

Donald van As
Jaap Klouwen

B Basisboek



kwantitatieve methoden

statistiek
met Exceltoepassingen

uitgeverij
coutinho

Basisboek kwantitatieve methoden

Statistiek

Basisboek kwantitatieve methoden

Statistiek
met Exceltoepassingen

Donald van As
Jaap Klouwen

Tweede, herziene druk

uitgeverij | C
c o u t i n h o

bussum 2024

www.coutinho.nl/statistiek

Bij dit boek hoort een website met de uitwerkingen van de opgaven uit het boek. Dit materiaal is te vinden via **www.coutinho.nl/statistiek**.

© 2007/2013 Uitgeverij Coutinho bv

Alle rechten voorbehouden.

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of op enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever. Voor zover het maken van reprografische verveelvoudigingen uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet 1912 dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan Stichting Reprorecht (www.reprorecht.nl). Voor de readerregeling kan men zich wenden tot Stichting UvO (Uitgeversorganisatie voor Onderwijslicenties, www.stichting-uvO.nl). Voor het gebruik van auteursrechtelijk beschermd materiaal in knipselkranten dient men contact op te nemen met Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, www.stichting-pro.nl).

Eerste druk 2007

Tweede, herziene druk 2013, vijfde oplage 2024

Uitgeverij Coutinho

Postbus 333

1400 AH Bussum

info@coutinho.nl

www.coutinho.nl

Zetwerk: The DocWorkers, Almere

Omslag: studio Pietje Precies bv, Hilversum

Noot van de uitgever

Wij hebben alle moeite gedaan om rechthebbenden van copyright te achterhalen. Personen of instanties die aanspraak maken op bepaalde rechten, wordt vriendelijk verzocht contact op te nemen met de uitgever.

ISBN 978 90 469 0356 8

NUR 916

Voorwoord

Voor u ligt het tweede deel van het *Basisboek kwantitatieve methoden*. Het eerste deel gaat over de basisbegrippen van de wiskunde en de financiële rekenkunde. Dit deel heeft de elementaire begrippen uit de beschrijvende en mathematische statistiek tot onderwerp: tabellen en grafieken, maatstaven, inleiding in de combinatoriek en kansrekening, de binomiale en normale kansverdeling en tot slot lineaire regressie.

Dit tweede deel kan gebruikt worden zonder uitputtende kennis van het eerste, wiskundige deel. Enige kennis van en vaardigheid in Excel is gewenst, zoals werken met eenvoudige formules, cellen kopiëren, absolute en relatieve verwijzingen gebruiken en grafieken maken.

In eerste instantie is dit boek geschreven als basis en herhaling voor studenten van (economische) hbo-opleidingen. Het kan echter ook goed dienen als 'opfriscursus' voor studies in het wo, mede omdat het boek zich leent voor zelfstudie.

Wij hebben met veel plezier aan dit boek gewerkt en hopen dat het resultaat door de gebruiker gewaardeerd wordt. We houden ons aanbevelen voor op- of aanmerkingen.

Maarn en Amersfoort, juni 2007

Donald van As

Jaap Klouwen

Bij de tweede, herziene druk

In deze editie zijn fouten uit de eerste editie verbeterd en zijn opgaven en delen van de leerstof geactualiseerd. Ook is in deze tweede editie een aantal nieuwe onderdelen te vinden. Zo wordt in hoofdstuk 1 aandacht besteed aan onderwerpen als de Paretoanalyse, Lorenzcurve en Gini-coëfficiënt. Verder is als aanvulling op de paragraaf over het prijsindexcijfer een tweetal paragrafen toegevoegd, waarin aandacht wordt besteed aan enkelvoudige en samengestelde indexcijfers (Laspeyres,

Paasche en Fisher). Daarnaast hebben we in hoofdstuk 1 en hoofdstuk 7 zowel leerzame als amusante nieuwe voorbeelden geplaatst die betrekking hebben op (lineaire) regressie.

Wij danken onze collega's Hanneke van de Velde en Rob Maas voor hun adviezen, die wij goed konden gebruiken bij het maken van deze nieuwe editie. In de productiefase zijn wij Eva van Wijk dankbaar voor de begeleiding vanuit Coutinho en waren de statistische opmerkingen bij de tekst zowel als bij de uitwerkingen van redacteur Guido Schmeits zeer waardevol.

Maarn en Amersfoort, mei 2013

Donald van As

Jaap Klouwen

Inhoud

Inleiding	11
------------------	-----------



1 Grafieken en tabellen in de beschrijvende statistiek	13
---	-----------

1.1	Inleiding	13
1.2	Dure Engelse treinreizen	15
1.3	Eten en drinken buiten de deur	16
1.4	De Big Mac als welvaartsstandaard	18
1.5	Dalende immigratie	19
1.6	Nederlandse kerncijfers	21
1.7	ABN wint, AEX verliest	22
1.8	Hebben en houden	24
1.9	Koopjes jagen	26
1.10	Nauwkeurigheid versus overzicht	27
1.11	Drankorgels	30
1.12	Het prijsindexcijfer van de gezinsconsumptie	32
1.13	Enkelvoudige indexcijfers	33
1.14	Samengestelde indexcijfers	36
1.15	Paretoanalyse	39
1.16	Lorenzcurve en Ginicoëfficiënt	42
1.17	De koopkrachtwolk van het CPB	44
1.18	Gemengde opgaven	46
	Antwoorden hoofdstuk 1	57



2 Maatstaven in de beschrijvende statistiek	67
--	-----------

2.1	Inleiding	67
2.2	Centrummaten	68
2.2.1	Centrummaten bij afzonderlijke waarnemingen	68
2.2.2	Centrummaten in Excel bij afzonderlijke waarnemingen	77
2.2.3	Centrummaten bij een klassenindeling	78
2.2.4	Centrummaten in Excel bij een klassenindeling	81

2.3	Spreidingsmaten	85
2.3.1	Spreidingsmaten bij afzonderlijke waarnemingen	86
2.3.2	Spreidingsmaten in Excel bij afzonderlijke waarnemingen	92
2.3.3	Spreidingsmaten bij een klassenindeling	93
2.3.4	Spreidingsmaten in Excel bij een klassenindeling	95
2.4	Vormmaten	95
2.5	Eigenschappen van maatstaven	99
2.5.1	Boxplot	99
2.5.2	Lineaire transformatie	100
2.5.3	2s-interval	102
2.6	Voorbeelden	103
2.7	Gemengde opgaven	110



3 Elementaire kansrekening 115

3.1	Inleiding	115
3.2	Basisprincipe van de combinatoriek	115
3.3	Permutaties	116
3.4	Combinaties	119
3.5	Permutaties en combinaties met identieke objecten	121
3.5.1	Permutaties met identieke objecten	121
3.5.2	Combinaties met identieke objecten	122
3.6	Excelfuncties voor combinatoriek	122
3.7	Kansrekening	125
3.8	Ontkeningsregel	129
3.9	Optelregel	130
3.10	Voorwaardelijke kansen	132
3.11	Onafhankelijkheid	136
3.12	Gemengde opgaven	143



4 Kansvariabelen 147

4.1	Inleiding	147
4.2	Definitie kansvariabele	147
4.3	Verwachtingswaarde en standaardafwijking van een kansvariabele	151
4.4	Combinaties van kansvariabelen	154
4.4.1	Lineaire transformaties bij kansvariabelen	154
4.4.2	Onafhankelijkheid van twee kansvariabelen	156
4.5	Gemengde opgaven	163



5 De binomiale kansverdeling

169

5.1	Inleiding	169
5.2	Twee alternatieven	169
5.3	De binomiale kansvariabele	170
5.4	De binomiale verdeling	171
5.5	De cumulatieve binomiale verdeling	172
5.6	De tabel van de cumulatieve binomiale verdeling	173
5.7	De binomiale verdeling en Excel	179
5.8	Verwachtingswaarde en standaardafwijking van de binomiale verdeling	180
5.9	Gemengde opgaven	184



6 De normale verdeling

187

6.1	Inleiding	187
6.2	De normale verdeling	187
6.3	De standaardnormale verdeling	197
6.4	Combinaties van normale verdelingen	205
6.4.1	Somvariabele met gelijke kansverdeling	205
6.4.2	Steekproefgemiddelde	207
6.4.3	Centrale limietstelling	210
6.5	Gemengde opgaven	212



7 Lineaire regressie

217

7.1	Inleiding	217
7.2	Regressielijn	217
7.2.1	Berekening van a en b met de Excelfuncties INTERCEPT en SLOPE	221
7.2.2	De trendlijnmethode in een spreidingsdiagram van Excel	222
7.3	Correlatiecoëfficiënt	225
7.3.1	Berekening van de correlatiecoëfficiënt met de Excelfunctie CORREL	226
7.3.2	Bepaling van de correlatiecoëfficiënt met behulp van de trendlijnmethode	227
7.4	Voorbeelden en toepassingen	228
7.5	Gemengde opgaven	240

Appendices	249
A1 Centrummaten en spreidingsmaten met de (grafische) rekenmachine	251
A2 Gebruikte Excelfuncties (Engels-Nederlands en Nederlands-Engels)	255
A3 Alternatieve formule voor percentielen	259
A4 Formules voor scheefheid en spitsheid	261
A5 Bewijs tweede regel voor onafhankelijkheid	263
A6 Optelregels voor varianties	265
A7 Tabellen cumulatieve binomiale verdeling voor $n = 20$ en $n = 50$	267
A8 Functievoorschrift normale verdeling	270
A9 Tabel van de standaardnormale verdeling (rechteroverschrijdingskansen)	271
A10 Andere berekeningsmethoden bij lineaire regressie	273
A11 De Ginicoëfficiënt en de oppervlakte van een trapezium	280
Literatuur	283
Register	284
Over de auteurs	288

Inleiding

In dit boek ligt in hoofdstuk 1 de nadruk op de wijze waarop tabellen en grafieken worden gebruikt om gegevens te sorteren en te expliciteren, en zo statistisch onderzoek te ondersteunen. Grafieken en tabellen komen in een zeer groot aantal verschijningsvormen voor in rapporten, krantenartikelen, websites en andere media-uitingen. Het noemen van al die vormen zou, indien al mogelijk, tot een saaie en weinig overzichtelijke opsomming leiden. Veel belangrijker vinden wij het dat de student een kritische houding ontwikkelt ten aanzien van dergelijk grafisch materiaal en de gevolgtrekkingen die naar aanleiding van het materiaal (kunnen) worden gemaakt. Om deze redenen is in hoofdstuk 1 voor een andere aanpak gekozen dan in de andere hoofdstukken, die veel gestructureerder van aard zijn. In hoofdstuk 1 wordt aan de hand van een groot aantal casussen en krantenartikelen, in 42 opgaven met uitwerkingen, onderzocht met welk doel de betreffende tabel of grafiek werd gebruikt en welk effect die zal hebben op het trekken van conclusies. Ook wordt aandacht besteed aan de doelstelling en de uitvoering van het statistische onderzoek waarvan de betreffende grafiek of tabel deel uitmaakt.

In hoofdstuk 2 worden de bekende en minder bekende maatstaven voor midden, onderlinge afwijking en vorm besproken. De voorbeelden en de opgaven moeten het begrip van deze maatstaven bevorderen. Excel is daarbij een belangrijk hulpmiddel. Er wordt gebruikgemaakt van een standaard rekenblad in Excel waarmee omslachtig rekenwerk wordt vereenvoudigd.

De hoofdstukken 3 en 4 gaan over telproblemen, elementaire kansrekening en het theoretische begrip kansvariabele. Deze hoofdstukken vormen een inleiding op de meest voorkomende kansvariabelen in de (economische) statistiek: de binomiale kansverdeling (hoofdstuk 5) en de normale kansverdeling (hoofdstuk 6). Het boek besluit met het bestuderen van mogelijk lineaire verbanden tussen twee variabelen, de lineaire regressieanalyse (hoofdstuk 7).

In elk hoofdstuk zijn – soms op meerdere plaatsen – opgaven opgenomen, waarvan de volledige uitwerkingen en eventuele opmerkingen op de website van de uitgever zijn te vinden, te bereiken via www.coutinho.nl/statistiek. Hoofdstuk 1 kent een alternatieve aanpak. Hierin zijn de uitwerkingen van de casussen en krantenartikelen achter in het hoofdstuk opgenomen. Elk hoofdstuk besluit met gemengde opgaven die soms teruggrijpen op eerdere hoofdstukken. De uitwerkingen daarvan zijn op de website te vinden. Op de website staan ook de Excelbestanden die gebruikt kunnen worden bij enkele opdrachten. We hebben in dit boek gebruikgemaakt van Excel 2007. (Voor de gebruiker die alleen over een oudere versie van Excel beschikt worden op de website ook xls-versies van deze bestanden aangeboden.)

Leidraad in dit boek is een niet al te technische aanpak, zonder echter de wiskundige en statistische inhoud te veronachtzamen. Het streven is om na een aantal, meestal economische, voorbeelden tot een algemeen geldende formule te komen. Exceltoepassingen spelen daarbij een belangrijke rol, hetgeen tot een andere didactische aanpak heeft geleid dan tot nu toe gebruikelijk was. Er is echter ook aandacht voor berekeningen op de twee meest voorkomende grafische rekenmachines. De meeste statistische toepassingen zijn terug te vinden in de accountancy, logistiek, bedrijfseconomie en commerciële economie. Soms zijn technische of verdiepende delen van de stof verplaatst naar een appendix.

1 Grafieken en tabellen in de beschrijvende statistiek



1.1 Inleiding

Uitdrukkingen als 'uit de statistieken is gebleken dat ...' en 'statistisch onderzoek heeft uitgewezen dat ...' worden in het dagelijks leven veelvuldig gebruikt. Als het goed is, bevindt de gebruiker van dergelijke uitdrukkingen zich in de fase van het trekken van een of meerdere conclusies of het bewijzen van een stelling of uitspraak.

De statistische weg die tot een verantwoorde conclusie dan wel een betrouwbare voorspelling leidt, kan in een aantal fasen worden opgedeeld:

- 1 formuleren van de doelstelling(en) van het onderzoek;
- 2 verzamelen van de gegevens (met een *steekproef* of door onderzoek van de gehele *populatie*);
- 3 sorteren en expliciteren van de gegevens (tabellen, grafieken);
- 4 analyseren van de gegevens;
- 5 trekken van conclusies en/of doen van voorspellingen.

Belangrijke aandachtspunten bij statistisch onderzoek kunnen als volgt worden geformuleerd:

- Het doel van het onderzoek dient scherp omlijnd te zijn.
- De te onderzoeken *populatie* moet zorgvuldig omschreven zijn.
- Het *aselecte* karakter en de representativiteit van de uit te voeren steekproef (*representatieve steekproef*) moeten gewaarborgd zijn.
- In voorkomende gevallen dient een *gestratificeerde steekproef* (ook: *gelaagde steekproef*) te worden getrokken: dit is een steekproef die uit deelsteekproeven bestaat die ieder uit een deel van de populatie worden getrokken.
- De wijze waarop de benodigde gegevens zullen worden verkregen, moet duidelijk zijn; daarbij kan worden gedacht aan bijvoorbeeld een rechtstreekse *enquête*, maar ook aan het gebruik van resultaten uit eerder onderzoek (*primaire* respectievelijk *secundaire statistiek*).



1 GRAFIEKEN EN TABELLEN IN DE BESCHRIJVENDE STATISTIEK

- De *steekproefomvang* dient vermeld te worden; bij verwijzingen naar statistische onderzoeksresultaten wordt in de praktijk lang niet altijd de steekproefomvang van dat onderzoek vermeld.
- De *non-respons* dient vermeld te worden: het is van belang om te weten hoe groot het percentage *respondenten* is van het totaal aantal ondervraagden; nog mooier is als duidelijk wordt welke motieven er achter de non-respons schuilen (geen tijd, geen zin, geen interesse et cetera).
- Het ordenen van de uitkomsten, door ze in een *tabel* of een *datamatrix* onder te brengen, en het uitbeelden ervan door middel van een *grafiek* of een *diagram* dient zorgvuldig te gebeuren; alleen dan is het mogelijk om een verantwoord overzicht te creëren van die uitkomsten.

Nadat de onderzoeksresultaten zijn gepresenteerd, kunnen ze geanalyseerd en *geïnterpreteerd* worden. Er moet, indien van toepassing, worden nagegaan of:

- de *statistische variabelen* structureel toe- of afnemen in de tijd;
- waargenomen veranderingen zich volgens een bepaald patroon voltrekken;
- er sprake kan zijn van een zeker verband tussen twee of meerdere variabelen;
- er *uitbijters* (of: *uitschieters*) onder de uitkomsten zijn (uitkomsten die extreem afwijken van het overgrote deel van de overige uitkomsten); er moet onderzocht worden waardoor de uitbijters zijn veroorzaakt en daarna moet besloten worden of ze in de conclusie of voorspelling worden meegewogen of niet.

De laatste stap is het trekken van een of meerdere conclusies respectievelijk het doen van voorspellingen die voortvloeien uit de geanalyseerde gegevens.

In dit hoofdstuk ligt de nadruk op de wijze waarop de verzamelde gegevens (kunnen) worden gerangschikt en/of uitgebeeld. Anders gezegd: hoe beschrijvende statistiek wordt ingezet in de diverse verschijningsvormen van communicatie.

Daarnaast is er aandacht voor de samenstelling van de betreffende populatie en de steekproef. Een enkele keer wordt ingegaan op de analyse van de onderzoeksgegevens en de mogelijke conclusies of voorspellingen die daaruit kunnen voortvloeien, maar dit onderwerp komt in latere hoofdstukken nadrukkelijker aan de orde. Aan de hand van casussen zal een groot aantal begrippen uit de beschrijvende statistiek behandeld worden. Ten slotte komen ook Exceltoepassingen aan bod.

1.2 Dure Engelse treinreizen

In de *NRC* van 6 januari 2007 stond de volgende tabel bij een artikel over de hoge kosten van het reizen per trein in Engeland.

Tabel 1.1 Dure Engelse treinreizen

Britse treinen duurste van Europa		
Prijsverschillen in Europa, standaardtarief enkele reis tweede klas:		
<i>Traject</i>	<i>Afstand</i>	<i>Tarief in euro</i>
Londen-Manchester	321 km	162,00
Londen-Newcastle	464 km	165,00
Londen-Birmingham	195 km	79,00
Amsterdam-Groningen	217 km	45,80
Madrid-Barcelona	633 km	63,60
Milaan-Rome	629 km	51,00
Berlijn-Hamburg	291 km	62,00
Berlijn-Dresden	191 km	34,00
Parijs-Calais	274 km	59,40

NRC Handelsblad 050107 / BG / Bron: Spoorwegmaatschappijen

Uit de tabel valt het volgende af te leiden:

- Het doel van het onderzoek is nagaan of de kosten van het reizen per trein in Engeland hoger zijn dan die in enkele andere Europese landen.
- Er is gekozen voor een steekproef uit de treintrajecten in Engeland, Frankrijk, Duitsland, Spanje, Italië en Nederland, met de bijbehorende afstanden en tarieven; men heeft zich beperkt tot enkele reizen, tweede klas; zie ook opgave 1.1.
- De uitkomsten zijn in een tabel gezet, waarin het traject, de bijbehorende afstand in kilometers en de prijs in euro's zijn vermeld.
- De analyse: zie hiervoor opgave 1.2.
- De conclusie: treinreizen in Engeland zijn aanmerkelijk duurder dan treinreizen in de overige onderzochte landen; zie hiervoor opgave 1.3.

OPGAVEN

- 1.1 Is er sprake van een representatieve steekproef? Zijn er redenen om te veronderstellen dat een grotere steekproef had moeten worden genomen?



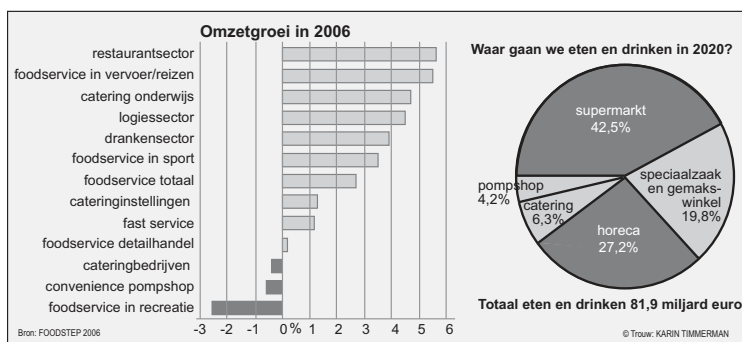
1 GRAFIEKEN EN TABELLEN IN DE BESCHRIJVENDE STATISTIEK

- 1.2 Is op grond van de gegevens uit de tabel een duidelijke analyse te maken? Is er aanleiding om (afgeleide) gegevens toe te voegen om een helderder beeld te krijgen?
- 1.3 Zijn er, behalve de prijs, andere criteria te bedenken die de conclusie beïnvloeden?
- 1.4 Maak een duidelijker overzicht, sorteer en analyseer de verkregen resultaten en trek vervolgens, indien mogelijk, een duidelijke conclusie.

1.3 Eten en drinken buiten de deur

In het dagblad *Trouw* van 11 januari 2007 stond een artikel geflankeerd door een tweetal grafieken (zie figuur 1.1). De vermelde gegevens zijn afkomstig van een onderzoek dat in opdracht van het FSIN (FoodService Instituut Nederland) werd uitgevoerd door de organisatie Foodstep met als doel om een overzicht te krijgen van de ontwikkelingen in de foodsector, zowel qua omzetten als eetgewoonten. Ook is blijkbaar een extrapolatie uitgevoerd naar het jaar 2020 met betrekking tot de dan verwachte marktaandelen van de consumptie van voedsel en drank.

Het verzamelen van de gegevens zal plaatsgevonden hebben via continu onderzoek binnen de gehele foodsector. De grafieken laten een staafdiagram zien met *liggende staven* en daarnaast een *cirkeldiagram*. Het staafdiagram toont de percentages omzetgroei over het jaar 2006 in de diverse categorieën van de foodservice, het cirkeldiagram de verwachte marktaandelen in 2020. Uit de tekst blijkt dat het onderzoek nog veel meer resultaten heeft opgeleverd dan in de grafieken te zien is.



Figuur 1.1 Eten en drinken buiten de deur

Surfen naar www.fsin.nl op 7 mei 2007 leverde onder andere de onderstaande extra informatie op met betrekking tot bovengenoemd krantenartikel. Tevens was vermeld dat de consumptie buitenshuis in 2006 ongeveer 17,5 miljard euro bedroeg.

Tabel 1.2 De markt voor buitenshuis eten en drinken

FOODSERVICE SEGMENTEN		
	2005	2006
Catering-segmenten	3076	3095
<i>On the Move-segmenten</i>	2913	2922
Sport en recreatie	1410	1428
Fastservicesector	2200	2226
Drankensector*	2858	2970
Restaurantsector*	4274	4514
Logiessector	1437	1501
Foodservice Totaal	17.020	17.481

* Deze cijfers zijn uitgezonderd de cijfers die ook bij recreatie, detailhandel en vervoer staan.

Bron: Foodstep 2006

Berekening: FSIN

OPGAVEN

- 1.5 Ga, waar mogelijk, door berekening na of de bedragen die in tabel 1.2 worden genoemd, overeenkomen met hetgeen uit het staafdiagram van figuur 1.1 af te lezen is. (Tip: breng de tabel over op een Excelwerkblad en voeg een kolom toe waarin de procentuele verandering wordt berekend.)
- 1.6 Stel, indien mogelijk, met behulp van de gegevens uit het krantenartikel en de afbeeldingen vast in welke geldeenheden de getallen uit tabel 1.2 zijn uitgedrukt.
- 1.7 Bereken met welk percentage de omzet uit de sector 'Catering' naar schatting in 2020 gestegen zal zijn in vergelijking met 2006.



- 1.8 Bereken met welk percentage het marktaandeel van de cateringsegmenten tot 2020 zal toe- of afnemen in vergelijking met 2006. Ga er hierbij van uit dat de totalen uit de tabel en die uit het cirkeldiagram de gehele markt omvatten.

1.4 De Big Mac als welvaartsstandaard

In het driemaal per jaar verschijnende blad *Wegwijs* (woningmarkt, hypotheek, beleggingen, verzekeringen, pensioenen) van november 2006, nr. 43, stond een artikel over de welvaart in een elftal landen, in beeld gebracht door een *staafdiagram* (zie figuur 1.2). Het criterium dat is genomen voor de mate van welvaart is het aantal minuten dat op basis van het gemiddelde salarisoniveau in een grote stad uit het betreffende land moet worden gewerkt om een Big Mac te kunnen kopen.

Big Mac als welvaartsstandaard

Volgens topbestuurders in de Amerikaanse financiële wereld is het welvaartspeil van een land goed te bepalen aan de hand van de prijs van de Big Mac van McDonald's. Op basis van het gemiddelde salarisoniveau van een land en de tijd die men nodig heeft om dit salaris te verdienen, heeft men op een grafiek uitgewerkt hoelang men in een land moet werken om een Big Mac te kunnen kopen. De plaats op de grafiek bepaalt, volgens deze bestuurders, het welvaartsniveau van een land. Volgens de grafiek heeft men in Japan al in 9 minuten een Big Mac verdiend. In Indonesië doet men daar gemiddeld 80 minuten over. Een Nederlander heeft de Big Mac na 18,5 minuten te pakken, terwijl Parijzenaars op dat moment nog zitten te 'kwijlen'. Verstandige Fransen hebben op dat moment de iets goedkopere Franse frietjes al verorberd.



Figuur 1.2 Dik kwartier werken is genoeg

- De doelstelling van het onderzoek is duidelijk: tot een overzicht komen van de verschillen in welvaart tussen de diverse landen.

- *Wegwijs* meldt: 'Volgens topbestuurders in de Amerikaanse financiële wereld is het welvaartspeil van een land goed te bepalen aan de hand van de prijs van een Big Mac van McDonald's.' Er is dus gekozen voor het verzamelen van gegevens via een steekproef onder restaurants van de genoemde hamburgerketen in een grote stad (meestal de hoofdstad) in het betreffende land; zie opgave 1.9.
- De prijs van de Big Mac is omgerekend naar de tijd die een inwoner met een gemiddeld inkomen moet werken om het broodje hamburger bij elkaar te verdienen. De uitkomsten van het onderzoek zijn verwerkt in een staafdiagram met *horizontale staven*, waarbij de keuze is gemaakt om de betreffende aantallen minuten van laag naar hoog te rangschikken; zie opgave 1.10.
- Analyse van de gegevens zou kunnen leiden tot een aantal voorzichtige conclusies, zoals:
 - 1 In de onderzochte West-Europese landen ontlopen de niveaus van welvaart elkaar weinig.
 - 2 De VS doen het iets beter dan West-Europa.
 - 3 Japan heeft van de onderzochte landen de hoogste welvaart.
 - 4 Met Indonesië is het vooralsnog slecht gesteld.
 - 5 Oost-Europa blijft ook nog flink achter in welvaart.
 De vraag is echter of uit de genoemde gegevens wel duidelijke conclusies vallen te trekken; zie opgave 1.11.

OPGAVEN

- 1.9 Zijn er redenen aan te voeren om te twifelen aan de representativiteit van de steekproef?
- 1.10 Wat zal als maatstaf zijn genomen: bruto- of netto-inkomens?
- 1.11 Zijn er wel eenduidige conclusies te trekken over de verschillende welvaartsniveaus?

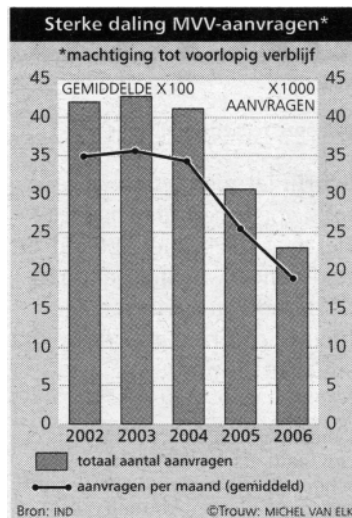
1.5 Dalende immigratie

In het dagblad *Trouw* van 19 januari 2007 stond een artikel dat de dalende immigratie in Nederland tot onderwerp had. Bij dit artikel werd een overzicht gebruikt van het aantal aanvragen van een machtiging tot voorlopig verblijf (mvv) in de periode 2002 tot en met 2006. In onderstaande grafiek is zowel een staafdiagram (met *verticale staven*) als een



1 GRAFIEKEN EN TABELLEN IN DE BESCHRIJVENDE STATISTIEK

lijndiagram te zien, waarbij twee verschillende *schaalverdelingen* langs de verticale assen zijn af te lezen.



Figuur 1.3 Sterke daling mvv-aanvragen (machtiging tot voorlopig verblijf)

- De doelstelling van het onderzoek is al aangegeven: nagaan hoe het aantal aanvragen van een mvv in de afgelopen jaren is verlopen.
- Bij de verzamelde gegevens wordt onderscheid gemaakt tussen het totaal aantal aanvragen per jaar (in duizendtallen) en het gemiddelde aantal aanvragen per maand gedurende datzelfde jaar (in honderdtallen); de gegevens zijn afkomstig van de IND (Immigratie- en Naturalisatiedienst), waardoor aangenomen mag worden dat dit een volledig overzicht in absolute aantallen is (en geen steekproef); zie opgaven 1.12 tot en met 1.15.
- De analyse en conclusies spreken hier voor zichzelf.

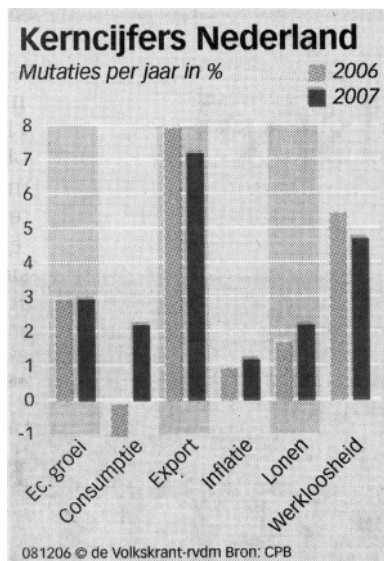
OPGAVEN

- 1.12 Welke as hoort bij de jaarlijkse aantallen aanvragen?
- 1.13 Waarom zijn de waarden op het lijndiagram (de dikke stippen) in het midden van de staven geplaatst?
- 1.14 Welk verband bestaat er tussen de waarden die door de staven worden aangegeven en de waarden die af te lezen zijn op het lijndiagram?

1.15 Welke redenen kunnen er geweest zijn om beide gegevens in de grafiek af te beelden?

1.6 Nederlandse kerncijfers

In de *Volkskrant* van 8 december 2006 stond een artikel met als kop 'Hoge looneisen dreigen inflatie op te drijven'. Ter ondersteuning van het artikel was de grafiek van figuur 1.4 geplaatst. Er is sprake van een staafdiagram, waarin de mutaties per jaar te zien zijn van een zestal economische indicatoren over de jaren 2006 en 2007. Duidelijk zal zijn dat de cijfers over 2007 berusten op schattingen. De cijfers zijn afkomstig van het *Centraal Bureau voor de Statistiek (CBS)*, het belangrijkste instituut op het gebied van statistisch onderzoek in Nederland.



Figuur 1.4 Kerncijfers Nederland

- Dit type onderzoek is van doorlopende aard: het CBS houdt van een groot aantal economische indicatoren de ontwikkeling continu bij. Het doel is uiteraard om goed in kaart te krijgen hoe het met de Nederlandse economie is gesteld.
- De gegevens worden doorlopend opgevraagd bij een groot aantal bedrijven en overheden; via complexe computermodellen komen dan de indicatoren tot stand, waarvan de zes in de grafiek een voorbeeld zijn.



1 GRAFIEKEN EN TABELLEN IN DE BESCHRIJVENDE STATISTIEK

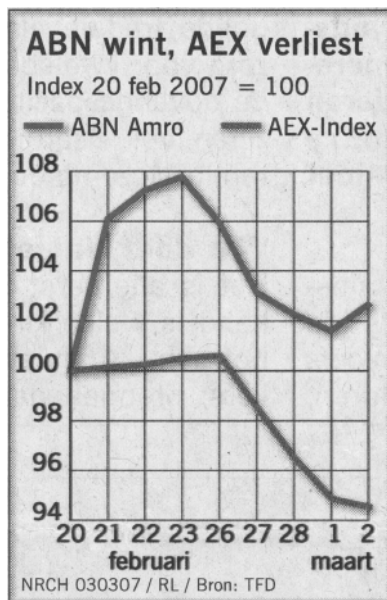
- In de grafiek zijn de mutaties af te lezen in procenten per jaar; zie opgave 1.16.
- Uit de grafiek is het nodige af te leiden; zie opgaven 1.17 tot en met 1.19.

OPGAVEN

- 1.16 Leg uit wat de betekenis is van de staaf bij 'Consumptie' in 2006.
- 1.17 Welke conclusie kan worden getrokken uit de gegevens met betrekking tot de export in 2006 en 2007?
- 1.18 Maak een berekening om aan te geven met welk percentage de werkloosheid in 2007 naar schatting zal zijn veranderd in vergelijking met 2005.
- 1.19 Ondersteunt de grafiek de dreiging die in de aanhef van het artikel wordt genoemd?

1.7 ABN wint, AEX verliest

In de *NRC* van 3 maart 2007 stond een artikel over de koersontwikkeling van ABN AMRO na de gebleken interesse van beleggingsfonds TCI (een zogenaamd *hedgefund*) voor ABN. Bijgaande grafiek (zie figuur 1.5) geeft informatie over de koersontwikkeling van ABN in relatie tot die van de AEX over de periode 20 februari tot en met 2 maart. Er is sprake van een lijndiagram, maar in dit geval worden niet de absolute gegevens van ABN en de AEX getoond, maar zijn die gegevens verwerkt met behulp van *enkelvoudige indexcijfers* met als *basis* 20 februari 2007. Op 20 februari stond de AEX op 508,02 punten en het aandeel ABN deed € 25,92.



Figuur 1.5 ABN wint, AEX verliest

- Het doel van het onderzoek is uiteraard nagaan hoe de koers van ABN AMRO zich heeft ontwikkeld in relatie tot de AEX.
- Het verzamelen van de gegevens is zeer eenvoudig, omdat alle benodigde informatie beschikbaar is. Het is alleen de vraag geweest over welke periode de gegevens moesten worden verzameld en welke gegevens precies (openingskoersen, hoogste koers, slotkoers et cetera).
- De analyse wordt door de compacte grafiek aanzienlijk vergemakkelijkt: met absolute getallen in plaats van indexcijfers had bijvoorbeeld gekozen moeten worden voor twee assen met verschillende schaalverdelingen; zie opgave 1.20.
- De conclusie: er lijkt een duidelijke aanwijzing te zijn dat de interesse van het *hedgfund* voor ABN een opwaartse druk op de koers van ABN veroorzaakt.

OPGAVEN

- 1.20 In tabel 1.3 zijn de koersen van ABN AMRO en de AEX gegeven in de periode 20 februari 2007 tot en met 2 maart 2007 (bron: www.behr.nl, 18 maart 2007).



1 GRAFIEKEN EN TABELLEN IN DE BESCHRIJVENDE STATISTIEK

Tabel 1.3

Datum	Koers ABN AMRO	Indexcijfer	Koers AEX	Indexcijfer
20/02/07	25,92	100	508,02	100
21/02/07	27,50	106,1	508,78	
22/02/07	27,80		509,33	
23/02/07	27,94		510,61	
26/02/07	27,47		511,17	
27/02/07	26,74		500,33	
28/02/07	26,51		490,21	
01/03/07	26,34		481,96	94,9
02/03/07	27,02		480,20	

Het indexcijfer voor de koers van ABN AMRO op 21 februari 2007 met als basis 20 februari 2007 wordt als volgt berekend:

$$\frac{27,50}{25,92} \times 100 = 106,1 \text{ (afgerond op één decimaal)}$$

Het indexcijfer voor de AEX van 1 maart 2007 met als basis 20 februari 2007 wordt dus als volgt berekend:

$$\frac{481,96}{508,02} \times 100 = 94,9 \text{ (afgerond op één decimaal)}$$

Bereken alle overige indexcijfers op één decimaal nauwkeurig en vergelijk deze met de waarden die uit de grafiek zijn af te lezen.

- 1.21 Neem de tabel uit opgave 1.20 over in Excel, voeg een nieuwe kolom in en bereken met behulp van een rekenoperatie in Excel de indexcijfers die ontstaan als de basis wordt verlegd naar 2 maart 2007 (één decimaal nauwkeurig).

1.8 Hebben en houden

In *Het Financieele Dagblad* van 12 april 2007 was een artikel gewijd aan de vraag hoe werkgevers het probleem van een groeiend personeelsverloop en moeilijk te vervullen vacatures het hoofd zouden kunnen bieden. De resultaten van een onderzoek door verzekeraar Centraal Beheer zijn in het volgende overzicht terug te vinden: