverdeling, die ook wel door het leven gaat als Gauss-curve, naar de Duitse wiskundige Friedrich Gauss, of in het Engels als *bell-curve*, naar de klokvorm van de grafiek. De normaalverdeling laat zich definiëren door het gemiddelde en de standaarddeviatie. Het is een continue, symmetrische kansverdeling. De populariteit heeft de normaalverdeling te danken aan het feit dat veel verschijnselen in de natuur maar ook in de menswetenschappen te beschrijven zijn met de normaalverdeling. Daarnaast speelt de normaalverdeling een cruciale rol in de centrale limietstelling die we al eerder hebben aangestipt. De centrale limietstelling stelt dat, onder normale omstandigheden, de gemiddelden die ik vind bij het nemen van een serie steekproeven normaal verdeeld zijn rond het echte gemiddelde van de populatie.

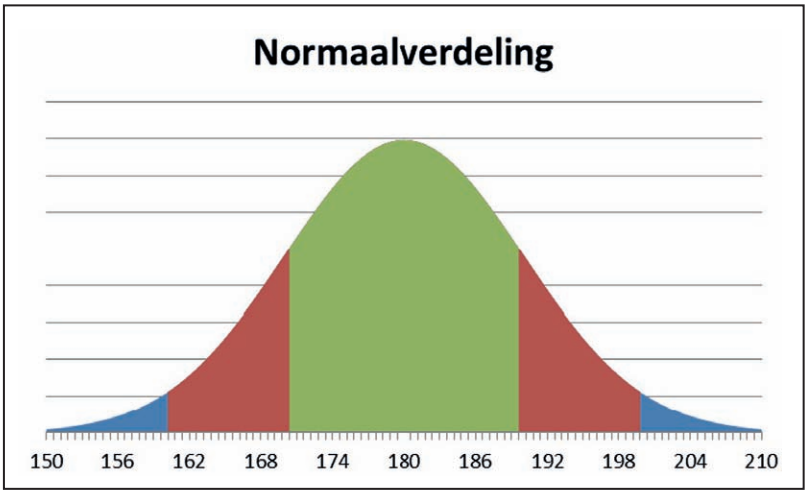
Hoe meer steekproeven ik neem, hoe dichter het gemiddelde komt bij het echte gemiddelde van de populatie.

Laten we eens naar het bekende voorbeeld kijken van de verdeling van lengtes van mensen. Deze zijn normaal verdeeld. Stel ik meet de lengte van duizend volwassen mannen in de Amsterdamse Kalverstraat. Ik neem het gemiddelde van deze duizend metingen en concludeer dat dit 1,80 m is en reken uit in Excel dat de standaarddeviatie 10 cm is. Als ik de lengte van volwassen mannen als variabele wil modelleren, kan worden volstaan met een normaalverdeling met 180 cm als gemiddelde en een standaardafwijking, vaak weergegeven met de Griekse letter σ, van 10 cm.



Figuur 17: Voorbeeld van een normaalverdeling.

De lengte van een willekeurige nieuwe volwassen man die langs komt zal dus waarschijnlijk in de buurt van de 1,80 m liggen met een 50% kans dan hij langer is dan 1,80 m en 50% kans dat hij kleiner is. Dit is ook het juiste moment om terug te komen op de standaarddeviatie: in ons voorbeeld is de standaardafwijking 10 centimeter. Velen hebben moeite om een gevoel te krijgen wat deze standaarddeviatie inhoudt. We weten dat de σ wordt weergegeven in de zelfde eenheden als het gemiddelde, in ons voorbeeld dus centimeters of meters. Als we nu naar de distributie kijken en vanaf het gemiddelde één standaarddeviatie naar links gaan, komen we uit bij 1,70 m. Gaan we vanaf het gemiddelde van 1,80 één σ naar rechts komen we uit bij 1,90 m. Nu is bij een normaalverdeling het gebied dat wordt begrensd door één sigma links, en één sigma rechts 68% van de totale oppervlakte van de grafiek. Dit wil zeggen dat 68% van de volwassen mannen langer is dan 1,70 m en kleiner dan 1,90 m.

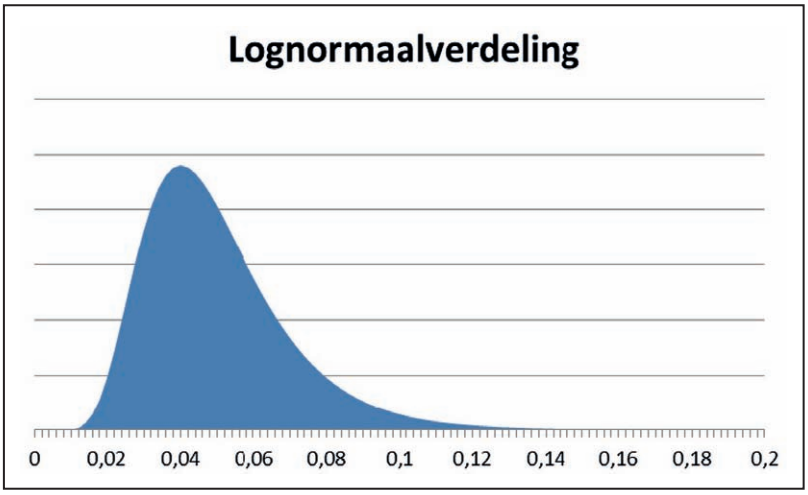


Figuur 18: Bij een normaalverdeling is er 68% kans om binnen één standaarddeviatie van het gemiddelde uit te komen (geel) en 95% kans om binnen twee standaarddeviaties van het gemiddelde uit te komen (rood + geel).

Herhalen we deze exercitie maar ditmaal met twee standaarddeviaties, dus tweemaal tien centimeter, dan komen we links van het midden uit op 1,60 m en rechts op 2,00 m. De oppervlakte van dit gebied beslaat 95% van de normaalverdeling. De interpretatie is dat 95% van de mannen tussen de 1,60 en 2,00 m lang is.

Lognormaal verdeling

De lognormale verdeling is een distributie die gebruikt kan worden voor asymmetrische variabelen. Dit zijn kansvariabelen waarbij de lengte van de rechterstaart van de verdeling langer is dan de linkerstaart. Aan de linkerkant is de distributie begrensd door een hard minimum, aan de rechterkant is er meestal geen plafond aan de mogelijke uitkomsten. De lognormale verdeling is vooral een goede keuze bij het modelleren van zaken als rentetarieven, reparatietijden, prijzen enzovoorts. Rentetarieven zijn bijvoorbeeld gemiddeld 5% en kunnen in een normaal functionerende economie niet lager zijn dan 0%. Echter, ze kunnen aan de andere kant wel zeer hoog zijn. Er zijn landen met hyperinflatie en rentetarieven van duizenden procenten. Hetzelfde geldt voor prijzen: deze kunnen niet negatief worden, dus nul is de ondergrens, maar een plafond voor prijzen bestaat niet. Wat is de maximale prijs voor olie in de wereldhandel?



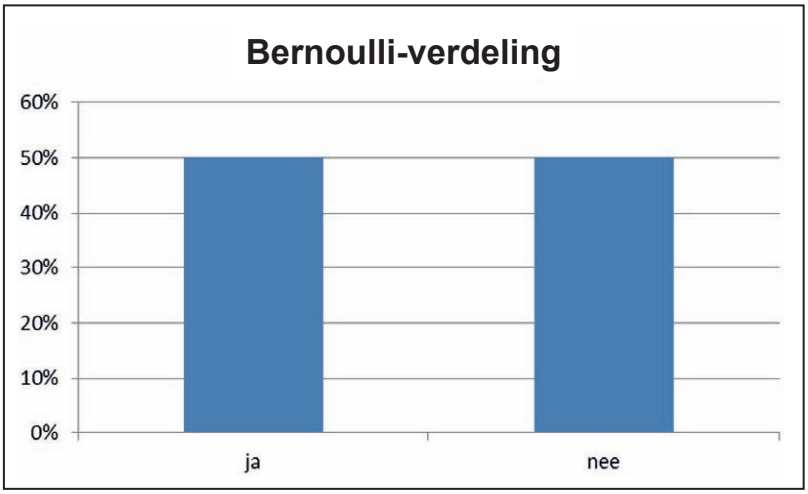
Figuur 19: Voorbeeld van een lognormaalverdeling.

De lognormale verdeling is een continue distributie. De parameters gemiddelde en de standaardafwijking bepalen de lognormale distributie. In de grafiek is duidelijk te zien dat de distributie scheef is naar rechts. Het valt ook op dat het gemiddelde niet samenvalt met de top van de grafiek. De top is in statistische termen de modus, dat wil zeggen de waarde die het meest voorkomt. Het gemiddelde ligt rechts van deze modus omdat het gewicht van de lange staart naar rechts meegenomen wordt in de berekening van dit gemiddelde.

Bernoulli of ja/nee-verdeling

De Bernoulli verdeling is genoemd naar de Zwitserse wiskundige Jacob Bernoulli. Deze verdeling wordt ook wel de ja/nee-verdeling genoemd omdat de kansvariabele onder deze distributie maar twee waarden kan aannemen. De kansen op ieder van deze waarden is opgeteld uiteraard weer honderd procent. Het voorbeeld dat ons onmiddellijk te binnen schiet, is uiteraard het gooien met een muntstuk. Er zijn twee mogelijke uitkomsten: kop en munt. De kans op kop is 50% en de kans op munt is ook 50%, wat optelt naar 100%.

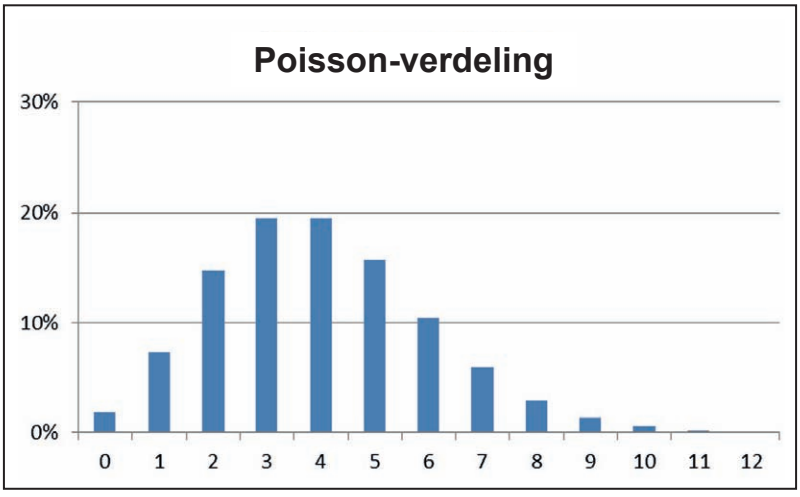
De Bernoulli-distributie is eenvoudig te definiëren aan de hand van één parameter: we hoeven alleen maar aan te geven wat de kans op 'ja' is. In het geval van het muntstuk is deze kans 50% voor 'ja'. Daarmee is de andere mogelijke uitkomst ook gedefinieerd want 100% - 50% = 50%. De kansen zijn niet altijd 50% / 50% verdeeld. Bij een examen is de slaagkans bijvoorbeeld 90%. We kunnen deze slaagkans dan modelleren als een Bernoulli-distributie met als 'ja' een kans van 90% en 'nee', dit wil zeggen gezakt, een waarschijnlijkheid van 10%. Deze distributie is discreet omdat er duidelijk maar twee mogelijke waarden zijn voor de kansvariabele.



Figuur 20: Voorbeeld van een Bernoulli-verdeling.

Poisson-verdeling

De Poisson-verdeling is een discrete verdeling die wordt gebruikt voor het modelleren van telbare aantallen. Voorbeelden zijn: het aantal klanten per uur, het aantal telefoontjes per dag, het aantal zetfouten per pagina enzovoorts. De verdeling is genoemd naar de Franse wiskundige Siméon Poisson. De Poisson-verdeling kent maar één parameter te weten de verwachtingswaarde. Deze ratio wordt meestal weergegeven met de kleine Griekse letter λ (lambda). Laten we weer eens een voorbeeld nemen. Stel iemand bestelt meestal vier consumpties tijdens etentjes in een restaurant. Dit is een discreet gegeven want men kan niet 2,73 glazen wijn bestellen, het is of nul, één, twee, drie enzovoorts glazen.



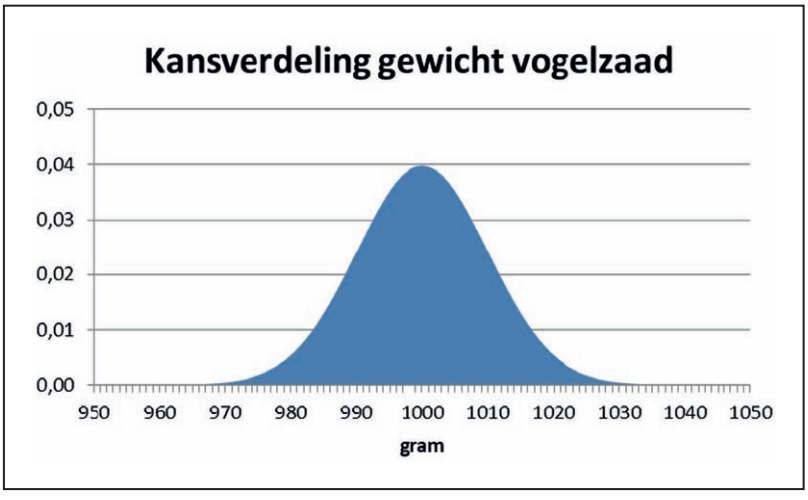
Figuur 21: Voorbeeld van een Poisson verdeling.

Het mooie van de Poisson-verdeling is dat men door het definiëren van deze ratio λ zowel het gemiddelde als de variantie en daarmee de standaarddeviatie vastlegt. Zowel het gemiddelde als de variantie is bij Poisson-verdeling gelijk aan de lambda. In het voorbeeld van de restaurantconsumpties is λ vier en daarmee de variantie ook vier. De standaardafwijking is dan de wortel uit vier: twee.

Rekenen met kansverdelingen

Wanneer we de kansverdeling van een toevalsvariabele kennen dan kunnen we hier interessante vraagstukken mee oplossen.

Laten we naar een voorbeeld kijken dat iedereen zal aanspreken: het afvullen van zakken parkietenzangzaad in een diervoederfabriek. De fabriek levert zakken van een kilogram. De afvulmachine levert niet precies 1.000 gram: soms is het iets meer, soms is het iets minder. Deze onzekerheid met betrekking tot de afleverhoeveelheid komt tot uitdrukking in de standaarddeviatie, die 10 gram bedraagt. De hoeveelheid die de machine levert varieert per verpakking en is dus een toevalsvariabele. We kunnen stellen dat de afgeleverde hoeveelheden normaal zijn verdeeld. De machine levert in de helft van de gevallen minder dan 1.000 gram en in de helft van de gevallen meer dan 1.000 gram. In onderstaande grafiek zien we hoe deze kansverdeling eruit ziet. Het gemiddelde definieert het midden van de kansverdeling, in ons geval één kilogram, de afleverhoeveelheden staan horizontaal, de kans staat op de verticale as. Als we dit plaatje bestuderen is duidelijk te zien dat de kans op een afleverhoeveelheid in de buurt van de 1.000 gram het hoogst is. Hoe verder je van het midden en dus ook van het gemiddelde kijkt, hoe minder waarschijnlijk het wordt dat dit zich voordoet. De staarten links en rechts van de normaalverdelingen naderen de x-as snel als je weggaat van het midden.



Figuur 22: Verdeling van het gewicht van een kiloverpakking vogelzaad.

Rekenen met kansverdelingen: Hoe doe je het in Excel?
Excel kent een aantal functies om te helpen rekenen met kansverdelingen. Zo is het mogelijk om, voor een gegeven kansverdeling, uit te rekenen hoe groot de kans is om op of onder een zekere waarde uit te komen. In onderstaande tabel staat weergegeven welke Excel functie gebruikt kan worden voor welke verdeling.

Verdeling	Oude Excel (NL)	Nieuwe Excel (NL)
Normaal	[=norm.verd(X*;μ;σ;T*)]	[=norm.verd.μ(X*;μ;σ;T*)]
Lognormaal	[=log.norm.verd(X*;μ;σ;T*)]	[=lognorm.verd.μ(X*;μ;σ;T*)]
Binomiaal	[binomiale.verd(X*;N;p;T*)]	[binom.verd(X*;N;p;T*)]
Poisson	[=poisson(X*;λ;T*)]	[=noisson.verd(X*;λ;T*)]

X* = de waarde waarvoor de kans uitgerekend moet worden.
T* = ONWAAR als de kans op waarde X uitgerekend moet worden,
WAAR als de kans op een waarde lager dan X uitgerekend moet worden.

Verdeling	Oude Excel (EN)	Nieuwe Excel (EN)
Normaal	[=normdist (X*;μ;σ;T*)]	[=norm.dist(X*;μ;σ;T*)]
Lognormaal	[=lognormdist(X*;μ;σ;T*)]	[=lognorm.dist(X*;μ;σ;T*)]
Binomiaal	[binomdist(X*;N;p;T*)]	[binom.dist(X*;N;p;T*)]
Poisson	[=poisson(X*;λ;T*)]	[=noisson.dist(X*;λ;T*)]

X* = de waarde waarvoor de kans uitgerekend moet worden.
T* = FALSE als de kans op waarde X uitgerekend moet worden,
TRUE als de kans op een waarde lager dan X uitgerekend moet worden.

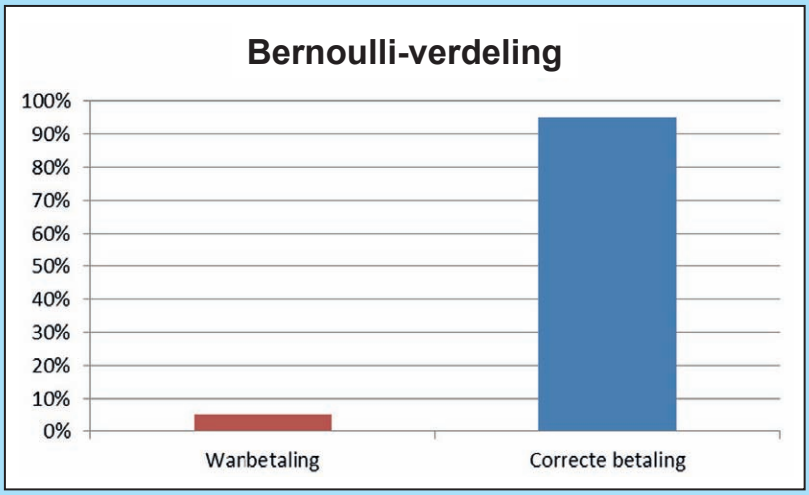
Zou je 50 gram, dus 5 standaarddeviaties, onder het gemiddelde gaan, dan is de kans dat de verpakking gevuld wordt met minder dan 950 gram kleiner dan 0,00003%. Dus minder dan eenmaal per drie miljoen zakken krijgt de klant minder dan 950 gram geleverd in zijn kiloverpakking. Dit is een verwaarloosbare kleine kans en het is duidelijk dat de grote meerderheid van verpakkingen een gewicht heeft dicht in de buurt van de beloofde 1.000 gram.

Stel dat de directie van de zangzaadfabriek stelt dat men het aan de afnemers verplicht is om minimaal 1.000 gram te leveren. Men wil deze norm met een zekerheid van 95% garanderen. Op welk gewicht stellen we de afvulmachine af om deze kwaliteitsnorm te halen?

Om deze norm te halen moet de machine gemiddeld iets teveel leveren, maar hoeveel teveel? Het antwoord is te vinden door het te leveren gemiddelde zover op te schuiven dat de kans om lager dan 1.000 gram uit te komen minder is dan iets minder dan vijf procent. Hiervoor moet de instelling van de machine

naar gemiddeld 1.017 gram. Met deze instelling kan de leverancier garanderen dat in 95% van de zakken ten minste één kilo zit. Dit voorbeeld laat zien dat kansverdelingen behulpzaam kunnen zijn bij het oplossen van bedrijfskundige en financiële vraagstukken.

Kansverdelingen: Het nut voor de credit manager
Kansverdelingen beschrijven verwachtingen voor de toekomst. De kansverdeling geeft de verschillende toekomstige waarden van een variabele aan en de bijbehorende kans. De ja/nee-verdeling kan bijvoorbeeld worden gebruikt om een model te maken dat aan de ene kant de kans op wanbetaling aangeeft en aan de andere kant de kans op correcte betaling.



Figuur 23: Bernoulli-verdeling wanbetalingen.

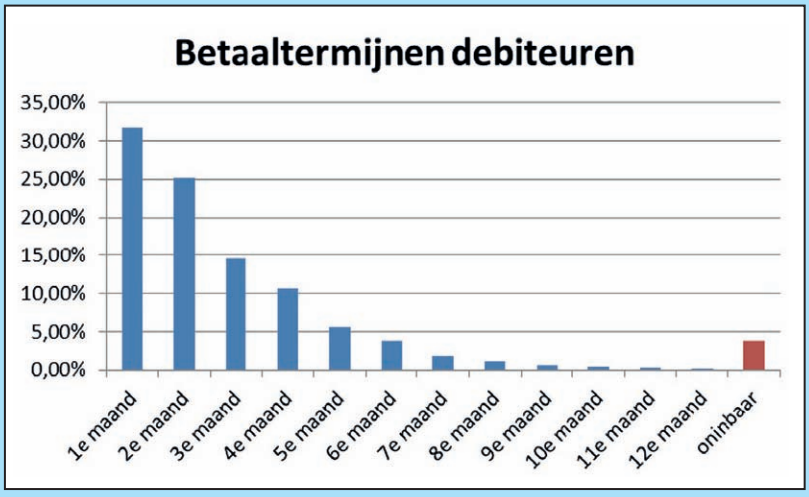
We kunnen ook kijken naar bijvoorbeeld het betaalgedrag en dit in een distributie weergeven. Stel we kijken naar een duizendtal facturen langer dan een jaar oud en we willen een model bouwen dat voorspelt in hoeveel maanden een factuur wordt voldaan. Omdat we ons model baseren op historische data gaan we ervan uit dat het betaalbedrag in de toekomst ongeveer gelijk zal zijn aan het debiteurengedrag uit het verleden. In een frequentietabel registeren we hoe snel facturen worden betaald: binnen één maand, twee maanden enzovoorts. Vorderingen die langer dan een jaar uitstaan schrijven we af. We krijgen zo dertien categorieën. De aantallen kunnen we omzetten naar percentages: als 318 van de duizend facturen binnen één maand worden betaald dan is de kans op deze uitkomst:

$$\frac{318}{1.000} = 31{,}8\%$$

Betaaldatum	#	%
1e maand	318	31,8%
2e maand	252	25,2%
3e maand	146	14,6%
4e maand	107	10,7%
5e maand	56	5,6%
6e maand	38	3,8%
7e maand	18	1,8%
8e maand	12	1,2%
9e maand	6	0,6%
10e maand	4	0,4%
11e maand	3	0,3%
12e maand	2	0,2%
	38	3,8%

Tabel 9: Debiteurengedrag.

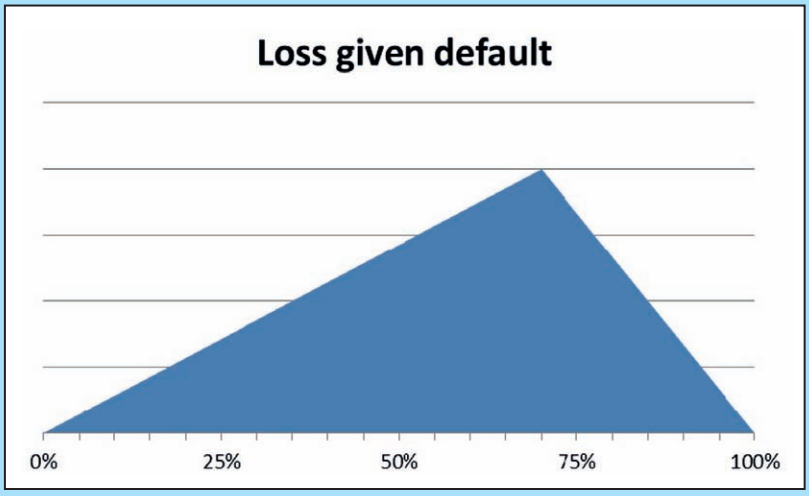
Deze frequentietabel is de basis voor een kansverdeling of distributie waarbij de verschillende betaaltermijnen met hun respectievelijke kansen het model vormen dat iets voorspelt over de toekomst. Voor een nieuwe factuur kunnen we het model gebruiken om te bepalen hoe snel deze betaald gaat worden.



Figuur 24: Verdeling debiteurengedrag.

Zouden we dit model laten draaien onder een Monte Carlo-simulatie, dan zal in ongeveer 32% van de gevallen de factuur binnen de eerste maand worden geïnd. Let op: het gaat om ongeveer 32% omdat een Monte Carlo-simulatie een stochastisch proces is dat voor een gedeelte op toeval is gebaseerd.

Een ander voorbeeld is het modelleren en simuleren van het te verwachten verlies bij wanbetaling of de Loss Given Default *(LGD). LGD is het omgedraaide of complement van het terugvorderingspercentage oftewel de . Het idee is dat bij wanbetaling een deel van de vordering kan worden teruggehaald door verkoop van onderpanden of claims of door een beroep te doen op garanties. De LGD ligt tussen 0% in het slechtste geval en 100% in het beste geval. Laten we aannemen dat uit de historische data blijkt dat de LGD doorgaans op 70% ligt. De LGD zouden we kunnen modelleren als een driehoeksverdeling. De hoeken van deze driehoek worden links en rechts gevormd door het minimum 0% en het maximum 100%. De top van de driehoek is de meest verwachte waarde: 70%. De driehoeksverdeling is nu het model van de LGD. In een Monte Carlo-simulatie laten we de LGD voor iedere vordering bepalen door de beschreven driehoeksverdeling. Iedere keer als we een simulatiestap uitvoeren zal het toeval een andere waarde toekennen aan de LGD. Dit is een stochastisch proces, maar dit toevalsproces wordt wel beheerst door de kansverdeling die in het model vastligt.*



Figuur 25: Verdeling Loss Given Default.

De variabele gedraagt zich in onze simulatie in lijn met deze driehoeksverdeling waarbij de waarden altijd in het bereik 0% tot 100% liggen en de kans dat waarden in de buurt van de top van 70% voorkomen het grootst is.

Modellen en simulaties

Een model is een vereenvoudigde weergave van de werkelijkheid. We gebruiken modellen om deze complexe werkelijkheid beter te begrijpen. Een model is dus een nabootsing die in feite nooit perfect is omdat het nooit alle details van de werkelijkheid kan omvatten. Een goed voorbeeld van een model is een landkaart die een vereenvoudigde weergave is van een landschap. Per definitie staat niet alles op de landkaart: kleine paadjes staan bijvoorbeeld niet weergegeven op kaart en de kaart of het model schiet tekort in het volledig weergeven van de werkelijkheid.

In de wiskunde en economie maken we volop gebruik van modellen. Deze modellen geven meestal de samenhang tussen invoervariabelen en uitvoervariabelen. Laten we eens kijken naar een eenvoudig model dat we allen meteen zullen herkennen:

Omzet = prijs x hoeveelheid

De uitvoervariabele is de 'omzet' omdat deze afhangt van de invoervariabelen 'prijs' en 'hoeveelheid'. Als prijs of hoeveelheid wijzigen verandert de omzet. De afhankelijke variabele omzet volgt dus de invoervariabelen prijs en hoeveelheid.

We kunnen de invoervariabelen variëren en kijken wat er gebeurt met de uitvoervariabele. We gaan als het ware spelen met de prijs en hoeveelheid en kijken wat voor een effect dit heeft op de omzet. Dit spelen met de variabelen noemen we ook wel simuleren en het toepassen van een model om de werkelijkheid na te bootsen heet daarom een simulatie.

Simuleren: hoe doe je het in Excel?
Om in Excel een eenvoudige simulatie te bouwen, heb je een beperkt aantal basisvaardigheden nodig. Zo moet je weten hoe je kunt optellen en vermenigvuldigen in Excel, en moet je verwijzingen naar andere cellen kunnen maken. Het is ook nuttig, hoewel niet absoluut noodzakelijk, om het verschil tussen absolute en relatieve verwijzingen te weten. Wanneer je deze basisvaardigheden beheerst, is de enige overgebleven uitdaging om de gegevens op een overzichtelijke manier te ordenen.

Simulaties kunnen, zeker als de modellen ingewikkelder worden, een flinke hoeveelheid rekenwerk vereisen en worden daarom vaak met behulp van een computer uitgevoerd, maar dit is geen vereiste.

Laten we eens kijken of we aan de hand van een aantal voorbeelden meer grip kunnen krijgen op simulaties.

Monte Carlo-simulaties

Modellen waarbij het toeval een rol speelt, noemen we stochastische simulaties. We hebben gezien dat door een experiment vele malen te herhalen we kunnen proberen lessen te trekken uit de resultaten van deze simulatie. Wanneer we dit principe toepassen op stochastische simulaties, dan spreken we van een Monte Carlo-simulatie. Dit is verwijst naar de badplaats aan de Côte d'Azur, beroemd om haar casino. Het roulettewiel staat dan symbool voor een kansexperiment.

Monte Carlo-simulaties werden voor het eerst ingezet door de wetenschappers die in de Tweede Wereldoorlog werkten aan het Manhattan project om de Verenigde Staten aan de atoombom te helpen. De kernfysici probeerden de baan van nucleaire deeltjes te voorspellen met behulp van simulaties. Lang werd deze techniek van Monte Carlo-simulaties beperkt tot technische toepassingen. Vanaf de jaren tachtig van de vorige eeuw zet men dergelijke simulaties ook in voor financiële berekeningen, vooral in risicomanagement. Kapitaalbuffers zoals economisch kapitaal en *Value-at-Risk* (VaR) worden veelal met deze simulaties berekend.

Simulatie 1: verliezen op debiteuren

Een eenvoudig, maar relevant voorbeeld: een bedrijf levert producten die € 1.000 kosten. De onderneming heeft zijn debiteuren in drie risicogroepen ingedeeld met een laag, gemiddeld en hoog risicoprofiel. Deze profielen corresponderen met uitvalkansen van 2%, 4% en 10%. Deze uitvalkansen kennen we uit ervaring en zijn meestal gebaseerd op historische cijfers of schattingen. Kort gezegd geeft deze uitvalkans of de kans dat afnemers niet aan hun verplichtingen voldoen en niet betalen. De jaarlijkse afzet van dit bedrijf is 300 producten met een gelijke verdeling over de risicoprofielen. Het gemiddelde verlies is in dit geval eenvoudig uit te rekenen:

Jaarlijks verlies = 2% x € 1.000 x 100 + 4% x € 1.000 x 100 + 10% x € 1.000 x 100

Jaarlijks verlies = € 16.000

Het zal duidelijk zijn dat dit bedrag van € 16.000 een gemiddeld verwacht verlies is: in de praktijk kan dit hoger of lager uitvallen. Stel nu dat het bedrijf in zijn budget voor het komende boekjaar een voorziening wil opnemen voor te verwachten debiteurenverliezen. De credit manager is voorzichtig van aard en wil de voorziening zo groot maken dat hij in 95% van de gevallen goed zit. Hij wil met andere woorden een budgetbuffer voorstellen die de kans dat de voorziening tekortschiet, beperkt tot vijf procent. Anders gezegd, eens in de twintig jaar is de buffer niet hoog genoeg, de overige negentien jaren voldoet de buffer. De hamvraag waarmee de credit manager wordt geconfronteerd is: hoe groot moet deze buffer zijn?

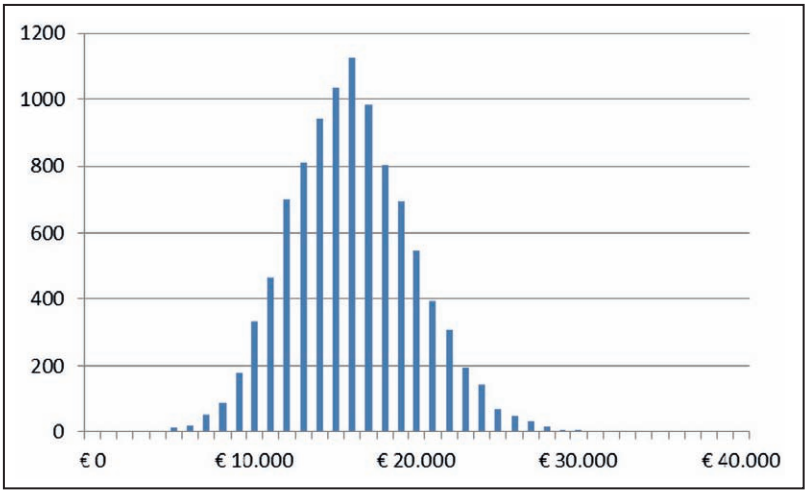
Dit soort vragen kunnen het beste met een simulatie worden uitgerekend, vooral wanneer de modellen ingewikkelder worden. In plaats van dit proberen uit te rekenen, simuleren we dit proces door de input variabele voor de wanbetaling te laten variëren. Voor ieder van de uitstaande vorderingen gooien we als het ware een vals muntstuk op waarbij de kansen niet 50% / 50% verdeeld zijn voor kop of munt maar 2% / 98%, waarbij de 2% staat voor de kans op wanbetaling en 98% voor een correcte betaling. Eenzelfde verdeling kunnen we ook maken voor de overige risicocategorieën.

Stochastisch simuleren: hoe doe je het in Excel?
In Excel kun je de random number generator *gebruiken om stochastische parameters te genereren. Deze gegenereerde parameters kun je vervolgens gebruiken in je berekeningen.*

Welke Excel-formules je gebruikt om stochastische parameters te genereren, hangt af van welke verdeling de parameter moet hebben. Kijk hiervoor in onderstaande tabel. Merk op dat vet en cursief gedrukte woorden en symbolen de parameters van de verdeling zijn.

Verdeling	**Oude Excel (NL)**	**Nieuwe Excel (NL)**
Uniforme	[=min+(max-min)*aselect()]	
Uniform, discreet	[=aselecttussen(min;max)]	
Driehoeks	[=als(aselect()<(mode-min)/(max-min); min+(mode-min)*sqrt(aselect()); max+(mode-max)*sqrt(aselect()))]	
Normaal	[=norm.inv(aselect();μ;σ)]	[=norm.inv.n(aselect();μ;σ)]
Lognormaal	[=log.norm.inv(aselect();μ;σ)]	[=lognorm.inv.n(aselect();μ;σ)]
Bernoulli	[=als(aselect()<p;'ja';'nee')]	
Binomiaal	Niet beschikbaar	[binomiale.inv(N;p;aselect())]
Verdeling	Oude Excel (EN)	Nieuwe Excel (EN)
Uniforme	[=min+(max-min)*rand()]	
Uniform, discreet	[=randbetween(min;max)]	
Driehoeks	[=if(rand()<(mode-min)/(max-min); min+(mode-min)*sqrt(rand()); max+(mode-max)*sqrt(rand()))]	
Normaal	[=norminv(rand();μ;σ)]	[=norm.inv(rand();μ;σ)]
Lognormaal	[=loginv(rand();μ;σ)]	[=lognorm.inv(rand();μ;σ)]
Bernoulli	[=if(rand()<p;'ja';'nee')]	
Binomiaal	Niet beschikbaar	[binom.inv(N;p;rand())]

Elk van de 300 vorderingen kunnen we simuleren met een model van een vals muntstuk met 2% / 98% kansen. Dit doen we door een kansverdeling te gebruiken die de relatie weergeeft tussen de mogelijke uitkomsten van een kansexperiment en de verschillende kansen op deze uitkomsten. In het geval van de afnemers-kredieten gaat het om een simpele ja/nee-verdeling. Immers, de afnemer betaalt of betaalt niet, waarbij we omwille van de eenvoud afzien van deelbetalingen. Een dergelijke ja/nee-verdeling wordt ook wel Bernoulli-verdeling genoemd. De simulatie bestaat uit het opgooien van de 300 valse muntstukken met de eerder omschreven kansverdelingen. We weten nu welke kredieten oninbaar zijn in deze ene proefronde. Doordat het om een stochastisch proces gaat dat van het toeval afhangt, zal de uitkomst bij herhaling van de proef waarschijnlijk verschillen. Steeds zullen andere combinaties van leningen falen. Echter, als we het experiment een groot aantal malen herhalen, wordt een patroon zichtbaar. Tienduizend experimenten leveren ook tienduizend uitkomsten op. De uitkomsten kunnen we weergeven in een frequentietabel, waarbij we op de horizontale as de mogelijke uitkomsten weergeven en op de verticale as de frequentie, of anders gezegd, hoe vaak een bepaalde uitkomst voorkomt in de serie van tienduizend proeven. Een dergelijk frequentieoverzicht wordt ook wel histogram genoemd.

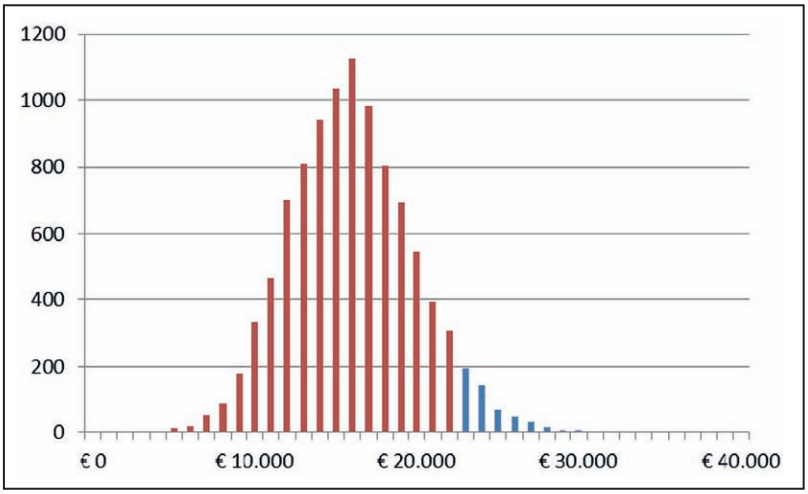


Figuur 26: Histogram bij simulatie 1.

De wet van de grote aantallen stelt dat bij het tienduizendmaal herhalen van de proef, de gemiddelde verliespercentages uitkomen in de buurt van de uitvalkansen die we hadden gedefinieerd. Dus is natuurlijk een zoektocht naar zelf verstopte paaseieren: we vinden wat we erin hadden verstopt. Toch levert deze simulatie nieuwe inzichten op die weliswaar ook met de hand uit te rekenen zijn maar eenvoudiger zijn te verkrijgen met een simulatie. Een van de belangrijkste inzichten is de spreiding van de verliezen. Door naar het histogram van de

uitkomsten te kijken stellen we onmiddellijk vast of er een beperkte spreiding is, waarbij de uitkomsten geclusterd zijn rond het gemiddelde, of dat er juist een grote spreiding is waarbij de uitkomsten zijn uitgesmeerd over een wijd gebied. Deze mate van spreiding is vervat in de standaardafwijking of de variantie en is uiteraard ook uit te rekenen.

Een tweede inzicht is de relatie tussen het budget en risico. Het is duidelijk dat naarmate we het budget groter maken, de kans dat het budget tekortschiet kleiner wordt. Dit brengt ons dicht bij de vraag die de credit manager opwierp: hoeveel budget moet ik reserveren om in 95% van de gevallen goed te zitten? Als we opnieuw een blik op de grafiek werpen, zien we dat deze grafiek tienduizend uitkomsten weergeeft van tienduizend kansexperimenten. De oppervlakte van de grafiek of van alle staafjes van de grafiek omvat 100% van alle mogelijke uitkomsten. Nu is de vraag of we ergens in de grafiek een scheiding kunnen aanbrengen bij de eerder genoemde 95%-knip. Links van deze lijn bevindt zich 95% van het oppervlak van de grafiek, wat ook 95% van de mogelijke uitkomsten representeert, rechts van deze lijn vinden we de 5% uitkomsten die niet worden gedekt door het budget. De scheidingslijn staat gelijk aan het 95^{e} percentiel. Percentielen staan gelijk aan procenten in een geordende dataset. De dataset in onze grafiek is geordend aangezien de laagste uitkomst links staat en de hoogste rechts. In ons geval willen we weten wat de uitkomst ofwel het benodigde budget is bij het 95^{e} percentiel. Dit blijkt € 23.000 te zijn.



Figuur 27: Histogram bij simulatie 1, met in het rood aangegeven het deel van het histogram dat onder het 95^{e} percentiel ligt.

Hoe moeten we dit getal interpreteren? Als we € 23.000 aan budget reserveren, kunnen we in 95% van de gevallen met onze reserveringen de verliezen afdekken. In 5% van de gevallen, dus eens in de twintig jaar, schiet ons budget tekort. Willen we meer zekerheid, dan moeten we het budget en het percentiel verhogen; van de andere kant, als we bereid zijn meer risico te accepteren, kunnen we volstaan met lager percentiel en budget. Ons model verschaft ons inzage in de verhouding tussen risico en budget of risico en geld. Dit is een van de belangrijkste mechanismen in de economie.

Zouden we bijvoorbeeld volstaan met 60% zekerheid voor ons budget, dan hebben we aan € 17.000 genoeg. In onderstaande tabel is een overzicht te vinden van de zekerheidspercentielen en de bijbehorende budgets.

Percentielen bij simulatie 1	
10e percentiel	**€ 11.000**
20e percentiel	**€ 13.000**
30e percentiel	**€ 14.000**
40e percentiel	**€ 15.000**
50e percentiel	**€ 16.000**
60e percentiel	**€ 17.000**
70e percentiel	**€ 18.000**
80e percentiel	**€ 19.000**
90e percentiel	**€ 21.000**

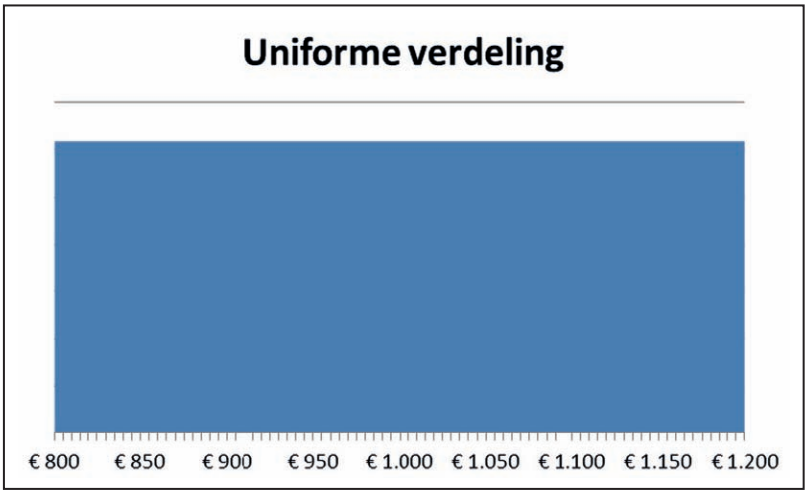
Tabel 10: Percentielen van de uitkomst bij simulatie 1.

Percentielen: hoe doe je het in Excel?
Percentielwaarden kun je eenvoudig uitrekenen met Excel door gebruik te maken van (NL)[=percentiel()] of (EN)[=percentile()], of in de nieuwere versies (NL)[=percentiel.exc()] of (EN)[=percentile.exc()].

Simulatie 2: verliezen op debiteuren met variabele bedragen

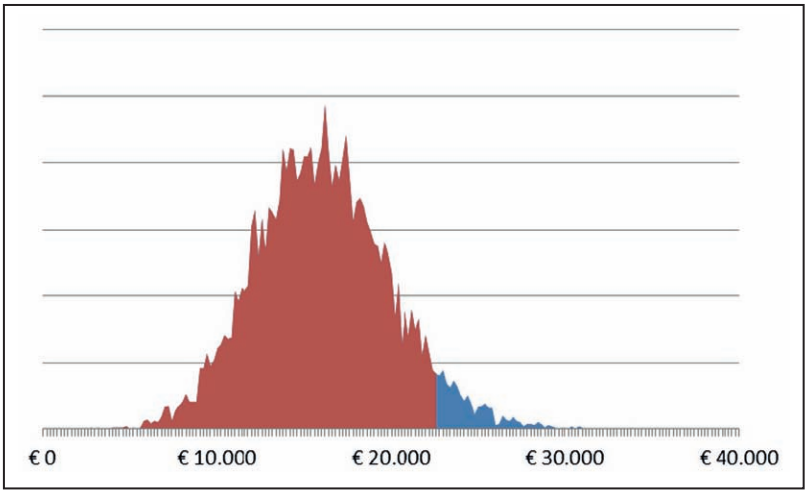
In ons eerste simulatiemodel gingen we ervan uit dat de bedragen van leveringen in alle gevallen duizend euro waren. We kunnen de simulatie verfijnen door te werken met variabele bedragen voor de leveringen. Laten we aannemen dat de waarde van de levering gemiddeld inderdaad duizend euro is maar dat het factuurbedrag kan variëren tussen € 800 en € 1.200. Het gaat hier om een uniforme continue kansverdeling. Als we naar de horizontale as kijken, zien we de waarden die het factuurbedrag kan aannemen. Dit zijn alle bedragen tussen enerzijds het minimum van € 800 en anderzijds het maximum van € 1.200. De

hoogte van de grafiek langs de Y-as is in alle gevallen gelijk omdat de kans voor de verschillende bedragen hetzelfde is, vandaar de term uniform.



Figuur 28: Verdeling van het factuurbedrag.

In onze aangepaste simulatie variëren niet alleen de uitvalkansen maar ook de hoogten van de bedragen. Dus, voor iedere openstaande vordering bepalen we per kansexperiment de of de vordering al dan niet uitvalt en wat de hoogte van de vordering is. Wederom herhalen we deze proef tienduizend maal en verwerken deze in een histogram. Dit keer is de grafiek continu omdat de hoogte van iedere uitstaande vordering verschillend kan zijn.



Figuur 29: Verdeling van de uitkomsten bij simulatie 2. Het rode gebied geeft alle mogelijke uitkomsten onder het 95ste percentiel aan.

Ook nu stellen we de vraag naar de hoogte van het budget waarbij we 95% zeker zullen zijn dat we de debiteurenverliezen kunnen opvangen. Wederom kunnen we dit bedrag vinden door naar het 95ste percentiel te kijken. In dit geval ligt het 95e percentiel op € 22.620.

Van spreiding naar risico

In de modellenbouw worden de spreidingsmaten vaak gebruikt om de volatiliteit van bijvoorbeeld aandelen weer te geven. Volatiliteit is de beweeglijkheid van de koers van een aandeel. Het zal duidelijk zijn dat een hoge volatiliteit meer risico met zich meebrengt dan een lage, omdat er meer onzekerheid is over de opbrengst van het aandeel. Risico zal men vaak proberen te vermijden en reduceren, en risico heeft ook een prijs. Je zou kunnen argumenteren dat als je baas steeds een muntstuk opgooit bij de maandelijkse uitbetalingen van de salarissen en de cheque verdubbelt bij kop en niet uitbetaalt bij munt dit niets uitmaakt. Op de lange termijn zal in de helft gevallen het dubbele worden uitbetaald, en de andere helft niet. Gemiddeld levert dit hetzelfde salaris op. Toch heeft dit niet dezelfde waarde voor de werknemer als het vaste salaris. De onzekerheid heeft een prijs. Hoe hoog die prijs is, is afhankelijk van de appetijt voor risico van de werknemer en van zijn financiële reserves.

Door risico gelijk te stellen aan een gemeten spreidingsmaat kan het misverstand ontstaan dat indien deze waarde laag of zelfs nul is er geen sprake zou zijn van risico. Dit is alleen waar wanneer de data waarmee gerekend is een voldoende compleet beeld van de werkelijkheid geven. Uit de volgende voorbeelden blijkt dat dit niet altijd het geval is:

Voorbeeld 1: brandverzekering

De meeste huiseigenaren hebben een brandverzekering. Zou men maandelijks bijhouden of er brand is geweest, dan zal hopelijk in alle maanden geen brand worden geregistreerd. De variantie en standaarddeviatie van deze waarnemingen zal dan ook nul zijn. Dit wil echter niet zeggen dat het risico op brandschade gelijk aan nul is!

In dit voorbeeld is er geen sprake van een representatieve steekproef: nog niet alle mogelijke situaties zijn aan bod geweest. Alleen bij een representatieve steekproef geeft de gemeten standaarddeviatie een realistische weergave van een risico. Hierbij is het vooral erg belangrijk dat men niet gericht geselecteerd heeft op 'mooi weer' uitkomsten.

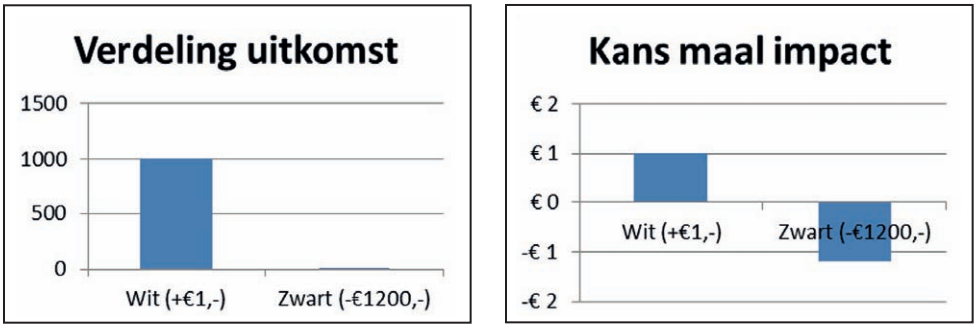
Voorbeeld 2: het knikkerspel

Stel dat een docent een bloempot met duizend witte knikkers en één zwarte knikker op zijn lessenaar plaatst en zijn studenten uitnodigt voor het volgende spel. Ieder van de collegebezoekers trekt bij binnenkomst een knikker en legt deze terug. In het geval van een witte knikker krijgt de student één euro om een blikje frisdrank te kopen. Mocht iemand de zwarte knikker trekken, dan is deze student twaalfhonderd euro verschuldigd aan de docent.

Slimme studenten spelen dit spel natuurlijk niet. Net zoals een loterij een speciale belasting is voor mensen die niet hebben opgelet bij wiskunde, zo is ook dit spel nadelig voor de spelers. Gemiddeld haalt de docent een voordeel, hoewel het lang kan duren voordat dit voordeel kan worden gerealiseerd.

Het kan gebeuren dat na tien colleges met vijftig deelnemers niemand de zwarte knikker heeft getrokken. Alle getrokken knikkers zijn wit en de standaarddeviatie van de observaties is nul. Zou men concluderen dat het spelen van dit spel zonder risico is, dan heeft men het mis! De docent had immers geen geheim gemaakt van de zwarte knikker.

De spelers van het spel kunnen verwachten dat ze ongeveer een kans hebben van één op duizend, om precies te zijn $1/1001 \approx 0{,}099\%$, om € 1.200 te verliezen. Hier staat tegenover een kans van $1000/1001 \approx 99{,}9\%$ om een euro te winnen. Hieruit volgt een gemiddeld (verwacht) verlies van €0,19 per spel, en een standaarddeviatie van €37,94!



Figuur 30: (links) Verdeling van mogelijke uitkomsten bij kansspel. Merk op dat de kans op een zwarte knikker verwaarloosbaar klein lijkt. (rechts) Door de kans op een uitkomst te vermenigvuldigen met de impact van deze uitkomst blijkt dat de kans op een zwarte knikker allesbehalve verwaarloosbaar is.

De les die we uit deze voorbeelden kunnen trekken, is dat het verregaande gevolgen kan hebben om voorspellingen te doen op basis van een te kleine of niet representatieve steekproef. Dit geld des te meer voor extreme risico's, die misschien niet vaak voorkomen, maar wel een grote impact hebben.

Monte Carlo-simulaties: het nut voor de credit manager
Monte Carlo-simulaties kunnen de credit manager inzicht geven in wat de toekomst kan brengen. Hierbij wordt niet alleen een voorspelling voor de toekomst gemaakt, maar ook de onzekerheid en risico die bij deze voorspelling horen worden afgeschat. Dit maakt de Monte Carlo-simulatie tot een krachtig stuk gereedschap in risicoanalyses, budgets en planningen.

We hebben gezien hoe we variabelen kunnen modelleren met kansverdelingen. De kansverdeling wordt gekozen en gedefinieerd op basis van historische data, op basis van inzicht en schattingen of op basis van economische theorie.

Als we het rijtje inzichten afgaan:

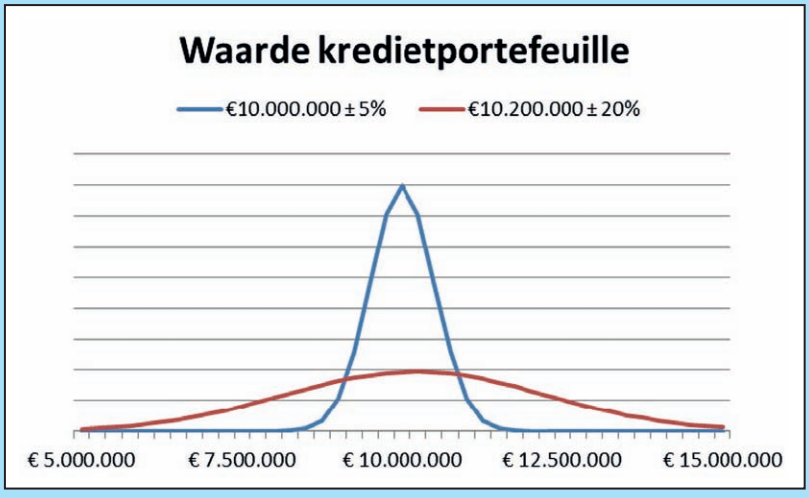
1. *Historische data zijn een goed begin omdat het bestuderen van het gedrag van een variabele in het verleden vaak iets zegt over het te verwachten gedrag voor de toekomst. Echter, we mogen hier geen wetmatigheid van maken. Er is geen garantie dat als een debiteur altijd binnen dertig dagen betaalt hij dit in de toekomst weer zal doen, het is echter wel waarschijnlijk dat hij op tijd gaat betalen. Veel van de kritiek op financiële modellenbouw richt zich op het falen van de modellen bij het voorspellen van de recente financiële crisis. Eén van de oorzaken is de foutieve aanname dat het verleden de toekomst voorspelt: tien jaren van economische voorspoed bieden geen garantie dat ook het elfde jaar rooskleurig verloopt…*
2. *Expertise en schattingen van deskundigen zijn daarom vaak een goed alternatief voor historische analyses. Bedrijven schrikken terug van expertise en schattingen omdat deze benadering te subjectief zou zijn terwijl de historische analyse de schijn geeft van statistische zekerheid. Dit is echter niet terecht. Ook al heeft een debiteur de laatste jaren steeds stipt op tijd betaald, als we weten dat deze klant in moeilijk vaarwater terecht is gekomen dan mogen we ook veronderstellen dat de betaaltermijnen hieronder gaan lijden.*
3. *De economische en wiskundige theorie biedt ook een handvat. Als men alom een lognormaal-verdeling gebruikt om rentestanden te modelleren dan is het als credit manager verstandig hierbij aansluiting te zoeken en niet opnieuw het wiel uit te vinden.*

Maar, terug naar de Monte Carlo-simulatie: waarom zou een credit manager hiermee aan de slag gaan, of anders gezegd, wat voegen Monte Carlo-simulaties toe aan de gebruikelijke modellen in Excel waarmee de meesten van ons tot dusver werken?

De voordelen van Monte Carlo-simulaties zijn:

1. *Soms zijn de uitkomsten van een Monte Carlo-simulatie anders dan men op basis van de oorspronkelijke aannamen had verwacht. Bedrijven gebruiken vaak een worst-case, best-case en een middle-of-the-road scenario om iets te zeggen over de toekomst. Deze benadering heeft als probleem dat geen van deze drie cases erg realistisch is. Het is onwaarschijnlijk dat alle variabelen tegelijkertijd de meest negatieve waarde of, aan de andere kant van het spectrum, de meest positieve waarde zouden aannemen. De kans dat worst-case-low en best-case-high zich voordoen, is in de meeste modellen vrijwel nihil. De waarde van deze scenario's is daarmee beperkt. Ook het in Excel berekende middle-of-the-road scenario hoeft niet altijd realistisch te zijn. In een Monte Carlo-simulatie kan men correlaties tussen variabelen en lange staarten meenemen. Hierdoor kan het resultaat van een simulatie afwijken van conventionele niet-stochastische modellen. De resultaten van Monte Carlo-simulaties zijn doorgaans betrouwbaarder.*

2. *Het tweede voordeel is dat de Monte Carlo-simulatie ook zicht geeft op het risico verbonden met de einduitkomst. Dit risico vindt men terug in de standaarddeviatie. We weten nu of er een grote of kleine spreiding is van mogelijke uitkomsten. Stel we bepalen de waarde van twee kredietportefeuilles: de eerste heeft een waarde van tien miljoen euro met een standaarddeviatie van 5%, de tweede heeft een waarde van*



Figuur 31: Waarde van twee fictieve kredietportefeuilles.

10,2 miljoen euro maar met een standaardafwijking van 20%. Welke portefeuille is nu meer waard? De keuze zal afhangen van de risicovoorkeuren van de beslisser maar het is duidelijk dat de hogere netto waarde niet automatisch de beste keuze is.

Monte Carlo-simulaties geven dus een beter inzicht in de risico's en helpen te komen tot betere keuzes en beslissingen.

3. *Hierbij sluit aan dat we inzicht krijgen in de relatie tussen risico en zekerheid aan de ene kant en de uitkomsten van ons model aan de andere kant. Hierbij kunnen we antwoord geven op vragen als: hoeveel geld verwacht ik met 80% zekerheid te innen op deze kredietportefeuille? Of hoeveel moet ik budgetteren zodat ik de te verwachten debiteurenverliezen in 95% van de gevallen kan afdekken?*

Kortom, Monte Carlo-simulaties geven waardevolle inzichten in planningen en budgets en helpen credit managers om te gaan met onzekerheid.

Afronding

Hiermee zijn we aan het einde gekomen van onze korte verkenning van de statistiek en kansrekening. De credit manager weet nu wat de belangrijkste grondbeginselen zijn, hoe hij deze in eenvoudige Excel-modellen kan toepassen en vooral hoe hij deze principes kan gebruiken in het nemen van betere beslissingen in de alledaagse werkpraktijk. Voor veel bedrijven en organisaties is er een wereld te winnen met modelleren. Te vaak worden beslissingen genomen op basis van onderbuikgevoel en valse percepties. Modelleren dwingt de credit manager na te denken over de problematiek en de zaken rationeel op een rijtje te zetten. We hebben gezien dat modellen maar een beperkte waarheid leveren, per definitie niet in staat zijn de complexe werkelijkheid perfect te vatten en soms zelfs de plank volledig misslaan. Dit is waar. Echter, het is ook waar dat nadenken over risico's, het onderzoeken van patronen in data en het schatten van kansen en risico's, leiden tot een beter begrip van de zakelijke problemen op het pad van de credit manager. Daarom is het zinvol te modelleren. Dit leidt zeker tot betere bedrijfsbeslissingen in het credit management. Statistiek, kansrekening, Excel, gezond verstand en een kop koffie zijn de materialen waarmee we werken. Aan de slag!

André Koch
Amsterdam, oktober 2014
Stachanov Solutions & Services BV
andre@stachanov.com

Definities van gebruikte termen

Bernoulli-verdeling
Een kansverdeling met twee mogelijke uitkomsten.

Categorische variabele
Zie discrete variabele

Continue variabele
Een variabele die tot op een oneindig detailniveau alle mogelijke waarden aan kan nemen. Tegenhanger van de discrete variabele.

Correlatie
Een maat voor de samenhang tussen twee variabelen.

Covariantie
Een maat voor de samenhang tussen twee variabelen.

Deterministisch proces
Een proces waarbij de uitkomst vast ligt.

Deterministische simulatie
Een simulatie waarbij geen rekening wordt gehouden met het toeval.

Discrete variabele
Een variabele die en beperkt aantal waarden binnen een bandbreedte aan kan nemen. Tegenhanger van de continue variabele.

Driehoeksverdeling
Een kansverdeling met een karakteristieke driehoekige vorm.

Gemiddelde
Een centrummaat in de statistiek.

Histogram
Een weergave van een gemeten verdeling.

Ja/nee-verdeling
Zie Bernoulli verdeling.

Kansexperiment
Zie stochastisch proces.

Kansrekening
De tak van de wiskunde die zich bezig houd met het berekenen van kansen en mogelijkheden.

Kansvariabele
Zie stochast.

Kansverdeling
Geeft de kans waarschijnlijkheid aan dat verschillende uitkomsten verkregen worden in een stochastisch proces.

Kolommendiagram
Zie histogram.

Kwalitatieve data
Data die in categorieën ingedeeld kan worden.

Kwantitatieve data
Data die in getallen weergegeven kan worden.

Kurtosis
Een vormmaat in de statistiek.

Lognormaal-verdeling
Een asymmetrische kansverdeling die alleen positieve uitkomsten geeft.

Mediaan
Een centrummaat in de statistiek.

Modus
Een centrummaat in de statistiek.

Moment
Andere naamgeving voor: gemiddelde, variantie, scheefheid en kurtosis.

Monte Carlo-simulatie
Een simulatietechniek voor het in kaar brengen van de invloed van het toeval.

Normaalverdeling
Een kansverdeling met een karakteristieke klokvorm. Komt vaak voor in de natuur en bij steekproeven.

Poisson-verdeling
Een discrete kansverdeling die gebruikt wordt voor telbare aantallen.

Scheefheid
Een vormmaat in de statistiek.

Simulatie
Een nabootsing van de werkelijkheid.

Staartafhankelijkheid
Het verschijnsel dat twee variabelen die onder normale omstandigheden onafhankelijk zijn correlatie vertonen in de extreme waarden, in de 'staart' van de verdeling.

Standaarddeviatie
Een spreidingsmaat in de statistiek.

Statistiek
De wetenschap van het verzamelen, verwerken en weergeven van grote hoeveelheden gegevens.

Stochast
Een toevalsvariabele. De uitkomst van een stochastisch proces.

Stochastisch proces
Een proces waarbij toeval een rol speelt.

Stochastische simulatie
Een simulatie waarbij het toeval gesimuleerd wordt.

Toevalsvariabele
Zie stochast.

Uniforme verdeling
Een kansverdeling waarbij alle mogelijke uitkomsten even waarschijnlijk zijn.

Variabele
Een grootheid die in waarde kan variëren.

Variantie
Een spreidingsmaat in de statistiek.

Verdeling
De relatieve frequentie waarmee een variabele verschillende waarden aanneemt.

Verwachtingswaarde
De verwachte uitkomst van een stochastisch proces. Is gelijk aan het gemiddelde van de bijbehorende kansverdeling.

Waarschijnlijkheidsleer
Zie kansrekening.

Risico en onzekerheid zijn vaste bestanddelen van het dagmenu van de credit manager. Grip op onzekerheid en risico vertaalt zich in lagere afschrijvingen op debiteuren, minder kapitaalsbeslag en betere financiële planning. Statistiek en kansrekening zijn voor de credit manager de professionele gereedschappen om zijn vak succesvol uit te oefenen. Als onderdeel van de Praktijkreeks voor Credit Management biedt dit handboek een beknopte en duidelijke inleiding in de statistiek en kansrekening toegespitst op het werkterrein van de credit manager. Duidelijke voorbeelden uit de praktijk wijzen de weg en beknopte Excel-instructies helpen de credit manager zelf aan de slag te gaan.

Over de auteur

André Koch is partner bij Stachanov Solutions & Services bv, een Amsterdams IT-bedrijf dat zich specialiseert in de bouw van databanken en softwaresimulaties. Als consultant heeft hij op het gebied van risicomodellering gewerkt voor tal van internationale opdrachtgevers in meer dan vijfentwintig landen. Daarnaast doceert hij als senior visiting lecturer aan Nyenrode, The Netherlands Business University. Hij schrijft regelmatig voor magazine 'De Credit Manager' van de Vereniging voor Credit Management (VVCM), hét onafhankelijke vakblad voor credit managers en debiteurenbeheerders in Nederland.

9 789082 305104 >