

# Basisboek kwantitatieve methoden

## Statistiek met Exceltoepassingen

Uitwerkingen van de opgaven

Tweede, herziene druk

Donald van As  
Jaap Klouwen

u i t g e v e r i j | C  
c o u t i n h o

bussum 2013

Deze uitwerkingen horen bij de tweede, herziene druk van *Basisboek kwantitatieve methoden – Statistiek met Exceltoepassingen* van Donald van As en Jaap Klouwen.

© 2007 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 16 h Auteurswet 1912 dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan Stichting Reprorecht (Postbus 3051, 2130 KB Hoofddorp, [www.reprorecht.nl](http://www.reprorecht.nl)). Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16h Auteurswet 1912) kan men zich wenden tot Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, Postbus 3060, 2130 KB Hoofddorp, [www.stichting-pro.nl](http://www.stichting-pro.nl)).

Eerste druk 2007

Tweede, herziene druk 2013

Uitgeverij Coutinho

Postbus 333

1400 AH Bussum

[info@coutinho.nl](mailto:info@coutinho.nl)

[www.coutinho.nl](http://www.coutinho.nl)


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

# Inhoud

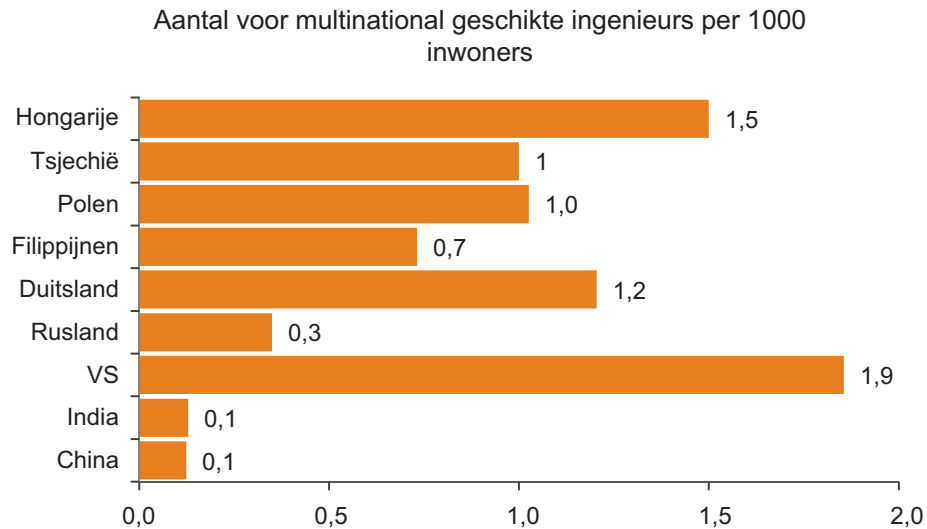
	1	Grafieken en tabellen in de beschrijvende statistiek	5
	2	Maatstaven in de beschrijvende statistiek	11
	3	Elementaire kansrekening	21
	4	Kansvariabelen	27
	5	De binomiale kansverdeling	33
	6	De normale verdeling	36
	7	Lineaire regressie	41



# 1 Grafieken en tabellen in de beschrijvende statistiek



- 1
  - a Het doel van het onderzoek is geweest: nagaan in hoeverre in 2012 een verband bestond tussen de koersen van de AEX en Dow Jones.
  - b De koers van de AEX zal de afhankelijke variabele zijn en de koers van de Dow Jones de onafhankelijke variabele: we gaan ervan uit dat de Amerikaanse economie leidend is.
  - c De waarden die de Dow Jones aanneemt zijn veel hoger dan de waarden die de AEX aanneemt, waardoor de ontwikkeling van de twee variabelen moeilijk in één grafiek is aan te geven.
  - d Het probleem van de zogenaamde *time lag*: de handel op de AEX vindt uren eerder plaats dan die op de Dow Jones, als we uitgaan van dezelfde datum. Het is dus van belang dat de gegevens uit Nederland worden vergeleken met die van de Dow Jones van de vorige dag of zelfs nog een dag eerder, teneinde uitspraken te kunnen doen over een verband.
  - e Uit de grafiek is duidelijk te zien dat de koersen van mei tot september stijgen, hetgeen inhoudt dat men de genoemde beurswijsheid beter *niet* kan opvolgen.
  - f Op het oog lijkt er een stevig verband te zijn tussen de koersen van de AEX en die van de Dow Jones. Het lijkt er ook op dat de koersschommelingen van de AEX heftiger zijn geweest dan die van de Dow Jones.
  
- 2
  - a Het bovenste gedeelte van de staven blijft qua oppervlakte vrijwel gelijk. Dit betreft de grotere uitvaartbedrijven. De groei zit uitsluitend in het onderste gedeelte van de staven, waarmee de aantallen kleine bedrijven worden aangegeven.
  - b Met een intervallschaal; er is hier immers niet echt sprake van een natuurlijk nulpunt, maar wel van gelijke afstanden tussen de jaartallen.
  - c De meest voor de hand liggende verklaring is de behoefte van de consument aan meer gespecialiseerde uitvaartplechtigheden, meer op speciale wensen toegesneden ceremonies.
  - d Met behulp van een lijndiagram. Een gemiste kans in de gegeven grafiek.
  
- 3
  - a 'Aantal' is een ratioschaal. (Merk op dat de variable 'land' een nominale schaal heeft: er is hier geen enkele ordening mogelijk.)
  - b De omvang van de categorie 'geschikt personeel' delen door de omvang van de betreffende bevolking.



c Een verhoudingsgetal is altijd een ratioschaal.

4 a Met één blik in de rechter grafiek is te zien dat Nederland van de acht genoemde landen op de zesde plaats staat met een groei van circa 2,2%. In de vs ligt dat percentage bijna een vol punt hoger.

b Over welk tijdsbestek is de groei gegeven? Gaat het over de gehele periode van vijf à zes jaar, of gaat het om een jaarlijkse groei?

c We gaan eerst uit van een productiviteitsgroei over de gehele periode en dat de linkergrafiek geldt voor het jaar 2003/2004. Tussen haakjes staan de rangnummers voor 2009.

vs (1): circa  $39 \times 1,032 \approx 40,2$

Nederland (2): circa  $33 \times 1,022 \approx 33,7$

Frankrijk (3): circa  $32 \times 1,018 \approx 32,6$

Zweden (4): circa  $31 \times 1,028 \approx 31,9$

Denemarken (6): circa  $30 \times 1,017 \approx 30,5$

Duitsland (5): circa  $30 \times 1,024 \approx 30,7$

Groot-Brittannië (7): circa  $25 \times 1,028 = 25,7$

Finland (8): circa  $23 \times 1,03 \approx 23,7$

Gaan we echter uit van een productiviteitsgroei *per jaar*, dan ziet het plaatje er al weer anders uit:

vs (1): circa  $39 \times 1,032^5 \approx 45,7$

Nederland (2): circa  $33 \times 1,022^5 \approx 36,8$

Frankrijk (4): circa  $32 \times 1,018^5 \approx 35,0$

Zweden (3): circa  $31 \times 1,028^5 \approx 35,6$

Denemarken (6): circa  $30 \times 1,017^5 \approx 32,6$

Duitsland (5): circa  $30 \times 1,024^5 \approx 33,8$

Groot-Brittannië (7): circa  $25 \times 1,028^5 \approx 28,7$

Finland (8): circa  $23 \times 1,03^5 \approx 26,7$

- d In de ranglijsten die in c. zijn ontstaan is niet echt een groot risico te zien. Een vergelijking met voornamelijk EU-landen is blijkbaar niet zo duidelijk als illustratie voor het genoemde risico.

- 5 a Fixed-income capital markets: circa  $100/395 \times 44,4$  ofwel circa 11,2 miljard dollar.

Equity capital markets: circa  $100/120 \times 27,0$  ofwel circa 22,5 miljard dollar.

b

	2000	2001	2002	2003	2004	2005	2006
Fixed-income	100	130	150	205	230	290	395
Equity	100	80	60	65	70	90	120

- c Fixed income:  $395/150 \times 100\% = 263\%$ , dus circa 163% groei.

Equity:  $120/60 \times 100\% = 200\%$ , dus circa 100% groei.

d

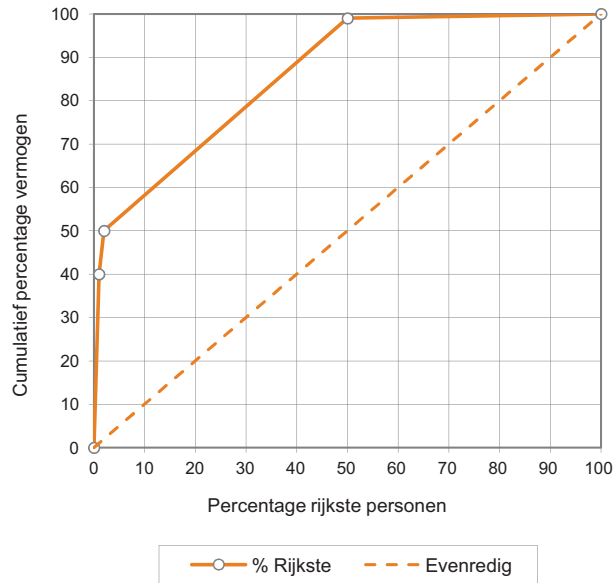
	2000	2001	2002	2003	2004	2005	2006
Fixed-income	100	130	150	205	230	290	395
	49	63	73	100	112	141	193
Equity	100	80	60	65	70	90	120
	154	123	92	100	108	138	185

- e Staafdiagram met steeds twee staven, voor elke grootheid één.

- 6 a Een lijndiagram, een cirkeldiagram en een staafdiagram. De wereldkaart is een afgeleide vorm van een zogeheten *choropleet*. Daarin worden statistische gegevens in verschillende tinten geografisch afgebeeld. Op de horizontale as van de grafiek staat het cumulatieve percentage van de wereldbevolking, gesorteerd van rijk naar arm. Op de y-as staat het gemiddeld vermogen in dollar.

- b Maak onderstaande tabel en daarmee de Lorenzcurve.

x	y
% rijkste personen	% vermogen
0	0
1	40
2	50
50	99
100	100



- c Zie onderstaande tabel. De oppervlakte onder de Lorenzcurve is dus 0,8616 (zie appendix A11). De Ginicoëfficiënt voor inkomen wordt daarmee  $2 \times 0,8616 - 1 = 0,72$ .

b	$\bar{h}$	$b \times \bar{h}$
0,01	0,2	0,002
0,01	0,45	0,0045
0,48	0,745	0,3576
0,50	0,995	0,4975
		0,8616

- d Zie onderstaande tabellen.

b	$\bar{h}$	$b \times \bar{h}$
0,01	0,1	0,001
0,04	0,275	0,011
0,06	0,405	0,0243
0,14	0,565	0,0791
0,25	0,77	0,1925
0,50	0,935	0,4675
		0,7754

De oppervlakte onder de Lorenzcurve is dus 0,78 (zie appendix A11). De Ginicoëfficiënt voor inkomen wordt daarmee  $2 \times 0,7754 - 1 = 0,55$ .

Op soortgelijke manier volgt de Ginicoëfficiënt voor de belastingopbrengst:  $2 \times 0,89045 - 1 = 0,78$ .



$b$	$\bar{h}$	$b \times \bar{h}$
0,01	0,19	0,0019
0,04	0,485	0,0194
0,06	0,645	0,0387
0,14	0,78	0,1092
0,25	0,915	0,22875
0,50	0,985	0,4925
		0,89045

- 7 a** (1) Regio, (2) consumptie dierlijk eiwit en (3) bevolking per regio.  
**b** Variabele (1) is discreet, (2) en (3) zijn continu. Variabele (1) is nominaal, (2) en (3) zijn ratio.  
**c** Op de horizontale as staat een cumulatieve grootheid.  
**d** De oppervlaktes van de regio's geven het aandeel in de dierlijke consumptie aan.  
**e** Lees de schatting uit onderstaande tabel af: 15%.

Regio	Eiwit (kg)	Aantal	Aandeel	Aandeel
	per hoofd bev.	miljard	mld kg	%
Noord-Amerika	26	0,3	7,8	11
Europa	22	0,5	11	15
Rijk Azië	19	0,2	3,8	5
Russische regio	16	0,3	4,8	7
Latijns-Amerika	14	0,6	8,4	12
China	12	1,4	16,8	23
Arm Azië	6	2,4	14,4	20
Afrika	5	1	5	7
<i>Totaal</i>		<i>6,7</i>	<i>72,0</i>	<i>100</i>

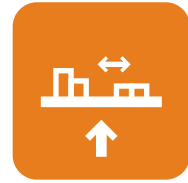
- f** Voor elke regio  $i$  geeft  $b_i \times d_i$  de totale eiwitconsumptie van die regio (tevens de oppervlakte van de staaf). De totale wereldconsumptie bedraagt dus  $\sum b_i \times d_i$ . Delen we dit door de totale wereldbevolking ( $\sum b_i$ , eenvoudig af te lezen op de horizontale as), dan vinden we de gemiddelde eiwitconsumptie per wereldburger. In 2007:  $\bar{d} = (72,0 \text{ mld kg}) / (6,7 \text{ mld personen}) = 10,7 \text{ kg p.p.}$

- g** Zie onderstaande tabel voor 2030. Toename consumptie dierlijk eiwit t.o.v. 2007:  $101,6/72,0 = 1,43$  dus 41%. Toename bevolking:  $8,3/6,7 = 1,24$  dus 24%. De toename van de consumptie van dierlijk eiwit kent dus de grootste stijging.

Regio	Eiwit (kg)	Aantal	Aandeel	Aandeel
(2030)	per hoofd bev.	miljard	mld kg	%
Noord-Amerika	28	0,4	11,2	11
Europa	23	0,5	11,5	11
Rijk Azië	19	0,2	3,8	4
Russische regio	17	0,3	5,1	5
Latijns-Amerika	19	0,8	15,2	15
China	14	1,5	21,0	21
Arm Azië	8	3,1	24,8	24
Afrika	6	1,5	7,8	9
<i>Totaal</i>		<i>8,3</i>	<i>102,6</i>	<i>100</i>

- 8 a** Lijndiagram.
- b**  $m = 100 - j$ ;  $m$  = percentage geslaagde meisjes;  $j$  = percentage geslaagde jongens.
- c** 1996.
- d** Als  $m = j$  geldt dus  $m = 100 - m$ , dus  $2m = 100$  ofwel  $m = 50$  (%).
- 9 a** Er is een toename van 50 naar 140 in indexcijfer, dus een groeifactor van  $140/50 = 2,8$ . Dit correspondeert met een procentuele toename van 180%.
- b** Het basisjaar is het jaar 2000, want daar zijn alle indexwaarden gelijk aan 100.
- c** Nee, want ter linkerzijde van de grafiek staat: index 2003 = 100. Dat zou betekenen dat het basisjaar 2003 is.
- 10 a** Het is niet gezegd dat het eten van chocolade het aantal Nobelprijswinnaars verhoogt. Het omgekeerde, dat de aanwezigheid van veel Nobelprijswinnaars ervoor zorgt dat er meer chocolade wordt gegeten, is ook niet bijzonder aanmerkelijk.
- b** Ja, want een correlatie tussen twee variabelen betekent nog geen oorzaak-gevolgrelatie.
- c** De keuze van het sterk stijgende model in figuur 1.29 (een exponentieel model) is suggestief.
- d** Ja, dat is zeker mogelijk. Dat zou op zijn beurt weer onderzocht kunnen worden met een spreidingsdiagram van die twee variabelen.
- e** Er moet een duidelijke, geaccepteerde *verklaring* bestaan die gevolg aan oorzaak koppelt.

## 2 Maatstaven in de beschrijvende statistiek



- 2.1**
- a** 36,39 (euro)
  - b** 36,90
  - c** 37,30 (deze waarde komt twee keer voor)
  - d** Eerste kwartiel is 35,68. Met Excel: QUARTILE(A1:J1;1)  
Derde kwartiel is 37,25. Met Excel: QUARTILE(A1:J1;3)
  - e**  $P_{60} = \text{PERCENTILE}(A1:J1;60\%) = 37,04$
  - f**  $\text{TRIMMEAN}(A1:J1;40\%) = 36,70$   
NB De twee laagste en de twee hoogste koersen worden verwijderd.
  - g** De negen rendementen, en daaruit bepaalde groeifactoren zijn:

Rendementen:

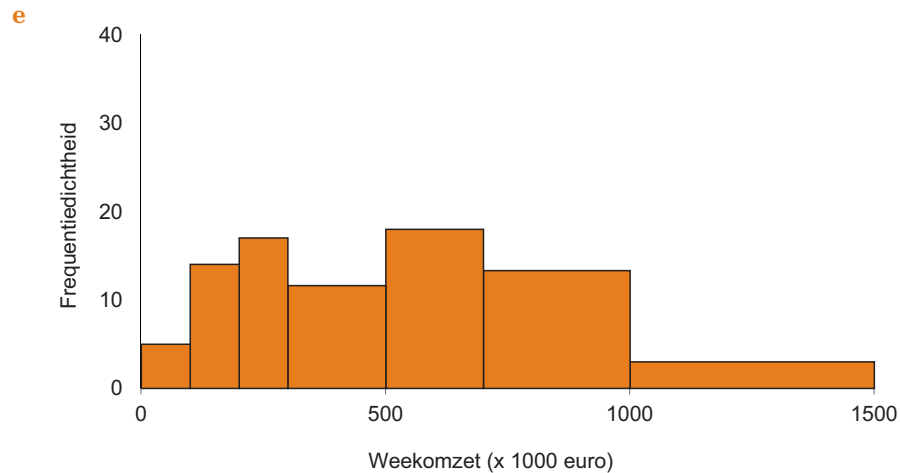
0,0179   0,0173   0,0600   -0,0048   0,0000   -0,0137   -0,0046   0,0104   0,0027

Groeifactoren:

1,0179   1,0173   1,0600   0,9952   1,0000   0,9863   0,9954   1,0104   1,0027

Het meetkundige gemiddelde is 1,0092, dus een gemiddeld percentage van Dit antwoord is ook te berekenen door  $\sqrt[9]{37,10/34,15} = 1,0092$ , ofwel de negendemachtswortel uit de laatste gedeeld door de eerste koers.

- 2.2**
- a** 111,9 (uur)
  - b**  $110 + (50 - 43,22)/(73,73 - 43,22) \times 10 = 112,2$  (lineaire interpolatie; zie paragraaf 2.2.1, deel 7 (kwartielen en percentielen))  
(De cumulatieve relatieve percentages van de klassengrenzen zijn 19,49%; 43,22%; 73,73%; 94,92% en 100%.)
  - c** De derde klasse is de modale klasse, met de grootste frequentie: 36.
  - d**  $Q_1 = 100 + (25 - 19,49)/(43,22 - 19,49) \times 10 = 102,3$   
 $Q_3 = 120 + (75 - 73,73)/(94,92 - 73,73) \times 10 = 120,6$
  - e**  $P_{30} = 100 + (30 - 19,49)/(43,22 - 19,49) \times 10 = 104,4$
- 2.3**
- a** 601,0 ( $\times \text{€ } 1000$ )
  - b** De vijfde klasse, met de grootste frequentiedichtheid: 18 (basis € 100.000).
  - c** Mediaan =  $P_{50} = 500 + (50 - 39,33)/(63,33 - 39,33) \times 200 = 588,9$
  - d** Derde kwartiel ( $Q_3$  oftewel  $P_{75}$ ) is  $700 + (75 - 63,33)/(90 - 63,33) \times 300 = 831,3$



- 2.4 a De frequenties zijn: 65, 75, 100, 125, 70, 40 en 25.  
 b De eerste klasse (zie de vijfde kolom in tabel hieronder), met de grootste relatiefrequentiedichtheid: 0,052.  
 c 19,5 belminuten.  
 d  $\text{Mediaan} = 10,5 + (50 - 48)/(73 - 48) \times 10 = 11,3$   
 e  $Q_1 = 2,5 + (25 - 13)/(28 - 13) \times 3 = 4,9$   
 $Q_3 = 20,5 + (75 - 73)/(87 - 73) \times 10 = 21,9$   
 f  $P_{35} = 5,5 + (35 - 28)/(48 - 28) \times 5 = 7,3$   
 g  $((5,5/10) \times 0,25 + 0,14 + (14,5/30) \times 0,08) \times 100\% = 31,6\%$

Aantal belminuten	Relatieve frequentie	Midden	Breedte	Relatieve-frequentie-dichtheid	Cumulatieve relatieve frequentie
0 t/m 2	0,13	1,25	2,5	0,052	0,13
3 t/m 5	0,15	4	3	0,050	0,28
6 t/m 10	0,20	8	5	0,040	0,48
11 t/m 20	0,25	15,5	10	0,025	0,73
21 t/m 30	0,14	25,5	10	0,014	0,87
31 t/m 60	0,08	45,5	30	0,003	0,95
61 t/m 180	0,05	120,5	120	0,0004	1

2.5 a

Bedrag in euro	Cumulatieve frequentie	Cumulatieve relatieve frequentie (%)
0 -< 20	57	14,25
20 -< 30	127	31,75
30 -< 40	211	52,75
40 -< 60	323	80,75
60 -< 100	400	100

- b** Mediaan =  $30 + 73 \times 10/84 = 38,69$  of  $30 + (50 - 31,75)/(52,75 - 31,75) \times 10 = 38,69$ ; afgerond: 39 (euro).
- c** Het aantal klanten dat minder dan € 45 per bezoek uitgeeft:  $211 + (45 - 40)/(60 - 40) \times 112 = 239$  (klanten). In een percentage dus  $(239/400) \times 100\% = 59,75\%$ .
- d**  $P_{45} = 30 + (45 - 31,75)/(52,75 - 31,75) \times 10 = 36,31$  (euro). Hier wordt gevraagd hoeveel euro de 180<sup>e</sup> klant (45<sup>e</sup> percentiel) uitgeeft en bij de vorige vraag betrof de vraag hoeveel mensen minder dan 45 euro uitgeven.
- e** De modale klasse is de klasse met de grootste frequentiedichtheid; de frequentiedichtheden per € 10 zijn hier respectievelijk 28,5; 70; 84; 56 en 19,25, dus de modale klasse is de derde klasse (30 -< 40).

**2.6 a**

Klasse	Bovengrens	Frequentie
1	1,77	19
2	1,79	30
3	1,81	36
4	1,83	34
5	1,85	28
6	1,87	3
7	> 1,87	0

- b** 1 AVERAGE(A1:F25) = 1,809  
 2 MEDIAN(A1:F25) = 1,810  
 3 MODE(A1:F25) = 1,810  
 4 QUARTILE(A1:F25;3) = 1,830  
 5 PERCENTILE(A1:F25;90%) = 1,841

**2.7 a**  $R = 37,48 - 34,15 = 3,33$ 

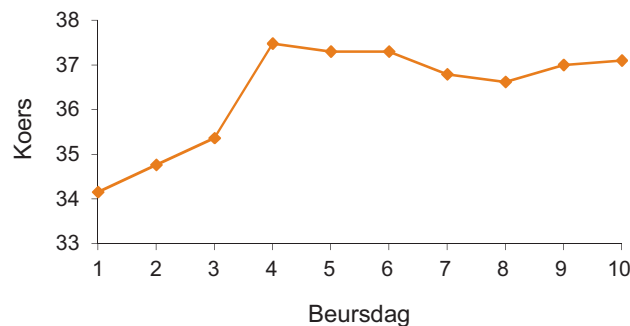
**b**  $Q_3 - Q_1 = 37,25 - 35,68 = 1,57$

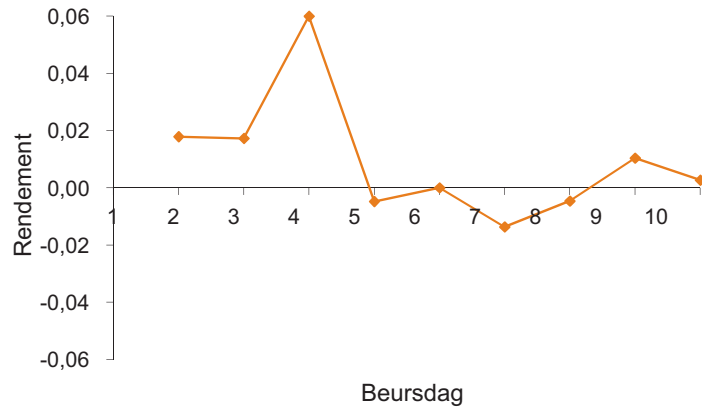
**c**  $GAA = 0,98$  (Excel: AVEDEV(A1:J1))

**d**  $s = 1,19$  (Excel: STDEV(A1:J1))

**e** scheefheid = SKEW(A1:J1) = -1,1

**f**





- g** De standaardafwijking en de GAA, omdat die beide alle waarnemingen hebben betrokken in een berekening van de spreiding van de koersen.

**2.8 a** 50 (uur)

**b** 11,6

**c** 134,53

**d**  $V = s/\bar{x} = 11,60/111,9 = 0,10$

NB Een variatiecoëfficiënt heeft geen 'dimensie' ('eenheid') (hier: uren gedeeld door uren).

**e** 0,07 (dimensieloos)

**f** -0,90 (dimensieloos)

**g** (88,67; 135,06). Ongeveer 95% van de batterijen heeft een levensduur die in dit interval ligt.

**2.9 a**  $Q_3 - Q_1 = 831,25 - 313,04 = 518,2$  ( $\times \text{€ } 1000$ )

**b**  $s = 329,61$

**c**  $V = 329,61/601 = 0,548$

**d**  $Q_1 = x_{37,5} = 300 + 1,5 \cdot 200/23 = 313,0$

Boxplot:



0 300 600 900 1200 1500

Uitschieters: boven  $2,5Q_3 - 1,5Q_1 = 1609$

**2.10 a**  $R = 180$  (NB Eigenlijk 180,5, omdat de laatste klasse formeel tot 180,5 loopt.)

**b**  $s = 26,08$

**c** scheefheid is 2,91, dus rechtsscheef

**d** spitsheid is 8,53, dus spitse verdeling

- 2.11 a**  $Q_1 = 20 + (100 - 57) \cdot (10/70) = 26,1$ ;  $Q_3 = 40 + (300 - 211) \cdot (20/112) = 55,9$   
De kwartielafstand is dus  $Q_3 - Q_1 = 29,8$  (euro).
- b** De standaarddeviatie (standaardafwijking) is 22,44.
- c** De variantie is 503,6 (kwadraat van antwoord b).
- d** De spitsheid is  $-0,79$  (met behulp van de file *maatstaven.xlsx*), dus spitsers dan een normale verdeling met hetzelfde gemiddelde en standaardafwijking (42,6 respectievelijk 22,44).
- e** De mate van symmetrie wordt bepaald door de scheefheid; deze is 0,38, dus rechtsscheef.
- f** De berekende grenzen van het 95%-interval zijn  $42,6 \pm 2 \cdot 22,44 = -2,3$  en 87,5. Het interval is dus (0; 87,5).

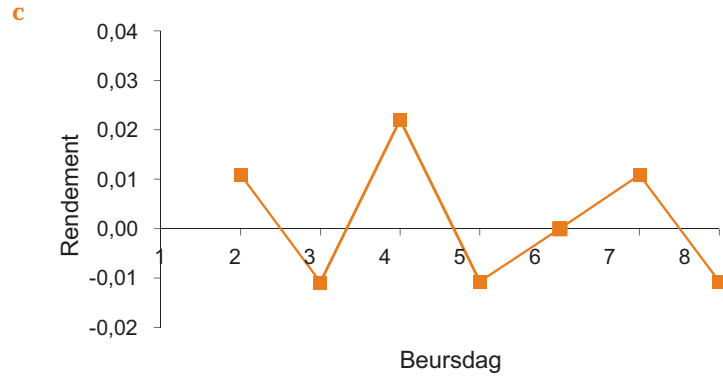
- 2.12 a**  $g = 0,3 \cdot v + 30$  ( $g$  = gasopbrengst, in euro;  $v$  = verbruik van gas in  $m^3$ )
- b** Gemiddelde is  $0,3 \cdot 1975 + 30 = 622,50$ ; standaardafwijking is  $0,3 \cdot 384 = 115,20$  (euro's per huishouden per jaar).
- c**  $(622,50 \pm 2 \cdot 115,20) = (392,10; 852,90)$

- 2.13 a** De klassengrenzen zijn niet bekend, alleen de gemiddelden van de tien klassen.
- b** Omdat elke inkomensgroep 10% vertegenwoordigt, kan het gemiddelde inkomen berekend worden door  $(9,7 + 19,3 + 25,3 + \dots + 158,7) / 10 = 56,83$  ( $\times \text{€ } 1000$ ). De standaardafwijking kan bijvoorbeeld met de Excelfunctie STDEV (A1:A10) worden gevonden:  $s = 44,43$  ( $\times \text{€ } 1000$ ). Merk op dat de laatste berekening een benadering is, omdat we daar alleen de gemiddelden per klasse hebben gebruikt.
- c** De mediaan van de bruto-inkomens is  $(40,6 + 50,3) / 2 = 45,45$  ( $\times \text{€ } 1000$ ). Dit kan ook met de Excelfunctie MEDIAAN. Net als in onderdeel b is hier sprake van een benadering.

- 2.14 a** Zie onderstaande tabel, in de tweede kolom.

Koers FGH	Rendement (FGH)	Groefactor
91	-	-
92	0,0110	1,0110
91	-0,0109	0,9891
93	0,0220	1,0220
92	-0,0108	0,9892
92	0,0000	1,0000
93	0,0109	1,0109
92	-0,0108	0,9892

- b** Het meetkundig gemiddelde van de zeven groeifactoren (derde kolom) is 1,0016, dus het gemiddeld rendement van FGH is 0,16%. NB Dit kan ook worden berekend door  $\sqrt[7]{92/91} = 1,0016$ , ofwel de zevendemachtswortel uit de laatste gedeeld door de eerste koers.  
Het gemiddeld rendement van BCD is 1,09% (zie paragraaf 2.2.1, deel 3 (meetkundig gemiddelde)).



d De (steekproef)standaardafwijking van de rendementen van BCD respectievelijk FGH is 1,6% en 1,3%.

**2.15 a** gemiddelde is 1,11 ( $\times \text{€ } 1000$ ), standaardafwijking is 2,02 ( $\times \text{€ } 1000$ )

**b** mediaan is  $0 + (50 - 31)/(53 - 31) = 0,864$  ( $\times \text{€ } 1000$ )

**c**  $(5/6) \cdot 17 + 4 = 18,2\%$

**d** 13,0%; 18,0%; 22,0%; 26,0%; 5,7% en 0,8%

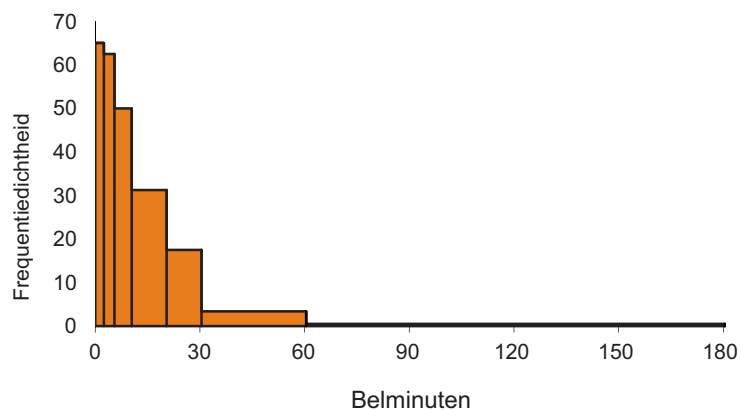
**2.16 a** 0,028 (cm)

**b**  $(1,809 \pm 2 \cdot 0,028) = (1,752; 1,865)$ . Vier waarnemingen zijn kleiner dan 1,752; twee zijn groter dan 1,865; samen dus 6 van de 150, dus 4%. Er ligt dus 96% in het 2s-interval.

**c**  $\text{KURT}(A1:F25) = -0,67$ , dus niet zo spits

**d**  $\text{SKEW}(A1:F25) = -0,20$ , dus linksscheef

**2.17 a** Met als basisbreedte de eerste klassen (2,5 belminuten), zijn de frequentiedichtheden respectievelijk 65; 62,5; 50; 31,25; 17,5; 3,33 en 0,52. Het histogram wordt dan:



**b** De steekproefstandaardafwijking is (met de Excelfile *maatstaven.xlsx* of de rekenmachine): 26,08 minuten.

**c** De populatiestandaardafwijking zou dan 26,06 belminuten zijn geweest. NB Bij de steekproefgrootte van 500 is al te zien dat er weinig verschil is tussen beide grootheden.

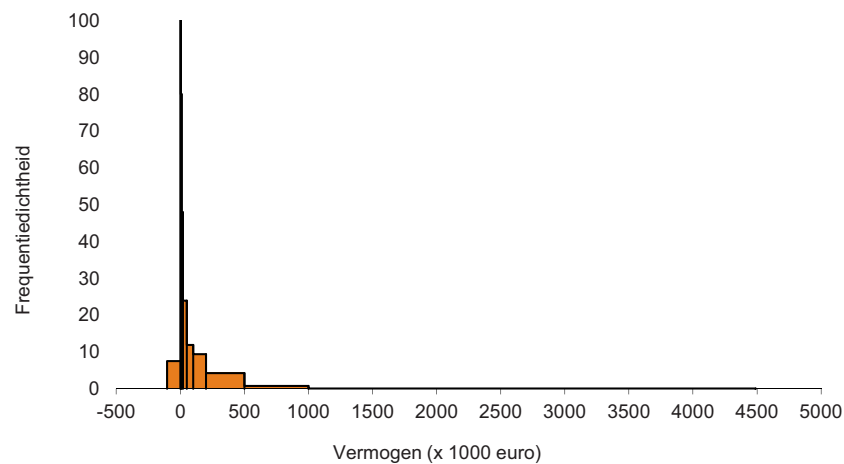
**d** Bereken met Excel de scheefheid (skewness): 2,91, dus rechtsscheef.



- e  $(19,47 \pm 2 \cdot 26,08) = (-32,7; 71,6)$ . De linkergrens is negatief, wat economisch niet kan, aangezien het om minuten gaat. Het interval mag daarom ook genoteerd worden als  $(0; 71,6)$ .

**2.18** De klassenindeling wordt:

Klasse	Linkergrens	Rechtergrens	Frequentie	Frequentiedichtheid
1	-104,0	0	781	7,5
2	0	5	1675	335
3	5	10	400	80
4	10	20	479	47,9
5	20	50	716	23,9
6	50	100	593	11,9
7	100	200	937	9,37
8	200	500	1255	4,2
9	500	1000	351	0,7
10	1000	4488,9	162	0,05



- 2.19** a RG = 3,85%, in perunages 0,0385.  
 b MG = 2,44%, in perunages 0,0244.

Rendement	Groefactor
-0,102	0,898
0,116	1,116
0,214	1,214
-0,253	0,747
0,157	1,157
0,099	1,099

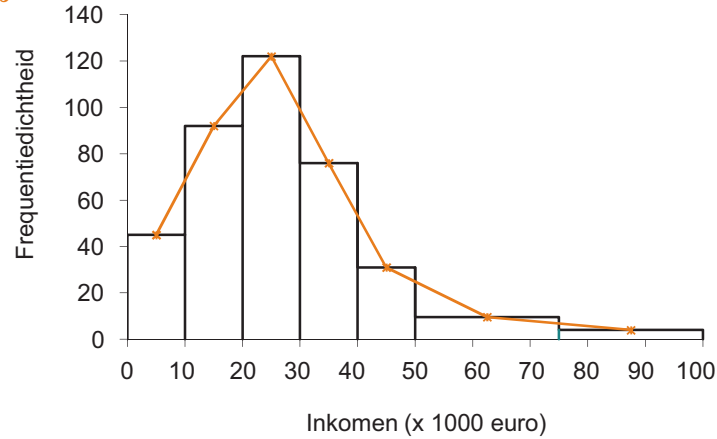
- c Meetkundig gemiddelde, omdat een rendement een groeivoet is en dus verbonden met een groefactor.

- d Standaardafwijking van de (rekenkundige) percentages is 0,1784.

$$\text{Checken of } MG \approx RG - \frac{\sigma^2}{2} \text{ ofwel } MG - RG + \frac{\sigma^2}{2} \approx 0$$

$$\text{Er volgt: } 0,0244 - 0,0385 + 0,1784^2 / 2 = 0,00181 \approx 0$$

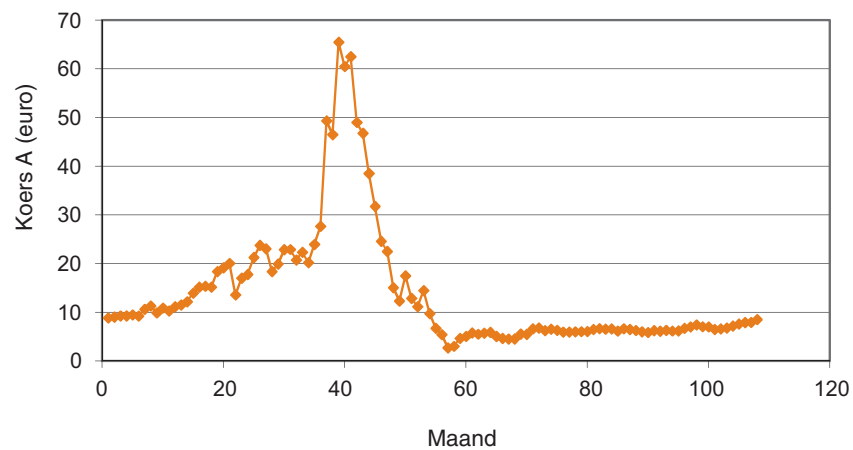
## 2.20 a, b



L	R	f	rel. f	cum. rel. f	fd. per 10
0	10	45	0,1125	0,1125	45
10	20	92	0,23	0,3425	92
20	30	122	0,305	0,6475	122
30	40	76	0,19	0,8375	76
40	50	31	0,0775	0,915	31
50	75	24	0,06	0,975	9,6
75	100	10	0,025	1	4
		400			

- c Het gemiddelde inkomen is 27,7 (× € 1000).  
 d  $Q_1 = 10 + 10 \times (25 - 11,25) / (34,25 - 11,25) = 16,0$   
 e  $s = 17,13$  (× € 1000)  
 f Ongeveer  $\frac{1}{2} \times 122 + 76 + \frac{1}{2} \times 31 = 153$  personen.

- 2.21 a De maandkoersen lopen van 1 januari 2004 tot en met 1 december 2012.



- b** De eerste vijf rendementen zijn: 0,0215; 0,0255; 0,0011; 0,0205 en –0,0245. Pas op alle 107 rendementen  $AVERAGE(\text{rendementen})$  toe.

NB We nemen hier het rekenkundig gemiddelde. (Het meetkundig gemiddelde, dat hier eigenlijk gebruikt moet worden, is –0,0004. In de berekening van de standaardafwijking moet echter gerekend worden met het rekenkundig gemiddelde. Deze methode wordt ook bedrijfseconomisch gehanteerd.)

Uitkomst: 0,0133 ofwel 1,33%.

- c,d** De volatiliteit is de (steekproef)standaardafwijking van de rendementen, berekend met  $STDEV(\text{rendementen})$ . In onderstaande tabel staan de gemiddelde (meetkundige) rendementen en hun volatiliteit van de vijf bedrijven over de periode 2004 tot en met 2012.

Merk op dat de volatiliteit (standaardafwijking) berekend is t.o.v. het *rekenkundig* gemiddelde, omdat daar automatisch mee gerekend wordt in de Excel-functie  $STDEV$ .

Bedrijf	A	B	C	D	E
Gemiddeld rendement	–0,0004	–0,0058	–0,0023	–0,0061	0,0093
Volatiliteit	0,1681	0,0994	0,1217	0,1216	0,1026

- 2.22** Noem de grafieken van Totaal, Inkomen uit arbeid en eigen onderneming, en Uitkering respectievelijk grafiek 1, 2 en 3

- a** 2  
**b** 2  
**c** 2  
**d** 2  
**e** 3  
**f** 2  
**g** 3

**Opmerking:** Bij rechtsscheve verdelingen, zoals inkomens, geldt meestal:  $\text{modus} < \text{mediaan} < \text{gemiddelde}$

**2.23** Te controleren is of het gewogen gemiddelde van de rubrieken '01000: Voedingsmiddelen en alcoholvrije dranken' tot en met '14000: Consumptie in het buitenland' inderdaad het indexcijfer 111,9 oplevert. Dit klopt niet:  $11.232.277,8/100.000 = 112,3$ .

Afdelingen	CPI	Wegingscoëfficiënt	Index × weging
00000 Totaal bestedingen	111,9		
01000 Voedingsmidd. en alc.vrije dranken	112,4	11.317	1.272.030,8
02000 Alcoholhoudende dranken en tabak	129,1	3111	401.630,1
03000 Kleding en schoeisel	102,1	4930	503.353,0
04000 Huisvesting, water en energie	112,4	24.529	2.757.059,6
05000 Stoffering en huish. apparaten	109,3	5993	655.034,9
06000 Gezondheid	104,2	1090	113.578,0
07000 Vervoer	117,8	11.641	1.371.309,8
08000 Communicatie	95,4	3263	311.290,2
09000 Recreatie en cultuur	100,9	10.285	1.037.756,5
10000 Onderwijs	107,8	141	15.199,8
11000 Hotels, cafés en restaurants	120,6	4199	506.399,4
12000 Diverse goederen en diensten	116,7	11.302	1.318.943,4
13000 Cons.geb. belast./overheidsdienst.	118,7	3670	435.629,0
14000 Consumptie in het buitenland	117,7	4529	533.063,3
<i>Totaal</i>		<i>100.000</i>	<i>11.232.277,8</i>

- 2.24**
- a** 28.176 / 4231,5 / 959,9 / 66,0 / 3,7 / 0,9
  - b**  $\bar{x} = € 20.095$
  - c**  $\sigma = € 75.439$
  - d**  $2\sigma$ -interval is  $(-130.783; 170.973)$ . (NB Linkergrens kan eigenlijk alleen maar 0 zijn.)
  - e**  $Q_1 \approx x_{16,676,25} = 592$  (zie appendix A3). 25% van de gefactureerde bedragen is kleiner dan 592 euro.
  - f**  $P_{80} \approx x_{53,364} = 5000 + (53.364 - 28.176 - 16.926) \cdot 15.000 / 14.399 = 13.607$ . 80% van de gefactureerde bedragen is kleiner dan 13.607 euro.
  - g**  $\bar{y} = 40 + 0,08 \cdot \bar{x} = 40 + 0,08 \cdot 20.095 = 1647,60$ . Totale boete is  $0,3 \cdot 66.705 \cdot 1647,60 = € 32.970.974$ .
  - h** Ja, de scheefheid is 7,23, dus rechtsscheef.
  - i** De spitsheid is 60,19, dus spitser dan de bijbehorende normale verdeling.

# 3 Elementaire kansrekening



**3.1** Op  $4 \cdot 2 \cdot 3 \cdot 6 = 144$  manieren.

**3.2**  $6^4 = 1296$ ;  $1/6^e$  deel daarvan begint met een 5, dus 216.

**3.3** Het aantal mogelijke diners is  $4 \cdot 8 \cdot 5 \cdot 2$  (wel of geen koffie) = 320.

**3.4 a**  $4^{10} = 1.048.576$

**b**  $4^5 \cdot 2^5 = 32.768$

**3.5**  ${}_{12}P_3 = 1320$  (want *met* volgorde)

**3.6 a** 2520; 840; 5040; 1; bestaat niet; 7

${}_7P_k$ : op hoeveel manieren zijn  $k$  verschillende cijfers van 1 tot en met 7 te rangschikken als de volgorde van die cijfers van belang is?

**b**  $r = 6$  en  $r = 7$ . Als  $r$  groter wordt, is ook het aantal mogelijke rangschikkingen groter.

**3.7 a**  ${}_{10}C_3 = 120$

**b**  ${}_{10}P_3 = 720$

**3.8**  ${}_{12}C_4 = 495$

**3.9**  $\frac{15!}{7! \times 5! \times 3!} = 360.360$

Merk op: kan ook met  ${}_{15}C_7 \cdot {}_8C_5$ ,  ${}_{15}C_5 \cdot {}_{10}C_7$  of  ${}_{15}C_3 \cdot {}_{12}C_5$ , die alle ook op 360.360 uitkomen.

**3.10 a** 21; 35; 1; 1; bestaat niet; 7

${}_7C_k$ : op hoeveel manieren zijn  $k$  verschillende cijfers te rangschikken als de volgorde van die cijfers *niet* van belang is?

**b**  $r = 3$  en  $r = 4$  geven beide de uitkomst 35. Wegens de symmetrie van de functie  ${}_7C_r$  (bijvoorbeeld  ${}_7C_2 = {}_7C_5$ ) moet het maximum 'in het midden liggen'.

**3.11** Het totaal aantal soorten pannenkoeken is het aantal met *maximaal* drie ingrediënten en is dus  ${}_{10}C_3 + {}_{10}C_2 + {}_{10}C_1 + {}_{10}C_0 = 120 + 45 + 10 + 1 = 176$ .

**3.12 a** geen enkele

**b** De enig mogelijke uitkomst is 1,1,1,1.

**c** 1,1,1,3 op  ${}_4C_1$  ofwel 4 manieren en 1,1,2,2 op  ${}_4C_2$  ofwel 6 manieren, in totaal 10 manieren.

- d** 1,1,1,4 op  ${}_4C_1$  ofwel 4 manieren en 1,1,2,3 op  ${}_4C_2 \cdot 2!$  ofwel 12 manieren en 1,2,2,2 op  ${}_4C_1$  ofwel 4 manieren, in totaal 20 manieren.
- e** 1,1,1,5: 4 manieren  
 1,1,2,4: 12 manieren  
 1,1,3,3: 6 manieren  
 1,2,2,3: 12 manieren  
 2,2,2,2: 1 manier  
 totaal: 35 manieren

**3.13** Op  ${}_6C_2$  ofwel 15 manieren respectievelijk op  ${}_6C_2 \cdot {}_6C_3$  ofwel  $15 \cdot 20 = 300$  manieren.

**3.14 a** kans =  $\frac{{}_6C_4 \times {}_{39}C_2}{{}_{45}C_6} = 0,0014$

**b** kans =  $\frac{{}_6C_3 \times {}_{39}C_3}{{}_{45}C_6} = 0,0224$

**c** kans =  $\frac{{}_6C_0 \times {}_{39}C_6}{{}_{45}C_6} = 0,4006$

**d** (uiteraard) gelijk aan 1

**e** kans (minstens 1 goed) =  $1 - \text{kans (geen enkele goed)} = 1 - 0,4006 = 0,5994$

**3.15 a** kans =  $\frac{51 + 43}{250} = 0,376$

**b** kans =  $\frac{250 - 80}{250} = 0,68$

**c** kans =  $\frac{12}{250} = 0,048$

**d**  $\frac{12}{43}$  ofwel circa 27,9%

**e** kans =  $\frac{170}{250} \times \frac{170}{250} = 0,4624$

**3.16 a**  $P(B \text{ of } S) = P(B) + P(S) - P(B \text{ en } S) = 0,9 + 0,3 - 0,25 = 0,95$

**b** Nu is  $P(B \text{ of } S) = P(B) + P(S) - P(B \text{ en } S) = 0,9 + 0,5 - 0,3 = 1,1$ . Dit is een kans groter dan 1, dus zijn de gegevens tegenstrijdig.

**c**  $P(B \text{ of } S) = P(B) + P(S) - P(B \text{ en } S) = 0,9 + 0,3 - 0,2 = 1$ . Alle klanten hebben een betaalrekening of een spaarrekening (of beide).

**d**  $P(B \text{ of } S) = P(B) + P(S) - P(B \text{ en } S) = 0,9 + 0,3 - 0,9 \cdot 0,3 = 0,93$ .

**3.17 a**  $P(X \text{ en } Y) = P(X) \cdot P(Y) = 0,75 \cdot 0,5 = 0,375$  ofwel 37,5%

**b**  $P(X | Y) = 0,8$  dus  $\frac{P(X \text{ en } Y)}{P(Y)} = 0,8$  dus  $P(X \text{ en } Y) = 0,8 \cdot P(Y) = 0,8 \cdot 0,5 = 0,4$  (40%)

**c**  $P(Y | X) = 0,7$  dus  $\frac{P(X \text{ en } Y)}{P(X)} = 0,7$  dus  $P(X \text{ en } Y) = 0,7 \cdot P(X) = 0,7 \cdot 0,75 = 0,525$  (52,5%)

**3.18 a** kans =  $\left(\frac{5}{6}\right)^5 \approx 0,4019$

**b** kans =  $\left(\frac{5}{6}\right)^{10} \approx 0,1615$

**c**  $\left(\frac{5}{6}\right)^n = 0,05$  leidt tot  $n = \frac{\log 0,05}{\log \frac{5}{6}} = 16,4$  dus kleinste steekproefgrootte is 17

**d** Kans =  ${}_5C_1 \times \frac{1}{6} \times \left(\frac{5}{6}\right)^4 = 0,4019$ . Er zijn vijf manieren ( ${}_5C_1$ ) waarop je in een steekproef van vijf klanten precies één ontevreden klant kunt aanwijzen (alleen de eerste klant, alleen de tweede, enz.). De kans op elk van deze vijf mogelijkheden is steeds  $(1/6) \times (5/6)^4$ . De gevraagde kans is dus  ${}_5C_1 \times (1/6) \times (5/6)^4 = 0,4019$ . Zie ook hoofdstuk 5.

- 3.19 a** stel  $g$  = 'goede auto',  $s$  = 'slechte auto',  $+$  = goedgekeurd en  $-$  = afgekeurd  
 $P(g \mid +) = 0,95$  dus  $P(g \text{ en } +) = P(+)$  ·  $P(g \mid +) = 0,8 \cdot 0,95 = 0,76$   
 $P(g \mid -) = 0,08$  dus  $P(g \text{ en } -) = P(-)$  ·  $P(g \mid -) = 0,2 \cdot 0,08 = 0,016$   
 De bijbehorende kruistabel is:

	g	s	
+	0,76	0,04	0,8
-	0,016	0,184	0,2
	0,776	0,224	1

**b**  $P(- \mid g) = \frac{P(- \text{ en } g)}{P(g)} = \frac{0,016}{0,776} \approx 0,0206$  dus circa 2,1%

**c** Aflezen uit kruistabel:  $P(g \text{ en } +) = 0,76$

**3.20 a**

	< 30	30 -< 60	60 -< 120	
M	0,1592	0,2694	0,1020	0,5306
V	0,2245	0,1837	0,0612	0,4694
	0,3837	0,4531	0,1632	1,0000

**b**  $\frac{55}{115} \times 100\% \approx 47,8\%$

**c**  $P(M \text{ en } < 30) = \frac{39}{245} \approx 0,1592$  en  $P(M) \times P(< 30) = \frac{130}{245} \times \frac{94}{245} \approx 0,2036$

De kansen zijn niet gelijk aan elkaar, dus de gebeurtenissen  $M$  en  $< 30$  zijn niet onafhankelijk. De kenmerken zijn niet onafhankelijk.

**3.21 a**  $P(\text{harten en plaatje}) = \frac{4}{52} = \frac{1}{13}$  en  $P(\text{harten}) \cdot P(\text{plaatje}) = \frac{13}{52} \cdot \frac{16}{52} = \frac{1}{13}$ .

Deze kansen zijn gelijk aan elkaar, dus er is sprake van onafhankelijkheid.

**b**  $P(\text{harten}) = \frac{13}{52} = \frac{1}{4}$  en  $P(\text{harten} \mid \text{plaatje}) = \frac{4}{16} = \frac{1}{4}$

- c Nu is  $P(\text{harten en plaatje}) = \frac{4}{54} = 0,0741$  en  $P(\text{harten}) \cdot P(\text{plaatje}) = \frac{13}{54} \cdot \frac{16}{54} = 0,0713$ , dus deze gebeurtenissen zijn niet onafhankelijk.

### 3.22

Geslacht	Studierichting			totaal
	BE	CE	MER	
Man	60	150	90	300
Vrouw	40	100	60	200
<b>totaal</b>	<b>100</b>	<b>250</b>	<b>150</b>	<b>500</b>

Geslacht	Studierichting			totaal
	BE	CE	MER	
Man	0,12	0,3	0,18	0,6
Vrouw	0,08	0,2	0,12	0,4
<b>totaal</b>	<b>0,2</b>	<b>0,5</b>	<b>0,3</b>	<b>1</b>

- 3.23** a  $0,17 + 0,12 + 0,20 + 0,02 = 0,51$   
 b  $0,11 + 0,12 = 0,23$   
 c  $0,13 + 0,17 + 0,09 + 0,02 = 0,41$   
 d 0  
 e  $1 - 0,36$  ofwel  $0,64$   
 f  $0,49 + 0,23 - 0,11$  dus  $0,61$
- 3.24** a  $P(M | \text{MER}) = \frac{P(M \text{ en MER})}{P(\text{MER})} = \frac{0,16}{0,36} \approx 0,4444$   
 b  $P(\text{MER} | M) = \frac{P(M \text{ en MER})}{P(M)} = \frac{0,16}{0,49} \approx 0,3265$   
 c Het uitgangspunt is steeds een andere deelpopulatie.  
 d  $P(\text{CE} | V) = \frac{P(\text{CE en V})}{P(V)} = \frac{0,17}{0,51} \approx 0,3333$   
 e  $P(V | \text{IF}) = \frac{P(V \text{ en IF})}{P(\text{IF})} = \frac{0,12}{0,23} \approx 0,5217$   
 f  $P(M \text{ en CE}) = 0,13$  en  $P(M) \cdot P(\text{CE}) = 0,49 \cdot 0,30 = 0,147$   
 De gebeurtenissen M en CE zijn niet onafhankelijk.

- 3.25** a  $0,51^2 = 0,2601$   
 b Onderscheid de volgorden: M-V en V-M.  
 kans =  $2 \cdot 0,49 \cdot 0,51 = 0,4998$   
 c kans =  $0,13^2 + 0,11^2 + 0,16^2 + 0,09^2 + 0,17^2 + 0,12^2 + 0,20^2 + 0,02^2 = 0,1464$

- 3.26** a aantal pakketten is  ${}_{24}C_{10} = 1.961.256$   
 b aantal pakketten is  ${}_{24}C_5 \cdot {}_{24}C_{10}$ , ofwel circa  $8,336 \cdot 10^{10}$



- c  $\beta$ (portefeuille) =  $0,45 \cdot 0,453 + 0,35 \cdot 0,826 + 0,2 \cdot 1,512$  ofwel circa 0,795 (dit is een gewogen gemiddelde)

**3.27 a** De kruistabel wordt:

		Werkelijke toestand jaarrekening		
		Bevat geen fouten	Bevat fout(en)	
Beslissing	Goedkeuring	0,69	0,01*	0,7
	Afkeuring	0,11	0,19	0,3
		0,8	0,2	1

\* 5% van de jaarrekeningen die fouten bevatten, wordt goedgekeurd; dus  $0,05 \times 0,2 = 0,01$ ; enzovoort

- b  $P(\text{fouten} \mid \text{goedkeuren}) = \frac{P(\text{fouten en goedkeuren})}{P(\text{goedkeuren})} = \frac{0,01}{0,7} \approx 0,014$  dus circa 1,4%
- c Nu geldt de onafhankelijkheidsregel:  $P(A \text{ en } B) = P(A) \times P(B)$ . Dus de tabel wordt:

		Werkelijke toestand jaarrekening		
		Bevat geen fouten	Bevat fout(en)	
Beslissing	Goedkeuring	0,56	0,14*	0,7
	Afkeuring	0,24	0,06	0,3
		0,8	0,2	1

\*  $0,14 = 0,2 \times 0,7$  enzovoort

**3.28 a** aantal volgorden =  $\frac{8!}{3! \times 2!} = 3360$

b aantal volgorden =  $\frac{8!}{4! \times 2! \times 2!} = 420$

- c De standaardafwijking is in alle gevallen gelijk.

**3.29 a** De gunstige uitkomsten zijn BCA en CAB; totaal aantal mogelijke uitkomsten:  $3! = 6$ , dus de kans =  $1/3$ .

b Gunstig zijn nu: BADC, BCDA, BDAC, CADB, CDAB, CDBA, DABC, DCAB, DCBA (9 stuks). Totaal aantal mogelijkheden is  $4!$  ofwel 24, dus de kans =  $9/24$ .

c Gunstig zijn: ACB, BAC, CBA; mogelijk:  $3!$  ofwel 6, dus de kans =  $1/2$ .

**3.30 a**

2 letters, 4 cijfers:	$26^2 \cdot 10^4$	=	6.760.000
3 letters, 3 cijfers:	$26^3 \cdot 10^3$	=	17.576.000
4 letters, 2 cijfers:	$26^4 \cdot 10^2$	=	45.697.600
Totaal:			$\frac{70.033.600}{+}$

$$\begin{aligned}
 \text{b } 2 \text{ letters, 4 cijfers: } & 26 \cdot 25 \cdot 10 \cdot 9 \cdot 8 \cdot 7 = 3.276.000 \\
 3 \text{ letters, 3 cijfers: } & 26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 = 11.232.000 \\
 4 \text{ letters, 2 cijfers: } & 26 \cdot 25 \cdot 24 \cdot 23 \cdot 10 \cdot 9 = 32.292.000 \\
 \text{Totaal:} & \underline{\hspace{1.5cm}} + \\
 & = 46.800.000
 \end{aligned}$$

**3.31 a** kans =  $\left(\frac{1}{6}\right)^4 \approx 0,00077$  (zie ook opgave 3.12)

**b** kans =  $4 \times \left(\frac{1}{6}\right)^4 \approx 0,00309$  (zie ook opgave 3.12)

**c** kans =  $4 \times \left(\frac{1}{6}\right)^4 + 6 \times \left(\frac{1}{6}\right)^4 \approx 0,00772$  (zie ook opgave 3.12)

**d** kans =  $2 \times 4 \times \left(\frac{1}{6}\right)^4 + 12 \times \left(\frac{1}{6}\right)^4 \approx 0,01543$  (zie ook opgave 3.12)

**e** kans =  $4 \times \left(\frac{1}{6}\right)^4 + 12 \times \left(\frac{1}{6}\right)^4 + 6 \times \left(\frac{1}{6}\right)^4 + 12 \times \left(\frac{1}{6}\right)^4 + \left(\frac{1}{6}\right)^4 \approx 0,02701$   
(zie ook opgave 3.12)

**3.32 a** kans =  $\frac{{}^6C_6 \times {}^{43}C_0}{{}^{49}C_6} \approx 7,15 \times 10^{-8}$  (d.i. circa 7 op 100 miljoen)

**b** kans =  $\frac{{}^6C_5 \times {}^{43}C_1}{{}^{49}C_6} \approx 0,00002$

**c** kans =  $\frac{{}^6C_0 \times {}^{43}C_6}{{}^{49}C_6} \approx 0,43596$

**d** kans =  $1 - \frac{{}^6C_6 \times {}^{43}C_0}{{}^{49}C_6} \approx 1 - 0,43596 = 0,56404$

**3.33 a**  ${}^7C_3 \times 10^4 \times 26^3 = 6.151.600.000$

**b**  ${}^7C_3 \times {}^{10}P_4 \times {}^{26}P_3 = 2.751.840.000$

### 3.34

	A	B	C	
U	120	60	70	250
V	380	90	280	750
	500	150	350	1000

$$P(B | U) = \frac{P(B \text{ en } U)}{P(U)} \rightarrow 0,24 = \frac{0,06}{P(U)} \rightarrow P(U) = \frac{0,06}{0,24} = 0,25 \text{ dus } P(V) = 0,75$$

$$P(B \text{ of } U) = P(B) + P(U) - P(B \text{ en } U) \rightarrow 0,34 = P(B) + 0,25 - 0,06 \rightarrow P(B) = 0,15$$

Dus  $P(C) = 1 - 0,5 - 0,15 = 0,35$

$$P(A | U) = \frac{P(A \text{ en } U)}{P(U)} \rightarrow 0,48 = \frac{P(A \text{ en } U)}{0,25} \rightarrow$$

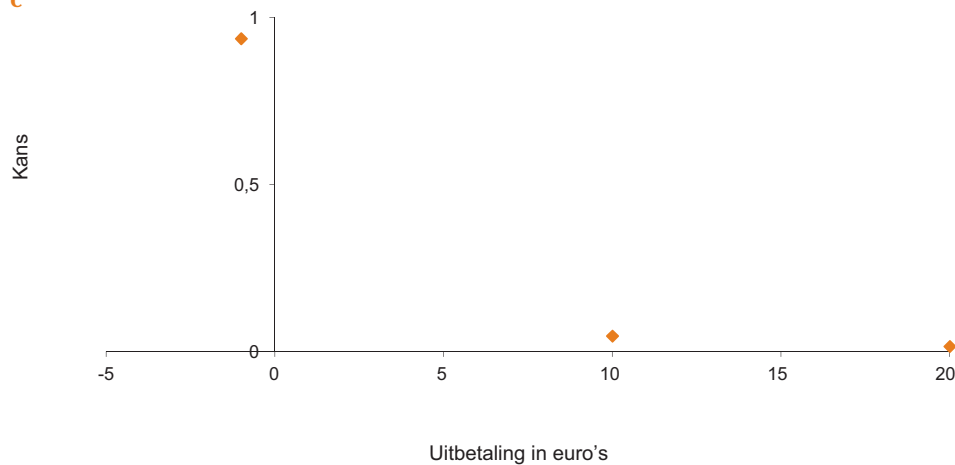
$$P(A \text{ en } U) = 0,48 \times 0,25 = 0,12; \text{ nu verder aanvullen.}$$

# 4 Kansvariabelen

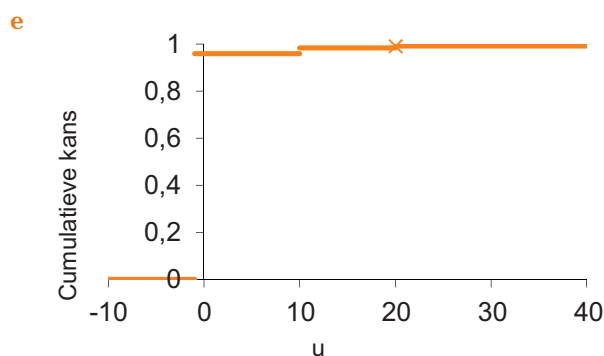


- 4.1 a** Met *maatstaven.xlsx* of (grafische) rekenmachine: verwachtingswaarde ( $E$ ) is € 57.000 en de standaardafwijking ( $\sigma$ ) is € 6403,12.  
**b**  $(E - 2 \cdot \sigma; E + 2 \cdot \sigma) = (44.193,75; 69.806,25)$   
**c** 0,9

- 4.2 a**  $P(\underline{u} = 20) = P(\text{AAA}) = 0,25^3 = 0,015625$ ;  $P(\underline{u} = 10) = P(\text{BBB, CCC of DDD}) = 3 \cdot 0,015625 = 0,046875$ ;  $P(\underline{u} = -1) = 1 - 0,015625 - 0,046875 = 0,9375$   
**b**  $E(\underline{u}) = -0,16$ ;  $\sigma(\underline{u}) = 3,44$  (euro)  
**c**



- d**  $P(\underline{u} < 5) = P(\underline{u} = -1) = 0,9375$



- 4.3 a**  $E(\underline{k}) = 1,47$  uur  
**b**  $\sigma(\underline{k}) = 0,877$  uur  
**c**  $\underline{v} = 50 + 80 \cdot \underline{k}$ , dus  $E(\underline{v}) = 50 + 80 \cdot 1,47 = € 167,60$   
**d**  $\sigma(\underline{v}) = 80 \cdot 0,877 = € 70,16$   
**e** De verwachting wordt 3 keer zo groot en de standaardafwijking wordt  $\sqrt{3}$  keer zo groot, dus respectievelijk € 502,80 en € 121,52.

- 4.4 a** Met de uitbetalingstabel hieronder volgt een verwachte opbrengst van € -0,15.

Opbrengst	Kans
8499	0,0001
-1	0,9999

- b** We gaan uit van een trekking met teruglegging. De kans op het winnende lot is dan  $2 \cdot 0,0001 \cdot 0,9999 \approx 0,0002$ . (Zonder teruglegging is de kans exact 0,002.)

Met de uitbetalingstabel hieronder volgt een verwachte opbrengst van -€ 0,30.

Opbrengst	Kans
8498	0,0002
-2	0,9998

- 4.5 a** Noem  $\underline{o}$  de onkosten van de machine na de garantietijd en binnen drie jaar. Dan geldt de volgende kansstabel voor  $\underline{o}$ :

Onkosten	Kans
0	0,95
100	0,05

De verwachte waarde hiervan is  $0 \cdot 0,95 + 100 \cdot 0,05 = € 5$ , kleiner dan € 15. De garantieverzekering is dus niet voordeliger voor de koper.

- b** Noem de reparatiekosten  $x$ . Er moet dan gelden:  $0 \cdot 0,95 + x \cdot 0,05 > 15$ . Hieruit volgt  $x > 300$ , dus bij reparatiebedragen boven € 300.
- 4.6 a**  $E(s) = 0,850$  en  $\sigma(s) = 1,236$ .
- b** Stel  $\underline{t}$  = aantal storingen per week. Dan is  $E(t) = 6 \cdot E(s) = 5,1$  en  $\sigma(t) = \sqrt{6} \cdot \sigma(s) = 3,03$ , onder de aanname dat de storingen onafhankelijk zijn van dag tot dag.
- 4.7 a** Verwachting en standaardafwijking zijn respectievelijk € 11.050.000 en € 194.700 per kwartaal.
- b** Verwachting en standaardafwijking zijn respectievelijk € 44.200.000 en € 389.400 per jaar.
- 4.8 a** 550.000 respectievelijk 34.928 euro
- b** 660.000 respectievelijk 37.269 euro
- 4.9** Verwachtingswaarde en standaardafwijking van de winst per maand zijn respectievelijk € 300.000 en € 264.197.

**4.10** Maak een kansverdeling van de klassenindeling met de klassenmiddens als uitkomsten.

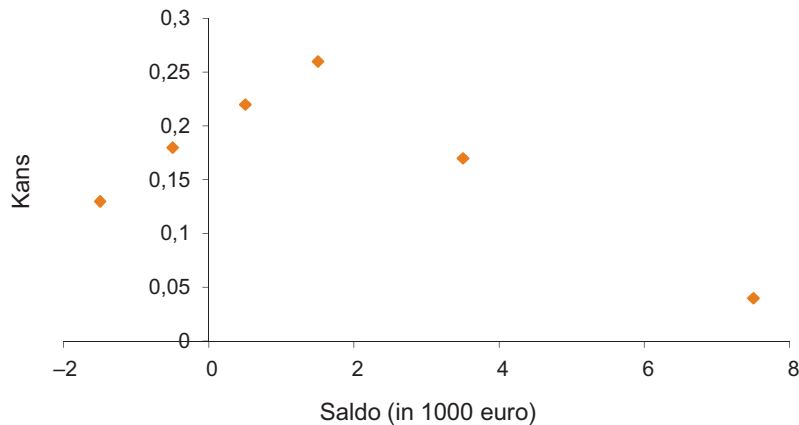
- a** Verwachtingswaarde is € 3460,00.
- b** Standaardafwijking is € 2511,65.
- c** Noem het totaal uit te keren schadebedrag  $t$ . Dan is  $E(t) = 1750 \cdot 3460 = 6.055.000$  euro en  $\sigma(t) = \sqrt{1750 \cdot 2511,65} = 105.070$  euro. Het  $1\sigma$ -interval voor het totale schadebedrag is dus:  $(6.055.000 \pm 105.070) = (5.949.930; 6.160.070)$ .
- d**  $2000 + 0,12 \cdot (2000/0,26) = € 2923$

**4.11 a**

Saldo*	Kans
-1,5	0,13
-0,5	0,18
0,5	0,22
1,5	0,26
3,5	0,17
7,5	0,04

\* in € 1000

**b**



**c** Met *maatstaven.xlsx* kunnen de verwachtingswaarde en standaardafwijking (voor één klant) worden bepaald:  $E(s) = 1,11$  en  $\sigma(s) = 2,02$ . Voor twee onafhankelijk gekozen klanten geldt:

$$E(\underline{s}_1 + \underline{s}_2) = E(\underline{s}_1) + E(\underline{s}_2) = 2 \cdot 1,11 = 2,22 (\times € 1000)$$

$$\sigma(\underline{s}_1 + \underline{s}_2) = \sqrt{s_1^2 + s_2^2} = \sqrt{2,02^2 + 2,02^2} = 2,86 (\times € 1000).$$

Of:

$$\sigma(\underline{s}_1 + \underline{s}_2) = 2,02 \times \sqrt{2} = 2,86 (\times € 1000).$$

**d**  $E(\underline{s}_1 + \dots + \underline{s}_{10}) = 10 \cdot 1,11 = 11,1 (\times € 1000)$

$$\sigma(\underline{s}_1 + \dots + \underline{s}_{10}) = \sqrt{2,02^2 + \dots + 2,02^2} = \sqrt{10 \cdot 2,02^2} = \sqrt{10} \cdot 2,02 = 6,39 (\times € 1000)$$

**e** De scheefheid is 1,17 (Excel), dus rechtsscheef.

**f** De spitsheid is 1,78, dus spitsser dan een vergelijkbare normale verdeling.

- 4.12 a**  $E(k) = 0,95$   
**b**  $\sigma(k) = 1,20$   
**c** Een totaal van 14 stringen of meer in 5 dagen kan alleen door de combinaties (3, 3, 3, 3, 3), (3, 3, 3, 3, 2), (3, 3, 3, 2, 3), (3, 3, 2, 3, 3), (3, 2, 3, 3, 3) en (2, 3, 3, 3, 3). De kans hierop is  $0,2^5 + 5 \cdot 0,2^4 \cdot 0,1 = 0,00112$ .

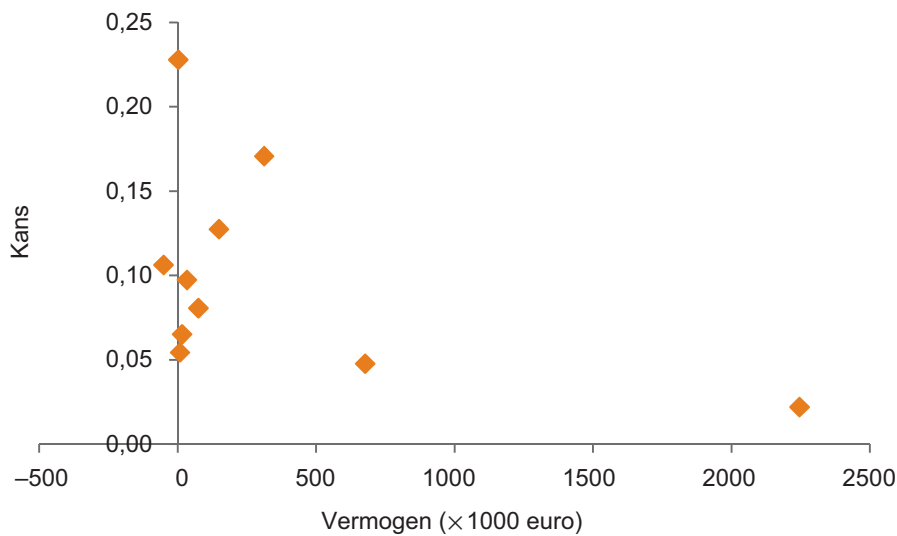
- 4.13 a** De gegeven tabel kent slechts tien mogelijke uitkomsten.  
**b** De tien frequenties zijn gedeeld door de totale frequentie (7349):

Gemiddeld vermogen	Kans
-51.985	0,1063
1254	0,2279
7250	0,0544
14.614	0,0652
32.402	0,0974
73.862	0,0807
148.026	0,1275
311.155	0,1708
675.783	0,0478
2.244.444	0,0220

1

- c**  $E(v) = 158.988$ ;  $\sigma(v) = 357.429$  (€) (Zie ook voorbeeld 2.8.)

**d**



- e** Respectievelijk 100 en 10 keer zo groot als de antwoorden bij c.

- 4.14 a** Schrijf alle zes lettercombinaties van de letters A, B en C uit en tel hoeveel er op hun 'oorspronkelijke' plaats staan. Men komt tot de volgende kansverdeling (merk op dat  $\underline{k}_3$  de waarde 2 niet kan aannemen):

$\underline{k}_3$	Kans
0	1/3
1	1/2
3	1/6

- b** Beide zijn exact gelijk aan 1.  
**c** Schrijf de 24 mogelijke permutaties van ABCD uit en tel de aantallen gelijkgeplaatste letters. Er volgt:

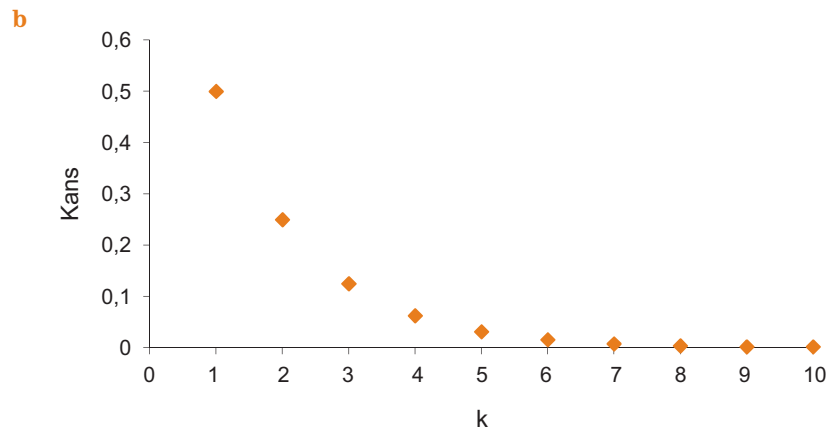
$\underline{k}_4$	Kans	
0	0,3750	(9/24)
1	0,3333	(8/24)
2	0,25	(6/24)
4	0,0714	(1/24)

- d**  $E(\underline{k}_4) = 1,12$  en  $\sigma(\underline{k}_4) = 1,11$ .

- 4.15 a**  $E(\underline{y}) = 15 + 15 = 30$ ;  $\sigma(\underline{y}) = \sqrt{3^2 + 3^2} = \sqrt{18} = 4,24$   
**b**  $E(\underline{y}) = 2 \cdot 15 = 30$ ;  $\sigma(\underline{y}) = 2 \cdot \sigma(\underline{x}) = 2 \cdot 3 = 6$  (NB Merk het verschil op met a!)  
**c**  $E(\underline{y}) = 17$ ;  $\sigma(\underline{y}) = 3$   
**d**  $E(\underline{y}) = 15 + 15 + 15 - 2 = 43$ ;  $\sigma(\underline{y}) = \sqrt{3^2 + 3^2 + 3^2} = \sqrt{27} = 5,20$   
**e**  $E(\underline{y}) = 0$ ;  $\sigma(\underline{y}) = 1$ . (Dit is standaardiseren; zie paragraaf 6.3.)

- 4.16 a**  $P(\underline{k} = 1) = 0,5$ ;  $P(\underline{k} = 2) = 0,25$ ;  $P(\underline{k} = 3) = 0,5^3$ , etc. Dit levert de volgende kans-tabel:

$\underline{k}$	Kans
1	0,5
2	0,25
3	0,125
4	0,0625
5	0,03125
6	0,015625
7	0,007813
8	0,003906
9	0,001953
10	0,001953



- c**  $E(\underline{k}) = 2,0$  en  $\sigma(\underline{k}) = 1,40$
- d** 1,99 (Excel), dus rechtsscheef.
- e** 5,00, dus zeer spits.



# 5 De binomiale verdeling



- 5.1 a** kans =  ${}_{50}C_5 \times 0,1^5 \times 0,9^{45} = 0,1849$ .  
In Excel: BINOMDIST(5;50;0,1;0) (Ned: BINOMIALE.VERD(5;50;0,1;0))
- b** kans = 0,6161 (tabel binomiale verdeling met  $n = 50$  en  $p = 0,1$ ; kijk bij  $k = 5$ )  
of:  
kans =  ${}_{50}C_0 \times 0,1^0 \times 0,9^{50} + {}_{50}C_1 \times 0,1^1 \times 0,9^{49} + \dots + {}_{50}C_5 \times 0,1^5 \times 0,9^{45}$  of:  
kans = BINOMDIST(5;50;0,1;1) = 0,6161
- c** kans =  $1 - \text{kans op ten hoogste } 5 = 1 - \text{BINOMDIST}(5;50;0,1;1) = 1 - 0,6161 = 0,3839$
- d** kans = kans op ten hoogste 7 – kans op ten hoogste 3 = 0,6276
- 5.2 a**  $P(\underline{x} < 38) = P(\underline{x} \leq 37) = 0,0299$
- b**  $P(\underline{x} \geq 42) = 1 - P(\underline{x} \leq 41) = 1 - 0,1684 = 0,8316$
- c**  $P(20 \leq \underline{x} \leq 40) = P(\underline{x} \leq 40) - P(\underline{x} \leq 19) = 0,1167 - 0,0000 = 0,1167$
- d**  $E(\underline{x}) = n \times p = 76 \times 0,6 = 45,6$
- 5.3 a** kans = BINOMDIST(0;120;0,03;0) = 0,0259
- b** kans = BINOMDIST(2;120;0,03;1) = 0,2984
- c** kans =  $1 - \text{BINOMDIST}(1;120;0,03;1) = 1 - 0,1218 = 0,8782$
- d** CRITBINOM(120;0,03;0,1) geeft uitkomst 1; het antwoord is dus 0
- 5.4 a**  $P(\underline{y} \geq 48) = 1 - P(\underline{y} \leq 47) = 1 - 0,2452 = 0,7548$
- b**  $P(\underline{y} \leq 40 \text{ of } \underline{y} \geq 60) = P(\underline{y} \leq 40) + 1 - P(\underline{y} \leq 59) = 0,0343 + 1 - 0,8766 = 0,1577$
- c** CRITBINOM geeft 42, dus het antwoord is 41
- d**  $P(\underline{y} \geq m) \leq 0,05$  geeft  $1 - P(\underline{y} \leq m - 1) \leq 0,05$ ; dus  $P(\underline{y} \leq m - 1) \geq 0,95$   
CRITBINOM geeft de kleinste waarde van  $m - 1$  waarvoor de kans nog groter is dan 0,95  
dus  $m - 1 = 63$ ; dus  $m = 64$
- 5.5 a** Er is hier sprake van een binomiale verdeling met  $n = 16$  en  $p = 0,65$ ; stel  $\underline{a}$  is het aantal monteurs dat aan het werk zal zijn (respectievelijk het aantal opdrachten dat wordt gegeven).  
Berekend moet worden:  $P(\underline{a} \geq 10)$  ofwel  $1 - P(\underline{a} \leq 9) = 1 - 0,3119 = 0,6881$ .
- b** Hier moet opgelost worden: voor welke waarde van  $n$  geldt  $P(\underline{a} \geq 10) \geq 0,95$ ?  
In onderdeel a zagen we dat 16 offertes er nog te weinig zijn, dus gaan we via 'trial and error' de juiste  $n$ -waarde zoeken:  $n = 24$  geeft  $P(\underline{a} \geq 10) = 0,9945$ ;  $n = 20$  geeft  $P(\underline{a} \geq 10) = 0,9468$ ; voor  $n = 21$  geldt dat de kans gelijk is aan 0,9687; de oplossing is dus: 21 offertes.

- 5.6**
- a**  $E(\underline{m}) = 600 \times 0,01 = 6$
- b**  $\text{var}(\underline{m}) = 600 \times 0,01 \times (1 - 0,01) = 5,94$  dus  $\sigma(\underline{m}) = \sqrt{5,94} \approx 2,44$
- c**  $P(\underline{m} \leq 2) = 0,0611$
- d** Probeer een aantal waarden voor n:  
 $n = 1000: P(\underline{m} \leq 2) = 0,0027$   
 $n = 800: P(\underline{m} \leq 2) = 0,0134$   
 $n = 700: P(\underline{m} \leq 2) = 0,0291$   
 dan bij  $n = 720: P(\underline{m} \leq 2) = 0,02497$  en bij  $n = 719: P(\underline{m} \leq 2) = 0,02516$ ; dus het antwoord is 720.
- 5.7**
- a**  $P(\underline{w} \geq 16) = 1 - P(\underline{w} \leq 15) = 1 - \text{BINOMDIST}(15;84;0,24;1) = 1 - 0,11471 = 0,88529$
- b**  $P(21 \leq \underline{w} \leq 25) = P(\underline{w} \leq 25) - P(\underline{x} \leq 20) = 0,91123 - 0,54335 = 0,36788$
- c** Met behulp van  $\text{CRITBINOM}(84;0,24;0,1)$  vinden we 15: dit is de kleinste waarde waarvoor de kans nog groter is dan 0,1; dus het antwoord is hier  $m = 14$ .
- d**  $P(\underline{w} \geq m) \leq 0,1$  kan worden herschreven als  $1 - P(\underline{w} \leq m - 1) \leq 0,1$  hetgeen weer kan worden omgezet in  $P(\underline{w} \leq m - 1) \geq 0,9$ ;  $\text{CRITBINOM}(84;0,24;0,9)$  geeft 25 hetgeen dus de kleinste waarde van  $m - 1$  is waarvoor de kans nog groter is dan 0,9, dus het antwoord is  $m = 26$ .
- e** Skewness (scheefheid) = 0,13 (dus is de verdeling rechtsscheef); de kurtosis (spitsheid) is -0,006, dus net zo spits als een vergelijkbare normale verdeling.
- 5.8**
- a** kans =  $0,95^5 = 0,77378$
- b** stel  $\underline{g}$  = aantal correct uitgevoerde keuringen;  $n = 10$   
 $P(\underline{g} = 8) = {}_{10}C_8 \cdot 0,95^8 \cdot 0,05^2 = 0,07463$
- c** kans =  $0,95^8 \cdot 0,05^2 = 0,00166$
- d**  $n = 100; P(\underline{g} = 80) = \text{BINOMDIST}(80;100;0,95;0) = 0,00000$  (calculator geeft  $8,44 \cdot 10^{-8}$ )
- 5.9**
- a** kans =  $\text{BINOMDIST}(7;10;0,6;0) = 0,21499$
- b** kans =  $\text{BINOMDIST}(7;10;0,6;1) = 0,83271$
- c** kans =  $\text{BINOMDIST}(10;10;0,6;0) = 0,00605$
- d** verwachtingswaarde =  $n \cdot p = 10 \cdot 0,6 = 6$
- e**  $SD = \sqrt{(n \cdot p \cdot (1 - p))} = \sqrt{(10 \cdot 0,6 \cdot 0,4)} = 1,55$
- f** verwachtingswaarde =  $10 \cdot 0,4 = 4$
- g**  $SD = \sqrt{(10 \cdot 0,4 \cdot 0,6)} = 1,55$
- h** scheefheid = -0,13; spitsheid = -0,18
- 5.10**
- a** Stel aantal overlijdensgevallen =  $\underline{d}$ ; verder geldt:  $n = 1600; p = 0,04$   
 $P(\underline{d} \geq 80) = 1 - P(\underline{d} \leq 79) = 1 - \text{BINOMDIST}(79;1600;0,04;1) = 1 - 0,97299 = 0,02701$  dus ca. 2,7%
- b** R voldoet niet aan de Europese norm.
- c**  $\text{CRITBINOM}(1600;0,04;0,95)$  geeft 77, dus  $P(\underline{d} \geq 77) \geq 0,95$  dus  $P(\underline{d} \leq 76) \leq 0,05$ ; dus bij ten hoogste 76 overlijdensgevallen zou R nog aan de norm voldoen en de conclusie dus anders zijn geweest dan in onderdeel b.

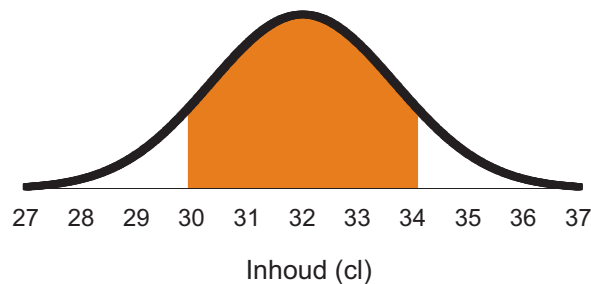
- 5.11**
- a** kans =  $0,9^{20} = 0,1216$
  - b** kans =  ${}_{20}C_2 \cdot 0,9^{18} \cdot 0,1^2 = 0,2852$
  - c** kans =  $1 - \text{BINOMDIST}(2;20;0,1;1) = 1 - 0,67693 = 0,3231$
  - d** verwachtingswaarde =  $20 \cdot 0,1 = 2$   
SD =  $\sqrt{(20 \cdot 0,1 \cdot 0,9)} = 1,34$
- 5.12**
- a**  $P(\underline{k} = 12) = 0,0967$
  - b**  $P(8 \leq \underline{k} \leq 14) = P(\underline{k} \leq 14) - P(\underline{k} \leq 7) = 0,7086$
  - c**  $P(\underline{k} \leq 5) = 0,062$  en  $P(\underline{k} \leq 6) = 0,124$  geeft  $L = 5$ .
  - d**  $\mu = n \cdot p = 10$ ;  $\sigma = \sqrt{(n \cdot p \cdot (1 - p))} = \sqrt{9,5} = 3,1$
  - e**  $P(\underline{k} \leq 13) = 0,87$  en  $P(\underline{k} \leq 14) = 0,92 > 0,90$ . Gevolg:  $R - 1 = 14$  en dus  $R = 15$ .
- 5.13** Er zijn duidelijk 7 staven te zien, dus  $n = 6$  (uitkomsten 0, 1, 2, 3, 4, 5 en 6). De top ligt bij 2, dus de kans zal circa 0,3 bedragen.

# 6 De normale verdeling



- 6.1**
- a  $\text{NORMDIST}(40;45;6;1) = 0,2023$
  - b  $1 - \text{NORMDIST}(48;45;6;1) = 0,3085$
  - c  $\text{NORMDIST}(50;45;6;1) - \text{NORMDIST}(42;45;6;1) = 0,4891$
  - d 0
  - e  $\text{NORMDIST}(63;45;6;1) - \text{NORMDIST}(27;45;6;1) = 0,9973$
  - f  $\text{NORMDIST}(54;45;6;1) - \text{NORMDIST}(36;45;6;1) = 0,8664$

- 6.2**
- a  $\text{NORMDIST}(30;32;1,6;1) = 0,1056$ .
  - b  $\text{NORMDIST}(34;32;1,6;1) - \text{NORMDIST}(30;32;1,6;1) = 0,7887$
  - c Zie onderstaande grafiek.



- d  $\text{NORMDIST}(24;32;1,6;1) = 0,0000003$ , dus de linkeroverschrijdingskans van  $\mu - 5\sigma = 32 - 5 \cdot 1,6 = 24$  is nagenoeg nul.

- 6.3**
- a  $\text{NORMDIST}(10;13;1,1;1) = 0,0032$ . NB De variabelen x (omzet),  $\mu$  en  $\sigma$  zijn een factor 1000 kleiner gemaakt. Dat is niet van invloed op het antwoord. Zie voor een verklaring paragraaf 6.3 over de standaardnormale verdeling.
  - b  $\text{NORMDIST}(15;13;1,1;1) - \text{NORMDIST}(10;13;1,1;1) = 0,9623$
  - c  $\text{NORMINV}(0,2;13;1,1) = 12,074$ , dus  $Y = \text{€ } 12.074$
  - d  $\text{NORMINV}(0,1;13;1,1) = 11,590$ ,  $Z = \text{€ } 11.590$

- 6.4**
- a  $1 - \text{NORMDIST}(25;30;4,25;1) = 0,8803$
  - b  $\text{NORMINV}(0,2;30;4,25) = 26,4$

- 6.5**
- a  $P(\underline{R} < 0) = \text{NORMDIST}(0;8,6;6,2;1) = 0,0827$
  - b  $P(0,05 < \underline{R} < 0,10) = \text{NORMDIST}(10;8,6;6,2;1) - \text{NORMDIST}(5;8,6;6,2;1) = 0,3086$
  - c  $(\mu \pm 1,96\sigma) = (8,6 \pm 1,96 \cdot 6,2) = (-3,6; 20,8)$ , dus tussen  $-3,6\%$  en  $20,8\%$ .

- 6.6**
- a  $1 - \text{NORMDIST}(5500;5000;350;1) = 0,0766$
  - b  $\text{NORMINV}(0,999;5000;350) = 6082$  liter

- 6.7**
- a**  $1 - \text{NORMDIST}(1,4) = 0,0808$
  - b**  $1 - \text{NORMDIST}(-0,123) = 0,5489$ ; of in z-tabel opzoeken, maar is dan een benadering:  $1 - P(z > 0,12) = 1 - 0,4522 = 0,5478$
  - c** 0,5, want het is precies de helft van de standaardnormale verdeling (of  $1 - \text{NORMDIST}(0) = 0,5$ )
  - d**  $\text{NORMDIST}(2,1) - \text{NORMDIST}(-0,1) = 0,5220$
  - e** Letterlijk in Excel:  $\text{NORMDIST}(-1,7) + 1 - \text{NORMDIST}(1,7) = 0,0891$ , maar ook bijvoorbeeld met z-tabel:  $2 \cdot P(\underline{z} > 1,7) = 2 \cdot 0,0446 = 0,0892$  (verschil door afronding in z-tabel).

- 6.8**
- a**  $p = \text{NORMSINV}(0,75) = 0,674$
  - b**  $q = \text{NORMSINV}(0,99) = 2,326$
  - c**  $r = \text{NORMSINV}(0,95) = 1,645$

- 6.9** Omdat 3% van 500 gram gelijk is aan 15 gram, is de voorwaarde:  $P(\underline{x} < 485) = 0,1$ , met  $\underline{x}$  een normale verdeling met onbekende  $\mu$  en standaardafwijking 10. De waarde is  $\text{NORMSINV}(0,1) = -1,282$ . Met behulp van voorbeeld 6.6 volgt:  $\mu = x - z \cdot \sigma = 485 + 1,282 \cdot 10 = 497,8$  gram.

- 6.10** We moeten controleren:  $P(\underline{x} < 485 \mid N(499,8;2,17)) < 0,1$ .  
(De notatie ' $N(499,8;2,17)$ ' betekent: 'gegeven een normale verdeling met  $\mu = 499,8$  en  $\sigma = 2,17$ .)  
Dit kan in Excel met:  $\text{NORMDIST}(485;499,8;2,17;1) = 4,5 \times 10^{-12}$  ( $= 4,5 / 10^{12}$ ). Dit is zeker kleiner dan 0,1.

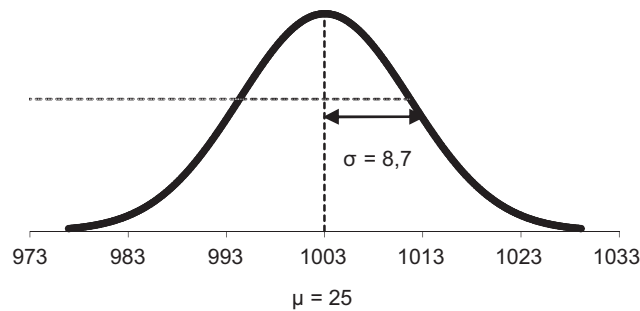
- 6.11** Gebruik het stelsel vergelijkingen of de Excelmethode (zie voorbeeld 6.7).  
Antwoord:  $\mu = 500,7$  en  $\sigma = 6,9$  (gram).

- 6.12** De somvariabele  $\underline{s}$  is normaal verdeeld met  $\mu = 550$  en  $\sigma = 34,928$  (beide in € 1000).  
Gevraagd wordt dus  $P(\underline{s} < 500)$ . Excel geeft:  $\text{NORMDIST}(500;550;34,928;1) = 0,0761$ .

- 6.13** De winstvariabele  $\underline{w}$  is normaal verdeeld met  $\mu = 0,3$  en  $\sigma = 0,264197$  (beide in miljoen euro).
- a**  $1 - \text{NORMDIST}(2,75;2,5;0,23;1) = 0,1385$
  - b**  $1 - \text{NORMDIST}(1;0,3;0,264197;1) = 0,0040$
  - c**  $\text{NORMDIST}(0;0,3; 0,264197;1) = 0,1281$

- 6.14**
- a**  $1 - \text{NORMDIST}(1,5;1,3;0,077519;1) = 0,0049$  (somvariabele met  $\mu = 13 \cdot 100.000$  en  $\sigma = \sqrt{13} \cdot 21.500$ , in miljoenen ingevuld in Excel)
  - b**  $\text{NORMDIST}(1,5;1,3;0,077519;1) - \text{NORMDIST}(1;1,3;0,077519;1) = 0,9950$
  - c**  $1 - \text{NORMDIST}(1,365;1,3;0,077519;1) = 0,2009$

- 6.15**
- a**  $1 - \text{NORMDIST}(120,7;120;2;1) = 0,3632$
  - b**  $0,3632^3 = 0,0479$
  - c**  $1 - \text{NORMDIST}(362,1;360;\sqrt{3} \cdot 2;1) = 0,2722$
  - d** Voor de somvariabele in c zijn meer combinaties mogelijk: niet alledrie hoeven meer te wegen dan 120,7 kg, slechts het gemiddelde moet meer zijn dan 120,7 kg; dus een grotere kans dan bij b.

**6.16 a**


**b**  $(\mu \pm 1,96\sigma) = (1003 \pm 1,96 \cdot 8,7) = (985,9; 1020,1)$

**c**  $\text{NORMDIST}(1000; 1003; 8,7; 1) = 0,3651$

**d**  $\text{NORMDIST}(12000; 12036; 8,7 \cdot \sqrt{12}; 1) = 0,1161$

**e**  $\text{NORMDIST}(24000; 24072; 8,7 \cdot \sqrt{24}; 1) = 0,0456$

**f** c:  $P(1 \text{ pak}) < 1 \text{ kg}$

d:  $P(12 \text{ pakken}) < 12 \text{ kg}$

e:  $P(24 \text{ pakken}) < 24 \text{ kg}$

Omdat het aantal ( $n$ ) in deze rij toeneemt, vermindert de standaardafwijking van het gemiddelde (' $\sqrt{n}$ -wet!'). Zware en lichte pakken heffen elkaar bij grote aantallen min of meer op en daardoor ligt het gemiddelde dichter rond  $\mu$  verspreid dan afzonderlijke waarnemingen, dus de kansen worden kleiner.

**6.17** Noem  $\underline{b}$  de bezoektijd,  $\underline{r}$  de reistijd tussen twee klanten en  $\underline{c}$  de cyclustijd ( $= \underline{b} + \underline{r}$ ).

**a** Er geldt:  $\underline{c}$  is een normale verdeling met  $\mu = 50 + 30 = 80$  en  $\sigma = \sqrt{(15^2 + 10^2)} = \sqrt{325}$ . Gemiddelde (verwachtingswaarde) en standaardafwijking zijn dus respectievelijk 80 en  $\sqrt{325} = 18,0$  minuten.

**c**  $P(\underline{c} > 100 \mid N(80; 18,0)) = P(\underline{z} > 1,11) = 0,1335$ ; of  $\text{normalcdf}(100, 10^{\wedge}99, 80, 18)$ ; of met Excel:

$$1 - \text{NORMDIST}(100; 80; \text{SQRT}(325); 1) = 0,1336$$

**6.18 a**  $P(\text{omzet} > 170.000) = 1 - \text{NORMDIST}(170; 150; 12; 1) = 0,0478$ . Of met grafische rekenmachine:  $\text{normalcdf}(170, 10^{\wedge}99, 150, 12) = 0,0478$ . Of met z-tabel:  $P(z > [170.000 - 150.000]/12.000) = P(\underline{z} > 1,67) = 0,0475$  (afwijkend antwoord wegens afronding van  $z$  op 2 decimalen).

**b**  $P(\underline{w} > 50.000 \mid N(40.000; \sqrt{(12.000^2 + 9.000^2)})) = 1 - \text{NORMDIST}(50; 40; 15; 1) = 0,2525$ . Kan ook met:  $P(\underline{w} > 50.000 \mid N(40.000; 15.000)) = P(\underline{z} > 0,67) = 0,2514$ .

**c**  $P(\underline{w} < 0 \mid N(40.000; 15.000)) = \text{NORMDIST}(0; 40; 15; 1) = 0,0038$ . De vuistregelintervallen ( $\mu \pm 2\sigma$ ) voor omzet respectievelijk kosten zijn (in duizenden):  $(150 \pm 2 \times 12)$  en  $(110 \pm 2 \times 9)$ , dus  $(126; 174)$  resp.  $(92; 128)$ . Deze intervallen hebben nauwelijks overlap, dus de kans dat de kosten hoger zijn dan de omzet is klein.

**6.19 a**  $1 - \text{NORMDIST}(25; 21; 3; 1) = 0,0912$

**b**  $1 - \text{NORMDIST}(240; 252; 3 \cdot \sqrt{12}; 1) = 0,8759$

**c** Maximaal  $(1/12) \cdot (240 - \text{NORMSINV}(0,75) \cdot 3 \cdot \text{SQRT}(12)) = 19,4$  minuten

**d** Dit is onmogelijk, want de bedoelde kans is altijd groter dan 0,5.

- 6.20**
- a  $\text{BINOMDIST}(50;200;0,3;1) = 0,0695$
  - b  $\mu = n \cdot p = 200 \cdot 0,3 = 60$  en  $\sigma = \sqrt{n \cdot p \cdot (1 - p)} = \sqrt{42} = 6,48$
  - c  $\text{NORMDIST}(50;60;6,48;1) = 0,0614$
  - d  $\text{NORMDIST}(50,5;60;6,48;1) = 0,0713$
  - e Bij ' $\leq$ ' moet er bij de normale benadering een 'halve binomiale klasse extra' worden genomen omdat de normale verdeling door het midden van die klasse loopt; bij  $\geq$  is het precies andersom.
  - f  $\text{var} = n \cdot p \cdot (1 - p) = 42 > 10$
  - g  $1 - \text{BINOMDIST}(74;200;0,3;1) = 0,0138$   
 $1 - \text{NORMDIST}(74,5;60;6,48;1) = 0,0126$
  - h Voor de binomiale verdeling geldt: scheefheid is 0,06; de spitsheid is  $-0,006$  (zie voor beide formules appendix A4). Conclusie: nauwelijks scheef of spits. Voor de normale verdeling zijn beide vormmaten gelijk aan 0.
- 6.21**
- a  $1 - \text{NORMDIST}(186;181,2;8,1;1) = 0,2767$
  - b  $\text{NORMDIST}(182;181,2;8,1;1) - \text{NORMDIST}(178;181,2;8,1;1) = 0,1929$
  - c Maak de verschilvariabele 'lengte man – lengte vrouw'. Deze is normaal verdeeld met  $\mu = 181,2 - 168,3 = 12,9$  en  $\sigma = \sqrt{(8,1^2 + 8,6^2)} = 11,74$ . Er volgt dan:  $\text{NORMDIST}(0;12,9;11,74;1) = 0,1360$
  - d De variatiecoëfficiënten zijn:  $V_{\text{man}} = 8,1/181,2 = 0,0447$ ;  $V_{\text{vrouw}} = 8,6/168 = 0,0512$ , dus bij de vrouwen.
  - e  $P(186) = \text{NORMDIST}(186,5;181,2;8,1;1) - \text{NORMDIST}(185,5;181,2;8,1;1) = 0,0413$
  - f Maak in Excel een tabel van de kansen van het type uit onderdeel e:  
 $P(150)^2 + P(151)^2 + \dots + P(200)^2 = 0,0348$
- 6.22**
- a  $Q_1 = \text{NORMINV}(0,25;70;10) = 63,3$ ;  $Q_3 = \text{NORMINV}(0,75;70;10) = 76,7$
  - b  $P_{20} = \text{NORMINV}(0,2;70;10)$  of  $\text{NORMINV}(20\%;70;10) = 61,6$   
 $P_{40} = \text{NORMINV}(0,4;70;10) = 67,5$ ; idem:  $P_{60} = 72,5$ ;  $P_{80} = 78,4$
  - c Bereken  $P(\underline{x} < 50 \text{ N}(70;10)) = \text{NORMDIST}(50;70;10;1) = 0,0228$ ;  
 $P(50 < \underline{x} < 60 \text{ N}(70;10)) = \text{NORMDIST}(60;70;10;1) - \text{NORMDIST}(50;70;10;1) = 0,1359$ ,  
 etc. Dit leidt de tabel linksonder met daarin de verwachte kansen, en vermenigvuldigd met 500, de verwachte frequenties.
  - d Idem, zie de tabel rechtsonder.

Onderdeel c	
kans	frequentie
0,0228	11
0,1359	68
0,3413	171
0,3413	171
0,1359	68
0,0228	11

Onderdeel d	
kans	frequentie
0,0062	4
0,0606	36
0,2417	145
0,3829	230
0,2417	145
0,0606	36
0,0062	4

- 6.23 a**  $P(\underline{b}_1 > 15) = 1 - \text{NORMDIST}(15;9;10;1) = 0,2743$
- b**  $P(\overline{b}_{10} > 15) = 1 - \text{NORMDIST}(15;9;10/\text{SQRT}(10);1) = 0,0289$   
( $\sqrt{n}$ -wet voor gemiddelde)
- c** v.i. 1 jaar: sparen  $(3 \pm 1,96 \cdot 0,5) = (2,0;4,0)$ ; beleggen  $(9 \pm 1,96 \cdot 10) = (-10,6;28,6)$
- d** v.i. 10 jaar: sparen  $(3 \pm 1,96 \cdot 0,5/\sqrt{10}) = (2,7;3,3)$ ; beleggen  $(9 \pm 1,96 \cdot 10/\sqrt{10}) = (2,8;15,2)$
- e**  $P(\overline{b}_{10} - \overline{s}_{10} > 0) = 1 - \text{NORMDIST}(0;6;\sqrt{(0,5^2+10^2)}/\sqrt{10};1) = 0,9710$
- f**  $n \geq 16$  ('solver' toepassen op:  $1 - \text{NORMDIST}(0;6;\sqrt{(0,5^2+10^2)}/\sqrt{n};1) = 0,99$ )
- g** Normale benadering; rendementen zijn eigenlijk groeifactoren, dus je mag ze niet rekenkundig middelen; onafhankelijkheid rendementen van jaar tot jaar.
- h** Bij b, d en f: de  $\sqrt{n}$ -wet voor gemiddelden ('centrale limietstelling') zegt dat deze verdelingen bij benadering normaal zijn voor grotere  $n$ .



# 7 Lineaire regressie

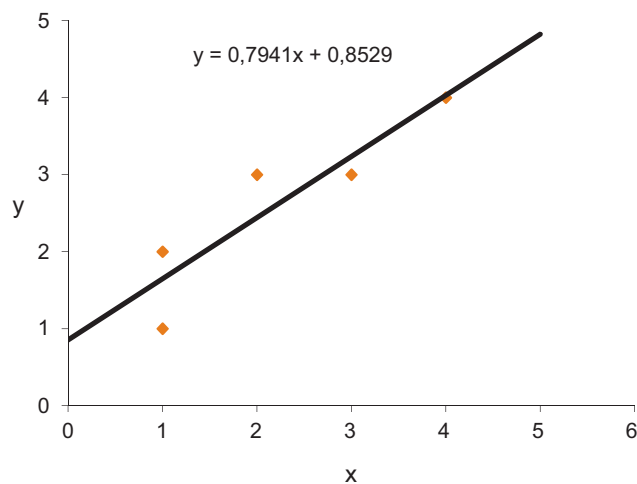


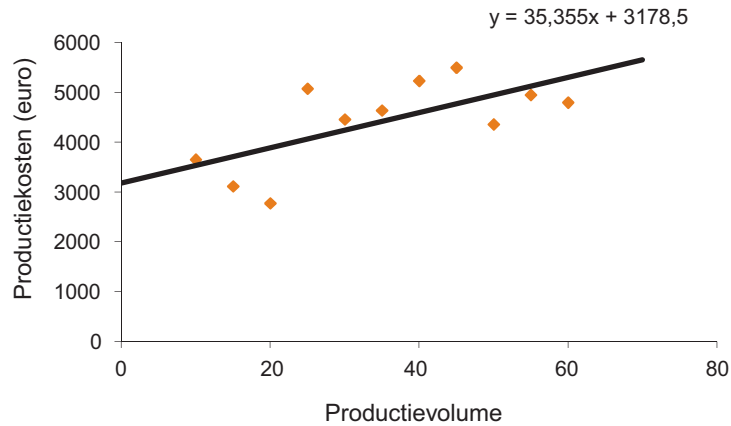
- 7.1 a** Bereken met Excel:  $\text{INTERCEPT}(B2:B8;A2:A8) = 11.492$  en  $\text{SLOPE}(B2:B8;A2:A8) = 64,459$ . De vergelijking van de regressielijn is dus  $TK = 64,459q + 11.492$
- b** Het zwaartepunt (gemiddelde van de q- respectievelijk TK-coördinaat) is (724,29; 58.179). Ingevuld in de regressielijn uit onderdeel a levert dit een ware bewering op:

$$TK = 64,459 \cdot 724,29 + 11.492 = 46.687,01 + 11.492 = (\text{afgerond}) 58.179$$

	A	B
1	q	TK
2	625	58.250
3	495	60.500
4	500	46.500
5	1050	67.500
6	1350	101.500
7	750	59.000
8	300	14.000

- 7.2 a, b, c**



**7.3 a**


- b** De vaste kosten zullen bij benadering € 3178 zijn; de variabele kosten kunnen geschat worden op € 53,35 per stuk.
- c**  $TK(70) = 3178,50 + 35,355 \cdot 70 = € 5653,35$

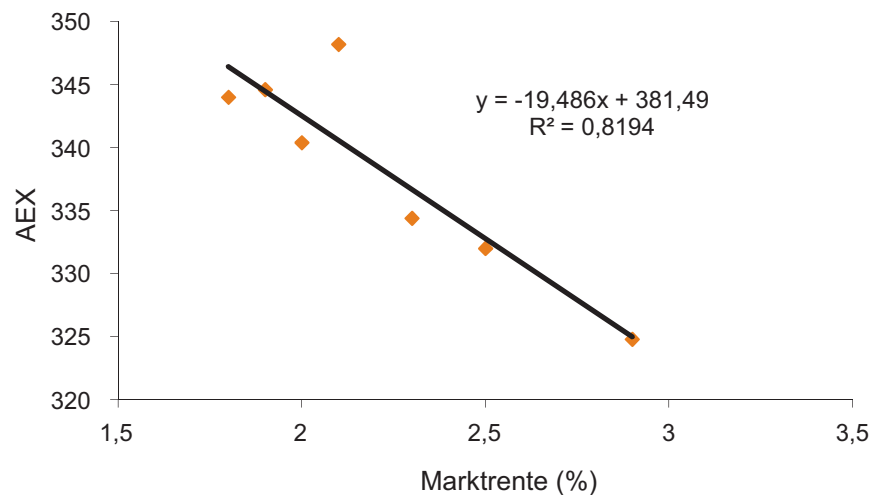
**7.4 a, b, c** De kwadratensommen zijn respectievelijk 3, 5,25 en  $11/9 = 1,22$ .

- d** De lijn  $y = x + 1/3$  heeft de kleinste kwadratensom en past dus het beste bij de vijf punten.

**7.5 a** De marktrente is de verklarende (onafhankelijke) variabele en de AEX is de verklaarde (afhankelijke) variabele.

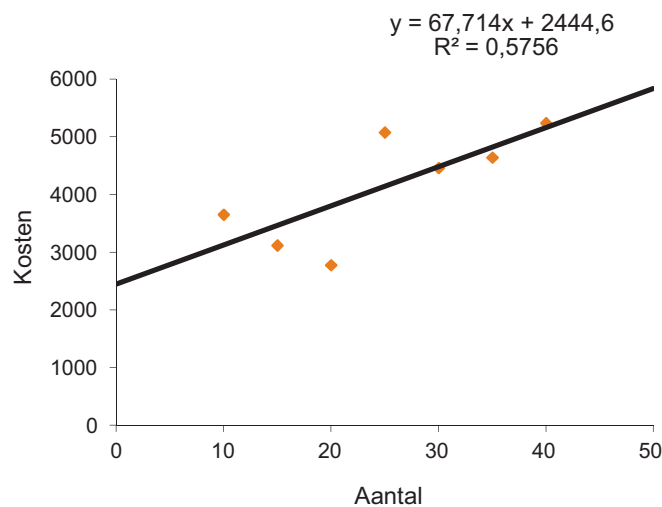
	A	B	C	D	E	F	G	H
1	Tijdstip	1	2	3	4	5	6	7
2	Marktrente (%)	2,5	1,9	1,8	2,9	2	2,1	2,3
3	Beursindex	332	344,6	344	324,8	340,4	348,2	334,4

- b**  $b = \text{SLOPE}(\text{Beursindex}; \text{Marktrente}) = \text{SLOPE}(B3:H3;B2:H2) = -19,486$ ;  $a = \text{INTERCEPT}(\text{Beursindex}; \text{Marktrente}) = 381,489$ . Dus  $y = 381,489 - 19,486 D x$ .
- c**  $r = \text{CORREL}(\text{Beursindex}; \text{Marktrente}) = \text{CORREL}(B3:H3;B2:H2) = -0,9052$
- d**  $AEX(5) = 381,49 - 19,486 \cdot 5 = 284,06$ , afgerond op hele cijfers (zoals de rij zelf ook): 284.
- e** De determinatiecoëfficiënt is 0,8194. De wortel hieruit is de correlatiecoëfficiënt.



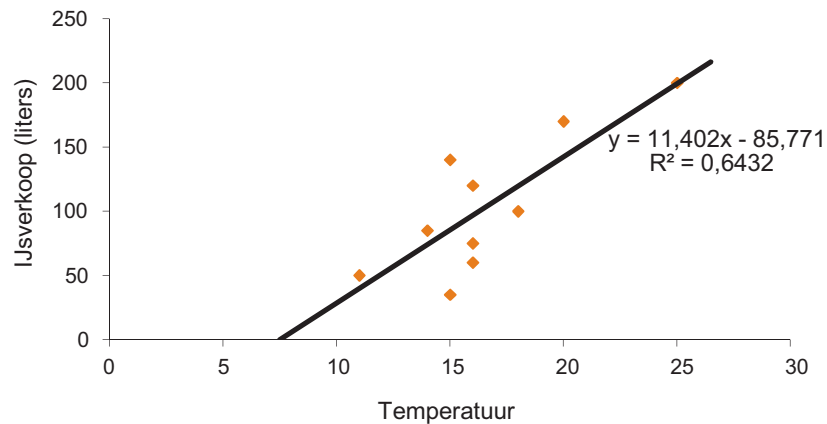
- 7.6 a** De waarde  $b = 1,772$  uit de regressievergelijking is de richtingscoëfficiënt van het lineaire verband  $y = a + b \cdot x$ .
- b** De correlatiecoëfficiënt is de mate van lineaire samenhang van de jaarrendementen van NV Haast en de markt. De uitkomst  $0,964$  betekent een zeer sterk verband.
- c** Een schatting van het jaarrendement van NV Haast als het marktrendement op jaarbasis  $10,0\%$  zou bedragen is  $y = -3,793 + 1,772 \cdot 10 = 13,93$ , een goede schatting aangezien de correlatiecoëfficiënt bijna  $1$  is.
- d** Het zwaartepunt is  $(5,3; 5,6)$ . Dit punt voldoet aan de regressievergelijking:  $5,6 = -3,793 + 1,772 \cdot 5,3$ .

- 7.7 a** Het productievolume is de onafhankelijke variabele en de grootte van de productiekosten is de afhankelijke variabele.
- b**  $\text{Kosten} = 67,714 \cdot \text{Aantal} + 2444,6$  (€)



- c**  $r = \sqrt{0,5756} = 0,75$ , of met  $\text{CORREL}(\text{Kosten}, \text{Aantal})$ , een redelijk goede positieve samenhang.
- d** De determinatiecoëfficiënt geeft het percentage verklaarde variatie in de productiekosten ten gevolge van het aantal aan. Er geldt:  $R^2 = 0,5756$  (uit grafiek), of met  $\text{RSQ}(\text{Kosten}, \text{Aantal})$ , dus ongeveer  $58\%$  van de kosten wordt verklaard door het aantal.

- 7.8 a** De (gemiddelde) dagtemperatuur is de oorzakelijke variabele en de ijsverkoop is de gevolgvariabele.
- b** De correlatiecoëfficiënt  $r$  is  $0,8020$ ; er is dus sprake van een matig positieve correlatie. De determinatiecoëfficiënt  $R^2$  heeft de waarde  $0,8020^2 = 0,64$ .
- c** De twee parameters  $a$  en  $b$  van de regressielijn zijn  $-85,77$  respectievelijk  $11,40$ .



- d** De negatieve waarde van  $a$  is niet vreemd, want de  $x$ -waarden (temperatuur) liggen tussen 11 en 25 (de *range*). In de buurt van de  $x$ -waarde 0 is het (blijkbaar) niet gezegd dat het lineaire verband nog steeds geldt.
- e** Een voorspelling van de verkoop bij een temperatuur van  $30^\circ\text{C}$  is  $y(30) = -85,77 + 11,40 \cdot 30 = 256$  liter ijs

- 7.9**
- a** groot, dus dicht bij 1
- b**  $\text{CORREL}(x_2;x_1) = 1$ , een perfecte correlatie dus.
- c** ongeveer 2 keer
- d**  $\text{SLOPE}(x_2;x_1) = 2$ . Dit betekent dat de reeks  $x_2$  twee keer zo snel beweegt als reeks  $x_1$ .
- e**  $\text{SLOPE}(x_1;x_2) = 0,5$

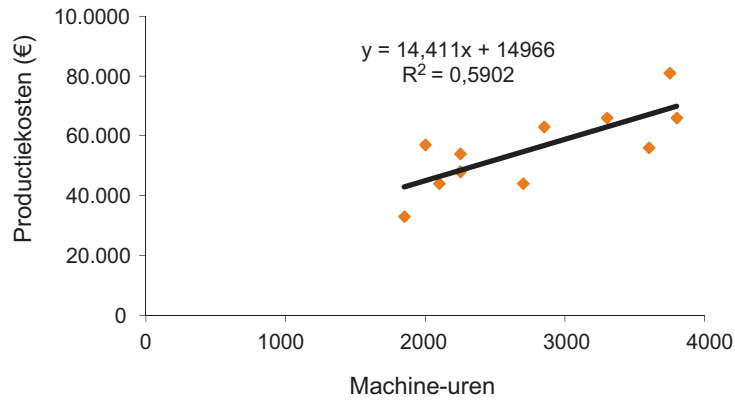
- 7.10**
- a**  $\approx 0,9$
- b**  $\text{CORREL}(x_2;x_1) = 0,82$
- c**  $\approx 1,5$
- d**  $\text{SLOPE}(x_2;x_1) = 1,25$ . Dit betekent dat reeks  $x_2$  1,25 keer zo hard schommelt als reeks  $x_1$ .
- e**  $\text{SLOPE}(x_1;x_2) = 0,54$  (dus niet  $1/1,25 = 0,8!$ )

- 7.11**

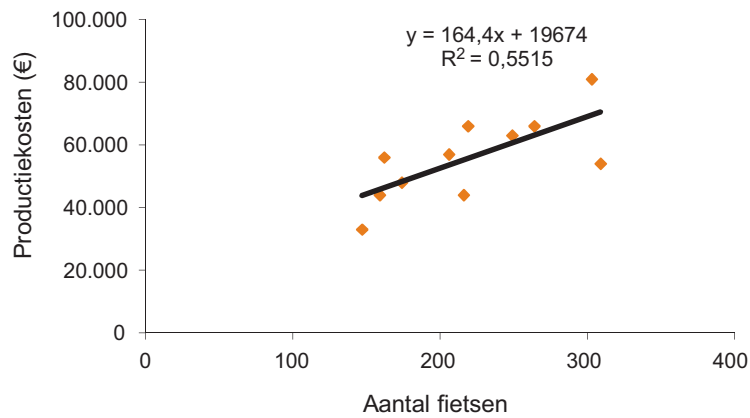
Uitgeverij C	Uitgeverij W
$a = 4,63$	$a = 22,82$
$b = 0,07$	$b = 0,062$
$r = 0,87$	$r = 0,88$

- b** Voor beide regressielijnen liggen de correlatiecoëfficiënten dicht bij 0,9, dus goede correlatie.
- c** De uitgeverij uit voorbeeld 7.4 heeft de laagste prijs per pagina.

7.12 a

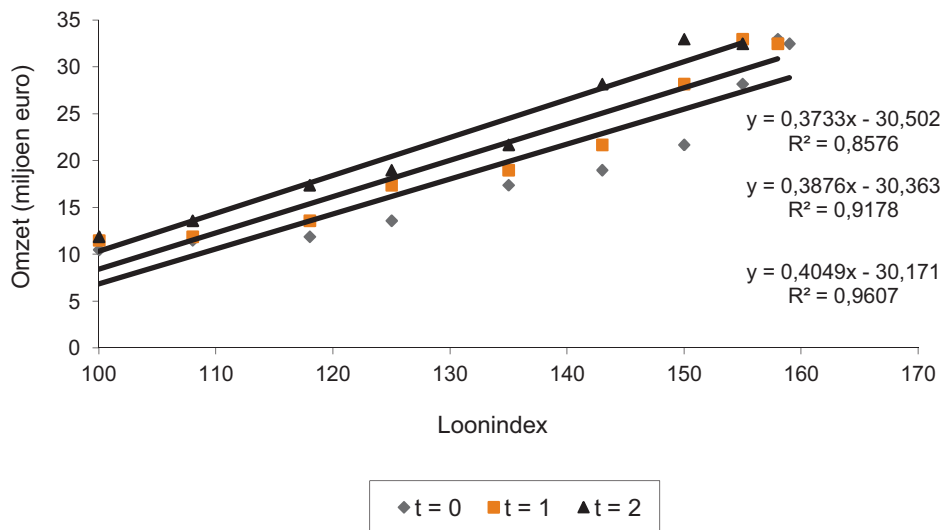


b



c Voor beide regressielijnen liggen de correlatiecoëfficiënten in de buurt van 0,6: matige correlatie. (NB Bij kostenschattingen wordt dit in de praktijk toch als voldoende samenhang beschouwd.)

7.13 a, b, c Met behulp van de trendlijnmethode in Excel.



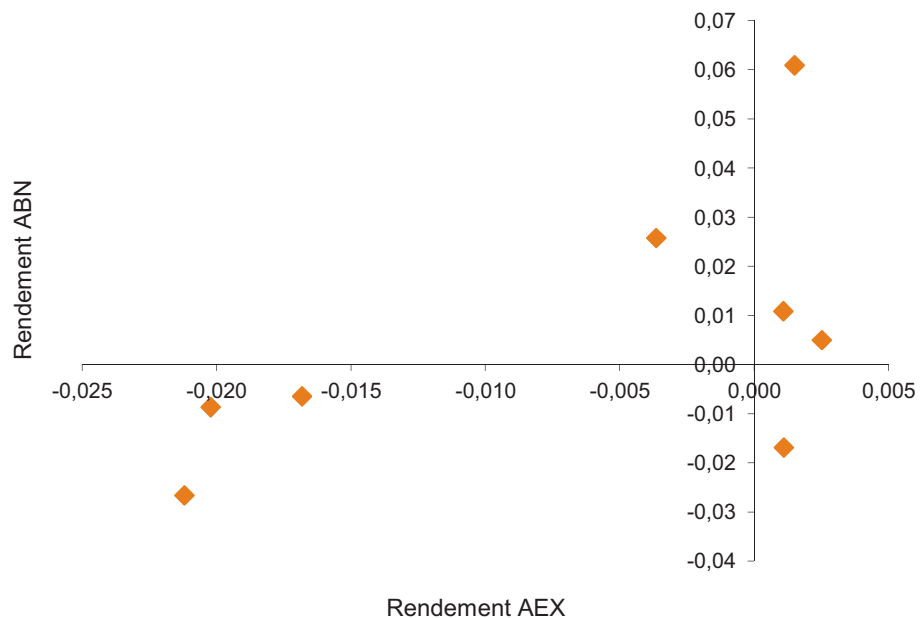
De correlatiecoëfficiënten (de wortel uit de  $R^2$ -waarden) zijn respectievelijk 0,93; 0,96 en 0,98.

d De regressielijn met een time lag van 2 jaar past dus het beste bij de gegevens.

**7.14 a** De rendementen zijn uit de koersen berekend (en niet uit de indexcijfers).

Dag	Rendementen	
	ABN	AEX
1		
2	0,061	0,001
3	0,011	0,001
4	0,005	0,003
5	-0,017	0,001
6	-0,027	-0,021
7	-0,009	-0,020
8	-0,006	-0,017
9	0,026	-0,004

**b**



- c**  $R(\text{ABN}) = 0,0159 + 1,4869 \cdot R(\text{AEX})$ , met  $r = 0,57$ . Matige correlatie.  
**d** Indexcijfers geven de groei van de koers aan ten opzichte van een vast gekozen koers; rendementen geven de groei van de koers van dag tot dag.

**7.15 a** Olieprijs is de oorzaak, winst is het gevolg.

**b**  $w = 0,0591 \cdot p - 0,185$ , met (grafische) rekenmachine.

**c** De correlatiecoëfficiënt  $r$  is 0,90; conclusie: goede positieve correlatie.

**d**  $w = 0,0591 \cdot 100 - 0,185 = 5,7$  miljard (euro).

**e** Een negatieve regressieconstante lijkt vreemd, maar het gebied rond  $p = 0$  (snijpunt met  $w$ -as) ligt buiten de 'relevant range'.

Of: Het snijpunt met de  $p$ -as ligt bij  $p = 3,13$  dollar per vat; dit is geen realistische situatie.

- 7.16** De stelling is waar. Sterker nog, er geldt bij lineaire regressie:  
 Als de correlatiecoëfficiënt  $r$  positief is, dan is de regressiecoëfficiënt  $b$  ook positief.  
 Als de correlatiecoëfficiënt  $r$  negatief is, dan is de regressiecoëfficiënt  $b$  ook negatief.  
 Als de correlatiecoëfficiënt  $r$  nul is, dan is de regressiecoëfficiënt  $b$  ook nul.

Ook het omgekeerde is waar, bijvoorbeeld:  
 Als de regressiecoëfficiënt  $b$  positief is, dan is de correlatiecoëfficiënt  $r$  ook positief.

Meer wiskundig (het symbool ' $\Leftrightarrow$ ' betekent dat de regel van links naar rechts, maar ook van rechts naar links gelezen kan worden: 'dan en slechts dan als'):

$$r > 0 \Leftrightarrow b > 0$$

$$r < 0 \Leftrightarrow b < 0$$

$$r = 0 \Leftrightarrow b = 0$$

**7.17 a**

Buurt	Inkomen*	% stijging huizenprijs
Oude Pijp	17	29
Diamantbuurt	17	32
Nieuwe Pijp	17,5	35
Duivelseiland	19	27,5
Stadionbuurt	19	16
Schinkelbuurt	20	26
Hoofddorpplein	20	25
Museumkwartier	27,5	17
Willemspark	29,5	16
Apollobuurt	34,5	13

\* in € 1000

- b** Lineaire regressie levert: % stijging huizenprijs =  $46 - 1,01 \cdot \text{Inkomen}$  (met inkomen in € 1000);  $r = -0,81$ : redelijk goede *negatieve* correlatie.

- 7.18** Nee; de regressielijn van  $y$  op  $x$  is dié rechte lijn waarbij de som van de *verticaal* gemeten afstanden van  $y$ -waarden tot die lijn zo klein mogelijk is ('kleinstekwadraten criterium'). Bij een regressie van  $x$  op  $y$  zullen *horizontale* afstanden moeten worden gemeten. Dit betekent dat er geen 'eenvoudige' omzetting van  $y = a + b \cdot x$  naar  $x = c + d \cdot y$  kan plaatsvinden.

- 7.19 a** Er moet een aselechte steekproef onder de personeelsleden worden genomen.  
**b**  $a = \bar{y} - b \cdot \bar{x} = 211,7 - 0,509 \cdot 339,3 = 39,0$  (zie appendix A10).  
**c** Geeft de toename van de kosten aan bij een toename van de afstand met 1 km.  
**d**  $k = 39,0 + 0,509 \cdot r = 39,0 + 0,509 \cdot 450 = 268,05$  (euro)

- 7.20** a, b en d worden alle 1,15 keer zo groot; c blijft gelijk.
- 7.21** a Het gemiddelde maandrendement van de beurs is de oorzaak; het gemiddelde maandrendement van Philips is het gevolg.  
 b  $y = 0,8783 \cdot x + 0,0226$  (met grafische rekenmachine)  
 c De correlatiecoëfficiënt  $r$  is 0,8978 ( $R^2 = 0,8061$ ). Deze is zo goed als gelijk aan 0,9, dus goede lineaire regressie.  
 d Rendement Philips =  $0,8783 \cdot 0,2 + 0,0226 = 0,1983$ ; dus 19,8%.
- 7.22** a regressieconstante =  $\bar{b} - \text{regressiecoëfficiënt} \cdot \bar{a} = 9,2 - 1,14 \cdot 7,7 = 0,422$   
 b Hoe goed het lineaire verband tussen beide variabelen is.  
 c  $b = 0,422 + 1,14 \cdot a = 0,422 + 1,14 \cdot 10,5 = 12,4$
- 7.23** a  $\text{CORREL}(A, B) = 0,4360$   
 b  $\text{CORREL}(A, C) = 0,2402$   
 $\text{CORREL}(A, D) = 0,1644$ , dus de kleinste  
 $\text{CORREL}(A, E) = 0,2363$
- 7.24** De correlatiecoëfficiënt is 0,24; er is dus nauwelijks een lineair verband tussen de variabelen.
- 7.25** a Winkeloppervlakte is oorzaak; index winkelbezoek is gevolg.  
 b  $y = -5,85 \cdot x + 234,45$   
 c De correlatiecoëfficiënt  $r$  is  $-0,986$  ( $R^2 = 0,9725$ ). Deze is kleiner dan  $-0,9$ , dus goede (negatieve) lineaire regressie.  
 d  $y = -5,85 \cdot 31 + 234,45 = 53,1$   
 e Het vreemde is dat ondanks het gestegen winkeloppervlak het bezoek is gedaald! Zou hier niet sprake zijn van een 'schijnverband'? Bijvoorbeeld:  
 – Het aantal shoppers daalt omdat steeds meer mensen op internet kopen.  
 – Desondanks stijgt het vloeroppervlak, bijvoorbeeld omdat huurprijzen zijn gedaald en winkels daardoor kunnen uitbreiden.