

# **Basisboek kwantitatieve methoden**

**Statistiek  
met Exceltoepassingen**

## **UITWERKINGEN VAN DE OPGAVEN**

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 2007

Deze uitwerkingen van de opgaven horen bij *Basisboek kwantitatieve methoden - Statistiek met Exceltoepassingen* van Donald van As en Jaap Klouwen.

Speciale dank gaat uit naar Rob Maas.

© 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 16 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)).

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 0069 7  
NUR        916

# 1 Grafieken en tabellen in de beschrijvende statistiek

- 1a Het doel van het onderzoek is tweeledig geweest:
1. nagaan in hoeverre in 2006 een verband bestond tussen de koersen van de AEX en Dow Jones;
  2. een overzicht geven van de slechtst en best scorende fondsen op de Nederlandse aandelenmarkten over het jaar 2006.
- b De koers van de AEX zal de afhankelijke variabele zijn en de koers van de Dow Jones de onafhankelijke variabele.
- 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 in de loop van mei dalen en in september weer stijgen, hetgeen inhoudt dat de genoemde beurswijsheid leidt tot verkoop tegen relatief hoge koers en weer kopen tegen een lage koers.
- 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.
- 2a 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.
- 3a Nominale schaal. 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 ratioschaal, nu is er immers een natuurlijk nulpunt, hoewel een omvang van 0 geschikt personeel op het totaal onwaarschijnlijk is.
- 4a 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.
- VS: circa  $39 \times 1,032$  geeft circa 40,2  
 Nederland: circa  $33 \times 1,022$  geeft circa 33,7  
 Frankrijk: circa  $32 \times 1,018$  geeft circa 32,6  
 Zweden: circa  $31 \times 1,028$  geeft circa 31,9  
 Denemarken: circa  $30 \times 1,017$  geeft circa 30,5  
 Duitsland: circa  $30 \times 1,024$  geeft circa 30,7  
 Groot-Brittannië: circa  $25 \times 1,028$  geeft circa 25,7  
 Finland: circa  $23 \times 1,03$  geeft circa 23,7

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

- VS: circa  $39 \times 1,032^5$  geeft circa 45,7
- Nederland: circa  $33 \times 1,022^5$  geeft circa 36,8
- Frankrijk: circa  $32 \times 1,018^5$  geeft circa 35,0
- Zweden: circa  $31 \times 1,028^5$  geeft circa 35,6
- Denemarken: circa  $30 \times 1,017^5$  geeft circa 32,6
- Duitsland: circa  $30 \times 1,024^5$  geeft circa 33,8
- Groot-Brittannië: circa  $25 \times 1,028^5$  geeft circa 28,7
- Finland: circa  $23 \times 1,03^5$  geeft circa 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.

5a 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
<b>Fixed -income</b>	100	130	150	205	230	290	395
<b>Equity</b>	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
<b>Fixed -income</b>	100	130	150	205	230	290	395
	49	63	73	100	112	141	193
<b>Equity</b>	100	80	60	65	70	90	120
	154	123	92	100	108	138	185

e Bijvoorbeeld een lijndiagram of een dubbel kolommendiagram.

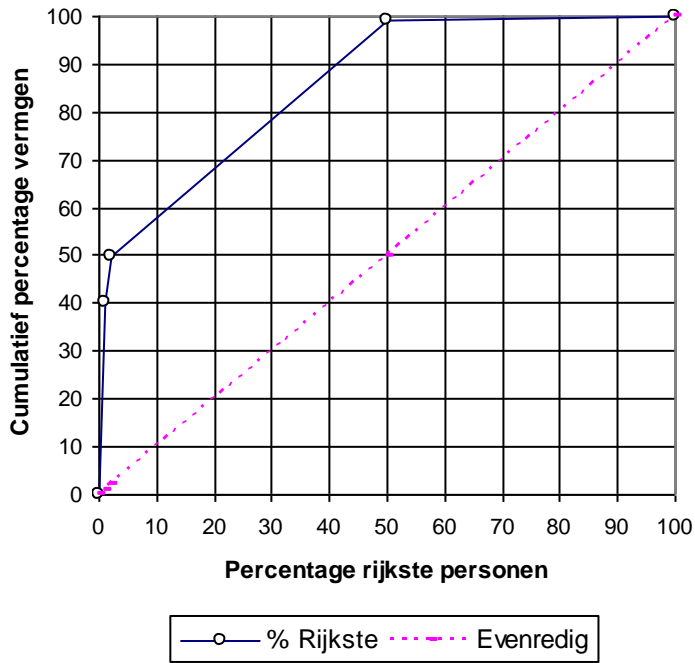
6a Een spreidingsdiagram, een cirkeldiagram en een staafdiagram.

b

<b>X</b> <b>% Rijkste</b> <b>personen</b>	<b>y</b> <b>% Vermogen</b>
0	0
1	40
2	50
50	99
100	100

c

### Concentratiecurve vermogen



d De verdeling van het vermogen is zeer scheef.

## 2 Maatstaven in de beschrijvende statistiek

- 2.1a 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 kleinste en de grootste 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

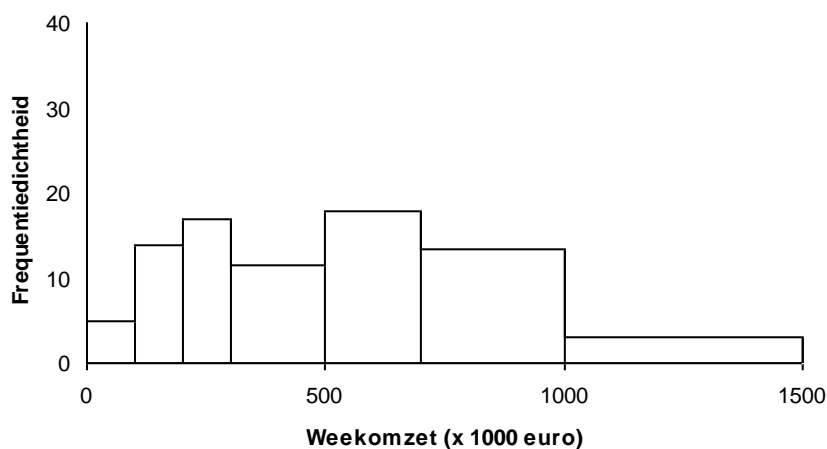
groeifactoren:

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 0,92%.

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.

- 2.2a 111,9 (uur)  
 b  $110 + (50 - 43,22)/(73,73 - 43,22) \times 10 = 112,2$  (lineaire interpolatie; zie p. 55)  
 (De cumulatieve relatieve percentages van de klassengrenzen zijn 19,49%; 43,22%; 73,73%; 94,92% en 1.)  
 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  $P_{30} = 100 + (30 - 19,49)/(43,22 - 19,49) \times 10 = 104,4$
- 2.3a 601,0 ( $\times \text{€ } 1000$ )  
 b De vijfde klasse, met de grootste frequentiedichtheid 18 (basis € 100.000)  
 c Mediaan =  $P_{50} = 500 + (50 - 39,3)/(63,3 - 39,3) \times 200 = 589,2$   
 d Derde kwartiel ( $Q_3$  oftewel  $P_{75}$ ) is  $700 + (75 - 63,3)/(90 - 63,3) \times 300 = 831,3$   
 e



- 2.4a De frequenties zijn: 65, 75, 100, 125, 70, 40 en 25.  
 b Eerste klasse (zie de vijfde kolom in tabel hieronder), met grootste frequentiedichtheid: 0,052.  
 c 19,5 belminuten  
 d 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  $((5,5/10) \times 0,25 + 0,14 + (24,5/30) \times 0,08) \times 100\% = 34,3\%$

Aantal belmin.	Relatieve frequentie	Midden	Breedte	Freq.-dichtheid	Cum.freq.
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	0,2	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

2.5a

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
50-<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:  $211 + (45 - 40)/(60 - 40) \times 112$  ofwel 239 (klanten).  
 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 het hoeveel de 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).

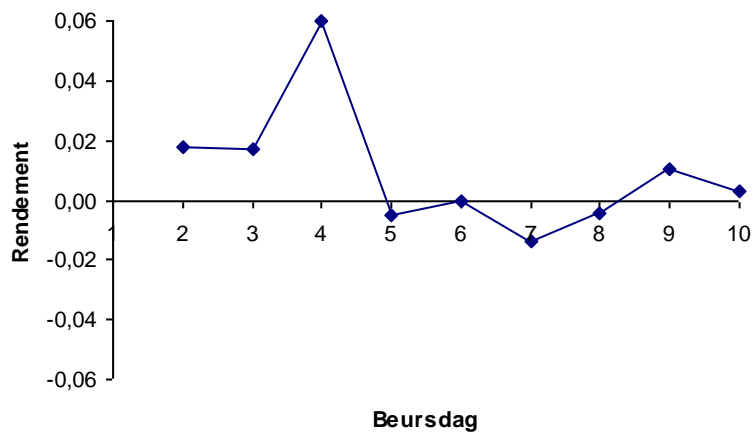
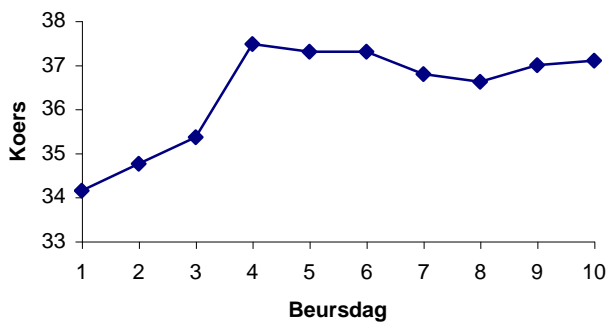
2.6a

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.  $\text{MEDIAN}(A1:F25) = 1,810$
3.  $\text{MODE}(A1:F25) = 1,810$
4.  $\text{QUARTILE}(A1:F25;3) = 1,830$
5.  $\text{PERCENTILE}(A1:F25;90\%)$

- 2.7a  $R = 37,48 - 34,15 = 3,33$
- b  $Q3 - Q1 = 37,25 - 35,68 = 1,57$
- c  $GAA = 0,98$  (Excel:  $\text{AVEDEV}(A1:J1)$ )
- d  $s = 1,19$  (Excel:  $\text{STDEV}(A1:J1)$ )
- e scheefheid =  $\text{SKEW}(A1:J1) = -1,1$
- f



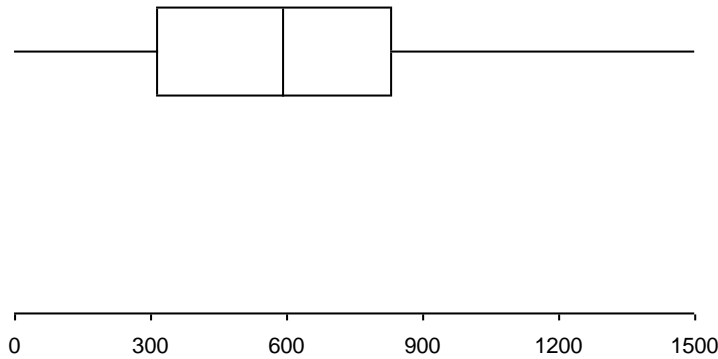
- g De standaardafwijking en de GAA, omdat die beide alle waarnemingen hebben betrokken in een berekening van de spreiding van de koersen.

- 2.8a 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' (hier: uren gedeeld door uren)
- e 0,07 (dimensieloos)
- f -0,90 (dimensieloos)
- g (88,67; 135,06)

- 2.9a  $Q_1 - Q_3 = 831,5 - 313,1 = 518,4$  ( $\times \text{€ } 1000$ )
- b  $s = 329,61$
- c  $V = 1,82$



d Boxplot:



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

2.10a  $R = 180$

b  $s = 26,08$

c scheefheid is 2,91, dus rechtsscheef

d spitsheid is 5,3, dus spitse verdeling

2.11a De kwartielafstand is 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, dus linksscheef.

e De mate van symmetrie wordt bepaald door de spitsheid; deze is 0,38, dus spitsser dan een normale verdeling met hetzelfde gemiddelde en standaardafwijking (42,6 respectievelijk 22,44).

f De berekende grenzen van het 95%-interval zijn -2,33 en 87,43. Het interval is dus (0; 87,43).

2.12a  $g = 0,3 \cdot v + 30$  ( $g$  = 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  $(622,50 \pm 2 \cdot 115,20) = (392,10; 852,90)$

d Vermenigvuldig de grenzen van het interval uit c met 945.000: (370.534,50; 805.990,50 (euro's per jaar)

2.13a De klassengrenzen zijn niet bekend, alleen de gemiddelden van de tien klassen.

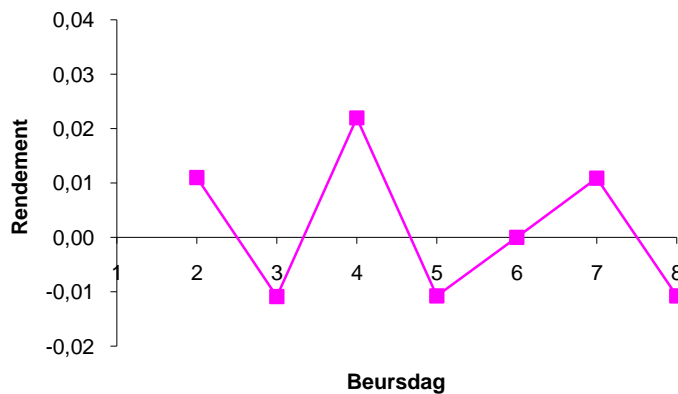
b Omdat elke inkomensgroep 10% vertegenwoordigt, kan het gemiddelde inkomen berekend worden door  $(-3 + 0,2 + 2,0 + \dots + 125)/10 = 37$  ( $\times$  €1000).

c € 38.427

2.14a Zie onderstaande tabel, in de tweede kolom.

FGH	r(FGH)	groefactoren
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 groefactoren (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 7demachtswortel uit de laatste gedeeld door de eerste koers. Het gemiddeld rendement van BCD is 1,09% (zie pagina 51).
- c

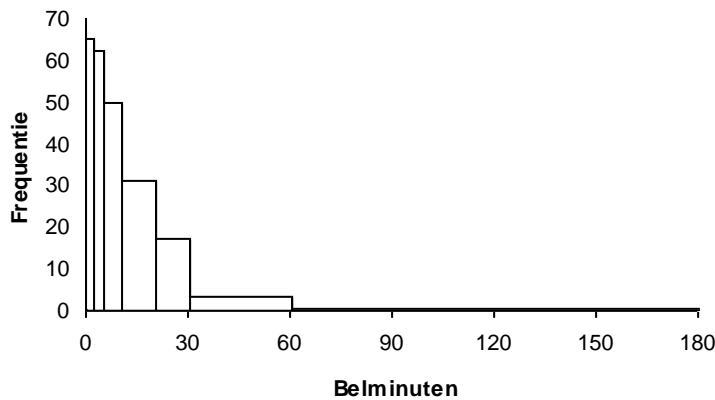


- 2.15a gemiddelde is 1,11 ( $\times \text{€}1000$ ), standaardafwijking is 2,02 ( $\times \text{€}1000$ )
- b mediaan is  $0 + (50 - 31)/(53 - 31) = 0,863,5$  ( $\times \text{€}1000$ )
- c  $(5/6) \cdot 17 + 4 = 18,2\%$
- d 13,0%; 18,0%; 22,0%; 26,0%; 5,7% en 0,8%

2.16a 0,028 (cm)

- b  $(1,809 \pm 2 \cdot 0,028) = 1,752; 1,865$
- c  $\text{KURT}(A1:F25) = -0,67$ , dus niet zo spits
- d  $\text{SKEW}(A1:F25) = -0,20$ , dus linksscheef

2.17a 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 als volgt:



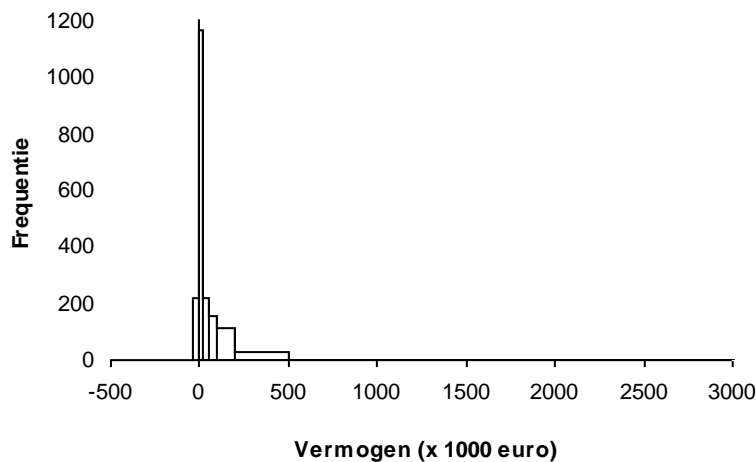
- b De steekproefstandaardafwijking is (met de Excelfile *maatstaven.xls* of de rekenmachine) is 26,08 minuten.
- c De populatiestandaardsafwijking zou dan 26,06 belminuten zijn geweest zijn. NB: Bij de steekproefgrootte van 500 is al te zien dat er weinig verschil is tussen beide grootheden.
- d Bereken 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	Midden	Freq.	Freq.dichtheid*
1	-42,8	0	-21,4	934	218,22
2	0	20	10	2.336	1168,00
3	20	50	35	667	222,33
4	50	100	75	792	158,40
5	100	200	150	1.108	110,80
6	200	500	350	794	26,47
7	500	1000	750	152	3,04
8	1000	3667,0	2333,5	72	0,27

\* basisbreedte € 1000



2.19a  $RG = 3,85\%$ , in perunages 0,0385.

b  $MG = 2,44\%$ , in perunages 0,0244.

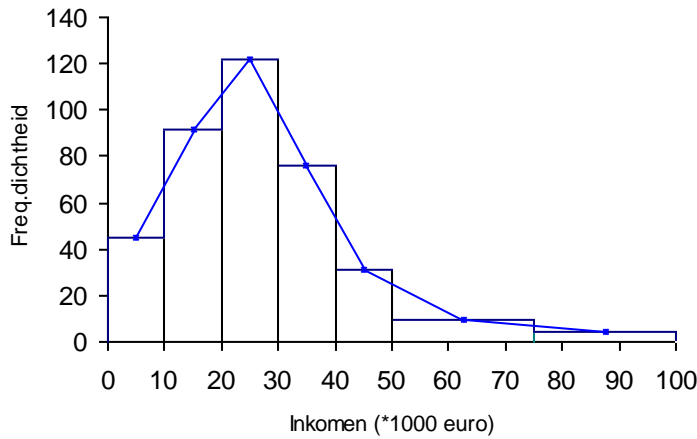
Rendement	Groefactor
-0,102	0,898
0,116	1,116
0,214	1,214
-0,253	0,747
0,157	1,157
0,099	1,099

- c Meetkundig gemiddelde, omdat een rendement een groeivoet is en dus verbonden met een groefactor.
- d Standaardafwijking van de (rekenkundige) percentages is 0,1784

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

Er volgt:  $0,0244 - 0,0385 + 0,1784^2/2 = 0,00143 \approx 0$

2.20a, b

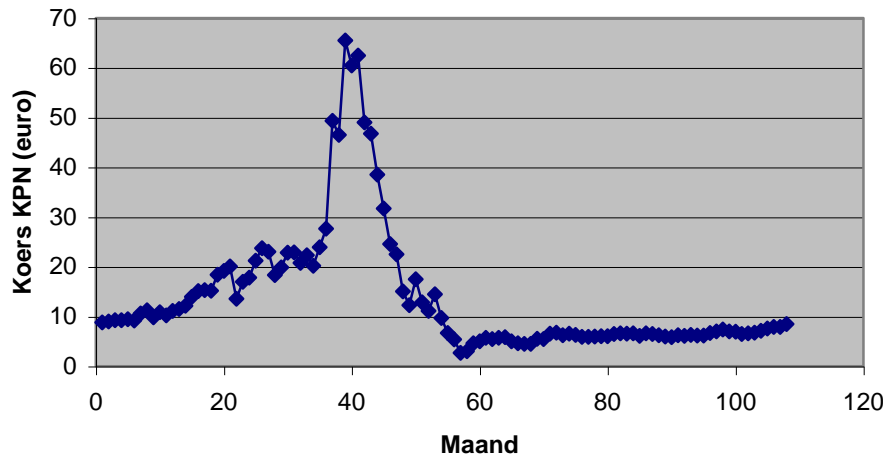


L	R	f	rel f	cum rel f	fd per 10
0	10	45	0,1125	0,1125	45
10	20	92	0,23	0,3425	92
20	30	122	0,305	0,6475	122
30	40	76	0,19	0,8375	76
40	50	31	0,0775	0,915	31
50	75	24	0,06	0,975	9,6
75	100	10	0,025	1	4

400

- c Het gemiddelde inkomen is 27,7 (\* €1000).
- d  $Q1 = 10 + 10*(25 - 11,25)/(34,25 - 11,25) = 16,0$
- e  $s = 17,13$  (\* €1000)
- f Ongeveer  $\frac{1}{2}*122 + 76 + \frac{1}{2}*31 = 153$  personen

2.21a De maandkoersen lopen van 1 jan 1997 tot en met 1 december 2005.



- b De eerste vijf rendementen zijn: 0,0215; 0,0255; 0,0011; 0,0205 en -0,0245. Pas op alle 107 rendement  $AVERAGE(\text{rendementen})$  toe.
- c, d De volatiliteit is de (steekproef)standaardafwijking van de rendementen, berekend met  $STEDEV(\text{rendementen})$ . In onderstaande tabel staan de gemiddelde rendementen en hun volatiliteit van de vijf bedrijven over de periode 1997-2005.

	KPN	BT-Group	Deutsche Telecom	France Telecom	Telefonica
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:

modus < mediaan < gemiddelde

2.23 Te controleren is of het gewogen gemiddelde van de rubrieken 0100: Voedingsmiddelen en alc.vrije dranken tot en met 14000: Consumptie in het buitenland, inderdaad het indexcijfer 114,4 oplevert. Dit klopt:  $11438056/100000 = 114,4$ .

(1) Indexcijfer	(2) Wegingsfactor	(1)*(2) Product
108,3	11082	1200180,6
130,1	3072	399667,2
97,8	6409	626800,2
126,1	21810	2750241
109,9	7139	784576,1
108,6	599	65051,4
119,2	12121	1444823,2
96,4	3643	351185,2
104,5	11471	1198719,5
127,4	106	13504,4
122,3	4425	541177,5
115,4	10161	1172579,4
101,2	3383	342359,6
119,5	4579	547190,5
Totaal	100000	11438056

## 2.24

De maatstaven gemiddelde, standaardafwijking, scheefheid en spitsheid zijn nu, berekend met behulp van *maatstaven.xls*, respectievelijk 11,9; 60,9; 13,6 en 207,8 (zie onderstaande tabel voor gegevens). De verdeling van werknemers is minder spits in 1996 dan in 2006, maar er zijn geen duidelijke verschillen zichtbaar.

Aantal werknemers.	1996	2006	midden
0	2390	2865	0
1 t/m 4	2275	1985	2,5
5 t/m 9	835	630	7
10 t/m 19	550	420	14,5
20 t/m 49	430	390	34,5
50 t/m 99	140	115	74,5
100 t/m 199	55	40	59,5
200 t/m 499	35	30	349,5
500 en meer	20	15	1000*
	6730	6490	

\* Aanname; zie voorbeeld 2.7.

### 3 Elementaire kansrekening

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

3.2  $6^4 = 1296$ ;  $1/6$  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.

3.4a  $4^{10} = 1.048.576$

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

3.5  ${}_5P_3 = 60$  (of  $5 \cdot 4 \cdot 3$ )

3.6a 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.

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

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 allen ook op 360.360 uitkomen.

3.10a 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  ${}_{10}C_3 + {}_{10}C_2 + {}_{10}C_1 + {}_{10}C_0 = 120 + 45 + 10 + 1 = 176$ .

3.12a 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.

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

$$b \quad \text{kans} = \frac{{}_6C_3 \times {}_{39}C_3}{{}_{45}C_6} = 0,0022$$

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

d (uiteraard) gelijk aan 1

$$e \quad \text{kans (minstens 1 goed)} = 1 - \text{kans (geen enkele goed)} = 1 - 0,40056 = 0,5994$$

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

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

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

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

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

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

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

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

$$d \quad \text{P(B of S)} = \text{P(B)} + \text{P(S)} - \text{P(B en S)} = 0,9 + 0,5 - 0,9 \cdot 0,3 = 0,93$$

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

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

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

$$3.18a \text{ kans} = \left(\frac{5}{6}\right)^5 \approx 0,40188$$

$$b \quad \text{kans} = \left(\frac{5}{6}\right)^{10} \approx 0,16151$$

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

$$d \quad \text{kans} = {}_5C_1 \times \frac{1}{6} \times \left(\frac{5}{6}\right)^4 = 0,4019$$



3.19a stel  $g = \text{goed}$ ,  $s = \text{slecht}$ , goedgekeurd = + en afgekeurd = -

$$P(g \mid +) = 0,95 \text{ dus } P(g \text{ en } +) = P(+)\cdot P(g \mid +) = 0,8 \cdot 0,95 = 0,76$$

$$P(g \mid -) = 0,95 \text{ dus } P(g \text{ en } -) = P(-)\cdot P(g \mid -) = 0,2 \cdot 0,08 = 0,016$$

De bijbehorende kruistabel is:

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

b  $P(- \mid g) = \frac{P(- \text{ en } g)}{P(g)} = \frac{0,016}{0,776} \approx 0,02062$  dus circa 2,1%

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

3.20a

	<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.

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

	Studierichting			
Geslacht	BE	CE	MER	totaal
Man	60	150	90	<b>300</b>
Vrouw	40	100	60	<b>200</b>
<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	<b>0,6</b>
Vrouw	0,08	0,2	0,12	<b>0,4</b>
<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  $0,61$

3.24a  $P(M | MER) = \frac{P(M \text{ en } MER)}{P(MER)} = \frac{0,16}{0,36} \approx 0,444$

b  $P(MER | M) = \frac{P(M \text{ en } MER)}{P(M)} = \frac{0,16}{0,49} \approx 0,327$

c Het uitgangspunt is steeds een andere deelpopulatie.

d  $P(CE | V) = \frac{P(CE \text{ en } V)}{P(V)} = \frac{0,17}{0,51} \approx 0,333$

e  $P(V | IF) = \frac{P(V \text{ en } IF)}{P(IF)} = \frac{0,12}{0,23} \approx 0,522$

f  $P(M \text{ en } CE) = 0,13$  en  $P(M) \cdot P(CE) = 0,49 \cdot 0,30 = 0,27$   
 De gebeurtenissen M en CE zijn niet onafhankelijk.

3.25a  $0,51^2 = 0,2601$

b Onderscheid de volgorden: M–V en V–M  
 kans =  $2 \cdot 0,49 \cdot 0,51 = 0,4998$

c kans =  $0,13^2 + 0,11^2 + 0,16^2 + 0,09^2 + 0,17^2 + 0,12^2 + 0,20^2 + 0,02^2 = 0,1464$

3.26a Aantal pakketten  ${}_{24}C_{10} = 1.961.256$

b Aantal pakketten is  ${}_{24}C_5 \cdot {}_{24}C_{10}$ , ofwel circa  $8,336 \cdot 10^{10}$

c bèta(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)

3.27a 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  $0,05 \times 0,2 = 0,01$ ; enz.

b  $P(\text{fouten} | \text{goedkeuren}) = \frac{P(\text{fouten en goedkeuren})}{P(\text{goedkeuren})} = \frac{0,01}{0,7} \approx 0,014$  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.

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

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

c SD is in alle gevallen gelijk

3.29a 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 =  $1/2$ .

3.30a 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:  $\text{-----} +$   
 70.033.600

b 2 letters, 4 cijfers:  $26 \cdot 25 \cdot 10 \cdot 9 \cdot 8 \cdot 7 = 3.276.000$   
 3 letters, 3 cijfers:  $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 = 11.232.000$   
 4 letters, 2 cijfers:  $26 \cdot 25 \cdot 24 \cdot 23 \cdot 10 \cdot 9 = 32.292.000$

Totaal:  $\text{-----} +$   
 = 46.800.000

3.31a 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)

$$3.32a \text{ kans} = \frac{{}_6C_6 \times {}_{43}C_0}{{}_{49}C_6} \approx 7,15 \times 10^{-8} \text{ (d.i. circa 7 op 100 miljoen)}$$

$$b \text{ kans} = \frac{{}_6C_5 \times {}_{43}C_1}{{}_{49}C_6} \approx 0,00002$$

$$c \text{ kans} = \frac{{}_6C_0 \times {}_{43}C_6}{{}_{49}C_6} \approx 0,43596$$

$$d \text{ kans} = 1 - \frac{{}_6C_6 \times {}_{43}C_0}{{}_{49}C_6} \approx 1 - 0,43596 = 0,56404$$

## 4 Kansvariabelen

4.1a Met *maatstaven.xls* of (grafische) rekenmachine: verwachtingswaarde (E) is € 57.000 en de standaardafwijking ( $\sigma$ ) is € 6403,12

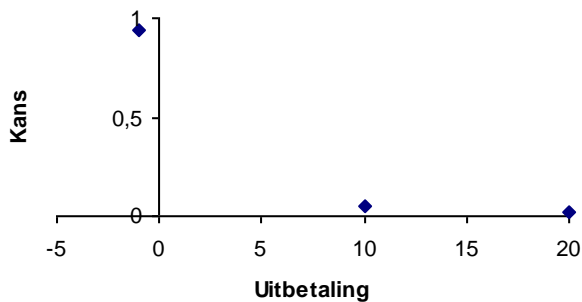
b  $(E - 2 \cdot \sigma; E + 2 \cdot \sigma) = (44.193,75; 69.806,25)$

c 0,9

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

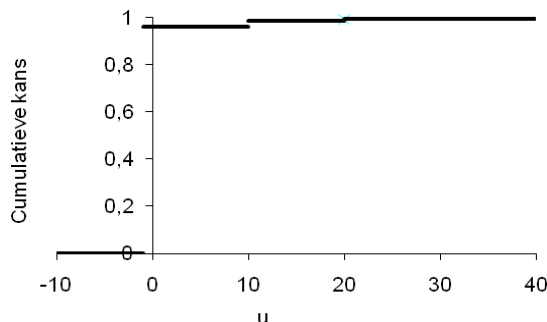
b  $E(\underline{u}) = -0,16$ ;  $\sigma(\underline{u}) = 3,44$  (euro)

c



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

e



4.3a  $E(k) = 1,47$  uur

b  $\sigma(k) = 0,88$  uur

c  $v = 50 + 80 \cdot k$ , dus  $E(v) = 50 + 80 \cdot 1,47 = € 167,6$  en

d  $\sigma(v) = 80 \cdot 0,88 = € 70,2$

e De verwachting wordt 3 keer zo groot en standaardafwijking wordt  $\sqrt{3}$  keer zo groot, dus respectievelijk € 502,8 en € 121,52.

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

Opbrengst	Kans
8499	0,0001
-1	0,9999

- b Kans op het winnende lot is nu  $2 \cdot 0,001 \cdot 0,999 \approx 0,0002$ .  
Met de uitbetalingstabel hieronder volgt een verwachte opbrengst van € -0,30.

Opbrengst	Kans
8498	0,0002
-2	0,9998

- 4.5 a De wasmachine met garantie kost € 615.

Bij het kopen van de wasmachine zonder garantie is de kanstabel voor prijs plus eventuele onkosten:

Prijs	Kans
600	0,95
700	0,05

De verwachte waarde hiervan is  $600 \cdot 0,95 + 700 \cdot 0,05 = € 605$ . De garantieverzekering is dus niet voordeliger voor de koper.

- b Er moet dan gelden:  $600 \cdot 0,95 + x \cdot 0,05 > 615$ . Hieruit volgt  $x > 900$ , dus bij reparatiebedragen boven € 300.

- 4.6a  $E(\underline{s}) = 0,85$  en  $\sigma(\underline{s}) = 1,24$

- b Stel  $\underline{t}$  = aantal storingen per week. Dan is  $E(\underline{t}) = 6 \cdot E(\underline{s}) = 5,1$  en  $\sigma(\underline{t}) = \sqrt{6} \cdot \sigma(\underline{s}) = 3,03$ , onder de aanname dat de storingen onafhankelijk zijn van dag tot dag.

- 4.7a Verwachting en standaardafwijking zijn respectievelijk € 11.050.000 en € 194.670 per kwartaal

- b respectievelijk € 44.200.000 en € 389.400 per jaar

- 4.8a 550.000 resp. 34.928 euro

- b 660.000 resp. 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 € 3.460,00.

- b Standaardafwijking is € 2.511,65.

- c  $(3.460 \pm 2.511,65) = (948,35; 5971,65)$  (euro)

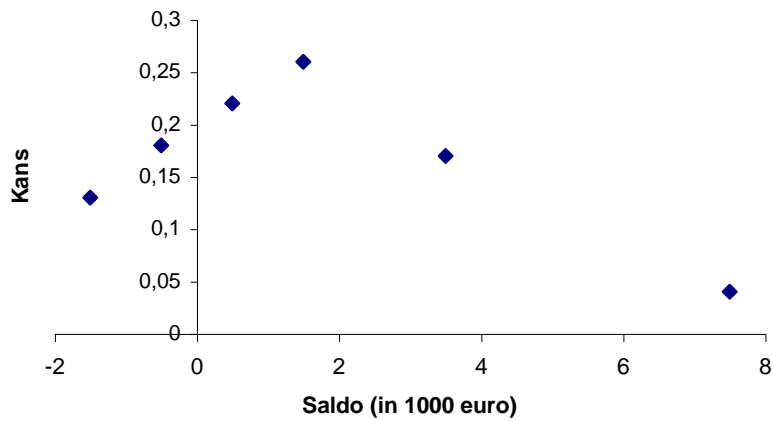
- d  $2000 + 0,12 \cdot (2000/0,26) = € 2923$

- 4.11a

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 Verwachtingswaarde is  $2 \cdot E(\underline{s}) = 2 \cdot 1,11 = 2,22$  ( $\times \text{€ } 1000$ ); standaardafwijking is  $2 \cdot \sigma(\underline{s}) = 2,02 = 4,04$  ( $\times \text{€ } 1000$ )
- d  $10 \cdot E(\underline{s}) = 10 \cdot 1,11 = 11,1$  ( $\times \text{€ } 1000$ );  $10 \cdot \sigma(\underline{s}) = \sqrt{10} \cdot 2,02 = 6,39$  ( $\times \text{€ } 1000$ )
- e De scheefheid is 1,17 (Excel), dus rechtsscheef.
- f De spitsheid is 1,78, dus spitsers dan een vergelijkbare normale verdeling.

4.12a  $E(k) = 0,95$

b  $\sigma(k) = 1,20$

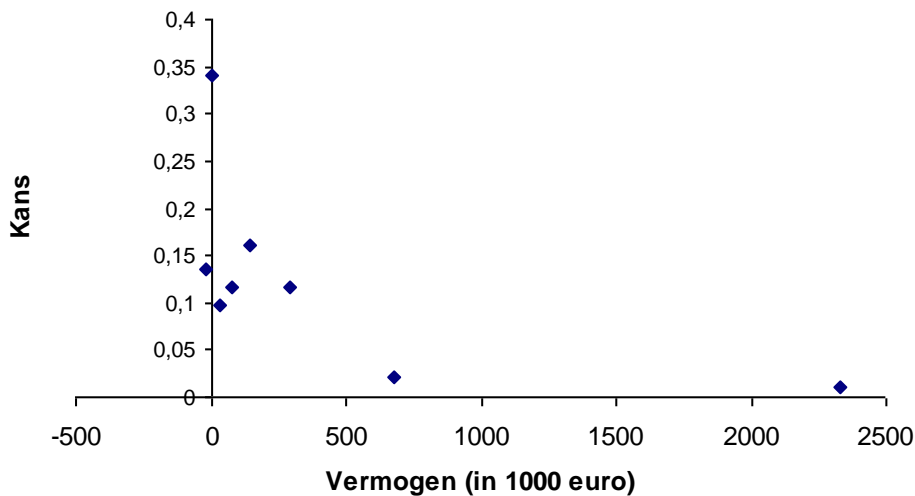
c Een totaal van 14 storingen of meer in 5 dagen kan alleen door de combinaties (3,3,3,3,3) en (3,3,3,3,2). De kans hierop is  $0,2^5 + 5 \cdot 0,2^4 \cdot 0,1 = 0,00112$ .

4.13a De gegeven tabel kent slechts acht mogelijke uitkomsten.

b De acht frequenties zijn gedeeld door de totale frequentie (6.855.000).

Gemidd. vermogen	Kans
-21,4	0,1363
5,6	0,3408
33	0,0973
74,5	0,1155
144,4	0,1616
294,7	0,1158
677,6	0,0222
2333,3	0,0105

- c  $E(v) = 107,82; \sigma(v) = 264,6$  (€ 1000)  
 d



- e Respectievelijk 100 en 10 keer zo groot als de antwoorden bij c, dus 10.782 en 2.646 ( $\times$  € 1000).

4.14a 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  $k_3$  de waarde 2 niet kan aannemen):

$k_3$	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:

$k_4$	Kans
0	0,3750 (9/24)
1	0,3333 (8/24)
2	0,25 (6/24)
4	0,0714 (1/24)

- d  $E(k_4) = 1,12$  en  $\sigma(k_4) = 1,11$ .

4.15a  $E(y) = 15 + 15 = 30; \sigma(y) = \sqrt{3^2 + 3^2} = \sqrt{18} = 4,24$

b  $E(y) = 2 \cdot 15 = 30; \sigma(y) = 2 \cdot \sigma(x) = 2 \cdot 3 = 6$  (NB: merk verschil op met a)

c  $E(y) = 17; \sigma(y) = 3$

d  $E(y) = 15 + 15 + 15 - 2 = 43; \sigma(y) = \sqrt{3^2 + 3^2 + 3^2} = \sqrt{27} = 5,20$

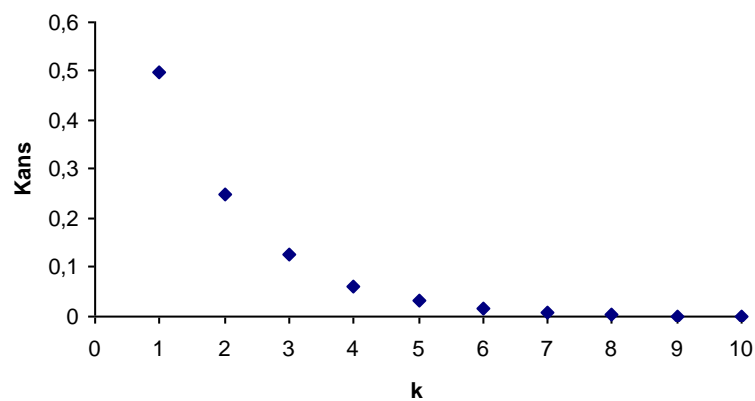
e  $E(y) = 0; \sigma(y) = 1$ . (Dit is standaardiseren; zie hoofdstuk 6.)

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



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

b



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

## 5 De binomiale verdeling

- 5.1a 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 bin. verd. met  $n = 50$  en  $p = 0,1$ ; kijken 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
- 5.2a  $P(\underline{x} < 38) = P(\underline{x} \leq 37) = 0,0299$
- 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  $E(\underline{x}) = n \times p = 76 \times 0,6 = 45,6$
- 5.3a 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.
- 5.4a  $P(\underline{y} \geq 48) = 1 - P(\underline{y} \leq 47) = 1 - 0,2452 = 0,7548$
- 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  $m = 64$ .
- 5.5a 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 (resp. 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.
- 5.6a  $E(\underline{m}) = 600 \times 0,01 = 6$
- 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.
- 5.7a  $P(\underline{w} \geq 16) = 1 - P(\underline{w} \leq 15) = 1 - \text{BINOMDIST}(15; 84; 0,24; 1) = 1 - 0,11471 = 0,88529$
- b  $P(21 \leq \underline{w} \leq 25) = P(\underline{w} \leq 25) - P(\underline{w} \leq 20) = 0,91123 - 0,54335 = 0,36788$
- c M.b.v. 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  $P(\underline{w} \leq m - 1) \geq 0,9$ ; 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 en kurtosis (spitsheid) = -0,006. De standaardafwijking is 3,91.

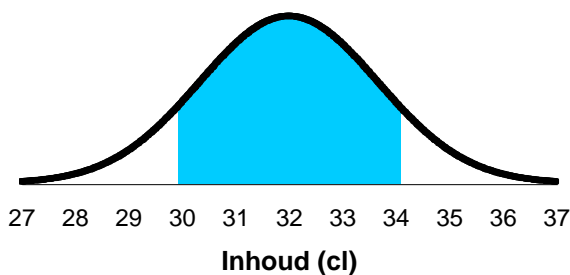
- 5.8a kans =  $0,95^5 = 0,77378$
- b Stel  $g$  = aantal correct uitgevoerde keuringen;  $n = 10$   
 $P(g = 8) = {}_{10}C_8 \cdot 0,95^8 \cdot 0,05^2 = 0,07463$
- c kans =  $0,95^8 \cdot 0,05^2 = 0,00166$
- d  $n = 100$ ;  $P(g = 20) = \text{BINOMDIST}(20; 100; 0,05; 0) = 0,00000$  (calculator geeft  $8,44 \cdot 10^{-8}$ )
- 5.9a 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 =  $-0,18$
- 5.10a 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.
- 5.11a kans =  $0,95^{20} = 0,35849$
- b kans =  ${}_{20}C_2 \cdot 0,95^{18} \cdot 0,05^2 = 0,18868$
- c kans =  $1 - \text{BINOM}(2; 20; 0,05; 1) = 1 - 0,92452 = 0,07548$
- d verwachtingswaarde =  $20 \cdot 0,05 = 1$   
 $SD = \sqrt{(20 \cdot 0,05 \cdot 0,95)} = 0,97468$
- 5.12 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

- 6.1a  $\text{NORMDIST}(40;45;6;1) = 0,2023$   
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  $0,9973$  (of  $\text{NORMDIST}(63;45;6;1) - \text{NORMDIST}(27;45;6;1)$ )  
f  $\text{NORMDIST}(54;45;6;1) - \text{NORMDIST}(36;45;6;1) = 0,8664$

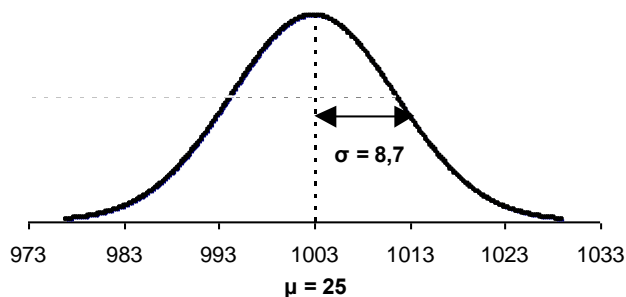
6.2

- 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



- 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.
- 6.3a  $\text{NORMDIST}(10;13;1,1;1) = 0,0032$ . NB: de variabelen  $x$ ,  $\mu$  en  $\sigma$  zijn een factor 1000 kleiner gemaakt. Dat is niet van invloed op het antwoord. Zie voor een verklaring paragraaf 63 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  $Y = \text{€ } 12.704$   
d  $\text{NORMINV}(0,9;13;1,1) = 14,410$ ,  $Z = \text{€ } 14.410$
- 6.4a  $P(\underline{R} < 0) = \text{NORMDIST}(0;8,6;6,2;1) = 0,0827$   
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  $(\mu \pm 1,96\sigma) = (0,086 \pm 1,96 \cdot 0,062) = (-0,036; 0,208)$ , dus tussen -3,6% en 20,8%.
- 6.5a  $1 - \text{NORMDIST}(5500;5000;400;1) = 0,1056$   
b  $\text{NORMINV}(0,999;5000;400) = 6236$  liter
- 6.6a  $1 - \text{NORMSDIST}(1,4) = 0,0808$   
b  $1 - \text{NORMSDIST}(-0,123) = 0,5489$ ; of in z-tabel opzoeken:  $1 - P(z > 0,123) = 1 - 0,4511$   
c 0,5, want het is precies de helft van de standaardnormale verdeling (of  $1 - \text{NORMSDIST}(0) = 0,5$ )  
d  $\text{NORMSDIST}(2,1) - \text{NORMSDIST}(-0,1) = 0,5220$   
e Letterlijk in Excel:  $\text{NORMSDIST}(-1,7) + 1 - \text{NORMSDIST}(1,7) = 0,0891$ , maar ook bijvoorbeeld met z-tabel:  $2 \cdot P(z > 1,7) = 2 \cdot 0,0446 = 0,0891$  (verschil door afronding in z-tabel)
- 6.7a  $p = \text{NORMSINV}(0,75) = 0,674$   
b  $q = \text{NORMSINV}(0,99) = 2,326$   
c  $r = \text{NORMSINV}(0,95) = 1,645$ .

- 6.8 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 z-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.
- 6.9 We moeten controleren:  $P(\underline{x} < 485 \mid N(499,8; 2,17)) < 0,1$ . Dit kan in Excel met:  $\text{NORMDIST}(485;499,8;2,17;1) = 4,6 \times 10^{-12}$  ( $= 4,6/10^{12}$ ). Dit is zeker kleiner dan 0,1.
- 6.10 Gebruik het stelsel vergelijkingen op p. 175, of de Excel-methode (p. 175/176).  
Antwoord:  $\mu = 500,7$  en  $\sigma = 6,9$  (gram).
- 6.11 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:  $\text{NORMDIST}(500;550;34,928;1) = 0,0761$
- 6.12 De winstvariabele  $\underline{w}$  is normaal verdeeld met  $\mu = 0,3$  en  $\sigma = 0,264197$  (beide in milj. euro's).
- $1 - \text{NORMDIST}(2,75;2,5;0,23;1) = 0,1385$
  - $1 - \text{NORMDIST}(1;0,3;0,264197;1) = 0,0040$
  - $\text{NORMDIST}(0;0,3; 0,264197;1) = 0,1281$
- 6.13a  $1 - \text{NORMDIST}(1,5;1,3; 0,077519;1) = 0,0049$  (somvariabele met  $\mu = 13 \cdot 100.000$  en  $\sigma = \sqrt{13} \cdot 21.500$ , in miljoenen ingevuld in Excel)
- $\text{NORMDIST}(1,5;1,3; 0,077519;1) - \text{NORMDIST}(1000;1300; 0,077519;1) = 0,9950$
  - $1 - \text{NORMDIST}(1,365;1,3; 0,077519;1) = 0,2009$
- 6.14a  $1 - \text{NORMDIST}(120,7;120;2;1) = 0,3632$
- $0,3632^3 = 0,0479$
  - $1 - \text{NORMDIST}(362,1;360;\sqrt{3} \cdot 2;1) = 0,2722$
- d Voor de somvariabele in c zijn meer combinaties mogelijk, niet alledrie hoeven meer te wegen dan 120,7 kg, slechts het gemiddelde hoeft meer te zijn dan 120,7 kg; dus een grotere kans dan bij b.
- 6.15a



- $(\mu \pm 1,96\sigma) = (1003 \pm 1,96 \cdot 8,7) = (985,9; 1020,1)$
  - $\text{NORMDIST}(1000;1003;8,7;1) = 0,3651$
  - $\text{NORMDIST}(12000;12036;8,7 \cdot \sqrt{12};1) = 0,1161$
  - $\text{NORMDIST}(24000;24072;8,7 \cdot \sqrt{24};1) = 0,0456$
- f
- c: P(1 pak) < 1 kg
  - d: P(12 pakken) < 12 kg
  - e: P(24 pakken) < 24 kg

Omdat het aantal ( $n$ ) in deze rij toeneemt, vermindert de standaardafwijking van het gemiddelde (' $\sqrt{n}$ -wet!'). Er zijn dus minder mogelijkheden, dus de kansen worden kleiner.

- 6.16 Noem  $b$  de bezoektijd,  $r$  de reistijd tussen twee klanten en  $c$  de cyclustijd ( $= b + r$ ).
- a, b Er geldt:  $c \sim N(50+30; \sqrt{(15^2+10^2)}) = N(80; \sqrt{325})$ . Gemiddelde (verwachtingswaarde) en standaardafwijking zijn dus respectievelijk 80 en  $\sqrt{325} = 18,0$  minuten.
- c  $P(c > 100 | N(80; 18,0)) = P(z > 1,11) = 0,1335$ ; of  $\text{normalcdf}(100, 10^10, 80, 18)$ ; of met Excel:  $1 - \text{NORMDIST}(100; 80; \text{sqrt}(325); 1) = 0,1336$
- 6.17a  $P(\text{omzet} > 170.000) = 1 - \text{NORMDIST}(170; 150; 12; 1) = 0,0478$ . Of met grafische rekenmachine:  $\text{normalcdf}(170, 10^10, 150, 12) = 0,0478$ . Of met z-tabel:  $P(z > [170.000 - 150.000]/12.000) = P(z > 1,67) = 0,0475$  (afwijkend antwoord wegens afronding van  $z$  op 2 decimalen)
- b  $P(\underline{w} > 50.000 | N(40.000; \sqrt{(12.000^2 + 9.000^2)}) = 1 - \text{NORMDIST}(50; 40; 15; 1) = 0,2525$ . Kan ook met:  $P(\underline{w} > 50.000 | N(40.000; 15.000)) = P(z > 0,67) = 0,2514$
- c  $P(\underline{w} < 0 | N(40.000; 15.000)) = \text{NORMDIST}(0; 40; 15; 1) = 0,0038$ . De vuistregelintervallen ( $\mu \pm 2\sigma$ ) voor omzet respectievelijk kosten zijn (in 1000en):  $(150 \pm 2 \times 12)$  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.
- 6.18a  $1 - \text{NORMDIST}(25; 21; 3; 1) = 0,0912$
- 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.
- 6.19a  $\text{BINOMDIST}(50; 200; 0,3; 1) = 0,0695$
- 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.
- 6.20a  $1 - \text{NORMDIST}(186; 181,2; 8,1; 1) = 0,2767$
- 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:  $V_{\text{man}} = 181,2/8,1 = 22,4$ ;  $V_{\text{vrouw}} = 168/8,6 = 19,8$ , dus bij de mannen.
- 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:  
 $P(150)^2 + P(151)^2 + \dots + P(200)^2 = 0,0348$
- 6.21a  $Q_1 = \text{NORMINV}(0,25; 70; 10) = 63,3$ ;  $Q_3 = \text{NORMINV}(0,75; 70; 10) = 76,7$
- b  $P_{20} = \text{NORMINV}(0,2; 70; 10)$  of  $\text{NORMINV}(20\%; 70; 10) = 61,6$   
 $P_{40} = \text{NORMINV}(0,470; 10) = 67,5$ ; idem:  $P_{60} = 72,5$ ;  $P_{80} = 78,4$
- c Bereken  $P(\underline{x} < 50 | N(70; 10)) = \text{NORMDIST}(50; 70; 10; 1) = 0,0228$ ;

$P(50 < \bar{x} < 60 | N(70;10)) = \text{NORMDIST}(60;70;10;1) - \text{NORMDIST}(50;70;10;1) = 0,1359$ , etc.  
Dit leidt tot onderstaande tabel, en vermenigvuldigd met 500, tot de verwachte frequenties.

d

Onderdeel c		Onderdeel d	
kans	frequentie	kans	frequentie
0,0228	11	0,0062	4
0,1359	68	0,0606	36
0,3413	171	0,2417	145
0,3413	171	0,3829	230
0,1359	68	0,2417	145
0,0228	11	0,0606	36
		0,0062	4

6.22a  $P(\bar{b}_1 > 15) = 1 - \text{NORMDIST}(15;9;10;1) = 0,2743$

b  $P(\bar{b}_{10} > 15) = 1 - \text{NORMDIST}(15;9;10/\sqrt{10};1) = 0,0289$  ( $\sqrt{n}$ -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  $P(\bar{b}_{10} - \bar{b}_{10} > 0) = 1 - \text{NORMDIST}(0;6;\sqrt{(0,5^2+10^2)}/\sqrt{10};1) = 0,9710$

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; onafhankelijkheid rendementen van jaar tot jaar.

h Bij b, d en g: de  $\sqrt{n}$ -wet voor gemiddelden zegt dat deze verdelingen bij benadering normaal zijn voor grotere n.

## 7 Lineaire regressie

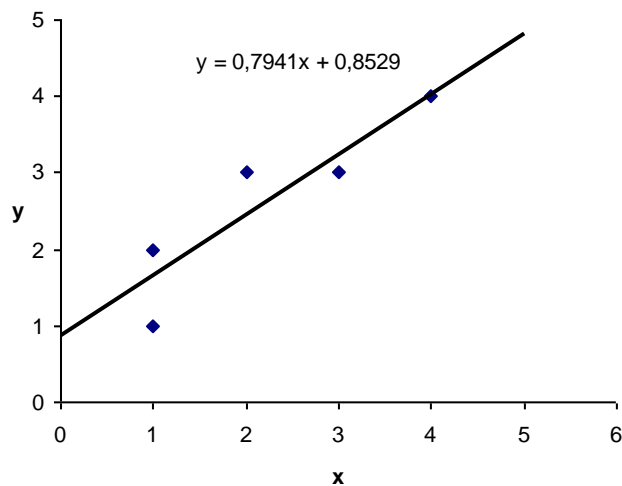
7.1a. Bereken met Excel:  $\text{INTERCEPT}(B2:B8;A2:A8) = 11.492$  en  $\text{SLOPE}(B2:B8;A2:A8) = 64,459$ . De vergelijking van de regressielijn is dus  $\text{TK} = 64,459q + 11.492$

b Het zwaartepunt (gemiddelde van q- respectievelijk TK-coördinaat) is (724,29; 58.179).

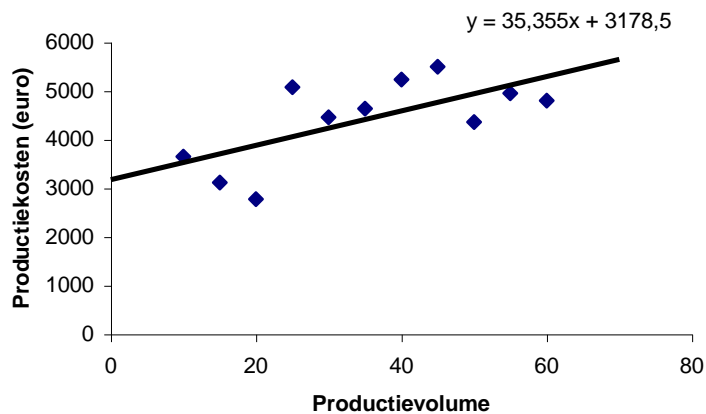
Ingevuld in de regressielijn uit onderdeel a levert dit een ware bewering op:  
 $\text{TK} = 64,459 \cdot 724,29 + 11.492 = 46.687,01 + 11.492 = (\text{afgerond}) 58.179$ .

	A	B
1	q	TK
2	625	58250
3	495	60500
4	500	46500
5	1050	67500
6	1350	101500
7	750	59000
8	300	14000

7.2a, b, c



7.3a





- b De vaste kosten zullen bij benadering € 3.178 zijn; de variabele kosten kunnen geschat worden door € 53,35 per stuk.
- c  $TK(70) = 3.178,50 + 35,354 \cdot 70 = € 5.653,32$

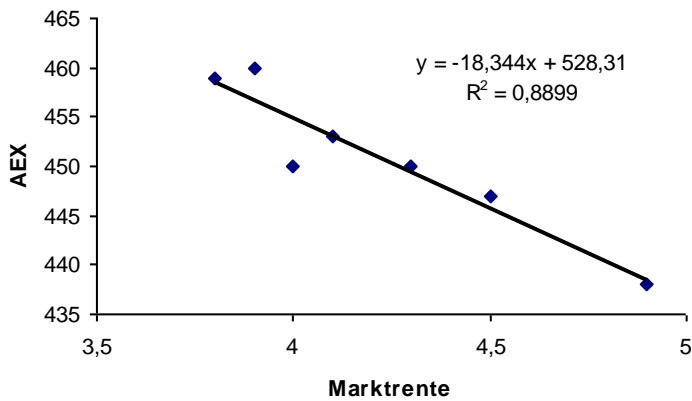
7.4abc De kwadratensommen zijn respectievelijk 3, 5,25 en  $11/9 = 1,22$ .

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

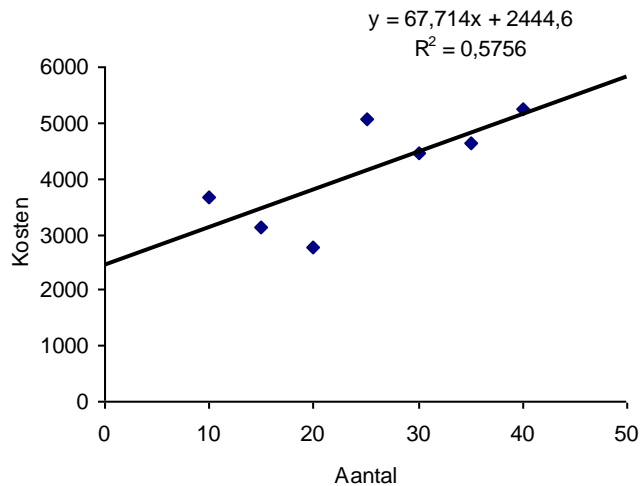
7.5a De marktrente is de verklarende (onafhankelijke) variabele en de AEX is de verklaarde (onafhankelijke) variabele.

	A	B	C	D	E	F	G	H
1	Tijdstip	1	2	3	4	5	6	7
2	Marktrente (%)	4,5	3,9	3,8	4,9	4	4,1	4,3
3	Beursindex	447	460	459	438	450	453	450

- b  $b = \text{SLOPE}(\text{Beursindex}; \text{Marktrente}) = \text{SLOPE}(B3:H3; B2:H2) = -18,344$ ;  $a = \text{INTERCEPT}(\text{Beursindex}; \text{Marktrente}) = 528,31$ .
- c  $r = \text{CORREL}(\text{Beursindex}; \text{Marktrente}) = \text{CORREL}(B3:H3; B2:H2) = -0,94$
- d  $\text{AEX}(5) = 528,31 - 18,344 \cdot 5 = 436,6$ , afgerond op hele cijfers (zoals de reeks zelf ook): 437.
- e De determinatiecoëfficiënt is 0,8899. De wortel hieruit is de correlatiecoëfficiënt.

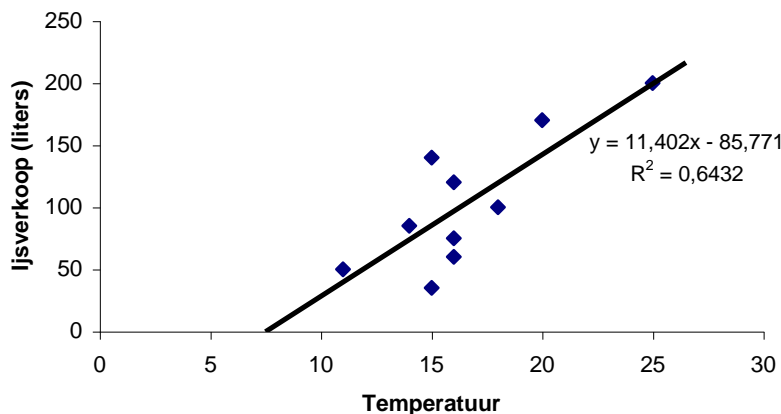


- 7.6a De waarde  $b = 1,772$  uit de regressievergelijking is de richtingscoëfficiënt van het lineaire verband  $y = a + b \cdot x$ .
- b De correlatiecoëfficiënt is mate van lineaire samenhang van de jaarrendementen van Haast NV en de markt. De uitkomst 0,964 betekent een zeer sterk verband.
- c Een 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:  $5,6 = -3,793 + 1,772 \cdot 5,3$ .
- 7.7a Het productievolume is de onafhankelijke variabele en de grootte van de productiekosten is de afhankelijke variabele.
- b  $\text{Kosten} = 67,714 \cdot \text{Aantal} + 2.444,6$  (€)



- 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.

- 7.8a 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 redelijk goede positieve 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 °C is  $y(30) = -85,77 + 11,40 \cdot 30 = 256$  liter ijs

- 7.9a 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  $\text{SLOPE}(x_1; x_2) = 0,5$

7.10a  $\approx 0,9$

b  $\text{CORREL}(x_2; x_1) = 0,82$ .

c  $\approx 1,5$

d  $\text{SLOPE}(x_2; x_1) = 0,85$ . Dit betekent dat de reeks  $x_2$  1,25 keer zo hard schommelt als reeks  $x_1$ .

e  $\text{SLOPE}(x_1; x_2) = 0,54$  (dus niet  $1/1,25 = 0,8!$ )

7.11a

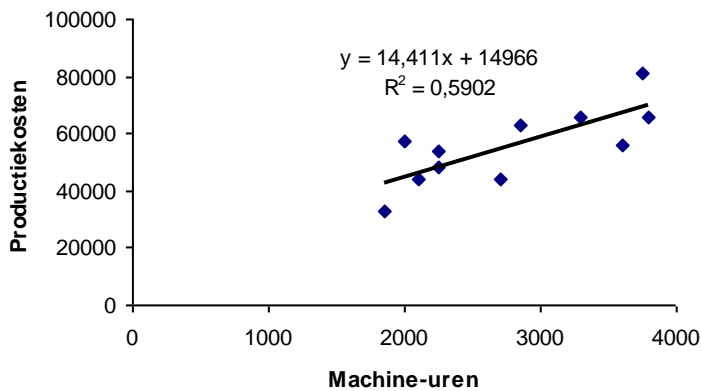
Uitgeverij C	
a =	4,63
b =	0,07
r =	0,87

Uitgeverij W	
a =	22,82
b =	0,062
r =	0,88

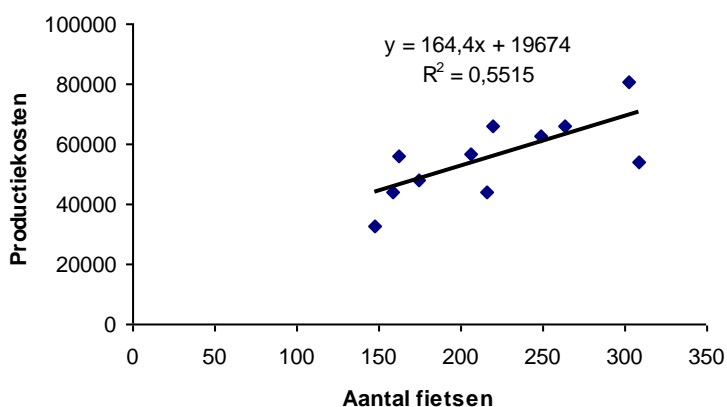
b Voor beide regressielijnen zijn de correlatiecoëfficiënten dicht bij 0,9, dus goede correlatie.

c Uitgeverij C heeft de laagste prijs per pagina.

7.12a



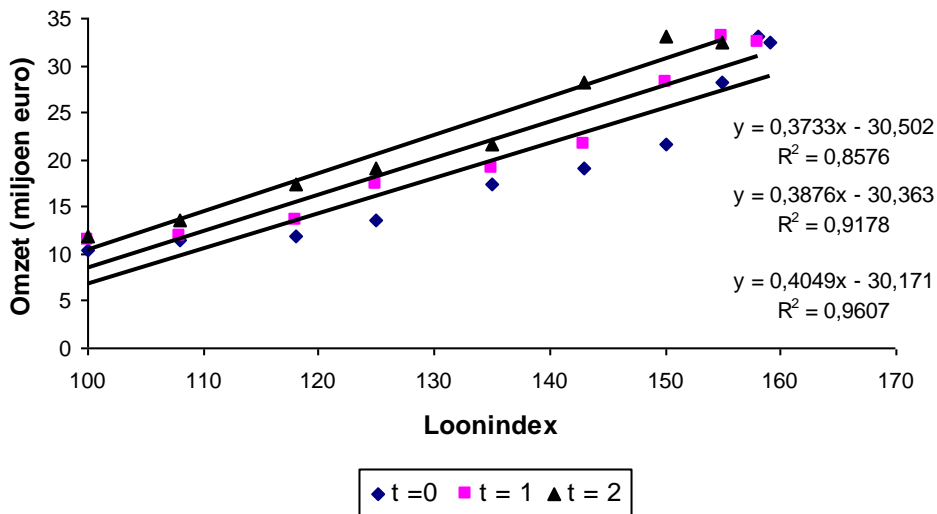
b



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

7.13a, b, c

Met behulp van de trendlijnmethode in Excel.



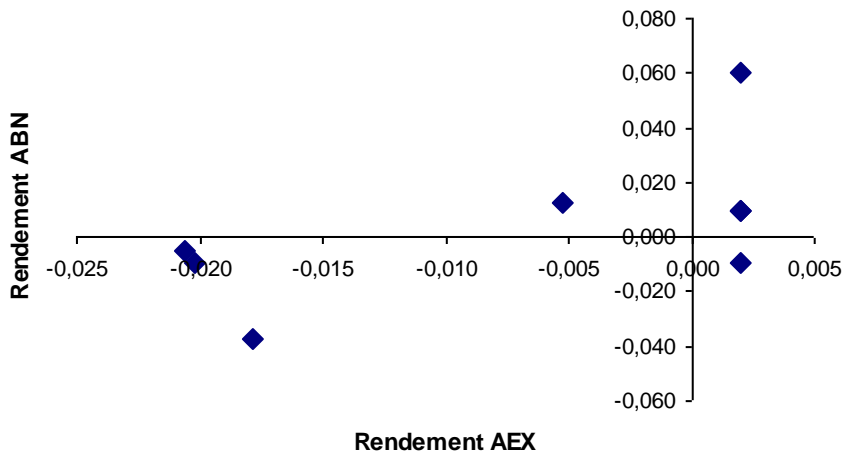
De correlatiecoëfficiënten (de wortel uit de  $R^2$ -waarden) zijn respectievelijk 0,93; 0,96 en 0,98.

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

7.14a

Dag	Index		Rendementen	
	ABN	AEX	ABN	AEX
1	100	100		
2	106	100,2	0,060	0,002
3	107	100,4	0,009	0,002
4	108	100,6	0,009	0,002
5	107	100,8	-0,009	0,002
6	103	99	-0,037	-0,018
7	102	97	-0,010	-0,020
8	101,5	95	-0,005	-0,021
9	102,8	94,5	0,013	-0,005

b



- c  $R(\text{ABN}) = 0,0147 + 1,5526 \cdot R(\text{AEX})$ , met  $r = 0,60$ . Matige correlatie.  
 d Indexcijfers geven de groei van de koers aan t.o.v. een vast gekozen koers; rendementen geven de groei van de koers van dag tot dag.

7.15a  $y = 50 + 5x$ , met de geschatte punten: (0,50) en (300;1500)

b  $y_0 = 50 + 5 \cdot 150 = 800$ ; 2,5s-interval is  $(800 \pm 2,5 \cdot 450,9) = (-357; 1927)$

c Het punt (150; 3400) is een uitschieter voor de y-waarde, want  $3400 > 1927$ .

d Het punt 'rechtsonder', (280;550), zou een uitschieter kunnen zijn voor de x-waarde. Daarvoor zou gecontroleerd moeten worden of 280 buiten het 2,5s-interval voor x ligt.

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'):

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

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

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

7.17a

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
Appollobuurt	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 ('kleinste kwadraten 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.

7.19a Er moet een aselechte steekproef onder de personeelsleden worden genomen.

b  $a = \bar{y} - b \cdot \bar{x} = 211,7 - 0,509 \cdot 339,3 = 39,0$  (ze appendix A10).

c geeft de toename van de kosten aan bij een toename van de afstand met 1 km.

d  $k = 39,0 + 0,509 \cdot r = 39,0 + 0,509 \cdot 450 = 268,50$  (euro)

7.20 a, b en d worden alle 1,5 keer zo groot; c blijkt gelijk

7.21a  $(230 - 218)/218 = 0,05504$ , dus 5,5%;  $(120 - 109)/109 = 0,1009$ , dus 10,1%

b 5,5; 7,8; 0,8; 12,8; 5,3; -3,0; 0,7; 10,3; 11,3; 4,2 (rendement beurs)  
10,1; 10,8; 5,3; 28,5; 10,6; -9,0; 9,4; 13,6; 13,3; 9,8 (rendement NKP)

c Het rendement van de beurs is de onafhankelijke variabele.

d  $r_{\text{NKP}} = 1,65 + 1,54 \cdot r_{\text{beurs}}$

e De  $\beta$  van het aandeel NKP is 1,54; het aandeel NKP is dus gevoelig voor marktschommelingen.

f  $r = 0,86$

g  $r$  geeft aan hoe goed de puntenparen bij lineaire regressielijn liggen (in relatieve zin).

7.22a regressieconstante =  $\bar{b} - \text{regressiecoëfficiënt} \cdot \bar{a} = 9,2 - 1,14 \cdot 7,7 = 0,422$

b Wat de toename van  $b$  is bij een toename van  $a$  met 1.

c  $b = 0,422 + 1,14 \cdot a = 0,422 + 1,14 \cdot 10,5 = 12,4$

7.23a  $\text{CORREL}(KPN, \text{British Telecom}) = 0,4360$

b  $\text{CORREL}(KPN, \text{Deutsche Telecom}) = 0,2402$   
 $\text{CORREL}(KPN, \text{France Telecom}) = 0,1644$ , dus de kleinste  
 $\text{CORREL}(KPN, \text{Telefonica}) = 0,2363$

7.24 De correlatiecoëfficiënt is 0,24; er is dus nauwelijks lineair verband tussen de variabelen.