1.1 Oplossingen

Oplossing oefening 2.1

1. De variabele ‘geslacht’ is een dichotome nominale variabele: nominaal omdat het kenmerk ongeordend categoriserend gemeten wordt en dichotoom omdat de veranderlijke slechts twee nominale waarden aanneemt.

2. De tweede vraag meet het ‘geboortejaar’. Het is een kwantitatieve variabele, meer bepaald op intervalniveau. In tegenstelling tot de voorgaande variabele is hier wel sprake van een ordening (de ene persoon kan later in de jaartelling geboren zijn dan een ander individu). Bovendien is er een betekenisvolle meeteenheid maar geen absoluut nulpunt.

3. De variabele is ‘het dagelijks aantal keren afleggen van de weg tussen woon- en werkplaats’ lijkt op het eerste zicht kwantitatief maar wordt hier ordinaal gemeten. We kunnen weliswaar een ordening aanbrengen maar we hebben geen gelijke intervallen: het verschil tussen de antwoordcategorieën ‘1’ en ‘2’ is niet gelijk aan het verschil tussen ‘2’ en ‘3 of meer’. Het is bovendien een discrete variabele: enkel 1, 2 en 3 zijn legitieme meetwaarden.

4. Met de vierde vraag wil men ‘het aantal dagen per week waarop de weg tussen woon- en werkplaats wordt afgelegd’ meten. Het gaat om een kwantitatieve variabele: het kenmerk ordening is aanwezig (we kunnen spreken in termen van meer of minder dagen) en er is een betekenisvolle meeteenheid.

De variabele is ‘de wijze van verplaatsen’. We hebben hier verschillende antwoordcategorieën die niet te ordenen zijn. Het betreft hier een nominale veranderlijke.

6.De ‘afgelegde afstand’ is het kenmerk waarnaar gepeild wordt in de zesde vraag. Het is een kwantitatieve variabele gemeten op het rationiveau: ordening is mogelijk, er is een betekenisvolle meeteenheid en tenslotte is er ook een absoluut nulpunt.

Oplossing oefening 2.2

1) $\sum_{i=2}^{5} x_i =$ 1 + 2 + 6 + 5 = 14

2) $\sum_{i=1}^{6} x_i y_i =$ (3 x 10) + (1 x 5) + (2 x 4) + (6x 7) + (5 x 12) + (7 x 5) = 180

3) $\sum_{i=2}^{5} (3x_i + 1) = \sum_{i=2}^{5} 3x_i + \sum_{i=2}^{5} 1$

$= 3 \sum_{i=2}^{5} x_i + 1 \text{ x } (5 - 2 + 1)$

= 3 (1 + 2 + 6 + 5) + 4

= 46

4) $\sum_{i=1}^{6} (x_i)^2 =$ $3^2 + 1^2 + 2^2 + 6^2 + 5^2 + 7^2 = 124$

$$\left(\sum_{i=1}^{6} x_i\right)^2 = \quad (3+1+2+6+5+7)^2 = 24^2 = 576$$

5) $$\sum_{i=2}^{5}\sum_{j=1}^{4}\left[(3i-2j^3)+2(3+j)^3\right]$$

$$= \sum_{i=2}^{5}\sum_{j=1}^{4}\left[(3i-2j^3)+2(3^3+3\times 3^2\times j+3\times 3\times j^2+j^3)\right]$$

$$= \sum_{i=2}^{5}\sum_{j=1}^{4}\left[(3i-2j^3)+2(27+27j+9j^2+j^3)\right]$$

$$= \sum_{i=2}^{5}\sum_{j=1}^{4}\left[3i-2j^3+54+54j+18j^2+2j^3\right]$$

$$= \sum_{i=2}^{5}\sum_{j=1}^{4}\left[3i+54j+18j^2+54\right]$$

$$= \sum_{i=2}^{5}\sum_{j=1}^{4}3i + \sum_{i=2}^{5}\sum_{j=1}^{4}54j + \sum_{i=2}^{5}\sum_{j=1}^{4}18j^2 + \sum_{i=2}^{5}\sum_{j=1}^{4}54$$

$$= 3\times 4\sum_{i=2}^{5} i + 54 \text{ x } 4 \sum_{j=1}^{4} j + 18 \text{ x } 4 \sum_{j=1}^{4} j^2 + 54 \text{ x } 4 \text{ x } 4$$

= 12 (2 + 3 + 4 + 5) + 216 (1 + 2 + 3 + 4) + 72 (1² + 2² + 3² + 4²) + 864

= 168 + 2160 + 2160 + 864

= 5352

6) $$\frac{a^2 y_2}{x_2^2} + \frac{a^2 y_3^{3/2}}{x_3} + a^2 y_4^2$$

$$= a^2 x_2^{-2} y_2^1 + a^2 x_3^{-1} y_3^{3/2} + a^2 x_4^0 y_4^2$$

$$= a^2\left(x_2^{-2} y_2^1 + x_3^{-1} y_3^{3/2} + x_4^0 y_4^2\right)$$

$$= a^2 \sum_{i=2}^{4} x_i^{i-4} y_i^{i/2}$$

Oplossing oefening 2.3

1. Geboortejaar = interval

2. Geslacht = nominaal

3. Gewicht = ratio

4. Euro = ratio

5. Werelddeel = nominaal

6. Aantal handlangers in 3 categorieën = ordinaal

Oplossing oefening 2.4

$$\sum_{i=1}^{4}(i-2)^2 = \sum_{i=1}^{4}(\mathrm{i}^2 - 2.2.i + 2^2)$$

$$\sum_{i=1}^{4}\mathrm{i}^2 - \sum_{i=1}^{4}2.2.i + \sum_{i=1}^{4}4 = \sum_{i=1}^{4}\mathrm{i}^2 - 2.2\sum_{i=1}^{4}i + 4\sum_{i=1}^{4}$$

$$\sum_{i=1}^{4}i^2 - 4\sum_{i=1}^{4}i + 4.4 = (1^2 + 2^2 + 3^2 + 4^2) - 4(1+2+3+4) + 16$$

$$30 - 40 + 16 = 6$$

$$\sum_{i=2}^{6}(2i-i)^3 = \sum_{i=2}^{6}((2\mathrm{i})^3 - 3.(i).(2i)^2 + 3.(i)^2.(2i) - (i)^3)$$

$$\sum_{i=2}^{6}(2\mathrm{i})^3 - 3\sum_{i=2}^{6}(i).(2i)^2 + 3\sum_{i=2}^{6}(i)^2.(2i) - \sum_{i=2}^{6}(i)^3$$

$$\sum_{i=2}^{6}8i^3 - 3\sum_{i=2}^{6}4i^3 + 3\sum_{i=2}^{6}2i^3 - \sum_{i=2}^{6}i^3$$

$$8\sum_{i=2}^{6}i^3 - 3.4\sum_{i=2}^{6}i^3 + 6\sum_{i=2}^{6}i^3 - \sum_{i=2}^{6}i^3$$

$$(8.440) + (12.440) + (6.440) - 440 = 440$$

$$\sum_{i=1}^{3}i^4 - 1 = (1^4 + 2^4 + 3^4) - 1$$

$$(1+16+81) - 1 = 98 - 1 = 97$$

$$\sum_{i=1}^{3}(2+3-3i)^2 = \sum_{i=1}^{3}(5-3i)^2$$

$$\sum_{i=1}^{3}(5^2 - 2.5.(3i) + (3i)^2)$$

$$\sum_{i=1}^{3}25 - \sum_{i=1}^{3}30i + \sum_{i=1}^{3}9i^2$$

$$(25.3) - 30(1+2+3) + 9(1^2+2^2+3^2)$$

$$75 - 180 + 126 = 21$$

$$\sum_{i=3}^{6}(2i+5-i^2) = \sum_{i=3}^{6}2i + 5\sum_{i=3}^{6} - \sum_{i=3}^{6}i^2$$

$$2\sum_{i=3}^{6}i + (5.4) - (3^3+4^3+5^3+6^3)$$

$$36 + 20 - 86 = -30$$

$$\sum_{i=1}^{6}2(i-2i^2)^2 = 2\left[\sum_{i=1}^{6}(i^2 - 2.(2i^2).(i) + (2\mathrm{i}^2)^2)\right]$$

$$2\left[\sum_{i=1}^{6}i^2 - 2.2\sum_{i=1}^{6}i^3 + 4\sum_{i=1}^{6}i^4\right]$$

$$2\left[(1^2+2^2+3^2+4^2+5^2+6^2) - 4(1^3+2^3+3^3+4^3+5^3+6^3) + 4(1^4+2^4+3^4+4^4+5^4+6^4)\right]$$

$$2\left[91 - 4(441) + 4(2275)\right] = 14854$$

$$\sum_{i=2}^{5}\left[(i+2i)^2-(i-2)\right]$$

$$\sum_{i=2}^{5}\left[(i^2+2.2i.i+(2i)^2)-i+2)\right]$$

$$\sum_{i=2}^{5}i^2+4\sum_{i=2}^{5}i^2+4\sum_{i=2}^{5}i^2-\sum_{i=2}^{5}i+2\sum_{i=2}^{5}$$

$$9\sum_{i=2}^{5}i^2-\sum_{i=2}^{5}i+2.4=9(54)-(2+3+4+5)+(2.4)$$

$$486-14+8=480$$

$$\sum_{i=1}^{3}\sum_{j=1}^{5}(i+2j)=\sum_{i=1}^{3}\sum_{j=1}^{5}i+2\sum_{i=1}^{3}\sum_{j=1}^{5}j$$

$$5\sum_{i=1}^{3}i+2.3\sum_{j=1}^{5}j=5(1+2+3)+6(1+2+3+4+5)$$

$$30+90=120$$

$$\sum_{i=1}^{3}\sum_{j=3}^{5}(2\mathrm{i}+3\mathrm{j})^2$$

$$\sum_{i=1}^{3}\sum_{j=3}^{5}\left[(2i)^2+2.(3j).(2i)+(3j)^2\right]$$

$$\sum_{i=1}^{3}\sum_{j=3}^{5}(2i)^2+2\sum_{i=1}^{3}\sum_{j=3}^{5}(3j).(2i)+\sum_{i=1}^{3}\sum_{j=3}^{5}(3j)^2$$

$$\sum_{i=1}^{3}\sum_{j=3}^{5}4i^2+2\left[\left(\sum_{j=3}^{5}3j\right).\left(\sum_{i=1}^{3}2i\right)\right]+\sum_{i=1}^{3}\sum_{j=3}^{5}9j^2$$

$$4.3\sum_{i=1}^{3}i^2+2\left[\left(3\sum_{j=3}^{5}j\right).\left(2\sum_{i=1}^{3}i\right)\right]+9.3\sum_{j=3}^{5}j^2$$

$$12(1^2+2^2+3^2)+2\left[3(3+4+5).2(1+2+3)\right]+27(3^2+4^2+5^2)$$

$$(12.14)+(2.432)+(27.50)=2382$$

$$\sum_{i=2}^{4}\sum_{j=1}^{4}\left[(i^2+4)+3(i^2-j)^2\right]$$

$$\sum_{i=2}^{4}\sum_{j=1}^{4}\left[i^2+4+3((i^2)^2-2.(\mathrm{i}^2).(\mathrm{j})+(\mathrm{j})^2\right]$$

$$\sum_{i=2}^{4}\sum_{j=1}^{4}\left[i^2+4+3i^4-6.\mathrm{i}^2.\mathrm{j}+3\mathrm{j}^2\right]$$

$$\sum_{i=2}^{4}\sum_{j=1}^{4}i^2+\sum_{i=2}^{4}\sum_{j=1}^{4}4+3\sum_{i=2}^{4}\sum_{j=1}^{4}i^4-6\sum_{i=2}^{4}\sum_{j=1}^{4}i^2.j+3\sum_{i=2}^{4}\sum_{j=1}^{4}j^2$$

$$4\sum_{i=2}^{4}i^2+(4.3.4)+3.4\sum_{i=2}^{4}i^4-6\left[\left(\sum_{i=2}^{4}i^2\right).\left(\sum_{j=1}^{4}j\right)\right]+3.3\sum_{j=1}^{4}j^2$$

$$4(2^2+3^2+4^2)+(48)+12(2^2+3^2+4^2)-6\left[(2^2+3^2+4^2).(1+2+3+4)\right]+9(1^2+2^2+3^2+4^2)$$

$$4(29)+(48)+12(353)-6(29.10)+9(30)$$

$$116+48+4236-1740+270=2930$$

$$\sum_{i=2}^{5}\sum_{j=1}^{3}\left[(i+2j)^3-3(4!+j^2)\right]$$

$$\sum_{i=2}^{5}\sum_{j=1}^{3}\left[(i^3+3.(i)^2.(2j)+3.(\mathrm{i}).(2\,\mathrm{j})^2+(2j)^3-3(4.3.2.1)-3j^2\right]$$

$$\sum_{i=2}^{5}\sum_{j=1}^{3}i^3+3\sum_{i=2}^{5}\sum_{j=1}^{3}i^2.2j+3\sum_{i=2}^{5}\sum_{j=1}^{3}i.4j^2+8\sum_{i=2}^{5}\sum_{j=1}^{3}j^3-\sum_{i=2}^{5}\sum_{j=1}^{3}3.24-3\sum_{i=2}^{5}\sum_{j=1}^{3}j^2$$

$$3\sum_{i=2}^{5}i^3+3\left[\left(\sum_{i=2}^{5}i^2\right).\left(2\sum_{j=1}^{3}j\right)\right]+3\left[\left(\sum_{i=2}^{5}i\right).\left(4\sum_{j=1}^{3}j^2\right)\right]+8.4\sum_{j=1}^{3}j^3-3.24.4.3-3.4\sum_{j=1}^{3}j^2$$

$$(3.224)+3(54.12)+3(14.4.14)+(32.36)-(864)-(1214)$$

$$672+1944+2352+1152-864-168=5088$$

$$\sum_{i=2}^{4}\sum_{j=3}^{5}\left[\frac{(j-4)^2}{4i}\right]$$

$$\sum_{i=2}^{4}\sum_{j=3}^{5}\left[\frac{\left(j^2+2.4.j+4^2\right)}{4i}\right]$$

$$\sum_{i=2}^{4}\sum_{j=3}^{5}\frac{j^2}{4i}-\sum_{i=2}^{4}\sum_{j=3}^{5}\frac{8j}{4i}+\sum_{i=2}^{4}\sum_{j=3}^{5}\frac{16}{4i}$$

$$\left(\sum_{j=3}^{5}j^2.\sum_{i=2}^{4}\frac{1}{4i}\right)-\left(\sum_{j=3}^{5}8j.\sum_{i=2}^{4}\frac{1}{4i}\right)+4\sum_{i=2}^{4}\sum_{j=3}^{5}\frac{1}{i}$$

$$\left(\sum_{j=3}^{5}j^2.\frac{1}{4}\sum_{i=2}^{4}\frac{1}{i}\right)-\left(8\sum_{j=3}^{5}j.\frac{1}{4}\sum_{i=2}^{4}\frac{1}{i}\right)+4.3\sum_{i=2}^{4}\frac{1}{i}$$

$$\left[50.\frac{1}{4}\left(\frac{1}{2}+\frac{1}{3}+\frac{1}{4}\right)\right]-\left[8(3+4+5).\frac{1}{4}\left(\frac{1}{2}+\frac{1}{3}+\frac{1}{4}\right)\right]+12\left(\frac{1}{2}+\frac{1}{3}+\frac{1}{4}\right)$$

$$\left[50.\frac{1}{4}\left(\frac{6}{12}+\frac{4}{12}+\frac{3}{12}\right)\right]-\left[8(12).\frac{1}{4}\left(\frac{6}{12}+\frac{4}{12}+\frac{3}{12}\right)\right]+12\left(\frac{6}{12}+\frac{4}{12}+\frac{3}{12}\right)$$

$$\left[50.\frac{1}{4}\left(\frac{13}{12}\right)\right]-\left[96.\frac{1}{4}\left(\frac{13}{12}\right)\right]+12.\frac{13}{12}$$

$$\left(50.\frac{13}{48}\right)-\left(96.\frac{13}{48}\right)+13$$

$$13,54166667-26+13=0,5417$$

$$\sum_{i=1}^{3}\sum_{j=1}^{4}\left[\frac{2j}{i}-3(3\mathrm{i}+\mathrm{j}^2)\right]=-449,3333$$

$$\sum_{i=1}^{3}\sum_{j=3}^{5}\left[\left(2j-i^2\right)+3\left\{\binom{5}{3}-j\right\}^2\right]=1020$$

$$\sum_{i=1}^{4}\sum_{j=1}^{3}\left\{\frac{3!i^2}{\binom{5}{2}j}-1\right\}=21$$

Oplossing oefening 3.1

	f_i	F_i
4|899	3	3
5|2457788	7	10
6|0012223455556899	16	26
7|0022233	7	33
8|01	2	35

Een stam-blad diagram laat toe op een relatief eenvoudige wijze de modus af te lezen. De modus bedraagt in dit geval 6,5. Door het diagram aan te vullen met de cumulatieve frequenties kunnen we eveneens de mediaan afleiden. De mediaan is de waarde van de $\frac{n+1}{2}$-de waarneming, hier de 18de waarneming. Die is hier gelijk aan 6,4.

De spreiding van de waarden in het stam-blad diagram kan gemeten worden met behulp van de variatiebreedte $x_{i\,(max)} - x_{i\,(min)} = 8{,}1 - 4{,}8 = 3{,}3$.

De vorm van de verdeling is unimodaal en bijna symmetrisch. Het aantal waarnemingen onder de modale stam is nauwelijks groter dan het aantal waarnemingen erboven. Als we 2 stammen per eenheid reserveren (één voor de decimalen 0 t.e.m. 4 en één voor de decimalen 5 t.e.m. 9), krijgen we een iets meer gedetailleerd beeld van de vorm van de verdeling.

	fi
4|899	3
5|24	2
5|57788	5
6|00122234	8
6|55556899	8
7|0022233	7
7|	0
8|01	2

Oplossing oefening 3.2

a) Om de kwartielen en de mediaan te berekenen, is het handig de cumulatieve frequenties te kennen:

Uitgaven	f_i	F_i
0	1	1
4	1	2
10	1	3
12	2	5
15	1	6
16	1	7
18	1	8
19	2	10
22	1	11
24	1	12
25	4	16
29	1	17
30	6	23
31	4	27
32	1	28
33	3	31

Uitgaven	f_i	F_i
34	1	32
35	3	35
36	2	37
37	5	42
39	3	45
40	1	46
41	2	48
42	3	51
43	2	53
45	1	54
47	1	55
49	2	57
50	1	58
51	1	59
125	1	60
n	60	

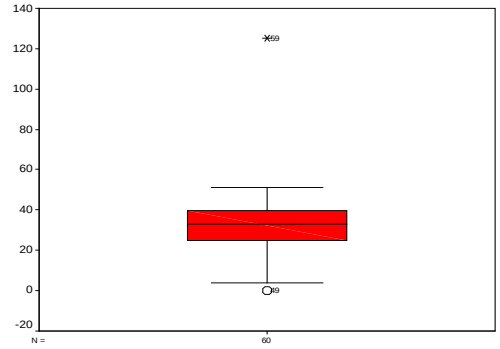
Q_1 is de $\left(\frac{60+1}{4}\right)^{de}$ of de 15,25^{ste} waarneming. Zowel de 15de als de 16de waarneming heeft de waarde 25. Q_1= 25. De doos wordt dus aan de onderzijde begrensd door 25.

Me is de $\left(\frac{60+1}{2}\right)^{de}$ en of de 30,5^{de} waarneming. Zowel de 30ste als 31ste waarneming heeft de waarde 33. Dus Me=33 en in de doos wordt een horizontale streep getrokken ter hoogte van de waarde 33.

Q_3 is de $\left(\frac{3(60+1)}{4}\right)^{de}$ of de 45,75^{ste} waarneming. De 45ste waarneming heeft waarde 39 en de 46ste waarneming heeft waarde 40. De 45,75^{ste} waarneming wordt berekend via lineaire interpolatie: Q_3=39+ 0,75.(40-39)= 39,75. De doos wordt dus aan de bovenzijde begrensd door 39,75.

De interkwartielafstand Q_3 - Q_1 is gelijk aan 14,75. De kleinste waarneming $x_{i\ (min)}$= 0 ligt op een afstand van 25 van Q_1 en is dus meer dan anderhalve interkwartielafstand (22,125) van de onderkant van de doos verwijderd. Het onderste snorhaar wordt daarom niet doorgetrokken tot de allerkleinste waarneming, maar tot de kleinste waarneming waarvan de waarde minder dan anderhalve interkwartielafstand van Q_1 verwijderd is. De waarde van deze waarneming is 4.

De grootste waarneming $x_{i\ (max)}$ = 125 is ook meer dan anderhalve interkwartielafstand van de doos verwijderd. Het bovenste snorhaar reikt tot de hoogste waarneming waarvan de waarde niet meer dan anderhalve interkwartielafstand van Q_3 verwijderd is. De waarde van deze waarneming is 51.



b. Me= 33

$$\text{Middenkwartiel}=\frac{25+39{,}75}{2}=32{,}38$$

$$\text{Trigemiddelde}=\frac{25+2.33+39{,}75}{4}=32{,}69$$

Het trigemiddelde neemt niet alleen de mediaan in rekening, maar ook de kwartielen. In dit geval is het trigemiddelde kleiner dan de mediaan omdat de afstand tussen het eerste kwartiel en de mediaan groter is dan de afstand tussen de mediaan en het derde kwartiel. De verdeling is met andere woorden links asymmetrisch. Het middenkwartiel is enkel en alleen gebaseerd op de kwartielen, en wordt daarom in grotere mate dan het trigemiddelde bepaald door de scheefheid van de verdeling. Bij een links scheve verdeling is het middenkwartiel daarom kleiner dan het trigemiddelde.

c. Variatiebreedte= $x_{i\,(max)} - x_{i\,(min)} = 125$

Interkwartielafstand= $Q_3 - Q_1 = 14{,}75$

Uit de doosdiagram blijkt dat de kleinste en grootste waarnemingen beiden uitschieters zijn. In dit geval is de variatiebreedte geen geschikte spreidingsmaat. De interkwartielafstand is veel resistenter voor uitschieters.

Oplossing oefening 3.3

Q_1: $25{,}25^{ste}$ wrn = 10

Me: $50{,}5^{de}$ wrn = 11

Q_3: $75{,}75^{ste}$ wrn = 13,5

Gemiddelde = 11,32

LET OP: indien je de kwartielen met het rekentoestel TI-83 of TI-84 berekent, dan krijg je als derde kwartiel de waarde 13, i.p.v. 13.5. De reden is dat een TI-83 of TI-84 een andere benadering kiest voor de bepaling van de kwartielen. Bij een even aantal waarnemingen (zoals hier 100) wordt eerst het aantal waarnemingen in twee gelijke blokken verdeeld (50 en 50). Daarna wordt de middelste waarde van elk blok als respectievelijk Q_1 en Q_3 genomen. Voor Q_1 geeft dit dus de $25{,}5^{de}$ waarneming (=10), voor Q_3 is dit de $75{,}5^{de}$ waarneming (= 13).

Oplossing oefening 3.4

Bij de berekening van scharnierwaarden vertrekken we van de rang van Q_1 en de rang van Q_3.

De rang van Q_1= (n+1)/4 = 18/4 = $4{,}5^{de}$ waarneming.

Rang van S_1 = 4,5 + 0.5 = 5^{de} wrn. S_1: 5^{de} wrn = 5

Me: 9^{de} wrn = 8

De rang van Q_3= 3(n+1)/4 = 3(18/4) = 13,5^{de} waarneming.

Rang van S_2 = 13,5 - 0.5 = 13de wrn

S_2: 13de wrn = 12

Trigemiddelde:

$$\frac{S_1 + 2Me + S_2}{4} =$$

$$\frac{5 + 2.8 + 12}{4} = 8,25$$

Oplossing oefening 3.5

n = 23

Q_1: 6de wrn = 56

Q_3: 18de wrn = 75

Middenkwartiel:

$$\frac{Q_1 + Q_3}{2} =$$

$$\frac{56 + 75}{2} = 65,5$$

Oplossing oefening 3.6

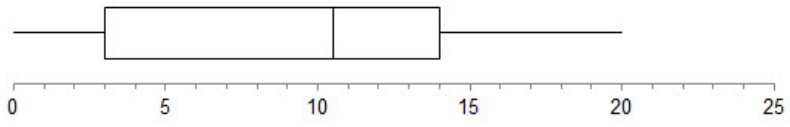
n = 38

S_1: 10de wrn = 3

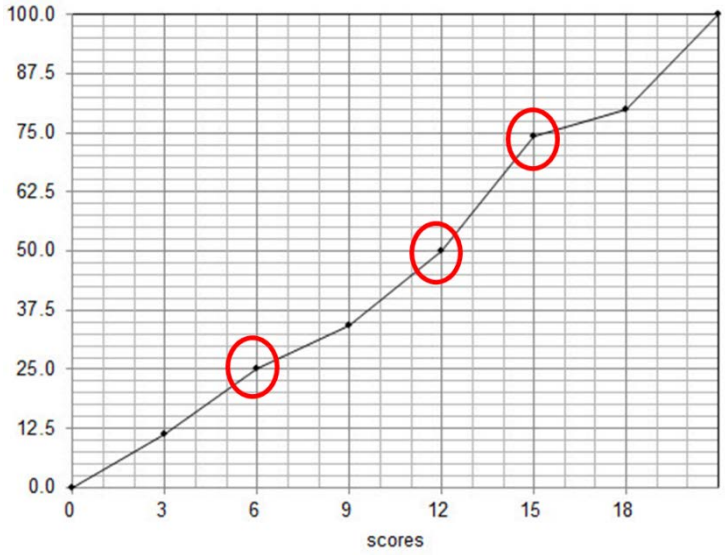
Me: 19,5^{de} wrn = 10,5

S_2: 29de wrn = 14

Boxplot A



Oplossing oefening 3.7



$Q_1 = 6$

Me = 12

$Q_3 = 15$

<u>Trigemiddelde:</u>

$$\frac{Q_1 + 2Me + Q3}{4} =$$
$$\frac{6 + (2 \times 12) + 15}{4} = 11,25$$

Oplossing oefening 3.8

Middenkwartiel:

$$\frac{Q_1 + Q_3}{2} =$$
$$\frac{6 + 15}{2} = 10,5$$

Oplossing oefening 3.9

x_i	F_i	f_i	diepte
1,5	100	100	100
2,5	140	40	115
3,5	170	30	90
4,5	180	10	60
5,5	230	50	50

Oplossing oefening 3.10

Me: 12^{de} wrn → maximale diepte = 12

Diepte van waarde 72 = 9

Oplossing oefening 3.11

$Q_1 = 56$

$Q_3 = 75$

Interkwartielafstand (IKA) = 19

Outlier = 75 + (1,5 x 19) = 103,5

Oplossing oefening 4.1

a) We hebben hier te maken met een kwantitatieve variabele, met andere woorden we kunnen de nominale, de ordinale en de kwantitatieve centrummaten berekenen.

1. op nominaal niveau

De modus (M_o) is de vaakst voorkomende waarneming. Bij deze resultaten levert de modus geen eenduidige waarde op, aangezien alle waarnemingen slechts één maal voorkomen. De modus is in deze situatie geen geschikte centrummaat.

2. op ordinaal niveau

a. Vooraleer we de mediaan kunnen berekenen, dienen we de dataset te ordenen:

9 – 20 – 22 – 36 – 38 – 39 – 43 – 45 – 50 – 57 – 59 – 60 –68 – 73 – 87

Me $= (\frac{(N+1)}{2} ste \quad x_i)$

$= (\frac{(15+1)}{2} ste \quad x_i)$

$= x_8$

= 45

(De helft van de waarnemingen heeft een score van 45 of minder)

b. Middenkwartiel: $\frac{Q_1 + Q_3}{2}$

met $Q_1 = (\frac{(N+1)}{4} ste \quad x_i) = (\frac{(15+1)}{4} ste \quad x_i) = x_4 = 36$

met $Q_3 = (\frac{3(N+1)}{4} ste \quad x_i) = (\frac{3(15+1)}{4} ste \quad x_i) = x_{12} = 60$

$$\Sigma \frac{Q_1 + Q_3}{2} = \frac{36 + 60}{2} = 48$$

(Opmerking: de score van het middenkwartiel is hoger dan de mediaan. Dit betekent dat de afstand tussen het eerste kwartiel en de mediaan in dit geval kleiner dan de afstand tussen de mediaan en het derde kwartiel)

c. Middenscharnier: $\frac{S_1 + S_2}{2}$

met $S_1 = \left[\frac{N+1}{4} + 0{,}5\right]^{ste}$ wn $= \left[\frac{(15+1)}{4} + 0{,}5\right]^{ste}$ wn = 4,5ste wn = 37

(De eerste scharnierwaarde (S_1) stemt overeen met het gemiddelde van de vierde waarneming (36) en de vijfde waarneming (38)).

met $S_2 = \left[\frac{3(N+1)}{4} - 0{,}5\right]^{ste}$ wn $= \left[\frac{3(15+1)}{4} - 0{,}5\right]^{ste}$ wn = 11,5ste wn = 59,5

(De tweede scharnierwaarde (S_2) stemt overeen met het gemiddelde van de elfde waarneming (59) en de twaalfde waarneming (60)).

$$\Sigma \frac{S_1 + S_2}{2} = \frac{37 + 59{,}5}{2} = 48{,}25$$

d. Trigemiddelde van Tukey: $\left[S_1 + (2 \times Me) + S_2\right] / 4$

$= \left[37 + (2 \times 45) + 59{,}5\right] / 4$

= 46,62

3. op kwantitatief niveau

$$\mu = \frac{\sum_{i=n}^{N} x_i}{N}$$

$$= \frac{(9 + 20 + 22 + 36 + 38 + 39 + 43 + 45 + 50 + 57 + 59 + 60 + 68 + 73 + 87)}{15}$$

= 706/15

= 47,07

b) Wat na toevoeging van een score van 186?

Me $= \left(\frac{(N+1)}{2} ste \quad x_i\right) = \left(\frac{(16+1)}{2} ste \quad x_i\right) = x_{8,5}$

De mediaan is het gemiddelde van de 8ste en de 9de waarneming, respectievelijk gelijk aan 45 en 50. De mediaan is dus gelijk aan 47,5.

μ = 55,8

De mediaan is een zeer resistente centrummaat die betekent dat dit kengetal weinig gevoelig is voor extreme waarden. Het gemiddelde daarentegen, houdt rekening met alle waarden waardoor het meer beïnvloed wordt door de toevoeging van een uitschieter en dit geval dus sterk toeneemt.

Oplossing oefening 4.2

a) We gebruiken een harmonisch gemiddelde omdat de meeteenheid van de variabele een samegenstelde grootheid is (euro/liter) en omdat er een vast ijkingspunt is (namelijk 1000 Euro).

$$HG = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i} f_i}$$ met N = 2, x_1 =0,50, f_1 = 1, x_2 =0,60 en f_2 = 1

dan is: HG = $\frac{2}{\frac{1}{0,50} + \frac{1}{0,60}}$ = 0,5455

De gemiddelde prijs bedraagt 54,55 eurocent per liter stookolie..

Oplossing oefening 4.3

b) Om een gemiddeld groeipercentage te berekenen, maken we gebruik van het meetkundig gemiddelde. De groeifactoren per decennium zijn respectievelijk gelijk aan: 1,5; 2 en 4.

Dan is de rekenkundige variant van het meetkundig gemiddelde:

MG = $\sqrt[3]{1,5 \times 2 \times 4}$ = 2,29

Per decennium nam de bevolking dus gemiddeld met 129 procent toe.

Oplossing oefening 4.4

Harmonisch gemiddelde!

$$\frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{7}{\frac{1}{3,8} + \frac{1}{4,2} + \frac{1}{4,2} + \frac{1}{5,3} + \frac{1}{5,1} + \frac{1}{5,8} + \frac{1}{4,9}} = 4,6648$$

Dit wijkt bijna niet af van Bobs gemiddelde tijdens de eerste 20 km van de dodentocht. Het verschil bedraag 0,0648 km/u.

Oplossing oefening 4.5

Meetkundig gemiddelde!

$$\sqrt[10]{\frac{4000}{2000}} = 1 \times 0718$$

Jaarlijks moet de gemiddelde groeivoet 7.18% bedragen.

Oplossing oefening 4.6

Meetkundig gemiddelde!

Gemiddelde 20-jaarlijkse bevolkingsgroei

$$\sqrt[5]{\prod x_i}$$

$$= \sqrt[5]{\frac{3,7}{2,5} \times \frac{5,3}{3,7} \times \frac{6,8}{5,3} \times \frac{8,1}{6,8} \times \frac{8,9}{8,1}} = \sqrt[5]{\frac{8,9}{2,5}} = 1,2891$$

De gemiddelde 20-jaarlijkse groeipercentage bedraagt 28,91%.

Jaarlijkse groeipercentage tussen 2010 en 2030

$$\sqrt[20]{\frac{8,1}{6,8}} = 1,0088$$

Het jaarlijkse groeipercentage bedraag 0,88% tussen 2010 en 2030.

Oplossing oefening 4.7

Gewogen gemiddelde!

$\mathbf{x_i}$	$\mathbf{f_i}$
1	2
2	14
5	8
7	32

$$\frac{\sum_{i=1}^{k} (f_i . x_i)}{\sum_{i=1}^{k} f_i} = 5,25$$

Oplossing oefening 4.8

$$1000 \times 1,025^5 = 1131,4082$$

Oplossing oefening 4.9

Gewogen gemiddelde!

$$\frac{\sum_{i=1}^{k}\left(f_i.x_i\right)}{\sum_{i=1}^{k} f_i} = 2,4$$

$$\frac{0,3 \times 1 + 0,7.x}{1} = 2,4$$

$$0,7x = 2,4 - 0,3$$

$$x = \frac{2,1}{0,7} = 3$$

Oplossing oefening 4.10

$$Q_3 - IKA = Q_1$$

$$7 - 5 = 2$$

Outliers aan de onderkant:

$$Q_1 - (1,5 \times IKA)$$

$$= 2 - (1,5 \times 5)$$

$$= -5,5$$

Oplossing oefening 4.11

Harmonisch gemiddelde!

$$\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{2}{\frac{1}{25,3} + \frac{1}{x}} = 28$$

X = 31,3452

Uitwerking

$$\frac{2}{\frac{1}{25,3}+\frac{1}{x}}=28$$

$$\frac{\frac{1}{25,3}+\frac{1}{x}}{2}=\frac{1}{28}$$

$$\frac{1}{25,3}+\frac{1}{x}=\frac{2}{28}=\frac{1}{14}$$

$$\frac{1}{x}=\frac{1}{14}-\frac{1}{25,3}=0,0319028$$

$$x=\frac{1}{0,0319028}=31,34521108\approx 31,3452$$

Oplossing oefening 4.12

Gemiddelde aangroei 2006 – 2015

$$\sqrt[9]{\prod x_i}=1,1840$$

Gemiddeld aantal tussen 2006 en eind 2015

Rekenkundig gemiddelde!

= 3702,6916

Oplossing oefening 4.13

Meetkundig gemiddelde!

Procentuele groei tussen 4e en 8e verjaardag:

$$\sqrt[4]{\frac{6000}{3000}}=1,1892$$

Procentuele groei tussen 8e en 12e verjaardag:

$$\sqrt[4]{\frac{17000}{6000}}=1,2974$$

De woordenschat van een kind groeit gemiddeld 10,82% meer van 8e tot de 12e verjaardag dan tussen de 4e en de 8e verjaardag.

Oplossing oefening 5.1

De volgende beweringen zijn correct: A, C, D, F, G en I.

Oplossing oefening 5.2

We kunnen de vraag beantwoorden door de vastgestelde spreiding bij beide onderzoekers met elkaar te vergelijken. We kunnen echter niet zomaar de standaardafwijkingen met elkaar vergelijken aangezien de onderzoekers een verschillende meeteenheid gebruikt hebben. Kortom, we zullen gebruik moeten maken van een relatieve spreidingsmaat, in dit geval de variatiecoëfficiënt.

VC = $\frac{s}{\bar{x}}$ met $\bar{x} = \frac{\sum_{i=1}^{k}(f_i \cdot x_i)}{\sum_{i=1}^{k} f_i}$ en $s = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n-1}}$

Eerste onderzoeker: VC = $\frac{s}{\bar{x}}$ = 0,45

Tweede onderzoeker: VC = $\frac{s}{\bar{x}}$ = 0,56

Dus: wanneer we de twee variatiecoëfficiënten vergelijken, stellen we vast dat de inkomensongelijkheid bij de tweede onderzoeker het grootst is.

Oplossing oefening 5.3

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1}} = 1,7233$$

Oplossing oefening 5.4

$$VC_x = \frac{s_X}{\bar{X}} = \frac{1,7233}{2,4160} = 0,7133 \quad VC_y = \frac{s_y}{\bar{Y}} = \frac{1,3842}{2,6765} = 0,5172$$

De spreiding is het grootst bij het aantal uren sport (X).

Oplossing oefening 5.5

$$VC_x = \frac{\sigma_X}{\bar{X}} = \frac{14,5497}{67,0435} = 0,2170$$

Oplossing oefening 5.6

$$oDi = 1 - \frac{\sum_{i=1}^{k-1}\left(F_i^* - \frac{1}{2}\right)^2}{\frac{k-1}{4}} = 1 - \frac{0,2794}{1} = 0,7206$$

Basisinkomen?	fi	Fi*
1 Helemaal oneens	10	0,0833
2 Oneens	25	0,2917
3 Noch eens, noch oneens	12	0,3917
4 Eens	40	0,725
5 Helemaal eens	33	1

Oplossing oefening 5.7

$$s_x^{\,2} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n-1} = 22,2353$$

Oplossing oefening 5.8

Wijn	F_i	f_i^*
Châteaux Cheval Blanc (2010)	15	0,3409
Châteaux Ausone (2010)	23	0,1818
Châteaux Pavie (2006)	29	0,1364
Châteaux Figeac (2010)	39	0,2273
Châteaux Troplong Mondot (2006)	44	0,1136

$$nDi = 1 - \frac{\sum_{k=1}^{k-1}(f_i * - \frac{1}{k})^2}{\frac{(k-1)}{k}} = 1 - \frac{0,0324}{4/5} = 0,9595$$

Oplossing oefening 5.9

$$Genormeerde\ entropie = \frac{-\sum_{k=1}^{k}\left[f_i^* \times \ln f_i^*\right]}{\ln k} = \frac{1,5005}{\ln 5} = 0,9323$$

Oplossing oefening 5.10

Dit is een lineaire transformatie van de vorm y = a + b*x

$$Y = 22 + \frac{X}{2} = 22 + (\frac{1}{2} \times X)$$

In het handboek staat dat wanneer we een constante a optellen bij alle waarnemingen, dit geen enkel effect heeft op de variantie. Enkel de multiplicator b moet in rekening worden gebracht.

$$VAR_{(a+bx)} = b^2 \times VAR(X)$$

$$VAR_{(a+bx)} = \left(\frac{1}{2}\right)^2 \times 4^2 = 4$$

Oplossing oefening 5.11

$$oDi = 1 - \frac{\sum_{i=1}^{k-1}\left(F_i^* - \frac{1}{2}\right)^2}{\frac{k-1}{4}} = 1 - \frac{0,3603}{1} = 0,6397$$

Oplossing oefening 5.12

$$oDi = 1 - \frac{\sum_{i=1}^{k-1}\left(F_i^* - \frac{1}{2}\right)^2}{\frac{k-1}{4}} = 1 - \frac{0,4053}{1} = 0,5947$$

Oplossing oefening 5.13

$$nDi = 1 - \frac{\sum_{i=1}^{k}(f_i* - \frac{1}{k})^2}{\frac{(k-1)}{k}} = 1 - \frac{0,0681}{9/10} = 0,9244$$

$$Genormeerde\ entropie = \frac{-\sum_{i=1}^{k}\left[f_i^* \times \ln f_i^*\right]}{\ln k} = \frac{2,0218}{\ln 10} = 0,8781$$

Oplossing oefening 5.14

Aantal glazen Adorp (X)	$\mathbf{F_i^*}$	$\mathbf{f_i^*}$	$\mathbf{f_i}$
10	0,3	0,3	180
20	0,5	0,2	120
30	0,6	0,1	60
40	0,9	0,3	180
50	1	0,1	60

$$VC_x = \frac{s_X}{\overline{X}} = \frac{14,1893}{27} = 0,5255$$

Aantal glazen Bdorp (Y)	$\mathbf{F_i^*}$	$\mathbf{f_i^*}$	$\mathbf{f_i}$
10	0,1	0,1	60
20	0,2	0,1	60
30	0,4	0,2	120
40	0,7	0,3	180
50	1	0,3	180

$$VC_y = \frac{s_y}{\overline{Y}} = \frac{12,8169}{36} = 0,3560$$

Oplossing oefening 5.15

Het verband tussen X en Y wordt voorgesteld door een rechte en heeft dus de vorm y = a + b * x. A is het intercept of de kruising met de x-as en is dus gelijk aan 16. B is de richtingscoëfficiënt en is gelijk aan 0,25 (telkens je op de x-as 4 waarden omhoog gaat, ga je er 1 omhoog op de y-as).

Net als in een van de voorgaande oefeningen houden we enkel rekening met de richtingscoëfficiënt.

$$VAR_{(a+bx)} = b^2 \times VAR(X)$$

$$VAR_{(a+bx)} = 0,25^2 \times 63 = 3,9375$$

Oplossing oefening 5.16

$$nDi = 1 - \frac{\sum_{i=1}^{k}(f_i * - \frac{1}{k})^2}{\frac{(k-1)}{k}} = 1 - \frac{0,0103}{4/5} = 0,9871$$

$$Genormeerde\ entropie = \frac{-\sum_{i=1}^{k}\left[f_i^* \times \ln f_i^* \right]}{\ln k} = \frac{1,5807}{\ln 5} = 0,9821$$

Oplossing oefening 6.1

Bruto jaarwedde	f_i	F_i	$x_i.f_i$	$f_i.(x_i - \mu)^2$	$f_i.(x_i - \mu)^3$
14.000	8	8	112.000	738.240.313	-7.091.721.001.953
17.500	18	26	315.000	671.153.203	-4.098.229.246.582
20.000	58	84	1.160.000	754.292.266	-2.720.166.482.910
24.000	35	119	840.000	5.426.367	2.136.632.080
27.000	18	137	486.000	207.315.703	703.577.667.480
31.000	13	150	403.000	710.678.008	5.254.575.520.264
36.000	7	157	252.000	1.075.235.273	13.326.197.170.166
55.000	2	159	110.000	1.971.135.078	61.881.321.858.887
99.000	1	160	99.000	5.684.217.539	428.554.476.085.693
Totaal	160		3.777.000	11.817.693.750	495.812.168.203.125
Totaal / N			23.606	73.860.586	3.098.826.051.270

Een zeer eenvoudige manier om de scheefheid van de verdeling van een kwantitatieve variabele te bepalen is de scheefheidscoëfficient van Pearson: $\left[3 \times \left(\mu - Me\right)\right] / \sigma$. Die is in dit geval gelijk aan 1,26 ($\left[3 \times \left(23606 - 20000\right)\right] / 8594$). Gezien de uitkomst veel groter is dan nul, kunnen we inderdaad besluiten dat deze loonverdeling rechtsscheef is.

Als de scheefheidscoëfficient van Pearson groter is dan nul geeft dit aan dat het rekenkundig gemiddelde groter is dan de modus of mediaan, wat wijst op de aanwezigheid van enkele zeer hoge lonen. Het gemiddelde is immers veel gevoeliger voor uitschieters dan de modus of mediaan, en wordt daarom sterker beïnvloed door de zeer hoge lonen.

Iemand die wil aantonen dat een onderneming vrij hoge lonen heeft, zal dus beroep doen op het rekenkundig gemiddelde. In het andere geval is bij een rechtsscheve verdeling de mediaan een meer geschikte centrummaat gezien deze geen rekening houdt met de extreem hoge lonen van enkelingen.

De scheefheidscoëfficient van Pearson vergelijkt de modus met het rekenkundig gemiddelde; de scheefheidscoëfficiënt γ_1 houdt rekening met de afwijking van elke x_i tot het rekenkundig gemiddelde.

$$\gamma_1 = \frac{(\sum f_i (x_i - \mu)^3) / N}{\left[\sum f_i (x_i - \mu)^2 / N\right]^{3/2}} = \frac{3.098.826.051.270}{(73.860.586)^{3/2}} = 4,88$$

Ook γ_1 geeft aan dat we te maken hebben met een rechtsscheve verdeling. Gezien γ_1 veel groter is dan 1 gaat het om een sterk rechts asymmetrische verdeling.

Oplossing oefening 6.2

Gegeven is dat n = 50, $\bar{x} = 25$, $\frac{s}{x} = 0{,}50$, $m_3 = 4$ en $m_4 = 68$

Daaruit volgt dat $s = 25 \times 0{,}5 = 12{,}5$.

Dan is $\gamma_1 = \frac{4}{12{,}5^3} = 0{,}002048$ en $\gamma_2 = \frac{68}{12{,}5^4} - 3 = -2{,}9972$

De frequentieverdeling is verdeling quasi symmetrisch (γ_1 is bijna nul) maar is tegelijk veel 'platter' (minder sterk gepiekt) dan de normaalverdeling (γ_2 is duidelijk kleiner dan 0).

Oplossing oefening 6.3

$$gamma1 = \frac{\mu_3}{\sigma_X^3} = \frac{-1947{,}1748}{18{,}0884^3} = -0{,}3290$$

$$kappa = \frac{\mu_4}{\sigma_X^4} = \frac{212087{,}9048}{18{,}0884^4} = 1{,}9811$$

$$gamma2 = kappa - 3 = -1{,}0189$$

Oplossing oefening 6.4

Vermits we enkel beschikken over rangorde kengetallen steunen we de berekening van scheefheid op de formule van Yule.

$Q_1 = 16$

Me = 25

$Q_3 = 56$

$$\frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)} = 0{,}55$$

Oplossing oefening 6.5

In de formule van Yule worden enkel verhoudingen gebruikt en doet de meeteenheid er niet toe. We mogen dus zelf waarden toekennen, zolang de verhoudingen maar gerespecteerd blijven.

Het correcte antwoord is alternatief a).

We gebruikten hierbij $Q_1 = 0$, Me = 7,5 en $Q_3 = 11$

$$\frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)} = -0,364$$

Oplossing oefening 6.6

Indien de afstand tussen Q3 – Me anderhalve keer de afstand van Me – Q1 bedraagt, dan is de afstand Q3 – Me = 36, en Me – Q1 = 24.

Vermits

$$\frac{36}{24} = 1,5$$
$$en \quad 36 + 24 = 60$$

Stellen we $Q_1 = 0$, dan is Me = 24 en $Q_3 = 60$.

De Yule formule levert ons dan: 0,20 als antwoord (alternatief b).

Oplossing oefening 6.7

Alternatief b)

Oplossing oefening 6.8

Berekening van de standaardafwijking op basis van de variatiecoëfficiënt:

$$VC = \frac{S}{\overline{X}} = 0,35 \Rightarrow S = 0,35 \times 50 = 17,5$$

Berekening van de mediaan op basis van Pearson 2 =

$$\frac{3 \times (50 - Me)}{17,5} = -0,25 \Rightarrow 150 - 3Me = -0,25 \times 17,5$$

$$-3Me = -4,375 - 150 \Rightarrow Me = \frac{154,375}{3}$$

$$Me = 51,4583$$

Berekening van de modus op basis van de relatie van Pearson:

$$\overline{X} - Mo = 3(\overline{X} - Me)$$
$$-Mo = 3\overline{X} - 3Me - \overline{X}$$
$$-Mo = 2\overline{X} - 3Me$$
$$-Mo = 100 - 3(51,4583) = -54,3749$$
$$Mo = 54,3749$$

Oplossing oefening 6.9

De 4 momenten zijn

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X}) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2 = 4 \quad \textit{Variantie}$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^3 = 10 \quad \textit{De teller van de Gamme}\,1\ \textit{formule}$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^4 = 46 \quad \textit{De teller van de Kappa formule}$$

Voor de berekening van de scheefheid:

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^3}{(SD)^3} = \frac{10}{(2)^3} = 1,25$$

Voor de berekening van kurtosis:

$$kappa = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^4}{(SD)^4} = \frac{46}{(2)^4} = 2,875$$

$$Gamma2 = 2,875 - 3 = -0,125$$

Oplossing oefening 6.10

Alternatief b)

$Q_3 = 5$

$Q_1 = 0$

Me = 3

Anderhalve IKA = 7,5

Te meten vanaf 0 = -7,5

Oplossing oefening 6.11

$$\mu_2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2 = 9$$

$$\mu_1' = \frac{1}{n}\sum_{i=1}^{n}(x_i - 0)^1 = \frac{\sum_{i=1}^{n} x_i}{n} = 0,5$$

$$VC = \frac{\sigma_X}{\bar{X}} = \frac{3}{0,5} = 6$$

Oplossing oefening 6.12

$Q_1 = 56$

Me = 69

$Q_3 = 75$

Scheefheidscoëfficiënt van Yule:

$$\frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)} =$$

$$\frac{(75 - 69) - (69 - 56)}{(75 - 56)} = -0,3684$$

Oplossing oefening 6.13

Aangezien Gamma 1 gelijk is aan 0, weten we dat de verdeling perfect symmetrisch is. Het middenkwartiel ligt met andere woorden perfect in het midden van de doos van het doosdiagram. Bij een perfect symmetrische verdeling ligt ook de modus perfect in het midden, ook de mediaan en ook het gemiddelde. We leiden dus af dat het gemiddelde ook gelijk is aan 6.

$$VC = \frac{\sigma_X}{\bar{X}}$$

$$0,5 = \frac{\sigma_X}{6}$$

$$\sigma_X = 3$$

$$VAR_x = 3^2 = 9$$

Oplossing oefening 7.1

De frequentiepolygoon leent zich goed tot het vergelijken van twee verdelingen omdat meerdere polygonen in één grafische voorstelling kunnen worden opgenomen. Daartoe moeten we beide verdelingen wel op een uniforme manier classificeren. We bepalen het aantal klassen voor beide verdelingen op basis van de verdeling met het grootst aantal elementen (jongens: N = 48). Voor de variatiebreedte nemen we beide groepen samen, zodat de eventuele uitschieters van zowel de jongens als de meisjes in onze classificatie ondergebracht kunnen worden. De laagste waarde is afkomstig van de jongens ($x_{i(min)} = 54$) en de hoogste waarde is gesitueerd bij de meisjes ($x_{i(max)} = 97$).

Opstellen geclassificeerde frequentietabel

Aantal klassen: $k = 1 + 3{,}32 \log 48 = 6{,}6$. We zullen de gegevens indelen in 7 klassen.

Variatiebreedte: $VB(x_i) = 97 - 54 = 43$

Klassebreedte: $v = 43 / 7 = 6{,}1$. Afgerond naar boven: $v = 7$ (meeteenheid data = 1)

Variatiebreedte klassenindeling: $VB(\text{klassenindeling}) = 7 * 7 = 49$

$VB(\text{klassenindeling}) - VB(x_i) = 6$ is even dus we werken met afgeronde klassegrenzen.

Afgeronde ondergrens eerste klasse: $l_1^a = 54 - (6 / 2) = 51$

Afgeronde bovengrens eerste klasse: $u_1^a = 51 + (7 - 1) = 57$

Exacte ondergrens eerste klasse: $l_1^e = 51 - 0{,}5 = 50{,}5$

Exacte bovengrens eerste klasse: $u_1^e = 57 + 0{,}5 = 57{,}5$

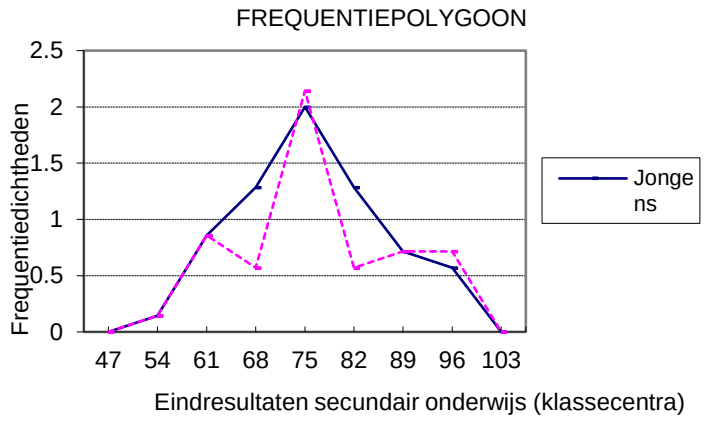
Klasse	Afgeronde ondergrens	Afgeronde bovengrens	Exacte ondergrens	Exacte bovengrens
2	$l_2^a = 57 + 1 = 58$	$u_2^a = 58 + (7 - 1) = 64$	$l_2^e = 57{,}5$	$u_2^e = 57{,}5 + 7 = 64{,}5$
3	$l_3^a = 64 + 1 = 65$	$u_3^a = 65 + (7 - 1) = 71$	$l_3^e = 64{,}5$	$u_3^e = 64{,}5 + 7 = 71{,}5$
4	$l_4^a = 71 + 1 = 72$	$u_4^a = 72 + (7 - 1) = 78$	$l_4^e = 71{,}5$	$u_4^e = 71{,}5 + 7 = 78{,}5$
5	$l_5^a = 78 + 1 = 79$	$u_5^a = 79 + (7 - 1) = 85$	$l_5^e = 78{,}5$	$u_5^e = 78{,}5 + 7 = 85{,}5$
6	$l_6^a = 85 + 1 = 86$	$u_6^a = 86 + (7 - 1) = 92$	$l_6^e = 85{,}5$	$u_6^e = 85{,}5 + 7 = 92{,}5$
7	$l_7^a = 92 + 1 = 93$	$u_7^a = 93 + (7 - 1) = 99$	$l_7^e = 92{,}5$	$u_7^e = 92{,}5 + 7 = 99{,}5$

11-12.

i	l_i^e - u_i^e	l_i^a - u_i^a	x_i^c	Jongens	f_i	$\frac{f_i}{v}$	Meisjes	f_i	$\frac{f_i}{v}$
1	50,5-57,5	51-57	54	I	1	0,14	I	1	0,14
2	57,5-64,5	58-64	61	IIIII I	6	0,86	IIIII	5	0,86
3	64,5-71,5	65-71	68	IIIII IIII	9	1,29	IIIII	5	0,57
4	71,5-78,5	72-78	75	IIIII IIIII IIII	14	2,00	IIIII IIIII IIIII	15	2,14
5	78,5-85,5	79-85	82	IIIII IIII	9	1,29	IIII	4	0,57
6	85,5-92,5	86-92	89	IIIII	5	0,71	IIIII	5	0,71
7	92,5-99,5	93-99	96	IIII	4	0,57	IIIII	5	0,71
				N=48			N=40		

Grafische voorstelling beide verdelingen

De verdeling van de eindresultaten lijkt niet erg te verschillen naargelang het geslacht. Beide verdelingen hebben één modale klasse, telkens de klasse met centrum 75 procent, en zijn tamelijk symmetrisch. De resultaten van de meisjes zijn in vergelijking met de jongens evenwel iets meer in de top en de staarten (vooral rechterstaart) van de verdeling gesitueerd.



Oplossing oefening 7.2

Exacte klassegrenzen	Absolute frequenties	Cumulatieve frequenties
53,75-57,15	2	2
57,15-60,55	3	5
60,55-63,95	2	7
63,95-67,35	6	13
67,35-70,75	15	28
70,75-74,15	14	42
74,15-77,55	9	51

Scheefheidscoëfficiënt van Yule = $\frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$

Berekening van Q_1, Me en Q_3:

v = 57,15 – 53,75 = 3,4

$Q_1 = (\frac{(51+1)}{4} ste \quad x_i) = x_{13}$

Q_1 is dus gesitueerd in de vierde klasse, meer precies $Q_1 = 63,95 + \frac{13 - 7}{6} \times 3,4 = 67,35$. In een kwart van de staten van de V.S. bezit hoogstens 67,4 procent van de inwoners een woning.

$Me = (\frac{(51+1)}{2} ste \quad x_i) = x_{26}$

Me is dus gesitueerd in de vijfde klasse, meer precies $Me = 67,35 + \frac{26 - 13}{15}.3,4 = 70,30$.

$Q_3 = (\frac{3.(51+1)}{4} ste \quad x_i) = x_{39}$

Q_3 is dus gesitueerd in de zesde klasse, meer precies $Q_3 = 70,75 + \frac{39 - 28}{14} \times 3,4 = 73,42$. In driekwart van de staten van de VS bezit hoogstens 73,9 procent van de inwoners een woning.

Scheefheidscoëfficiënt van Yule:

$$\frac{(73,4 - 70,3) - (70,3 - 67,4)}{(73,4 - 70,3) + (70,3 - 67,4)} = 0.03$$

De verdeling is quasi symmetrisch.

Oplossing oefening 7.3

Allereerst wordt de cumulatieve frequentie (F_i) berekend.

Lengte (in inch)	aantal	F_i
57,5 – 58,5	10	10
58,5 – 59,5	18	28
59,5 – 60,5	30	58
60,5 – 61,5	42	100
61,5 – 62,5	35	135
62,5 – 63,5	28	163
63,5 – 64,5	16	179
64,5 – 65,5	8	187

De rang van de mediaan = (187+1)/2 = 94ste waarneming. Dat is de 4de klasse (60,5 – 61,5).

De geïnterpoleerde formule:

$$Me = l_4^e + \frac{rangMe - F_3}{f_4} \times v$$

$$Me = 60,5 + \frac{94 - 58}{42} \times 1 = 61,3571$$

De rang van Q_1 = (187+1)/4 = 47ste waarneming. Dat is de 3de klasse (59,5 – 60,5).

$$Q_1 = l_3^e + \frac{rangQ_1 - F_2}{f_3} \times v$$

$$Q_1 = 59,5 + \frac{47 - 28}{30} \times 1 = 60,1333$$

De rang van Q_3 = 3(187+1)/4 = 141ste waarneming. Dat is klasse 6 (62,5 – 63,5).

$$Q_3 = l_6^e + \frac{rangQ_3 - F_5}{f_6} \times v$$

$$Q_3 = 62,5 + \frac{141 - 135}{28} \times 1 = 62,7143$$

De geïnterpoleerde scheefheid van Yule:

$$\frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)} =$$

$$\frac{(62,7143 - 61,3571) - (61,3571 - 60,1333)}{(62,7143 - 60,1333)} = 0,0517$$

Oplossing oefening 7.4

Allereerst wordt de klassecentra berekend. De klassecentra mag je gebruiken om gamma 1 te berekenen.

Lengte (in inch)	aantal	x_i^c
57,5 – 58,5	10	58
58,5 – 59,5	18	59
59,5 – 60,5	30	60
60,5 – 61,5	42	61
61,5 – 62,5	35	62
62,5 – 63,5	28	63
63,5 – 64,5	16	64
64,5 – 65,5	8	65

Met behulp van ons grafisch rekentoestel vullen we de lijsten in.

L1	L2
58	10
59	18
60	30
61	42
62	35
63	28
64	16
65	8
------	------

Voor het verdere verloop van de bewerking in het rekentoestel verwijzen we naar paragraaf 6.4. Het uiteindelijke resultaat voor Gamma 1 = 0,0469

Oplossing oefening 8.1

$(A \cap B) = \{2,4,6\}$

$(A \cap C) = \{3,6\}$

$(A \cap B \cap C) = \phi$

$(A \cup B) = \{1,2,3,4,5,6,8,10,12,14\}$

$(B \cup C) = \{2,3,4,6,8,9,10,12,14,15\}$

$(A \cup B \cup C) = \{1,2,3,4,5,6,8,9,10,12,14,15\}$

Oplossing oefening 8.2

a) Fout
b) Juist
c) Juist
d) Fout

Oplossing oefening 8.3

$B \subset C$

$B \bigcap C = B$

$B \bigcup C = C$

$B \bigcap D = \phi$

$C \bigcup A = A$

$D \bigcap E = D$

$E \setminus D = \{stijn\}$

$D \setminus E = \phi$

$Bruno \notin C \bigcap B$

$D \bigcap E \subset A$

Oplossing oefening 8.4

a) A∪B : (nog aan te vullen)
b) A∪C :
c) B∪C :
d) A∪B∪C :
e) A∩B :
f) A∩C :
g) B∩C :
h) A∩B∩C :
i) A\B :

Oplossing oefening 8.5

Totaal aantal mogelijke toevalsgebeurens: 1309816

Aantal gunstige toevalsgebeurens: aantal studenten die nog in het secundair, het lager onderwijs of het kleuteronderwijs zitten = 259399 + 410208 + 444307.

$\frac{1113914}{1309816}$ = 0,8504

Oplossing oefening 8.6

2 taalvakken * 3 wiskundige vakken * 5 keuzevakken * 1 seminarie = 30

Oplossing oefening 8.7

	Man	Vrouw	
Verpl.	1/13	7/13	8/13
Arts	2/13	3/13	5/13
	3/13	10/13	1

$(V \cup Man) = P(V) + P(Man) - P(V \cap Man) = \frac{8}{13} + \frac{3}{13} - \frac{1}{13} = \frac{10}{13} = 0.7692$

Oplossing oefening 8.8

$P(Pl \cup SH) = P(Pl) + P(SH) - P(Pl \cap SH) = \frac{192}{566} + \frac{39}{566} - \frac{11}{566} = \frac{220}{566} = 0{,}3887$

Oplossing oefening 8.9

$$P(SpH|Pl) = \frac{P(SpH \cap Pl)}{P(Pl)} = \frac{34/566}{192/566} = \frac{34}{192} = 0{,}1771$$

Oplossing oefening 8.10

Je kan hier de productregel bij statistische onafhankelijkheid gebruiken (trekking 1 is onafhankelijk van trekking 2). De trekking gebeurt wel zonder teruglegging, waardoor de kans op een inbraak uit 2014 de tweede keer lager ligt dan de eerste keer.

$$P(2014(1) \cap 2014(2)) = P(2014(1)) \text{ x } P\big(2014(2)\big) = \frac{16}{53} \times \frac{15}{52} = 0{,}0871$$

Oplossing oefening 8.11

Gegeven:

$P(Man) = 0{,}64 \quad P(G|Man) = 0{,}3$

$P(D) = 0{,}12 \qquad\qquad P(D|Man) = 0{,}15$

$P(G) = 0{,}24$

Gevraagd: $P(D|Vrouw) = ?$

Het antwoord vind je door gebruik te maken van de regel van de totale kans:

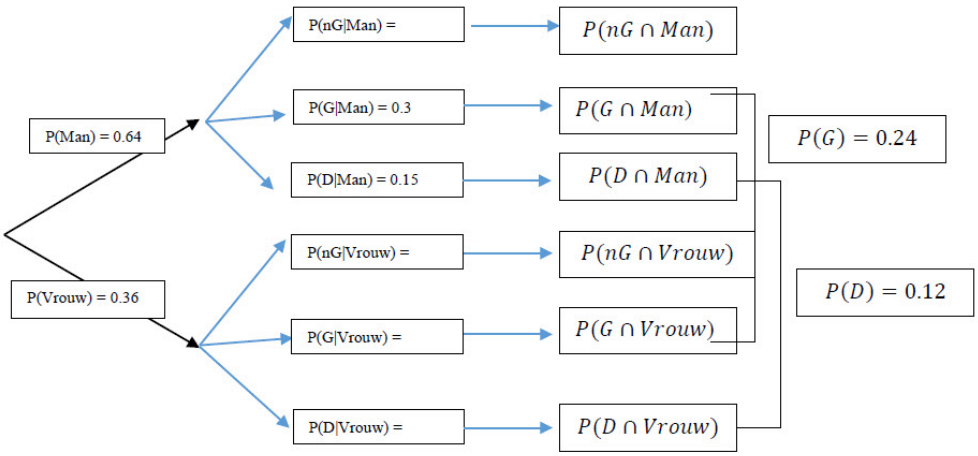
$$P(D) = P(D \cap Man) + P(D \cap Vrouw)$$

$$P(D) = P(D|Man)\ x\ P(Man) + P(D|Vrouw)\ x\ P(Vrouw)$$

$$0{,}12 = 0{,}15 \text{ x } 0{,}64 + P(D|Vrouw) \text{x } 0{,}36$$

$$\frac{0{,}024}{0{,}36} = P(D|Vrouw) = 0{,}0667$$

Aan de hand van de gegeven kansen, kan je ook het hele dendrogram invullen en zo tot de juiste uitkomst komen. In onderstaand dendrogram staan voorlopig enkel de gegeven kansen. Probeer de rest van het dendrogram zelf aan te vullen.



Oplossing oefening 8.12

Gegeven:

$P(St.Vr|Man) = 0{,}3$

$P(St.Vr|Vrouw) = 0{,}8$

$P(Vrouw) = 0{,}35$

$P(Man) = 0{,}65$

Gevraagd: $P(St.Vr)$

Antwoord: maak gebruik van de regel van de totale kans:

$P(St.Vr) = P(St.Vr|Man)\ \text{x}\ P(Man) + P(St.Vr|Vrouw)\ \text{x}\ P(Vrouw)$

$P(St.Vr) = 0{,}3\ \text{x}\ 065 + 0{,}8\ \text{x}\ 0{,}35 = 0{,}475$

Oplossing oefening 8.13

Gegeven:

$P(T+|B+) = 0{,}85$

$P(T-|B-) = 0{,}88$

$P(B+) = 0{,}004$

Gevraagd:

$P(B+|T+) = ?$

Antwoord:

Maak gebruik van de regel van Bayes.

$$P(B+|T+) = \frac{P(T+|B+)\ \text{x}\ P(B+)}{P(T+|B+)\ \text{x}\ P(B+) + P(T+|B-)\ \text{x}\ P(B-)}$$

$$P(B+|T+) = \frac{0{,}85\ \text{x}\ 0{,}004}{0{,}85\ \text{x}\ 0{,}004 + 0{,}12\ \text{x}\ 0{,}996} = \frac{0{,}0034}{0{,}12292} = 0{,}02766$$

Oplossing oefening 8.14

Gegeven:

$P(A|I) = 0{,}85$

$P(A|gnI) = 0{,}01$

$P(I) = 0{,}0001$

Gevraagd:

$P(I|A) = ?$

Antwoord:

Maak wederom gebruik van de regel van Bayes.

$$P(I|A) = \frac{P(A|I) \text{ x } P(I)}{P(A|I) \text{ x } P(I) + P(A|gnI) \text{ x } P(gnI)}$$

$$P(I|A) = \frac{0{,}85 \,.\, 0{,}0001}{0{,}85 \text{ x } 0{,}0001 + 0{,}01 \text{ x } 0{,}9999} = \frac{0{,}000085}{0{,}010084} = 0{,}008429$$

Oplossing oefening 9.1

We schetsen eerst alle mogelijke resultaten (=uitkomstenruimte) van dit experiment in een kruistabel.

	Eerste dobbelsteen					
Tweede dobbelsteen	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Vervolgens gebruiken we deze uitkomstenruimte om de simultane kansverdeling op te stellen voor de som van de twee dobbelstenen (X).

X	P
2	**1/36**
3	**2/36**
4	**3/36**
5	4/36
6	5/36
7	6/36
8	5/36
9	**4/36**
10	**3/36**
11	**2/36**
12	**1/36**

De elementaire uitkomsten die je 10 Euro rijker maken zijn in het rood gearceerd. Wanneer we de kansen van al deze elementaire uitkomsten bij elkaar optellen krijgen we een kans van 16/36. Het complement daarvan, 20/36, geeft de kans dat je 10 Euro moet betalen. Dus, je hebt meer kans dat je moet betalen, dan dat je kan winnen.

Oplossing oefening 9.2

We vullen allereerst onze tabel aan (gearceerde waarden).

Y\X	0	1	P(X=x)
0	0,36	0,24	0,60
1	0,18	0,12	0,30
2	0,06	0,04	0,10
P(Y=y)	0,60	0,40	1

De voorwaardelijke kans

$$P(Y = 2|X = 1) = \frac{P(Y = 2 \cap X = 1)}{P(X = 1)} = \frac{0{,}04}{0{,}40} = 0{,}10$$

De som

$$P(Y = 0 \cup X = 0) = P(Y = 0) + P(X = 0) - P(Y = 0 \cap X = 0) = 0{,}60 + 0{,}60 - 0{,}36 = 0{,}84$$

De standaardafwijking en de verwachte waarde van de stochast (X * Y)

Hiervoor bouwen we eerste de simultane kansverdeling

(X * Y)	P
0	0,84
1	0,12
2	0,04

Voor het verdere verloop van de berekening verwijzen we naar paragraaf 9.3.

Het resultaat is een verwachte waarde van 0,2 en een standaardafwijking van 0,48989

Oplossing oefening 9.3

Bepaal de kansverdeling van het totaal aantal maaltijden dat mevrouw X die twee zondagen zal nuttigen.

Eerst maken we een overzicht van de 6 gezinnen.

Gezin	Aantal maaltijden
A	1
B	2
C	2
D	2
E	3
F	3

Mevrouw X kiest op twee zondagen 2 gezinnen. We kijken hoeveel mogelijke keuzes ze kan maken. Dat komt neer op de berekening van een combinatie, mb. op hoeveel manieren kan je 2 gezinnen kiezen uit een groep van 6. De berekening hiervoor is :

$$C_n^k = \frac{n!}{k!.(n-k)!} = \frac{6!}{2!.4!} = \frac{6.5.4.3.2.1}{2.1.4.3.2.1} = 15$$

Je kan de situatie ook schetsen in een simultane kanstabel, waarbij de zwarte cellen niet kunnen voorkomen.

		Eerste keuze						
		A (1)	B (2)	C (2)	D (2)	E (3)	F (3)	
Tweede keuze	A (1)		3	3	3	4	4	1/6
	B (2)	3		4	4	5	5	1/6
	C (2)	3	4		4	5	5	1/6
	D (2)	3	4	4		5	5	1/6
	E (3)	4	5	5	4		6	1/6
	F (3)	4	5	5	5	6		1/6
		1/6	1/6	1/6	1/6	1/6	1/6	1

We zetten alle mogelijke keuzes (combinaties) uit in een tabel met daarnaast het aantal maaltijden voor de twee zondagen. Omdat de volgorde waarin de gezinnen worden bezocht geen rol speelt zijn er maar 15 combinaties.

Combinatie	Aantal maaltijden	Som aantal maaltijden
AB	1+2	3
AC	1+2	3
AD	1+2	3
AE	1+3	4
AF	1+3	4
BC	2+2	4
BD	2+2	4
BE	2+3	5
BF	2+3	5
CD	2+2	4
CE	2+3	5
CF	2+3	5
DE	2+3	5
DF	2+3	5
EF	3+3	6

De waarden van onze stochast (aantal maaltijden) loopt dus tussen 3 en 6. Samen met de respectievelijke kansen ziet onze kansverdeling er als volgt uit:

Aantal maaltijden (X)	P
3	3/15
4	5/15
5	6/15
6	1/15
	1

Het te verwachten aantal maaltijden berekenen we door de verwachte waarde te berekenen. Deze maal zonder rekentoestel.

Aantal maaltijden (X)	P	P*X
3	3/15	9/15
4	5/15	20/15
5	6/15	30/15
6	1/15	6/15
Som	1	65/15

De verwachte waarde E[X] = $\sum P * X$ = 65/15 = 4.33 maaltijden.

Variant: bepaal ook de kansverdeling van het totaal aantal maaltijden wanneer mevrouw X ook voor de bepaling van het tweede gezin opnieuw aselect kiest (dus, eerste gezin opnieuw mogelijk).

We schetsen de situatie in een kruistabel.

		Eerste gezin			
	aantal maaltijden	1	2	3	P
Tweede gezin	1	2	3	4	1/6
	2	3	4	5	3/6
	3	4	5	6	2/6
	P	1/6	3/6	2/6	1

Omdat er tweemaal na elkaar aselect wordt gekozen zijn de kansen onafhankelijk van elkaar. Door middel van de productregel voor onafhankelijkheid kunnen we nu de respectievelijke kansen in elke cel bijschrijven (= de waarde tussen haakjes).

		Eerste gezin			
	aantal maaltijden	1	2	3	P
Tweede gezin	1	2 (1/36)	3 (3/36)	4 (2/36)	1/6
	2	3 (3/36)	4 (9/36)	5 (6/36)	3/6
	3	4 (2/36)	5 (6/36)	6 (4/36)	2/6
	P	1/6	3/6	2/6	1

Tenslotte bouwen we de simultane kansverdeling. De stochastische waarden lopen deze keer tussen 2 en 6.

Een andere werkwijze is opnieuw gebruik te maken van de simultane kanstabel, waarbij deze keer de diagonaal situaties wel kunnen voorkomen.

		Eerste keuze						
		A (1)	B (2)	C (2)	D (2)	E (3)	F (3)	
Tweede keuze	A (1)	2	3	3	3	4	4	1/6
	B (2)	3	4	4	4	5	5	1/6
	C (2)	3	4	4	4	5	5	1/6
	D (2)	3	4	4	4	5	5	1/6
	E (3)	4	5	5	4	6	6	1/6
	F (3)	4	5	5	5	6	6	1/6
		1/6	1/6	1/6	1/6	1/6	1/6	1

Antal maaltijden	P
2	1/36
3	6/36
4	13/36
5	12/36
6	4/36
Som	1

De verwachte waarde hier is 156/36 = 4.33

Antal maaltijden	P	P*X
2	1/36	2/36
3	6/36	18/36
4	13/36	52/36
5	12/36	60/36
6	4/36	24/36
Som	1	156/36

Oplossing oefening 9.4

Het antwoord is 0,3. In formele kansnotatie tellen we twee producttermen bij elkaar op:

$$P(Vierkant \cap Rood) + P(Vierkant \cap Groen) = \frac{1}{10} + \frac{2}{10} = \frac{3}{10}$$

Oplossing oefening 9.5

We schetsen eerste de situatie van 4 gezinnen.

Gezin	Aantal kinderen
A	1
B	2
C	2
D	5

Er moeten 2 gezinnen worden gekozen uit een groep van 4. Dat is de berekening van een combinatie van 4 boven 2.

$$C_n^k = \frac{n!}{k!.(n-k)!} = \frac{4!}{2!.2!} = \frac{4.3.2.1}{2.1..2.1} = 6$$

De 6 mogelijke combinaties zijn:

Combinatie	Aantal kinderen
AB	3
AC	3
AD	6
BC	4
BD	7
CD	7

In een kansverdeling zijn de waarden van onze stochast (X) dus 3, 4, 6 en 7.

Aantal kinderen (X)	P
3	2/6
4	1/6
6	1/6
7	2/6

Oplossing oefening 9.6

Geef E[X] en σ voor stochast S.

Breng de kansverdeling in in je rekentoestel, zoals reeds getoond in opdracht 2.

De verwachte waarde = 5,57, de standaardafwijking = 2,44

Geef de cumulatieve kansverdeling van A.

Hiervoor moeten we eerst het aantal eerste plaatsen zien te vinden die aanwezig zijn in de som van de drie keuzes.

S	P[S=s]	Eerste plaatsen (patronen)	Eerste plaatsen (aantal)
3	10/35	1 -1- 1	3
4	10/35	2 -1 -1 (of 121 of 112)	2
8	10/35	6 -1 -1 (of 116 of 161)	2
9	5/35	6 -2 -1 (of 261 of 216 of 162)	1

De kansverdeling voor het aantal eerste plaatsen is dan:

Aantal eerste plaatsen (A)	P	Cumulatieve P
1	5/35	5/35
2	20/35	25/35
3	10/35	35/35

Geef de simultane kansverdeling voor de stochasten S en A.

De stochasten zijn afhankelijk en sommige combinaties zijn onmogelijk (grijs). De producten in de tabel worden dan ook berekend d.m.v. de afhankelijkheidsregel.

	S				
A	3	4	8	9	
1				5/35	5/35
2		10/35	10/35		20/35
3	10/35				10/35
	10/35	10/35	10/35	5/35	1

Geef de marginale kansverdeling van A.

Aantal eerste plaatsen (A)	P
1	5/35
2	20/35
3	10/35

Zijn S en A onafhankelijk? Waarom wel/niet?

De stochasten zijn afhankelijk omdat sommige waarden in één stochast, automatisch waarden van een andere stochast uitsluiten. Zo kan je onmogelijk een som van 9 bekomen, wanneer je 3 eerste plaatsen hebt.

$P(A \cap S) \neq P(A)$ x $P(B)$

$P(A \cap S) = P(A|S)$ x $P(S) = P(S|A)$ x $P(A)$

Oplossing oefening 9.7

Smul & Vul	$P(X=x_i)$	X.P
20	0,6	12
30	0,4	12
		E(X) = 24

Je verwacht 24 euro uit te geven.

Oplossing oefening 9.8

X	$P(X=x_i)$	X.P
30-20 = 10	0,1	1
90-20 = 70	0,4	28
60-20 = 40	0,3	12
50-20 = 30	0,1	3
80-20 = 60	0,1	6
		E(X) = 50

De dagprijs per persoon moet 50 euro bedragen.

Oplossing oefening 10.1

We gebruiken hier de Poisson verdeling met een λ van 1 (per vijf pagina's). Omdat het "interval" in de vraag 1 pagina is, moeten we de λ waarde aanpassen naar 0,2. De kans op meer dan 1 wordt dan:

Met de tabellen

$$P(x \geq 2) = 1 - P(x \leq 1)$$
$$= 1 - (0{,}1637 + 0{,}8187)$$
$$= 0{,}0176$$

Met het grafisch rekentoestel

1-POISSONCDF(0,2,1)

=0,01752….dus afgerond 0,0176

Oplossing oefening 10.2

We gebruiken hier de POISSON verdeling met λ= 25 (per jaar).

1ste vraag:
$$P(x \leq 24) = POISSONCDF(25,24)$$
$$= 0,4734$$

2de vraag
$$P(x \geq 31) = 1 - P(x \leq 30)$$
$$1 - POISSONCDF(25,30)$$
$$= 0,1367$$

Oplossing oefening 10.3

De verjaardag van de eerste persoon mag op elke dag van het jaar vallen. Neem een willekeurig ander persoon uit de 500. De verjaardag van deze persoon mag niet op die van de eerste vallen, dus dat zijn 364 van de 365 dagen (364/365). Dit geldt voor elke willekeurige andere persoon uit de resterende 498 personen.

Dus wordt de berekening:

$$\frac{365}{365}.\left(\frac{364}{365}\right)^{499} = 0,25436$$

Of ongeveer 25,44 procent kans.

Oplossing oefening 10.4

Dit is een binomiaal probleem. Wat we nodig hebben is de kans op ontevredenheid P(Tc) bij de concurrentie. Die kans kunnen we afleiden uit de tabel. Dat zijn de gearceerde cellen in onderstaande tabel:

tevreden	Telenet	Proximus	Belgacom
Ja	305	175	145
nee	125	100	150
n=	430	275	295

Kortom, 250/1000 of 25 %.

De noodzakelijke binomiale verdeling is B(10; 0,25) met als cumulatieve vraag P(x >= 7).

In onze tabel kijken we bij n=10 en r=7, in de kolom 0,25 en vinden een waarde van 0,0035 of 0,35%.

Oplossing oefening 10.5

We moeten om te beginnen de kans weten op bijwerkingen in de drie producten. Welke soort van bijwerking maakt niets uit, dus kunnen we onze tabel vereenvoudigen naar:

	n=48	n=48	n=64
Bijwerking	Product A	Product B	Placebo
JA	0,25	0,34	0,06
NEEN	0,75	0,66	0,94

We berekenen vervolgens de kans op bijwerking in de volledige groep, in de vraag wordt er immers geen onderscheid gemaakt tussen de 3 groepen.

De kans op bijwerking is dus de kans op bijwerking wanneer product A wordt genomen, plus de kans op bijwerking bij product B, plus de kans op bijwerking bij placebo. Formeel:

$P(Bw) = P(Bw \cap A) + P(Bw \cap B) + P(Bw \cap Plc)$

En dat kan worden berekend door middel van de productregel voor afhankelijkheid

$P(Bw) = P(Bw|A)\ \text{x}\ P(A) + P(Bw|B)\ \text{x}\ P(B) + P(Bw|Plc)\ \text{x}\ P(Plc)$

Dat is gelijk aan

$(0.25\ \text{x}\ 0.3) + (0.34\ \text{x}\ 0.3) + (0.06\ \text{x}\ 0.4) = 0{,}201$

Afgerond is onze kans op bijwerkingen in de groep van 160 deelnemers = 20%.

Opmerking: je kan de bovenstaande berekening ook d.m.v. een dendrogram visualiseren.

De uiteindelijke vraag:

Een meerderheid in een groep van 14 = 8 of meer.

Onze binomiaalvraag wordt P(X>=8) bij een verdeling B(14; 0,2)

In onze tabel kijken we bij n=14 en r = 8, bij een kans van 0,2

Resultaat 0.,0024 of 0,24%

Oplossing oefening 10.6

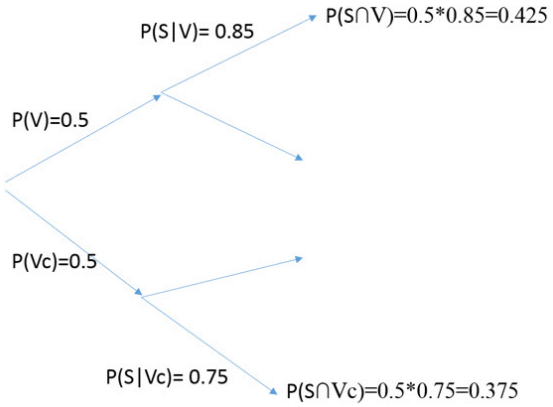
De kans dat iemand niet komt opdagen is hier 15%. Om geconfronteerd te worden met te weinig plaatsen moeten alle 14 de passagiers komen opdagen; In binomiaal termen zoeken we dus de kans bij B(14; 0,15) van P(X=0) of geen enkele persoon komt NIET opdagen.

In de tabel is dit P(X>=0) – P(X>=1) = 1 – 0,8972 = 0,1028 of 10,28%.

Oplossing oefening 10.7

We weten dat de kans op voorbereiding P(V) = 0,5 (de helft).

De rest van de waarden gaan we ditmaal illustreren d.m.v. een dendrogram



Dus, de kans op succes in het algemeen = 0,425 + 0,375 = 0,8

Vraag 1: wanneer we enkel kiezen uit de voorbereidingsgroep is de kans op succes = 0,85.

Bij een binomiaal B(10; 0,85) met als vraag P(X>=6). Omdat we de waarde van p=0,85 niet in onze tabel hebben draaien we de vraag om en zoeken we de kans dat er 4 of minder falen P(X<=4) bij een binomiaal B(10; 0,15) .

Via de BINOMCDF() functie geeft dit 0,9901

Via de tabel zoeken we 1 – P(X>=5): we kijken bij n=10 en r=5. In de kolom 0,15. We lezen een waard af van 0,0014. Hiervan moeten we nog het complement nemen

1 – 0,0099 = 0,9901

Vraag 2: het algemeen succes is 0,80

Alle stappen blijven gelijk maar we werken hier met een kans van 0,80.

Resultaat is: 0,9672

Oplossing oefening 10.8

Elke staaf stelt in feite een binomiale situatie voor. In de leeftijdsgroep van

45-54 is P(F) = 0,35 en de P(Fc) = 0,65.

55-64 = P(F)=0,20 en P(Fc)=0,80 en

65+ P(F)=0,07 en P(Fc)=0,93.

Vraag 1:

De berekening van de verwachte waarde E[X] in de drie groepen is steeds n.p. LET OP we zoeken personen ZONDER Facebook account, dus:

45-54: E[X] = 100 x 0.65 = 65

55-64: E[X] = 100 x 0,80 = 80

65+: E[X] = 100 x 0,93 = 93

Samen is dit 238 personen uit de groep van 300

MERK OP dat dit evengoed kan worden gezien als een berekening van een gewogen gemiddelde:

(100 x 0,65) + (100 x 0,8) + (100 x 0,93)

Vraag 2:

Hier gebruiken we de regel VAR[X + Y + Z] = VAR[X] + VAR[Y] + VAR[Z], waarbij de VAR steeds wordt berekend d.m.v. n.p.q

(100 x 0,35 x 0,65) + (100 x 0,20 x 0,80) + (100 x 0,07 x 0,93) = 45,26

Oplossing oefening 10.9

De oplossing loopt analoog aan die uit opdracht 5.

De kans op weekenddode

$P(Dw) = P(Dw \cap 15_17j) + P(Dw \cap 18_24j) + P(Dw \cap 25_29j)$

Dat is gelijk aan

$P(Dw) = P(Dw|15_17j)$ x $P(15_17j) + P(Dw|18_24j)$ x $P(18_24j) + P(Dw|25_{29j})$ x $P(25_29j)$

=(0,37 x 0,028) + (0,7 x 0,053) + (0,39 x 0,046) = 0,05321

De verwachte waarde voor de volledige groep van 900 = np = 900 x 0,05321 = 47,889 weekenddoden.

Oplossing oefening 10.10

Het gemiddelde van een binomiale verdeelde stochast = np

De variantie van een binomiale verdeelde stochast = npq

$$VC = \frac{\sigma}{\mu}$$

$$VC = \frac{\sqrt{npq}}{np}$$

$$VC = \frac{\sqrt{30 \times 0,6 \times 0,4}}{30 \times 0,6}$$

$$VC = 0,1491$$

Oplossing oefening 10.11

Dit is een hypergeometrisch probleem, telkens je een getal trekt (verandert) daalt immers de kans op een even nummer.

Van de 42 nummers zijn er 21 even. De eerste keer dat je een getal trekt heb je een kans van 21/42, de tweede keer een kans van 20/41, enz.

X~H(6;21;42)

$$\frac{\binom{21}{6}\binom{21}{0}}{\binom{42}{6}} = \frac{54264 \times 1}{5245786} = 0,0103$$

Oplossing oefening 11.1.

!!! Hoewel dit in de onderstaande oplossingen niet gebeurd is, verdient het aanbeveling om de gezochte grenswaarde en/of kansdichtheid te visualiseren aan de hand van een tekening.

X = IQ leerlingen secundair onderwijs in Vlaanderen

X~N(100;20)

$P(X < 130) = ?$

We zoeken de kans dat X kleiner is dan 130. Mogelijke berekeningswijzen:

a)We berekenen de oppervlakte links van de x-waarde = 130 onder de curve van de

normaalverdeling met gemiddelde 100 en standaardafwijking 20.

Deze oppervlakte zouden we kunnen berekenen met behulp van integraalrekening. De oppervlakte van -∞ tot 130 onder de curve van de normaalverdeling met μ = 100 en σ = 20 is gelijk aan:

$$\int_{-\infty}^{130} \frac{1}{20\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-100}{20}\right)^2} dx$$

b)Maar we kunnen deze kans gelukkig ook via de tabel van de standaardnormale verdeling berekenen. Elke normaal verdeelde kansvariabele X~N(μ,σ) kan immers via standaardisering $Z = \frac{X - \mu}{\sigma}$ tot een standaardnormaal verdeelde variabele Z~N(0;1) herleid worden.

Uitgaande van de verdeling van het IQ van de leerlingen: X~N(100,20), bekomen we dus een standaardnormaal verdeelde variabele $Z = \frac{X - 100}{20}$ met Z~N(0,1). Indien we de kans zoeken dat X<130, komt dit overeen met het zoeken naar de kans dat $Z < \frac{130 - 100}{20}$ of P(Z<1,5). Tabel 8.1 leert ons dat de gecumuleerde kansdichtheid bij de z-score 1,50 = 0,9332. De kans dat een leerling een IQ lager dan 130 heeft, bedraagt dus 93,32%. Of: 93,32% van de leerlingen heeft een IQ lager dan 130.

X = IQ leerlingen secundair onderwijs in Vlaanderen

X~N(100;20)

We zoeken eerst de kans dat het IQ tussen 90 en 110 ligt.

$P(90 < X < 110) = ?$

Om gebruik te kunnen maken van de tabel van de gestandaardiseerde normale verdeling werken we met de z-scores en zoeken we bijgevolg $P\left(\frac{90-100}{20} < Z < \frac{110-100}{20}\right)$ of $P(-0{,}5 < Z < 0{,}5)$ met Z~N(0;1). Aangezien tabel 8.1 gecumuleerde kansdichtheden geeft, met andere woorden kansdichtheden van de vorm $P(Z < z)$ dienen we $P(-0{,}5 < Z < 0{,}5)$ om te zetten als volgt:

$P(-0{,}5 < Z < 0{,}5) = P(Z < 0{,}5) - P(Z < -0{,}5)$ (1)

$= P(Z < 0{,}5) - [1 - P(Z < 0{,}5)]$ (2)

$= 2 \times P(Z < 0{,}5) - 1$

= 2 x 0,6915 – 1

= 0,383

a)Indien je dit grafisch zou voorstellen: de oppervlakte onder de curve van de standaardnormale verdeling tussen de waarden –0,5 en 0,5 is gelijk aan de oppervlakte onder deze curve links van 0,5 min de oppervlakte links van –0,5.

b)Omdat de N(0;1)-tabel in dit handboek enkel positieve z-waarden behelst, kunnen we $P(Z<-0{,}5)$ niet rechtstreeks aflezen. Omwille van de symmetrie van de standaardnormale verdeling is $P(Z<-0{,}5) = P(Z>0{,}5)$. Wetende dat de totale oppervlakte onder de curve van de gestandaardiseerde normale verdeling 100% of de totale kansdichtheid 1 bedraagt, kunnen we $P(Z>0{,}5)$ gelijkstellen aan $1-P(Z<0{,}5)$.

In een school van 500 leerlingen zullen 500 x 0,383 = 191,5 of 192 (volgens eerder geziene afrondingsregel) leerlingen een IQ tussen 90 en 110 hebben

Het derde kwartiel is de x-waarde waarvoor geldt $P(X<x)=0{,}75$. Opdat we de tabel van de standaardnormale verdeling zouden kunnen gebruiken, werken we met z-scores. We zoeken bijgevolg de z-waarde waarvoor geldt: $P(Z<z)=0{,}75$, om hieruit vervolgens de gezochte x-waarde te berekenen. Aangezien z een lineaire transformatie van x is, komt dit overeen met het zoeken naar de x-waarde waarvoor geldt dat $P\left(Z<\frac{x-100}{20}\right)=0{,}75$.

In de tabel van de gestandaardiseerde normaalverdeling vinden we dat de z-waarde waarvoor de gecumuleerde kansdichtheid P(Z<z)= 0,75 om en bij de 0,67 is, met andere woorden P(Z<0,67)= 0,75. Hieruit volgt dat $\frac{x-100}{20}=0{,}67$. Het oplossen van deze vergelijking naar x levert een IQ-score (x-waarde) op van 113,4.

Hoe kunnen we dit 3de kwartiel interpreteren? 75% van de leerlingen heeft een IQ van hoogstens 113,4. Leerlingen met een IQ hoger dan 113,4 behoren dus tot de 25% meest intelligente leerlingen.

Oplossing oefening 11.2

Om de overlap te berekenen moeten we voor de grijze verdeling (links) het stuk boven de waarde 17 berekenen, en voor de witte verdeling (rechts) het stuk onder de waarde 17.

We bereken eerst de z-waarde voor de ruwe waarde 17 voor beide verdelingen:

Voor de linkse verdeling: Z_1 = (17-12)/3 = 1,67

Voor de rechtse verdeling: Z_2 = (17-20)/2 = -1,5

Door middel van de tabellen zoeken we nu op hoeveel procent er respectievelijk boven en onder de berekende z-waarden ligt.

P(Z > 1,67) = 1 – 0,9525 = 4,75%

P(Z < -1,5) = 1 – 9332 = 6,68%

Samen geeft dit een overlap van 11,46%.

Oplossing oefening 11.3

Vraag a.

We moeten de kans berekenen op een lengte hoger dan 205 cm in een verdeling met gemiddelde 175 cm en standaarddeviatie 7 cm.

We berekenen eerst onze z-score: (205 – 175)/7 = 4,2857

In onze tabel lopen de waarden tot maximaal 3,99. Een waarde van 4,28 of hoger is dus zeer uitzonderlijk, met een kans kleiner dan 0,01%.

We kunnen dit ook berekenen d.m.v. de NormalCDF functie van je grafisch rekentoestel:

NORMALCDF(205, 20000, 175,7)

We gebruiken de waarde 20000 (of een andere extreem hoge waarde) om oneindig aan te geven.

Resultaat is 0,00000911

Vraag b

Hier moeten we de hoogte vinden waarboven nog 5% van de lengtes ligt. Daarvoor gebruiken we de invers-normale functie met als input percentiel 95.

INVNORM(0.95, 175, 7) = 186,51 cm

Gebruiken we de tabel, dan zoeken we 'binnen' in de tabel de waarde dichtst bij .9500. Dat is 1,64 of 1,65. Daarna rekenen we deze z-waarde om naar ruwe lengtes in cm:

(1.64 x 7) + 175 = 186,48

Oplossing oefening 11.4

We zoeken het percentage boven de ruwe waarde 70 in een verdeling met gemiddelde 50 en standaarddeviatie 8. We berekenen eerst de z-waarde voor de ruwe waarde 70:

Z = (70 – 50)/8 = 2,5

In de tabel zien we dat onder een waarde van z = 2,5 er 99,38% van de waarden liggen. Dus, boven deze grens liggen er nog 1 – 0,9938 = 0,6% van de waarden.

Oplossing oefening 11.5

We zoeken de z-waarden die overeen komen met het punt waaronder 97,5 % van de waarden ligt en het punt waaronder 2,5% van de waarden ligt. Dit is een inverse-normaal handeling.

We zoeken 'binnen' in de tabel naar de waarde die het dichtst bij de waarde 0,9750 ligt.

Dat is exact z = 1,96. Voor de z-waarde langs de linkerzijde is dit uiteraard -1,96.

Vervolgens rekenen we deze z-waarden om naar breedtes in cm:

Bovengrens = (1,96 x 2.5) + 15,5 = 20,4 cm

Ondergrens = (-1,96 x 2,5)/15,5 = 10,6

Oplossing oefening 11.6

Vraag a

De kans om te slagen op proef A: Z = (70 – 60)/10 = 1

De kans om boven z-waarde 1 uit te komen is (via tabel): 1 – 0,8413 = 15,87%

De kans om te slagen op proef B: Z = (50 – 40)/12 = 0,83

De kans om boven een z-score van 0,83 uit te komen (via tabel) = 1 - 0,7967 = 20,33%

Indien beide proeven onafhankelijk zijn, is de kans om op beide delen te slagen = 0,1587 * 0,2033 = 0,0322 of 3,22%

Op een groep van 500 kandidaten maakt dat 500 x 0,0322 = 16,10 dus ong. 16 kandidaten.

Vraag b

125 van de 500, geeft een selectiepercentage van 25%. Dat kan enkel worden bereikt indien de slaagkansen op proef A en op proef B 50% bedragen. Immers 0,5 x 0,5 = 0,25.

Als 50% moet slagen op elke proef, moet de gehanteerde minimumscore gelijk zijn aan het gemiddelde. Dus, 60/100 op proef A en 40/100 op proef B.

Oplossing oefening 11.7

Vraag a

We zien op de figuur dat percentiel 50 (= gemiddelde) overeen komt met de ruwe waarde 25. Verder zien we dat percentiel 80 perfect op de ruwe score van 30 uitkomt. De z-score voor percentiel 80 geeft 0,84 (in de tabel). Voor het gemiddelde is dit uiteraard 0. Kortom, het verschil van 5 ruwe punten (tussen de ruwe score van 25 en 30) is 0,84 standaarddeviaties lang.

5 cm = 0,84 standaardafwijkingen

1 standaardafwijking = (5/0,84) = 5,95 cm

Vraag b

70% van een maximale score van 45 is gelijk aan een score van 31,5/45.

Vraag c

De kans om een score van 31,5 of hoger te behalen berekenen we door eerst de z-waarde van 31,5 te bepalen:

Z = (31,5 -25)/5,94 = 1,09. In de tabel lezen we bij deze waarde 0,8621. Dus, 1 – 0,8621 = 13,79% kans om hoger te scoren.

D.m.v. het rekentoestel berekenen we dit als volgt: NORMALCDF(31,5, 5000, 25, 5,94) = 13,69. De waarde 5000 dient hier opnieuw als oneindigheid.

De kans om in een dergelijke groep te slagen is dus (afgerond) 14 procent. Om het aantal in een groep van 10 te berekenen gebruiken we de binomiale verdeling ~B(10; 0,14).

Met als vraag P(X ≥ 3).

Via BINOMCDF (let op: werkt enkel in de richting ≤):

1 – P(X ≤ 2) = 1 – BINOMCDF(10, 0.14, 2) = 0,1545 of 15,45%

De tabel van de binomiale verdeling geeft geen p-kans van 0,14, dus gebruiken we 0,15.

P(X ≥ 3) bij n=10 en p = 0,15 → 17,98%

Oplossing oefening 11.8

Uitgedrukt in Z-scores is 5% van de z-scores kleiner dan -1,64 of 1,65. De ruwe waarde 160 moet met andere woorden overeenkomen met de z-waarde -1,64 of -1,65.

$$-1,64 = \frac{160-175}{\sigma} \qquad -1,65 = \frac{160-175}{\sigma}$$

$$\sigma = \frac{-15}{-1,64} = 9,1463 \qquad \sigma = \frac{-15}{-1,65} = 9,0909$$

Oplossing oefening 11.9

De z-score waarvoor 25% groter is (en dus 75% kleiner) is 0,67 of 0,68. We gaan een continue verdeling gebruiken om een discreet probleem op te lossen, dus we hebben een continuïteitscorrectie nodig. In de opgave staat dat we de kans moeten berekenen op meer dan 320 eenheden. De CC bedraagt steeds een halve eenheid en aangezien 320 er niet meer bij hoort ('méér dan') tellen we de CC erbij op.

De vraag wordt dus: $P(X \geq 320,5) = 75\%$ ~N(μ;52) → μ = ?

$$0,67 = \frac{320,5-\mu}{52}$$
$$0,67 \times 52 - 320,5 = -\mu$$
$$\mu = 285,66$$

Oplossing oefening 11.10

We moeten dus de ruwe waardes van Q_1 en Q_3 berekenen. De z-scores voor Q1 en Q3 komen respectievelijk overeen met -0,67 en +0,67.

$$-0,67 = \frac{x_i - 63}{4} \qquad 0,67 = \frac{x_i - 63}{4}$$
$$-0,67 \times 4 + 63 = x_i \qquad 0,67 \times 4 + 63 = x_i$$
$$x_i = 60,32 \qquad x_i = 65,68$$

$$IKA = Q_3 - Q_1$$
$$IKA = 65,68 - 60,32$$
$$IKA = 5,36$$

Oplossing oefening 12.1

Vraag a) Tenminste 60 personen kiezen zoute chips.

Een binomiaal probleem:

n = 150

P(zoute chips) = 0,33

waarbij np = 49,5 en nq = 100,5, dus mag een benadering d.m.v. de normaalverdeling worden doorgevoerd. Hierbij moet wel rekening worden gehouden met **de continuïteitscorrectie**.

De vraag is P(x ≥ 60), met een continuïteitscorrectie wordt dat ≥ 59,5

We berekenen de z-waarde:

$$\frac{x - np}{\sqrt{npq}} = \frac{59{,}5 - 49{,}5}{5{,}7589} = 1{,}736$$

We zoeken vervolgens de kans dat een z-waarde groter is dan 1,736

In de normaaltabel kunnen we aflezen dat de kans dat een z-waarde KLEINER is dan 1,73 = 0,9582, dus nemen we hiervan het complement:

1 – 0,9582 = 0,0418

Indien men gebruik maakt van de NormalCDF functie: NORMALCDF(1,736, 100). De waarde 100 (= 100 standaarddeviaties) gebruiken we als vervanger voor plus oneindig.

Resultaat = 0,0413

Deze vraag hadden we ook kunnen oplossen door gebruik te maken van **proporties** i.p.v. absolute waarden. Voor P hoedje moet je ook wel de CC gebruiken.

$$\frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0{,}397 - 0{,}33}{0{,}03839} = 1{,}746$$

Als controle berekenen we de **exacte binomiale kans** d.m.v. ons rekentoestel. Let wel op, hier heb je een complement nodig aangezien ons rekentoestel enkel in de richting van kleiner of gelijk aan werkt

1 - BINOMCDF(150, .33, 59) = 0,04277

Vraag b) Exact 60 zullen zoute chips kiezen

Wederom mogen we dit binomiaal probleem benaderen d.m.v. de normale verdeling. We houden rekening met de continuïteitscorrectie en bouwen dus eerst een mini-interval rond de waarde 60. Dat wordt een interval tussen 59,5 en 60,5. We berekenen vervolgens de beide z-waarden.

Voor de waarde van 59,5 hebben we dit reeds gedaan in punt a. Resultaat was 1,736. P(z ≤ 1,73) = 0,9582 (via tabel).

Voor de waarde 60.5:

$$\frac{x - np}{\sqrt{npq}} = \frac{60{,}5 - 49{,}5}{5{,}7589} = 1{,}91$$

P(z ≤ 1,91) = 0,9719

De oppervlakte voor dit kleine interval bekomen we nu door de kans op 59,5 of minder, weg te rekenen bij de kans van 60,5 of minder. Dus:

0,9719 – 0,9582 = 0,0137

Als controle berekenen we de exacte uitkomst met de BINOMDPF(150, .33, 60) = 0,013254

Vraag c) Minder dan 60 personen zoute chips zullen kiezen

Dat is de kans op X ≤ 59,5 (met CC)

= 0,958718

Vraag d) Wanneer er van elk product 70 stuks aan boord zijn, wat is de kans dat een passagier niet zijn product van keuze zal kunnen krijgen.

Deze situatie doet zich voor wanneer er meer dan 70 stuks worden gevraagd van een bepaald product. Met CC wordt dit ≥ 70,5

$$\frac{x - np}{\sqrt{npq}} = \frac{70{,}5 - 49.5}{5{,}7589} = 3{,}6465$$

Deze z-waarde is zo groot dat er in onze z-tabel de waarde 0,9999 wordt gegeven. De kans dat er dus nog hogere waarden worden gevonden is kleiner dan 0,001%.

Vraag e) Hoeveel stuks moeten er aan boord worden meegenomen wil de maatschappij maar 1% kans lopen de wens van een passagier niet te kunnen inwilligen

De z-waarde waarboven nog 1% van de gegevens te vinden zijn is (via de tabel) 2,33.

We zoeken nu de ruwe x-waarde in onze vergelijking

$$\frac{x - np}{\sqrt{npq}} = \frac{x - 49{,}5}{5{,}7589} = 3{,}6465$$

Die vinden we door de vergelijking uit te werken:

X = (2,33 x 5,7589) + 49,5 = 62,91 of afgerond 63 stuks.

Oplossing oefening 12.2

De gegevens uit deze vraag zijn:

P(elektr fiets) = 0.125

P(geen elektr fiets) = 0,875

n = 500

np en nq zijn hoger dan 10, dus we mogen dit binomiaal probleem benaderen d.m.v. de normale verdeling.

De vraag is: p(x ≥ 50) → met CC wordt dit 49,5

49,5 zetten we om in een Z-waarde.

Formule:

$$\frac{x - np}{\sqrt{npq}} = \frac{49{,}5 - 62{,}5}{7{,}3951} = -1{,}7579$$

P(z $\geq$ -1,75) = P(z $\leq$ 1,75) = 0,9599

Controle d.m.v. BINOMCDF (complement noodzakelijk want de vraag is $\geq$):

1 – BINOMCDF(500, 0.125, 49) = 0,96389

Oplossing oefening 12.3

De gegevens zijn

N = 200

n = 50

kans op baars (p) = 0,4

kans op karper(q) = 0,6

steekproef is groter dan 10% van de populatie, dus een **hypergeometrisch probleem**. Dat kan men benaderen, mits gebruik te maken van zowel CC als van de eindigheidscorrectie.

Er wordt gevraagd naar een meerderheid in een steekproef van 50, dat wordt dus groter of gelijk aan 26. Met **CC** is dit 25,5.

De formule:

$$\frac{x - np}{\sqrt{npq}.\sqrt{\frac{N-n}{N-1}}} = \frac{25{,}5 - 20}{3{,}4641\ X\ 0{,}8682} = 1{,}8287$$

De kans op een z-waarde groter dan 1,82 (in tabel) geeft = 1 – 0,9656 = 0,0344

De exacte waarde d.m.v. een hypergeometrische applet (kan ook in Excel) is hieronder voorgesteld

Population size	200
Number of successes in population	80
Sample size	50
Number of successes in sample (x)	25
Hypergeometric Probability: P(X = 25)	0.0334294881613854
Cumulative Probability: P(X < 25)	0.932551080832414
Cumulative Probability: P(X ≤ 25)	0.965980568993799
Cumulative Probability: P(X > 25)	0.034019431006201
Cumulative Probability: P(X ≥ 25)	0.0674489191675863

Oplossing oefening 12.4

Dit een binomiaal probleem met volgende parameters:

n = 54

p = 0,2

Vraag: P ($X \geq 4$)

np = 10,8 en nq = 43,2 dus je mag dit binomiaal probleem benaderen d.m.v. de normale verdeling.

Met continuïteitscorrectie wordt de vraag dan: P ($X \geq 3,5$)

We zetten 3,5 om in een Z-score:

$$\frac{x - np}{\sqrt{npq}} = \frac{3,5 - 10,8}{2,9394} = -2,4835$$

P ($Z \geq$ -2,4835) = P ($Z \leq 2,48$) = 99,34%

Oplossing oefening 12.5

De gegevens uit deze vraag zijn:

P = 0,10

q = 0,90

n = 100

np =10 en nq = 90, dus benaderen mag.

De vraag komt neer op de kans dat er 4 of minder overlijdens zijn.

Met CC wordt dat 4,5.

Formule:

$$\frac{x - np}{\sqrt{npq}} = \frac{4,5 - 10}{3} = -1,833$$

In onze z-tabel zien we bij 1,83 de waarde 0,9664. Dat is de kans op kleiner dan PLUS 1,83. Wij zijn dus op zoek naar het deel erboven. Dat is 1 – 0,9664 = 0,0336 → antwoord A

Als controle de berekening d.m.v. de BINOMCDF(100, 0.10, 4) = 0,0237

Oplossing oefening 13.1

De formule die we hierbij nodig hebben is:

$$m = Z^* \times \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{z^* \times \sigma}{m} \right)^2$$

In cijfers wordt dat:

$$n = \left(\frac{(2,58)^2 \times (800)^2}{(200)^2} \right) = 106,5024$$

Afgerond is dit een steekproef van 107 personen

Oplossing oefening 13.2

De kritieke z-waarde is hier 1,65.

De meetfout = $\sqrt{npq}$ = 7,07107 OF $\sqrt{\frac{pq}{n}} =$ 0,0353

De foutmarge = 1,65 x 7,07107 = 11,66726 OF 1,65 x 0,0353 = 0,0583

In aantallen loopt het betrouwbaarheidsinterval loopt van 82,3327 tot 105,667.

In proporties loopt het betrouwbaarheidsinterval van 0,4117 tot 0,5283.

Oplossing oefening 13.3

Gegeven

N = 400

n = 21

$\hat{p} = \frac{7}{21}$

Onze steekproef bedraagt minder dan 10% van de populatie, dus we hebben geen eindigheidscorrectie nodig.

Het gaat om een proportietoets, dus we gebruiken de z-tabel.

De kritieke z-waarde bij een 90% betrouwbaarheidsinterval is 1,64.

$$m = Z^* \times \sqrt{\frac{\hat{p}\hat{q}}{n}}$$
$$m = 1,64 \times \sqrt{\frac{0,3333 \times 0,6667}{21}}$$
$$m = 0,1687$$

Bovengrens van het betrouwbaarheidsinterval

$$\hat{p} + m = 0,3333 + 0,1687 = 0,5020$$

Oplossing oefening 14.1

Vraag a) Toets de hypothese dat de schatting van 18000 Euro te laag is (alfa = 5%).

Gegeven

n= 100

steekproefgemiddelde = 18250

populatie gemiddelde = 18000

s = 700

sigma = 800

Hypothese

$H0: \mu = 18000 \qquad H1: \mu > 18000$

Dit is een rechts-eenzijdige hypothese. Een alfa van 5% geeft hierbij een kritieke z-waarde van 1,65.

Berekening van de z-waarde

$$z = \frac{18250 - 18000}{\frac{800}{\sqrt{100}}} = 3{,}125$$

Een z-waarde van 3,125 ligt ruim voorbij de kritieke z-waarde van 1,65. De waarde van 18000 is wel degelijk en te lage schatting. De gemiddelde schuld ligt hoger.

Vraag b) Wat is de overschrijdingskans van deze toets?

$$P(z \geq 3{,}125) = 0{,}0008$$

Vraag c) Welke hypothese zal je verwerpen H_0 of H_1?

Je hebt slechts 0,08% kans dat de nulhypothese klopt. Je verwerpt dus de nulhypothese dat de gemiddelde schuld gelijk is aan 18000 euro en je aanvaardt de alternatieve hypothese dat de gemiddelde schuld groter is dan 18000 euro.

Oplossing oefening 14.2

Gegeven

n = 200

x = 94 (of $\hat{p} = \frac{94}{200}$)

N = 3000

Een eindigheidscorrectie is dus niet nodig want de steekproef is kleiner dan 10% van de populatie.

Continuïteitscorrectie hebben we wel nodig want het is een proportietoets en aantallen zijn altijd discreet!

De hypothese

$H_0: p = 0{,}5 \quad H_1: p < 0{,}5$

Dit is een links-eenzijdige hypothesetoets. Met alfa = 1% is de kritieke z-waarde = -2,33 (volgens de tabel).

Berekening van de z-waarde

Met aantallen: $z = \frac{x(+CC) - np}{\sqrt{npq}} = \frac{94{,}5 - 100}{7{,}07107} = -0{,}7778$

OF met proporties: $z = \frac{\hat{p} - P}{\sqrt{PQ}} = \frac{\frac{94,5}{200} - \frac{100}{200}}{\sqrt{\frac{0,5 \times 0,5}{200}}} = -0,7778174593$

De z-waarde -0,7778 ligt niet voorbij de kritieke z-waarde van -2,33 en ligt dus in het aanvaardingsgebied.

Conclusie: We aanvaarden de nulhypothese, 94 verschilt niet significant van 100. Er zijn dus voldoende argumenten om een maandblad op te richten.

Oplossing oefening 14.3

Gegeven

n = 51

steekproefgemiddelde = 12,7

s = 5

$\mu = 14$

$\bar{x} = 12,7$

Alpha = 2,5%

Sigma is onbekend, de scores zijn in de populatie niet normaal verdeeld en we hebben een steekproef met meer dan 30 personen. Dus we gebruiken de T-tabel!

Hypothese (links-eenzijdig) $H0\ \mu = 14$ $H1\ \mu < 14$

Kritieke t- waarde bij 50 vrijheidsgraden en 2,5% staartkans = -2,009

Berekening van de t-waarde

$t = \frac{12.7 - 14}{\frac{5}{\sqrt{51}}} = -1,8568$

De t-waarde -1,8568 is groter dan de kritieke t-waarde (valt binnen het aanvaardingsgebied), dus wordt de nulhypothese aanvaard. Het niveau van de eerstejaars ligt niet significant lager dan de vooropgestelde norm.

Oplossing oefening 14.4

Gegeven

n = 51

x = 25 (of $\hat{p} = \frac{25}{51}$)

We werken nu met aantallen (discreet) en we hebben dus een **continuïteitscorrectie** nodig.

Dit is een proportietoets, dus mogen we steeds de **z-tabel** gebruiken.

De hypothese H_0: p = 0,6 H_1: p < 0,6

Met alfa = 2,5% is de kritieke z-waarde = -1,96

Berekening van de z-waarde

In aantallen: $z = \frac{x(+CC)-np}{\sqrt{npq}} = \frac{25,5-30,6}{3,498571} = -1,4577$

OF in proporties: $z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} = \frac{\frac{25,5}{51}-0,6}{\sqrt{\frac{0,6 \text{ x } 0,4}{51}}} = -1,4577$

De z-waarde -1,4577 is groter dan de vooropgestelde kritieke z-waarde -1,96 en ligt dus in het aanvaardingsgebied. We aanvaarden dus de nulhypothese. Als 25 van de 51 studenten een score halen boven de 14, verschilt dit niet significant van de norm van 60%.

Oplossing oefening 14.5

Gegeven

N = 400

n = 21

p = 0,27

$\hat{p} = \frac{7}{21}$

Onze steekproef bedraagt minder dan 10% van de populatie, dus we hebben geen eindigheidscorrectie nodig. Eigenlijk hebben we een continuïteitscorrectie nodig, maar in de opgave staat dat we deze mogen negeren.

Het gaat om een proportietoets, dus we gebruiken de z-tabel.

Hypothese

H_0: p = 0,27 H_1: p > 0,27

Berekening van de z-waarde

$z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} = \frac{0,3333-0,27}{\sqrt{\frac{0,27 \text{ x } 0,73}{21}}} = 0,6537$

Overschrijdingskans

P(Z > 0,6537) = 0,2578 (gebaseerd op de tabelwaarden)

Oplossing oefening 14.6

Gegeven

n = 64

$\overline{X}$ = 48,9

$\boldsymbol{\sigma}$ = 10

Sigma is gekend, dus z-tabel!

Hypothesetoets

H_0: $\mu = 50$ vs. H_1: $\mu < 50$

Berekening van de z-waarde

$$z = \frac{48{,}9 - 50}{\frac{10}{\sqrt{64}}} = -0{,}88$$

Overschrijdingskans

P(Z < -0,88) = 0,1894 (gebaseerd op tabelwaarden)