

Statistiek om mee te werken

Opgaven en
uitwerkingen



Noordhoff Uitgevers

Arie Buijs

10^e druk

Statistiek om mee te werken

Opgaven en uitwerkingen

Prof. dr. A. Buijs

Tiende druk

Noordhoff Uitgevers, Groningen/Utrecht

Ontwerp omslag: G2K (Groningen-Amsterdam)

Omslagillustratie: Unsplash/Vladimir Kudinov

Eventuele op- en aanmerkingen over deze of andere uitgaven kunt u richten aan:
Noordhoff Uitgevers bv, Afdeling Hoger Onderwijs, Antwoordnummer 13, 9700 VB
Groningen, e-mail: info@noordhoff.nl

Aan de totstandkoming van deze uitgave is de uiterste zorg besteed. Voor informatie die desondanks onvolledig of onjuist is opgenomen, aanvaarden auteur(s), redactie en uitgever geen aansprakelijkheid. Voor eventuele verbeteringen van de opgenomen gegevens houden zij zich aanbevolen.



0 / 17

© 2017 Noordhoff Uitgevers bv Groningen/Utrecht, The Netherlands.

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden veeleelvoudigd, 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. Voor zover het maken van reprografische veeleelvoudingen uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet 1912 dient men de daarvoor verschuldigde vergoedingen te voldoen aan Stichting Reprorecht (postbus 3060, 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) kan men zich wenden tot Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, postbus 3060, 2130 KB Hoofddorp, www.stichting-pro.nl).

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 prior written permission of the publisher.

ISBN (ebook) 978-90-01-87810-8

ISBN 978-90-01-87809-2

NUR 916

Woord vooraf

Voor iemand die beschikt over statistische vaardigheden zijn de vooruitzichten op studiesucces heel gunstig. Statistiek is een vakgebied dat voor vrijwel alle studierichtingen van belang is. Het levert een breed scala van technieken waarmee men uit een verzameling gegevens gefundeerde conclusies naar boven kan halen. Bij het vakgebied statistiek gaat het niet alleen om de theorie. Ook het praktisch werken aan statistische problemen is van groot belang. Door vraagstukken te oefenen met een statistische component, ontwikkelt men bovendien het eigen analytisch denkvermogen.

Op de huidige arbeidsmarkt hebben personen met analytische skills een heel kansrijke positie. Dat heeft onder andere te maken met de opkomst van Big Data. Het gaat dan om de beschikbaarheid van grote gegevensbestanden die kunnen worden geanalyseerd voor wetenschappelijke of commerciële toepassingen. De functie van data-analist is op dit moment het meest gevraagde beroep op de arbeidsmarkt voor net-afgestudeerden. En dan maakt het eigenlijk niet eens uit welke studierichting iemand heeft afgerond, maar gaat het vooral om het logisch kunnen denken gecombineerd met een solide basiskennis van statistische technieken.

Dit opgavenboek biedt de student de mogelijkheid om aan beide competenties te werken.

De herziene editie van het theorieboek *Statistiek om mee te werken* vormt de basis van dit boek. Hierin is een aantal nieuwe opgaven opgenomen en een aantal oude opgaven werd aangepast. Achterin zijn voor alle opgaven de antwoorden te vinden.

Van de opgaven zijn ook uitwerkingen beschikbaar. Deze worden op enkele verschillende manieren beschikbaar gesteld aan de gebruikers.

Er zijn drie categorieën:

- Van een aantal van de opgaven is de uitwerking in het laatste deel van dit boek afgedrukt. Deze opgaven zijn te herkennen aan het blokje rond het nummer.
- Een tweede groep uitwerkingen komt beschikbaar via een website (www.statistiekommeetewerken.noordhoff.nl) waarop studenten kunnen inloggen.
- Een derde groep uitwerkingen is exclusief beschikbaar voor docenten.

Een fors aantal van die uitwerkingen is gemaakt door ir. Koen de Bont uit Eindhoven. Hij is heel wat jaren mijn steun en toeverlaat geweest bij het uitwerken van de opgaven. Hiervoor ben ik Koen heel veel dank verschuldigd. Ik hoop dat deze tiende editie tot genoeg is van studenten en docenten.

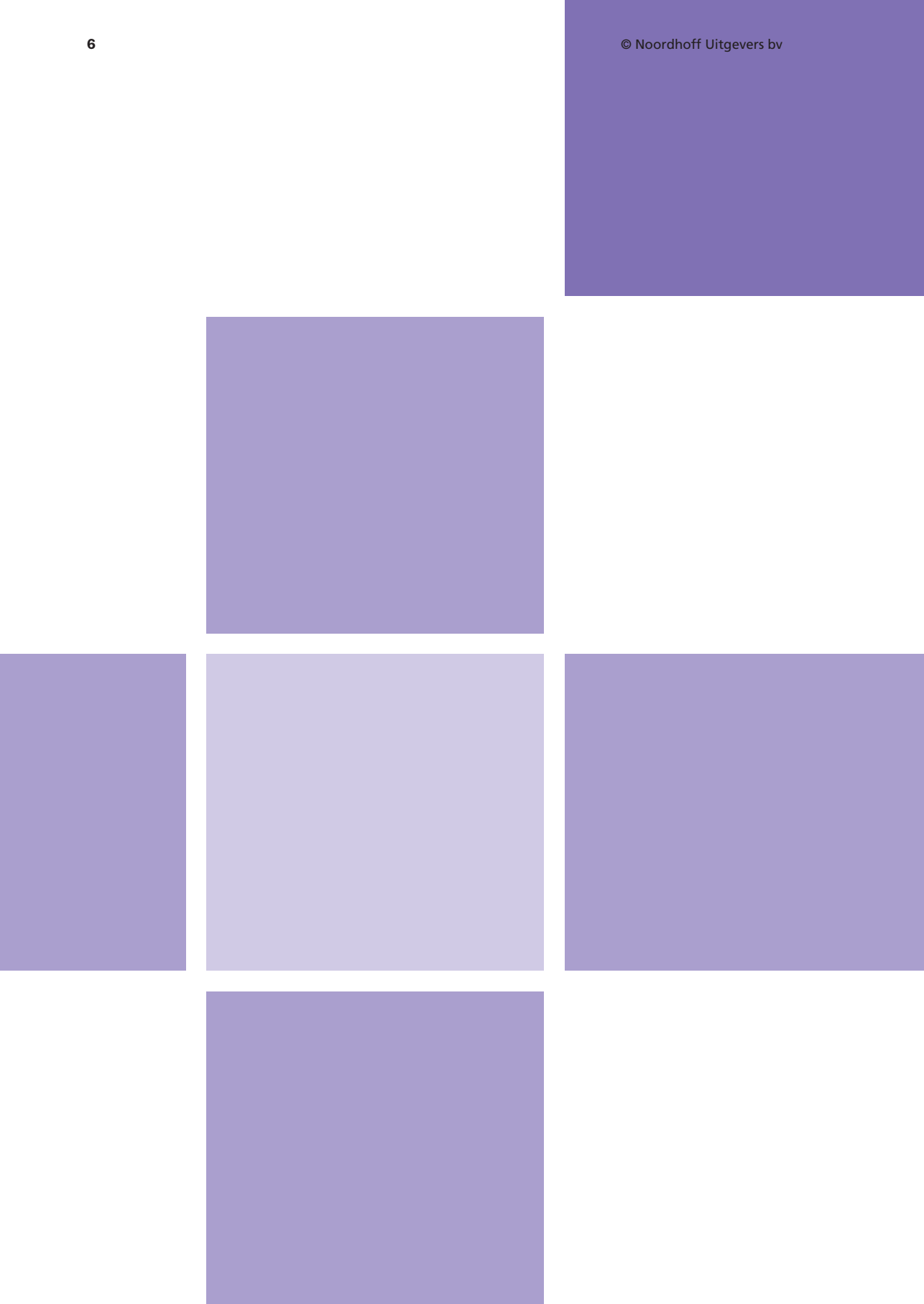
Voor de studenten van nu hoop ik dat zij veel baat zullen hebben van de aangeleerde technieken bij hun afstudeerprojecten en ook later in hun loopbaan.

Bilthoven, voorjaar 2017

Arie Buijs

Inhoud

- 1 Inleiding tot de beschrijvende statistiek 7
- 2 Maatstaven voor ligging en spreiding 23
- 3 Kansrekening 37
- 4 Kansvariabelen 51
- 5 Normale verdeling 65
- 6 Binomiale verdeling 75
- 7 De Poissonverdeling 85
- 8 Schatten 95
- 9 Toetsen 107
- 10 De chikwadraatverdeling 119
- 11 Verschiltoetsen 129
- 12 Variantieanalyse 139
- 13 Regressie en correlatie 149
- 14 Indexcijfers 163
- 15 Tijdreeksen 175
- 16 Gemengde opgaven 189
- 17 Werken met gegevensbestanden 195
- 18 Casestudies 205
- 19 Antwoorden 213
- 20 Uitwerkingen van geselecteerde opgaven 251



1

Inleiding tot de beschrijvende statistiek

Deel A Meerkeuzevragen

- m1** Bij een verkeersonderzoek is een van de grootheden die wordt genoteerd het merk van de passerende auto's. Dit merk is ...
- a** een ratiovariabele.
 - b** een kwantitatieve variabele.
 - c** een nominale variabele.
 - d** geen variabele.

Gegevens voor de vragen m2 en m3

Bij een stratenquête werden 250 voorbijgangers gevraagd naar hun mening over een aantal door de gemeente voorgestelde verkeersmaatregelen. De resultaten staan in de volgende tabel:

	<u>Mee eens</u>	<u>Geen mening</u>	<u>Oneens</u>	<u>Totaal</u>
Man	25	35	50	110
Vrouw	35	55	50	140
Totaal	60	90	100	250

- m2** Zie tabel. Bij de mannen is het percentage dat het eens is met de voorgestelde maatregelen gelijk aan ...
- a** 41,7%.
 - b** 10%.
 - c** 25%.
 - d** 22,7%.
- m3** Zie tabel. De groep mensen die het oneens is met de maatregelen bestaat voor ... uit vrouwen.
- a** 20%
 - b** 56%
 - c** 50%
 - d** 40%

Gegevens voor de vragen m4, m5 en m6

Bij de reorganisatie P-Tours heeft men bijgehouden hoeveel geboekte passagiers kort voor het vertrek van een busreis hun reis afzeggen. Voor 80 busreizen leverde dit de volgende tabel:

Aantal passagiers per busreis dat annuleert

Nummer	Aantal afzeggers	Frequentie
1	0 - < 5	36
2	5 - < 8	24
3	8 - < 12	8
4	12 - < 16	6
5	16 - < 20	4
6	20 en hoger	2
Totaal		80

Bron: administratie P-Tours 2016

m4 De klassenbreedte bij $0 - < 5$ bedraagt ...

- a 5.
- b 4.
- c 4,5.
- d 36.

m5 De relatieve frequentie van de klasse $16 - < 20$ bedraagt ...

- a 0,05.
- b 4.
- c 16.
- d 0,20.

m6 Het klassenmidden van de klasse $8 - < 12$ bedraagt ...

- a 2.
- b 10.
- c 10,5.
- d 9,5.

Deel B Open vragen

1.1 In een fabriek staat een machine opgesteld die stalen kogeltjes vervaardigt. Omdat deze kogeltjes aan bepaalde specificaties moeten voldoen, zijn er 60 kogeltjes nagemeten. Deze kogeltjes bleken de volgende diameter te hebben:

Diameter van 60 kogeltjes (in mm)

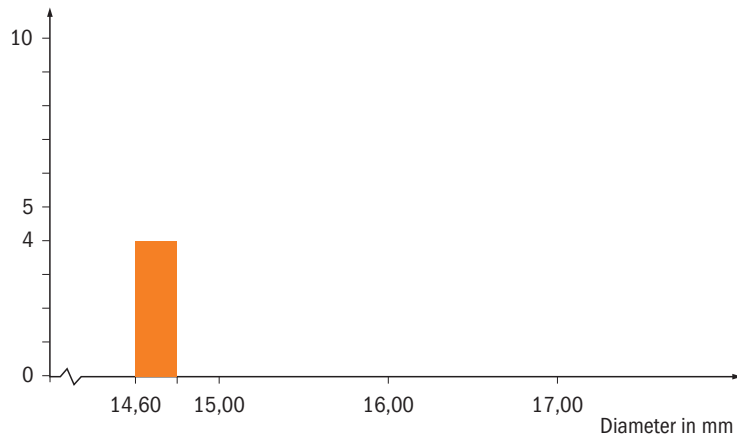
15,42	14,78	15,24	16,05	15,70	16,78	15,58	17,02	15,90	16,12
15,70	16,58	15,62	15,96	15,50	15,42	15,44	15,12	15,96	15,70
16,26	15,78	15,24	16,80	15,06	15,38	14,62	15,20	15,14	15,73
15,96	14,90	14,84	15,30	15,48	15,02	16,08	15,62	14,83	17,10
16,28	16,10	15,76	15,02	16,60	15,52	15,60	15,26	15,85	15,81

- a Verwerk de waargenomen uitkomsten tot een frequentieverdeling met klassen van 0,2 mm breed.

Klasse	Turven	Aantal
14,60 - < 14,80
14,80 - <
..... - <
..... - <
..... - <
..... - <
..... - <
..... - <
..... - <
..... - <
..... - <
..... - <
..... - <
..... - <
Totaal

ANTWOORD

FIGUUR 1.1



- b Verwerk de waargenomen gegevens in een histogram.
De klassengrenzen zijn gegevens als bijvoorbeeld: 14,60 - < 14,80.
Dit houdt in dat een waarneming 14,60 wél en een waarneming 14,80 niet in deze klasse wordt geplaatst. Op grond hiervan is het beter de klassengrenzen iets te verschuiven (namelijk met 0,005 mm), en wel zodanig dat de klasse 14,60 - < 14,80 wordt getekend als 14,595 - < 14,795.

1.2

Een groep van 200 schouwburgbezoekers is gevraagd naar het jaarincome. De inkomens leverden de volgende frequentieverdeling op (zie de tabel):

- a Teken een histogram van de hier gegeven frequentieverdeling.

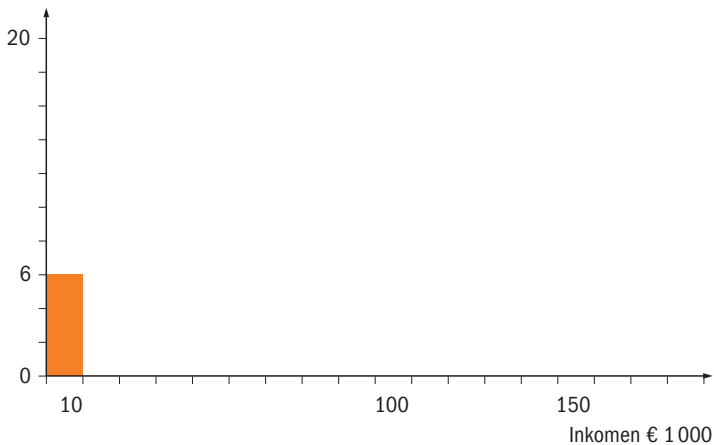
ANTWOORD

Om het histogram te kunnen tekenen, moeten eerst de frequentiedichtheden worden berekend. Allereerst moet er een standaardklassenbreedte worden gekozen. Hiervoor kiezen we een breedte van €5 000.

Inkomen (in euro)	Aantal
0 - < 10 000	12
10 000 - < 15 000	8
15 000 - < 20 000	12
20 000 - < 25 000	16
25 000 - < 30 000	20
30 000 - < 40 000	36
40 000 - < 50 000	32
50 000 - < 60 000	24
60 000 - < 100 000	30
100 000 - < 150 000	10
Totaal	200

De kolommen kunnen nu worden getekend.

Klasse	Aantal maal de standaardbreedte	Frequentie	Frequentiedichtheid
0 - < 10 000	2	12	$\frac{12}{2} = 6$
10 000 - < 15 000	...	8
15 000 - < 20 000	...	12
20 000 - < 25 000	...	16
25 000 - < 30 000	...	20
30 000 - < 40 000	...	36
40 000 - < 50 000	...	32
50 000 - < 60 000	...	24
60 000 - < 100 000	...	30
100 000 - < 150 000	...	10

FIGUUR 1.2

- b** Bereken de relatieve frequenties van de aangegeven klassen.

ANTWOORD

De relatieve frequenties worden berekend door deling van de absolute frequenties door het totaal aantal waarnemingen.

We vinden dan:

Inkomen (in euro)	Frequentie	Relatieve frequentie
0 - < 10 000	12	$\frac{12}{200} = 0061$
10 000 - < 15 000	8
15 000 - < 20 000	12
20 000 - < 25 000	16
25 000 - < 30 000	20
30 000 - < 40 000	36
40 000 - < 50 000	32
50 000 - < 60 000	24
60 000 - < 100 000	30
100 000 - < 150 000	10
Totaal	200

- 1.3** Geef voor elk van de volgende gevallen aan of de genoemde verzameling als een steekproef of als een populatie mag worden beschouwd:
- de commissarissen van de koning van de 12 Nederlandse provincies
 - de 200 personen die zijn geïnterviewd bij een straatenquête
 - de 150 automobilisten die moesten stoppen voor een alcoholcontrole
 - de 740 leden van een studentenvereniging
 - de 38 klanten die tussen 11:00 en 12:00 uur een postkantoor binnenkomen
 - de 12 000 verzekerden bij een verzekeringsmaatschappij
 - de 20 nummers die worden gedraaid in een muziekprogramma op de radio
- 1.4** Geef voor de volgende variabelen aan of deze een nominale, ordinale, interval- of ratioschaal heeft:
- de speelduur van een dvd
 - de kleur van tulpen
 - de industrietak waarin werknemers een baan hebben
 - de jaaromzet (in euro) van bedrijven
 - het aantal sterren dat de moeilijkheidsgraad van puzzelboekjes aangeeft
 - de hoogte boven de zeespiegel van wintersportdorpen
- 1.5** De tentamencijfers voor het vak Statistiek van 60 eerstejaarsstudenten zijn weergegeven in de volgende tabel:

Cijfers van 60 eerstejaarsstudenten te X

6	8	4	5	2	2	7	10	9	6	5	6	3	4	6
7	5	6	2	8	5	4	3	6	6	7	5	1	6	8
4	9	6	6	7	6	6	5	6	5	7	6	4	1	3
5	4	7	5	9	5	8	7	4	2	3	1	6	7	6

- Tel de frequenties per cijfer.
- Bereken de relatieve frequenties.
- Teken van de frequentieverdeling een staafdiagram.

- 1.6** Voor een aantal beursfondsen is op 31 december van een bepaald jaar vastgesteld hoe hoog de aandelenkoers is ten opzichte van de in het verstreken jaar bereikte bedrijfswinst per aandeel (de zogenoemde koers-winstverhouding K/W). Verder is berekend met welk percentage de omzet van het bedrijf is gegroeid in het verstreken jaar. Dit leverde de volgende cijfers op:

Bedrijf	K/W	Groei (%)	Bedrijf	K/W	Groei (%)
A	18,0	22	K	22,5	23
B	9,5	10	L	14,1	6
C	9,3	3	M	16,2	15
D	10,6	7	N	9,6	2
E	15,6	17	O	9,9	6
F	8,7	5	P	16,0	18
G	11,4	4	Q	17,7	24
H	12,7	14	R	11,8	5
I	10,4	8	S	13,3	12
J	13,9	15	T	9,2	2

- a** Maak een onderscheid tussen snelle groeiers (vanaf 10%) en langzame groeiers. Verdeel de koers-winstverhoudingen in klassen van 10 eenheden breed. Stel aan de hand van deze indeling een kruistabel op. Hoe groot is het percentage bedrijven dat een snelle groeier is en een lage K/W heeft? Hoeveel procent van de bedrijven met een K/W tussen 10% en 20% kan worden beschouwd als een snelle groeier?
- b** Teken een spreidingsdiagram voor de relatie tussen K/W en het groeipercentage.
- 1.7** Een groep van dertig eerstejaarsstudenten is een aantal vragen voorgelegd. Dit betrof:
- leeftijd
 - woonsituatie (z = zelfstandig, o = bij ouders)
 - geslacht (m = man, v = vrouw)
 - de maandelijkse bestedingen aan voedsel en drank
 - de score voor het tentamen statistiek
- In de volgende tabel staan de resultaten:

Gegevens van 30 eerstejaarsstudenten

Naam	Leeftijd	Woonsituatie	Geslacht	Besteding	Score
C. Been	18	z	v	510	72
S. Everwijn	23	o	m	280	62
Y. Anjema	23	z	v	440	49
J. van den Boom	21	o	m	490	58
T. van Tuyl	20	o	v	330	86
L. Herber	19	o	m	260	67
J. Derickx	22	o	m	420	95
L. Tielens	21	z	m	570	68
H. Eijgenhuijsen	20	z	m	530	72
L. de Wilde	21	z	v	590	63
H. van Woensel	20	o	m	460	52
O. Segers	19	z	v	480	73
K. de Bont	23	z	m	550	92
N. Morel	22	z	v	470	72

Gegevens van 30 eerstejaarsstudenten (vervolg)

Naam	Leeftijd	Woonsituatie	Geslacht	Besteding	Score
H. Moerland	18	o	v	380	45
R. Broersen	19	z	m	540	68
J. Beuving	22	z	v	390	46
E. Franken	26	o	m	450	38
A. de Groot	21	o	m	270	83
G. Rietkerk	22	o	m	390	86
L. Verhees	24	o	m	540	47
J. Buhrman	20	z	m	360	96
D. van Kraay	19	o	m	290	72
M. Eijsbouts	25	z	v	460	48
J. Wijbenga	20	o	m	310	81
R. de Leeuwe	21	z	m	570	84
C. Arends	19	o	m	280	51
M. Tijmstra	20	z	v	410	62
M. Jansen	22	z	v	510	78
C. zur Lage	21	o	v	330	46

Beantwoord nu de volgende vragen (eventueel nadat het gegevensbestand is ingevoerd in de computer):

- Geef aan op welk type schaal de vijf variabelen worden gemeten.
- Maak een frequentieverdeling van de leeftijden.
- Teken een histogram van de bestedingen. Begin met ondergrens €250.
- Maak een klassenindeling van de scores voor mannen en vrouwen afzonderlijk. Kies voor de klassen 10 eenheden en verwerk de resultaten in een kruistabel. Bereken ook de procentuele frequenties van de scoreklassen voor de mannen en vrouwen afzonderlijk.
- Maak een kruistabel waarin de waarnemingen worden verdeeld naar geslacht en woonsituatie.
- Teken een spreidingsdiagram met 'leeftijd' langs de horizontale as en 'score' langs de verticale as.

- 1.8** Bij een onderzoek naar de jaarinkomens van personen van 30 jaar oud met een voltijdse baan is de relatie bekeken tussen inkomen en opleidingsniveau. De volgende tabel was het resultaat:

Opleidingsniveau	Inkomen (* €1 000)				
	Minder dan 20	20 - < 30	30 - < 40	40 - < 50	50 of meer
Vmbo of lager	40	32	16	8	4
Mbo	18	22	14	10	6
Vwo	12	16	30	18	14
Hbo	4	8	16	34	18
Wo	0	6	14	26	34

- Bereken voor de tabel de totalen per rij en per kolom.
- Hoeveel procent van de hbo'ers verdient minstens 30 000 euro?
- Hoeveel procent van de hoogste inkomenscategorie heeft hbo of vwo als opleiding?
- Een advertentiecampagne vindt plaats in een blad dat uitsluitend wordt gelezen door hbo'ers en vwo'ers. Hoeveel procent van deze doelgroep heeft een inkomen lager dan 40 000 euro?

- 1.9** Er zijn twee nieuwe middelen op de markt gekomen ter bestrijding van hooikoorts. Een groot gezondheidscentrum heeft 120 patiënten geselecteerd aan wie middel A of middel B werd voorgeschreven. Deze keuze werd bepaald door het toeval (dus niet op basis van bewuste keuze door de arts). Na verloop van tijd werd de deelnemers gevraagd naar het resultaat. De resultaten waren als volgt:

Medicijn	Wel hooikoorts	Geen hooikoorts	Totaal
A	12	48	60
B	18	42	60

- a** Hoeveel procent van de patiënten krijgt hooikoorts bij gebruik van middel A? En bij B?
b Bereken het relatief risico op hooikoorts bij vergelijking van A en B.

- 1.10** Op de jaarlijkse Koningsmarkt in ons dorp verkopen vrijwilligers zelfgemaakte gerechten zoals erwtensoep en vissalade. Na afloop van de jaarmarkt bleek een fors aantal bezoekers last te hebben van een lichte voedselvergiftiging. De gemeentelijke gezondheidsdienst deed daarom onderzoek. Door 100 bezoekers werden vragenlijsten ingevuld. Daaruit bleek onder andere het volgende. Door 48 bezoekers werd zowel soep als salade geconsumeerd, 14 aten uitsluitend de soep, 22 aten uitsluitend de vissalade en door 16 bezoekers werd geen van beide producten genuttigd. Voor deze vier groepen werd vervolgens vastgesteld hoeveel personen er na de jaarmarkt fysieke klachten hadden. Het bleken er 32 te zijn.

Aantal ziektegevallen binnen groepen consumenten

	Vissalade	Geen vissalade	Totaal
Soep	20 / 48	2 / 14	62
Geen soep	8 / 22	2 / 16	38

Als vergelijkingsmaatstaf kiezen we de jaarmarktbezoekers die geen van beide producten hebben gegeten.

- a** Maak een tabel waarin de groep die uitsluitend soep at wordt vergeleken met de 16 bezoekers die geen van beide aten en bereken het relatief risico.
b Doe hetzelfde voor de groep die uitsluitend vissalade nuttigde.
c Doe hetzelfde voor de groep die beide gerechten consumeerde. Wat is je conclusie?
- 1.11** Bij de marathon van Rotterdam 2016 bereikten 12 814 deelnemers de finish. De deelnemerslijst met de finishtijden is helaas onvindbaar, maar wél is een tabel bekend met de aantallen deelnemers in klassen gedeeld. Met name is iets bekend over een aantal percentielen van de verdeling¹.

¹ Een percentiel is een punt op de x-as waar beneden zich een gegeven percentage van de waarnemingen bevindt. Dus als het 16^e percentiel 188 minuten zou bedragen, dan heeft 16% van de deelnemers een finishtijd lager of gelijk aan 188 minuten.

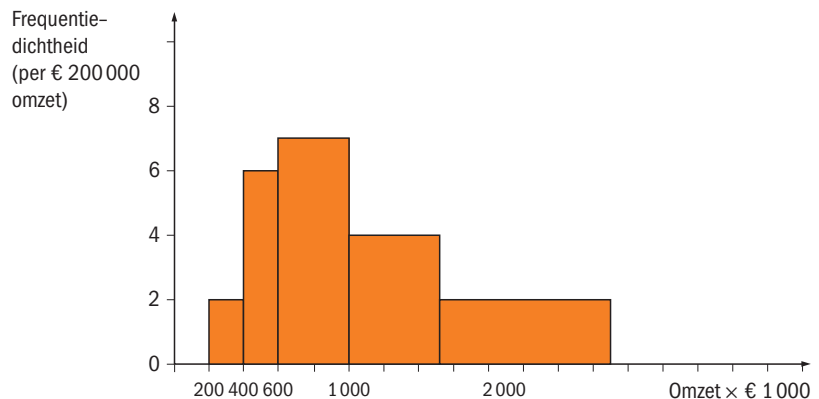
Enkele resultaten van de marathon van Rotterdam 2016

percentiel	Tijd in minuten
0 ^e (= winnaar)	126
10 ^e	196
20 ^e	212
30 ^e	224
40 ^e	235
50 ^e	243
60 ^e	253
70 ^e	264
80 ^e	277
90 ^e	293
95 ^e	306
100 ^e (= laatste plaats)	359

Construeer een histogram voor deze verdeling. (Hint: bedenk dat elke groep van 10% bestaat uit 1281 waarnemingen. De kolombreedtes in minuten verschillen echter).

- 1.12** De omzetcijfers van 42 bedrijven uit een bedrijfstak zijn weergegeven in het volgende histogram (figuur 1.3). Zet de gegevens om in een frequentie-tabel.

FIGUUR 1.3



- 1.13** De branduur van een bepaald merk lamp is geregistreerd voor 50 exemplaren. De resultaten zijn als volgt:

Branduur in uren

1 248	1 150	1 215	1 380	1 495	1 325	1 418	1 060	1 176	1 204
1 020	1 292	1 350	1 553	1 236	1 310	1 464	1 280	1 017	1 360
1 527	1 140	1 238	1 347	1 134	1 264	1 198	1 618	1 372	1 278
1 642	1 394	1 090	1 407	1 082	1 262	1 514	1 430	1 255	1 424
1 368	1 646	1 103	1 385	1 258	1 485	1 380	1 270	1 525	1 100

- a** Maak een indeling in klassen van 50 eenheden breed, te beginnen met de klasse 1 000 - < 1 050 en bereken de frequentie per klasse.
- b** Verwerk de onder vraag a ontstane frequentieverdeling in een histogram. Bedenk hierbij dat de brandduur (= tijd) een continue variabele is. De in de tabel gegeven waarden moeten in dit kader worden beschouwd als afrondingen naar het dichtstbijzijnde gehele urenaantal. (Dus 1 247 uur en 35 minuten komt als afgeronde waarde 1 248 in de tabel.)
- c** Teken een frequentiepolygoon van de verdeling.

- 1.14** Bij een bedrijf zijn in een bepaald jaar totaal 100 hoogopgeleide werknemers in dienst getreden, 50 mannen en 50 vrouwen. De maandsalarissen van deze werknemers zijn weergegeven in de volgende tabel:

Maandsalarissen van 100 nieuwe werknemers (in euro)

Mannen					Vrouwen				
2 850	4 850	3 910	2 270	4 980	1 850	2 380	2 870	3 290	5 100
3 400	4 780	3 370	3 850	3 260	2 350	3 870	3 360	2 130	1 970
3 750	5 900	3 740	3 490	4 760	2 720	2 410	2 350	2 800	1 890
2 250	4 000	4 130	2 350	5 500	3 270	2 560	3 280	2 650	4 210
3 150	3 670	4 590	4 200	5 100	3 850	3 140	2 100	2 910	3 620
2 960	5 980	2 250	3 240	4 350	2 610	3 650	2 480	1 990	4 410
2 230	2 840	4 480	2 580	2 350	2 370	2 830	3 720	2 180	4 220
3 555	2 170	3 450	5 280	5 420	2 660	4 670	1 950	4 150	3 250
2 430	3 390	2 560	5 650	2 740	4 140	3 560	2 980	3 900	3 170
3 610	2 490	5 750	5 360	2 690	4 850	5 350	3 490	3 360	4 280

Construeer voor deze gegevens een verticaal tweezijdig stamdiagram:

- a** met klassen van €1 000 breed
- b** met klassen van €500 breed

- 1.15** Bij een verkeerscontrole zijn de banden van 200 auto's gecontroleerd. Hierbij werd het profiel gemeten in mm. Een en ander leidde tot de volgende frequentieverdeling (zie de tabel).

Gemeten profiel (in mm)	Aantal auto's
0,0 - < 2,0	4
2,0 - < 4,0	34
4,0 - < 6,0	82
6,0 - < 8,0	66
8,0 - < 10,0	14
Totaal	200

- a** Geef de waargenomen verdeling weer door middel van een cumulatieve frequentiecurve.
- b** Probeer met de grafiek te schatten hoe groot het aantal auto's is met een profiel van minder dan 5,00 mm. (NB In deze opgave mag men veronderstellen dat de klassengrenzen exact zijn, dus er hoeven geen correcties met 'halfjes' uitgevoerd te worden.)

- 1.16** Bij een grote onderneming hebben veel werknemers een deeltijdaanstelling. Bij een onderzoek werd het verband bestudeerd tussen de leeftijd van werknemers en het percentage deeltijdaanstelling. (Dus 100% betekent fulltimewerknemers, 60% betekent een aanstelling voor drie dagen per week enzovoort). Er werden twee categorieën werknemers onderscheiden, namelijk HO (hoger opgeleiden) en LO (lager opgeleiden).

Voor de 200 hoger opgeleide werknemers leverde dat de volgende tabel:

Leeftijd	Deeltijd	100% werktijd
20 - < 30	2	8
30 - < 40	25	35
40 - < 50	30	40
50 - < 60	27	33

Voor de 202 lager opgeleide werknemers leverde dat:

Leeftijd	Deeltijd	100% werktijd
20 - < 30	55	40
30 - < 40	42	28
40 - < 50	20	10
50 - < 60	5	2

- a Bereken bij de hoger opgeleiden voor iedere leeftijdscategorie het percentage werknemers dat voltijds (100%) werkt.
- b Idem voor de lager opgeleiden.
- c Voeg de tabellen samen tot een tabel die over alle werknemers de aantallen aangeeft verdeeld naar leeftijd en werktijd. Bereken opnieuw per leeftijdscategorie het percentage dat voltijds werkt. Wat valt op als je dit vergelijkt met de antwoorden op vraag a en b?
- 1.17** Het bedrijf voor het schildersbedrijf heeft onderzocht hoe het is gesteld met het buitenschilderwerk van de koopwoningen in een grote gemeente. Voor een steekproef van 250 huizen werd vastgesteld hoe lang het geleden is dat het buitenschilderwerk was gedaan. Dat leverde de volgende tabel, waarvan helaas ten gevolge van een computerstoring nogal wat gegevens onleesbaar waren:

Aantal jaren sinds de laatste schilderbeurt	Waargenomen aantal	Relatieve frequentie	Cumulatieve frequentie
0 - < 2 jaar	?	?	?
2 - < 4 jaar	?	0,14	80
4 - < 6 jaar	75	?	?
6 - < 8 jaar	60	?	?
8 - < 10 jaar	?	?	235
10 - < 12 jaar	?	?	?
Totaal	250	1,00	

- a** Vul de ontbrekende getallen in.
b Teken een histogram en een frequentiepolygoon van de verdeling.

1.18

Voor echte liefhebbers van wielrennen is het beklimmen van de Alpe d'Huez een van de grootste uitdagingen die men zich kan indenken. Het is een ideaal waar sommigen lange tijd naar toe werken, door veel te trainen. Een wielclub is met 50 leden deze berg gaan beklimmen. Voor alle deelnemers werd het aantal minuten genoteerd dat men nodig had voor de klim. Verder werd vermeld of een deelnemer man of vrouw is. Dat leverde de volgende gegevens:

Tijd m/v	Tijd m/v	Tijd m/v	Tijd m/v	Tijd m/v
58 m	73 m	79 v	86 m	96 v
63 m	73 m	80 v	87 v	98 v
64 m	74 v	81 m	88 m	99 m
64 m	75 m	81 m	89 v	101 v
66 m	76 v	82 v	90 m	101 m
67 v	76 m	82 m	91 m	104 v
67 m	76 m	82 m	92 m	107 v
68 m	77 v	83 v	92 m	109 v
69 m	77 m	83 v	93 v	126 v
72 m	79 m	84 m	94 v	138 m

- a** Maak een frequentieverdeling van alle 50 tijden. Maak klassen van 10 eenheden breed.
b Maak frequentieverdelingen voor mannen en vrouwen afzonderlijk.
c Bereken de relatieve frequenties voor beide verdelingen van vraag **b**. Teken in één figuur de histogrammen van de relatieve frequentieverdelingen.

1.19

Een bank onderzoekt het spaargedrag van haar cliënten. Van 500 willekeurige cliënten is vastgesteld hoeveel middelen door hen aan de bank zijn toevertrouwd. De resultaten zijn te zien in de tabel.

- a** Bereken voor de klassen de relatieve frequenties en de frequentiedichtheden.
b Teken een histogram van de verdeling.
c Teken een frequentiepolygoon.
d Teken een cumulatieve frequentiecurve.

Inleg (in euro)	Aantal
0 - < 500	80
500 - < 1 000	65
1 000 - < 2 000	75
2 000 - < 5 000	60
5 000 - < 10 000	100
10 000 - < 25 000	120
Totaal	500

1.20

Een groot ziekenhuis houdt een klanttevredenheidsonderzoek onder patiënten die minstens twee dagen aaneengesloten in het ziekenhuis hebben doorgebracht.

De top 10 van klachten is weergegeven in de volgende tabel:

Klacht	Aantal respondenten
Andere patiënten gedragen zich storend	12
Artsen verschijnen op onvoorspelbare tijden	6
Bedden worden niet goed verschoond	4
Bezoektijden zijn te beperkt	32
Herrie vanwege een verbouwing	18
Maaltijden zijn onsmakelijk	14
Onvriendelijke verplegers	8
Radio/tv voldoet niet	22
Verplegers reageren niet op alarmsignaal	4
Vroeg gewekt worden	44
Totaal	164

- Maak een Paretdiagram van deze gegevens.
- Geef aan door welke klachten de belangrijkste groep van ongeveer 80% van de verdeling wordt gevormd.
- Welke zaken zou de directeur van het ziekenhuis met voorrang moeten aanpakken teneinde de tevredenheid te bevorderen?

1.21 Voor een bedrijfstak is onderzocht hoe de verdeling is van de totale omzet van de bedrijfstak over de bedrijven. Het resultaat wordt weergegeven in de cumulatieve tabel, waarbij de bedrijven zijn geordend naar omzetgrootte.

- Geef de informatie uit de tabel weer door middel van een Lorenzcurve.
- Hoe verhoudt zich de gemiddelde omzet van de 10% grootste bedrijven (qua omzet) tot de gemiddelde omzet van de 10% kleinste bedrijven?

Percentage van de bedrijven	Percentage van de totale omzet
10	1
20	5
30	10
40	17
50	25
60	35
70	46
80	58
90	70
100	100

1.22 Open in Excel het bestand 'MULTIFOOD'². Maak een grafiek van de omzet (Y_t) als functie van de tijd (T). Selecteer de eerste twee kolommen van het bestand en gebruik de 'Wizard Grafieken'. Kies bij de eerste stap als grafiektype 'Spreiding' en als subtype de vierde optie: 'Spreiding met gegevenspunten' verbonden via lijnen. In de derde stap als grafiektitel opgeven 'Tijddiagram Multifood' en bij de assen respectievelijk *tijd (kwartaalnummer)* en *omzet (in mln euro)*.

1.23 Open in Excel het bestand 'BANKIERS'.

- Beschrijf van werknemer 19 het opleidingsniveau en het functieniveau.
- Maak een histogram van de variabele *dienstjaren*. Ga als volgt te werk:

² Zie hoofdstuk 17 voor een beschrijving van de bestanden.

- Voeg een nieuw werkblad in en vermeld hier de klassengrenzen (0, 5, 10 enzovoort). Plaats de tekst *dienstjaren* erboven.
- Kies in het menu 'Extra → Gegevensanalyse' de optie 'Histogram'.
- Vul het dialoogvenster in en vergeet niet om de optie 'Grafiek maken' aan te vinken.
- Klik op de rand van het grafiekgebied en vergroot dit door te slepen met de rechteronderhoek. Zorg dat de grafiek een bruikbare grootte en vorm krijgt.
- Klik met de rechtermuisknop op een van de staven en kies in het snelmenu de optie 'Gegevensreeks opmaken'. Op het tabblad 'Opties' kan de tussenruimte van de kolommen worden ingesteld. Vanwege de vorm van het histogram moet de breedte tussenruimte op 0 worden ingesteld.

Het histogram is nu klaar!

- c** Maak een kruistabel waarin het opleidingsniveau en het functieniveau tegen elkaar zijn afgezet. Ga als volgt te werk:
- Start de 'Wizard Draaitabellen' vanuit het menu 'Data → Draaitabelrapport'
 - Plaats (bij stap 3 van de wizard) de variabelen *functie* en *niveau* in respectievelijk het rijvak en het kolomvak. Sleep een van de variabelen in het vak gegevens en verander *Som van ...* in *Aantal van ...* (door te dubbelklikken op het blokje).
 - Laat de draaitabel op een nieuw werkblad maken (stap 4 van de wizard). De draaitabel is nu gereed! Beantwoord met behulp van de draaitabel de volgende vragen:
 - Wat is binnen de groep van personen met opleidingsniveau 4 de verdeling van de functieniveaus?
 - Wat is binnen de groep met functieniveau 4 de verdeling van opleidingsniveaus?
- d** Verander in de kruistabel waarin het opleidingsniveau en het functieniveau tegen elkaar zijn afgezet de samenvattingsfunctie in *Gemiddelde van ...* en kies de variabele *salaris* om weer te geven. Kun je het resultaat verklaren?

1.24 Open in Excel het bestand 'KOERSEN'.

- a** Maak via de 'Wizard Grafieken' een spreidingsdiagram, waarbij de variabelen *dollar* en *olie* worden gekozen. Welke globale conclusie kan men hier trekken over de relatie tussen deze variabelen?
- b** Beantwoord dezelfde vragen voor de relatie tussen de variabelen *olie* en *dowjones*.
- c** Beantwoord dezelfde vragen voor de relatie tussen de variabelen *dollar* en *dowjones*.

1.25 Open in Excel het bestand 'WINKELS'.

- a** Beschrijf nauwkeurig de gegevens voor winkel 34.
- b** Maak een staafdiagram van de verdeling over de verschillende branches.
Hint: maak eerst met behulp van het menu 'Data → Draaitabelrapport' een frequentietabel van de branches.
- c** Maak een spreidingsdiagram waarin *vloeroppervlak* en *jaarhuur* tegen elkaar worden afgezet.
Hint: vervang eerst de waarde M in de kolom jaarhuur door niets.

1.26 Open in Excel het bestand 'CURSUS'.

- a Hoeveel punten behaalde kandidaat 29 bij de twee tentamens en hoeveel uren heeft deze student besteed?
- b Teken een histogram van de behaalde cijfers. Kies als grenzen 40, 50 enzovoort. (NB Excel telt altijd frequenties inclusief de bovengrens.)
- c Maak met een draaitabel een uitsplitsing van de scores voor evaluatie 1 en 2.
- d Bekijk of de verdelingen van ev_1 en ev_2 verschillen voor mannen en vrouwen.

1.27 Open in Excel het bestand 'WONINGEN'.

- a Maak een spreidingsdiagram met op de horizontale as het *bouwjaar* en langs de verticale as de *grondoppervlakte*. Wat valt op?
- b Maak een kruistabel van *wijk* en *garage*. In welke wijk hebben relatief de meeste huizen een garage?
- c Bekijk (aansluitend op vraag b) het gemiddeld aantal kamers, uitgesplitst naar wijk en wel of geen garage. (Dus kies in de draaitabel voor het gemiddelde.) Beschrijf in alledaagse bewoordingen je conclusie.

