Statistiek: huiswerkopgaven

Dr. ir. D.A.M.P. Blom

18 april 2025

Week 1: grondslagen statistiek en datavisualisatie

Hoofdstuk 1

Opdracht 1.m1: Bij een verkeersonderzoek is een van de grootheden die wordt genoteerd het merk van de passerende auto's. Dit merk is ...

(a) een ratiovariabele

(b) een kwantitatieve variabele

(c) een nominale variabele.

(d) geen variabele

Uitwerking

Het juiste antwoord is (c)

Opdracht 1.m2: Zie tabel. Bij de mannen is het percentage dat het eens is met de voorgestelde maatregelen gelijk aan ...

(a) $41,7\%$

(b) 10%

(c) 25%

(d) $22,7\%$

Uitwerking

Het juiste antwoord is (d)

Opdracht 1.m3: Zie tabel. De groep mensen die het oneens is met de maatregelen bestaat voor ... uit vrouwen.

(a) 20%

(b) 56%

(c) 50%

(d) 40%

Uitwerking

Het juiste antwoord is (c)

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

Tabel 1: *
Aantal passagiers per busreis dat annuleert

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

Opdracht 1.m4: De klassenbreedte bij 0 -< 5 bedraagt ...

(a) 5

(b) 4

(c) $4,5$

(d) 36

Uitwerking

Het juiste antwoord is (a)

Opdracht 1.m5: De relatieve frequentie van de klasse 16 -< 20 bedraagt ...

(a) $0,05$

(b) 4

(c) 16

(d) $0,20$

Uitwerking

Het juiste antwoord is (a)

Opdracht 1.m6: Het klassenmidden van de klasse 8 -< 12 bedraagt ...

(a) 2

(b) 10

(c) $10,5$

(d) $9,5$

Uitwerking

Het juiste antwoord is (d)

Opdracht 1.3: Geef voor elk van de volgende gevallen aan of de genoemde verzameling
als een steekproef of als een populatie mag worden beschouwd:

(a) de commissarissen van de koning van de 12 Nederlandse provincies

(b) de 200 personen die zijn geïnterviewd bij een straatenquête

(c) de 150 automobilisten die moesten stoppen voor een alcoholcontrole

(d) de 740 leden van een studentenvereniging

(e) de 38 klanten die tussen 11.00 en 12.00 uur een postkantoor binnenkomen

(f) de 12000 verzekerden bij een verzekeringsmaatschappij

(g) de 20 nummers die worden gedraaid in een muziekprogramma op de radio

Uitwerking

(a) Populatie

(b) Steekproef

(c) Steekproef

(d) Populatie

(e) Steekproef

(f) Populatie

(g) Steekproef

Opdracht 1.4: Geef voor de volgende variabelen aan of deze een nominale, ordinale, interval- of ratioschaal heeft:

(a) de speelduur van een dvd

(b) de kleur van tulpen

(c) de industrietak waarin werknemers een baan hebben

(d) de jaaromzet (in euro) van bedrijven

(e) het aantal sterren dat de moeilijkheidsgraad van puzzelboekjes aangeeft

(f) de hoogte boven de zeespiegel van wintersportdorpen

Uitwerking

(a) Ratio

(b) Nominaal

(c) Nominaal

(d) Ratio

(e) Ordinaal

(f) Ratio

Opdracht 1.7: Een groep van dertig eerstejaarsstudenten is een aantal vragen voorgelegd. Dit betrof:

- Leeftijd
- Woonsituatie (z=zelfstandig, o=bij ouders)
- Geslacht (m=man, v=vrouw)
- De maandelijke bestedingen aan voedsel en drank
- De score voor het tentamen statistiek

In het boek staan de resultaten in een tabel.

(a) Geef aan op welk type schaal de vijf variabelen worden gemeten

Uitwerking

Naam: nominaal, leeftijd: ratio, woonsituatie: nominaal, geslacht: nominaal, besteding: ratio, score: interval

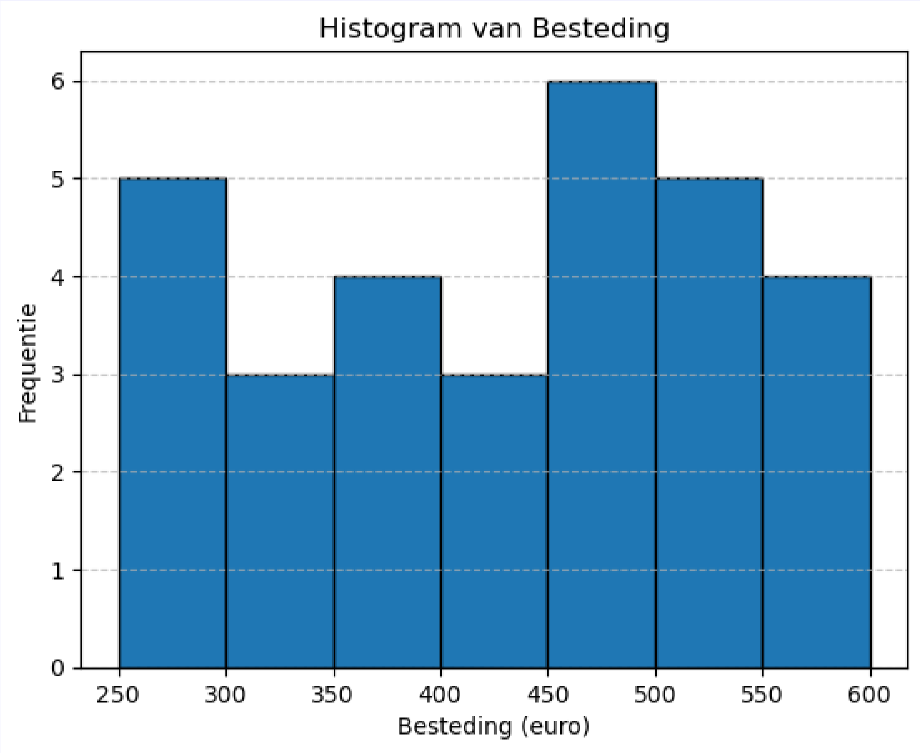
(b) Maak een frequentieverdeling van de leeftijden

Uitwerking

Leeftijd	**Frequentie**
18	2
19	5
20	6
21	6
22	5
23	3
24	1
25	1
26	1
Totaal	30

(c) Teken een histogram van de bestedingen. Begin met ondergrens €250.

Uitwerking



(d) Maak een klassenindeling van de scores voor mannen en vrouwen afzonderlijk. Kies voor de klassen 10 eenheden en verwerk de resultaten in een kruistabel. Bereken ook de procentuele frequenties van de scoreklassen voor de mannen en vrouwen afzonderlijk.

Uitwerking

De absolute frequentieverdeling van de scoreklassen naar geslacht is gelijk aan

Scoreklasse	Man	Vrouw	Totaal
$[30, 40)$	1	0	1
$[40, 50)$	1	5	6
$[50, 60)$	3	0	3
$[60, 70)$	4	2	6
$[70, 80)$	2	4	6
$[80, 90)$	4	1	5
$[90, 100)$	3	0	3
Totaal	18	12	30

De relatieve frequentieverdeling van de scoreklassen naar geslacht is gelijk aan

Scoreklasse	Man	Vrouw	Totaal
$[30, 40)$	5.6	0.0	5.6
$[40, 50)$	5.6	41.7	47.3
$[50, 60)$	16.7	0.0	16.7
$[60, 70)$	22.2	16.7	38.9
$[70, 80)$	11.1	33.3	44.4
$[80, 90)$	22.2	8.3	30.5
$[90, 100)$	16.7	0.0	16.7
Totaal	60	40	100

(e) Maak een kruistabel waarin de waarnemingen worden verdeeld naar geslacht en woonsituatie.

Uitwerking

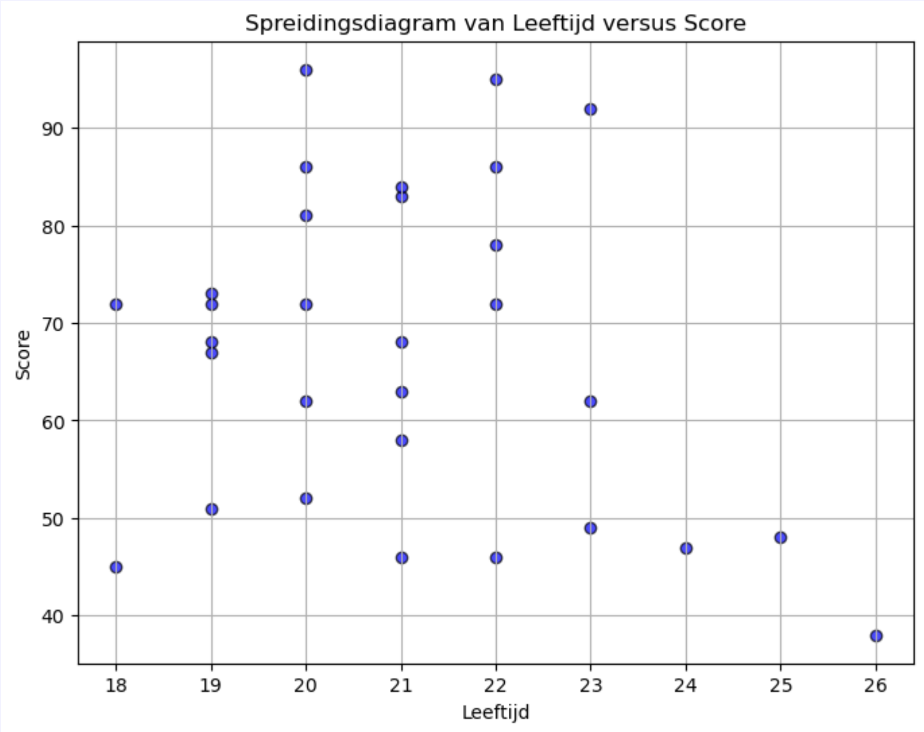
De kruistabel voor geslacht en woonsituatie is gelijk aan

Geslacht	Zelfstandig	Bij ouders	Totaal
Man	6	12	18
Vrouw	9	3	12
Totaal	15	15	30

(f) Teken een spreidingsdiagram met “leeftijd” langs de horizontale as en “score” langs de verticale as.

Uitwerking

De spreidingsdiagram van leeftijd (X) versus score (Y) ziet er als volgt uit:



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

Gemeten profiel (in mm)	**Aantal auto's**
$0{,}00 - < 2{,}00$	4
$2{,}00 - < 4{,}00$	34
$4{,}00 - < 6{,}00$	82
$6{,}00 - < 8{,}00$	66
$8{,}00 - < 10{,}00$	14
Totaal	200

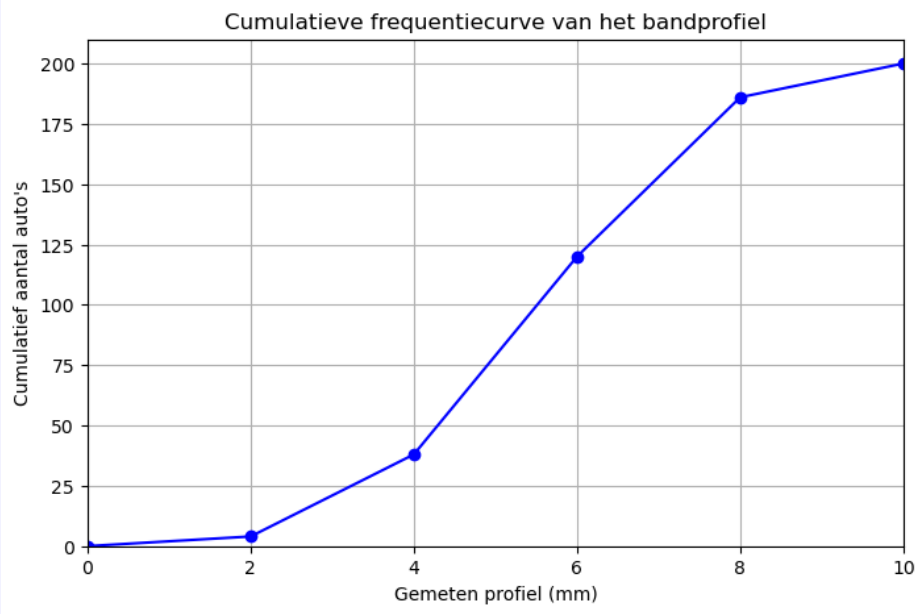
(a) Geef de waargenomen verdeling weer door middel van een cumulatieve frequentiecurve.

Uitwerking

De cumulatieve frequentieverdeling is als volgt:

Gemeten profiel (in mm)	**Cumulatief aantal auto's**
$0{,}00 - < 2{,}00$	4
$2{,}00 - < 4{,}00$	38
$4{,}00 - < 6{,}00$	120
$6{,}00 - < 8{,}00$	186
$8{,}00 - < 10{,}00$	200
Totaal	200

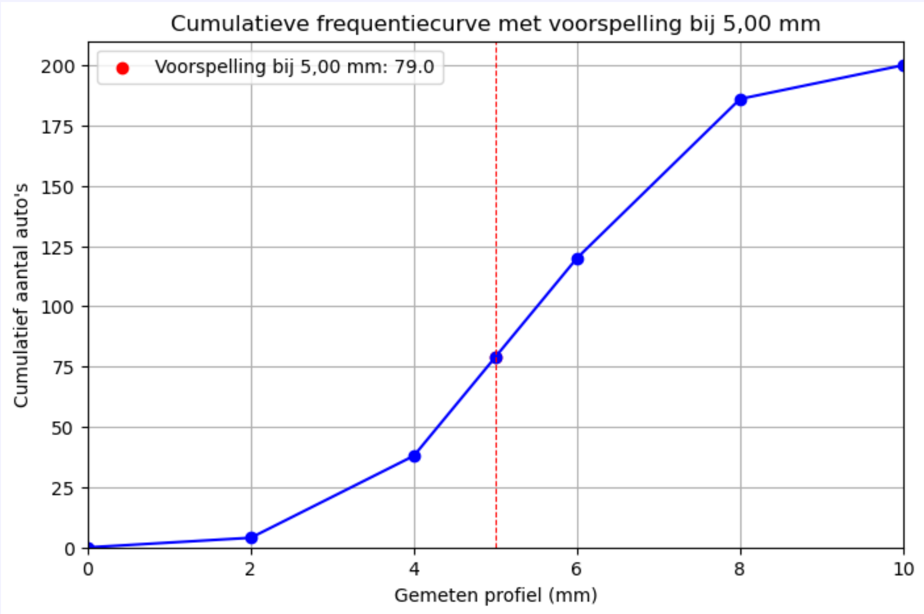
De cumulatieve frequentiecurve ziet er als volgt uit:



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

Uitwerking

De schatting voor het aantal auto's metg hoogstens een profiel van 5,00mm vinden we door te interpoleren op de cumulatieve frequentiecurve van opgave (a). Dit ziet er als volgt uit:



Naar schatting zullen er dus 79 auto's zijn met een profiel van hoogstens

5,00mm dikte.

Hoofdstuk 2

Opdracht 2.m1: De mediaan van een even aantal waarnemingen is gelijk aan ...

(a) de gemiddelde uitkomst.

(b) de hoogste minus de laagste waarneming gedeeld door twee.

(c) het gemiddelde van de middelste twee waarnemingen na ordening.

(d) het getal dat het meeste blijkt te zijn waargenomen.

Uitwerking

Het juiste antwoord is (c)

Opdracht 2.m2: Het gemiddelde maandinkomen van een groep van 15 net-afgestudeerde hbo'ers is 1.850 euro per maand. Er blijkt nog een zestiende persoon tot die groep te behoren. Zijn maandinkomen bedraagt 2.170 euro. Van de totale groep van zestien personen is het gemiddelde inkomen dus

(a) 2.010 euro

(b) 1.870 euro

(c) 2.150 euro

(d) niet te berekenen

Uitwerking

Het juiste antwoord is (b)

Opdracht 2.m3: Bij een frequentieverdeling van de gewichten van schapen wordt gewerkt met klassen van 10kg breed. De gewichten worden op gehele kilogrammen afgerond. De klasse 20 -< 30 heeft dus als werkelijke bovengrens ...

(a) 10 kg

(b) 29,5 kg

(c) 30 kg

(d) 30,5 kg

Uitwerking

Het juiste antwoord is (b)

Opdracht 2.m4: De variantie is ...

(a) slecht hanteerbaar als maatstaf voor spreiding.

(b) de wortel uit de standaarddeviatie.

(c) een grootheid die wordt opgebouwd uit gekwadrateerde afwijkingen van waarneming ten opzicht van het gemiddelde.

(d) het kwadraat van de gemiddelde afwijking.

Uitwerking

Het juiste antwoord is (c)

Opdracht 2.m5: Voor een steekproef van 500 werknemers van een groot concern is in 2001 de verdeling van de jaarinkomens onderzocht. Hierbij bleek dat de standaarddeviatie van de inkomens 7.500 guldens bedroeg. Ter wille van een internationale publicatie moeten de gegevens worden uitgedrukt in euro's (1 euro = afgerond 2,20 gulden). De standaarddeviatie gemeten in euro's bedraagt dus ...

(a) 16.500

(b) 36.300

(c) 3.409

(d) 1.100

Uitwerking

Het juiste antwoord is (c)

Opdracht 2.m6: In een boxplot kan men diverse kenmerken van een frequentieverdeling aflezen. Wat kan men hierin echter *niet* aflezen?

(a) de mediaan

(b) het rekenkundig gemiddelde

(c) het eerste kwartiel

(d) de hoogste waarneming

Uitwerking

Het juiste antwoord is (b)

Opdracht 2.9: Een groep van 10 studenten behaalde de volgende scores bij een examen:

$$45, 60, 82, 32, 25, 75, 65, 66, 70, 80$$

(a) Bereken het rekenkundig gemiddelde, de mediaan en de standaarddeviatie voor deze gegevens.

Uitwerking

Het rekenkundig gemiddelde wordt berekend met:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{45 + 60 + \ldots + 80}{10} = \frac{600}{10} = 60$$

De mediaan wordt gevonden door eerst de scores te sorteren:

$$25, 32, 45, 60, 65, 66, 70, 75, 80, 82$$

Omdat er een even aantal scores is, bekijken we het gemiddelde van de middelste twee waarden:

$$\text{Mediaan} = \frac{x_5 + x_6}{2} = \frac{65 + 66}{2} = 65.5$$

Omdat we te maken hebben met een steekproef, wordt de variantie berekend met:

$$\begin{aligned} s^2 &= \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2 \\ &= \frac{1}{10-1} \cdot \left((45-60)^2 + (60-60)^2 + \ldots + (80-60)^2\right) \\ &= \frac{1}{9} \cdot 3504 \\ &= \frac{3504}{9} \end{aligned}$$

De standaarddeviatie vinden we door de wortel uit de variantie te nemen:

$$s = \sqrt{s^2} = \sqrt{\frac{3750}{9}} \approx 19.7315$$

(b) Er wordt besloten alle studenten 10 punten extra te geven. Wat is het gevolg van deze maatregel voor de bij vraag (a) berekende maatstaven.

Uitwerking

Rekenkundig gemiddelde: als alle scores met 10 worden verhoogd, dan wordt het rekenkundig gemiddelde ook met 10 verhoogd.
Mediaan: als alle scores met 10 worden verhoogd, dan geldt dit specifiek ook voor de middelste twee scores (merk op dat de volgorde gelijk blijft!). De mediaan is het gemiddelde van de twee middelste scores, dus de mediaan wordt ook verhoogd met 10
Standaarddeviatie: als alle scores met 10 worden verhoogd, dan blijft de spreiding van de scores gelijk, alleen met 10 naar rechts verschoven. De standaarddeviatie blijft in dat geval dus hetzelfde.

(c) In plaats van 10 punten voor iedereen erbij (vraag b) wordt besloten de kandidaten allemaal een 10% hogere score te geven dan zij oorspronkelijk gehaald hebben. Wat is de invloed van deze maatregel op de onder a berekende grootheden?

Uitwerking

Rekenkundig gemiddelde: als alle scores met 10% worden verhoogd, dan wordt het rekenkundig gemiddelde ook met 10% verhoogd.
Mediaan: als alle scores met 10% worden verhoogd, dan geldt dit specifiek ook voor de middelste twee scores (merk op dat de volgorde gelijk blijft!). De mediaan is het gemiddelde van de twee middelste scores, dus de mediaan wordt ook verhoogd met 10%
Standaarddeviatie: als alle scores met 10% worden verhoogd, dan wordt de spreiding van de scores groter. Merk op dat voor elke term $(x_i - \bar{x})^2$ in de formule van de variantie in deze nieuwe situatie moet worden vervangen door

$$(1.1 \cdot x_i - 1.1 \cdot \bar{x})^2 = \left(1.1 \cdot (x_i - \bar{x})^2\right) = 1.1^2 \cdot (x_i - \bar{x})^2 = 1.21 \cdot (x_i - \bar{x})^2$$

De variantie stijgt dus met een factor 1.21, en de standaarddeviatie met een factor $\sqrt{1.21} = 1.1$ (10%).

Opdracht 2.10: In een gemeente is bijgehouden hoe de prijsontwikkeling is van woningen. Hierbij wordt een onderscheid gemaakt naar het type woning. De volgende resultaten zijn verzameld over het afgelopen jaar. Prijsverandering geeft het verschil aan tussen huidige verkoopprijzen en de prijzen van een jaar geleden.

Type woning	Aantal verkocht	Aantal aanwezig	Prijsverandering
Villa	14	280	$+5\%$
Rijtjeshuis	130	1.600	$+2,4\%$
Appartement	48	450	$+1,6\%$
Twee onder een kap	26	500	$+7,2\%$
Serviceflat	28	140	$-2,5\%$

(a) Bereken voor deze gemeente de gemiddelde prijsverandering van de verkochte huizen.

Uitwerking

De gemiddelde prijsverandering van de verkochte huizen wordt berekend als een gewogen gemiddelde:

$$\begin{aligned}\bar{x} &= \frac{14 \cdot 5 + 130 \cdot 2,4 + 48 \cdot 1,6 + 26 \cdot 7,2 + 28 \cdot (-2,5)}{14 + 130 + 48 + 26 + 28} \\ &= \frac{576}{246} \\ &\approx 2,3415\end{aligned}$$

(b) Bereken op basis van de gegevens de gemiddelde waardeverandering van het totale woningbestand op basis van weging met het aantal aanwezige huizen.

Uitwerking

De gemiddelde waardeverandering van het totale woningbestand wordt berekend als een gewogen gemiddelde:

$$\begin{aligned}\bar{x} &= \frac{280 \cdot 5 + 1600 \cdot 2,4 + 450 \cdot 1,6 + 500 \cdot 7,2 + 140 \cdot (-2,5)}{280 + 1600 + 450 + 500 + 140} \\ &= \frac{9210}{2970} \\ &\approx 3,1010\end{aligned}$$

Opdracht 2.20: Bij een onderzoek in opdracht van Horeca Nederland werd onderzocht hoeveel geld mensen uitgeven aan eten buiten de deur. Twee groepen werden onderscheiden, namelijk studenten die zelfstandig wonen en afgestudeerden met minstens twee jaar werkervaring. De resultaten waren als volgt:

Uitgaven voor 30 studenten (euro per maand)
$72, 51, 146, 30, 28, 88, 92, 47, 52, 68, 80, 34, 28, 105, 76, 93, 55, 40, 62, 37, 88, 30, 122, 46, 35, 29, 77, 40, 71, 57$
Uitgaven voor 40 afgestudeerden (euro per maand)
$88, 130, 255, 56, 0, 38, 167, 188, 132, 147, 78, 80, 40, 170, 280, 46, 174, 182, 75, 89, 103, 230, 380, 57, 55, 90, 96, 102, 78, 69, 53, 160, 195, 245, 60, 94, 145, 115, 225, 71$

(a) Bereken voor beide groepen het gemiddelde en de mediaan

Uitwerking

Het gemiddelde wordt berekend als:

$$\bar{x} = \frac{\sum x_i}{n}$$

Voor de groep studenten is het gemiddelde gelijk aan:

$$\bar{x}_{\text{studenten}} = \frac{72 + 51 + \cdots + 57}{30} \approx 62,6333.$$

Voor de groep afgestudeerden is het gemiddelde gelijk aan:

$$\bar{x}_{\text{afgestudeerden}} = \frac{88 + 130 + \cdots + 71}{40} \approx 125,9500.$$

Aangezien voor beide groepen geldt dat het aantal waarnemingen even is, bekijken we het gemiddelde van de twee middelste waarden. Voor de groep studenten is de gesorteerde lijst van waarnemingen gelijk aan:

$$28, 28, 29, 30, 30, 34, 35, 37, 40, 40, 46, 47, 51, 52, 55, 57, 62, 68, 71, 72, 76, 77, 80$$
$$88, 88, 92, 93, 105, 122, 146$$

De mediaan is gelijk aan het gemiddelde van de twee middelste waarden, oftewel:

$$\text{Mediaan}_{\text{studenten}} = \frac{55 + 57}{2} = 56.$$

Voor de groep afgestudeerden is de gesorteerde lijst van waarnemingen gelijk aan:

$$0, 38, 40, 46, 53, 55, 56, 57, 60, 69, 71, 75, 78, 78, 80, 88, 89, 90, 94, 96, 102, 103, 115, 130, 132, 145, 147, 160, 167, 170, 174, 182, 188, 195, 225, 230, 245, 255, 280, 380$$

De mediaan is gelijk aan het gemiddelde van de twee middelste waarden, oftewel:

$$\text{Mediaan}_{\text{afgestudeerden}} = \frac{96 + 102}{2} = 99.$$

(b) Bereken voor beide groepen de standaarddeviatie.

Uitwerking

De variantie wordt berekend als

$$s^2 = \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2$$

Voor de groep studenten is de variantie gelijk aan

$$\begin{aligned} s^2_{\text{studenten}} &= \frac{1}{30-1} \cdot \left((72 - 62,6333)^2 + (51 - 62,6333)^2 + \ldots + (57 - 62,6333)^2\right) \\ &\approx 896,5161. \end{aligned}$$

De standaarddeviatie is de wortel uit de variantie, oftewel

$$s_{\text{studenten}} = \sqrt{s^2_{\text{studenten}}} = \sqrt{896,5161} \approx 29,9419.$$

Voor de groep afgestudeerden is de variantie gelijk aan

$$\begin{aligned} s^2_{\text{afgestudeerden}} &= \frac{1}{40-1} \cdot \big((88 - 125,9500)^2 + (130 - 125,9500)^2 + \ldots \\ &\quad + (71 - 125,9500)^2\big) \\ &\approx 6255,9974. \end{aligned}$$

De standaarddeviatie is de wortel uit de variantie, oftewel

$$s_{\text{afgestudeerden}} = \sqrt{s^2_{\text{afgestudeerden}}} = \sqrt{6255,9974} \approx 79,0949.$$

(c) Bepaal voor beide groepen een boxplot. Zijn er uitbijters?

Uitwerking

De interkwartielafstand (interquartile range) IQR wordt berekend als:

$$IQR = Q_3 - Q_1$$

Voor studenten:

$$Q_1 = 37, \quad Q_3 = 80 \Rightarrow IQR_{\text{studenten}} = 43$$

Voor afgestudeerden:

$$Q_1 = 70, \quad Q_3 = 172 \Rightarrow IQR_{\text{afgestudeerden}} = 102$$

Uitbijters worden gedefinieerd als waarden buiten:

$$[Q_1 - 1,5 \times IQR, Q_3 + 1,5 \times IQR]$$

Voor studenten:

$$[37 - 1,5 \times 43, 80 + 1,5 \times 43] = [-27,5, 144,5] \quad \text{Geen uitbijters}$$

Voor afgestudeerden:

$$[70 - 1,5 \times 102, 172 + 1,5 \times 102] = [-83, 325] \quad \text{Uitbijters: 380}$$

(d) Schrijf een korte notitie waarin het verschil wordt toegelicht tussen beide groepen.

Uitwerking

Afgestudeerden geven gemiddeld significant meer geld uit aan eten buiten de deur dan studenten. Dit verschil kan worden verklaard door hun hogere inkomen, stabiliteit en veranderde eetgewoonten. Daarnaast is de spreiding bij afgestudeerden aanzienlijk groter, wat betekent dat sommige individuen zeer hoge uitgaven hebben. De boxplot-analyse bevestigt dat er een uitbijter in de groep van afgestudeerden aanwezig is (380 euro), wat mogelijk iemand is met uitzonderlijke uitgavenpatronen.

Week 2: discrete kansvariabelen

Hoofdstuk 3

Opdracht 3.11: Bij een exameninstituut voor rijexamens zijn drie examinatoren in dienst: Dirk, Judith en Ilse. Van alle kandidaten doet 40% examen bij Dirk, 40% doet examen bij Judith en 20% bij Ilse. Bekend is dat het slagingspercentage voor de drie examinatoren verschilt. Dat is 40% bij Dirk, 60% bij Judith en 70% bij Ilse.

(a) Kandidaat Peter van der Meer gaat rijexamen doen bij een van de drie examinatoren. Hoe groot is de kans dat hij slaagt?

Uitwerking

Definieer de gebeurtenissen $S = \{\text{Peter slaagt}\}$, $D = \{\text{Peter doet examen bij Dirk}\}$, $J = \{\text{Peter doet examen bij Judith}\}$ en $I = \{\text{Peter doet examen bij Ilse}\}$. Uit de vraag volgt dan dat $P(D) = 0,4, P(J) = 0,4$ en $P(I) = 0,2$, evenals de conditionele kansen $P(S|D) = 0,4$, $P(S|J) = 0,6$ en $P(S|I) = 0,7$. In dat geval is de kans dat Peter slaagt gelijk aan

$$\begin{aligned} P(S) &= P(S|D) \cdot P(D) + P(S|J) \cdot P(J) + P(S|I) \cdot P(I) \\ &= 0,4 \cdot 0,4 + 0,6 \cdot 0,4 + 0,7 \cdot 0,2 \\ &= 0,54 \end{aligned}$$

Peter slaagt dus met 54% kans voor zijn rijexamen.

(b) Kandidaat Peter blijkt geslaagd te zijn. Hoe groot is de kans dat hij examen heeft afgelegd bij Ilse?

Uitwerking

In dit geval willen we de kans $P(I|S)$ bepalen. Uit de regel van Bayes volgt dan:

$$P(I|S) = \frac{P(S|I) \cdot P(I)}{P(S)} = \frac{0,7 \cdot 0,2}{0,54} \approx 0,2593$$

Met 25,93% kans heeft Peter zijn examen afgelegd bij Ilse, gegeven dat hij is geslaagd.

Sommige gezakte kandidaten dienen een klacht in over het examen. Op basis van ervaring weten we dat 10% van de gezakten van examinator Dirk een klacht indient. Bij Judith is dat 6% en bij Ilse is dat 3%.

(d) Er komt een klacht binnen over het examen zonder vermelding van de naam van de examinator. Hoe groot is de kans dat de klacht over Dirk gaat?

Uitwerking

De relatieve frequenties van combinaties van examinatoren en al dan niet geslaagd zijn zijn als volgt:

Examinator	wel geslaagd	niet geslaagd	**Totaal**
Dirk	40% van 40% = 16%	60% van 40% = 24%	40%
Judith	60% van 40% = 24%	40% van 40% = 16%	40%
Ilse	70% van 20% = 14%	30% van 20% = 6%	20%
Totaal	54%	46%	100%

Het percentage deelnemers dat zakt en een klacht over Dirk instuurt is dus 10% van 24% = 2,4%. Daarnaast is het percentage deelnemers dat zake en een klacht instuurt (over alle examinatoren) gelijk aan

$$0,1 \cdot 0,24 + 0,06 \cdot 0,16 + 0,03 \cdot 0,06 = 0,0354 \rightarrow 3,54\%$$

De kans dat een klacht over Dirk gaat is dus gelijk aan $\frac{2{,}4}{3{,}54} \approx 0,68 \rightarrow 68\%$.

Opdracht 3.24: In een fabriek wordt de kwaliteit gecontroleerd van de uitgaande producten. De employé die de controle verrichtte, blijkt 1% van alle goede producten af te keuren en verder keurt hij 5% van alle slechte producten goed. De totale productie bestaat voor 90% uit goede producten.

(a) Bereken de kans dat een willekeurig product goed is en wordt goedgekeurd.

Uitwerking

We zetten de relatieve frequenties even in een tabel:

	goedgekeurd	afgekeurd	**Totaal**
goed product	99% van 90% = 89,1%	1% van 90% = 0,9%	90%
slecht product	5% van 10% = 0,5%	95% van 10% = 9,5%	10%
Totaal	89,6%	10,4%	100%

Aflezen uit de cel voor “goed productën “goedgekeurd” geeft een kans van $89,1\% \rightarrow 0,891$.

(b) Hoe groot is de kans dat de controleur voor een willekeurig product de verkeerde beslissing neemt?

Uitwerking

Deze kans kunnen we opnieuw aflezen uit de tabel, namelijk de som van de cellen ("goed product", "afgekeurd") en ("slecht product", "goedgekeurd"). Dit geeft een kans van $0,9\% + 0,5\% = 1,4\% \to 0,014$.

(c) Hoe groot is het percentage goedgekeurde producten dat de fabriek verlaat?

Uitwerking

Deze kans kunnen we opnieuw aflezen uit de tabel, namelijk het kolomtotaal voor de kolom "goedgekeurd". Hieruit volgt dat 89,6% van de producten wordt goedgekeurd.

Hoofdstuk 4

Opdracht 4.m1: We besluiten driemaal een muntstuk op te gooien en we tellen het aantal malen dat de uitkomst "kop" verschijnt. Dit aantal malen "kop" ...

(a) is geen kansvariabele, omdat de uitkomsten "kop" en "munt" geen getallen zijn.

(b) is een continue kansvariabele, omdat dit spelletje eindeloos vaak kan worden herhaald.

(c) laat altijd een waarde tussen 1 en 3 zien.

(d) is een discrete variabele, omdat slechts eindig veel uitkomsten hiervoor mogelijk zijn.

Uitwerking

Het juiste antwoord is (d)

Opdracht 4.m2: Voor de populatie Nederlandse hbo-studenten is vastgesteld dat het aantal verschillende statistiekboeken waaruit zij wel eens hebben gestudeerd, kan worden beschreven door een kansvariabele X die in de volgende tabel is weergegeven

Aantal boeken (k)	0	1	2	3	4
$P(X = k)$	$0,15$	$0,40$	$0,30$	$0,10$	$0,05$

De kans dat een willekeurige student heeft gestudeerd uit minstens twee statistiekboeken is daardoor gelijk aan ...

(a) $0,30$

(b) $0,85$

(c) $0,45$

(d) $0,15$

Uitwerking

Het juiste antwoord is (c)

Opdracht 4.m3: Een kansvariabele X kan n verschillende waarden aannemen, namelijk $k_1, k_2, \ldots, k_n$. Voor de kansfunctie van de variabele X moet dan gelden:

$$\sum f(k) = \sum P(X = k) = \ldots$$

(a) 0

(b) 1

(c) $\frac{1}{n}$

(d) n

Uitwerking

Het juiste antwoord is (b)

Opdracht 4.m4: Van de kansvariabele X is de verdelingsfunctie als volgt:

Uitkomst k	10	11	12	13	14	15	16
$F(k)$	$0,16$	$0,28$	$0,46$	$0,66$	$0,84$	$0,94$	$1,00$

Bereken $P(X = 14)$. Wat is de uitkomst?

(a) $0,18$

(b) $0,84$

(c) $0,10$

(d) $0,16$

Uitwerking

Het juiste antwoord is (a)

Opdracht 4.m5: Het aantal koelkasten dat wekelijks wordt verkocht op de afdeling "witgoed" van een supermarkt, kan worden weergegeven door de kansvariabele X waarvan de kansfunctie is weergegeven in de volgende tabel:

Aantal uitkomsten k	0	1	2	3	4
$P(X = k)$	$0,25$	$0,40$	$0,20$	$0,10$	$0,05$

De standaarddeviatie van de variabele X bedraagt dus ...

(a) $1,3$ koelkasten

(b) $1,1$ koelkasten

(c) $0,93$ koelkasten

(d) $1,21$ koelkasten

Uitwerking

Het juiste antwoord is (b)

Opdracht 4.3: Gegeven is dat een kansvariabele X alleen de waarden 10, 20, 30 en 40 kan aannemen.

(a) Verifieer voor de volgende vier gevallen of de geformuleerde waarden van $f(k)$ zodanig zijn dat f als een kansfunctie kan worden beschouwd.

(A) $f(10) = 0,30$; $f(20) = 0,40$; $f(30) = 0,10$ en $f(40) = 0,10$.

(B) $f(10) = 0,50$; $f(20) = 0,30$; $f(30) = 0,30$ en $f(40) = 0,10$.

(C) $f(10) = 0,05$; $f(20) = 0,05$; $f(30) = 0,10$ en $f(40) = 0,80$.

(D) $f(10) = 0,00$; $f(20) = 0,00$; $f(30) = 1,00$ en $f(40) = 0,00$.

Uitwerking

Om een kansfunctie te zijn moeten alle waarden voor $f(k)$ kansen zijn, dus getallen tussen 0 en 1. Verder moet de som van alle kansen gelijk zijn aan 1. Dit is alleen het geval voor C en D.

(b) Kansen kunnen ook worden weergegeven door de cumulatieve kansfunctie (verdelingsfunctie) $F(k)$. Geef voor de kansfunctie zoals beschreven bij punt C de verstrekte informatie weer door middel van $F(k)$.

Uitwerking

We beschreven F, de cumulatieve verdelingsfunctie (CDF), door middel van een tabel:

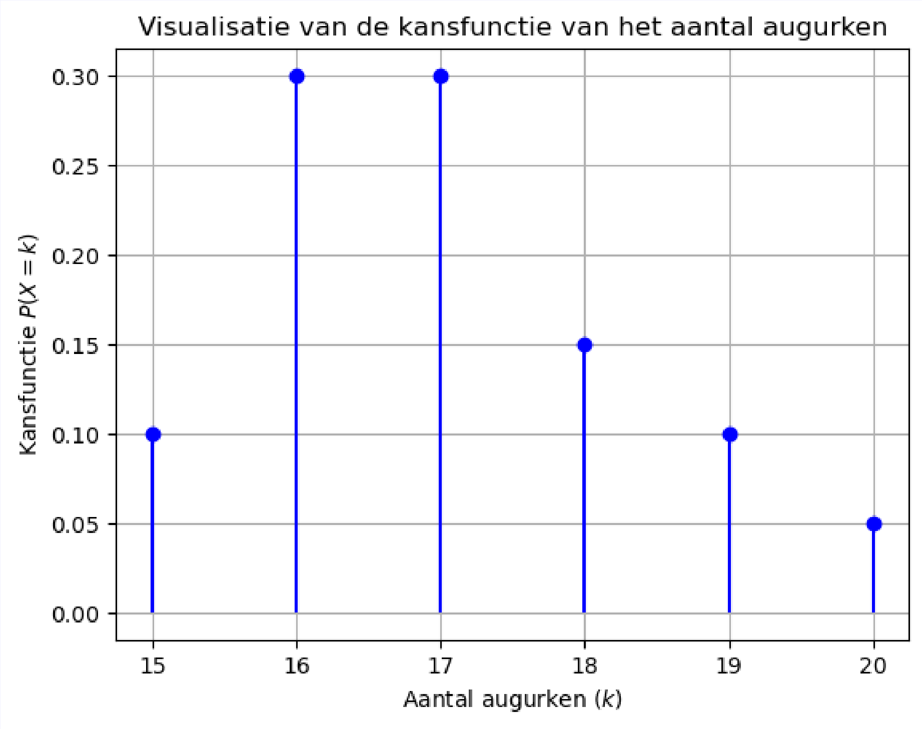
Uitkomst k	10	20	30	40
$F(k) = P(X \leq k)$	$0,05$	$0,10$	$0,20$	$1,00$

Opdracht 4.4: Het aantal augurken dat in een halveliterpot zit, blijkt enigszins te variëren. Uitvoerig onderzoek heeft aangetoond dat dit aantal tussen 15 en 20 ligt, waarbij de volgende kansen van toepassing zijn:

Aantal augurken k	15	16	17	18	19	20
$P(X = k)$	$0,10$	$0,30$	$0,30$	$0,15$	$0,10$	$0,05$

(a) Teken de kansfunctie

Uitwerking



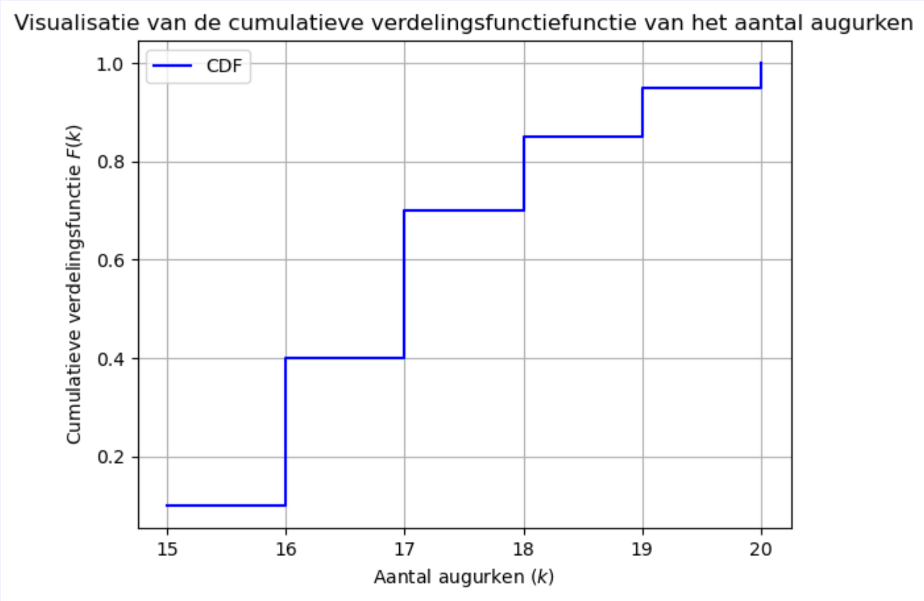
(b) Bereken en teken de cumulatieve kansfunctie.

Uitwerking

We beschreven F, de cumulatieve verdelingsfunctie (CDF), door middel van een tabel:

Aantal augurken k	15	16	17	18	19	20
$F(k) = P(X \leq k)$	$0,10$	$0,40$	$0,70$	$0,85$	$0,95$	$1,00$

In een grafiekvorm ziet dit er als volgt uit:



(c) Hoe groot is de kans om in een willekeurige pot meer dan 16 maar minder dan 20 augurken aan te treffen?

Uitwerking

We willen de kans $P(16 < X < 20)$ bepalen. Omdat we met een discrete verdeling werken, kunnen we die als volgt berekenen:

$$P(16 < X < 20) = P(17 \leq X \leq 19) = F(19) - F(16) = 0,95 - 0,40 = 0,55.$$

(d) Er worden twee willekeurige potten augurken geselecteerd. Hoe groot is de kans dat beide potten meer dan 18 augurken bevatten?

Uitwerking

Als er twee willekeurige potten geselecteerd zijn, dan geldt dat beide met kans $P(X > 18) = 1 - P(X \leq 18) = 1 - 0,85 = 0,15$ meer dan 18 augurken bevatten. Omdat de potten onafhankelijk van elkaar gevuld zijn, is de totale kans gelijk aan $0,15 \cdot 0,15 = 0,0225$.

(e) Hoe groot is de kans dat twee willekeurige potten augurken in totaal 38 augurken bevatten.

Uitwerking

Laat nu X_1 het aantal augurken zijn in pot 1, en X_2 het aantal augurken in pot 2. We willen de kans $P(X_1 + X_2 = 38)$ bepalen. Dit doen we als volgt:

$$\begin{aligned} P(X_1 + X_2 = 38) &= P(X_1 = 18, X_2 = 20) + P(X_1 = 19, X_2 = 19) \\ &\quad + P(X_1 = 20, X_2 = 18) \\ &= 0,15 \cdot 0,05 + 0,10 \cdot 0,10 + 0,05 \cdot 0,15 \\ &= 0,025. \end{aligned}$$

Met $2,5\%$ kans is het totaal aantal augurken in twee willekeurige potten samen gelijk aan 38.

Opdracht 4.6: Bij een schoolreisje worden vier bussen gebruikt met elk een eigen chauffeur. Van de in totaal 130 leerlingen hebben deze bussen respectievelijk 45, 35, 30 en 20 leerlingen vervoerd.

(a) Op basis van toeval kiest men één van de vier chauffeurs. Men vraagt aan de chauffeur hoe groot X is, het aantal leerlingen in zijn bus. Bereken $E[X]$.

Uitwerking

X is het aantal leerlingen in de bus van een willekeurige chauffeur. De onderliggende kansverdeling is dan

Aantal leerlingen k	20	30	35	45
$P(X = k)$	$0,25$	$0,25$	$0,25$	$0,25$

Het is immers zo dat ieder van de vier chauffeurs gekozen wordt met kans $\frac{1}{4}$. De verwachtingswaarde $E[X]$ is dan

$$E[X] = \sum_k k \cdot P(X = k) = 20 \cdot 0,25 + 30 \cdot 0,25 + 35 \cdot 0,25 + 45 \cdot 0,25 = 32,5.$$

(b) Op basis van toeval kiest men één van de 130 leerlingen en vraagt aan hem of haar hoeveel leerlingen Y er in de bus zaten waarmee deze leerling werd vervoerd. Bereken $E[Y]$.

Uitwerking

X is het aantal leerlingen in de bus van een willekeurige leerling. Het totaal aantal leerlingen is $20+30+35+45 = 130$. De onderliggende kansverdeling is dan

Aantal leerlingen k	20	30	35	45
$P(Y = k)$	$\frac{20}{130}$	$\frac{30}{130}$	$\frac{35}{130}$	$\frac{45}{130}$

Het is immers zo dat de kans op een uitkomst evenredig is met het aantal leerlingen dat in een bus met een aantal leerlingen zit. De verwachtingswaarde $E[Y]$ is dan

$$E[Y] = \sum_k k \cdot P(Y = k) = 20 \cdot \frac{20}{130} + 30 \cdot \frac{30}{130} + 35 \cdot \frac{35}{130} + 45 \cdot \frac{45}{130} = 35.$$

Opdracht 4.7: Het aantal personen dat in een willekeurig uur opbelt naar de klantenservice van een groot bedrijf, wordt beschreven door de kansvariabele X, waarvan de gegevens zijn geplaatst in de volgende tabel:

Aantal k	0	1	2	3	4	5
$f(k) = P(X = k)$	$0,20$	$0,30$	$0,20$	$0,15$	$0,10$	$0,05$

(a) Hoe groot is $P(X \geq 3)$ en $P(X \leq 3)$

Uitwerking

We berekenen deze kansen als volgt:

$$\begin{aligned}
P(X \geq 3) &= P(X = 3) + P(X = 4) + P(X = 5) \\
&= 0,15 + 0,10 + 0,05 = 0,30 \\
P(X \leq 3) &= 1 - P(X \geq 4) \\
&= 1 - (P(X = 4) + P(X = 5)) \\
&= 1 - (0,10 + 0,05) \\
&= 1 - 0,15 = 0,85
\end{aligned}$$

(b) Bereken de verwachtingswaarde van X.

Uitwerking

We berekenen de verwachtingswaarde van X als volgt:

$$\begin{aligned}
E[X] &= \sum_k k \cdot P(X = k) \\
&= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + \ldots + 5 \cdot P(X = 5) \\
&= 0 \cdot 0,20 + 1 \cdot 0,30 + \ldots + 5 \cdot 0,05 \\
&= 1,8
\end{aligned}$$

(c) Bereken de standaarddeviatie van X.

Uitwerking

We berekenen de standaarddeviatie van X door eerst de variantie te berekene:

$$\begin{aligned}
Var(X) &= \sum_k (k - E[X])^2 \cdot P(X = k) \\
&= (0 - 1,8)^2 \cdot P(X = 0) + (1 - 1,8)^2 \cdot P(X = 1) + \ldots \\
&\quad + (5 - 1,8)^2 \cdot P(X = 5) \\
&= (0 - 1,8)^2 \cdot 0,20 + (1 - 1,8)^2 \cdot 0,30 + \ldots + (5 - 1,8)^2 \cdot 0,05 \\
&= 2,06
\end{aligned}$$

De standaarddeviatie vinden we dan door de wortel uit de variantie te nemen:

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{2,06} \approx 1,4353$$

Opdracht 4.9: Een huiseigenaar in Amsterdam besluit drie kamers te huur aan te bieden via Airbnb. Het dagelijks aantal verhuurde kamers op een willekeurige dag kan worden weergegeven door de kansvariabele X, zoals weergegeven in de tabel:

Aantal boekingen	$P(X = k)$
0	$0,50$
1	$0,25$
2	$0,15$
3	$0,10$
Totaal	$1,00$

(a) Hoe groot is de kans dat op een willekeurige dag minstens twee kamers zijn verhuurd.

Uitwerking

We berekenen deze kans als volgt:

$$\begin{aligned}
P(X \geq 2) &= P(X = 2) + P(X = 3) \\
&= 0,15 + 0,10 \\
&= 0,25
\end{aligned}$$

(b) De aantallen boekingen op twee opeenvolgende dagen kan men beschouwen als onderling onafhankelijke kansvariabelen. Maak een kanstabel voor de variabele X_{som} waarmee het totaal aantal boeking per twee dagen wordt aangegeven. Hoe groot is de kans dat in twee dagen meer dan 3 boekingen plaatsvinden?

Uitwerking

De kanstabel voor de somvariabele X_{som} ziet er als volgt uit:

Aantal boekingen	$P(X_{\text{som}} = k)$
0	$0,25$
1	$0,25$
2	$0,2125$
3	$0,175$
4	$0,0725$
5	$0,03$
6	$0,01$
Totaal	$1,00$

De kans dat er in twee dagen meer dan 3 boekingen plaatsvinden is dus gelijk aangeeft

$$\begin{aligned} P(X_{\text{som}} > 3) &= P(X_{\text{som}} = 4) + P(X_{\text{som}} = 5) + P(X_{\text{som}} = 6) \\ &= 0,0725 + 0,03 + 0,01 \\ &= 0,1125 \end{aligned}$$

Opdracht 4.12: Een doe-het-zelfzaak verhuurt apparaten waarmee oud behang van een muur kan worden afgestoomd. De vraag per dag naar dergelijke apparaten kan worden weergegeven door een kansvariabele X die de volgende waarden aanneemt:

Aantal gevraagde machines (k)	0	1	2	3
$P(X = k)$	$0,20$	$0,35$	$0,25$	$0,20$

(a) Bereken de verwachtingswaarde van het aantal gevraagde machines.

Uitwerking

We berekenen deze verwachtingswaarde van het aantal gevraagde machines X als volgt:

$$\begin{aligned} E[X] &= \sum_k k \cdot P(X = k) \\ &= 0 \cdot P(X = 0) + \ldots + 3 \cdot P(X = 3) \\ &= 0 \cdot 0,20 + 1 \cdot 0,35 + 2 \cdot 0,25 + 3 \cdot 0,20 \\ &= 1,45 \end{aligned}$$

(b) De bedrijfsleiding vraagt zich af hoeveel machines in voorraad moeten worden gehouden om aan de (onzekere) vraag te kunnen voldoen. Het beschikbaar hebben van zo'n apparaat kost 4 euro per dag, terwijl de huurprijs 6 euro per dag bedraagt. Bereken de opbrengst per dag Y indien twee apparaten beschikbaar zijn bij de verschillende waarden van X. Doe dit ook bij drie beschikbare apparaten. Bereken voor beide gevallen de verwachtingswaarde van Y. Welke beslissing levert de hoogste verwachtingswaarde.

Uitwerking

Indien er twee apparaten beschikbaar zijn, beschrijft de volgende tabel de opbrengsten per dag. De opbrengsten zijn berekend op basis van opbrengst is huurinkomsten minus beschikbaarheidskosten.

Aantal gevraagde machines (k)	0	1	2	3
Opbrengst ℓ	-8	-2	4	4
$P(X=k)$	$0,20$	$0,35$	$0,25$	$0,20$

De verwachte opbrengst bij twee beschikbare apparaten is dus

$$\begin{aligned}
E[Y] &= \sum_{\ell} \ell \cdot P(X=k) \\
&= -8 \cdot P(X=0) + \ldots + 4 \cdot P(X=3) \\
&= -8 \cdot 0,20 + -2 \cdot 0,35 + 4 \cdot 0,25 + 4 \cdot 0,20 \\
&= -0,5
\end{aligned}$$

Indien er drie apparaten beschikbaar zijn, beschrijft de volgende tabel de opbrengsten per dag.

Aantal gevraagde machines (k)	0	1	2	3
Opbrengst ℓ	-12	-6	0	6
$P(X=k)$	$0,20$	$0,35$	$0,25$	$0,20$

De verwachte opbrengst bij twee beschikbare apparaten is dus

$$\begin{aligned}
E[Y] &= \sum_{\ell} \ell \cdot P(X=k) \\
&= -12 \cdot P(X=0) + \ldots + 6 \cdot P(X=3) \\
&= -12 \cdot 0,20 + -6 \cdot 0,35 + 0 \cdot 0,25 + 6 \cdot 0,20 \\
&= -3,3
\end{aligned}$$

De beslissing om twee apparaten beschikbaar te hebben levert dus naar verwachting de hoogste opbrengst op.

Opdracht 4.13: In een bedrijf staat een aantal machines opgesteld die soms vanwege een storing moeten worden stilgelegd. Op grond van ervaring is vastgesteld dat het aantal storingen X per week in de kanstabel wordt beschreven.

Aantal storingen (k)	$P(X = k)$
0	$0,50$
1	$0,26$
2	$0,12$
3	$0,08$
4	$0,04$
Totaal	$1,00$

(a) Bereken het verwachte aantal storingen per week.

Uitwerking

Het verwachte aantal storingen per week berekenen we als volgt:

$$
\begin{aligned}
E[X] &= \sum_k k \cdot P(X = k) \\
&= 0 \cdot P(X = 0) + \ldots + 4 \cdot P(X = 4) \\
&= 0 \cdot 0,50 + 1 \cdot 0,26 + 2 \cdot 0,12 + 3 \cdot 0,08 + 4 \cdot 0,04 \\
&= 0,9
\end{aligned}
$$

(b) Bereken de variantie van het aantal storingen per week.

Uitwerking

Het variantie van het aantal storingen per week berekenen we als volgt:

$$
\begin{aligned}
Var(X) &= \sum_k (k - E[X])^2 \cdot P(X = k) \\
&= (0 - 0,9)^2 \cdot P(X = 0) + \ldots + (4 - 0,9)^2 \cdot P(X = 4) \\
&= (0 - 0,9)^2 \cdot 0,50 + \ldots + (4 - 0,9)^2 \cdot 0,04 \\
&= 0,81 \cdot 0,50 + \ldots + 9,61 \cdot 0,04 \\
&\approx 1,29
\end{aligned}
$$

(c) Bereken het verwachte aantal storingen per jaar (=50 weken).

Uitwerking

We kunnen het aantal storingen in 50 weken zien als de som X_{som} van 50 onderling onafhankelijke kansvariabelen X_1, X_2, $\ldots X_{50}$. Aangezien de verwachtingswaarde een lineaire functie is, geldt er dat:

$$
\begin{aligned}
E[X_{\text{som}}] = E[X_1 + X_2 + \ldots + X_{50}] &= E[X_1] + E[X_2] + \ldots + E[X_{50}] \\
&= 0,9 + 0,9 + \ldots + 0,9 \\
&= 50 \cdot 0,9 = 45
\end{aligned}
$$

(d) Een storing kost het bedrijf aan reparatie en gemiste productie €500 per geval. Bereken de verwachte storingskosten per jaar.

Uitwerking

De verwachte storingskosten per jaar berekenen we simpelweg door het verwachte aantal storingen per jaar te vermenigvuldigen met de kosten per storing, oftewel $45\cdot$ €$500 =$ €22500.

(e) Bereken de standaarddeviatie van het aantal storingen per jaar.

Uitwerking

Aangezien de wekelijkse aantallen storingen onderling onafhankelijke kansvariabelen zijn, geldt

$$\begin{aligned} Var(X_{\text{som}}) &= Var(X_1 + X_2 + \ldots + X_{50}) \\ &= Var(X_1) + Var(X_2) + \ldots + Var(X_{50}) \\ &= 1,29 + 1,29 + \ldots + 1,29 \\ &= 50 \cdot 1,29 \\ &= 64,5 \end{aligned}$$

De standaarddeviatie van het aantal storingen per jaar vinden we door de wortel hiervan te nemen:

$$\sigma(X_{\text{som}}) = \sqrt{Var(X_{\text{som}})} = \sqrt{64,5} \approx 8,03$$