

STATISTIEK IN DE PRAKTIJK

Opgavenboek

David S. Moore
George P. McCabe

5e herziene druk



Oorspronkelijke titel: *Introduction to the Practice of Statistics, fifth edition.*

First published in the United States by W.H. Freeman and Company, New York and Basingstoke
Copyright © 2005 by W.H. Freeman and Company. All rights reserved.

Meer informatie over deze en andere uitgaven kunt u verkrijgen bij:

Sdu Klantenservice
Postbus 20014
2500 EA Den Haag
Telefoon (070) 37 89 880
www.sdu.nl/service

1e druk 1994
2e druk 1997
3e druk 2001
5e druk, 1e oplage februari 2006
2e oplage september 2007

Copyright Nederlandse vertaling © 2006–2008 Sdu Uitgevers bv, Den Haag
Academic Service is een imprint van Sdu Uitgevers bv.

Vertaling: Vertaalbureau Transvorm, Carola Bouman, Josefien Bruijn, Sietske Tol, Theo Tromp,
Marc Wiersma

Omslagontwerp: Scherphuis | Snijder BNO
Omslagillustratie: Corbis
Zetwerk: Elvenkind, Dordrecht
Druk- en bindwerk: De Groot Drukkerij, Goudriaan

ISBN 90 395 2361 4
NUR 123 / 916

Alle rechten voorbehouden. Alle auteursrechten en databankrechten ten aanzien van deze uitgave worden uitdrukkelijk voorbehouden. Deze rechten berusten bij Sdu Uitgevers bv.

Behoudens de in of krachtens de Auteurswet 1912 gestelde uitzonderingen, mag niets uit deze uitgave worden veelevoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voorzover het maken van reprografische veelevoudigingen uit deze uitgave is toegestaan op grond van artikel 16 h Auteurswet 1912, dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (postbus 3051, 2130 KB Hoofddorp, www.reprorecht.nl). Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16 Auteurswet 1912) dient men zich te wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro). Voor het overnemen van een gedeelte van deze uitgave ten behoeve van commerciële doeleinden dient men zich te wenden tot de uitgever.

Hoewel aan de totstandkoming van deze uitgave de uiterste zorg is besteed, kan voor de afwezigheid van eventuele (druk)fouten en onvolledigheden niet worden ingestaan en aanvaarden de auteur(s), redacteur(en) en uitgever deswege geen aansprakelijkheid voor de gevolgen van eventueel voorkomende fouten en onvolledigheden.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the publisher's prior consent.

While every effort has been made to ensure the reliability of the information presented in this publication, Sdu Uitgevers neither guarantees the accuracy of the data contained herein nor accepts responsibility for errors or omissions or their consequences.

Inhoud

1	Kijken naar gegevens – verdelingen	1
	Opgaven bij paragraaf 1.1	1
	Opgaven bij paragraaf 1.2	16
	Opgaven bij Paragraaf 1.3	24
	Opgaven bij hoofdstuk 1	35
2	Kijken naar gegevens – relaties	43
	Opgaven bij paragraaf 2.1	43
	Opgaven bij paragraaf 2.2	53
	Opgaven bij paragraaf 2.3	58
	Opgaven bij paragraaf 2.4	65
	Opgaven bij paragraaf 2.5	75
	Opgaven bij hoofdstuk 2	78
3	Gegevens verwerven	87
	Opgaven bij paragraaf 3.1	87
	Opgaven bij paragraaf 3.2	88
	Opgaven bij paragraaf 3.3	97
	Opgaven bij paragraaf 3.4	103
	Opgaven bij hoofdstuk 3	109
4	Kansrekening: de studie van het toeval	115
	Opgaven bij paragraaf 4.1	115
	Opgaven bij paragraaf 4.2	117
	Opgaven bij paragraaf 4.3	123
	Opgaven bij paragraaf 4.4	128
	Opgaven bij paragraaf 4.5	135
	Opgaven bij hoofdstuk 4	140
5	Steekproefverdelingen	145
	Opgaven bij paragraaf 5.1	145
	Opgaven bij paragraaf 5.2	152
	Opgaven bij hoofdstuk 5	159
6	Inleiding tot inferentie	163
	Opgaven bij paragraaf 6.1	163
	Opgaven bij paragraaf 6.2	168
	Opgaven bij paragraaf 6.3	176
	Opgaven bij paragraaf 6.4	179
	Opgaven bij hoofdstuk 6	181

7	Inferentie voor verdelingen	189
	Opgaven bij paragraaf 7.1	189
	Opgaven bij paragraaf 7.2	204
	Opgaven bij paragraaf 7.3	216
	Opgaven bij hoofdstuk 7	219
8	Inferentie voor fracties	231
	Opgaven bij paragraaf 8.1	231
	Opgaven bij paragraaf 8.2	236
	Opgaven bij hoofdstuk 8	241
9	Analyse van kruistabellen	251
	Opgaven bij hoofdstuk 9	251
10	Inferentie voor regressie	273
	Opgaven bij hoofdstuk 10	273
11	Meervoudige lineaire regressie	289
	Opgaven bij hoofdstuk 11	289
12	Eén-factor variantie-analyse	305
	Opgaven bij hoofdstuk 12	305
13	Twee-factor variantie-analyse	325
	Opgaven bij hoofdstuk 13	325
14	Bootstrap methoden en permutatietoetsen	339
	Opgaven bij paragraaf 14.1	339
	Opgaven bij paragraaf 14.2	340
	Opgaven bij paragraaf 14.3	344
	Opgaven bij paragraaf 14.4	345
	Opgaven bij paragraaf 14.5	350
	Opgaven bij hoofdstuk 14	357
15	Niet-parametrische toetsen	363
	Opgaven bij paragraaf 15.1	363
	Opgaven bij paragraaf 15.2	367
	Opgaven bij paragraaf 15.3	371
	Opgaven bij hoofdstuk 15	375
16	Logistische regressie	379
	Opgaven bij hoofdstuk 16	379

17 Statistiek in de kwaliteitszorg: stabiliteit en capaciteit	389
Opgaven bij paragraaf 17.1	389
Opgaven bij paragraaf 17.2	395
Opgaven bij paragraaf 17.3	401
Opgaven bij paragraaf 17.4	405
Opgaven bij hoofdstuk 17	408
 Bijlage	 413
 Tabellen	 423
Tabel A Standaardnormale kansen	424
Tabel B Toevalsgetallen	426
Tabel C Binomiale kansen	428
Tabel D Kritieke waarden voor de t -verdeling	433
Tabel E Kritieke waarden voor de F -verdeling	434
Tabel F Kritieke waarden voor de χ^2 -verdeling	442
 Antwoorden bij de oneven opgaven	 443
 Woordenlijst Engels–Nederlands	 489

1 Kijken naar gegevens – verdelingen

Opgaven bij paragraaf 1.1

- 1.1 Een klein deel van een gegevensverzameling beschrijft het brandstofverbruik (in miles per gallon) van auto's van model 2004:

Merk en model	Voertuigtype	Versnellingsbaktype	Aantal cilinders	Stad MPG	Snelweg MPG
Acura NSX	Tweezitter	Automatisch	6	17	24
BMW 3301	Compact	Hand	6	20	30
Cadillac Seville	Middenklasse	Automatisch	8	18	26
FordF150 2WD	Standard pickup	Automatisch	6	16	19

- (a) Wat zijn de elementen in deze gegevensverzameling?
(b) Welke variabelen zijn er per element? Welke variabelen zijn kwalitatief en welke kwantitatief?
- 1.2 De eerste regels van een gegevensverzameling van een docent aan het eind van een cursus statistiek:

NAME	MAJOR	POINTS	GRADE
ADVANI, SURA	COMM	397	B
BARTON, DAVID	HIST	323	C
BROWN, ANNETTE	BIOL	446	A
CHIU, SUN	PSYC	405	B
CORTEZ, MARIA	PSYC	461	A

Wat zijn de elementen en de variabelen in de gegevens? Welke variabelen zijn kwalitatief en welke kwantitatief?

- 1.3 Hoe kunnen we houten oppervlakken weerbestendig maken, in het bijzonder wanneer we historische huizen restaureren? In een onderzoek werden houten panelen geprepareerd en vervolgens aan het weer blootgesteld. Hier volgt een aantal variabelen. Welke variabelen zijn kwalitatief en welke kwantitatief?
- (a) Houtsoort (populier, grenen, cederhout)
(b) Waterafstotend (met oplosmiddel, op waterbasis)
(c) Dikte van de verf (mm)

- (d) Verfkleur (wit, grijs, lichtblauw)
 (e) Verwerkingstijd (maanden)

- 1.4** U bereidt een onderzoek voor naar het kijkgedrag van studenten. Beschrijf twee kwalitatieve variabelen en twee kwantitatieve variabelen die u per student meet. Geef de meeteenheden aan voor de kwantitatieve variabelen.
- 1.5** U wilt de 'lichamelijke conditie' van studenten meten. Beschrijf verschillende variabelen die u gaat gebruiken om de conditie te meten. Wat voor instrument of instrumenten heeft u voor elke meting nodig?
- 1.6** Populaire tijdschriften rangschikken hbo's en universiteiten naar 'academische kwaliteit' bij de opleiding van studenten. Beschrijf vijf variabelen die u per instituut zou willen meten als u voor het onderzoek werd gevraagd. Beredeneer elke keuze.
- 1.7** Wordt autorijden steeds gevaarlijker? Het aantal verkeersdoden nam gedurende een aantal jaren af. In 1992 kwamen nog maar 39.250 mensen om. Daarna begon het aantal weer te stijgen. In 2002 vonden 42.815 mensen de dood bij verkeersongevallen. Maar meer auto's leggen met elkaar meer mijlen af in 2002 dan in 1992. De (Amerikaanse) overheid stelt vast dat het aantal voertuigmijlen in 1992 2247 miljard bedroeg en in 2002 2830 miljard. Vervoersrapportages hanteren het aantal doden per 100 miljoen mijlen als een maat voor risico's. Vergelijk de cijfers voor 1992 en 2002. Wat is uw conclusie?
- 1.8** Een formele omschrijving van afval is 'huishoudelijk afval'. Hier volgt een uitsplitsing van materialen voor het huishoudelijk afval in Amerika in 2000:

Materialen	Gewicht (miljoenen tonnen)	Percentage van het totaal
Voedselresten	25,9	11,2
Glas	12,8	5,5
Metalen	18,0	7,8
Papier, karton	86,7	37,4
Plastic	24,7	10,7
Rubber, leer, textiel	15,8	6,8
Hout	12,7	5,5
Tuinafval	27,7	11,9
Overig	7,5	3,2
Totaal	231,9	100,0

- (a) Tel het gewicht op van de negen materialen, inclusief 'overig'. Iedere regel, ook het totaal, is afzonderlijk afgerond op het nabijgelegen tiental. Dus uw optelling en het aangegeven totaal kunnen enigszins verschillen door **afroundingsfouten**.

- (b) Maak een staafdiagram van de percentages. Het diagram geeft een duidelijker beeld van de belangrijkste bijdragen aan afval als u de staven van groot naar klein ordent.
- (c) Gebruikt u software, maak dan een taartdiagram van de percentages. Vergelijk de twee diagrammen en merk op dat het gemakkelijker is kleine verschillen tussen voedselresten, plastic en tuinafval te traceren in het staafdiagram.

1.9 Wat is de kleur van uw auto? Hier volgen de populairste kleuren van auto's die in Noord-Amerika zijn gefabriceerd in het modeljaar 2003:

Kleur	Percentage auto's
Zilver	20,1
Wit	18,4
Zwart	11,6
Middel/donkergrijs	11,5
Lichtbruin	8,8
Middel/donkerblauw	8,5
Middelrood	6,9

Wat is het percentage auto's met een andere kleur? Zet deze gegevens in een staafdiagram. Is het ook correct om een taartdiagram te gebruiken als een categorie 'overig' wordt vermeld? Gebruikt u software, maak dan een taartdiagram als dit correct is om te doen.

1.10 De favoriete kleur bij auto's verschilt per type. Hier volgen de gegevens van de populairste kleuren in 2003 voor personenauto's en voor SUV's, trucks en bestelwagens. Een '—' betekent minder dan 1%. Wees creatief: maak een staafdiagram dat zowel de twee voertuigtypen als de te vergelijken kleuren aangeeft. Richt uw diagram zo in dat het gemakkelijk is om de twee typen voertuigen te vergelijken.



Kleur	personenauto %	SUV/truck/bestelwagens %
Zwart	10,9	11,6
Lichtbruin	—	6,3
Middel/donkerblauw	3,8	9,3
Middel/donkergrijs	23,3	8,8
Middel/donkergroen	—	7,0
Middelrood	3,9	6,2
Wit	30,4	22,3
Zilver	18,8	17,0

- 1.11** Het (Amerikaanse) ministerie van onderwijs raamt het gemiddelde budgettekort van studenten – de collegegelden minus de geraamde gezinsbijdragen en overige bijdragen. Hier volgen de gemiddelden voor dagstudenten aan de vier instellingen in het academische jaar 1999-2000:

Openbaar 2-jarig	Openbaar 4-jarig	Particulier non-profit 4-jarig	Particulier
\$2747	\$2369	\$4931	\$6548

Maak een staafdiagram van deze gegevens. Schrijf een conclusie in een zin over het budgettekort van studenten. Leg duidelijk uit waarom het niet correct is om een taartdiagram te maken.

- 1.12** De JD Power Initial Quality Survey ondervraagt meer dan 50.000 kopers van een nieuwe auto 90 dagen na hun aankoop. In een enquête van twee pagina's wordt gevraagd over 'de dingen die verkeerd gingen'. Hier volgen gegevens over problemen per 100 auto's in de afgelopen jaren voor auto's van Toyota en General Motors. Toyota stond aan de top voor wat betreft kwaliteit. Maak twee tijdgrafieken in hetzelfde diagram om Toyota en GM te vergelijken. Wat zijn de belangrijkste conclusies die u uit uw diagram haalt?

Jaar	1998	1999	2000	2001	2002	2003	2004
GM	187	179	164	147	130	134	120
Toyota	156	134	116	115	107	115	101

- 1.13** Email spam is de vloek van het internet. Hier volgt een compilatie van de gangbare soorten spam:

Soort spam	Perc.
Sex	14,5
Financieel	16,2
Gezondheid	7,3
Vrije tijd	7,8
Producten	21,0
Zwendel	14,2

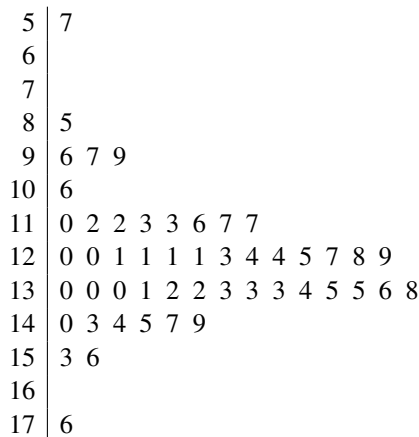
Maak twee tabelstaafdiagrammen van deze percentages, een volgens de tabel (alfabetisch) en de ander met staven van groot naar klein. Vergelijkingen zijn gemakkelijker te maken als u de staven naar hoogte sorteert. Een staafdiagram gesorteerd op lengte wordt vaak een **Pareto-diagram** genoemd, naar de Italiaanse econoom die deze procedure aanraadde.

- 1.14** Hier volgen de percentages vrouwelijke studenten in de verschillende studierichtingen in het academische jaar 1999-2000:

Graad	Percentage vrouwen
Bedrijfskunde (master)	39,8
Onderwijs (master)	76,2
Kunsten (master)	59,6
Natuurkunde (master)	53,0
Onderwijs (doctor)	70,8
Overig gepromoveerd	54,2
Geneeskunde	44,0
Recht	50,2
Theologie	20,2

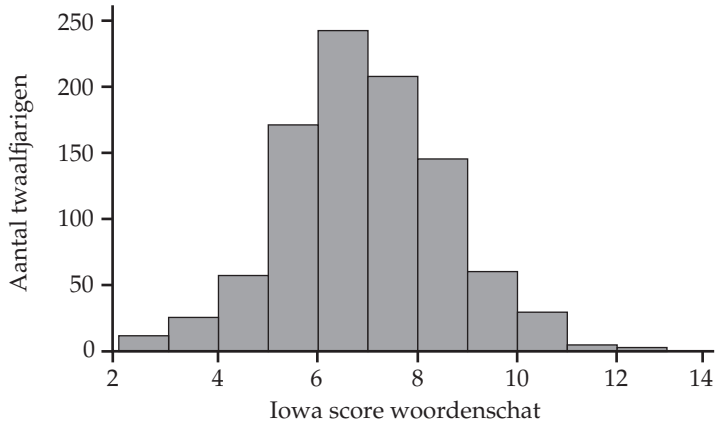
- (a) Leg duidelijk uit waarom we geen taartdiagram voor deze gegevens gebruiken.
- (b) Maak een staafdiagram van deze gegevens. (Vergelijkingen zijn gemakkelijker als u de staven naar hoogte sorteert.)

1.15 De bevolking in de Verenigde Staten vergrijst, hoewel minder snel dan in andere ontwikkelde landen. Hier volgt een stamdiagram van de percentages inwoners van 65 jaar of ouder in 50 staten volgens de volkstelling van 2000. De stammen bestaan uit hele procenten en de bladeren uit tienden van een procent.



- (a) Er zijn twee uitschieters: Alaska heeft het laagste percentage aan oudere inwoners en Florida het hoogste percentage. Wat zijn de percentages van deze twee staten?
- (b) Laat Alaska en Florida weg en beschrijf de vorm, het centrum en de spreiding van de verdeling.

1.16 Maak nog een stamdiagram van het percentage inwoners van 65 jaar en ouder in de staten zonder Alaska en Florida, door de stammen 8 t/m 15 te splitsen in de grafische voorstelling van de vorige opgave. Welke grafische voorstelling ziet u liever? Waarom?



Figuur 1.1 Histogram van de gegevens over de woordenschat van twaalfjarigen in Gary, Indiana, in de Iowa Test of Basic Skills (ITBS); (opgave 1.19)

- 1.17** Mensen met diabetes moeten hun bloedglucosespiegel onder controle houden. Het doel is de nuchtere plasmagluucose tussen de 90 en 130 milligram per deciliter (mg/dl) te handhaven. Hier volgen de niveaus voor 18 diabetici die meededen aan groepstraining, vijf maanden na afloop van de training:

141	158	112	153	134	95	96	78	148
172	200	271	103	172	359	145	147	255

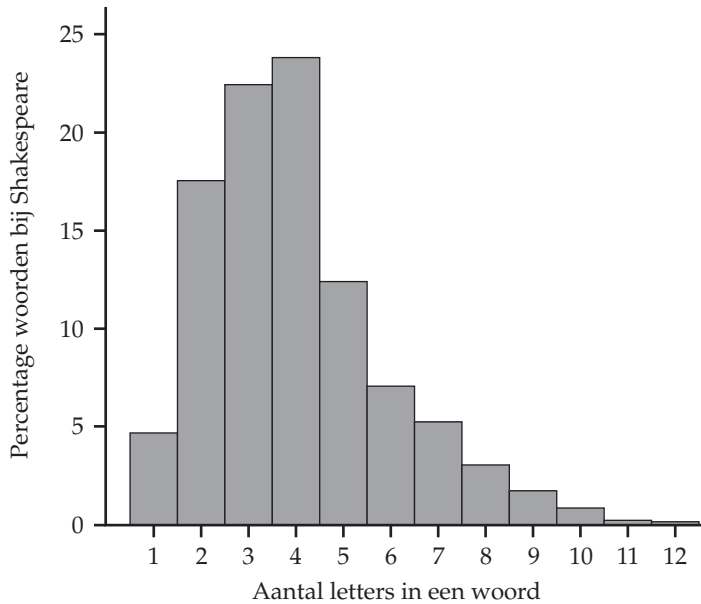
Maak een stamdiagram van deze gegevens en beschrijf de belangrijkste kenmerken van de verdeling. (U mag de stammen inkorten en splitsen.) Zijn er uitschieters? Hoe goed slaagt de groep als geheel er in het gestelde doel te behalen?

- 1.18** Het onderzoek uit de vorige opgave mat ook het nuchtere plasmagluucose van 16 diabetici die een persoonlijke instructie kregen over diabetescontrole. Hier volgen de gegevens:

128	195	188	158	227	198	163	164
159	128	283	226	223	221	220	160

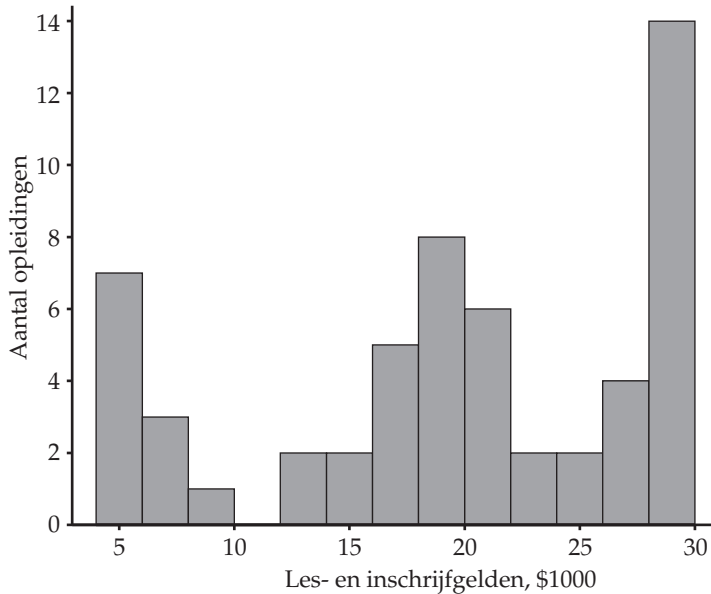
Maak een rug-aan-rug stamdiagram om de twee groepen die op verschillende wijze werden geïnstrueerd, te vergelijken. Hoe valt de vergelijking uit tussen de vorm van de verdelingen en het succes in bereiken van het gewenste glucoseniveau?

- 1.19** Figuur 1.1 toont de gegevens van alle 947 twaalfjarige leerlingen van openbare scholen van Gary in de staat Indiana, voor wat betreft het onderdeel woordenschat in de Iowa Test of Basic Skills (ITBS). Geef een korte beschrijving van het globale patroon (vorm, centrum, spreiding) van deze verdeling.



Figuur 1.2 Histogram van de lengte van woorden die voorkomen in toneelstukken van Shakespeare (opgave 1.20)

- 1.20** Figuur 1.2 is een histogram van de lengte van woorden die voorkomen in toneelstukken van Shakespeare. Omdat er zoveel woorden in de toneelstukken voorkomen, gebruiken we een histogram met procenten. Wat is de globale vorm van deze verdeling? Wat zegt de vorm over de woordlengte bij Shakespeare? Verwacht u bij andere auteurs verdelingen van woordlengten met dezelfde hoofdvorm? Waarom?
- 1.21** Jeanna wil gaan studeren in Massachusetts omdat zij daar woont. Ze zoekt de les- en inschrijfgelden op voor alle 56 vierjarige opleidingen in Massachusetts (met weglating van kunst- of andere bijzondere opleidingen). Figuur 1.3 is een histogram van de gegevens. Voor openbare scholen gebruikte Jeanna het door de staat vastgestelde tarief. Wat is het belangrijkste aspect van het globale patroon van deze verdeling? Waarom ziet dit patroon er zo uit?
- 1.22** De staten verschillen onderling aanmerkelijk qua zware weersomstandigheden. Tabel 1.1 toont de gemiddelde schade veroorzaakt door tornado's per jaar in de periode 1950 t/m 1999 in elk van de 50 staten en Puerto Rico. (Het verschil in koopkracht van de dollar is gecorrigeerd door de schade te berekenen naar de koopkracht in 1999.)
- Welke staten behoren tot de top vijf qua schade? Welke vijf lijden de minste schade?
 - Maak een histogram van de gegevens, handmatig of met software, met klassen ' $0 \leq \text{schade} < 10$ ', ' $10 \leq \text{schade} < 20$ ', enzovoort. Beschrijf vorm, centrum en spreiding van de verdeling. Welke staten zijn uitschieters? (Voor de beoordeling van uitschieters moet u erop letten dat de meeste tornado's in grotendeels landelijke



Figuur 1.3 Histogram van les- en inschrijfgelden van vierjarige opleidingen in Massachusetts (opgave 1.21)

staten zoals Kansas weinig schade aan onroerend goed veroorzaken. Schade aan de oogst wordt hier niet toe gerekend.)

- (c) Gebruikt u software, laat dan ook het standaard histogram zien dat uw software maakt als u geen instructies geeft. Wat zijn de verschillen met uw histogram in (b)?

1.23 Het *One Variable Statistical calculator*-applet op de website maakt stamdiagrammen en histogrammen. Dit is meer bedoeld als oefening dan als vervanging voor statistische software. De histogramfunctie is vooral nuttig omdat u het aantal klassen kunt veranderen door met de muis te slepen. De schadegegevens van de tornado's in tabel 1.1 zijn beschikbaar in het applet. Kies deze gegevensverzameling en ga dan naar tabblad histogram.



- (a) Beschrijf het histogram dat het applet als eerste toont. Als het standaard diagram geen negen klassen bevat, maak met de muis dan negen klassen en beschrijf het resultaat. Dit moet overeenkomen met uw histogram in deel (b) van de vorige opgave.
- (b) Maak een histogram met een klasse en ook een histogram met het maximale aantal klassen dat het applet aan kan. Beschrijf de resultaten.
- (c) Sleep met het diagram tot u het histogram vindt waarvan u denkt dat het de gegevens het beste weergeeft. Hoeveel klassen koos u? Beschrijf uw laatste histogram.

Staat	Schade (\$miljoenen)	Staat	Schade (\$miljoenen)	Staat	Schade (\$miljoenen)
Alabama	51,88	Louisiana	27,75	Ohio	44,36
Alaska	0,00	Maine	0,53	Oklahoma	81,94
Arizona	3,47	Maryland	2,33	Oregon	5,52
Arkansas	40,96	Massachusetts	4,42	Pennsylvania	17,11
California	3,68	Michigan	29,88	Puerto Rico	0,05
Colorado	4,62	Minnesota	84,84	Rhode Island	0,09
Connecticut	2,26	Mississippi	43,62	South Carolina	17,19
Delaware	0,27	Missouri	68,93	South Dakota	10,64
Florida	37,32	Montana	2,27	Tennessee	23,47
Georgia	51,68	Nebraska	30,26	Texas	88,60
Hawaii	0,34	Nevada	0,10	Utah	3,57
Idaho	0,26	New Hampshire	0,66	Vermont	0,24
Illinois	62,94	New Jersey	2,94	Virginia	7,42
Indiana	53,13	New Mexico	1,49	Washington	2,37
Iowa	49,51	New York	15,73	West Virginia	2,14
Kansas	49,28	North Carolina	14,90	Wisconsin	31,33
Kentucky	24,84	North Dakota	14,69	Wyoming	1,78

Tabel 1.1 Gemiddelde schade per jaar vanwege tornado's

- 1.24** De uitstoot van krachtcentrales en auto's bevat kooldioxide (CO_2) dat bijdraagt aan de opwarming van de aarde. Tabel 1.2 toont de CO_2 uitstoot per persoon in landen met een bevolking van minimaal 20 miljoen.
- (a) Waarom denkt dat we kozen voor uitstoot per persoon in plaats van de totale uitstoot CO_2 per land?
- (b) Laat de gegevens van tabel 1.2 in een diagram zien. Beschrijf vorm, centrum en spreiding van de verdeling. Welke landen zijn uitschieters?
- 1.25** Tabel 1.3 bevat gegevens over de gemiddelde jaarlijkse temperatuur (in graden Fahrenheit) in de jaren 1951 tot en met 2000 in Californië voor Pasadena en Redding. Maak tijdgrafieken van beide tijdreeksen en vergelijk de belangrijkste kenmerken. U ziet waarom discussies over klimaatverandering vaak onenigheid tweebrengen.
- 1.26** Maak een histogram van de gemiddelde jaarlijkse temperatuur in Pasadena in de jaren 1951 tot en met 2000. (gegevens staan in tabel 1.3). Verklaar vervolgens waarom dit histogram zeer belangrijke feiten mist over de temperatuur in Pasadena.
- 1.27** De indruk die een tijdgrafiek geeft hangt af van de schaal die u gebruikt voor de beide assen. Als u de verticale as uitrekt en de tijdas comprimeert, lijken de veranderingen veel sneller. Als u de verticale as comprimeert en de tijdas uitrekt, lijken de veranderingen veel trager. Maak nog twee tijdgrafieken van de gegevens voor Pasadena in tabel 1.3, een die zorgt dat de gemiddelde temperatuur zeer snel toeneemt en een die alleen een trage toename toont. De moraal van deze opgave is: *schenk veel aandacht aan de schalen wanneer u een tijdgrafiek bestudeert.*



Land	CO ₂	Land	CO ₂	Land	CO ₂
Algeria	2,3	Italy	7,3	Poland	8,0
Argentina	3,9	Iran	3,8	Romania	3,9
Australia	17,0	Iraq	3,6	Russia	10,2
Bangladesh	0,2	Japan	9,1	Saudi Arabia	11,0
Brazil	1,8	Kenya	0,3	South Africa	8,1
Canada	16,0	Korea, North	9,7	Spain	6,8
China	2,5	Korea, South	8,8	Sudan	0,2
Colombia	1,4	Malaysia	4,6	Tanzania	0,1
Congo	0,0	Mexico	3,7	Thailand	2,5
Egypt	1,7	Morocco	1,0	Turkey	2,8
Ethiopia	0,0	Myanmar	0,2	Ukraine	7,6
France	6,1	Nepal	0,1	United Kingdom	9,0
Germany	10,0	Nigeria	0,3	United States	19,9
Ghana	0,2	Pakistan	0,7	Uzbekistan	4,8
India	0,9	Peru	0,8	Venezuela	5,1
Indonesia	1,2	Philippines	0,9	Vietnam	0,5

Tabel 1.2 CO₂-uitstoot, metrieke tonnen per persoon

- 1.28** ‘Aanwas’, de toename van de vispopulatie, is een belangrijke maatstaf voor de toestand van de ecosystemen in de oceaan. Hier volgen de gegevens van de aanwas van tong in de Beringzee tussen 1973 en 2000:

Aanwas		Aanwas		Aanwas		Aanwas	
Jaar	(miljoenen)	Jaar	(miljoenen)	Jaar	(miljoenen)	Jaar	(miljoenen)
1973	173	1980	1411	1987	4700	1994	505
1974	234	1981	1431	1988	1702	1995	304
1975	616	1982	1250	1989	1119	1996	425
1976	344	1983	2246	1990	2407	1997	214
1977	515	1984	1793	1991	1049	1998	385
1978	576	1985	1793	1992	505	1999	445
1979	727	1986	2809	1993	998	2000	676

- (a) Laat met een diagram de verdeling van de aanwas van tong zien. Beschrijf het patroon en enkele opvallende afwijkingen die u ziet.
- (b) Maak een tijdgrafiek van de aanwas en beschrijf het patroon. Zoals vaak het geval is met gegevens van tijdreeksen, is een tijdgrafiek nodig om te begrijpen wat er gebeurt.

- 1.29** In de Aziatische cultuur wordt de nadruk niet op de slanke lijn gelegd, maar jonge Aziaten worden vaak door de westerse cultuur beïnvloed. In een onderzoek naar zorgen

Gemiddelde temperatuur			Gemiddelde temperatuur		
Jaar	Pasadena	Redding	Jaar	Pasadena	Redding
1951	62,27	62,02	1976	64,23	63,51
1952	61,59	62,27	1977	64,47	63,89
1953	62,64	62,06	1978	64,21	64,05
1954	62,88	61,65	1979	63,76	60,38
1955	61,75	62,48	1980	65,02	60,04
1956	62,93	63,17	1981	65,80	61,95
1957	63,72	62,42	1982	63,50	59,14
1958	65,02	64,42	1983	64,19	60,66
1959	65,69	65,04	1984	66,06	61,72
1960	64,48	63,07	1985	64,44	60,50
1961	64,12	63,50	1986	65,31	61,76
1962	62,82	63,97	1987	64,58	62,94
1963	63,71	62,42	1988	65,22	63,70
1964	62,76	63,29	1989	64,53	61,50
1965	63,03	63,32	1990	64,96	62,22
1966	64,25	64,51	1991	65,60	62,73
1967	64,36	64,21	1992	66,07	63,59
1968	64,15	63,40	1993	65,16	61,55
1969	63,51	63,77	1994	64,63	61,63
1970	64,08	64,30	1995	65,43	62,62
1971	63,59	62,23	1996	65,76	62,93
1972	64,53	63,06	1997	66,72	62,48
1973	63,46	63,75	1998	64,12	60,23
1974	63,93	63,80	1999	64,85	61,88
1975	62,36	62,66	2000	66,25	61,58

Tabel 1.3 Gemiddelde jaarlijkse temperatuur (°F) in twee steden in Californië

van jonge Koreaanse vrouwen over hun gewicht paste de onderzoekers de Drive for Thinness (DfT) schaal (een vragenlijst) toe op 264 vrouwelijke studenten in Seoul. De gegevens staan in het bestand `ex01-029.dat` op de website. DfT meet overmatige bezorgdheid over gewicht en dieet en de vrees voor gewichtstoename. Globaal genomen is een score van 15 kenmerkend voor westerse vrouwen met eetstoornissen, maar ongebruikelijk hoog (90^e percentiel) voor andere westerse vrouwen. Zet de gegevens in een diagram en beschrijf vorm, centrum en spreiding van de verdeling van de DfT-score voor deze Koreaanse studenten. Zijn er uitschieters?

- 1.30** Een verandering in de keuze van de klassen leidt tot een ander histogram. Hier volgt een voorbeeld waarbij een kleine verschuiving in klassen, zonder verandering van het aantal klassen, een belangrijk effect heeft op het histogram. Het gaat om gegevens van de zuurgraad (de pH) in 105 monsters regenwater. Gedistilleerd water heeft een pH 7,00. Als het water zuurder wordt gaat de pH omlaag. De pH van regenwater is belangrijk voor milieudeskundigen vanwege het probleem van zure regen.



4,33	4,38	4,48	4,48	4,50	4,55	4,59	4,59	4,61	4,61
4,75	4,76	4,78	4,82	4,82	4,83	4,86	4,93	4,94	4,94
4,94	4,96	4,97	5,00	5,01	5,02	5,05	5,06	5,08	5,09
5,10	5,12	5,13	5,15	5,15	5,15	5,16	5,16	5,16	5,18
5,19	5,23	5,24	5,29	5,32	5,33	5,35	5,37	5,37	5,39
5,41	5,43	5,44	5,46	5,46	5,47	5,50	5,51	5,53	5,55
5,55	5,56	5,61	5,62	5,64	5,65	5,65	5,66	5,67	5,67
5,68	5,69	5,70	5,75	5,75	5,75	5,76	5,76	5,79	5,80
5,81	5,81	5,81	5,81	5,85	5,85	5,90	5,90	6,00	6,03
6,03	6,04	6,04	6,05	6,06	6,07	6,09	6,13	6,21	6,34
6,43	6,61	6,62	6,65	6,81					

- Maak een histogram van de pH met 14 klassen, gebruik de klassengrenzen 4,2, 4,4, ..., 7,0. Hoeveel pieken laat uw histogram zien? Meer dan een piek suggereert dat de gegevens groepen met verschillende verdelingen bevatten.
- Maak een tweede histogram, ook met 14 klassen, gebruik als grenzen nu 4,14, 4,34, ..., 6,94. De klassen zijn ten opzichte van (a) met 0,06 naar links opgeschoven. Hoeveel pieken laat het nieuwe histogram zien?
- Gebruik de histogramfunctie van uw software om een histogram te maken zonder het aantal klassen of hun grenzen te specificeren. Hoe ziet het histogram met de standaardwaarde eruit in vergelijking tot die van (a) en (b)?

1.31 In een onderzoek onder een grote groep studenten werden de volgende vragen gesteld:

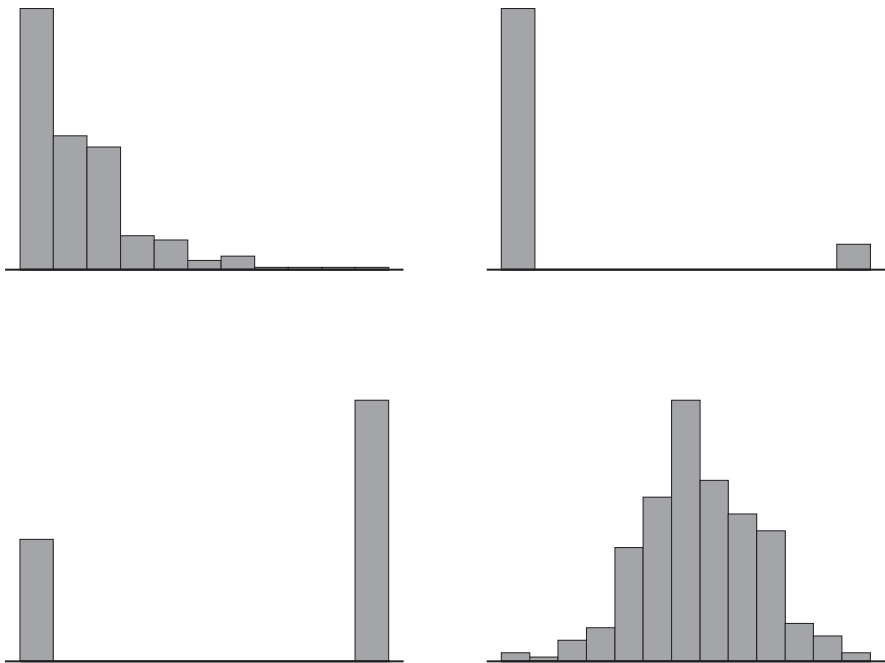


- Bent u vrouw of man? (In de gegevens: man = 0, vrouw = 1.)
- Bent u rechts- of linkshandig? (In de gegevens: rechts = 0, links = 1.)
- Hoe lang bent u (in inches)?
- Hoeveel minuten studeert u op een doordgewone avond in de week?

Figuur 1.4 toont histogrammen van de antwoorden, niet gesorteerd en zonder schaal-aanduiding. Welk histogram hoort bij welke variabele? Beredeneer uw antwoorden.

1.32 Maak een histogram voor een verdeling die linksscheef is. Veronderstel dat u en uw vrienden de zakken leegmaken en van de verzamelde munten steeds het jaar noteren. De verdeling van de jaren zou linksscheef moeten zijn. Leg uit waarom dat het geval is.

1.33 Hoeveel olie de bronnen in een bepaald veld uiteindelijk produceren is sleutel informatie voor de beslissing of er nog meer bronnen worden aangeboord. Hier volgen de geschatte totale aantallen olie afkomstig van 64 bronnen in het gebied van Devonian Richmond Dolomite (Michigan) per duizend vaten:



Figuur 1.4 Koppel ieder histogram aan zijn variabele (opgave 1.31)

21,71	53,2	46,4	42,7	50,4	97,7	103,1	51,9	43,4	69,5
156,5	34,6	37,9	12,9	2,5	31,4	79,5	26,9	18,5	14,7
32,9	196,0	24,9	118,2	82,2	35,1	47,6	54,2	63,1	69,8
57,4	65,6	56,4	49,4	44,9	34,6	92,2	37,0	58,8	21,3
36,6	64,9	14,8	17,6	29,1	61,4	38,6	32,5	12,0	28,3
204,9	44,5	10,3	37,7	33,7	81,1	12,1	20,1	30,5	7,1
10,1	18,0	3,0	2,0						

Maak een diagram, geef de verdeling weer en beschrijf de belangrijkste kenmerken.

- 1.34** In 1798 heeft de Engelse geleerde Henry Cavendish de dichtheid van de aarde gemeten met een torsiebalans. De waargenomen variabele was de dichtheid van de aarde als veelvoud van de dichtheid van water. Hier volgen de 29 metingen van Cavendish:

5,50	5,61	4,88	5,07	5,26	5,55	5,36	5,29	5,58	5,65
5,57	5,53	5,62	5,29	5,44	5,34	5,79	5,10	5,27	5,39
5,42	5,47	5,63	5,34	5,46	5,30	5,75	5,68	5,85	

Geef de metingen grafisch weer hetzij met een stamdiagram hetzij met een histogram en beredeneer uw keuze. Geef daarna een korte beschrijving van de belangrijkste

43	45	53	56	56	57	58	66	67	73
74	79	80	80	81	81	81	82	83	83
84	88	89	91	91	92	92	97	99	99
100	100	101	102	102	102	103	104	107	108
109	113	114	118	121	123	126	128	137	138
139	144	145	147	156	162	174	178	179	184
191	198	211	214	243	249	329	380	403	511
522	598								

Tabel 1.4 Overlevingsduur (in dagen) van proefkonijnen in een medisch experiment

kenmerken van de verdeling. Wat is op grond van deze metingen uw inschatting van de dichtheid van de aarde?

- 1.35** Studeren vrouwen meer dan mannen? We vroegen aan een grote groep eerstejaars hoeveel minuten zij op een gewone avond in de week studeerden. Hier volgen de antwoorden van een aselechte steekproef van 30 vrouwen en 30 mannen uit de groep

Vrouwen					Mannen				
180	120	180	360	240	90	120	30	90	200
120	180	120	240	170	90	45	30	120	75
150	120	180	180	150	150	120	60	240	300
200	150	180	150	180	240	60	120	60	30
120	60	120	180	180	30	230	120	95	150
90	240	180	115	120	0	200	120	120	180

- (a) Bestudeer de gegevens. Bent u niet verrast dat de meeste antwoorden veelvouden van 10 minuten zijn? We lieten een student weg die opgaf 30.000 minuten per avond te studeren. Zijn er andere antwoorden die u verdacht vindt?
- (b) Maak een rug-aan-rug stamdiagram van deze gegevens. Beschrijf bij benadering de centra van de beide groepen. Is het zo dat vrouwen meer studeren dan mannen (of dat althans beweren)?
- 1.36** Tabel 1.4 toont de overlevingsduur in dagen van 72 proefkonijnen die ten behoeve van een medisch experiment met tuberkelbacillen zijn geïnjecteerd. Maak een geschikte grafiek en beschrijf vorm, centrum en spreiding voor de verdeling van de overlevings-tijden. Zijn er ook uitschieters?
- 1.37** In tabel 1.5 worden gegevens weergegeven van 78 studenten in het zevende leerjaar van een plattelandsschool in het Midwesten. De onderzoeker was geïnteresseerd in de relatie tussen het ‘zelfbeeld’ van de studenten en hun intellectuele prestaties. De data in de tabel 1.5 bestaan uit een cijfergemiddelde (GPA), de score voor een standaard IQ-test en het geslacht, gegevens die uit de schooldossiers zijn overgenomen. Het geslacht is gecodeerd met een V voor vrouwelijk en een M voor mannelijk. De studenten zelf

OBS	GPA	IQ	Geslacht	Zelfbeeld	OBS	GPA	IQ	Geslacht	Zelfbeeld
001	7,940	111	M	67	043	10,760	123	M	64
002	8,292	107	M	43	044	9,763	124	M	58
003	4,643	100	M	52	045	9,410	126	M	70
004	7,470	107	M	66	046	9,167	116	M	72
005	8,882	114	F	58	047	9,348	127	M	70
006	7,585	115	M	51	048	8,167	119	M	47
007	7,650	111	M	71	050	3,647	97	M	52
008	2,412	97	M	51	051	3,408	86	F	46
009	6,000	100	F	49	052	3,936	102	M	66
010	8,833	112	M	51	053	7,167	110	M	67
011	7,470	104	F	35	054	7,647	120	M	63
012	5,528	89	F	54	055	0,530	103	M	53
013	7,167	104	M	54	056	6,173	115	M	67
014	7,571	102	F	64	057	7,295	93	M	61
015	4,700	91	F	56	058	7,295	72	F	54
016	8,167	114	F	69	059	8,938	111	F	60
017	7,822	114	F	55	060	7,882	103	F	60
018	7,598	103	F	65	061	8,353	123	M	63
019	4,000	106	M	40	062	5,062	79	M	30
020	6,231	105	F	66	063	8,175	119	M	54
021	7,643	113	M	55	064	8,235	110	M	66
022	1,760	109	M	20	065	7,588	110	M	44
024	6,419	108	F	56	068	7,647	107	M	49
026	9,648	113	M	68	069	5,237	74	F	44
027	10,700	130	F	69	071	7,825	105	M	67
028	10,580	128	M	70	072	7,333	112	F	64
029	9,429	128	M	80	074	9,167	105	M	73
030	8,000	118	M	53	076	7,996	110	M	59
031	9,585	113	M	65	077	8,714	107	F	37
032	9,571	120	F	67	078	7,833	103	F	63
033	8,998	132	F	62	079	4,885	77	M	36
034	8,333	111	F	39	080	7,998	98	F	64
035	8,175	124	M	71	083	3,820	90	M	42
036	8,000	127	M	59	084	5,936	96	F	28
037	9,333	128	F	60	085	9,000	112	F	60
038	9,500	136	M	64	086	9,500	112	F	70
039	9,167	106	M	71	087	6,057	114	M	51
040	10,140	118	F	72	088	6,057	93	F	21
041	9,999	119	F	54	089	6,938	106	M	56

Tabel 1.5 Onderwijsgegevens voor 78 studenten in het zevende leerjaar

worden alleen met een waarnemingsnummer (OBS) aangeduid. Aan de ontbrekende OBS-nummers is te zien dat sommige studenten voortijdig zijn afgehaakt. De laatste

variabele is de score van elke student op de Piers–Harris Children’s Self–Concept Scale, een psychologische test die door de onderzoeker is uitgevoerd.

- (a) Hoeveel variabelen bevat deze gegevensverzameling? Welke zijn kwalitatieve variabelen en welke zijn kwantitatieve variabelen?
- (b) Maak een stamdiagram van de verdeling van de GPA, nadat u deze op tientallen hebt afgerond.
- (c) Beschrijf vorm, centrum en spreiding van de GPA–verdeling. Identificeer eventuele uitschieters in het globale patroon.
- (d) Maak een rug–aan–rug stamdiagram van de afgeronde GPA’s voor mannelijke en vrouwelijke studenten. Beschrijf de overeenkomsten en verschillen tussen de beide verdelingen.
- 1.38** Maak een grafiek van de verdeling van de IQ–scores voor de zevendejaarsstudenten in tabel 1.5. Beschrijf vorm, centrum en spreiding voor de verdeling en eventuele uitschieters. Voor IQ–scores geldt gewoonlijk dat het centrum honderd is. Ligt het centrum voor deze studenten rond de honderd, duidelijk daarboven of duidelijk daaronder?
- 1.39** Beschrijf op basis van een geschikte grafiek kort de verdeling van de zelfbeeldscores voor de studenten in tabel 1.5. Geef eventuele uitschieters aan.
- 1.40** Sinds 1972 mogen vrouwen deelnemen aan de marathon van Boston. Hier volgen de tijden (in minuten, afgerond op hele minuten) voor de winnende vrouwen voor de jaren 1959 t/m 2003:

Jaar	Tijd	Jaar	Tijd	Jaar	Tijd	Jaar	Tijd
1972	190	1980	154	1988	145	1996	147
1973	186	1981	147	1989	144	1997	146
1974	167	1982	150	1990	145	1998	143
1975	162	1983	143	1991	144	1999	143
1976	167	1984	149	1992	144	2000	146
1977	168	1985	154	1993	145	2001	144
1978	165	1986	145	1994	142	2002	141
1979	155	1987	146	1995	145	2003	145

Maak een grafiek die de veranderingen in de loop van de tijd laat zien. Wat is het globale patroon? Zijn de tijden de laatste jaren niet meer verbeterd? Zo ja, sinds wanneer niet meer?

Opgaven bij paragraaf 1.2

- 1.41** Figuur 1.3 is een histogram van de collegelden en andere kosten van vierjarige opleidingen in Massachusetts. Hier volgen de kosten (in dollars) in oplopende volgorde gerangschikt: