Statistiek voor MBW / KW
Uitwerkingen huiswerkopgaven

Dr. ir. D.A.M.P. Blom & Dr. J.B.M. Melissen

2025

Week 9: schatten en betrouwbaarheid (deel 2)

Hoofdstuk 8

Opdracht 8.m4: Bij een schattingsprobleem moet men de t-verdeling gebruiken als voor de te onderzoeken kansvariabele geldt dat er een steekproef is genomen van:

(a) een normale verdeling met onbekende μ en onbekende variantie.

(b) een binomiale verdeling met onbekende succeskans π.

(c) een normale verdeling met gegeven standaarddeviatie en onbekende verwachtingswaarde.

(d) een gegeven normale verdeling met minder dan 30 vrijheidsgraden.

Uitwerking

De t-verdeling moet gebruikt worden in het geval dat de te onderzoeken kansvariabele normaal verdeeld is met onbekende verwachtingswaarde μ en onbekende standaarddeviatie σ. Het juiste antwoord is dus (a).

Opdracht 8.m6: Men wil een 99%-betrouwbaarheidsinterval voor μ berekenen voor een normale verdeling met een onbekende σ. Er worden acht waarnemingen gedaan. Het interval wordt dan berekend met ...

(a) $z = 2,58$

(b) $t = 3,499$

(c) $t = 2,998$

(d) $t = 2,306$

Uitwerking

Gegeven is dat we te maken hebben met een kansvariabele $X \sim N(\mu =?; \sigma =?)$ en een steekproefomvang $n = 8$. Omdat de gewenste betrouwbaarheid gelijk is aan 99%, maken we gebruik van $\alpha = 0,01$ en de t-verdeling met $df = n - 1 = 7$ vrijheidsgraden.

$$t = \mathrm{InvT}(opp = 1 - \alpha/2; df = n - 1) = \mathrm{InvT}(opp = 0,995; df = 7) \approx 3,4995.$$

Het juiste antwoord is dus (b).

Opdracht 8.2: Van een bepaalde oplossing worden tien even grote monsters genomen. Van elk monster bepaalt men het aantal milligram eiwit, om hiermee het eiwitgehalte (μ) van de gehele oplossing te kunnen schatten. De waargenomen uitkomsten zijn als volgt:

$36,5$	$45,2$	$40,5$	$42,0$	$37,7$	$39,4$	$41,6$	$38,8$	$39,0$	$43,3$

(a) Een zuivere schatter voor μ is het steekproefgemiddelde $\overline{x}$. Bereken $\overline{x}$.

Uitwerking

Het steekproefgemiddelde $\overline{x}$ berekenen we als volgt:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_{10}}{10} = \frac{36,5 + 45,2 + \ldots + 43,3}{10} = 40,4$$

Gemiddeld zit er $40,4$ mg eiwit in een monster.

(b) Hier kennen we σ niet, dus voordat er een schattingsinterval kan worden berekend voor μ, moet er een schatting van de variantie gemaakt worden. Bereken deze variantie.

Uitwerking

De steekproefvariantie s^2 berekenen we als volgt:

$$\begin{aligned} s^2 &= \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_{10} - \overline{x})^2}{10 - 1} \\ &= \frac{(36,5 - 40,4)^2 + (45,2 - 40,4)^2 + \ldots + (43,3 - 40,4)^2}{10 - 1} \\ &\approx 7,0533 \end{aligned}$$

Gemiddeld zit er $40,4$ mg eiwit in een monster.

(c) Bereken voor μ een betrouwbaarheidsinterval dat een betrouwbaarheid heeft van 95%.

Uitwerking

Noteer met X de kansvariabele die het aantal milligram eiwit in een willekeurig monster meet. Gegeven is dat $X \sim N(\mu = ?; \sigma = ?)$. Volgens de centrale limietstelling geldt dat het gemiddelde aantal milligram eiwit $\overline{X}$ in 10 willekeurige monsters normaal verdeeld is met verwachtingswaarde μ en een standaarddeviatie $\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{10}}$.
We hebben een geobserveerde steekproef van 10 zakken genomen met steekproefgemiddelde $\overline{x} = 40,4$ milligram en steekproefstandaardafwijking $s = \sqrt{s^2} = \sqrt{7,0533} \approx 2,6558$ milligram. Aangezien we een betrouwbaarheid van 95% willen, is $\alpha = 0,05$. Verder, omdat de standaardafwijking onbekend is en de steekproefomvang $n < 30$, moeten we de t-verdeling met $df = n - 1$ vrijheidsgraden gebruiken. In het bijzonder gebruiken we

$$t = \mathrm{InvT}(opp = 1 - \alpha/2; df = n - 1) = \mathrm{InvT}(opp = 0,975; df = 9) \approx 2,2622.$$

Het 95%-betrouwbaarheidsinterval voor μ is dan gelijk aan

$$\begin{aligned}
&[\overline{x} - t \cdot \frac{s}{\sqrt{n}}; \overline{x} + t \cdot \frac{s}{\sqrt{n}}] \\
&= [40,4 - 2,2622 \cdot \frac{2,6558}{\sqrt{10}}; 20142 + 2,2622 \cdot \frac{2,6558}{\sqrt{10}}] \\
&= [38,50; 42,30]
\end{aligned}$$

Met 95% betrouwbaarheid ligt de gemiddelde hoeveelheid eiwit in een monster tussen $38,50$ en $42,30$ milligram.

Opdracht 8.6: Bij de opleiding Logistiek Management wordt de vraag onderzocht hoeveel tijd per week door studenten wordt besteed aan bijbaantjes. Een steekproef van 18 studenten leverde als gemiddelde op 346 minuten per week en een standaarddeviatie van $s = 175$ minuten.

(a) Bereken een 95%-betrouwbaarheidsinterval voor μ: het gemiddelde aantal werkuren van de studentenpopulatie.

Uitwerking

Laat X de tijd meten die een willekeurige student per week besteedt aan bijbaantjes. Gegeven is dat $X \sim N(\mu =?; \sigma =?)$. Volgens de centrale limietstelling geldt dat de gemiddelde tijd $\overline{X}$ die 18 willekeurige studenten wekelijks aan bijbaantjes besteden normaal verdeeld is met verwachtingswaarde μ en een standaarddeviatie $\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{18}}$.
We hebben een geobserveerde steekproef van 18 studenten genomen met steekproefgemiddelde $\overline{x} = 346$ minuten en steekproefstandaardafwijking $s = 175$ minuten. Aangezien we een betrouwbaarheid van 95% willen, is $\alpha = 0,05$. Verder, omdat de standaardafwijking onbekend is en de steekproefomvang $n = 18 < 30$, moeten we de t-verdeling met $df = n - 1$ vrijheidsgraden gebruiken. In het bijzonder gebruiken we

$$t = \mathrm{InvT}(opp = 1 - \alpha/2; df = n - 1) = \mathrm{InvT}(opp = 0,975; df = 17) \approx 2,1098.$$

Het 95%-betrouwbaarheidsinterval voor μ is dan gelijk aan

$$\begin{aligned}
&[\overline{x} - t \cdot \frac{s}{\sqrt{n}}; \overline{x} + t \cdot \frac{s}{\sqrt{n}}] \\
&= [346 - 2,1098 \cdot \frac{175}{\sqrt{18}}; 346 + 2,1098 \cdot \frac{175}{\sqrt{18}}] \\
&= [258,97; 433,03]
\end{aligned}$$

Met 95% betrouwbaarheid ligt de gemiddelde tijd die studenten aan bijbaantjes besteden tussen 258 en 433 minuten per week.

(b) Nadere inspectie van de gegevens leverde dat drie studenten helemaal geen bijbaantje hebben, dus er zijn maar 15 werkenden. Bereken voor de groep werkenden het steekproefgemiddelde en de standaarddeviatie. *(Hint: doe dit door te manipuleren met de formule van s^2.)*

Uitwerking

De totale tijd die de 18 studenten bij elkaar aan bijbaantjes besteden is gelijk aan $18 \cdot 346 = 6228$ minuten. Aangezien slechts 15 van de 18 studenten een bijbaantje heeft, is het steekproefgemiddelde eigenlijk gelijk aan $\overline{x}_{\text{werkend}} = \frac{6228}{15} = 415,2$ minuten.
Voor de steekproefvariantie hebben we wat ingewikkeldere redeneringen nodig. Ten eerste werd de steekproefvariantie bepaald als

$$s^2 = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_{18} - \overline{x})^2}{18 - 1}$$

Stel nu dat $x_{16} = x_{17} = x_{18} = 0$, oftewel student $16, 17, 18$ zijn de drie niet-werkende studenten. De term $(x_i - \overline{x})^2$ kun je anders opschrijven als:

$$(x_i - \overline{x})^2 = x_i^2 - 2x_i\overline{x} + \overline{x}^2.$$

In andere woorden, we kunnen de formule van de steekproefvariantie s^2 omschrijven als:

$$\begin{aligned} s^2 &= \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_{18} - \overline{x})^2}{18 - 1} \\ &= \frac{(x_1^2 - 2\overline{x}x_1 + \overline{x}^2) + (x_2^2 - 2\overline{x}x_2 + \overline{x}^2)^2 + \ldots + (x_{18}^2 - 2\overline{x}x_{18} + \overline{x}^2)}{18 - 1} \\ &= \frac{(x_1^2 + x_2^2 + \ldots + x_{18}^2) - 2\overline{x}(x_1 + x_2 + \ldots + x_{18}) + 18 \cdot \overline{x}^2}{18 - 1} \end{aligned}$$

Merk op dat de som $x_1 + x_2 + \ldots + x_{18}$ gelijk is aan $18 \cdot \overline{x}$, dus de laatste formule kun je herschrijven tot

$$s^2 = \frac{(x_1^2 + x_2^2 + \ldots + x_{18}^2) - 18 \cdot \overline{x}^2}{18 - 1}$$

Gegeven is dat $\overline{x} = 346$ en $s^2 = 175^2 = 30625$. Hierdoor kunnen we $(x_1^2 + x_2^2 + \ldots + x_{18}^2)$, bepalen, namelijk:

$$\begin{aligned} x_1^2 + x_2^2 + \ldots + x_{18}^2 &= s^2 \cdot (18 - 1) + 18 \cdot \overline{x}^2 \\ &= 30625 \cdot 17 + 18 \cdot 346^2 \\ &= 2675513 \end{aligned}$$

We kunnen de steekproefvariantie s^2_{werkend} ook schrijven op basis van de 15 werkenden, namelijk

$$\begin{aligned} s^2_{\text{werkend}} &= \frac{(x_1^2 + x_2^2 + \ldots + x_{15}^2) - 15 \cdot \overline{x}^2_{\text{werkend}}}{15 - 1} \\ &= \frac{2675513 - 15 \cdot 415,2^2}{14} \\ &\approx 6403,3857 \end{aligned}$$

De steekproefstandaardafwijking s_{werkend} is de wortel van de steekproefvariantie, oftewel $s_{\text{werkend}} = \sqrt{s^2_{\text{werkend}}} = \sqrt{6403,3857} \approx 80,02$ minuten.

(c) Bereken het 95%-betrouwbaarheidsinterval voor μ van de populatie werkende studenten.

Uitwerking

De berekening kunnen we op eenzelfde manier uitvoeren als bij vraag **a**. Volgens de centrale limietstelling geldt dat de gemiddelde tijd $\overline{X}$ die 15 willekeurige studenten wekelijks aan bijbaantjes besteden normaal verdeeld is met verwachtingswaarde μ en een standaarddeviatie $\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{15}}$.
We hebben een geobserveerde steekproef van 15 studenten genomen met steekproefgemiddelde $\overline{x}_{\text{werkend}} = 415,2$ minuten en steekproefstandaardafwijking $s = 80,02$ minuten. Aangezien we weer een betrouwbaarheid van 95% willen, is $\alpha = 0,05$. Verder, omdat de standaardafwijking onbekend is en de steekproefomvang $n = 15 < 30$, moeten we de t-verdeling met $df = n - 1$ vrijheidsgraden gebruiken. In het bijzonder gebruiken we

$$t = \text{InvT}(opp = 1 - \alpha/2; df = n-1) = \text{InvT}(opp = 0,975; df = 14) \approx 2,1448.$$

Het 95%-betrouwbaarheidsinterval voor μ is dan gelijk aan

$$\begin{aligned} &[\overline{x} - t \cdot \frac{s}{\sqrt{n}}; \overline{x} + t \cdot \frac{s}{\sqrt{n}}] \\ &= [415,2 - 2,1448 \cdot \frac{80,02}{\sqrt{15}}; 415,2 + 2,1448 \cdot \frac{80,02}{\sqrt{15}}] \\ &= [370,89; 459,51] \end{aligned}$$

Met 95% betrouwbaarheid ligt de gemiddelde tijd die werkende studenten aan bijbaantjes besteden tussen $370,89$ en $459,51$ minuten per week.

Opdracht 8.7: In een aselecte steekproef van 400 kiesgerechtigde Nederlanders blijken 160 personen aanhanger van een bepaalde politieke partij te zijn.

Geef een 99%-betrouwbaarheidsinterval voor p: de fractie aanhangers in de totale populatie.

Uitwerking

Laat X het aantal kiesgerechtigde Nederlanders in een steekproef dat aanhanger is van een bepaalde politieke partij. Aangezien de steekproefomvang gelijk is aan 400 kiesgerechtigde Nederlanders, is X binomiaal verdeeld met $n = 400$ en nog onbekende succeskans p.
We bepalen een 99%-betrouwbaarheidsinterval voor p met behulp van de Clopper-Pearson methode. Omdat de gewenste betrouwbaarheid 99% is, geldt dat $\alpha = 0,01$.
De Clopper-Pearson methode werkt als volgt:

1. Bepaal de succeskans p_1 waarvoor geldt dat de linkeroverschrijdingskans van de uitkomst $k = 160$ gelijk is aan $\alpha/2$, oftewel $P(X \leq 160) = \alpha/2 = 0,005$. Voer hiervoor in het functiescherm van de grafische rekenmachine:

$$\begin{aligned} &y_1 = \text{binomcdf}(n = 400; p_1 = X; k = 160) \\ &y_2 = 0,005 \end{aligned}$$

De solver optie geeft een waarde van $p_1 \approx 0,4653$.

2. Bepaal de succeskans p_2 waarvoor geldt dat de rechteroverschrijdingskans van de uitkomst $k = 160$ gelijk is aan $\alpha/2$, oftewel $P(X \geq 160) = 1 - P(X \leq 159) = \alpha/2 = 0,005$. Voer hiervoor in het functiescherm van de grafische rekenmachine:

$$y_1 = 1 - \text{binomcdf}(n = 400; p_1 = X; k = 159)$$
$$y_2 = 0,005$$

De solver optie geeft een waarde van $p_2 \approx 0,3372$.

We vinden het Clopper-Pearson interval door de twee gevonden waarden als grenzen te nemen, oftewel het 99%-betrouwbaarheidsinterval voor de fractie aanhangers van de politieke partij is gelijk aan $[0,3372; 0,4653]$.

Opdracht 8.8: In een fabriek maakt men badkamertegels. Bekend is dat het productieproces 20% ondeugdelijke tegels oplevert. Na wat wijzigingen in het productieproces, werd bijgehouden hoe groot het aantal afgekeurde tegels is in de nieuwe situatie. Men vond de volgende aantallen:

Dag	Productieomvang	Aantal afgekeurd
1	1200	120
2	1800	210
3	1000	80
4	3000	450
5	1500	120
6	1700	170
7	2000	220
8	1400	80
9	1800	200
10	1500	130

(a) Stel dat alleen de gegevens van dag 1 bekend zouden zijn. Wat is op basis daarvan een 95%-betrouwbaarheidsinterval voor π: de populatiefractie van afgekeurde tegels?

Uitwerking

Laat X de kansvariabele zijn die het aantal afgekeurde tegels telt op een bepaalde dag. Omdat iedere tegel onafhankelijk van elkaar ondeugdelijk is of niet, hebben we te maken met een binomiale verdeling. Op dag 1 is er een productieomvang van 1200, waarvan er 120 zijn afgekeurd. Op basis hiervan geldt dat $X \sim \text{Binomiaal}(n = 1200; p = ?)$.
We willen een 95%-betrouwbaarheidsinterval voor p bepalen. Dit doen we met behulp van de Clopper-Pearson methode. Omdat de gewenste betrouwbaarheid 95% is, geldt dat $\alpha = 0,05$.
De Clopper-Pearson methode werkt als volgt:

(a) Bepaal de succeskans p_1 waarvoor geldt dat de linkeroverschrijdingskans van de uitkomst $k = 120$ gelijk is aan $\alpha/2$, oftewel $P(X \leq 120) = \alpha/2 = 0,025$. Voer hiervoor in het functiescherm van de grafische rekenmachine:

$$y_1 = \text{binomcdf}(n = 1200; p_1 = X; k = 120)$$
$$y_2 = 0,025$$

De solver optie geeft een waarde van $p_1 \approx 0,1184$.

(b) Bepaal de succeskans p_2 waarvoor geldt dat de rechteroverschrijdingskans van de uitkomst $k = 120$ gelijk is aan $\alpha/2$, oftewel $P(X \geq 120) = 1 - P(X \leq 119) = \alpha/2 = 0,005$. Voer hiervoor in het functiescherm van de grafische rekenmachine:

$$y_1 = 1 - \text{binomcdf}(n = 1200; p_1 = X; k = 119)$$
$$y_2 = 0,025$$

De solver optie geeft een waarde van $p_2 \approx 0,0836$.

We vinden het Clopper-Pearson interval door de twee gevonden waarden als grenzen te nemen. Het 95%-betrouwbaarheidsinterval voor de fractie ondeugdelijke tegels is dus gelijk aan $[0,0836; 0,1184]$.

(b) Bereken voor iedere dag de fractie afgekeurde tegels. Bereken op basis van deze dagelijkse steekproeffracties een interval van π_D: de gemiddelde dagelijkse fractie afgekeurde tegels.

Uitwerking

Deze vraag is niet van toepassing

(c) Bereken de totalen over de tien dagen van productieomvang en aantal afgekeurde tegels. Bereken op basis hiervan een interval voor π. Wat is de betekenis van deze π?

Uitwerking

Laat X nu de kansvariabele zijn die het aantal afgekeurde tegels telt over alle tien de dagen. Omdat iedere tegel onafhankelijk van elkaar ondeugdelijk is of niet, hebben we weer te maken met een binomiale verdeling. De totale productieomvang is 16900 geproduceerde tegels, waarvan 1780 afgekeurd. Op basis hiervan geldt dat $X \sim \text{Binomiaal}(n = 16900; p = ?)$.
We willen een 95%-betrouwbaarheidsinterval voor p bepalen. Dit doen we met behulp van de Clopper-Pearson methode. Omdat de gewenste betrouwbaarheid 95% is, geldt dat $\alpha = 0,05$.
De Clopper-Pearson methode werkt als volgt:

(a) Bepaal de succeskans p_1 waarvoor geldt dat de linkeroverschrijdingskans van de uitkomst $k = 1780$ gelijk is aan $\alpha/2$, oftewel $P(X \leq 1780) = \alpha/2 = 0,025$. Voer hiervoor in het functiescherm van de grafische rekenmachine:

$$y_1 = \text{binomcdf}(n = 16900; p_1 = X; k = 1780)$$
$$y_2 = 0,025$$

De solver optie geeft een waarde van $p_1 \approx 0,1101$.

(b) Bepaal de succeskans p_2 waarvoor geldt dat de rechteroverschrijdingskans van de uitkomst $k = 1780$ gelijk is aan $\alpha/2$, oftewel $P(X \geq 1780) = 1 - P(X \leq 1779) = \alpha/2 = 0,005$. Voer hiervoor in het functiescherm van de grafische rekenmachine:

$$y_1 = 1 - \text{binomcdf}(n = 16900; p_1 = X; k = 1779)$$
$$y_2 = 0,025$$

De solver optie geeft een waarde van $p_2 \approx 0,1007$.

We vinden het Clopper-Pearson interval door de twee gevonden waarden als grenzen te nemen. Het 95%-betrouwbaarheidsinterval voor de fractie afgekeurde tegels is dus gelijk aan $[0,1007; 0,1101]$.
Merk op dat dit interval al een stuk smaller is dan het interval dat we bij **a** hebben berekend. Dit komt doordat we nu een veel grotere steekproefomvang hebben waarop het interval is gebaseerd.

Opdracht 8.13: De regering besluit een referendum te houden over een kwestie aangaande de Europese Unie. Mensen kunnen alleen ja of nee stemmen. Door een bureau dat opiniepeilingen doet, wordt een representatieve steekproef getrokken uit de Nederlandse kiesgerechtigden. Van de 5000 ondervraagden stemden 3082 personen tegen en 1918 stemden voor.

(a) Geef een $99,74\%$-betrouwbaarheidsinterval voor π: de fractie voorstemmers in de populatie.

Uitwerking

Laat X de kansvariabele zijn die het aantal voorstemmers telt bij het referendum. Omdat iedere stemmer onafhankelijk van elkaar stemt, hebben we te maken met een binomiale verdeling. Op basis van de steekproef geldt dat $X \sim \text{Binomiaal}(n = 5000; p = ?)$.
We willen een $99,74\%$-betrouwbaarheidsinterval voor p bepalen. Dit doen we met behulp van de Clopper-Pearson methode. Omdat de gewenste betrouwbaarheid $99,74\%$ is, geldt dat $\alpha = 0,0026$.
De Clopper-Pearson methode werkt als volgt:

(a) Bepaal de succeskans p_1 waarvoor geldt dat de linkeroverschrijdingskans van de uitkomst $k = 3082$ gelijk is aan $\alpha/2$, oftewel $P(X \leq 3082) = \alpha/2 = 0,0013$. Voer hiervoor in het functiescherm van de grafische rekenmachine:

$$y_1 = \text{binomcdf}(n = 5000; p_1 = X; k = 3082)$$
$$y_2 = 0,0013$$

De solver optie geeft een waarde van $p_1 \approx 0,6370$.

(b) Bepaal de succeskans p_2 waarvoor geldt dat de rechteroverschrijdingskans van de uitkomst $k = 3082$ gelijk is aan $\alpha/2$, oftewel $P(X \geq 3082) = 1 - P(X \leq 3081) = \alpha/2 = 0,0013$. Voer hiervoor in het functiescherm van de grafische rekenmachine:

$$y_1 = 1 - \text{binomcdf}(n = 5000; p_1 = X; k = 3081)$$
$$y_2 = 0,0013$$

De solver optie geeft een waarde van $p_2 \approx 0,5955$.

We vinden het Clopper-Pearson interval door de twee gevonden waarden als grenzen te nemen. Het 95%-betrouwbaarheidsinterval voor de fractie ondeugdelijke tegels is dus gelijk aan $[0,5955; 0,6370]$.

(b) Vind je het houden van een referendum nog zinvol, gegeven het antwoord bij vraag **a**?

Uitwerking

Het antwoord op deze vraag is afhankelijk van het kiessysteem waarmee gewerkt wordt. Indien het gaat om een strikte meerderheid van $50\% + 1$ stem, dan is het niet heel waardevol omdat vrijwel zeker is dat de voorstemmers gaan winnen. Als het quotum op een andere waarde ligt of als er andere zwaarwegende politieke redenen in het spel zijn, kan het uiteraard wel van meerwaarde zijn om een referendum te houden.

Opdracht 8.17: De handvatten van koffers moeten van zodanige kwaliteit zijn dat ze niet breken bij een forse belasting van de koffers. Ter controle van de kwaliteit van de handvatten werden er proeven genomen waarbij het gewicht dat aan een handvat hing net zolang werd opgevoerd tot het handvat brak. Er werden negen handvatten getest waarbij de maximale belasting werd vastgesteld. De resultaten staan in de volgende tabel:

Handvat	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Breuk bij een gewicht (in kg) van	84	87	81	85	90	93	86	88	80

(a) Geef een schatting van de variantie.

Uitwerking

We berekenen een schatting s^2 voor de variantie door allereerst het steekproefgemiddelde te bepalen:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{x_1 + x_2 + \ldots + x_9}{9} = \frac{84 + 87 + \ldots + 90}{9} = 86.$$

Nu kunnen we de steekproefvariantie bepalen:

$$\begin{aligned} s^2 &= \frac{(x_1 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n - 1} \\ &= \frac{(x_1 - \overline{x})^2 + \ldots + (x_9 - \overline{x})^2}{9 - 1} \\ &= \frac{(84 - 86)^2 + \ldots + (80 - 86)^2}{8} \\ &= 17,0 \end{aligned}$$

De puntschatting voor de variantie σ^2 is dus gelijk aan $s^2 = 17$.

(b) Geef een betrouwbaarheidsinterval voor μ_x: de verwachtingswaarde van de maximumbelasting van een handvat. Kies een 99%-betrouwbaarheidsinterval.

Uitwerking

Omdat we hier te maken hebben met een onbekende standaardafwijking σ en een kleine steekproefomvang $n = 9 < 30$, moeten we de t-verdeling gebruiken. Gegeven is een betrouwbaarheidsniveau van 99%, oftewel $\alpha = 0,01$. Verder hebben we bij vraag (a) een schatting van de variantie $s^2 = 17$ berekend, oftewel de standaardafwijking kan worden geschat met $s = \sqrt{s^2} = \sqrt{17}$. De t-waarde die hierbij hoort is gelijk aan

$$t = \text{InvT}(opp = 1 - \alpha/2; df = n - 1) = \text{InvT}(opp = 0,995; df = 8) \approx 3,3554.$$

Het 99%-betrouwbaarheidsinterval voor μ_x is in dit geval gelijk aan

$$\begin{aligned} & [\overline{x} - t \cdot \frac{s}{\sqrt{n}}; \overline{x} + t \cdot \frac{s}{\sqrt{n}}] \\ &= [86 - 3,3554 \cdot \frac{\sqrt{17}}{\sqrt{9}}; 86 + 3,3554 \cdot \frac{\sqrt{17}}{\sqrt{9}}] \\ &= [81,3885; 90,6115] \end{aligned}$$

Met 99% betrouwbaarheid zit de gemiddelde maximumbelasting van een handvat tussen ongeveer $81,39$ en $90,61$ kg.

Opdracht 8.18: De Rijksdienst voor het Wegverkeer doet onderzoek naar het percentage auto's dat geregistreerd staat zonder dat er een geldige WA-verzekering is afgesloten.

(a) Voor een steekproef van 1000 kentekens werd nauwkeurig nagetrokken of op de auto een WA-verzekering was afgesloten. Er bleken 80 auto's niet verzekerd te zijn. Geef een 95%-betrouwbaarheidsinterval voor de fractie onverzekerde auto's in de gehele populatie.

Uitwerking

Laat X de kansvariabele zijn die het aantal onverzekerde auto's telt bij een steekproef van 1000 kentekens. Omdat iedere stemmer onafhankelijk van elkaar stemt, hebben we te maken met een binomiale verdeling. Op basis van de steekproef geldt dat $X \sim$ Binomiaal$(n = 1000; p = ?)$, waarbij we een uitkomst van 80 onverzekerde auto's hebben geobserveerd,
We willen een 95%-betrouwbaarheidsinterval voor p bepalen. Dit doen we met behulp van de Clopper-Pearson methode. Omdat de gewenste betrouwbaarheid 95% is, geldt dat $\alpha = 0,05$.
De Clopper-Pearson methode werkt als volgt:

(a) Bepaal de succeskans p_1 waarvoor geldt dat de linkeroverschrijdingskans van de uitkomst $k = 80$ gelijk is aan $\alpha/2$, oftewel $P(X \leq 80) = \alpha/2 = 0,025$. Voer hiervoor in het functiescherm van de grafische rekenmachine:

$$\begin{aligned} y_1 &= \text{binomcdf}(n = 1000; p_1 = X; k = 80) \\ y_2 &= 0,025 \end{aligned}$$

De solver optie geeft een waarde van $p_1 \approx 0,0986$.

(b) Bepaal de succeskans p_2 waarvoor geldt dat de rechteroverschrijdingskans van de uitkomst $k = 80$ gelijk is aan $\alpha/2$, oftewel $P(X \geq 80) = 1 - P(X \leq 79) = \alpha/2 = 0,025$. Voer hiervoor in het functiescherm van de grafische rekenmachine:

$$y_1 = 1 - \text{binomcdf}(n = 1000; p_1 = X; k = 79)$$
$$y_2 = 0,025$$

De solver optie geeft een waarde van $p_2 \approx 0,0639$.

We vinden het Clopper-Pearson interval door de twee gevonden waarden als grenzen te nemen. Het 95%-betrouwbaarheidsinterval voor de fractie onverzekerde auto's is dus gelijk aan $[0,0639; 0,0986]$.

(b) Stel dat de Rijksdienst het percentage onverzekerde auto's wil schatten met een marge van plus of min 1%. Hoeveel auto's moeten dan voor een steekproef worden opgenomen? Geef aan welke veronderstellingen u hanteert.

Uitwerking

Bij $n = 1000$ is de foutmarge gelijk aan $\frac{0,0986-0,0639}{2} = 0,01735$, oftewel $1,73\%$. Dat betekent dus dat de steekproefomvang nog groter moet worden om een foutmarge van slechts 1% te krijgen.
Omdat n heel groot is, mag de normale benadering gebruikt worden en geldt de formule

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sqrt{p \cdot (1-p)}}{a}\right)^2,$$

waarbij $a = 0,01$ de gewenste foutmarge (afwijking van het steekproefgemiddelde) is. Omdat het betrouwbaarheidsniveau 95% is, geldt $\alpha = 0,05$ en dus $z_{\alpha/2} = \text{InvNorm}(opp = 1 - \alpha/2 = 0,975) \approx 1,9600$. Verder nemen we aan dat $\hat{p} = 0,08$, dit is de puntschatter voor de binomiale succeskans p. De minimale steekproefomvang om een foutmarge van maximaal plus of min 1% te krijgen is gelijk aan

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sqrt{p \cdot (1-p)}}{a}\right)^2 = \left(\frac{1,9600 \cdot \sqrt{0,08 \cdot 0,92}}{0,01}\right)^2 \approx 2827,4176.$$

Afronden naar boven op het eerstvolgende gehele getal geeft een minimale steekproefomvang van $n = 2828$ auto's om een foutmarge van slechts $\pm 1\%$ te krijgen.

(c) Voor een steekproef van 25 onverzekerde auto's werd de ouderdom bepaald. Dat leverde een gemiddelde van $8,4$ jaar en een standaarddeviatie van $1,7$ jaar. Geef een 95%-betrouwbaarheidsinterval voor de gemiddelde leeftijd van alle onverzekerde auto's.

Uitwerking

Gegeven is een steekproef van $n = 25$ onverzekerde auto's, waarvoor van de leeftijd een steekproefgemiddelde $\overline{x} = 8,4$ jaar en steekproefstandaarddeviatie $s = 1,7$ jaar is bepaald. Omdat de steekproefomvang $n = 25 < 30$ is en de

standaarddeviatie σ onbekend is, moeten we de t-verdeling gebruiken. Omdat het gewenste betrouwbaarheidsniveau 95% is, geldt dat $\alpha = 0,05$.
De bijbehorende t-waarde is gelijk aan

$$t = \mathrm{InvT}(opp = 1 - \alpha/2; df = n - 1) = \mathrm{InvT}(opp = 0,975; df = 24) \approx 2,0639.$$

Het 95%-betrouwbaarheidsinterval is gelijk aan

$$\begin{aligned}
&[\overline{x} - t \cdot \frac{s}{\sqrt{n}}; \overline{x} + t \cdot \frac{s}{\sqrt{n}}] \\
&= [8,4 - 2,0639 \cdot \frac{1,7}{\sqrt{25}}; 8,4 + 2,0639 \cdot \frac{1,7}{\sqrt{25}}] \\
&= [7,6983; 9,1017]
\end{aligned}$$

Met 95% betrouwbaarheid zit de gemiddelde leeftijd van een onverzekerde auto tussen $7,7$ en $9,1$ jaar.

Opdracht 8.19: Een technicus van een bedrijf dat oefenapparaten ontwikkelt voor fitnesstrainingen heeft een nieuw ontwerp gemaakt van een bepaalde machine. Bedoeling is dat dit apparaat spierversterkend werkt voor personen met rugklachten. Helaas was maar één prototype van het apparaat beschikbaar om het te kunnen testen. Hierdoor konden slechts de gegevens van twaalf patiënten worden verzameld. Deze twaalf patiënten gingen oefenen met dit apparaat en het aantal dagen werd opgetekend dat nodig was om volledig het gewenste resultaat te behalen. Van de twaalf patiënten werden de volgende aantallen dagen waargenomen:

$$15, 18, 41, 27, 23, 16, 29, 30, 32, 22, 25, 22.$$

(a) Schat de standaarddeviatie.

Uitwerking

We berekenen een schatting s voor de standaarddeviatie door allereerst het steekproefgemiddelde te bepalen:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{x_1 + x_2 + \ldots + x_{12}}{12} = \frac{15 + 18 + \ldots + 22}{12} = 25.$$

Nu kunnen we de steekproefvariantie bepalen:

$$\begin{aligned}
s^2 &= \frac{(x_1 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n - 1} \\
&= \frac{(x_1 - \overline{x})^2 + \ldots + (x_{12} - \overline{x})^2}{12 - 1} \\
&= \frac{(15 - 25)^2 + \ldots + (22 - 25)^2}{11} \\
&= 54.7272
\end{aligned}$$

De puntschatting s voor de standaarddeviatie vinden we door de wortel hiervan te nemen:

$$s = \sqrt{54,7272} \approx 7,3978.$$

(b) Bereken een 95%-betrouwbaarheidsinterval voor μ: de gemiddelde hersteltijd voor de populatie.

Uitwerking

Gegeven is een steekproef van $n = 12$ patiënten, waarvoor het aantal dagen tot het gewenste resutaat het steekproefgemiddelde $\overline{x} = 25$ dagen en de steekproefstandaarddeviatie $s = 7,3978$ dagen is bepaald. Omdat de steekproefomvang $n = 12 < 30$ is en de standaarddeviatie σ onbekend is, moeten we de t-verdeling gebruiken. Omdat het gewenste betrouwbaarheidsniveau 95% is, geldt dat $\alpha = 0,05$.
De bijbehorende t-waarde is gelijk aan

$$t = \mathrm{InvT}(\mathit{opp} = 1 - \alpha/2; \mathit{df} = n - 1) = \mathrm{InvT}(\mathit{opp} = 0,975; \mathit{df} = 11) \approx 2,2010.$$

Het 95%-betrouwbaarheidsinterval is gelijk aan

$$\begin{aligned} &[\overline{x} - t \cdot \frac{s}{\sqrt{n}}; \overline{x} + t \cdot \frac{s}{\sqrt{n}}] \\ &= [25 - 2,2010 \cdot \frac{7,3978}{\sqrt{12}}; 25 + 2,2010 \cdot \frac{7,3978}{\sqrt{12}}] \\ &= [20,2997; 29,7003] \end{aligned}$$

Met 95% betrouwbaarheid zit het gemiddelde aantal dagen tot het gewenste resultaat tussen 20 en 30 dagen.

(c) In een vervolgstudie werden 90 personen getest. Er werd een standaarddeviatie gevonden van $6,5$ dagen en een gemiddelde hersteltijd van 27 dagen. Bereken opnieuw een 95%-betrouwbaarheidsinterval voor μ.

Uitwerking

Gegeven is nu een steekproef van $n = 90$ patiënten, waarvoor het aantal dagen tot het gewenste resutaat het steekproefgemiddelde $\overline{x} = 27$ dagen en de steekproefstandaarddeviatie $s = 6,5$ dagen is bepaald. Omdat nu geldt dat de steekproefomvang $n = 90 > 30$ is, mogen we wel de normale verdeling gebruiken en σ schatten met s. Omdat het gewenste betrouwbaarheidsniveau 95% is, geldt dat $\alpha = 0,05$.
De bijbehorende z-waarde is gelijk aan

$$z = z_{\alpha/2} = \mathrm{InvNorm}(\mathit{opp} = 1 - \alpha/2) = \mathrm{InvNorm}(0,975) \approx 1,9600.$$

Het 95%-betrouwbaarheidsinterval is gelijk aan

$$\begin{aligned} &[\overline{x} - z \cdot \frac{s}{\sqrt{n}}; \overline{x} + z \cdot \frac{s}{\sqrt{n}}] \\ &= [27 - 1,9600 \cdot \frac{6,5}{\sqrt{90}}; 27 + 1,9600 \cdot \frac{6,5}{\sqrt{90}}] \\ &= [25,6571; 28,3429] \end{aligned}$$

Met 95% betrouwbaarheid zit het gemiddelde aantal dagen tot het gewenste resultaat tussen 25 en 28 dagen. Merk op dat dit betrouwbaarheidsinterval al een stuk kleiner is, wat het effect van een grotere steekproefomvang onderschrijft!

(d) Men wenst het bij vraag **c** gevraagde interval zodanig te bepalen dat het resulterende interval voor μ niet breder is dan twee dagen. Hoe groot moet de steekproefomvang worden gekozen om dat te bereiken, als de standaarddeviatie van vraag **a** als vertrekpunt mag worden gekozen?

Uitwerking

Als het resulterende interval voor μ niet breder is dan twee dagen, dan is ook de foutmarge (afwijking van het gemiddelde) a niet groter dan 1 dag ($a = 1$). Verder mogen we aannemen dat $\sigma = 7,3978$. In formulevorm ziet dit er als volgt uit:

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq a = 1,$$

oftewel

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{a}\right)^2 = \left(\frac{1,9600 \cdot 7,3978}{1}\right)^2 \approx 210,2410$$

Afgerond naar boven naar het eerstvolgende gehele getal geeft een minimale steekproefomvang van $n = 211$ personen om een interval te krijgen voor μ van maximaal 2 dagen breed.