Kwantitatieve methoden

voor zorglogistiek

Auteurs:
Prof. Dr. G.M. Koole
Dr. R. Bekker

Inhoud

1 Inleiding

Casus

In een ziekenhuis is er sprake van een te hoog percentage afzeggingen van open-hartoperaties (OHO's). In de discussie hierover worden verschillende oorzaken geopperd die samenhangen met de beschikbare capaciteit en werkwijze van de operatiekamers, maar ook met capaciteitsproblemen in de zorgeenheden die de patiënten doorlopen. Het gesprek leidt niet tot een eenduidige conclusie. Afgesproken wordt dat er eerst uitgezocht moet worden waardoor en in welke mate operaties afgezegd worden. Daarna gaat men proberen uit te zoeken wat de consequenties zouden zijn, zowel op de prestaties van het ziekenhuis als financieel, van verschillende mogelijke oplossingen zoals het verruimen van de capaciteit of een andere wijze van werken.

Een thuiszorgorganisatie worstelt met het probleem dat ze te weinig cliënten hebben om zorg op afroep te leveren. De zorg op afroep voor niet planbare zorg is 24 uur per dag beschikbaar. Overdag gebeurt dit door medewerkers die voor reguliere activiteiten ingeroosterd zijn, maar bij een oproep direct weg kunnen. 's Avonds en 's nachts vindt invulling plaats door specifiek voor deze oproepen vrijgestelde medewerkers. Met name in de nacht lijken er onvoldoende oproepen te zijn om de dienst rendabel te houden. Wat is het minimaal aantal cliënten nodig om de dienst rendabel te houden. Afgesproken wordt dat er voor de avond en de nacht uitgezocht moet worden wat de aard en de omvang van de oproepen precies is.

Kader 1.1 Casus 'Meten'

"Meten is weten". Dit deel gaat over het hoe, wat en waarom van het meten. En ook over hoe, in geval we niet kunnen meten, we kunnen voorspellen wat de prestaties van een systeem zouden zijn. Het is namelijk niet altijd mogelijk te meten: voor meten heeft men een bestaand systeem nodig, waar ook nog gegevens over beschikbaar zijn. Voorspellen voorkomt dat we iets uitproberen en dat later, na implementatie, blijkt dat de uitkomsten niet naar wens zijn. We richten ons dus eerst op het meten van relevante parameters in medische processen, waarna we verder gaan met het voorspellen van diezelfde parameters in (nog) niet bestaande systemen.

1.1 Variabiliteit

Er zijn weinig systemen, in de zorg en daarbuiten, die geen variabiliteit vertonen. De zorgduur of opnameduur van een patiënt of cliënt, het aantal aankomsten op de spoedeisende hulp, het aantal beschikbare verzorgenden of verpleegkundigen: het verschilt van dag tot dag, van patiënt/cliënt tot patiënt/cliënt, enz. Dit brengt met zich mee dat we in onze metingen rekening moeten houden met deze fluctuaties. Het betekent niet dat we niet kunnen meten, maar dat we de verschillen in metingen op een adequate manier moeten interpreteren. Meten, voorspellen en sturen, kortweg de onderwerpen van dit deel, staan dus geheel in het teken van het interpreteren van en omgaan met variabiliteit.

Opdracht 1.1

Kies een vijftal soorten van capaciteit, van soorten zorgvraag (kies zowel kortdurende als langdurende zorgvragen) en zorgduren. Bepaal voor elk of er sprake is van fluctuaties en op welke *tijdschaal* deze fluctuaties zich afspelen, d.w.z. zien we de fluctuaties op het niveau van minuten, uren, dagen, maanden, of misschien zelfs jaren? Bespreek de antwoorden met enkele medestudenten.

1.2 Verkrijgen van data

In de laatste decennia heeft ICT een enorme vlucht genomen en is er nauwelijks een proces denkbaar waarvan, al of niet geautomatiseerd, geen gegevens worden opgeslagen. Deze gegevens worden voor verschillende doeleinden gebruikt: denk aan registratie van verrichtingen ten behoeve van facturering en het digitaal opslaan van röntgenfoto's voor medische analyse. Deze informatiesystemen zijn in de regel niet ontworpen voor het analyseren en verbeteren van de logistieke processen. Om deze reden hebben data vaak, logistiek gezien, tekortkomingen. De belangrijkste soorten problemen die we vaak tegenkomen zijn:

- Er zijn geen historische gegevens voorhanden, er zijn bijvoorbeeld alleen gegevens over huidige patiënten/cliënten;
- Er is sprake van een veelheid aan systemen, en data uit alle systemen gecombineerd is nodig;
- Een deel van de data (vaak data die niet noodzakelijk zijn voor het primaire doel van het systeem en die handmatig wordt ingevoerd) is onnauwkeurig, bijvoorbeeld tijdstippen waarop een handeling wordt afgerond omdat de registratie later plaatsvindt;
- De voor de logistiek wenselijke gegevens worden niet bijgehouden, bijvoorbeeld aankomstmomenten van patiënten die naar een ander ziekenhuis zijn doorverwezen of welke zorgactiviteiten er in een verzorgingshuis daadwerkelijk verleend worden.

Voor het analyseren van data voor management en planning zijn er informatiesystemen ontworpen waarin alle data bijeen wordt gebracht en waarin het analyseren gefaciliteerd wordt. Een voorbeeld van een dergelijk systeem is IBM Cognos. Dit ondervangt uiteraard niet dat data onbetrouwbaar of simpelweg afwezig zijn, maar wel de andere nadelen van reguliere financiële en patiëntinformatiesystemen.

Opdracht 1.2

Wordt er in uw organisatie bijgehouden:

Wanneer een patiënt wordt doorverwezen?	JA / NEE
Hoelang een cliënt in zorg is?	JA / NEE
Wanneer een patiënt op de "verkeerde" afdeling terechtkomt?	JA / NEE
Heeft u gegevens over de wachttijden van alle patiënten/cliënten?	JA / NEE
Worden de wachttijden wellicht steekproefsgewijs bepaald?	JA / NEE
Is er ook historische data beschikbaar?	JA / NEE

Op welke wijze verkrijgt u data en van welk systeem?

Bespreek uw antwoorden met enkele medestudenten.

1.3 Meten

Voor een algemene beeldvorming is het niet altijd noodzakelijk gedetailleerde informatie te hebben. Ook bij het vergelijken van verschillende alternatieven kan de relatieve prestatie een belangrijk inzicht geven. Toch geeft het analyseren van data over het algemeen wel een essentieel inzicht in het huidige verloop en de prestatie van processen. Voor een meer gedegen analyse zijn data daarom meestal gewenst. Wanneer ICT niet de gewenste gegevens kan leveren, dan kunnen in sommige gevallen de relevante gegevens uit papieren dossiers worden gehaald. Anders kan worden overgegaan om metingen te verrichten. Men dient zich wel te realiseren dat dit een tijdrovend proces is; het is daarom raadzaam eerst na te gaan in hoeverre de huidige informatievoorziening binnen de instelling kan worden ingezet.

Het uitvoeren van metingen is nauw verwant met de onderwerpen informatiemanagement en projectmanagement uit blok 4. We benoemen hier slechts de punten die specifiek betrekking hebben op het meten. Het succesvol uitvoeren van een meting start met een goede voorbereiding en een eenduidige vraagstelling. Bij het starten van een meting dient men zich de volgende vragen te stellen:

- Wat is het doel van de meting?
- Welke gegevens zijn nodig voor de analyse?
- Wie voert welke gegevens waar in?
- Hoe kan worden gewaarborgd dat de gegevens worden ingevuld en betrouwbaar zijn?
- Wat is de duur van de meting en/of hoeveel gegevens zijn nodig?

Op de laatste vraag wordt later in dit deel ingegaan. De derde en vierde vraag zijn vaak nauw met elkaar verwant. Het inzetten van medewerkers die uitsluitend gegevens verzamelen levert volledige en betrouwbare data op, maar dit is niet altijd mogelijk of gewenst. Dit is bijvoorbeeld het geval wanneer men het proces van iedere patiënt op de SEH (spoed eisende hulp) of in zorg bij de thuiszorg wil registreren. In dat geval dienen gegevens naast de overige werkzaamheden geregistreerd te worden. Om dit succesvol te laten verlopen, moet de extra benodigde inspanning van medewerkers zo klein mogelijk worden gehouden. Vaak kan de registratie gecombineerd worden met andere momenten waarop cliëntgegevens worden ingevoerd. Gegevens verkregen door het creëren van extra registratiemomenten zijn over het algemeen niet volledig en onbetrouwbaar. Verder dient het doel van de meting voor de medewerkers duidelijk te zijn en moet de meting veelvuldig onder de aandacht worden gebracht. Voor een succesvolle meting zijn de principes van projectmanagement van toepassing, zie blok 4.

1.4 Vraag en aanbod

Capaciteitsmanagement gaat over het afstemmen van vraag en aanbod, het zodanig beïnvloeden van vraag en/of aanbod dat de aanwezige capaciteit optimaal benut wordt. Zonder fluctuaties in vraag en/of aanbod hoeft men slechts eenmalig af te stemmen en kan men 100% gebruik maken van de aanwezige capaciteit. Dit is echter een puur hypothetische situatie, in de praktijk heeft men eigenlijk altijd te maken met fluctuaties, zowel in vraag als in aanbod. Zo varieert de vraag op een SEH van dag tot dag en van uur tot uur, en ook de behandelcapaciteit kan enorm fluctueren door bijvoorbeeld ziekte van personeel of beschikbaarheid van medische apparatuur.

Vanuit het oogpunt van de patiënt is het wenselijk om de capaciteit af te stemmen op de vraag. Echter, vaak ligt de capaciteit grotendeels vast en is het niet flexibel genoeg om fluctuaties in vraag te volgen. Soms leidt dit tot overcapaciteit, vaak zijn we verplicht de vraag te beïnvloeden. Dit kan bijvoorbeeld door vraag te weigeren of door behandelingen of zorg uit te stellen tot een later tijdstip dan door de patiënt/cliënt gewenst. Zowel weigeren als uitstellen van zorg kan leiden tot patiëntontevredenheid en/of –onveiligheid.

Een belangrijk onderscheid is of men op het moment van afstemming bekend is met vraag en aanbod; d.w.z. zijn vraag en aanbod op het moment dat men plant bekend of (nog) onzeker? De managementdenker Galbraith definieert onzekerheid dan ook als volgt: "Onzekerheid is informatie die je nog niet hebt maar die wel nodig is" (Galbraith (1973)). Indien vraag en capaciteit onzeker zijn kan men dus vraag en aanbod niet afstemmen. Maar zelfs al is de vraag bekend, dan nog is het mogelijk dat de inzet van capaciteit niet voldoende flexibel is om vraag en aanbod af te stemmen. Het is dus belangrijk dat men de hoeveelheid flexibiliteit afstemt op de mate van onzekerheid. Vaak neemt zowel de onzekerheid als de mate van flexibiliteit in de loop van de tijd af.

Het alternatief voor flexibiliteit is ofwel overcapaciteit ofwel weigeringen/uitstel. De concepten kunnen goed worden geïllustreerd aan de hand van traumapatiënten die mogelijkerwijs 's nachts aankomen of de zorg op afroep. Bij het OK team of thuiszorgteam dat 's nachts standby is, is er sprake van overcapaciteit. De specialist of oproepkracht die indien noodzakelijk wordt opgeroepen van thuis is een goed voorbeeld van flexibele inzet. Indien men wacht met een operatie tot de volgende dag is er sprake van uitstel. Indien er toevallig meerdere patiënten zijn kan men besluiten een patiënt door te sturen naar een ander traumacentrum. In dat geval is er sprake van een weigering. Uitstel of weigering zal in de thuiszorg slechts in beperkte mate voorkomen, gezien de aard van de vraag (bijvoorbeeld hulp bij de toiletgang, of een valincident).

Als vraag en aanbod geen fluctuaties vertonen is het eenvoudig uit te rekenen wat de vereiste capaciteit is. In het geval van fluctuaties zijn de berekeningen veel lastiger en vereisen vaak gespecialiseerde wiskundige kennis. Voor verschillende situaties zijn er zogenaamde wiskundige *modellen* beschikbaar. Deze helpen ons te voorspellen wat de prestaties zijn van (nog) niet bestaande systemen.

Opdracht 1.3

Bepaal, voor de vijf soorten vraag uit de opdracht van paragraaf 1.1 op welke wijze er wordt ingespeeld op de fluctuaties in vraag. Denk daarbij zowel aan het mogelijk uitstellen of weigeren van het leveren van de zorg en aan de wijze waarop de capaciteit inspeelt op de vraag.

Samenvattend schema 1.1

Begrip	**Betekenis**
Data analyse en wiskundige modellen	Hulpmiddel voor afstemmen vraag en aanbod
Fluctuaties	Spelen cruciale rol in vraag-aanbod afstemming Mate van fluctuatie is soms bekend op moment van afstemming, soms echter nog onzeker
Effectieve afstemming	Mogelijk indien: - Onzekerheid beperkt is - Voldoende flexibiliteit is in capaciteitsinzet

2 Data analyse

Casus

Bij het inplannen van ingrepen bij de OK voert de chirurg de verwachte tijdsduur van de operatie in. Vanwege onvoorziene fluctuaties in operatieduur zal de gerealiseerde duur meestal afwijken van de geplande duur. Maar is er sprake van een systematische afwijking? Dit zou het veel voorkomen van uitloop op de OK's (deels) kunnen verklaren. Om te onderzoeken of er sprake is van een dergelijk systematische afwijking besluit men een analyse te doen van historische data.

Bij het gebruik van routes als planningstool voor de organisatie van de zorg van een zorggroep voor VVT-zorg, lijken een paar routes meer dan de geplande tijd in beslag te nemen. De vraag is of hier werkelijk sprake is van een structurele afwijking. Om te onderzoeken of dit inderdaad het geval is besluit men een analyse te doen van de verzamelde data met betrekking tot de betreffende cliënten over de afgelopen maand.

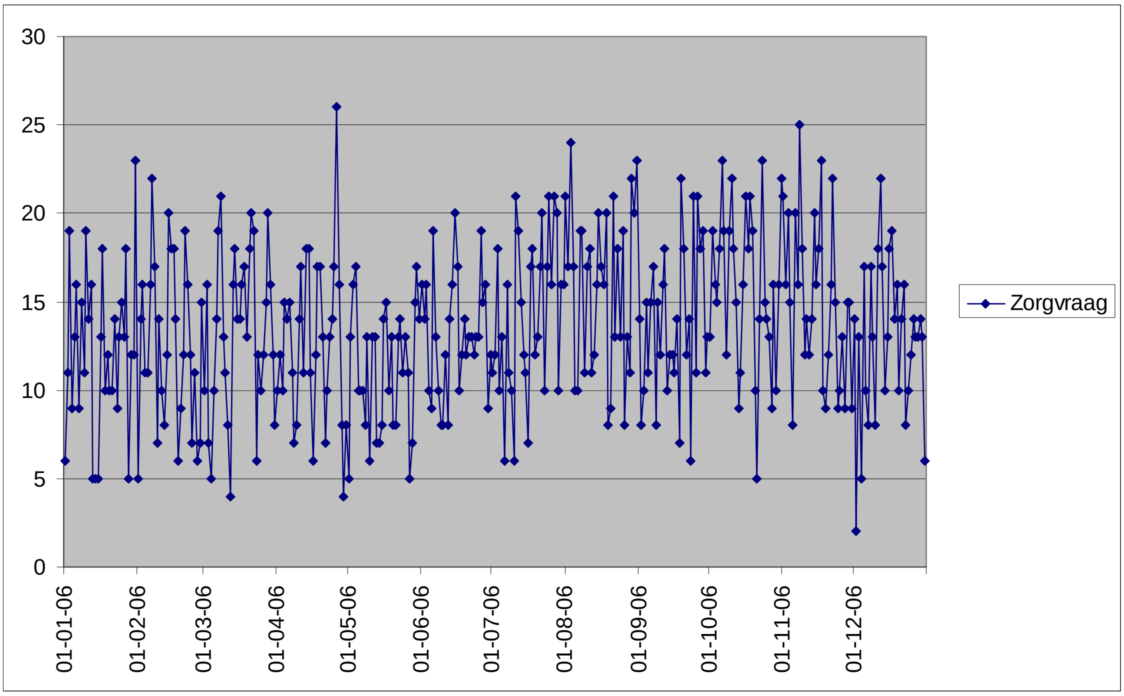
Kader 2.1 Data analyse

Data treffen we aan in een grote verscheidenheid. Een belangrijk verschil is of er sprake is van univariate of multivariate data: is er per patiënt 1 getal (bijvoorbeeld zorgduur) of zijn er meerdere getallen (bijvoorbeeld zorgduur, leeftijd, duur operatie). Een bijzonder geval is wanneer één van de variabelen de tijd betreft en men geïnteresseerd is hoe de andere variabelen zich in de loop van de tijd ontwikkelen. We spreken dan van tijdreeksen. Een voorbeeld is het aantal preparaten dat voor onderzoek van dag tot dag bij de afdeling Pathologie binnenkomt. Een analyse van deze data kan belangrijk zijn voor een analyse van doorlooptijden van pathologisch onderzoek. Een thuiszorgorganisatie zou geïnteresseerd kunnen zijn in hoe de vraag er gedurende een gemiddelde week uitziet, om op een verantwoorde manier keuzes te maken ten aanzien van de personele inzet.

2.1 Grafieken

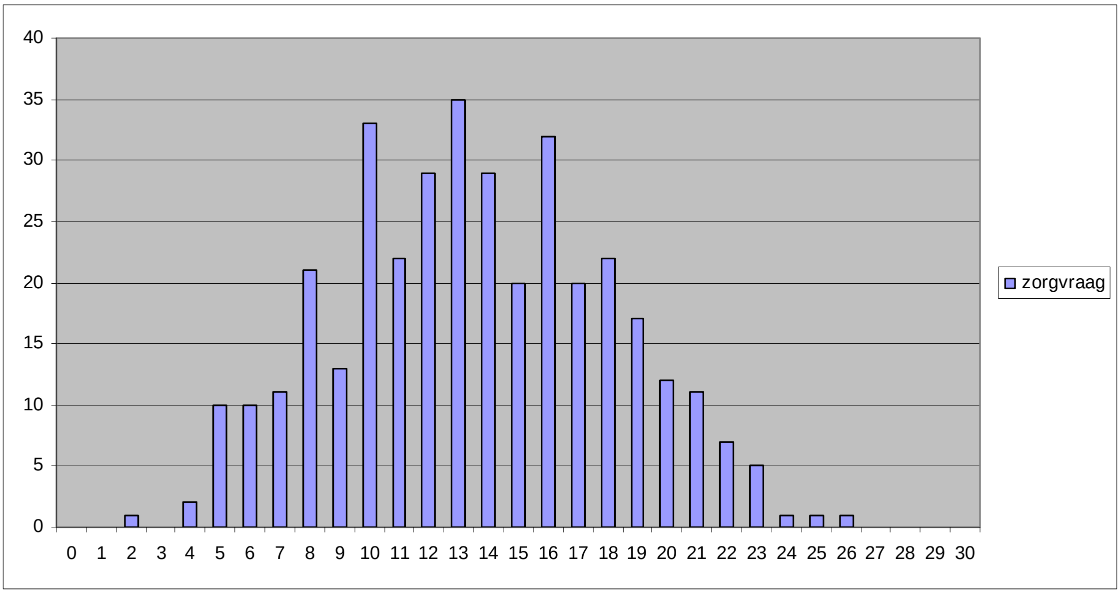
Men start data-analyse in de regel met het maken van enkele grafieken, al was het alleen al maar om gevoel te krijgen voor de data. Bij tijdsreeksen maakt men natuurlijk een plot met de tijd op de horizontale as. Een illustratie is te zien in de grafiek hieronder.[1]

[1] Alle gebruikte data zijn vanuit privacyoverwegingen hetzij fictief (maar representatief voor de zorg) hetzij zodanig gemanipuleerd dat ze niet meer herleidbaar zijn tot de oorspronkelijke instelling.



Figuur 2.1 Zorgvraag bij een zorgeenheid gedurende een jaar

Bij gewone univariate data is het gebruikelijk een *histogram* te maken. Bij een histogram staan de mogelijke waardes op de horizontale as en de frequenties op de verticale. De hoogste balk bevindt zich dus bij de meest voorkomende waarde. Een histogram geeft snel een indruk van plaatsing en spreiding van de getallen in een verzameling data.



Figuur 2.2 Histogram van de zorgvraag gedurende 2006

2.2 Univariate data

Vaak is men echter geïnteresseerd in één of enkele getallen die de data karakteriseren. Hierbij geeft meestal één getal het "midden" van de getallen aan en een andere is een maat voor de spreiding. Het meest gebruikt is het gemiddelde van een verzameling data. Een nadeel van het gemiddelde is dat het gevoelig voor is extreme waardes ("uitbijters"): 1 of enkele uitbijters kan een grote invloed hebben op het gemiddelde. In sommige gevallen is het essentieel de uitbijters in de analyse mee te nemen. De "lange liggers of verblijvers" op een zorgafdeling nemen een groot deel van de capaciteit in beslag. Niet het gemiddelde nemen maar de uitbijters weglaten in een analyse naar capaciteitsgebruik zou verkeerde resultaten opleveren. In sommige gevallen is dit echter wel wenselijk of zelfs noodzakelijk omdat de uitbijters verkeerd geregistreerde gegevens betreffen. Een voorbeeld is het aantal aankomsten bij een afdeling Pathologie of het aantal contactmomenten met bewoners in een verzorgingshuis: men is geïnteresseerd in een "gemiddelde" dag en niet in de uitzonderingen die vaak ook een speciale (en vooraf bekende) oorzaak hebben. Een methode is de laagste en hoogste *n* waardes weglaten, het zogenaamd *getrimd* gemiddelde. Een andere methode is de *mediaan* nemen, de middelste waarde van de data. Een laatste methode, die minder wordt gebruikt, is het nemen van de *modus*: de meest voorkomende waarde.

In Excel kunnen met behulp van ingebouwde functies snel deze getallen berekend worden. De macros heten in het Engels AVERAGE, MEDIAN, TRIMMEAN, en MODUS. In de samenvatting van deze paragraaf worden ook de Nederlandstalige namen genoemd. Voor bovenstaande dataset van Figuur 2.37 geldt dat het gemiddelde gelijk is aan 13.56, mediaan en modus zijn gelijk aan 13, het getrimd gemiddelde is afhankelijk van het aantal getallen dat we weglaten maar verschuift van gemiddelde naar mediaan naarmate we meer getallen weglaten.

De meest bekende maat voor de spreiding is de *standaard deviatie*. Het berekenen van de standaard deviatie vereist wat wiskundige kennis (of het gebruik van een hulpmiddel als Excel): het is de wortel van het gemiddelde van de kwadraten van de afwijkingen van het gemiddelde. Indien men niet de wortel neemt spreekt men van de *variantie*. Een andere mogelijkheid, die conceptueel wat eenvoudiger is, is het meten van de gemiddelde absolute afwijking van het gemiddelde. In onderstaande tabel zijn deze concepten aan de hand van enkele getallen toegelicht. De gemiddelde kwadratische afwijking is de variantie. De wortel daarvan, 2, is de standaarddeviatie.

	waardes	afwijking t.o.v. gemiddelde	absolute afwijking	Kwadratische afwijking
	1	-3	3	9
	3	-1	1	1
	5	1	1	1
	7	3	3	9
	4	0	0	0
gemiddelde	4	0	1.6	4

Figuur 2.3 Rekenvoorbeeld voor gemiddelde, gemiddelde absolute afwijking en variantie

Een geheel andere methode die zeer aansprekende resultaten oplevert, is het nemen van *percentielen*. Vaak wordt het 10 en 90% percentiel genomen, gedefinieerd als de punten waar 10 respectievelijk 90% een lagere waarde heeft. In feite is de mediaan het 50% percentiel. Bij de eerder gebruikte dataset ligt het 10% percentiel op 8 en het 90% percentiel op 20, dwz 10% van de data is kleiner of gelijk aan 8, 90% groter of gelijk aan 20.
Soms worden van een verzameling data ook minimum en maximum berekend. Dit is echter onverstandig: hoe meer getallen, hoe aannemelijker het is dat de minima en maxima ver uit

elkaar liggen. In de regel zoeken we naar grootheden die niet te zeer door de hoeveelheid data worden beïnvloed. Minimum en maximum komen dan niet in aanmerking.

Opdracht 2.1

In de file zorgduren.xls[2] zijn 2 tabbladen met zorgduren van verschillende zorgeenheden.

a. Maak histogrammen van beide dataverzamelingen.
b. Bereken gemiddelde, standaard deviatie, enz. van beide dataverzamelingen.

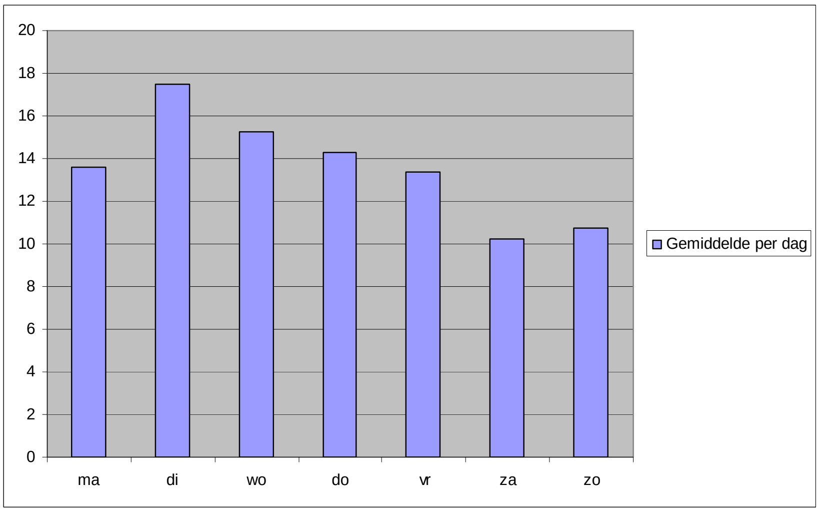
De volgende opgaven zijn ter verdieping en vereisen meer inzicht en meer kennis van Excel.

c. Sorteer de patiënten/cliënten op oplopende zorgduur en bepaal van elke patiënt/cliënt de fractie van de totale zorgduur die hij/zij heeft gebruikt.
d. Bepaal de partiële sommen, d.w.z., bepaal voor elke patiënt/cliënt welke fractie van de totale zorgvraag patiënten/cliënten met een lagere of gelijke zorgduur hebben gebruikt.
e. Gebruik deze gegevens om een grafiek te maken met horizontaal de fractie van het aantal patiënten en verticaal de door hen gebruikte fractie van de totale zorgduur.

Een dergelijke grafiek heet een *Lorenz-curve*. Hoe meer spreiding, hoe "dieper" de curve. Dit soort grafieken wordt veel in de economie gebruikt om de verdeling van middelen in een land weer te geven.
Verder lezen: http://nl.wikipedia.org/wiki/Lorenz-curve.

2.3 Tijdreeksen

Tijdsreeksen vereisen een aparte analyse. Bij tijdsreeksen waarbij menselijk handelen een rol speelt, zoals zorgprocessen, is er vaak sprake van seizoensinvloeden en cycli. Zo zijn er in veel zorginstellingen minder opnames in het weekend, is het aantal behandelingen lager in de zomer, is het aantal spoedopnames 's nachts lager dan overdag, enzovoorts. Het beschrijven van deze fluctuaties is belangrijk, met name als men later in staat wil zijn de toekomst te voorspellen. In de eerder gebruikte data blijkt er een duidelijke wekelijkse cyclus te zijn. We zien dat dinsdag de drukste dag is, zaterdag de rustigste.

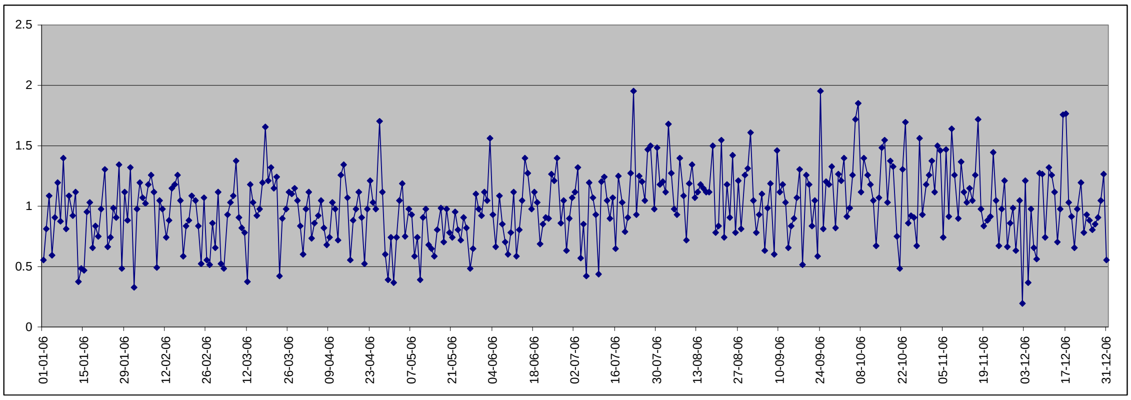


Figuur 2.4 Gemiddelde zorgvraag per dag

[2] De documenten voor de opgaven zijn (voorlopig) te vinden op http://www.math.vu.nl/~koole/LogiZ

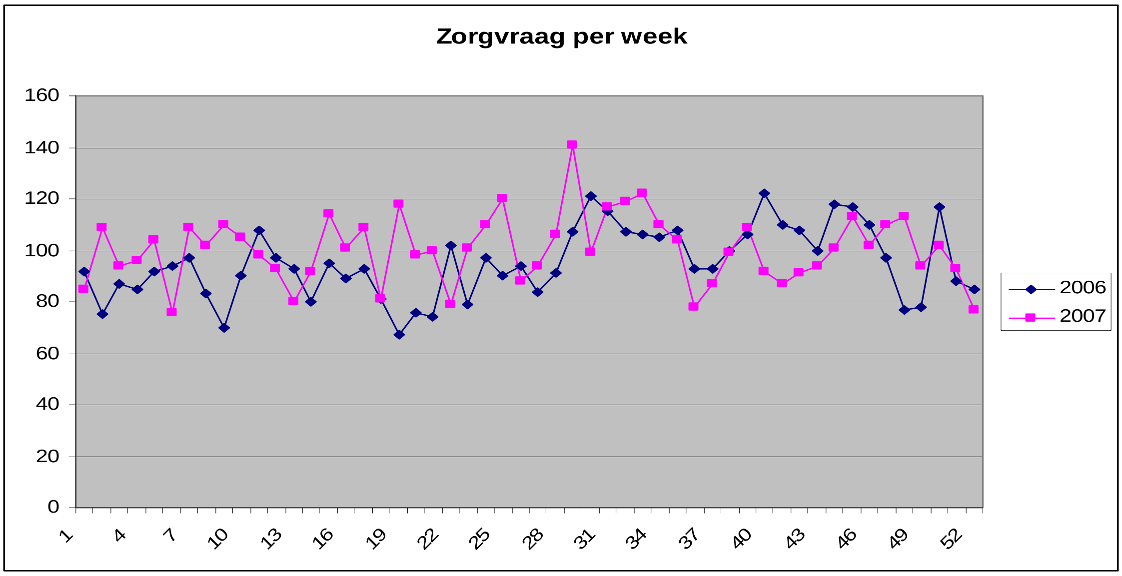
Soms is het interessant om nog dieper te gaan dan het dagniveau, bijvoorbeeld bij de aankomsten van een SEH, of bij de 24-uurs zorg op afroep waar het dag/nachtritme een belangrijke rol speelt.

Aan het hand van het weekritme kunnen we nu de volgende stap doen: de weekinvloed uit de data halen. Dit doen we door het getal voor elke dag door het daggemiddelde te delen. Zo zijn er gemiddeld 13.6 op maandag. Door alle maandagen door 13.6 te delen komt de maandag gemiddeld op 1. Door dit te doen voor elke dag is de weekinvloed verwijderd. We krijgen dan onderstaande grafiek. Het is nu veel eenvoudiger om andere factoren (zoals trend en seizoensinvloeden) te onderzoeken.



Figuur 2.5 Zorgaanbod zonder weekinvloeden gedurende een jaar

Ook in deze grafiek zijn nog veel fluctuaties te vinden. De volgende effecten spelen nog een rol: toeval, bijzondere gebeurtenissen, seizoensinvloeden, trend. Toeval en bijzondere gebeurtenissen (zoals feestdagen) veroorzaken de korte-termijn fluctuaties. Het seizoen is de jaarlijkse cyclus, de trend is de cyclus-overstijgende dalende of stijgende tendens. Het lijkt alsof kwartaal 3 iets hoger ligt dan de rest van het jaar en het feit dat begin en eind van het jaar ongeveer gelijk zijn suggereert de afwezigheid van trend. Om dit goed te kunnen onderzoeken hebben we meer dan 1 jaar data nodig. In onderstaande grafiek staan data van twee jaren, per week gesommeerd.



Figuur 2.6 Zorgvraag per week gedurende twee jaar

We constateren dat de een niet systematisch boven de ander ligt: er is geen sprake van een trend, en de seizoenseffecten bestaan inderdaad uit een dip rond de kerst en hogere waardes aan het begin van het derde kwartaal.

Opdracht 2.2

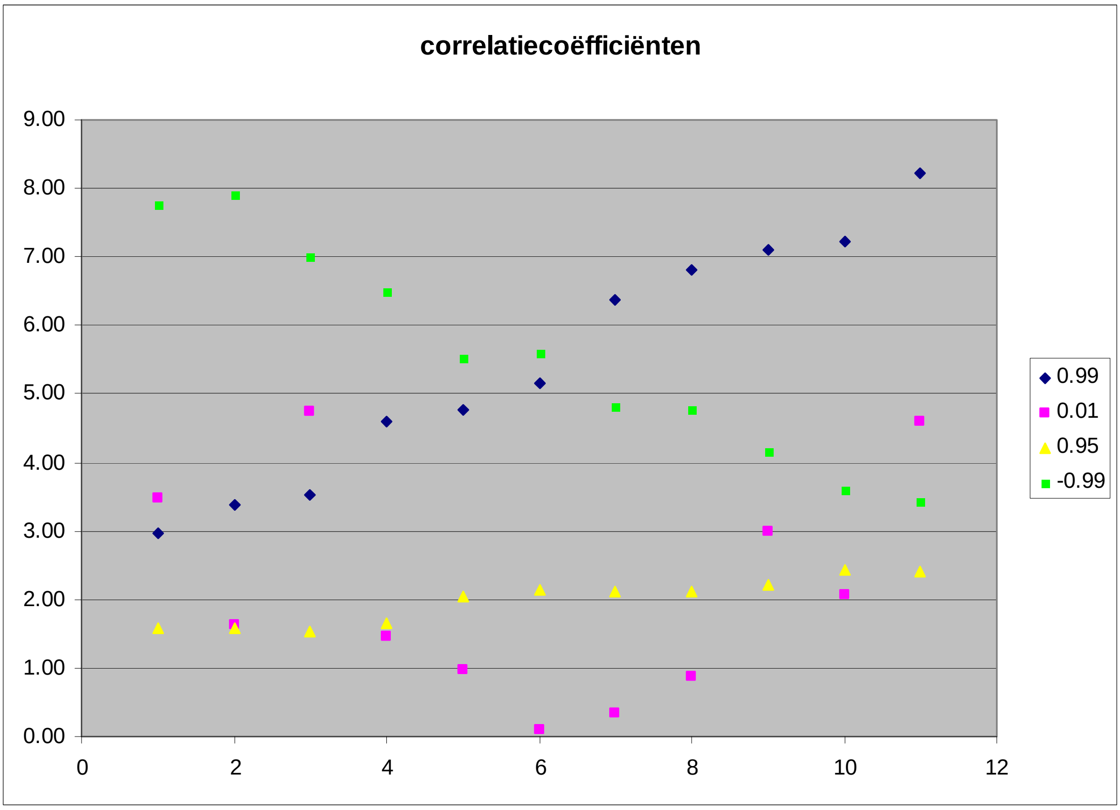
In de file aankomsten.xls staan het aantal aankomsten bij een poliklinische afdeling over 2008 en een deel van 2009. Regelmatig staan er nullen doordeweeks: dit betreft feestdagen. Bepaal het wekelijks patroon en haal de wekelijkse invloed uit de data. Is er sprake van seizoensinvloeden en trend? Hoe zou je deze toepassing kunnen gebruiken in de care-sector? Bespreek uw antwoorden met medestudenten.

2.4 Multivariate data

Er is sprake van multivariate data als er per geval sprake is van meer dan een gegeven of *attribuut*. Dus niet alleen lig- of verblijfsduur, maar bijvoorbeeld ook leeftijd en andere kenmerken. We hebben al kennisgemaakt met een speciaal geval van multivariate data, waar een van de attributen de tijd is: tijdreeksen. Bij multivariate data zijn we vaak geïnteresseerd in het verband tussen de verschillende variabelen: gaat een hogere leeftijd gemiddeld samen met een langere lig- of verblijfsduur? Eerder zagen we al de relatie tussen de dag van de week en de zorgvraag.

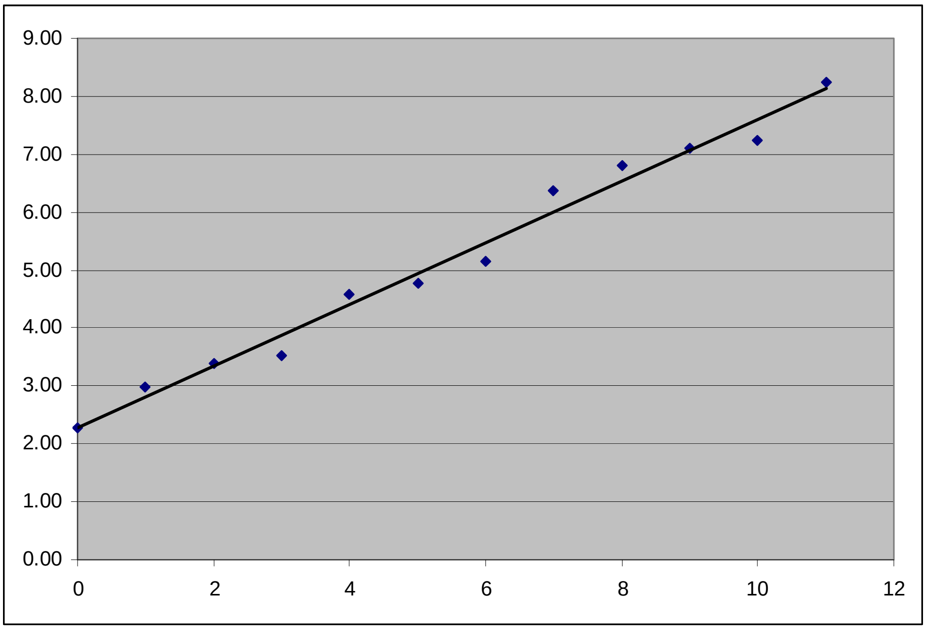
In sommige gevallen kunnen we het verband tussen twee variabelen uitrekenen. Dit is het geval als er sprake is van *kwantitatieve* data zoals lig- of verblijfsduren. *Kwalitatieve* data zoals de dagen van de week of het geslacht van personen komen hier niet voor in aanmerking. De *correlatiecoëfficiënt* geeft aan in hoeverre de datapunten op een rechte lijn liggen, dat wil zeggen, in hoeverre er sprake is van een lineair verband. Indien alle punten exact op een rechte liggen is de correlatiecoëfficiënt -1 of 1, als er geen lineair verband is 0. Enkele voorbeelden staan in onderstaande grafiek. Merk op dat de correlatiecoëfficiënt alleen rechtlijnige verbanden constateert; een U-vormig verband leidt tot een lage correlatiecoëfficiënt. Het is ook mogelijk om een dergelijk verband te beschrijven, maar dan dient men eerst een *transformatie* toe te passen. Het zou hier te ver gaan daarop in te gaan.

Verder lezen: http://en.wikipedia.org/wiki/Data_transformation_(statistics)



Figuur 2.7 Enkele kleine datasets met hun correlatiecoëfficiënten

De correlatiecoëfficiënt is onder Excel te vinden onder CORREL(). Indien de correlatie relatief hoog of laag (dicht bij 1 of -1) is het interessant de lijn te bepalen die het dichtst bij de punten ligt. Het bepalen van een dergelijke lijn noemen we lineaire *regressie*. Deze kwantificeert dan in belangrijke mate de relatie tussen de variabele op de horizontale as, de onafhankelijke of *verklarende* variabele, en de variabele op de verticale as, de *afhankelijke* variabele. De waarde op de horizontale as verklaart de waarde op de verticale as, die daarmee afhangt van de waarde op de horizontale as. Excel helpt ons weer met het vinden van de regressielijn met de functies INTERCEPT() en SLOPE(). INTERCEPT geeft het punt aan waar de lijn de verticale as kruist, SLOPE geeft de richting aan van de lijn. In onderstaande grafiek staat de regressielijn aangegeven voor een van de datasets van de voorgaande grafiek.



Figuur 2.8 Voorbeeld van lineaire regressie

In dit geval gaf INTERCEPT als waarde 2.28, en SLOPE 0.53. 2.28 is inderdaad het punt waar de regressielijn de verticale as snijdt; 0.53 is de toename van de regressielijn per eenheid op de horizontale lijn. Bij waarde 10 op de horizontale as verwachten we bij de regressielijn dus een waarde van 2.28 + 10 x 0.53 = 7.38, hetgeen we terugvinden in de grafiek.

In de volgende paragraaf over statistiek gaan we verder in op het nut van het bepalen van regressielijnen. Lineaire regressie in een variabele zoals hier besproken is slechts een eerste stap in de analyse van correlaties in data. Uiteraard zijn boeken over data analyse en statistiek een goede bron voor verdere informatie, maar tegenwoordig is er ook een schat van informatie op het internet te vinden. Met name wikipedia is een uitstekende bron. Een goed startpunt is http://en.wikipedia.org/wiki/Regression_analysis (welke uitgebreider is dan de Nederlandse pagina over dit onderwerp).

Opdracht 2.3

Beschouw de data in leeftijden-duren.xls waarin zorgduren en leeftijden van patiënten/cliënten te vinden zijn. Voor studenten die minder bekend met de mogelijkheden van Excel is het aan te raden eerst Bijlage A door te werken.

a. Bereken per leeftijd het aantal patiënten, de gemiddelde zorgduur en de standaard deviatie.
b. Maak een grafiek van de gemiddeldes en verklaar de fluctuaties bij (zeer) hoge leeftijden.
c. Maak een lineaire benadering voor zorgduren voor de groep tot 40 jaar en de groep ouder dan 40 jaar.

De volgende vraag is ter verdieping.

d. Doe hetzelfde als bij c, maar nu uitgaande van de originele data. Verklaar het verschil.

Samenvattend schema 2.1

Data analyse	**Werkwijze**
Grafiek	Goede start om gevoel te krijgen voor de data
Univariate data	Gemiddelde en standaard deviatie meest gebruikte maatstaven voor beschrijving
Excel functies	NL: GEMIDDELDE, STDEV, PERCENTIEL, CORRELATIE, SNIJPUNT, RICHTING UK: AVERAGE, STDEV, PERCENTILE, CORREL, INTERCEPT, SLOPE
Tijdreeksen	Kenmerken: - tijdscomponent - gelijktijdig optreden van cyclische effecten (week, jaar), trend, bijzondere gebeurtenissen en toeval.
Multivariate data	Lineaire regressie helpt in vinden relaties

-

3 Statistiek

Casus

Een ziekenhuis wil capaciteit op de CT-scan beschikbaar houden voor spoedpatiënten. De vraag is hoe veel slots dit moet zijn? Enerzijds wil men voldoende slots reserveren zodat het zelden (d.w.z., in niet meer dan 5% van de gevallen) voorkomt dat er niet voldoende slots zijn, anderzijds wil men de CT-scan goed benutten en de kans zo klein mogelijk houden dat er slots onbenut blijven. Men heeft data betreffende de vraag naar spoedscans. Maar hoe kan men het beste het aantal te reserveren slots berekenen?

In een verpleeghuis wordt op basis van een ruwe analyse over de afgelopen jaren, opgemerkt dat het aantal PG-cliënten dat met spoed moet worden opgenomen, vanwege een onhoudbare thuissituatie, de laatste jaren is toegenomen. Steeds vaker moet 'nee' verkocht worden bij de vraag om een dergelijke opname. Om de geldverstrekker te overtuigen van de noodzaak tot het creëren van een spoedvoorziening, wil men inzicht hebben in wanneer, hoeveel extra verblijfscapaciteit nodig is om in deze vraag te kunnen voorzien. Welke berekeningen kan men hiervoor gebruiken?

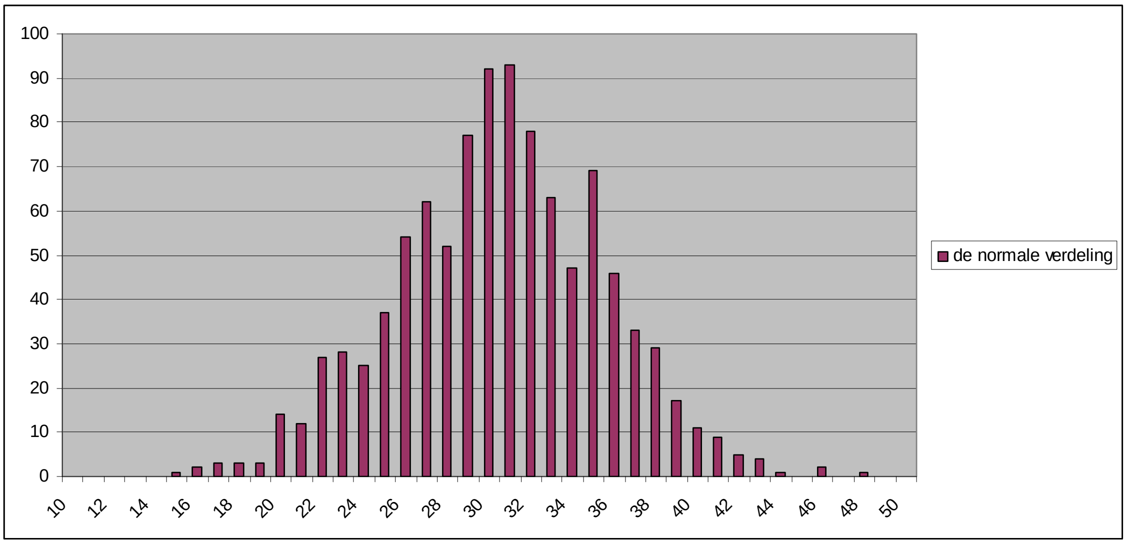
Kader 3.1 Casus 'statistiek'

De uitkomsten van data-analyses worden vaak gebruikt om uitspraken te doen die verder gaan dan de geanalyseerde data. Het kan zijn dat de gegevens van een deel van de patiënten/cliënten zijn geanalyseerd, en dat men uitspraken wil doen over de gehele populatie, of dat men op basis van huidige patiënten/cliënten iets wil zeggen over de toekomst. Om dit te kunnen veronderstelt men (vaak impliciet) dat de geanalyseerde data representatief is voor het geheel. De data die beschikbaar zijn voor analyse geven dus slechts een beperkte blik en dus beperkte informatie over de achterliggende onbekende gegevens. Het doen van verantwoorde uitspraken over een onbekende grootheid aan de hand van representatieve data heet *statistiek*. In het algemeen kan men zeggen: hoe meer gegevens en hoe lager de variantie van de steekproef (de *steekproefvariantie*), des te dichter ligt een schatting als het steekproefgemiddelde bij de werkelijke waarde.

Een belangrijk deel van de statistiek betreft het toetsen van hypotheses. Dit is bijvoorbeeld relevant voor geneeskundig onderzoek, waar vaak verschillende behandelmethodes met elkaar worden vergeleken. Dit is minder eenvoudig dan het op het eerste gezicht lijkt, want als een bepaalde methode (laten we zeggen methode A) betere resultaten geeft dan een andere (B) voor een zekere groep patiënten, betekent dat niet altijd dat methode A ook beter dan methode B is. De afwijking in de steekproef kan samenhangen met de opzet van de steekproef (bijvoorbeeld de selectie van de patiënten) of simpelweg met toeval. Een substantieel deel van de statistiek behandelt het opzetten en uitvoeren van deze zogenaamde *clinical trials*.
Bij procesanalyse speelt statistisch toetsen een minder grote rol. Enkele relevantere statistische onderwerpen komen in het vervolg ter sprake.

3.1 De normale verdeling

Statistisch worden data vaak gezien als realisaties van een onderliggende, onbekende grootheid. Schattingen voor vorm en parameters zoals gemiddelde zijn de uitkomsten van de data analyse. Een klasse van grootheden met een bepaalde vorm verdient bijzondere aandacht: de *Gaussische* of *normale verdeling*. Deze verdeling wordt gekenmerkt door een klokcurve als histogram. De exacte vorm wordt vastgelegd door twee parameters: het gemiddelde en de standaarddeviatie. Hieronder staat een histogram met 1000 trekkingen uit een normale verdeling met gemiddelde 30 en standaarddeviatie 5. De pieken die er voor zorgen dat de grafiek niet mooi glad is (zoals bij 35) komen door toeval. 1000 Andere trekkingen zouden tot andere onregelmatigheden leiden. Naarmate er meer trekkingen worden gedaan verdwijnen langzaamaan de onregelmatigheden.



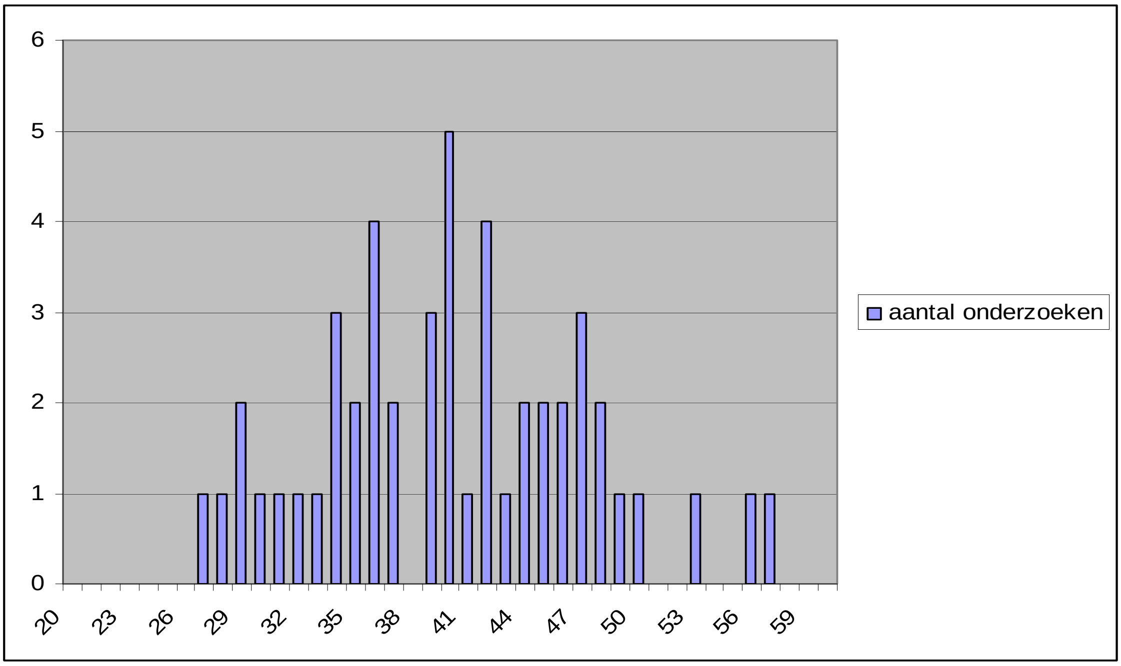
Figuur 3.1 Histogram met 1000 trekkingen uit een normale verdeling

Net zoals we voor een dataverzameling gemiddelde en standaarddeviatie uit kunnen rekenen, kunnen we dat ook voor de normale verdeling. Het gemiddelde ligt bij het hoogste punt (30 in de grafiek), de standaarddeviatie bepaalt de breedte van de grafiek. Een standaarddeviatie van 5 (zoals in bovenstaande grafiek) betekent dat 68% van alle trekkingen binnen 5 onder en boven het gemiddelde zullen vallen, in de grafiek dus in het gebied tussen 25 en 35.

Er zijn veel situaties in de praktijk waar de normale verdeling naar voren komt. We bespreken er twee die relevant zijn voor de zorg. De eerste hangt samen met de vraag naar een bepaald soort behandeling of ondersteuning. In de zorg dient vraag zich op twee wijzen aan: op afspraak of willekeurig, dus gepland of ongepland. Indien patiënten op afspraak komen, of geplande zorg ontvangen, hebben we het aantal precies in de hand (op no-shows na). Willekeurige aankomsten of ongeplande ondersteuningsactiviteiten vertonen fluctuaties die in hoge mate onvoorspelbaar zijn. Deze volgen de zogenaamde Poisson verdeling. De Poisson verdeling lijkt vanaf een gemiddelde van zeg 10 sterk op een naar het dichtstbijzijnde gehele getal afgeronde normale verdeling. De Poisson verdeling heeft de volgende bijzondere eigenschap: de variantie is gelijk aan het gemiddelde (of de standaarddeviatie is gelijk aan de wortel van het gemiddelde). In de praktijk ligt door fluctuaties in seizoen en week de variantie vaak hoger. Wanneer de variantie lager is, is er meestal sprake van (deels) planbare zorg.

Laten we als voorbeeld het aantal onderzoeken nemen bij een pathologie afdeling van een ziekenhuis. Het voorbeeld zou ook kunnen gaan over het aantal douchemomenten in een verzorgingshuis. Het histogram van een jaar data is te vinden in onderstaande grafiek. De grafiek ziet er redelijk symmetrisch als een klokcurve uit, dus er is geen reden aan te nemen dat het niet normaal verdeeld is. Het gemiddelde is 40.75. De wortel daarvan is 6.4, aanzienlijk minder dan 9.2, de standaarddeviatie van de data. Er is dus sprake van extra spreiding t.o.v. de Poisson verdeling.

De tweede eigenschap van de normale verdeling hangt samen met sommen van toevalsgrootheden en komt in de volgende paragraaf ter sprake.



Figuur 3.2 Histogram van aantal onderzoeken op een bepaalde dag van de week

Opdracht 3.1

Neem de dagtotalen uit aankomsten.xls, en plot deze per dag. Zou hier sprake kunnen zijn van normale verdelingen? Bereken de gemiddeldes en standaard deviaties.

Opdracht 3.2

Bekijk de datasets zorgduren.xls en leeftijd-duren.xls en maak histogrammen van de zorgduren. Is hier sprake van normale verdelingen?

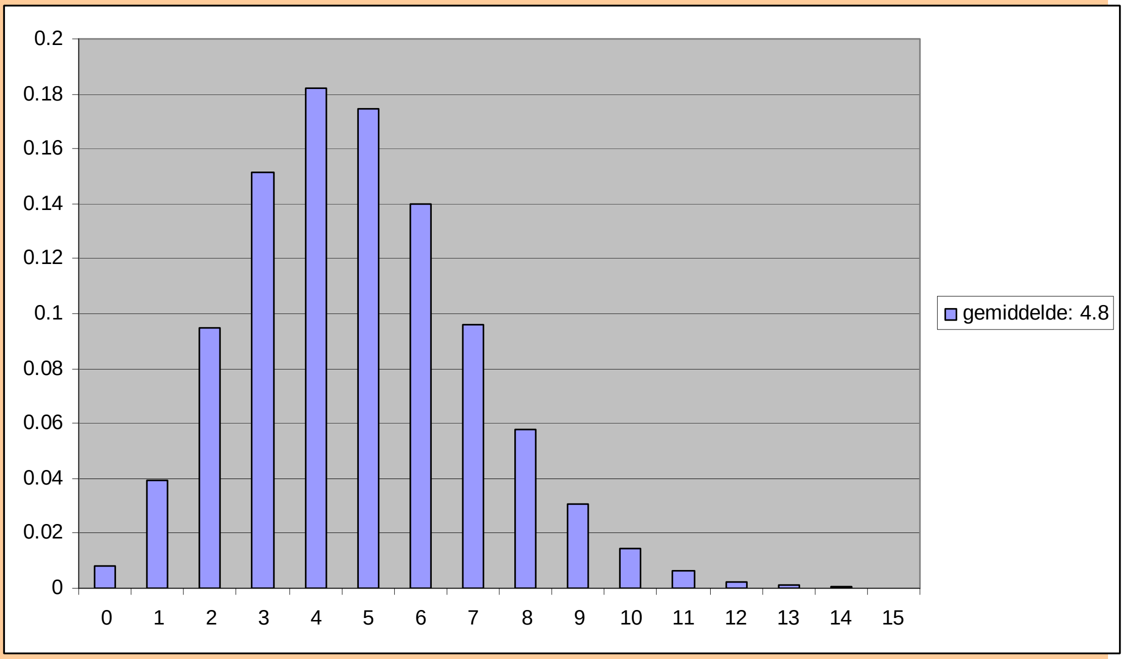
Verder lezen: http://nl.wikipedia.org/wiki/Poissonverdeling

3.2 Rekenen met toevalsgrootheden

In de vorige vraag hebben we kennisgemaakt met de normale verdeling en gezien hoe we de normale verdeling kunnen herkennen. We hebben ook geleerd dat de normale verdeling gekarakteriseerd wordt door twee getallen: gemiddelde en standaarddeviatie. Hoe kunnen we rekenen met de normale verdeling en andere verdelingen? Het volgende voorbeeld, vergelijkbaar met de casus aan het begin van dit hoofdstuk, zal dit duidelijker maken.

Voorbeeld

Voor een zorgpad wil men op de vaste wekelijkse dag dat de arts de patiënten ziet meteen de in veel gevallen voorgeschreven MRI uitvoeren. Daarvoor reserveert men een aantal slots. Hoeveel slots moet men reserveren zodat vraag en aanbod zoveel mogelijk in balans zijn? Gemiddeld schrijft de arts voor 4.8 patiënten een MRI scan voor. 5 Slots reserveren zou de fluctuaties in vraag negeren. Uit de data analyse blijken er weinig seizoensinvloeden te zijn. Vanuit theoretische overwegingen verwacht men dat de wekelijkse vraag Poisson verdeeld is; de standaarddeviatie (2.2) wijst daar inderdaad op. Nu kan uitgerekend worden wat de kans is, als er daadwerkelijk 5 slots gereserveerd worden, dat er meer dan 5 slots nodig zijn. Een histogram van de Poisson verdeling met gemiddelde 4.8 staat hieronder.



Uit bovenstaande grafiek zien we dat waardes van 6 en hoger vaak voorkomen: de kans van 6 en hoger bij elkaar opgeteld is 0.35. Dus in 35% van de gevallen, meer dan 1 op de 3 weken, is er dus een tekort aan slots. Dit pleit er voor meer slots te reserveren: Bij 6 slots is de kans op een tekort al teruggebracht tot 21%. Hoe meer slots we reserveren, hoe kleiner de kans op een tekort. Echter, het aantal ongebruikte slots neemt ook toe: gemiddeld 0.97 bij 5 gereserveerde slots, naar 1.62 slots bij 6 reserveringen.

Kader 3.2 Voorbeeld berekening aantal slots

Opdracht 3.3

Bovenstaande grafiek is gemaakt met de Excel functie POISSON, het eerste getal is bijvoorbeeld verkregen met POISSON(0,4.8,FALSE).

a. Reproduceer de grafiek
b. Bereken de kans op tekort voor verschillende aantallen gereserveerde slots
c. Verdiepingsvraag: reproduceer het gemiddeld aantal ongebruikte slots en bereken het gemiddeld tekort.

Merk op dat het gemiddeld tekort wat anders is dan de kans op tekort!

Bovenstaand voorbeeld laat goed zien hoe we de Poisson verdeling in de praktijk kunnen gebruiken. De normale verdeling kan op dezelfde wijze worden ingezet. De Excel functie NORMDIST kan gebruikt worden om de percentielen uit te rekenen. Het is handig om een paar percentielen te onthouden: de kans dat de afwijking van het gemiddelde (naar boven en naar beneden) minder dan 1 maal (2 maal) de standaardafwijking is is 68% (95%).

Opdracht 3.4

a. Maak op basis van onderstaande gegevens zelf een vergelijkbaar histogram als in kader 3.2.

Het aantal PG-cliënten dat voor een crisis-opname (per week) wordt aangemeld in een verpleeghuis is als volgt verdeeld:

Aanmeldingen crisis-opname	Kans
0	0,06
1	0,23
2	0,27
3	0,38
5	0,02
6	0,02
8	0,02

b. Wat is het gemiddeld aantal crisis-aanmeldingen?

Het verpleeghuis wil met betrekking tot de crisis-opname in 94% van de gevallen over voldoende capaciteit beschikken.

c. Hoeveel verblijfscapaciteit moet het verpleeghuis per week reserveren om aan bovenstaande doelstelling te voldoen?

Opdracht 3.5

a. Een bepaald soort behandeling of zorgactiviteit, bijvoorbeeld een operatie, heeft een gemiddelde lengte van 70 minuten met een standaardafwijking van rond de 10 minuten. Wat is de kans dat een dergelijke behandeling of zorgactiviteit meer dan 90 minuten duurt?
b. Wekelijks dienen zich gemiddeld 100 nieuwe patiënten/cliënten aan van een bepaald type. Wat is de kans dat er meer dan 110 zijn? En minder dan 80?

Net zoals we getallen op willen kunnen tellen, willen we ook toevalsvariabelen bij elkaar op kunnen tellen, om dan bijvoorbeeld de percentielen van de som uit te kunnen rekenen. Een goed voorbeeld is de lengte van een OK sessie of van een zorgroute.

Voorbeeld cure

Voor een OK sessie zijn een aantal operaties ingepland. Van de operaties zijn betrouwbare schattingen van gemiddelde lengte en standaarddeviatie bekend. Voor de gehele sessie is een tijdsblok van 08.00 tot 15.30 uur gereserveerd. Men is geïnteresseerd in de kans dat de OK sessie uitloopt tot na 15.30 uur.

Voorbeeld care

In een zorgroute zijn een aantal zorgactiviteiten ingepland. Per zorgactiviteit is een betrouwbare schatting van de gemiddelde duur in de vorm van normtijden bekend. Ook de standaarddeviaties zijn bekend. Voor de zorgroute is een tijdsblok van 07.00 tot 11.00 uur gereserveerd. Men is geïnteresseerd in de kans dat de zorgroute uitloopt tot na 11.00 uur.

Kader 3.3 Voorbeeld OK-sessie duur / zorgroute duur

De volgende rekenregels zijn er voor sommen van toevalsgrootheden: gemiddeldes mogen worden opgeteld, maar ook varianties. Laten we dit toelichten aan de hand van het voorbeeld: als er twee operaties zijn met standaarddeviatie 10, dan is de som van de varianties $10^2+10^2=200$, en de standaarddeviatie van de som dus $\sqrt{200} = 14.14$. We zien dus dat de standaarddeviatie relatief gezien afneemt. Dit geldt in het algemeen. Hoe lager de (relatieve) standaarddeviatie, hoe makkelijker het is vraag en aanbod af te stemmen. Anders gezegd: het vergroten van de schaal leidt tot een betere, efficiëntere, afstemming van vraag en aanbod. Schaalvoordelen zijn dus deels terug te voeren op eigenschappen van sommen van toevalsgrootheden.

Voor sommige soorten toevalsgrootheden zijn er nog meer eigenschappen bekend. Dit is ook het geval voor de normale en de Poisson verdeling. Er geldt namelijk dat sommen van normale en Poissonverdelingen weer normaal en Poisson verdeeld zijn. Voor de normale verdelingen gelden eigenschappen die nog sterker zijn, namelijk dat gemiddeldes (en ook sommen) van willekeurige verdeelde grootheden bij benadering normaal verdeeld zijn. Dit resultaat staat in de wiskunde bekend als de *Centrale Limietstelling*. In de zorg passen we dit bijvoorbeeld toe in het zojuist beschreven voorbeeld van lengtes van OK sessies of duur zorgroutes. Operatieduren of afzonderlijke zorgactiviteiten zijn in de regel niet normaal verdeeld, maar duren van de hele sessie of route kunnen vaak heel goed met een normale verdeling benaderd worden, vanwege de Centrale Limietstelling.

3.3 Trekken uit toevalsgrootheden

Een belangrijk nog te bespreken onderwerp is het in de computer kunnen nabootsen van al dan niet bestaande systemen, het zogenaamde *simuleren*. De basale stap in deze simulaties is het herhaaldelijk kunnen trekken uit toevalsgrootheden zoals zorgduren of aankomstmomenten, aangezien dit de bouwstenen van de simulaties van gehele systemen zijn. De stap die hier weer aan te grondslag is de functionaliteit van de meeste computertalen om een willekeurig getal tussen 0 en 1 te genereren. Excel heeft deze functionaliteit ook, middels de functie RAND(). Voor sommige verdelingen heeft Excel ook nog een functie die RAND vertaalt in een trekking uit die verdeling. Dit is onder andere het geval bij de normale verdeling: NORMINV(RAND(),30,10) geeft een trekking uit een normale verdeling met gemiddelde 30 en standaarddeviatie 10. Ook voor andere verdelingen bestaan er soortgelijke functies, maar helaas niet voor de Poisson verdeling. Een oplossing is de Data Analysis add-in te gebruiken (keuze "Random Number Generation"), of natuurlijk de functie zelf te programmeren.

Opdracht 3.6

In de vorige paragraaf was sprake van een Poisson verdeelde vraag naar MRI slots. Herhaal onderdelen b en c van de bijbehorende opgave door een aantal trekkingen uit een Poisson verdeelde grootheid te doen. Neem hiervoor voldoende trekkingen, bijvoorbeeld 1000.

Opdracht 3.7

Trek een aantal malen uit een normale verdeling, gebruikmakend van de Excel NORMINV functie. Gebruik dit om de percentielen op 1 en 2 standaarddeviaties van het gemiddelde mee te berekenen. Merk op dat met F9 we het hele sheet opnieuw doorrekenen. Gebruik F9 om de berekeningen een aantal malen overnieuw te doen.

3.4 Nauwkeurigheid en steekproefomvang

Bij het opzetten van een meting speelt nauwkeurigheid een belangrijke rol. Stel dat we de verblijftijd van 50 cliënten in een verzorgingshuis kennen. Met behulp van Excel kunnen we hieruit eenvoudig het gemiddelde berekenen. Vanwege de variatie zal de gemiddelde verblijftijd van de 50 cliënten (*steekproefgemiddelde*) niet overeenkomen met de werkelijk gemiddelde verblijftijd. Hoe minder data men heeft, hoe groter de mogelijke fout zal zijn. Aangezien metingen tijdrovend zijn, is er een spanningsveld tussen de nauwkeurigheid en de benodigde inzet in termen van geld en tijdsbesteding en duur van de meting. Vragen over de tijdspanne van een meting of het benodigde aantal datapunten (*steekproefomvang*) komen dan automatisch naar voren. Wanneer relevante logistieke data uit informatiesystemen kan worden gehaald, dan speelt nauwkeurigheid over het algemeen een minder belangrijke rol. Dit neemt niet weg dat data wel onbetrouwbaar kan zijn.

Allereerst dient opgemerkt te worden dat het bij het meten over een kortere tijdsperiode niet mogelijk is om seizoensinvloeden te ontdekken. Het aantal ambulanceritten en het aantal opnames op de SEH zijn vaak relatief hoog in november en relatief laag in augustus. Wanneer men slechts data heeft gedurende een maand, dan worden dergelijke patronen niet gevonden. Bovendien zorgen seizoensinvloeden ervoor dat er meer geschat dient te worden; we zijn nu bijvoorbeeld geïnteresseerd in het gemiddelde aantal ritten en opnames per maand in plaats van het gemiddelde aantal over een heel jaar. Bij tijdsafhankelijke patronen, of andere vormen van afhankelijkheid, zijn dus meer gegevens nodig dan wanneer deze patronen niet aanwezig zijn.

Laten we er vanuit gaan dat seizoensinvloeden niet van belang zijn en dat we geïnteresseerd zijn in een enkele grootheid. We kunnen hierbij denken aan de *gemiddelde* verblijftijd en de *fractie* cliënten met een vervolgtraject. Een belangrijke term uit de statistiek is de *betrouwbaarheid*. De betrouwbaarheid is de mate waarin de meetresultaten een afspiegeling zijn van de werkelijke waarde van de te meten variabele. Een veel gebruikte waarde is 95%. Stel dat de steekproefomvang n is, ofwel we hebben n gegevens. De maximale afwijking tussen de waarneming uit de steekproef, zoals het steekproefgemiddelde, en de werkelijke waarde wordt dan bij benadering gegeven door $z\frac{st.dev}{\sqrt{n}}$, waarbij *st.dev* de standaarddeviatie van de grootheid is en *z* een waarde die afhangt van de gekozen betrouwbaarheid; hoe hoger de betrouwbaarheid, hoe hoger de waarde van *z* en hoe groter de maximale fout. Als vuistregel wordt gebruikt dat deze benadering goed is wanneer het aantal waarnemingen niet te laag is, zeg tenminste 30 tot 50. Het getal *z* wordt bepaald door het kwantiel van de normale verdeling; bij een betrouwbaarheid van 95% is *z* gelijk aan 1.96. Uit de formule voor de maximale afwijking volgt ook direct dat deze kan kleiner wordt naarmate het aantal waarnemingen *n* groter is. Verder wordt ook de maximale afwijking groter wanneer we een variabele waarnemen met veel variatie (hoge standaarddeviatie). In sommige gevallen is de standaarddeviatie al bekend, bijvoorbeeld vanwege eerdere experimenten. Anders kan deze geschat worden door het bepalen van de standaarddeviatie van de *n* waarnemingen. Het invullen in bovenstaande formule geeft nu bij benadering de maximale afwijking.

Met het bovenstaande is het ook mogelijk terug te rekenen hoeveel waarnemingen nodig zijn om een bepaalde maximale afwijking te verkrijgen bij een gekozen betrouwbaarheid. Zo wordt het aantal waarnemingen bepaald door $\left(z\dfrac{st.dev}{afwijking}\right)^2$, waarbij *afwijking* de maximale afwijking is. Uit de formule volgt direct dat naarmate de maximale afwijking kleiner is, het benodigde aantal waarneming stijgt. Vooral bij een zeer kleine afwijking zal het aantal extra gegevens snel toenemen.

Voorbeeld

Bij een thuiszorginstelling werd de activiteit `hulp bij wassen' gemeten bij 50 cliënten. Met Excel rekende men na dat de gemiddelde duur 20 minuten was en de standaarddeviatie 15 minuten. De maximale afwijking tussen het gemeten gemiddelde en de werkelijk gemiddelde activiteitsduur, bij een betrouwbaarheid van 95%, is 1.96 x 15 / √50 ≈ 4,2 minuten. Ofwel, bij een betrouwbaarheid van 95%, ligt de gemiddelde duur van `hulp bij wassen' tussen de 20 – 4,2 = 15,8 en 20 + 4,2 = 24,2 minuten.

Deze gegevens vond men niet nauwkeurig genoeg en men besloot de meting voort te zetten. Hierbij kwam direct de vraag op tafel hoe lang nog gemeten diende te worden. Er werd in ieder geval overeen gekomen om een maximale afwijking van het gemeten gemiddelde ten opzichte van het werkelijk gemiddelde toe te staan van 2 minuten. Door gebruik te maken van de reeds geschatte standaarddeviatie van 15 minuten, werd het aantal benodigde gegevens geschat op (1,96 x 15 / 2) x (1,96 x 15 / 2) = 14,7 x 14,7 = 216,09. In totaal zijn dus 217 gegevens nodig, waardoor de activiteitsduur bij in ieder geval nog 167 cliënten gemeten dient te worden.

Opdracht 3.8

Beschouw het bovenstaande voorbeeld.

a. Reproduceer de bovengenoemde getallen in Excel
b. Ga na wat het geschatte benodigde aantal waarnemingen zou zijn wanneer men een maximale afwijking van drie, één en een halve minuut wenselijk vindt.

3.5 Voorspellen

In de voorgaande secties hebben we gezien hoe statistiek ingezet kan worden om uitspraken te doen over grootheden die zich nog moeten realiseren, op grond van representatieve informatie. In een zeker opzicht is er dus sprake van een voorspelling. Bij tijdsreeksen is men ook vaak geïnteresseerd in schattingen van toekomstige waardes. Hiervoor wordt ook vaak de Engelse term *forecasting* gebruikt.

De belangrijkste stap in forecasting is het analyseren van de historische data. Als eenmaal trend en seizoensinvloeden bepaald zijn kan de trend worden doorgetrokken en de seizoensinvloed weer worden ingebracht.

Een belangrijke vraag is hoe nauwkeurig de aldus verkregen forecast is. In veel situaties worden forecasts als "waar" aangenomen en elke afwijking wordt gezien als een fout. Echter, het is hoogst toevallig als een forecast precies goed is!
Er zijn twee soorten afwijkingen: systematische afwijkingen en ruis. Er is sprake van een systematische afwijking als trend of seizoensinvloeden niet overeenkomen met de werkelijkheid en de forecast systematisch te hoog of te laag is. Ruis is het feit dat niet alle fluctuaties voorspelbaar zijn. Dit zagen we al eerder: het verwijderen van seizoensinvloeden zorgt niet voor een volledig vlakke lijn, het is de ruis die overblijft. Bij korte-termijn forecasts kan men er van uitgaan dat trend en seizoen weinig veranderen, en dat de afwijking dus hoofdzakelijk bepaald wordt door de ruis. Soms kan de ruis verder worden gereduceerd door nieuwe verklarende factoren op te sporen zoals de invloed van feestdagen. Op de langere termijn neemt de afwijking toe. De verklaring is een systematische fout die toeneemt in de tijd, omdat trend en in mindere mate seizoensinvloeden wijzigen na verloop van tijd. Om die redenen moet men bij het interpreteren van forecasts altijd uitgaan van een foutmarge. Deze foutmarge kan ook gekwantificeerd worden, met name als het gaat om de ruis. Dit is zeer belangrijk voor de korte-termijn forecasts die vaak voor operationele doeleinden worden gebruikt. Aangezien de afwijking bij langere termijn voornamelijk bepaald wordt door wijzigingen in trend, is deze wijze van onder- en bovengrenzen schatten minder geschikt. Wat voor de langere termijn (meerdere jaren of zelfs decennia) geschikter is is het analyseren van verschillende scenario's. In alle gevallen dient men zich te realiseren dat lange-termijn forecasts vaak zeer onbetrouwbaar zijn.

We hebben zojuist een eenvoudige wiskundige methode voor forecasting besproken waarbij we expliciet seizoensfactoren en trend bepalen. Er bestaan ook geavanceerdere methodes, maar deze blijken in de praktijk nauwelijks beter te werken, en zij missen het voordeel van transparantie van de eenvoudiger methodes.
In veel situaties worden forecasts gemaakt op basis van menselijk inschattingsvermogen, zonder hulp van wiskundige methodes. Onderzoek naar de kwaliteit van forecastingmethodes laat zien dat dit vaak slechtere resultaat geeft dan puur objectieve wiskundige methodes. Een van de redenen dat mensen verschillende *biases* hebben, waaronder de neiging forecasts te maken op basis van wat wenselijk is en minder op wat werkelijk te verwachten is.

Opdracht 3.9

Beschouw de oefening over tijdsreeksen van paragraaf 2.3.

a. Bepaal een forecast voor de rest van 2009.
b. Verdiepingsvraag: Geef onder- en bovengrenzen waarbinnen de voorspelde waardes zich (bij gelijkblijvende trend) met 95% zal bevinden

c. Geef een forecast op kwartaalniveau voor 2010 t/m 2020. Is dit een realistische forecast? Bespreek uw antwoordt met medestudenten.

Verder lezen Een zeer interessant boek met een kritische en deskundige blik op lange-termijn forecasting is S.G. Makridakis, *Forecasting, Planning, and Strategies for the 21st Century*, The Free Press, 1990.

3.6 Scenario-analyse

Het doel van zorglogistiek is het verbeteren van de prestaties van zorgprocessen. Soms volstaat het aankomstprocessen of zorgduren te analyseren. Vaak echter is het wenselijk om de prestaties van soms zeer complexe processen te kunnen voorspellen om zo ideeën ter verbetering te evalueren. Dit voorkomt dure implementaties van nieuwe manieren van werken die bij nader inzien niet presteren zoals gehoopt. Scenario-analyse biedt ook een belangrijke objectieve en kwantitatieve ondersteuning voor strategische capaciteitsbeslissingen. De consequenties van verschillende logistieke investeringsbeslissingen kunnen vooraf worden doorgerekend en ingeschat. Tevens kunnen lange termijn ontwikkelingen in de zorg in de scenario's worden verwerkt.

Voorbeeld

Bij verpleeghuis `Sterk in Zorg' komen steeds meer cliënten die zorg behoeven. Vanwege deze aanhoudende groei in vraag wil men een vleugel van het verpleeghuis verbouwen. Er is nog onduidelijkheid over het aantal verdiepingen, de indeling en het aantal kamers op de nieuwe vleugel. Om de capaciteit juist in te schatten wordt een scenario-analyse uitgevoerd, waarmee het benodigde aantal kamers kan worden bepaald. Op basis van deze analyse worden de bouwplannen ontwikkeld.

Kader 3.4 Voorbeeld capaciteit verpleeghuis

Voor het kwantitatief analyseren van processen zijn verschillende methodes. Soms is het mogelijk een nieuwe manier van werken statistisch te analyseren, daaruit conclusies te trekken voor een betere manier van werken en dit ook aannemelijk te maken.

Voorbeeld

Bij een polikliniek werden gedurende enige tijd wachttijden gemeten en ook andere factoren zoals het aantal zieke medewerkers, aantal artsen dat te laat startte, enzovoorts. Een statistische analyse liet zien dat de wachttijden sterk gecorreleerd waren met artsen die te laat begonnen. En men kon de reductie in wachttijd schatten indien dit niet meer het geval zou zijn. Nieuw beleid ten aanzien van het op tijd beginnen leidde inderdaad tot de gewenste effecten.

Kader 3.5 Voorbeeld wachttijden polikliniek

Bij complexe processen schiet een statistische analyse vaak tekort. Zo is een statistische analyse vaak wel geschikt voor het schatten van het toekomstig aantal preparaten bij pathologie; de consequenties op de doorlooptijden zijn veel moeilijker statistisch te schatten. Hiervoor gebruiken we een andere methode, namelijk *wiskundig modelleren*. Een wiskundig model is een vereenvoudigde beschrijving van het werkelijk systeem op basis waarvan we uitspraken doen over de werkelijkheid. Typische componenten van een dergelijk systeem zijn aankomstprocessen, wachtrijen, verwerkingscapaciteiten van allerlei soorten, enzovoorts. Het inzetten van dit soort modellen voor het doorrekenen van verschillende scenario's noemen we wiskundig modelleren. Er bestaan verschillende soorten wiskundige modellen, wachtrijsystemen zijn het meest relevant voor zorglogistiek. Ook zijn er verschillende manieren om wiskundige modellen en in het bijzonder wachtrijsystemen op te lossen. Voor de meest eenvoudige wachtrijsystemen bestaan er formules. Een algemene oplossingsmethode, die om die reden ook veel gebruikt wordt, is *simulatie*. In wat volgt gaan we in meer detail op simulatie en wachtrijmodellen in.

Samenvattend schema 3.1

Statistiek	**Werkwijze**
Omschrijving	Statistiek is het doen van verantwoorde uitspraken over een onbekende grootheid aan de hand van representatieve data.
Wetmatigheden	De normale verdeling speelt een centrale rol in de statistiek Bij sommen van grootheden mogen gemiddeldes en varianties worden opgeteld Voor het simuleren van zorgprocessen moeten we uit toevalsgrootheden kunnen trekken
Forecasting	Eenvoudige forecasting methodes zijn goed bruikbaar voor korte-termijn forecasts, lange-termijn forecasts zijn in de regel zeer onbetrouwbaar
Excel functies	NORMDIST, POISSON, RAND, NORMINV
Tijdreeksen	Kenmerken: - tijdscomponent - gelijktijdig optreden van cyclische effecten (week, jaar), trend, bijzondere gebeurtenissen en toeval.

4 Modellen

Casus

Mevrouw de Boer heeft al tijden last van haar lies en bilstreek, waarbij de pijn steeds verder haar bovenbeen in trekt. Door de huisarts is zij verwezen naar de poli orthopedie. Zij is ruim op tijd voor haar afspraak van 10.00 uur, maar is uiteindelijk pas om 10.34 uur aan de beurt. Tijdens haar afspraak wordt besloten een foto te nemen van de heup. Het radiologisch onderzoek vindt op inloopbasis plaats, maar mevrouw de Boer heeft geluk en er kan direct een foto worden gemaakt. De heup blijkt versleten en er wordt besloten tot een operatie voor het plaatsen van een heupprothese. Er is een wachtlijst, dus ze kan pas 9 weken later terecht voor de geplande ingreep. De operatie zou op maandag plaatsvinden, maar vanwege uitloop van het dagprogramma van de OK wordt de operatie een dag uitgesteld. Mevrouw de Boer is erg tevreden over de zorg tijdens haar verblijf op de afdeling na afloop van de OK en de fysiotherapie verloopt voorspoedig. Op zaterdag zou zij naar huis mogen, maar vanwege de voor haar noodzakelijke thuiszorg gaat zij pas op maandag naar huis. Het herstel verloopt prima en na enige tijd is zij weer geheel op de been en de thuiszorg wordt gestopt.

Kader 4.1 Casus 'wiskundige modellen'

In bovenbeschreven casus maakt de patiënt gebruik van diverse soorten capaciteiten met eigen karakteristieken. In haar zorgproces zijn ook verschillende momenten en soorten van wachten te onderscheiden. Wachttijden en vertragingen in het zorgproces komen voort uit een (tijdelijk) tekort aan capaciteit; dit zijn perioden waarin vraag en aanbod niet geheel op elkaar zijn afgestemd. Vaak spelen fluctuaties en onzekerheid hierbij een rol. Hiermee zijn wachttijden en capaciteiten onlosmakelijk met elkaar verbonden. Wiskundige modellen zijn dan nodig om de benodigde of gewenste capaciteit te bepalen.

Een wiskundig model is een wiskundige, kwantitatieve beschrijving van een al dan niet bestaand systeem of proces. Eenvoudige modellen kunnen worden gebruikt om bijvoorbeeld de duur van een OK-sessie te voorspellen, en bestaat dan in zijn eenvoudigste vorm uit de som van de operatieduren. Hiermee kan bijvoorbeeld de kans op een afgezegde OK worden bepaald. Complexere modellen bestaan uit vele componenten zoals aankomstprocessen, wachtrijmodellen, verschillende soorten capaciteit die al dan niet gelijktijdig ingezet moeten worden, enzovoorts. Hieronder zullen we verschillende modellen toelichten aan de hand van bovenstaande casus.

4.1 Het berekenen van sessielengtes

Het plannen en bijsturen van OK-sessies is een zeer complex probleem waar veel ziekenhuizen mee worstelen, en waar tegelijkertijd nog veel verbeteringen mogelijk zijn. Het gebeurt in meerdere stappen waarbij meestal voor langere tijd sessies op vaste dagen worden toegewezen aan specialismes, die zij dan, gegeven zekere randvoorwaarden, naar eigen goeddunken kunnen invullen. Een deel van die randvoorwaarden heeft betrekking op de benutting van de OK's, met name op mogelijke uitloop.
De meeste ziekenhuizen hebben meerdere OK's. Om uitloop te voorkomen worden er operaties van OK verplaatst en afgezegd. Daarnaast is er wellicht sprake van spoedoperaties die een plek in de planning moeten krijgen. Dit bijsturen is een uiterst complex probleem wat ook zeer uitdagend is voor specialisten op het gebied van planning.

Om het afzeggen en van OK verplaatsen zoveel mogelijk te vermijden is het belangrijk dat het initiële plan zo goed mogelijk uitvoerbaar is. Dit plan bestaat uit, per OK, een lijst verrichtingen.

In deze paragraaf analyseren we dit eenvoudigst mogelijke probleem: een enkele operatiekamer waarin een gegeven aantal operaties gepland zijn gedurende een sessie met een bepaalde lengte. We willen graag weten wat de kans is dat het programma uitloopt en hoeveel. We nemen aan dat we een aantal operaties willen plannen waarvan historische data bekend zijn die representatief zijn voor de te verwachten operatieduren. We nemen ook aan dat omschakeltijden onderdeel zijn van de operatieduren zodat we hier geen rekening mee hoeven te houden. Tenslotte nemen we aan dat we elke operatie in aansluiting op de voorgaande kunnen beginnen, alle patiënten en operatiepersoneel zijn dus op tijd aanwezig.

Er zijn ruwweg twee soorten aanpakken voor dit probleem. De eerste aanpak is puur gebaseerd op simulatie. Neem aan dat elk van de geplande operaties zeg al 1000 maal is uitgevoerd en dat de tijden hiervan bekend zijn. Voor elk van de geplande operaties wordt er dan willekeurig een tijd getrokken; deze tijden worden bij elkaar opgeteld en er wordt gekeken of de totale duur de sessietijd overschrijdt. Dit wordt 1000 keer herhaald: de fractie overschrijdingen geldt als een schatting voor de kans dat er bij een volgende sessie een overschrijding optreedt. Dit soort simulatie heet *Monte Carlo simulatie* en kan in Excel worden uitgevoerd, eventueel met behulp van een speciale simulatie *add-in*.

Een modelmatige aanpak zou als volgt zijn. Uit de data wordt voor elke operatie het gemiddelde en de variantie van de duur geschat. De gemiddeldes en de varianties worden opgeteld, en zo komen we tot de verwachte lengte en variantie van de sessieduur. Nu nemen we aan dat de totale duur bij benadering normaal verdeeld is. En gebruikmakend van bijvoorbeeld een Excel functie kunnen we uitrekenen wat de overschrijdingskans is. Het voordeel van de modelmatige aanpak is de eenvoud, en het inzicht dat we verkrijgen door bijvoorbeeld de variantie van een van de duren te veranderen. De simulatie is tijdrovender maar nauwkeuriger. Daarnaast is de simulatie eenvoudig uit te breiden naar allerlei andere situaties, zoals het afzeggen van operaties als ze veel te laat beginnen, het verschuiven van operaties van de ene OK naar de andere, enzovoorts.

Welke methode we ook gebruiken, het blijft noodzakelijk te controleren of de resultaten van een analyse overeenkomen met de werkelijkheid. Het is bijvoorbeeld denkbaar dat operaties sneller verlopen als er eerder vertraging is ontstaan. Dit is nu niet in het model meegekomen. Of dit daadwerkelijk het geval is, zou statistisch kunnen worden onderzocht. Ook de aanname dat patiënten altijd op tijd aanwezig zijn is weinig realistisch.
Toch kan het heel goed het geval zijn dat het model goed overeenkomt met de werkelijkheid. Controleren of dit het geval is heet *validatie* en is essentieel voor elk modelleringsproject.

Opdracht 4.1

a. Maak een Excel sheet waarin de gebruiker voor een 10-tal zorgactiviteiten, zoals operaties, gemiddeldes en standaarddeviaties en de totale sessieduur of duur van de zorgroute in kan voeren.
b. Bereken de standaarddeviatie van de som en bereken de kans dat de sessie uitloopt.

Uit historische gegevens blijkt dat artsen de operatieduur met gemiddeld 10% onderschatten.

c. Bereken nu opnieuw de uitloopkans.

d. Trek een groot aantal maal uit een normale verdeling om zo het gemiddeld aantal minuten uitloop te bepalen.

4.2 Simuleren van processen

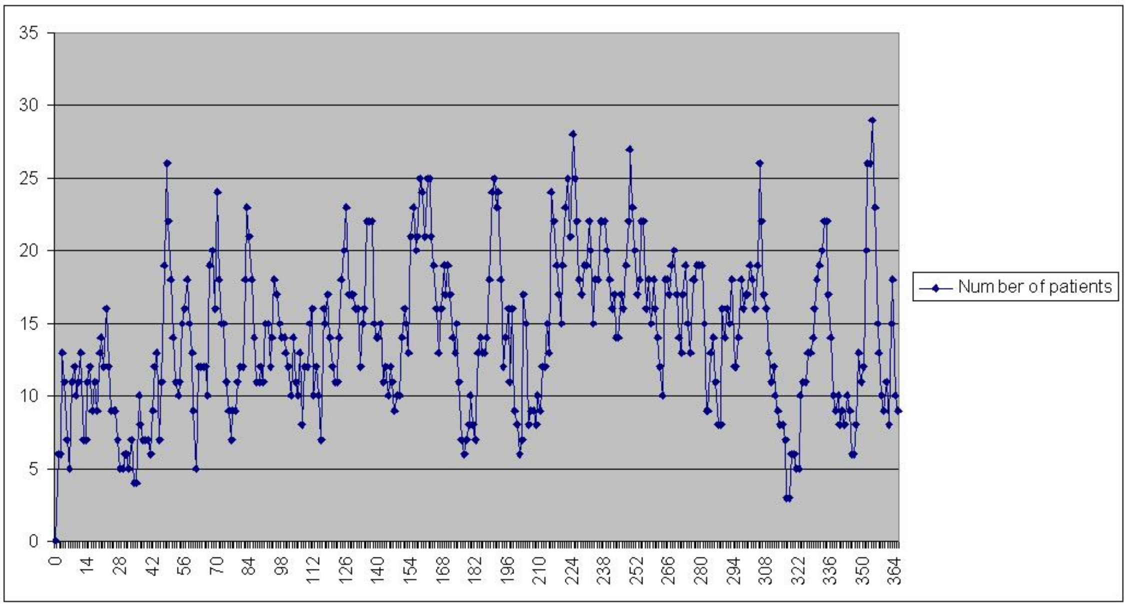
Vaak is men geïnteresseerd in processen die meer dan enkele gebeurtenissen betreffen. Uitgestelde operaties nemen op een later moment capaciteit in beslag, wat wellicht leidt tot wachttijden bij andere operaties. Het kan dus zinvol zijn OK-sessies in samenhang te beschouwen, zowel parallelle sessies op een dag, als die op achtereenvolgende dagen. Dit geldt ook voor ketenzorg. Simulaties van dit soort processen noemen we *discrete-event simulation*. Het aantal te modelleren verrichtingen wordt daarbij al gauw zodanig dat het gebruik van Excel niet meer praktisch en misschien zelfs niet meer mogelijk is. Voor dit soort langere-termijn simulaties bestaan speciale software tools met een grafische interface waarin de gebruiker het te modelleren proces grafisch op kan bouwen. Een andere mogelijkheid is om het zelf te (laten) programmeren in een programmeertaal als Java of C++. In principe elk zorgproces, hoe ingewikkeld ook, kan hiermee nagebouwd en gesimuleerd worden. Toch moet men grote terughoudendheid betrachten bij het simuleren van complexe processen. Aangezien elk model een vereenvoudiging is van de werkelijkheid zullen de resultaten van de simulatie nooit volledig sporen met de werkelijkheid. Soms is het verschil zo groot dat de resultaten niet echt bruikbaar zijn. Welke factoren essentieel zijn hangt af van de situatie, en sommige vaak essentiële factoren zijn heel moeilijk te modelleren. Dit betreft met name menselijk handelen: dit is vaak moeilijk in regels te vangen en daarmee moeilijk te modelleren. In het geval van complexe systemen is het moeilijk te achterhalen waarom het systeem niet de gewenste uitkomsten geeft. Voordat men dus aan simuleren begint moet men zich eerst afvragen of het echt nodig is en of er misschien alternatieven bestaan.

Simuleren kan wel vaak nuttig zijn voor onderdelen van complexere processen, die bijvoorbeeld geselecteerd zijn omdat zij de belangrijkste stap in een zorgpad vormen. Het valideren is vaak eenvoudiger, en de resultaten geven meer inzicht. Door de beperkte omvang van het model is het vaak ook mogelijk het systeem op een andere wijze te analyseren dan met simulatie. Zo is het in sommige gevallen mogelijk voor de resultaten zoals gemiddelde bezettingsgraad of weigeringspercentage een formule af te leiden. Dit is onder andere het geval voor enkele zogenaamde *wachtrijmodellen*, die we nu zullen bespreken, te beginnen met het *Erlang B model.*

4.3 Capaciteitsmanagement van zorgeenheden: Het Erlang B model

Een veel voorkomend probleem in ziekenhuizen en andere zorginstellingen is het bepalen van het aantal bedden of zorgplaatsen per zorgeenheid. Dit probleem is niet zo eenvoudig als het op het eerste gezicht lijkt, vanwege fluctuaties van aantallen opnames of in zorg genomen mensen en duur van de zorg. Als gevolg daarvan kan de benodigde capaciteit om aan de volledige zorgvraag te kunnen voldoen enorm fluctueren.

Een deel van de fluctuatie in zorgvraag hangt samen met wekelijkse en jaarlijkse fluctuaties. Er zijn ook fluctuaties die in het geheel niet voorspelbaar zijn, maar volledig veroorzaakt worden door het toeval. Door deze onvoorspelbare fluctuaties ontstaat er een aanzienlijk verschil tussen de minimaal en maximaal benodigde capaciteit in bijvoorbeeld een jaar. In onderstaande grafiek staat het typische capaciteitsbeslag op een zorgeenheid gedurende een jaar, er van uitgaande dat er altijd genoeg beddencapaciteit is.



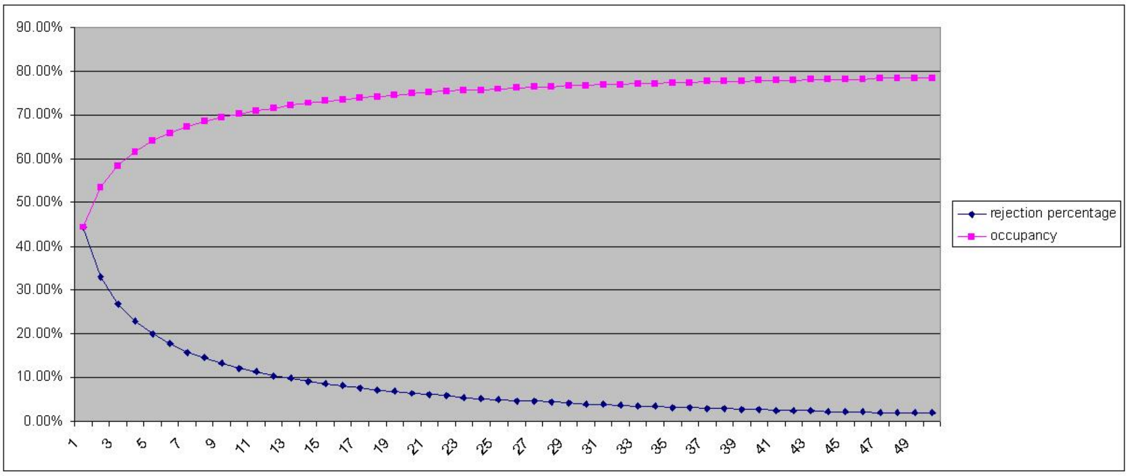
Figuur 4.1 Typische figuur met capaciteitsbeslag

Kiest een zorginstelling als aantal bedden het minimum in de grafiek, dan is de bezetting altijd 100% maar er zijn veel weigeringen (er kan niet voldaan worden aan de vraag). Kiest men voor het maximum dan is er een lage bezetting maar geen weigeringen (er kan volledig voldaan worden aan de vraag). Normaal plant een instelling het aantal beschikbare bedden / plaatsen zo, dat het tussen het minimale en maximaal benodigde aantal valt. Het resultaat daarvan is tweeledig: enerzijds blijven soms bedden leeg of plaatsen onbenut, anderzijds worden er op andere momenten patiënten/cliënten geweigerd. Om een goed onderbouwde beslissing over het aantal bedden of plaatsen te kunnen nemen, is het wenselijk het gemiddeld aantal lege bedden of plaatsen en weigeringen te kunnen berekenen. Dit is precies wat de zogenaamde *Erlang B formule* doet.

De Erlang B formule berekent op basis van zorgvraag, zorgduur en aantal bedden/plaatsen het gemiddeld aantal lege bedden/plaatsen en het gemiddeld aantal weigeringen. Merk op dat we het aantal weigeringen en daarmee ook de totale vraag vaak niet waarnemen: meestal zijn er alleen gegevens over de opnames.

Om de formule te gebruiken hoeft men zelf de formule niet door te rekenen, op internet en als Excel add-in zijn er verschillende tools beschikbaar die de uitkomsten van de formule geven. Met de in paragraaf 3 besproken POISSON functie is het zelfs mogelijk zelf een Excel calculator te maken. Zie de oefening aan het eind van de paragraaf.

In veel zorginstellingen wordt gestuurd op bezetting: bijvoorbeeld men bepaalt het aantal bedden/plaatsen zo dat men verwacht een bezetting van rond de 90% te halen. Kijkt men dan naar het percentage weigeringen, dan ziet men hevig fluctuerende getallen, ook op de wat langere termijn. Daar zijn een aantal redenen voor aan te wijzen, een daarvan is de *schaal*. In het algemeen kan men stellen dat grotere zorgeenheden, dus grotere schaal, percentueel gezien minder weigeringen hebben. In de volgende grafiek zien we wat er gebeurt als we de zorgvraag en het aantal bedden/plaatsen evenredig toe laten nemen, terwijl de gemiddelde zorgduur constant blijft. We zien duidelijk hoe het percentage weigeringen omlaag gaat en dat het bezettingspercentage omhoog gaat. Dit is een duidelijke illustratie van de schaalvoordelen. Er is ook sprake van afnemende meeropbrengsten: dit blijkt uit het feit dat de grafieken afvlakken naarmate de schaal groter is.



Figuur 4.2 Een grafiek met horizontaal het aantal bedden en verticaal weigeringspercentage en bezetting; de zorgvraag groeit met het aantal bedden mee, de gemiddelde zorgduur is constant.

Het bestaan van schaalvoordelen pleit voor grote zorgeenheden en dus voor het samenvoegen van kleinere zorgeenheden. Deze schaalvoordelen blijken echter niet altijd op te gaan, zeker niet als de gemiddelde zorgduren van mogelijk samen te voegen verpleegeenheden (zeer) verschillend zijn. Door het mengen van patiënten/cliënten met de kortste zorgduren met patiënten/cliënten met langste zorgduren kunnen de laatste de eerste wel eens in de weg gaan zitten en in totaal tot meer weigeringen leiden! De bezettingsgraad gaat wel altijd omhoog bij samenvoegen, of de zorgduren nu wel of niet verschillend zijn. De volgende tabel illustreert dit fenomeen.

	VE 1	VE 2	Gemiddeld	Samengevoegd
Zorgvraag /dag	1	5		6
Gem. zorgduur	5	1		1.67
Aantal bedden	3	7		10
Weigerings %	53%	12%	19%	21%
Bezettings %	78%	63%	67%	79%

Figuur 4.3 Getallenvoorbeeld van samenvoegen van verpleegeenheden

We zien de gegevens van twee verpleegeenheden, waarbij VE1 klein is en overbezet. VE1 heeft tevens de langste zorgduren. In de kolom "gemiddeld" staat het gemiddelde over de beide zorgeenheden, waarbij er rekening is gehouden met het feit dat VE2 meer aankomsten en meer bedden/plaatsen heeft. Als de beide VE's worden samengevoegd zien we dat het weigeringspercentage omhoog gaat van 19 naar 21%.

In de tabel lopen de zorgduren zeer uiteen; voor meer realistische getallen is het moeilijk een voorbeeld te vinden waar het overall weigerpercentage omlaag gaat door samenvoegen. Dit bezwaar blijkt dus relatief onbelangrijk te zijn. Wat belangrijker is, is het feit dat samenvoegen kan betekenen dat sommige soorten patiënten meer profiteren dan andere. Zo kan het bijvoorbeeld voorkomen dat het gemiddelde weigeringspercentage wel daalt, maar dat één patiëntengroep een hoger weigeringspercentage krijgt. Deze groep had in de oorspronkelijke situatie dus relatief veel overcapaciteit. Het kan dan ook zinvol zijn in een samengevoegde verpleegeenheid patiënten zodanig toe te laten dat één van de patiënten/cliënten een bepaalde manier van voorrang krijgt boven anderen. Op deze wijze kan elke patiënt/cliëntgroep op de gewenste wijze profiteren van het samenvoegen.

Of de Erlang B formule geschikt is om te gebruiken hangt van de situatie af. Het geeft met name een goede benadering in situaties waarin het verschil in aanbod tussen de dagen relatief klein is en waarin het aantal beschikbare bedden weinig varieert. Met name spoedopnames laten een aankomstpatroon zien dat goed overeenkomt met de Erlang formule. We merken op dat data-analyse heeft aangetoond dat in verschillende ziekenhuizen het opnamepatroon van geplande zorg grote overeenkomsten vertoont met die van spoedzorg. Deze grote variatie in de electieve stroom lijkt zelfs meer regel dan uitzondering. Met andere woorden, de Erlang B is vaak ook goed toepasbaar in het geval van een flinke electieve stroom.

Opdracht 4.2

Laat A de gemiddelde zorgvraag per dag zijn, en B de gemiddelde zorgduur. Met s geven we het aantal bedden of plaatsen aan. Dan is het weigeringspercentage in Excel als volgt te berekenen:

=POISSON(s,A*B,FALSE)/POISSON(s,A*B,TRUE)

a. Implementeer dit in Excel en vergelijk de resultaten met calculators die op internet te vinden zijn.
b. Verdiepingsvraag: Bereken ook de bezettingsgraad. Houdt er hierbij rekening mee dat de bezettingsgraad wordt berekend aan de hand van de geleverde zorg, niet aan de hand van de zorgvraag.

Opdracht 4.3

Gebruik een Excel calculator om de getallen in de tabel te reproduceren.

4.4 Modellen voor zorgprocessen met wachtrijen

In het geval van zorgeenheden worden patiënten die alle capaciteit 'bezet' aantreffen, vaak doorverwezen naar andere zorgeenheden of zorginstellingen. Slechts in enkele gevallen wordt het leveren van de zorg vertraagd; een voorbeeld is het uitstellen van een openhartoperatie als er geen IC bed beschikbaar is. In andere situaties dan zorgeenheden is er vaak sprake van vertraging van zorg indien er geen capaciteit meteen beschikbaar is; een voorbeeld is het niet direct in zorg nemen als er geen verzorgenden beschikbaar zijn. Dit leidt tot wachtrijen. Voor electieve patiënten kunnen we hierbij grofweg twee soorten van wachten onderscheiden. De eerste wachttijd is de tijd tussen het maken van een afspraak en het moment waarop de afspraak plaatsvindt. Dit wordt meestal aangeduid als toegangstijd en bevindt zich op de schaal van enkele dagen of weken. In de casus bedraagt de toegangstijd voor de heupoperatie bijvoorbeeld 9 weken. Wanneer de patiënt reeds in het systeem aanwezig is, dan is er een wachttijd tussen het moment van aankomst bij het systeem en het moment waarop de patiënt geholpen wordt. Deze wachttijd wordt gemeten in minuten. Enkele voorbeelden van deze vorm zijn wachttijden voor de SEH, röntgenafdeling, inlooppoli, bloedprikken en het anesthesiespreekuur. In de casus blijkt er bijvoorbeeld geen wachttijd te zijn voor het maken van een foto van de heup.

Toegangstijden

Wat betreft de toegangstijden zijn er twee effecten die ervoor zorgen dat wachtrijen in de zorg zich anders gedragen dan in de logistiek van bijvoorbeeld productiesystemen. Beide hangen samen met het gedrag van patiënten. Zo hebben korte wachtrijen een aanzuigende werking, waardoor de lengte van de wachtrij snel toeneemt; lange wachtrijen hebben tot gevolg dat minder patiënten/cliënten zich aanmelden. Hierdoor zal de lengte van de wachtrij de tendens hebben naar een situatie te gaan waar de instroom met de verwerkingscapaciteit in evenwicht is. Hierdoor kan over langere tijd de wachttijd even lang zijn, zelfs als de capaciteit uitgebreid wordt. Dit is funest voor zogenaamde "Werken Zonder Wachtlijst" (Advanced Access) projecten: zodra de wachtrij is weggewerkt is er vaak sprake van een dergelijke aanzuigende werking dat deze weer snel op de oude lengte terug is. Vandaar ook dat wordt geadviseerd waar mogelijk met een vaste *panel size* te werken: door de groep potentiële patiënten/cliënten te beperken voorkomt men de aanzuigende werking.

Een tweede vorm van patiënt/cliëntgedrag dat invloed heeft op de lengte van de wachtrij (toegangstijd) is afhaakgedrag: patiënten/cliënten verdwijnen uit de wachtrij, bijvoorbeeld omdat ze de zorg elders sneller kunnen krijgen, omdat de medische reden verdwenen of juist omdat ze niet zo lang kunnen wachten en daardoor prioriteit krijgen boven andere patiënten, of omdat ze overlijden. (In 2000 overleed 10-15% van de wachtenden voor opname in een verzorgingshuis/verpleeghuis tijdens het wachten.)

Wachttijden

Op diverse plaatsen in zorginstellingen ontstaan wachtrijen wanneer de hoeveelheid capaciteit (tijdelijk) niet toereikend is om aan de vraag te voldoen, zoals op de SEH of de röntgenafdeling in een ziekenhuis of voor opname in een verzorgings- of verpleeghuis. Wanneer de hoeveelheid ingezette capaciteit te laag is, kan de wachttijd een ongewenste omvang aannemen, terwijl te veel capaciteit leidt tot een slechte bezetting. De wiskundige theorie van de wachtrijen gaat terug naar het begin van de vorige eeuw, met als grondlegger de Deense wiskundige A.K. Erlang. Het bekendste wachtrijsysteem, samen met het Erlang B model, is het Erlang C model. Het verschil met de Erlang B is dat klanten die alle capaciteit bezet aantreffen in een wachtrij wachten tot er capaciteit beschikbaar is. Wanneer alle SEH verpleegkundigen bezet zijn, zal een patiënt plaats nemen in de wachtkamer. En wanneer alle appartementen in een verzorgingshuis bezet zijn, zal de toekomstige bewoner op een wachtlijst worden geplaatst.

Het Erlang C model heeft feitelijk dezelfde parameters nodig als invoer als het Erlang B model: (i) de gemiddelde zorgvraag, (ii) de gemiddelde behandelduur en (iii) het aantal eenheden beschikbare capaciteit. Essentieel hierbij is dat dezelfde tijdseenheid gebruikt wordt voor de zorgvraag en de zorgduur, bijvoorbeeld beide in uren of dagen. Op basis van deze drie parameters kan het Erlang C model belangrijke prestatie-indicatoren van het systeem berekenen, zoals de:

- bezettingsgraad
- kans op wachten
- gemiddelde wachttijd
- kans dat een patiënt/cliënt langer dan *x* tijdseenheden moet wachten voordat hij geholpen wordt

De bezettingsgraad is een belangrijke maat voor de mate waarin de capaciteit benut worden. Vanwege de wetmatigheid van Little kan deze hier direct worden bepaald door

$$bezettingsgraad = \frac{gem.zorgvraag \times gem.behandelduur}{capaciteit} \times 100\%$$

Bij het Erlang B model is het bepalen van de bezettingsgraad minder eenvoudig, omdat in dat geval patiënten/cliënten geweigerd kunnen worden en niet alle patiënten/cliënten van de capaciteit gebruik maken. De overige drie prestatie-indicatoren hebben betrekking op de verleende service naar de patiënt/cliënt en de doorstroom door het systeem. Over het algemeen wordt de gemiddelde wachttijd gebruikt om de verleende (logistieke) service uit te drukken. Voor alle genoemde wachttijd indicatoren zijn ook gesloten uitdrukkingen bekend, maar deze zijn minder eenvoudig. Op internet zijn voor ziekenhuiszorg diverse tools beschikbaar die de gewenste prestatie-indicatoren berekenen aan de hand van de drie ingegeven invoerparameters.
Zie bijvoorbeeld http://www.vumc.nl/afdelingen/pica/Software/Calc3/.

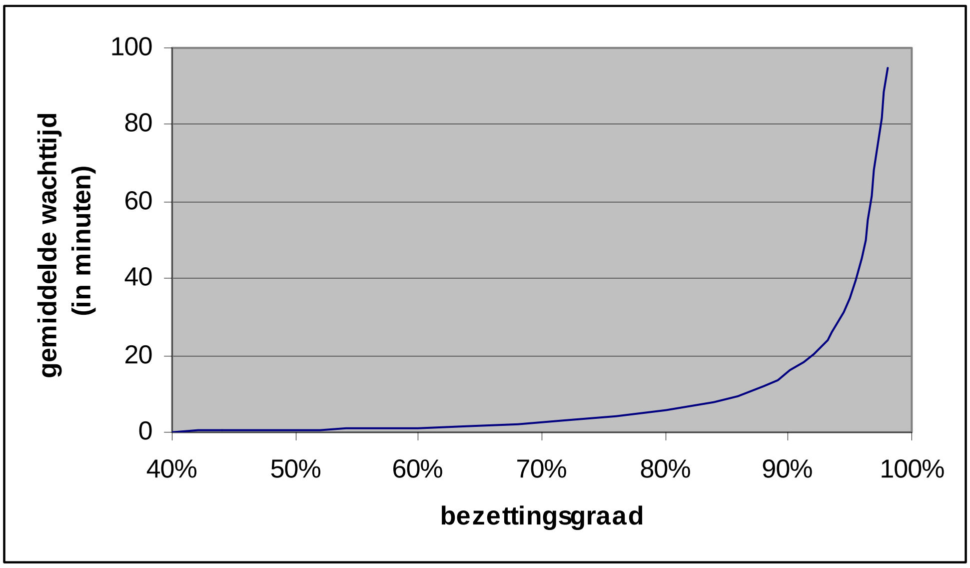
Voorbeeld

Het maken van röntgenfoto's vindt plaats op inloop. Gedurende een periode op de dag zijn 5 röntgenapparaten bemand. De behandelduur varieert uiteraard, maar op basis van historische data heeft men berekend dat de gemiddelde behandelduur 10 minuten bedraagt. Wanneer er gedurende een uur gemiddeld 24 patiënten gebruik maken van het apparaat, dan is de bezettingsgraad gelijk aan (24/60) x 10 / 5 x 100% = 80%. Het gemiddeld aantal aankomsten hebben we hier door 60 gedeeld om alles uit te drukken in minuten. Het invullen van een online tool geeft:
gemiddelde wachttijd: 5,54 minuten (ruim vijf en een halve minuut)
92,5% wacht minder dan 20 mintuten, ofwel 7,5% wacht langer dan 20 minuten

Kader 4.2 Voorbeeld wachttijd radiologie

Wanneer we het aantal aankomsten in het voorbeeld van radiologie variëren, terwijl we de behandelduur en het aantal bemande röntgenapparaten gelijk houden, dan krijgen we onderstaande grafiek. Hierin is de bezettingsgraad uitgezet tegen de gemiddelde wachttijd (in minuten). Wat direct opvalt, is dat de bezettingsgraad niet hoger kan zijn dan 100%. In dat geval is er sprake van een structureel capaciteitstekort waarbij de wachtrij alleen maar toe zal nemen en onbegrensd kan groeien. Verder heeft de grafiek een elleboog vorm, wat kenmerkend is voor dit soort wachtrijsystemen. Wanneer de bezettingsgraad toeneemt richting de 100%, dan zal de gemiddelde wachttijd snel groeien. In dat geval wordt de overcapaciteit dusdanig klein dat alle fluctuaties in vraag en aanbod niet meer goed opgevangen kunnen worden.

De exacte locatie van de elleboog in de grafiek hangt af van twee grootheden: de omvang van het systeem (ofwel de capaciteitsomvang) en de mate van variabiliteit. Naarmate de capaciteit groter is, dan zal de elleboog in de grafiek dichter bij de 100% liggen. Dit betekent dat systemen met meer capaciteit op een hogere bezettingsgraad kunnen functioneren bij een zelfde wachttijd dan kleinere systemen. Deze schaalvoordelen kwamen we eerder bij het Erlang B model al tegen. Maar ook bij systemen met een grote capaciteitsomvang brengt een hoge bezettingsgraad van dicht bij de 100% risico's met zich mee. Wanneer een periode onverwachts structureel drukker blijkt dan verwacht, dan is uit de grafiek te zien dat de gemiddelde wachttijd zeer snel toe zal nemen. Daarom zijn systemen met veel capaciteit en een hoge bezettingsgraad niet erg robuust.



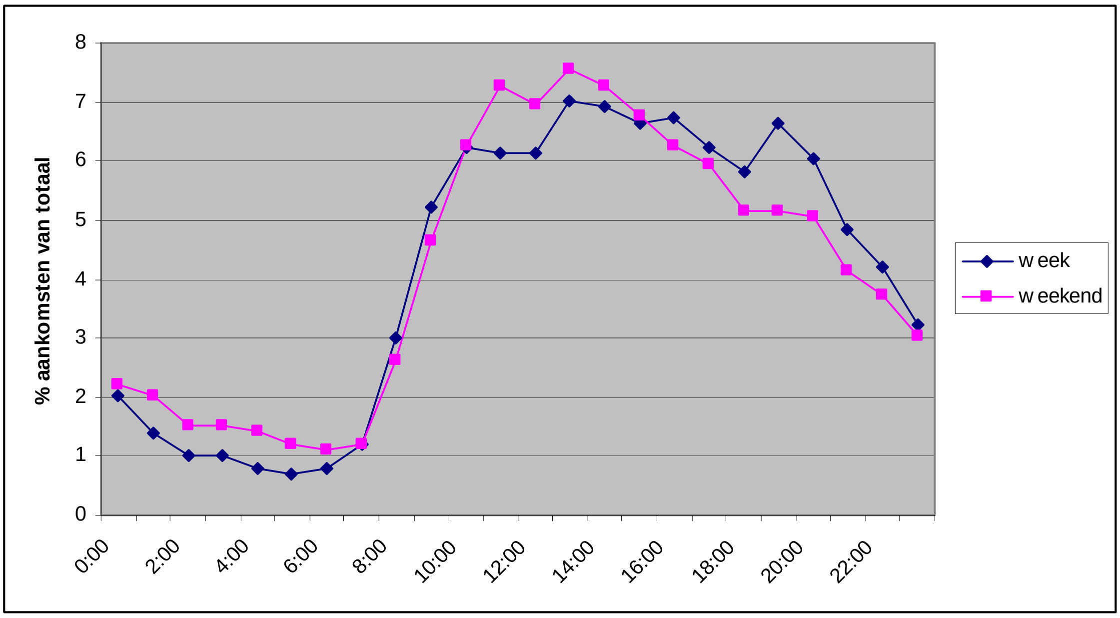
Figuur 4.4 De bezettingsgraad en gemiddelde wachttijd voor verschillende waarden van de zorgvraag; de behandelduur en de capaciteit zijn constant.

Met behulp van het Erlang C model kan inzicht worden gekregen in de gevolgen, d.w.z. bezetting en wachttijden, van een bepaalde capaciteitsinzet. Voor een goed inzicht is het uiteraard zeer nuttig om verschillende capaciteitsscenario's door te rekenen. Hiermee kan bijvoorbeeld ook de minimale capaciteit worden berekend zodanig dat er nog wel aan een bepaald service niveau wordt voldaan, zoals een maximaal gewenste gemiddelde wachttijd. Deze modellen kunnen beslissingen over de gewenste capaciteitsinzet ondersteunen door het kwantitatief inzicht dat hieruit verkregen wordt. Daarmee kan men tot een objectief onderbouwde beslissing komen.

Of het Erlang C model geschikt is hangt van de situatie af. Een essentiële aanname binnen het model betreft de variatie in het aankomstpatroon. Het model zal beter functioneren wanneer aankomsten ongepland zijn. In geval van geplande aankomsten, dan is een afsprakenmodel, zoals in de volgende paragraaf aan de orde komt, meer op zijn plaats. Verder gaat het model uit van een behandelduur waarvoor het gemiddelde en de standaarddeviatie gelijk zijn (zelfs exponentieel verdeeld). Een zorgduur met een andere mate van variatie zal een andere locatie van de kromming in de figuur te zien geven. Een laatste aanname betreft de onvoorspelbaarheid van het aankomstenpatroon. In de praktijk zal vaak sprake zijn van vooraf bekende drukkere en minder drukke perioden. Te denken valt bijvoorbeeld aan het verschil tussen weekdagen of de ochtend en de middag. Hoe we hier mee omgaan is het onderwerp van het volgende deel.

Voorspelbare fluctuaties in zorgvraag

Bij zorgprocessen zal in de praktijk meestal sprake zijn van een deels voorspelbare zorgvraag. Hierbij valt in het bijzonder te denken aan het onderscheid tussen week en weekend en een structureel dagpatroon. Van dergelijke patronen is bij vrijwel alle zorgprocessen, waarbij wachtrijen gevormd kunnen worden, sprake; bij de SEH, röntgenafdeling, inlooppoli, bloedprikken, anesthesiespreekuren, voor start zorg Thuiszorg of opname in een verpleeg- of verzorgingshuis. Een typisch voorbeeld van een dergelijk voorspelbaar zorgvraagpatroon voor de SEH is te zien in onderstaande grafiek. Hierin is het aandeel van de patiënten te zien dat in een bepaald uur van de dag bij de SEH aankomt.



Figuur 4.5 Typisch patroon in het aandeel patiënten dat in een bepaald uur van de dag aankomt.

In dergelijke situaties is een flexibele personeelsinzet gewenst. Aangezien het Erlang C model uitgaat van een constante capaciteit en een geheel onvoorspelbaar vraagpatroon zou het rechtstreeks toepassen van deze formules verkeerde conclusies opleveren. De eenvoudigste en meest inzichtelijke manier om met deze voorspelbare fluctuaties en flexibele capaciteitsinzet om te gaan is door de gehele periode, bijvoorbeeld een dag, op te knippen in kleinere intervallen waarin de grootheden die het zorglogistieke proces bepalen min of meer constant zijn. Dat wil zeggen, we zoeken perioden waarin (i) het aantal aankomsten, (ii) de behandelduur en (iii) de capaciteit geen structurele daling of stijging vertoont. In het voorbeeld van de SEH is bijvoorbeeld het opknippen van de dag in uren een geschikte optie. Bij het opdelen van de dag is het uiterst raadzaam om rekening te houden met de shifts binnen het personeelsrooster.

Wanneer de dag verdeeld is in perioden waarin de drie invoerparameters min of meer constant zijn, dan kan voor iedere periode de Erlang C formule worden toegepast. De invoerparameters kunnen dan eenvoudig worden bepaald:

- (i) het gemiddeld aantal aankomsten wordt over het gehele interval genomen (let wel op dat de tijdseenheid overeenkomt met die van de behandelduur);
- (ii) de behandelduur of zorgduur zal over het algemeen nauwelijks of niet afhangen van het tijdstip van de behandeling of in zorg nemen;
- (iii) houd bij het opdelen in perioden rekening met de shifts van het personeelsrooster.

Door het Erlang C model voor iedere periode toe te passen wordt een gewenste capaciteit per periode bepaald.

De beschreven aanpak wordt in de literatuur ook wel SIPP (stationary independent period-by-period) genoemd. Er moet wel worden opgemerkt dat dit geen exacte analyse is, maar een benadering. Dit is eenvoudig in te zien wanneer een drukke periode direct wordt opgevolgd door een rustige periode (of andersom). Het effect van de drukke periode zal doorwerken in de volgende periode omdat waarschijnlijk met een relatief lange wachtrij gestart wordt.

Wanneer de behandelduren of zorgduren niet te lang zijn in verhouding tot de lengtes van de perioden en de verschillen in invoerparameters tussen opeenvolgende perioden niet te groot, dan werkt de SIPP aanpak goed. In de context van zorgprocessen met wachtrijen is dit vaak het geval. Voor een meer gedetailleerde beschrijving van deze modellen verwijzen we graag naar de literatuur, zie bijvoorbeeld Green (2006).

Opdracht 4.4

Gebruik een online Erlang C calculator om de getallen uit het voorbeeld te reproduceren.

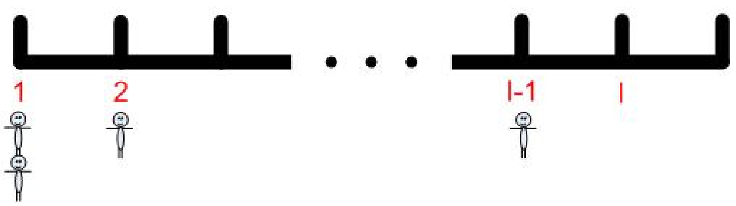
Opdracht 4.5

De doorstroom bij de poli Anesthesie loopt niet naar wens. Op dit moment worden bij deze poli slots gereserveerd per zorgketen. Dit zorgt voor veel kleine wachtrijen. Om de doorstroom te verbeteren overweegt men over te gaan naar een inlooppoli (één wachtrij). De behandelduur (hier de afspraak met de anesthesioloog) varieert erg; het gemiddelde en de standaarddeviatie van de behandeltijden liggen rond de 12 minuten. Op een donderdag bezoeken gemiddeld 8 patiënten per uur de poli Anesthesie, en er zijn twee anesthesiologen. Wat is de gemiddelde wachttijd? Hoeveel patiënten zijn er binnen een kwartier wachten aan de beurt?

4.5 Afsprakenmodellen

Een vorm van aankomstproces die specifiek voor de zorg is, is het maken van afspraken. We beschouwen hier dus niet de tijd tussen het moment van afspraak maken en de feitelijke afspraak, maar de momenten waarop de feitelijk afspraken worden gepland. Hier is ook vaak sprake van wachttijd, de zogenaamde *wachtkamertijd*. Daarnaast komt het ook voor dat degene met wie de afspraak is gemaakt moet wachten of het apparaat waarop capaciteit geboekt is (bijvoorbeeld een MRI) gedurende enige tijd ongebruikt blijft. De voor de hand liggende vraag is: op welke momenten moeten afspraken gepland worden zodat wachttijd van arts en patiënten zoveel mogelijk beperkt blijven? Complicerende factoren zijn de variabiliteit van de duur van de afspraak, de mogelijkheid dat de patiënt niet op de afspraak komt (een *no-show*) en dat de patiënt niet op tijd is. Een aanvullende factor zijn eventuele spoedpatiënten die het afsprakenschema verstoren.

Een analyse van een dergelijk afsprakensysteem kan met behulp van simulatie. In de wetenschappelijke literatuur zijn al veel afsprakenschema's op die wijze geanalyseerd. Daar komt een regel uit naar voren als een goede robuuste regel: de *Bailey-Welch rule*. Deze bestaat uit de reguliere individuele afsprakenregel (zeg 1 patiënt elke 15 minuten) met als wijziging dat de patiënt aan het eind wordt weggehaald en er aan het begin een patiënt wordt toegevoegd. Dit is geïllustreerd in onderstaande figuur: Aan het begin van de sessie zijn er twee patiënten gepland, aan het begin van de laatste slot geen.



Figuur 4.6 De Bailey-Welch regel

Als alles regelmatig verloopt is er dus altijd 1 patiënt in “voorraad”. Dit zorgt ervoor dat de arts bij onregelmatigheden niet meteen hoeft te wachten zonder dat de wachttijd van de patiënten te zeer oploopt. Men kan zich afvragen of het nog beter kan dan de Bailey-Welch rule. Dit kan met behulp van het online tool om optimale afsprakenschema’s te vinden dat beschikbaar is op http://www.vumc.nl/afdelingen/pica/Software/Calc4/. In de meeste gevallen blijkt dat Bailey-Welch aanzienlijk beter is dan de standaard individuele afsprakenregel, maar dat er verder niet veel verbeteringen meer mogelijk zijn. Dergelijke modellen zijn specifiek voor de cure-sector ontwikkeld.

Opdracht 4.6

Open de tool direct via http://obp.math.vu.nl/healthcare/software/ges/schedulecalc.php. Bereken het optimale rooster en vergelijk dit met de Bailey-Welch regel. Variëer het relatieve belang van wachtkamertijd en leegtijd (de alpha’s) en bestudeer het effect. Bespreek de resultaten met medestudenten.

4.6 Ketens

Tot nu hebben we het slechts over de analyse en planning van een enkele capaciteitseenheid gehad. Zorg bestaat vaak uit een complexe aaneenschakeling van handelingen, zelfs van verschillen zorginstellingen (ketenzorg). De traditionele manier van plannen is de volgende stap in het zorgproces te plannen op het moment dat de voorgaande is afgerond. Dit is voor de patiënt onaangenaam, hij of zij moet vaak terugkomen en de totale wachttijd is lang. Verder blijkt dat die wachttijd zich vaak concentreert op 1 of enkele stappen in de keten: de bottleneck(s). Door eerst de bottleneck te plannen en aan de hand daarvan de andere stappen, stroomlijnt men het proces en maakt men de wachttijd korter en het aantal afspraken kleiner. Dit idee van plannen vanuit de bottleneck staat ook centraal in de zogeheten Theory of Constraints, waarbij de bottleneck uiteraard de beperkende factor (constraint) is.

Voor het kwantitatief analyseren van ketens kan men simulatie gebruiken. Volgens de Theory of Constraints volstaat het vaak de analyse te beperken tot de bottleneck en de instroom ervan. Soms is het mogelijk de bottleneck op andere wijze te analyseren, bijvoorbeeld met behulp van een wachtrijmodel.

4.7 Plannen van shared resources

Veelal worden bottlenecks gedeeld door verschillende soorten patiënten of cliënten. Dit is te verklaren uit het feit dat bottleneck resources vaak duur zijn, anders was hun capaciteit geen beperking. De prijs van de capaciteit (materieel of personeel) verklaart ook de centrale locatie in de zorgorganisatie en het gebruik door verschillende specialismen. Dit veroorzaakt een lastig planningsprobleem, wat nog lastiger wordt als verschillende zorgpaden de gedeelde capaciteit gebruiken: de beperkte beschikbaarheid van de bottleneck is moeilijk te rijmen met de gewenste beschikbaarheid op vooraf voorgestelde momenten voor de verschillende soorten patiënten of zorgpaden. Dit leidt vaak tot het delen van de beschikbare tijd op de bottleneck in verschillende slots. Het gevaar is dat dit leidt tot versnippering door het niet aansluitend plannen van behandelingen op de bottleneck en daarmee tot verlies van kostbare capaciteit op de bottleneck. Ook is het moeilijk het juiste aantal slots te reserveren, omdat het op voorhand vaak niet duidelijk is hoeveel slots een zorgpad nodig heeft.

Om deze versnippering tegen te gaan is het verstandig meerdere soorten slots bijeen te voegen: hierdoor vallen er minder gaten, en de relatieve variabiliteit in aantal benodigde slots is lager waardoor er ook hierdoor minder gaten vallen.

Opdracht 4.7

Een drietal zorgpaden hebben elk op een bepaalde dag in de week gemiddeld 5 slots op de CT-scan nodig.

a. Bereken de consequenties van het apart reserveren van slots en het gezamenlijk reserveren van slots.
b. Geef enkele praktische bezwaren aan tegen het gezamenlijk reserveren van slots.

Geef een voorbeeld van een bottleneck in een care-omgeving. Welke consequenties heeft een dergelijke bottleneck voor de capaciteitsplanning?

Samenvattend schema 4.1

Modellen	**Werkwijze**
Omschrijving	Elk proces kan middels discrete-event simulatie geanalyseerd worden. Voor sommige processtappen bestaan er formules of andere berekenmethodes.
OK sessielengtes	Lengtes van OK sessies kan men simuleren of benaderen met behulp van een normale verdeling
Erlang B model	Vooral geschikt voor capaciteitsbeslissingen van verpleegeenheden Illustreert goed schaalvoordelen
Erlang C model	Model met wachtrij Houdt geen rekening met aanzuigende werking
Afsprakenmodellen	Gebruikelijke individuele planning kan verbeterd worden Laatste patiënt helemaal naar voren halen (Bailey-Welch regel) werkt vaak heel goed
Ketens	Focus op bottleneck
Shared resources	Aandacht voor versnippering bij reserveren slots

5 Literatuur

- J. Galbraith. Designing Complex Organizations. Addison-Wesley, 1973.
- L.V. Green. Queueing analysis in health care. In Patient Flow: Reducing Delay in Healthcare Delivery, Springer, 2006.
- Statistiek: Practical statistics for medical research, DG Altman, Chapman & Hall
- Erlang B & C: Numerous books on queueing theory contain the formula; on the internet Erlang B can be found on href=http://en.wikipedia.org/wiki/Erlang_(unit)#Erlang_B_formula
- S.G. Powell, K.R. Baker, and B. Lawson. Impact of errors in operational spreadsheets. Decision Support Systems, 7:126-132, 2009.

6 Geavanceerd gebruik van Excel (Bijlage A)

Voor de analyse van kleine en middelgrote datasets en het doen van veel berekeningen worden vaak spreadsheets vaak ingezet. Excel is het meest bekende en gebruikte spreadsheet en we zullen ons daartoe beperken. Deze bijlage gaat in op sommige van de meer geavanceerde aspecten van Excel. Voordat we dat doen gaan we in op een belangrijk algemeen aspect van Excel. Excel wordt namelijk gekarakteriseerd door een gebrek aan structuur. Dit gebrek aan structuur is zowel een voordeel als een nadeel. Het is een voordeel want het stelt ons in staat snel complexe berekeningen uit te voeren zonder dat we gedwongen zijn dat in een bepaald stramien te doen of eerst te programmeren. Het nadeel is dat het leidt tot spreadsheets die slecht onderhoudbaar en overdraagbaar zijn en waarin fouten moeilijk te detecteren zijn. Dit pleit ervoor geen spreadsheets te gebruiken voor structurele taken waar meerdere mensen bij betrokken zijn. Neem bijvoorbeeld het opstellen van een personeelsplanning: indien dit eenmalig moet gebeuren is het het meest efficiënt dat in Excel te doen, indien het week op week moet gebeuren is het beter een meer gestructureerde met speciale functionaliteit zoals Harmony van Ortec te gebruiken.
Een belangrijke vraag is in hoeverre spreadsheets in de praktijk fouten bevatten zonder dat men zich daarvan bewust is: dit geeft belangrijke informatie over in welke mate het verstandig is Excel te gebruiken. Onderzoek naar fouten in spreadsheets is complex en tijdrovend. Het vereist een diepgaande analyse van door anderen gemaakte spreadsheets en het is moeilijk representatieve uitspraken te doen. Toch zijn er enkele interessante studies in de wetenschappelijke literatuur. Daarin wordt het aannemelijk gemaakt dat in meer de helft van de spreadsheets met formules fouten te vinden zijn, die soms zeer vergaande consequenties kunnen hebben (zie Powell e.a. (2009)).

Excel heeft drie mogelijkheden voor het doen van geavanceerde berekeningen: Ingebouwde functies zoals COUNTIF, add-ins en tools zoals draaitabellen en data-analyse, en zelfgebouwde functies met behulp van Visual Basic for Applications (VBA). We bespreken de eerste twee, voor het gebruik van VBA is programmeerkennis nodig. Het moet opgemerkt dat door de mogelijkheden van VBA vrijwel alle mogelijkheden van een gewone programmeertaal tot de beschikking staan. Naarmate meer functionaliteit in VBA wordt geïmplementeerd kan zich natuurlijk afvragen wat de toegevoegde waarde is van Excel en of men niet beter alles in een programmeertaal kan implementeren.

In dit deel hebben we al kennis gemaakt met verschillende Excel functies. Sommige hadden enkele getallen als invoer, zoals NORMDIST; andere hadden een rij getallen (in de informatica meestal *array* genoemd) als invoer, zoals SUM en CORREL. We bepreken hier een aantal andere functies, die nuttig zijn om zelf histogrammen op te stellen of bijvoorbeeld aantallen met een bepaalde eigenschap te bepalen. Neem bijvoorbeeld de file leeftijden-duren.xls. Als we de gemiddelde verpleegduur voor 0-jarigen willen berekenen, kunnen we dit doen door gebruik te maken van de functies COUNTIF en SUMIF. In onderstaande grafiek staat hoe Excel het aantal maal dat "0" voorkomt uitrekent in de array A2:A14650. In de grafiek daaronder staat de totale verpleegduur voor 0-jarigen, waaruit de gemiddelde verpleegduur kan worden berekend. Dit kan uitgebreid worden tot een tabel met achtereenvolgens leeftijd, aantal, verpleegduur, en gemiddelde verpleegduur. Het kopiëren van cellen zorgt ervoor dat dit heel snel gaat. Een goed gebruik van de "$" in de arrays voorkomt fouten met de onder- en bovengrenzen van arrays bij het kopiëren, een veel voorkomende fout in Excel.

D1 | f_x =COUNTIF(A2:A14650,0)

	A	B	C	D	E
1	**Leeftijd**	**Verpleegduur**		980	
2	0	4			
3	0	4			
4	0	11			
5	0	5			
6	0	1			
7	0	5			
8	0	2			

Figuur 6.1 Het aantal 0-jarigen

E2 | f_x =SUMIF(A$2:A$14650,0,B2:B14650)

	A	B	C	D	E
1	**Leeftijd**	**Verpleegduur**		aantal	980
2	0	4		totale duur	4927
3	0	4		gemiddelde duur	5.027551
4	0	11			
5	0	5			
6	0	1			
7	0	5			

Figuur 6.2 De totale zorgduur van 0-jarigen en het gemiddelde

Er zijn nog twee manieren om (deels) dezelfde berekeningen te doen. Om het aantal patiënten per leeftijdsgroep te vinden kan men binnen de data-analyse add-in (onder Tools) kiezen voor histogram. Daarmee kan men zowel de histogram zelf als de onderliggende getallen verkrijgen. Een veelzijdiger mogelijkheid is het gebruik van pivot tables (draaitabellen), te vinden onder Data. Voor draaitabellen zijn prima handleidingen op internet te vinden. Draaitabellen zijn ook zeer geschikt om met data met meer dan twee attributen om te gaan.

Opdracht 6.1

Bereken, gebruikmakend van een draaitabel, per leeftijd het aantal patiënten/cliënten en de gemiddelde verpleegduur voor de gegevens in de file leeftijden-duren.xls.