Uitwerkingen opdrachten en opgaven

Statistiek in Business

voor gevorderden

Rob Erven,
Zwolle 2014.

Inhoudsopgave

Uitwerkingen hoofdstuk 1…….. .2

Uitwerkingen hoofdstuk 2……5

Uitwerkingen hoofdstuk 3……..13

Uitwerkingen hoofdstuk 4……..17

Uitwerkingen hoofdstuk 5……..25

Uitwerkingen hoofdstuk 6……..32

Uitwerkingen hoofdstuk 7……..36

Uitwerkingen hoofdstuk 8……..46

Uitwerkingen hoofdstuk 9……..51

Uitwerkingen hoofdstuk 10……57

Uitwerkingen hoofdstuk 11……65

Uitwerkingen hoofdstuk 12……75

Uitwerkingen hoofdstuk 1

Opdrachten hoofdstuk 1

Opdracht 1

$$\bar{x} \pm z_{\alpha/2} . \frac{s}{\sqrt{n}} = 12 \pm 1{,}645 . \frac{2}{\sqrt{500}} = 12 \pm 0{,}15 .$$

Het 90% betrouwbaarheidsinterval voor de gemiddelde wijnconsumptie per kwartaal varieert van **11,85** liter per persoon tot **12,15** liter per persoon.

Opdracht 2

$$n \geq \frac{z_{\alpha/2} . s^2}{F^2} = \frac{1{,}96^2 . 10^2}{0{,}5^2} = 1.536{,}64 .$$

De minimale steekproefgrootte bedraagt **1.537** schadeformulieren.

Opdracht 3

$$\bar{x} \pm z_{\alpha/2} . \frac{s}{\sqrt{n}} . \sqrt{\frac{N-n}{N-1}} = 6 \pm 1{,}96 . \frac{3}{\sqrt{350}} . \sqrt{\frac{1.200-350}{1.200-1}} = 6 \pm 0{,}26 .$$

Het 95% betrouwbaarheidsinterval voor de gemiddelde besteding per week aan het financieel softwarepakket geleverd door Dutch Compusteel varieert van **5,74** uur tot **6,26** uur.

Opdracht 4

$$n \geq \frac{N . z_{\alpha/2}{}^2 . s^2}{z_{\alpha/2}{}^2 . s^2 + (N-1) . F^2} = \frac{2.000 * 1{,}645^2 * 10.000^2}{1{,}645^2 * 10.000^2 + (2.000-1) * 1.200^2} = 171{,}86.$$

De minimale steekproefgrootte bedraagt n = **172** klanten.

Kennisopgaven hoofdstuk 1

Opgave 1

Een aselecte steekproef is een steekproef waarbij ieder element van een populatie een even grote kans heeft om in de steekproef terecht te komen.

Opgave 2

- De betrouwbaarheid geeft aan in hoeveel procent van alle gevallen het populatiekengetal correct geschat wordt.
- De foutmarge geeft aan hoever je er met een zekere waarschijnlijkheid maximaal naast zit met de gegeven schatting.

Opgave 3

Een tweezijdig betrouwbaarheidsinterval voor een gemiddelde vertelt welke getallen waarschijnlijk zijn voor het populatiegemiddelde. Dit interval wordt bepaald met een bovengrens en een ondergrens.

Opgave 4

Bij een éénzijdig betrouwbaarheidsinterval heb je alleen een bovengrens of een ondergrens en bij een tweezijdig betrouwbaarheidsinterval heb je beiden.

Opgave 5

In de formule van de minimale steekproefgrootte voor de beperkt eindige populatie houd je expliciet rekening met de populatiegrootte. Bij de formule van de minimale steekproefgrootte bij een zeer grote populatie doe je dat niet.

Toegepaste opgaven hoofdstuk 1

Opgave 1

a. $n \geq \frac{z_{\alpha/2}.s^2}{F^2} = \frac{1{,}645^2.0{,}300^2}{0{,}0355^2} = 198{,}81$. De minimale steekproefgrootte is **199** internetgebruikers.

b. $\bar{x} \pm z_{\alpha/2}.\frac{s}{\sqrt{n}} = 1{,}500 \pm 1{,}645.\frac{0{,}278}{\sqrt{200}} = 1{,}500 \pm 0{,}032$. Het 90% betrouwbaarheidsinterval loopt van **1,468** seconde aandacht voor een zoekresultaat tot **1,532** seconde.

Opgave 2

a. $\bar{x} - z_{\alpha}.\frac{s}{\sqrt{n}} = 500 - 2{,}33.\frac{150}{\sqrt{200}} = 475{,}29$ euro. De minimale gemiddelde netto besteding aan de boodschappen per maand bedraagt **475,29** euro.

b. $\bar{x} - z_\alpha . \frac{s}{\sqrt{n}} . \sqrt{\frac{N-n}{N-1}} = 500 - 2{,}33 . \frac{150}{\sqrt{200}} . \sqrt{\frac{1.800-200}{1.800-1}} = 476{,}69$ euro. De minimale gemiddelde netto besteding aan boodschappen per maand bedraagt met 99% betrouwbaarheid **476,69** euro.

Opgave 3

$n \geq \frac{z_{\alpha/2} . s^2}{F^2} = \frac{1{,}96^2 . 400^2}{45^2} = 303{,}53$. De minimale steekproefgrootte bedraagt **304** huishoudens.

Opgave 4

$$n \geq \frac{N . z_{\alpha/2}{}^2 . s^2}{z_{\alpha/2}{}^2 . s^2 + (N-1) . F^2} = \frac{2.000 * 2{,}576^2 * 500^2}{2{,}576^2 * 500^2 + (2.000-1) * 50^2} = 498{,}45 .$$

De minimale steekproefgrootte bedraagt **499** huishoudens met een villa met een waarde vanaf 4 miljoen euro.

Opgave 5

$\bar{x} \pm z_\alpha . \frac{s}{\sqrt{n}} = 12.500 \pm 1{,}645 . \frac{8.000}{\sqrt{2.000}} = 12.500 \pm 294{,}27$. Dit levert een 90% betrouwbaarheidsinterval op van **12.203,77 euro tot 12.794,27 euro** bereid om te besteden aan een auto komend jaar.

Opgave 6

a. $n \geq \frac{N . z_{\alpha/2}{}^2 . s^2}{z_{\alpha/2}{}^2 . s^2 + (N-1) . F^2} = \frac{1.200 * 1{,}645^2 * 100^2}{1{,}645^2 * 100^2 + (1.200-1) * 10^2} = 247{,}42$ mensen van 16 jaar en ouder.

b. $\bar{x} \pm z_\alpha . \frac{s}{\sqrt{n}} . \sqrt{\frac{N-n}{N-1}} = 18.000 \pm 2{,}576 . \frac{100}{\sqrt{248}} . \sqrt{\frac{1.200-248}{1.200-1}} = 18.000 \pm 14{,}58$. Dit levert een 99% betrouwbaarheidsinterval op van **17.985,02 euro tot 18.014,58 euro** voor het gemiddelde bedrag dat mensen kwijt zijn geraakt door de crisis.

Competentieprikkel

a. $n \geq \frac{z_{\alpha/2} . s^2}{F^2} = \frac{1{,}96^2 * 2.000^2}{80^2} = 2.401$. De minimale steekproefgrootte bedraagt **2.401** huishoudens.

b. $\bar{x} + z_{\alpha} . \frac{s}{\sqrt{n}} = 4.289 + 1{,}645 . \frac{1.935}{\sqrt{2.500}} = 4.352{,}66$ euro. Met 95% betrouwbaarheid bedraagt de maximale besteding aan voeding per jaar door een huishouden **4.352,66** euro.

Uitwerkingen hoofdstuk 2

Opdrachten hoofdstuk 2

Opdracht 1

Net als bij een eenzijdig betrouwbaarheidsinterval heb je hier slechts met één grenswaarde te maken. In de case is dit een bovengrens. Aan de rechteroppervlakte van deze grens verbind je de onbetrouwbaarheid en dus niet de helft van de onbetrouwbaarheid.

Opdracht 2

Stap 1

$H_0 : \mu \leq 7{,}6$

$H_1 : \mu > 7{,}6$

Met α = 5%.

Stap 2

- n = 2.000,
- $\bar{x} = 7{,}7$,
- s = 2,0.

Stap 3

De bovengrens is 7,6 + 1,645 . 2,0/√2.000 = **7,67.**

Stap 4

Conclusie: 7,7 > 7,67, dus je verwerpt de nulhypothese nog steeds. Met 95% betrouwbaarheid kun je zeggen dat het gemiddelde rapportcijfer van het imago van Philips hoger is dan 7,6.

Opdracht 3

De conclusie luidt: **1,50 < 1,645**, dus verwerp de nulhypothese niet. Het gemiddelde rapportcijfer van het imago van Philips is niet significant hoger dan 7,5.

Opdracht 4

$P(\bar{x} \geq 7{,}57 | \mu = 7{,}8) = P(z > \frac{7{,}57 - 7{,}8}{2{,}0/\sqrt{2.000}}) = P(z > -5{,}14) = 1 =$ **100%**. Met een steekproef van 2.000 mensen herkent de toets prima dat het gemiddelde significant hoger is dan 7,5.

Opdracht 5

De p-waarde is **10,1% > 10%** betekent dat je de nulhypothese niet moet verwerpen. Het gemiddelde is niet significant hoger zoals aangegeven in de alternatieve hypothese.

Kennisopgaven hoofdstuk 2

Opgave 1

- Formulering van de hypothesen en bepaling onbetrouwbaarheid.
- Bepaling toetsingsgrootheid en uitrekenen met behulp van de steekproefgegevens van de waarde van de toetsingsgrootheid.
- Bepaling kritieke gebied.
- Trekken van conclusie.

Opgave 2

De eerste hypothese heet de nulhypothese en de tweede hypothese heet de alternatieve hypothese.

Opgave 3

Bij éénzijdig toetsen maak je gebruik van een bovengrens of een ondergrens, bij tweezijdig toetsen maak je gebruik van een bovengrens en een ondergrens.

Opgave 4

$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$. In deze formule vergelijk je het steekproefgemiddelde met het populatiegemiddelde. Door te delen door de standaardfout kom je terecht in de standaard normale verdeling.

Opgave 5

Als de toetsingsgrootheid in het kritieke gebied ligt, luidt de conclusie dat er een significantie is.

Opgave 6

Significant betekent dat het steekproefresultaat dermate afwijkt van de nulhypothese dat dit niet meer verklaard kan worden door toeval.

Opgave 7

Het onderscheidend vermogen is de kans om de nulhypothese te verwerpen als de alternatieve hypothese waar is.

Opgave 8

a. De p-waarde is bij eenzijdig toetsen de rechter overschrijdingskans bij een rechtseenzijdige toets en de linker overschrijdingskans bij een links eenzijdige toets. Bij tweezijdig toetsen is de p-waarde 2* de rechter (of linker) overschrijdingskans.
b. De p-waarde vergelijk je met α.
c. Als de p-waarde groter is dan α, dan heb je geen significantie; als de p-waarde kleiner is dan of gelijk aan α, dan is er significantie.

Toegepaste opgaven hoofdstuk 2

Opgave 1

a. $\bar{x} \pm z_{\alpha/2} . \frac{s}{\sqrt{n}} = 220 \pm 1{,}645 . \frac{45}{\sqrt{400}} = 220 \pm 3{,}70$. Het 90% betrouwbaarheidsinterval van de gemiddelde potentiële besteding aan een dvd-recorder loopt van **216,30** euro tot **223,70** euro.
b.

Stap 1

$H_0 : \mu = 215$
$H_1 : \mu \neq 215$

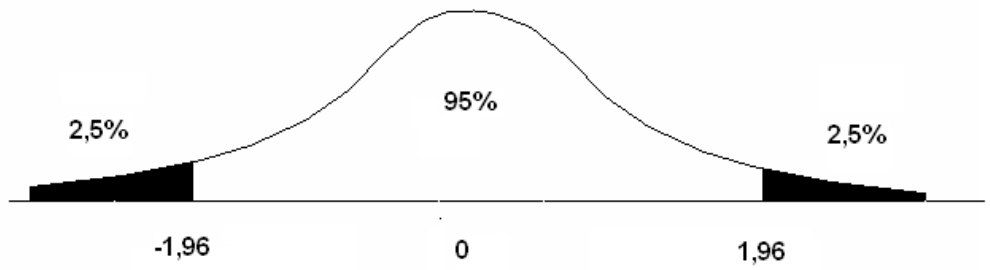
Met α = 5%.

Stap 2

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{220 - 215}{45 / \sqrt{400}} = 2{,}22 .$$

Stap 3

De ondergrens is **-1,96** en de bovengrens **1,96**.



Stap 4

Conclusie: z = **2,22 > 1,96**, verwerp de nulhypothese. Met 95% betrouwbaarheid kun je concluderen dat de gemiddelde potentiële besteding aan een dvd-recorder hoger is 215 euro.

c. De p-waarde = 2 * P(z > 2,22) = 2 * 0,0132 = 0,0264 = **2,64% < 5%.** Verwerp de nulhypothese. Met 95% betrouwbaarheid kun je concluderen dat de gemiddelde potentiële besteding aan een dvd-recorder hoger is 215 euro.

Opgave 2

a. $\bar{x} \pm z_{\alpha/2} . \frac{s}{\sqrt{n}} = 35.653 \pm 1{,}96 . \frac{5.702}{\sqrt{300}} = 35.653 \pm 645$.

Het 95% betrouwbaarheidsinterval voor het gemiddelde inkomen voor vrouwen die fulltime werken in New York varieert van **35.008** dollar tot **36.298** dollar.

b.

Stap 1

$H_0 : \mu = 35.000$

$H_1 : \mu > 35.000$

Met α = 1%.

Stap 2

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{35.653 - 35.000}{5.702 / \sqrt{300}} = 1{,}98 .$$

Stap 3

De bovengrens is **2,33**.

Stap 4

Conclusie: z = **1,98 < 2,33**, verwerp de nulhypothese niet. Met 99% betrouwbaarheid kun je niet concluderen dat vrouwen die fulltime werken in New York gemiddeld meer verdienen dan 35.000 dollar.

c. De p-waarde = P(z >1,98) = 0,0239 = **2,39% > 1%.** Wederom verwerp je de nulhypothese niet. Met 99% betrouwbaarheid kun je niet concluderen dat vrouwen die fulltime werken in New York gemiddeld meer verdienen dan 35.000 dollar.

Opgave 3

Stap 1

$$H_0 : \mu \geq 2.500$$
$$H_1 : \mu < 2.500$$

Met α = 10%.

Stap 2

$$z = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}} = \frac{2.460 - 2.500}{\frac{872}{\sqrt{600}} \cdot \sqrt{\frac{4.000-600}{4.000-1}}} = -1{,}22 .$$

Stap 3

De ondergrens is **-1,28**.

Stap 4

Conclusie: z = **-1,22 > -1,28**, verwerp de nulhypothese niet. Je kunt niet beweren met 90% betrouwbaarheid dat de Kamer van Koophandel gelijk heeft.

Opgave 4

a. $\begin{matrix} H_0 : \mu = 100 \\ H_1 : \mu > 100 \end{matrix}$.

b. α = 5%, dus de bovengrens is **1,645**.

c. $z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{102{,}75 - 100}{12{,}23/\sqrt{400}} = 4{,}50$. Deze waarde is veel groter dan 1,645. De conclusie is dat het gemiddelde bedrag dat te weinig opgegeven wordt door particulieren aan de belasting significant hoger is dan 100 euro.

d. In het oorspronkelijke probleem is de bovengrens 100 + 1,645 * 12,23/√400 = 101,01 euro. Het onderscheidend vermogen is

$P(\bar{x} > 101{,}01) = P(z > \frac{101{,}01 - 103}{12{,}23/\sqrt{400}}) = P(z > -3{,}25) = 1 - P(z < -3{,}25) = 1 - P(z > 3{,}25)$

=0,9994 = **99,94%.** Je hebt bij deze toets een groot onderscheidend vermogen.

Opgave 5

a. $\bar{x} \pm z_{\alpha/2} . \frac{s}{\sqrt{n}} = 8.500 \pm 2{,}576 . \frac{4.000}{\sqrt{2.000}} = 8.500 \pm 230{,}40$.

Het 99% betrouwbaarheidsinterval voor het gemiddeld bedrag dat men bereid is te betalen voor een tweede auto komend jaar loopt van **8.269,60 euro tot 8.730,40 euro**.

b. **Stap 1**

$H_0 : \mu = 8.400$

$H_1 : \mu > 8.400$

Met α onbekend.

Stap 2

$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{8.500 - 8.400}{4.000/\sqrt{2.000}} = 1{,}12$.

Stap 3

De bovengrens is **2,33** als α = 1%, **1,645** als α = 5%, **1,28** als α = 10%.

Stap 4

Conclusie: z = **1,12 < 1,28**, verwerp de nulhypothese niet. Je kunt niet concluderen dat vrouwen die fulltime werken in New York gemiddeld significant meer verdienen dan 35.000 dollar.

c. P-waarde = P(z > 1,12) = 0,1314 = 13,14%

Opgave 6

a. $\bar{x} \pm z_{\alpha/2} . \frac{s}{\sqrt{n}} = 450 \pm 1{,}645 . \frac{110}{\sqrt{200}} = 450 \pm 12{,}80$.

Het 90% betrouwbaarheidsinterval voor het gemiddeld bedrag van alle facturen loopt van **437,20 euro tot 462,80 euro**.

b. Stap 1

$H_0 : \mu = 475$

$H_1 : \mu \neq 475$

Met α = 5%.

Stap 2

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{450 - 475}{1100 / \sqrt{200}} = -3{,}21 .$$

Stap 3

De bovengrens is **1,96** en de ondergrens **-1,96**.

Stap 4

Conclusie: z = **-3,21 < -1,96**, verwerp de nulhypothese. Je kunt concluderen dat het gemiddelde factuurbedrag significant verschilt (lager is) van 475 euro.

Competentieprikkel

a. $\bar{x} \pm z_{\alpha/2} . \frac{s}{\sqrt{n}} = 5{,}2 \pm 1{,}645 . \frac{1{,}6}{\sqrt{800}} = 5{,}2 \pm 0{,}1$. Het 90% betrouwbaarheidsinterval voor het gemiddelde verfgebruik per jaar in China loopt van **5,1 liter tot 5,3 liter** per persoon.

b.

Stap 1

$H_0 : \mu = 5$

$H_1 : \mu > 5$

Met α = 5%.

Stap 2

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{5{,}2 - 5}{1{,}6 / \sqrt{800}} = 3{,}54 .$$

Stap 3

De bovengrens is **1,645**.

Stap 4

Conclusie: z = **3,54 > 1,645**, verwerp de nulhypothese. Met 95% betrouwbaarheid kun je concluderen dat het gemiddelde verfgebruik in China hoger is dan 5 liter per jaar per persoon.

c. De p-waarde = P(z > 3,54) = 0,0002 = **0,02% < 5%.** Wederom verwerp je de nulhypothese. Met 95% betrouwbaarheid kun je concluderen dat het gemiddelde verfgebruik in China hoger is dan 5 liter per jaar per persoon.

d. $n \geq \frac{z_{\alpha/2}.s^2}{F^2} = \frac{1{,}96^2.1{,}6^2}{0{,}1^2} = 983{,}45$. De minimale steekproefgrootte bedraagt **984** personen.

Uitwerkingen hoofdstuk 3

Opdrachten hoofdstuk 3

Opdracht 1

$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^{\ 2}}{n_1} + \frac{s_2^{\ 2}}{n_2}}} = \frac{30}{\sqrt{\frac{180^2}{200} + \frac{165^2}{200}}} = 1{,}74$. De conclusie luidt dat **1,74 < 1,96** en dus is er geen aantoonbaar verschil in de gemiddelde kantoorkosten van Amsterdam en Brussel.

Opdracht 2

De p-waarde is 2 * P(z >7,79) = **0% < 10%.** De conclusie luidt dat merk B significant langer meegaat dan merk A.

Opdracht 3

$z = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{0{,}2100 - 0}{1{,}1341 / \sqrt{300}} = 3{,}21$. De conclusie luidt dat **3,21 > 1,96**, verwerp dus de nulhypothese. Merk B gaat nog steeds significant langer mee dan merk A.

Kennisopgaven hoofdstuk 3

Opgave 1

Homogeen betekent dat de standaarddeviatie klein is in verhouding tot het gemiddelde.

Opgave 2

Twee steekproeven zijn onafhankelijk als de uitkomsten van de elementen uit de ene steekproef geen invloed hebben of geen relatie hebben met de uitkomsten van de elementen uit de andere steekproef. Heel formeel zijn A en B onafhankelijk als P(A en B) = P(A).P(B), waarbij A een uitkomst is van de ene steekproef en B van de andere steekproef.

Opgave 3

Voor twee steekproeven geldt dat ze gepaard zijn als aan ieder element van de steekproef uit het onderzoek twee metingen verricht zijn met dezelfde variabele.

Opgave 4

Je bepaalt twee aan twee paarsgewijze verschillen bij de gepaarde z-toets om het verschil tussen de elementen in de steekproef op te heffen. Dit verschil heeft op deze wijze geen invloed meer op de toetsingsgrootheid.

Opgave 5

Dat betekent dat de noemer van de z-toetsingsgrootheid groot wordt en dus de toetsingsgrootheid dus een kleinere waarde aanneemt. Dat laatste betekent dat er minder snel een significantie gevonden wordt. Bij heterogenere groepen is het lastiger om significanties te vinden tussen de gemiddelden.

Toegepaste opgaven hoofdstuk 3

Opgave 1

Stap 1

$$H_0 : \mu_{AH} - \mu_{Jumbo} = 0$$
$$H_1 : \mu_{AH} - \mu_{Jumbo} > 0$$

Met α = 10%.

Stap 2

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{3{,}75 - 0}{2{,}67 / \sqrt{250}} = 22{,}21.$$

Stap 3

De bovengrens is **1,28**.

Stap 4

De conclusie: **22,21 > 1,28**, dus AH is gemiddeld significant duurder dan Jumbo.

Opgave 2

Stap 1

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 < \mu_2$

Met α = 1%.

Stap 2

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^{\ 2}}{n_1} + \frac{s_2^{\ 2}}{n_2}}} = \frac{41 - 43}{\sqrt{\frac{14^2}{200} + \frac{16^2}{200}}} = -1{,}33.$$

Stap 3

De bovengrens is **-2,33**.

Stap 4

De conclusie: **-1,33 > -2,33**, je kunt niet beweren met 99% betrouwbaarheid dat de digitale spaarkaart effect heeft gehad..

Opgave 3

a. De formule voor het betrouwbaarheidsinterval luidt $\bar{x} \pm z_{\alpha/2} . \frac{s}{\sqrt{n}}$. Voor jongens volgt nu $\bar{x} \pm z_{\alpha/2} . \frac{s}{\sqrt{n}} = 87 \pm 1{,}96 . \frac{19}{\sqrt{1.000}} = 87 \pm 1{,}18$, een interval van **85,82** euro tot **88,18**

euro. Voor meisjes zien we $\bar{x} \pm z_{\alpha/2} . \frac{s}{\sqrt{n}} = 86 \pm 1{,}96 . \frac{19}{\sqrt{1.000}} = 86 \pm 1{,}24$, een interval van **84,76** euro tot **87,24** euro. Beide intervallen overlappen elkaar. Het lijkt er op dat jongens en meisjes niet significant verschillen in de gemiddelde besteding aan mobiele telefonie per maand.

b.

Stap 1

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Met α = 10%.

Stap 2

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{87 - 86}{\sqrt{\frac{19^2}{1.000} + \frac{20^2}{1.000}}} = 1{,}15 .$$

Stap 3

De ondergrens is **-1,645** en de bovengrens is **1,645**.

Stap 4

De conclusie: **-1,645 < 1,15 < 1,645,** je kunt niet beweren met 90% betrouwbaarheid dat er een verschil is tussen jongens en meisjes ten aanzien van de gemiddelde besteding per maand aan mobiele telefonie.

c. De p-waarde is 2 * P(z > 1,15) = 2 * 0,1251 = 0,2502 = **25,02%** > 10%, je kunt niet beweren met 90% betrouwbaarheid dat er een verschil is tussen jongens en meisjes ten aanzien van de gemiddelde besteding per maand aan mobiele telefonie.

Opgave 4

Stap 1

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Met α = 10%.

Stap 2

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{13{,}25 - 14{,}25}{\sqrt{\frac{2{,}36^2}{300} + \frac{2{,}32^2}{300}}} = -5{,}23 .$$

Stap 3

De ondergrens is **-1,645** en de bovengrens is **1,645**.

Stap 4

De conclusie: **-5,23 < -1,645**, je kunt met 90% betrouwbaarheid beweren dat de groep van 12 tot 18 jaar gemiddeld minder besteedt aan downloads dan de groep van 18 tot 24 jaar.

Competentieprikkel

a. Je moet een gepaarde toets gebruiken. Er zijn twee metingen verricht op dezelfde dag. Deze kun je twee aan twee met elkaar vergelijken.
b.

Stap 1

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

Met α = 10%.

Stap 2

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^{\ 2}}{n_1} + \frac{s_2^{\ 2}}{n_2}}} = \frac{1.320 - 1.280}{\sqrt{\frac{325^2}{230} + \frac{245^2}{230}}} = 1{,}49 .$$

Stap 3

De ondergrens is **-1,645** en de bovengrens is **1,645**.

Stap 4

De conclusie: **-1,645 < 1,49 < 1,645**, je kunt niet beweren met 90% betrouwbaarheid dat er een significant verschil is tussen de gemiddelde omzet per dag van de twee verkopers.

c. **Stap 1**

$H_0 : \mu_1 - \mu_2 = 0$
$H_1 : \mu_1 - \mu_2 \neq 0$

Met α = 10%.

Stap 2

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{40 - 0}{27 / \sqrt{230}} = 22{,}47 .$$

Stap 3

De ondergrens is **-1,645** en de bovengrens is **1,645**.

Stap 4
De conclusie: **22,47 > 1,645**, dus er is een significant verschil tussen verkoper 1 en verkoper 2 ten aanzien van de gemiddelde dagomzet.

d. Je neemt verkoper 1, hij verkoopt significant meer.

Uitwerkingen hoofdstuk 4

Opdrachten hoofdstuk 4

Opdracht 1

$\bar{x} + t_{\alpha;n-1} . \frac{s}{\sqrt{n}}$ = 62,33 + 1,729 * 13,88/√20 = 62,33 + 5,37 = 67,70. De bovengrens is 67,70 euro. Het interval is **[0, 67,70>.**

Opdracht 2

Bij 90% betrouwbaarheid hoort de tabelwaarde 1,328. De berekende t-waarde zit nog steeds niet in het kritieke gebied. De nulhypothese wordt ook nu niet verworpen.

Opdracht 3

Als s = 109,75, dan $t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{462{,}50 - 650}{109{,}75 * \sqrt{\frac{1}{4} + \frac{1}{6}}} = -2{,}647$**.**

Omdat **t = -2,647 < -2,306** wordt de nulhypothese nu wel verworpen. Er is een significant verschil gevonden tussen de twee provincies voor wat betreft de gemiddelde uitgave aan kinderkleding per huishouden per jaar.

Opdracht 4

De p-waarde = P(t > 3,800). Als we kijken in de t-tabel bij 5 vrijheidsgraden, dan P(t > 3,365) = 0,010 en P(t > 4,032) = 0,005. Dit betekent dat 0,005 < P(t > 3,800) < 0,010. Met de GRM vind je dat **P(t > 3,800) = 0,006**.

Kennisopgaven hoofdstuk 4

Opgave 1

Je spreekt van kleine steekproefaantallen als n < 200; je spreekt van grote steekproefaantallen als n ≥ 200.

Opgave 2

Gegevens dienen afkomstig te zijn uit de normale verdeling, als voorwaarde voor de t-toets op één gemiddelde.

Opgave 3

De t-verdeling kent de rechteroppervlakte en het aantal vrijheidsgraden als ingang.

Opgave 4

De gepoolde standaarddeviatie is de wortel van een combinatie van twee steekproefvarianties. Als voorwaarde geldt dat $\sigma_1 = \sigma_2$.

Opgave 5

a. Het aantal vrijheidsgraden bedraagt n – 1.
b. Het aantal vrijheidsgraden bedraagt $n_1 + n_2 - 2$.

Toegepaste Opgaven hoofdstuk 4

Opgave 1

a. $\bar{x} \pm t_{\alpha/2;n-1} . \frac{s}{\sqrt{n}} = 31{,}50 \pm 2{,}262 * \frac{17{,}33}{\sqrt{10}} = 31{,}50 \pm 12{,}40$.

Het 95% betrouwbaarheidsinterval loopt van **19,10** euro tot aan **43,90** euro. Je vindt de tabelwaarde met de ingangen α/2 = 0,025 en het aantal vrijheidsgraden n – 1 = 10 – 1 = 9.

b.

<u>Stap 1</u>

$H_0 : \mu = 30$

$H_1 : \mu \neq 30$

Met α = 10%.

Stap 2

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{31{,}50 - 30}{17{,}33/\sqrt{10}} = 0{,}274 .$$

Stap 3

Bij een rechteroppervlakte van α/2 = 0,05 en het aantal vrijheidsgraden n – 1 = 10 – 1 = 9 vind je de tabelwaarde 1,833. De ondergrens is **-1,833** en de bovengrens **1,833**.

Stap 4

Conclusie: t = **0,274 < 1,833**, de gemiddelde besteding aan biologische voeding per persoon per maand niet significant verschilt van 30 euro.

Opgave 2

a. $\bar{x} \pm t_{\alpha/2;n-1} . \frac{s}{\sqrt{n}} = 473{,}75 \pm 2{,}861 * \frac{167{,}06}{\sqrt{20}} = 473{,}75 \pm 106{,}87 .$

Het 99% betrouwbaarheidsinterval loopt van **366,80** euro tot aan **580,62** euro. Je vindt de tabelwaarde met de ingangen α/2 = 0,005 en het aantal vrijheidsgraden n – 1 = 20 – 1 = 19.

b.

Stap 1

$H_0 : \mu = 450$

$H_1 : \mu > 450$

Met α = 10%.

Stap 2

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{473{,}75 - 450}{167{,}06/\sqrt{20}} = 0{,}636 .$$

Stap 3

Bij een rechteroppervlakte van α = 0,10 en het aantal vrijheidsgraden n – 1 = 20 – 1 = 19 vind je de tabelwaarde 1,328. De ondergrens is **-1,328** en de bovengrens **1,328**.

Stap 4

Conclusie: t = **0,636 < 1,328**, de gemiddelde bijverdienste per maand van een student is niet significant groter dan 450 euro.

c. De p-waarde P(t > 0,636) ligt tussen de 10% en 50% in als je kijkt naar bijlage 2. Via de GRM vind je dat de p-waarde 0,2662 = **26,62%** is.

Opgave 3

a.
Stap 1

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

Met α = 5%.

Stap 2

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{122{,}17 - 124{,}71}{3{,}25\sqrt{\frac{1}{6} + \frac{1}{7}}} = -1{,}405.$$

Hierbij is $s^2 = \frac{(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2}}{n_1 + n_2 - 2} = \frac{(6-1)*3{,}06^2 + (7-1)*3{,}40^2}{6+7-2}$, je vindt s = 3,25.

Stap 3

- α/2 = 0,025
- Het aantal vrijheidsgraden is 6 + 7 – 2 = 11
- De tabelwaarde is 2,201

De ondergrens is **-2,201** en de bovengrens is **2,201**.

Stap 4

De conclusie: **-2,201 < -1,405 < 2,201,** je kunt niet beweren met 95% betrouwbaarheid dat het gemiddeld aantal auto's per dag dat van productielijn 1 en 2 rolt, verschillend is.

b. De voorwaarden zijn dat de gegevens afkomstig dienen te zijn uit de normale verdeling en dat de populatiestandaarddeviaties aan elkaar gelijk zijn.

Opgave 4

Stap 1

$H_0 : \mu_{na} = \mu_{voor}$
$H_1 : \mu_{na} > \mu_{voor}$

Met α = 10%.

Stap 2

Je berekent het verschil tussen na en voor:

2
0
5
0
2
2
10
0
3
0
5
5

Dan bepaal je het steekproefgemiddelde $\bar{x}$ = 2,833 en de steekproefstandaarddeviatie s = 3,010. Daarna komt

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{2{,}833 - 0}{3{,}010 / \sqrt{12}} = 3{,}257$$

Stap 3

- α = 0,10
- n – 1 = 12 – 1 = 11

Je vindt de tabelwaarde **1,363**.

Stap 4

Conclusie: **3,257 > 1,363**, er is sprake van een significant effect van de reclamecampagne van Axe.

Opgave 5

a. $\bar{x} \pm t_{\alpha/2;n-1} . \frac{s}{\sqrt{n}} = 225 \pm 1{,}796 * \frac{75}{\sqrt{12}} = 225 \pm 38{,}88$.

Het 90% betrouwbaarheidsinterval loopt van **186,12** euro tot aan **263,88** euro. Je vindt de tabelwaarde met de ingangen α/2 = 0,05 en het aantal vrijheidsgraden n – 1 = 12 – 1 = 11.

b.

<u>Stap 1</u>

$H_0 : \mu = 200$

$H_1 : \mu > 200$

Met α = 5%.

Stap 2

$$t = \frac{\overline{x} - \mu}{s / \sqrt{n}} = \frac{225 - 200}{75 / \sqrt{12}} = 1{,}155 .$$

Stap 3

Bij een rechteroppervlakte van α = 0,05 en het aantal vrijheidsgraden n – 1 = 12 – 1 = 11 vind je de tabelwaarde 1,796. De ondergrens is **-1,796** en de bovengrens **1,796**.

Stap 4

Conclusie: t = **1,115 < 1,796**, de gemiddelde online uitgave aan kleding per kwartaal is niet significant groter dan 200 euro.

Opgave 6

a.

Stap 1

$$H_0 : \mu_{2012} = \mu_{2013}$$
$$H_1 : \mu_{2012} > \mu_{2013}$$

Met α = 5%.

Stap 2

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{7.000 - 6.800}{1.200\sqrt{\frac{1}{13} + \frac{1}{10}}} = 0{,}396 .$$

Hierbij is $s^2 = \frac{(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2}}{n_1 + n_2 - 2} = \frac{(13-1)*1.200^2 + (10-1)*1.200^2}{13 + 10 - 2}$, je vindt s = 1.200 euro.

Stap 3

Bij een rechteroppervlakte van α = 0,05 en het aantal vrijheidsgraden 13 + 10 – 2 = 21 vind je de tabelwaarde 1,721. De bovengrens **1,721**.

Stap 4

Conclusie: t = **0,396 < 1,721**, de gemiddelde uitgave aan zonnepanelen in 2012 is niet significant groter dan die in 2013.

b.

$n \geq \frac{z_{\alpha/2}.s^2}{F^2} = \frac{1{,}96^2 * 1.200^2}{100^2} = 553{,}19$, dus er hadden minimaal **554** huishoudens betrokken moet zijn.

Competentieprikkel

a. Voor: $\bar{x}_1 \pm t_{\alpha/2;n_1-1}.\frac{s_1}{\sqrt{n_1}} = 7{,}50 \pm 2{,}262 * \frac{2{,}76}{\sqrt{10}} = 7{,}50 \pm 1{,}97$. Interval <**5,53; 9,47**>.

Na: $\bar{x}_2 \pm t_{\alpha/2;n_2-1}.\frac{s_2}{\sqrt{n_2}} = 5{,}86 \pm 2{,}447 * \frac{2{,}54}{\sqrt{7}} = 5{,}86 \pm 2{,}35$. Interval <**3,51; 8,21**>.

b.

Stap 1

$H_0 : \mu_{na} = \mu_{voor}$

$H_1 : \mu_{na} > \mu_{voor}$

Met α = 5%.

Stap 2

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{7{,}50 - 5{,}86}{2{,}67\sqrt{\frac{1}{10} + \frac{1}{7}}} = 1{,}246 .$$

Hierbij is $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(10-1) * 2{,}76^2 + (7-1) * 2{,}54^2}{10 + 7 - 2} = 7{,}15$ en s = 2,67.

Stap 3

- α = 0,05
- Het aantal vrijheidsgraden is 10 + 7 – 2 = 15
- De tabelwaarde is **1,753**.

Stap 4

De conclusie: 1**,246 < 1,753**, er is geen significante vermindering van de gemiddelde besteding aan dit product per persoon per week.

c. P(t > 1,246) > P(t > 1,3461) = 0,10. De p-waarde ligt tussen de 10% en 50%. Met de GRM vind je 0,1159 = **11,59%.**

d. Er had beter gebruik gemaakt kunnen worden van een gepaarde proefopzet. In dat geval is het persoonseffect uitgeschakeld. Nu verstoren de persoonsverschillen een mogelijke significantie.

Uitwerkingen hoofdstuk 5

Opdrachten hoofdstuk 5

Opdracht 1

De steekproefstandaarddeviatie is **s = 18,58** euro.

Opdracht 2

Bij een betrouwbaarheid van 99% en 6 vrijheidsgraden vind je de tabelwaarde **16,812**. De berekende waarde van de toetsingsgrootheid is 9,206. Deze is kleiner dan 16,812. Dit betekent dat je niet kunt beweren dat de populatiestandaarddeviatie significant groter is dan 15 euro per persoon per maand.

Opdracht 3

Bij α = 10% zijn de ingangen voor de F-tabel:

- $\alpha/2$ = 5% = 0,05;
- het aantal vrijheidsgraden van de teller is 10;
- het aantal vrijheidsgraden van de noemer is 8.

Dit geeft de tabelwaarde **3,35**. Omdat de berekende toetsingsgrootheid F = 3,59, kun je zien dat de populatiestandaarddeviaties significant verschillen.

Kennisopgaven

Opgave 1

De gegevens dienen afkomstig te zijn uit een normale verdeling.

Opgave 2

Het aantal vrijheidsgraden is **n – 1**.

Opgave 3

Dit wordt gedaan om het tabelgebruik te vereenvoudigen, zodoende heb je alleen de rechtergrens uit de F-verdeling nodig.

Opgave 4

Het aantal vrijheidsgraden van de teller is $n_1 - 1$ en het aantal vrijheidsgraden van de noemer is $n_2 - 1$.

Toegepaste Opgaven hoofdstuk 5

Opgave 1

Stap 1

$H_0 : \sigma = 80.000$
$H_1 : \sigma < 80.000$

Hierbij is α = 5%.

Stap 2

Je berekent de chikwadraatwaarde:

$$\chi^2 = \frac{(n-1).s^2}{\sigma^2} = \frac{(20-1)*72.578^2}{80.000^2} = 15{,}638$$

Stap 3

De ingangen voor de tabel zijn:

- α = 5% = 0,05;
- n – 1 = 20 – 1 = 19 is het aantal vrijheidsgraden.

De ondergrens is **10,117**.

Stap 4

Conclusie: de berekende chikwadraatwaarde **15,638** is groter dan de ondergrens **10,117**, hetgeen betekent dat de populatiestandaarddeviatie niet significant kleiner is dan 80.000 euro.

Opgave 2

a.
Stap 1

$H_0 : \sigma = 80$
$H_1 : \sigma \neq 80$

Hierbij is α = 1%.

Stap 2

Je berekent de chikwadraatwaarde:

$$\chi^2 = \frac{(n-1).s^2}{\sigma^2} = \frac{(10-1)*176{,}86^2}{80^2} = 43{,}987$$

Stap 3

De ingangen voor de tabel zijn:

- α/2 = 1%/2 = 0,005;
- n – 1 = 10 – 1 = 9 is het aantal vrijheidsgraden.

De bovengrens is **23,589**, de ondergrens is **1,735**.

Stap 4

Conclusie: de berekende chikwadraatwaarde **43,987** is groter dan de bovengrens **23,589**, hetgeen betekent dat de populatiestandaarddeviatie significant veranderd is.

b. $\overline{x} \pm t_{\alpha/2;n-1}.\frac{s}{\sqrt{n}} = 304 \pm 1{,}833 * \frac{176{,}86}{\sqrt{10}} = 304 \pm 102{,}52$.

Het 90% betrouwbaarheidsinterval loopt van **201,48** euro tot **406,52** euro.

Opgave 3

a.

Stap 1

$H_0 : \sigma_1 = \sigma_2$

$H_1 : \sigma_1 \neq \sigma_2$

Hierbij is α = 10%.

Stap 2

Je berekent de F-waarde met de grootste standaarddeviatie in de teller:

$$F = \frac{{s_1}^2}{{s_2}^2} = \frac{1{,}4^2}{1{,}2^2} = 1{,}36$$

Stap 3

De ingangen voor de tabel zijn:

- α/2 = 10%/2 = 5% = 0,05;

- $n_1 - 1$ = 11 – 1 = 10 is het aantal vrijheidsgraden van de teller.
- $n_2 - 1$ = 10 – 1 = 9 is het aantal vrijheidsgraden van de noemer.

De tabelwaarde is **3,14**.

Stap 4

Conclusie: de berekende F-waarde 1,36 is kleiner dan de bovengrens 3,14, hetgeen betekent dat de populatiestandaarddeviaties niet significant verschillen.

b.
Stap 1

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

Met α = 1%.

Stap 2

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{8-7}{1{,}309\sqrt{\frac{1}{10} + \frac{1}{11}}} = 1{,}748.$$

Hierbij is $s^2 = \frac{(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2}}{n_1 + n_2 - 2} = \frac{(10-1)*1{,}2^2 + (11-1)*1{,}4^2}{10+11-2}$ = 1,714, je vindt s = 1,309.

Stap 3

- α/2 = 0,005;
- het aantal vrijheidsgraden is 10 + 11 – 2 = 19;
- de tabelwaarde is **2,861**.

De ondergrens is **-2,861** en de bovengrens is **2,861**.

Stap 4

De conclusie: **-2,861 < 1,748 < 2,861**, je kunt niet beweren met 99% betrouwbaarheid dat er een verschil is in het gemiddeld aantal invulfouten in Venlo en Utrecht.

Opgave 4

a.
Stap 1

$H_0 : \sigma_1 = 35$
$H_1 : \sigma_1 > 35$

Hierbij is α = 5%.

Stap 2

Je berekent de chikwadraatwaarde:

$$\chi^2 = \frac{(n-1).s^2}{\sigma^2} = \frac{(10-1)*40^2}{35^2} = 11{,}755$$

Stap 3

De ingangen voor de tabel zijn:

- α = 5% = 0,05;
- n – 1 = 10 – 1 = 9 is het aantal vrijheidsgraden.

De bovengrens is **16,919**.

Stap 4

Conclusie: de berekende chikwadraatwaarde **11,755** is kleiner dan de bovengrens **16,919**, hetgeen betekent dat de populatiestandaarddeviatie van de groep van 16 tot 25 jaar niet significant groter is dan 35 euro.

b.
Stap 1

$H_0 : \sigma_1 = \sigma_2$
$H_1 : \sigma_1 \neq \sigma_2$

Hierbij is α = 10%.

Stap 2

Je berekent de F-waarde met de grootste standaarddeviatie in de teller:

$$F = \frac{s_1^{\ 2}}{s_2^{\ 2}} = \frac{45^2}{40^2} = 1{,}27$$

Stap 3

De ingangen voor de tabel zijn:

- α/2 = 10%/2 = 5% = 0,05;
- $n_1 - 1$ = 10 – 1 = 9 is het aantal vrijheidsgraden van de teller.
- $n_2 - 1$ = 10 – 1 = 9 is het aantal vrijheidsgraden van de noemer.

De tabelwaarde is **3,18**.

Stap 4

Conclusie: de berekende F-waarde 1,27 is kleiner dan de bovengrens 3,18, hetgeen betekent dat de populatiestandaarddeviaties niet significant verschillen.

c.
Stap 1

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 < \mu_2$

Met α = 10%.

Stap 2

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{125 - 155}{42{,}573\sqrt{\frac{1}{10} + \frac{1}{10}}} = -1{,}576.$$

Hierbij is $s^2 = \frac{(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2}}{n_1 + n_2 - 2} = \frac{(10-1)*40^2 + (10-1)*45^2}{10 + 10 - 2}$ = 1.812,500, je vindt s = 42,573.

Stap 3

- α = 10% = 0,10;
- het aantal vrijheidsgraden is 10 + 10 – 2 = 18;
- de tabelwaarde is **1,330**.

De ondergrens is **-1,330**.

Stap 4

De conclusie: **-1,576 < -1,330**, je kunt beweren met 90% betrouwbaarheid dat de gemiddelde uitgave aan tablets in de leeftijdsgroep van 25 tot 40 jaar significant groter is dan die in de leeftijdsgroep van 16 tot 25 jaar!

Competentieprikkel

a.
Stap 1

$H_0 : \sigma = 25$
$H_1 : \sigma > 25$

Hierbij is α = 5%.

Stap 2

Je berekent de chikwadraatwaarde:

$$\chi^2 = \frac{(n-1).s^2}{\sigma^2} = \frac{(13-1)*30^2}{25^2} = 17{,}280$$

Stap 3

De ingangen voor de tabel zijn:

- α = 5% = 0,05;
- n – 1 = 13 – 1 = 12 is het aantal vrijheidsgraden.

De bovengrens is **21,026**.

Stap 4

Conclusie: de berekende chikwadraatwaarde **17,280** is kleiner dan de bovengrens **21,026**, hetgeen betekent dat de populatiestandaarddeviatie niet significant groter is dan 25 dollar.

b.
Stap 1

$H_0 : \sigma_1 = \sigma_2$
$H_1 : \sigma_1 \neq \sigma_2$

Hierbij is α = 10%.

Stap 2

Je berekent de F-waarde met de grootste standaarddeviatie in de teller:

$$F = \frac{s_1^{\ 2}}{s_2^{\ 2}} = \frac{35^2}{30^2} = 1{,}36$$

Stap 3

De ingangen voor de tabel zijn:

- α/2 = 10%/2 = 5% = 0,05;
- $n_1 - 1$ = 13 – 1 = 12 is het aantal vrijheidsgraden van de teller.
- $n_2 - 1$ = 13 – 1 = 12 is het aantal vrijheidsgraden van de noemer.

De tabelwaarde is **2,69**.

Stap 4

Conclusie: de berekende F-waarde **1,36** is kleiner dan de bovengrens **2,69**, hetgeen betekent dat de populatiestandaarddeviaties niet significant verschillen.

c.
Stap 1

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

Met α = 10%.

Stap 2

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{120 - 150}{32{,}596\sqrt{\frac{1}{13} + \frac{1}{13}}} = -2{,}346.$$

Hierbij is $s^2 = \frac{(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2}}{n_1 + n_2 - 2} = \frac{(13-1)*30^2 + (13-1)*30^2}{13+13-2}$ = 1.062,50, je vindt s = 32,596.

Stap 3

- α/2 = 10%/2 = 5% = 0,05;
- het aantal vrijheidsgraden is 13 + 13 – 2 = 24;
- de tabelwaarde is 1,711.

De ondergrens is **-1,711** en de bovengrens is **1,711**.

Stap 4

De conclusie: **-2,246 < -1,711**, er is dus een significant verschil tussen de gemiddelde besteding aan G-STAR jeans per jaar door jongeren in Nederland en Zweden.

Uitwerkingen hoofdstuk 6

Opdrachten hoofdstuk 6

Opdracht 1

$\hat{p} - z_{\alpha} \cdot \sqrt{\frac{\hat{p}.(1-\hat{p})}{n}} = 70\% - 1{,}28 * \sqrt{\frac{70\%.30\%}{1.000}} = 70\% - 1{,}85\% = 68{,}15\%$, dus minimaal **68,15%** jongeren heeft een bijbaantje.

Opdracht 2

- Neem de tabel van bijlage 7 bij de hand.
- Kijk in de eerste rij bij 95% eenzijdig betrouwbaarheidsinterval.
- Kijk in de eerste kolom bij k = 2 (niet correcte posten).
- Je vindt de bovengrens 6,30.
- Deel deze waarde door het steekproefaantal n = 800 en vermenigvuldig met 100%: 6,30/800 * 100% = 78,75%.
- Je hebt het 95% betrouwbaarheidsinterval bepaald.

Met 95% betrouwbaarheid kun je concluderen dat het percentage niet correcte posten maximaal **78,75%** is.

Opdracht 3

$n \geq \frac{z_{\alpha/2}{}^{2}.p.(1-p)}{F^{2}} = \frac{1{,}96^{2}.50\%.50\%}{3\%^{2}} = 1.067{,}11$. De minimale steekproefgrootte bedraagt **1.068** personen.

Opdracht 4

$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}.(1-\hat{p})}{n}} . \sqrt{\frac{N-n}{N-1}} = 40\% \pm 1{,}645 . \sqrt{\frac{40\%.60\%}{200}} . \sqrt{\frac{1.500-200}{1.500-1}} = 40\% \pm 5{,}31\%$. De foutmarge wordt groter. Bij een steeds grotere populatie wordt de foutmarge bij een vast staande steekproefgrootte groter.

Opdracht 5

De eindigheidscorrectie wordt toegepast als n > 0,1N. Op het moment dat N groot is, ligt het voor de hand om de formule zonder eindigheidscorrectie voor de minimale steekproefgrootte te gebruiken. Je krijgt dan:

$n \geq \frac{z_{\alpha/2}{}^2.p.(1-p)}{F^2} = \frac{1{,}96^2.50\%.50\%}{3\%^2} = 1.067{,}11$, dus **1.068** werknemers in plaats van de gevonden 697 werknemers.

Kennisopgaven hoofdstuk 6

Opgave 1

Een steekproefpercentage wordt berekend op basis van een steekproef en een populatiepercentage op basis vaneen populatie. Het populatiepercentage is het percentage dat je graag wilt weten, maar in de praktijk ga je vanwege tijd en geld een steekproef. Je schat dan het populatiepercentage met het steekproefpercentage.

Opgave 2

Je gebruikt de eindigheidscorrectie als n > 0,1N.

Opgave 3

De standaardfout is $\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}}$. Dit is de standaarddeviatie van het steekproefpercentage. Om verwarring te voorkomen spreek je van de standaardfout.

Opgave 4

Je gebruikt een betrouwbaarheidsinterval op basis van de Poissonverdeling als je bijlage 7 kunt toepassen. Anders gebruik je het betrouwbaarheidsinterval op basis van de z-verdeling.

Opgave 5

De minimale steekproefgrootte voor een percentage bij een zeer grote populatie wordt bepaald door:

- De betrouwbaarheid.
- Het populatiepercentage.
- De foutmarge.

De minimale steekproefgrootte voor een percentage bij een beperkt eindige populatie wordt bepaald door:

- De populatiegrootte.
- De betrouwbaarheid.
- Het populatiepercentage.
- De foutmarge.

Opgave 6

Je vult p = 50% in, dit is de waarde waarbij p(1 – p) maximaal is.

Toegepaste opgaven hoofdstuk 6

Opgave 1

a. $\hat{p} \pm z_{\alpha/2}.\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}} = 12\% \pm 1{,}645.\sqrt{\frac{12\%.88\%}{400}} = 12\% \pm 2{,}67\%$. Het 90% betrouwbaarheidsinterval voor het percentage mensen dat zich het spotje spontaan kan herinneren loopt van **9,33%** tot **14,67%.**

b. $n \geq \frac{z_{\alpha/2}{}^2.p.(1-p)}{F^2} = \frac{1{,}96^2.50\%.50\%}{4\%^2} = 600{,}25$. De minimale steekproefgrootte bedraagt **601** respondenten.

Opgave 2

a. $n \geq \frac{z_{\alpha/2}{}^2.p.(1-p)}{F^2} = \frac{1{,}645^2.50\%.50\%}{4{,}5\%^2} = 334{,}08$. De minimale steekproefgrootte bedraagt **335** huishoudens.

b. $n \geq \frac{N.z_{\alpha/2}{}^2.p.(1-p)}{z_{\alpha/2}{}^2.p.(1-p)+(N-1).F^2} = \frac{1.800.1{,}96^2.10\%.90\%}{1{,}96^2.10\%.90\%+1.799.2{,}5\%^2} = 423{,}33$. De minimale steekproefgrootte bedraagt **424** huishoudens.

Opgave 3

a. $\hat{p} \pm z_{\alpha/2}.\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}} = 30\% \pm 1{,}96.\sqrt{\frac{30\%.70\%}{500}} = 30\% \pm 4{,}02\%$. Het 95% betrouwbaarheidsinterval voor het percentage personen dat voorstander is van een verbod van alcoholreclame op televisie varieert van **25,98%** tot **34,02%**.

b. $\hat{p} \pm z_{\alpha/2}.\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}}.\sqrt{\frac{N-n}{N-1}} = 30\% \pm 1{,}96.\sqrt{\frac{30\%.70\%}{500}}.\sqrt{\frac{2.500-500}{2.500-1}} = 30\% \pm 3{,}59\%$. Het 95% betrouwbaarheidsinterval voor het percentage personen dat voorstander is van een verbod van alcoholreclame op televisie varieert van **26,41%** tot **33,59%**.

Opgave 4

a. Je doorloopt de volgende stappen:

- Neem de tabel van bijlage 7 bij de hand.
- Kijk in de eerste rij bij 95% tweezijdig betrouwbaarheidsinterval.
- Kijk in de eerste kolom bij k = 4 (niet correcte posten).
- Je vindt 1,09 en 10,24.
- Deel deze waarde door het steekproefaantal n = 400 en vermenigvuldig met 100%: 1,09/400 * 100% = 0,27% en 10,24/400 * 100% = 2,56%.
- Je hebt het 95% betrouwbaarheidsinterval bepaald.

Met 95% betrouwbaarheid kun je concluderen dat het percentage niet correcte posten varieert van **0,27%** tot **2,56%.**

b. Analoog als bij onderdeel a: Neem de tabel van bijlage 7 bij de hand.

- Kijk in de eerste rij bij 99% eenzijdig betrouwbaarheidsinterval.
- Kijk in de eerste kolom bij k = 4 (niet correcte posten).
- Je vindt de bovengrens 11,61.
- Deel deze waarde door het steekproefaantal n = 400 en vermenigvuldig met 100%: 11,61/400 * 100% = 2,90%.
- Je hebt het 99% betrouwbaarheidsinterval bepaald.

Met 99% betrouwbaarheid kun je concluderen dat het percentage niet correcte posten maximaal **2,90%** is.

Opgave 5

a. Je doorloopt de volgende stappen:

- Neem de tabel van bijlage 7 bij de hand.
- Kijk in de eerste rij bij 95% eenzijdig betrouwbaarheidsinterval.
- Kijk in de eerste kolom bij k = 1 (ondeugdelijke kabelboom).
- Je vindt de ondergrens 0,05.
- Deel deze waarde door het steekproefaantal n = 1.000 en vermenigvuldig met 100%: 0,05/1.000 * 100% = 0,005%
- Je hebt het 95% betrouwbaarheidsinterval bepaald.

Met 95% betrouwbaarheid kun je concluderen dat het percentage ondeugdelijke kabelbomen minimaal **0,005%** bedraagt.

b. Allereerst bepaal je de bovengrens voor het percentage ondeugdelijke kabelbomen:

- Neem de tabel van bijlage 7 bij de hand.
- Kijk in de eerste rij bij 95% tweezijdig betrouwbaarheidsinterval.
- Kijk in de eerste kolom bij k = 1 (niet correcte posten).
- Je vindt de bovengrens 4,74.

- Deel deze waarde door het steekproefaantal n = 1.000 en vermenigvuldig met 100%: 4,74/1.000 * 100% = 0,47%.

  Vermenigvuldig nu dit percentage met 100.000 kabelbomen en met 20 euro:
  Het maximale verlies bedraagt **0,47% * 100.000 * 20 = 9.400 euro**.

Competentieprikkel

a. $\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}.(1-\hat{p})}{n}} = 17\% \pm 2{,}576.\sqrt{\frac{17\%.83\%}{10.000}} = 17\% \pm 0.97\%$. Het 99% betrouwbaarheidsinterval voor het percentage mensen dat zegt wel vertrouwen te hebben in de kwaliteit van mobiele telefonie varieert van **16,03%** tot **17,97%**.

b. De $\hat{p} + z_{\alpha} \cdot \sqrt{\frac{\hat{p}.(1-\hat{p})}{n}} = 25\% + 1{,}645.\sqrt{\frac{25\%.75\%}{3.100}} = 25\% + 1{,}28\%$. Je vindt de 3.100 door 31% te nemen van 10.000! De gevraagde bovengrens bedraagt **26,28%.**

c. De $n \geq \frac{z_{\alpha/2}{}^2.p.(1-p)}{F^2} = \frac{1{,}96^2.50\%.50\%}{1\%^2} = 9.604$. De minimale steekproefgrootte bedraagt **9.604** huishoudens.

Uitwerkingen hoofdstuk 7

Opdrachten hoofdstuk 7

Opdracht 1

De p-waarde is P(z > 1,94) = 0,0262. Je vergelijkt de p-waarde met de onbetrouwbaarheid 0,05. Je ziet dat **0,0262 < 0,05**. Dat betekent dat de nulhypothese verworpen wordt. Het percentage huishoudens dat digitale televisie heeft is significant hoger dan 27,5%!

Opdracht 2

P(k ≤ 2) bij μ = 3,5 in de Poisson verdeling levert op **0,3208**. Dit is groter dan 0,05. Je verwerpt de nulhypothese niet, de verantwoording wordt nader onderzocht.

Opdracht 3

$n \geq \frac{\log \alpha}{\log(1-p)} = \frac{\log 0{,}01}{\log 0{,}99} = 458{,}21$, dus n = **459** posten.

Opdracht 4

De p-waarde is P(z > 0,72) = **0,2358 > 0,10**. Verwerp de nulhypothese niet, er is geen bewijsmateriaal dat de spontane naamsbekendheid van Marga significant is toegenomen.

Opdracht 5

Stap 1

$$H_0 : p_1 - p_2 = 25\%$$
$$H_1 : p_1 - p_2 \neq 25\%$$

Met α = 5%.

Stap 2

$$z = \frac{\hat{p}_1 - \hat{p}_2 - verschil}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} = \frac{30{,}8\% - 4{,}0\% - 23\%}{\sqrt{\frac{30{,}8\% * 69{,}2\%}{750} + \frac{4{,}0\% * 96{,}0\%}{750}}} = 2{,}07$$

Stap 3

De ondergrens is **-1,645** en de bovengrens **1,645**.

Stap 4

Je ziet dat **2,07 > 1,645**. De conclusie luidt dat het procentuele verschil tussen mannen en vrouwen significant verschilt van 23% ten aanzien van het kijken naar de ondernemerszender.

Kennisopgaven hoofdstuk 7

Opgave 1

Je gebruikt de z-benadering als np ≥ 5 en n(1 – p) ≥ 5. Als np < 5 of n(1 – p) < 5, dan gebruik je de Poisson benadering.

Opgave 2

De onbetrouwbaarheidsdrempel is de maximale waarde voor de onbetrouwbaarheid die je acceptabel vindt. Bij een toetsingsgrootheden met een continue steekproefverdeling geldt dat de onbetrouwbaarheidsdrempel en de onbetrouwbaarheid samenvallen. Bij toetsingsgrootheden met een discrete steekproefverdelingen geldt dat de onbetrouwbaarheid maximaal gelijk is aan de onbetrouwbaarheidsdrempel.

Opgave 3

Als alle posten correct zijn, wordt de verantwoording goedgekeurd. Als minimaal één post niet correct is, dan wordt de verantwoording niet goedgekeurd en nader onderzocht.

Opgave 4

Een gepoold steekproefpercentage is een gewogen gemiddelde van twee percentages uit twee steekproeven.

Opgave 5

Toetsen op een verondersteld procentueel verschil:

$H_0 : p_1 - p_2 = verschil$
$H_1 : p_1 - p_2 \neq verschil$

Toetsen op gelijkheid van twee populatiepercentages:

$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$

Opgave 6

Alle verwachte frequenties moeten minimaal vijf zijn.

Toegepaste opgaven hoofdstuk 7

Opgave 1
a.
Stap 1

H_0: p ≤ 25%
H_1: p > 25%

Verder is $\alpha = 1\%$.
Stap 2

- In de steekproef is n = 500 bedrijven.
- Verder is k = 33,33% * 500 = 167.
- Onder de nulhypothese is p = 25% en np = 500 * 25% = 125 ≥ 5.
- 1- p = 1 – 0,25 = 0,75 = 75%.

Invullen levert op:

$$z = \frac{k - np}{\sqrt{np(1-p)}} = \frac{167 - 125}{\sqrt{500 * 25\% * 75\%}} = 4{,}34$$

Stap 3

We controleren de voorwaarden van toepassing van de binomiale verdeling:

- n = 500 bedrijven. Er is sprake van herhaling.
- Er zijn twee uitkomsten: wel of geen voorstander van meer vakantiedagen.
- De succeskans p is constant, omdat de totale populatie alle Nederlandse bedrijven betreft en dus zeer groot is.

We controleren de voorwaarden om de normale verdeling toe te passen als benadering:

- np = 500 * 25% = 125 ≥ 5.
- n(1 – p) = 500 * 75% = 375 ≥ 5.

Er is voldaan aan de voorwaarden van de benadering met de normale verdeling. Dit betekent dat de toetsingsgrootheid

$$z = \frac{k - np}{\sqrt{np(1-p)}}$$

standaard normaal verdeeld is. De gevraagde bovengrens is **2,33**.

Stap 4

Omdat **4,34 groter is dan 2,33**, verwerp je de nulhypothese. Je hebt aangetoond dat er een significant groter percentage dan 25% van de Nederlandse bedrijven meer vakantiedagen wil voor zijn werknemers.
b. De p-waarde is P(z > 4,34) = 0% < 1%. Verwerp de nulhypothese. Het percentage Nederlandse bedrijven dat meer vakantiedagen wil voor zijn werknemers is significant hoger dan 25%.

Opgave 2

a**. Stap 1**

H_0: p ≤ 2,25%
H_1: p > 2,25%
We werken met α = 5%.

Stap 2

De toetsingsgrootheid is het aantal bedrijven dat meedoet aan de levensloopregeling.

Uit de gegevens blijkt dat k = 8,5% * 200 = 17 zijn.

Stap 3

De toetsingsgrootheid is binomiaal verdeeld, want

- je werkt met herhaling, je onderzoekt n = 200 bedrijven;
- er zijn twee uitkomstmogelijkheden, wel of niet meedoen met de levensloopregeling;
- het percentage bedrijven dat meedoet met de levensloopregeling is constant omdat we te maken hebben met zeer veel bedrijven.

Zodoende is het aantal bedrijven dat meedoet aan de levensloopregeling binomiaal verdeeld met n = 200 en onder de nulhypothese gebruiken we p = 2,25% als uiterste waarde om de toets uit te kunnen voeren.

Omdat n groot is, kijken we of we een benadering toe mogen passen. We controleren de voorwaarden van de benadering van de binomiale verdeling met de normale verdeling:

- np = 200 * 2,25% = 4,5 < 5.
- n(1 – p) = 200 * 97,5% = 195,5 ≥ 5.

Er is niet voldaan aan de voorwaarden voor benadering met de normale verdeling! Gebruik nu de Poisson verdeling met μ = 4,5.

- Kijk in bijlage 6 bij μ = np = 4,5.
- P(k ≥ 9) = 1 – P(k ≤ 8) = 0,0403 < 0,05.
- P(k ≥ 8) = 1 – P(k ≤ 7) = 0,0866 > 0,05.

De bovengrens is **9**.

Stap 4

Omdat **17 > 9** wordt de nulhypothese verworpen. Het percentage bedrijven dat meedoet aan de levensloopregeling is significant hoger dan 2,25%.

b. . **Stap 1**

H_0 : p ≤ 2,25%
H_1 : p > 2,25%

We werken met α = 5%.

Stap 2

Uit de gegevens blijkt dat k = 8,5% * 800 = 68 is en np = 800 * 2,25% = 18 ≥ 5.

$$z = \frac{k - np}{\sqrt{np(1-p)}} = \frac{68 - 800 * 2{,}25\%}{\sqrt{800 * 2{,}25\% * 97{,}75\%}} = 11{,}92$$

Stap 3

De toetsingsgrootheid is binomiaal verdeeld, want

- je werkt met herhaling, je onderzoekt n = 800 bedrijven;
- er zijn twee uitkomstmogelijkheden, wel of niet meedoen met de levensloopregeling;
- het percentage bedrijven dat meedoet met de levensloopregeling is constant omdat we te maken hebben met zeer veel bedrijven.

Zodoende is het aantal bedrijven dat meedoet aan de levensloopregeling binomiaal verdeeld met n = 800 en onder de nulhypothese gebruiken we p = 2,25% als uiterste waarde om de toets uit te kunnen voeren.

Omdat n groot is, kijken we of we een benadering toe mogen passen. We controleren de voorwaarden van de benadering van de binomiale verdeling met de normale verdeling:

- np = 800 * 2,25% = 18 ≥ 5.
- n(1 – p) = 800 * 97,5% = 782 ≥ 5.
- Er is voldaan aan de voorwaarden voor benadering met de normale verdeling!

De bovengrens is 1,645.

Stap 4

Omdat **11,92 > 1,645** wordt de nulhypothese verworpen. Het percentage bedrijven dat meedoet aan de levensloopregeling is significant hoger dan 2,25%.

Opgave 3

a. $\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}.(1-\hat{p})}{n}} = 11{,}54\% \pm 1{,}645 . \sqrt{\frac{11{,}54\%.88{,}46\%}{260}} = 11{,}54\% \pm 3{,}26\%$. Het betrouwbaarheidsinterval voor het percentage 3-en loopt van **8,28% tot 14,80%**.

b.

Stap 1

$H_0 : p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9 = 10\%$

$H_1 :$ Minstens één percentage is niet gelijk aan 10%

Verder is $\alpha = 5\%$.

Stap 2

De toetsingsgrootheid die gebruikt wordt is:

$$X^2 = \sum \frac{(O-E)^2}{E}$$

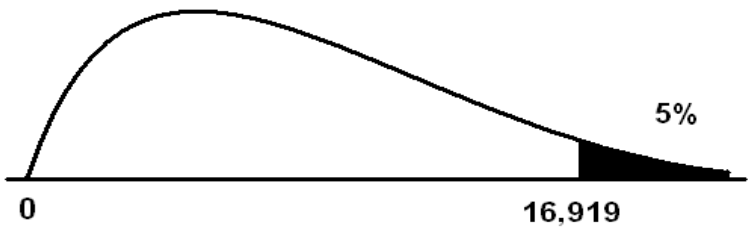
Cijfers	**O**	**E**	**O - E**	$(O-E)^2$	$\frac{(O-E)^2}{E}$
0	24	26	-2	4	0,1538
1	23	26	-3	9	0,3462
2	27	26	1	1	0,0385
3	30	26	4	16	0,6154
4	24	26	-2	4	0,1538
5	25	26	-1	1	0,0385
6	27	26	1	1	0,0385
7	26	26	0	0	0
8	27	26	1	1	0,0385
9	27	26	1	1	0,0385
Totaal	260	260	0		1,4617

Stap 3

Je vindt de tabelwaarde als volgt:

- Neem bijlage 3 bij de hand.
- Kijk bij de eerste tabel met een bovengrens.
- De eerste ingang is in de eerste rij $\alpha = 0{,}05$.
- De tweede ingang zijn de vrijheidsgraden vhg. Deze zijn gelijk aan het aantal percentages waarop getoetst wordt minus één. De vrijheidsgraden zijn 10 – 1 = 9.
- Je vindt de bovengrens 16,919.

Stap 4



Je ziet dat de berekende chikwadraatwaarde van **1,462** kleiner is dan de bovengrens **16,919**. De conclusie is dat je niet kunt aantonen dat er minstens één percentage significant afwijkt van 10%. De accountant kan niet beweren dat er frauduleuze handelingen gepleegd zijn.

De voorwaarde die hoort bij deze toets is dat alle verwachte frequenties E minimaal vijf zijn. Omdat alle verwachte frequenties 26 zijn, is hier aan voldaan.

Opgave 4

a. De hypothesen luiden $H_0 : p \geq 1{,}5\%$ tegenover $H_1 : p < 1{,}5\%$. De toetsingsgrootheid is $z = \frac{k - np}{\sqrt{np(1-p)}}$ met np = 1.000 * 1,5% ≥ 15 en n(1 – p) = 1.000 * 98,5% = 985 ≥ 5.
De ondergrens is z = **-2,33**.

b. De minimale steekproefgrootte is $n \geq \frac{\log \alpha}{\log(1-p)} = \frac{\log 0{,}01}{\log 0{,}985} = 304{,}7$, dus **305** posten.

c. $z = \frac{k - np}{\sqrt{np(1-p)}} = \frac{4 - 15}{\sqrt{1.000 * 1{,}5\% * 98{,}5\%}} = -2{,}86$ < -2,33. De nulhypothese wordt verworpen, de boekhouding wordt goedgekeurd.

Opgave 5

Stap 1

$H_0 : p_1 - p_2 = 30\%$
$H_1 : p_1 - p_2 \neq 30\%$

Verder is α = 5%.

Stap 2

$$z = \frac{\hat{p}_1 - \hat{p}_2 - verschil}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} = \frac{77{,}4\% - 26{,}8\% - 30\%}{\sqrt{\frac{77{,}4\% * 69{,}2\%}{500} + \frac{26{,}8\% * 73{,}2\%}{500}}} = 5{,}42$$

Stap 3

Bij een onbetrouwbaarheid van 5% gebruik je bij tweezijdig toetsen de z-waarden z = **-1,96** en z = **1,96**.

Stap 4

De berekende z-waarde **5,42 ligt boven 1,96**.Verwerp de nulhypothese. Je hebt aangetoond dat het procentuele verschil tussen mannen en vrouwen ten aanzien van minimaal drie maal winkelen per week significant verschilt van 30%.

Opgave 6

Stap 1

$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$

Bij deze toets is α = 1%.

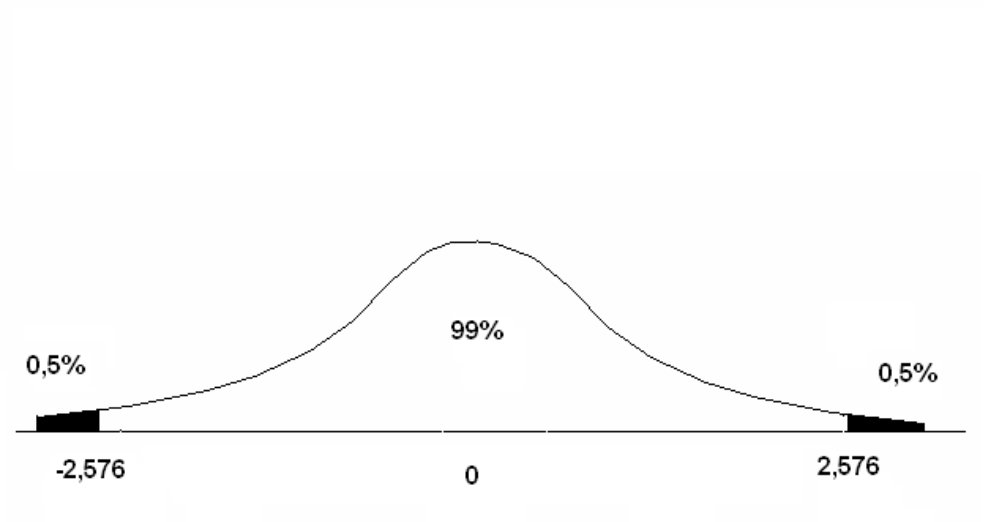
Stap 2

Je berekent allereerst het gepoolde percentage (78 + 60)/(300 + 300) = 138/600 = 23%. Daarna bereken je

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}.(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{26\% - 20\%}{\sqrt{23\% * 77\% * (\frac{1}{300}+\frac{1}{300})}} = 1{,}75$$

Stap 3

De hier geformuleerde toets is tweezijdig. Bij een α van 1% hoort de ondergrens **-2,576** en de bovengrens **2,576**.



Stap 4

Conclusie: de berekende z-waarde **1,75 is kleiner dan de bovengrens 2,576**, dus er is geen bewijsmateriaal gevonden dat het percentage jongeren van 8 tot 16 jaar dat minimaal één maal per maand Twix eet significant verschilt van het percentage jongeren van 16 tot 24 jaar dat minimaal één maal per maand Twix eet.

Competentieprikkel

a. $\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}.(1-\hat{p})}{n}} = 10\% \pm 2{,}576.\sqrt{\frac{10\%.90\%}{260}} = 10\% \pm 4{,}46\%$. Het betrouwbaarheidsinterval loopt van **5,54% tot 14,46%**.

b.

Stap 1

H_0 : p ≤ 8%
H_1 : p > 8%

We werken met α = 5%.

Stap 2

Uit de gegevens blijkt dat k = 30 is en np = 300 * 8% = 24 ≥ 5.

$$z = \frac{k - np}{\sqrt{np(1-p)}} = \frac{30 - 24}{\sqrt{300 * 8\% * 92\%}} = 1{,}28$$

Stap 3

De toetsingsgrootheid is binomiaal verdeeld, want

- je werkt met herhaling, je onderzoekt n = 300 adviseurs ;
- er zijn twee uitkomstmogelijkheden, wel of niet ernstige overtredingen hebben begaan;
- het percentage adviseurs dat ernstige overtredingen heeft begaan is constant omdat we veronderstellen te maken te hebben met veel adviseurs.

Zodoende is het aantal adviseurs dat ernstige overtredingen heeft begaan binomiaal verdeeld met n = 300 en onder de nulhypothese gebruiken we p = 8% als uiterste waarde om de toets uit te kunnen voeren.

Omdat n groot is, kijken we of we een benadering toe mogen passen. We controleren de voorwaarden van de benadering van de binomiale verdeling met de normale verdeling:

- np = 300 * 8% = 24 ≥ 5.
- n(1 – p) = 300 * 92% = 276 ≥ 5.
- Er is voldaan aan de voorwaarden voor benadering met de normale verdeling!

De bovengrens is **1,645**.

Stap 4

Omdat **1,28 < 1,645** wordt de nulhypothese niet verworpen. Je kunt niet met 95% betrouwbaarheid beweren dat het percentage financiële adviseurs hoger is dan 8%.

c. De toets verloopt analoog als bij onderdeel b, maar nu dient een eindigheidscorrectie uitgevoerd te worden omdat n = 1.100 = 0,10 * 11.000 = 0,10N. De nieuwe z-waarde wordt dan $z = \frac{k - np}{\sqrt{np(1-p)} * \sqrt{\frac{N-n}{N-1}}} = \frac{110 - 1.100 * 8\%}{\sqrt{1.100 * 8\% * 92\%} * \sqrt{\frac{11.000 - 1.100}{11.000 - 1}}} = 2{,}58$. Vergelijking van de waarde **2,58** met de grenswaarde **1,645** levert als conclusie op dat het percentage financiële adviseurs dat ernstige overtredingen begaat significant hoger is dan 8%.

Uitwerkingen hoofdstuk 8

Opdrachten hoofdstuk 8

Opdracht 1

De verwachte frequenties zijn 66,13% * 300 = **198,39** en 33,87% * 300 = **101,61**.

Opdracht 2

De verwachte frequenties zijn 132**,26, 165,33, 198,39, 67,74, 84, 67** en **101,61**. Deze waarden zijn allemaal groter of gelijk aan 5. Er is voldaan aan de voorwaarde van toepassing van de chikwadraat toets.

Kennisopgaven hoofdstuk 8

Opgave 1

Je moet het groepstotaal op 100% zetten en per groep percentages berekenen.

Opgave 2

Alle verwachte frequenties moeten minimaal 5 zijn.

Opgave 3

De chikwadraat toets bij de analyse van een kruistabel is een rechtseenzijdige toets.

Opgave 4

Je berekent per cel het kolomtotaal maal het rijtotaal gedeeld door de steekproefgrootte.

Toegepaste opgaven hoofdstuk 8

Opgave 1

Stap 1

H_0 : Mannen en vrouwen hebben dezelfde mening over de nieuwe chocolade.
H_1 : Mannen en vrouwen hebben een verschillende mening over de nieuwe chocolade.

Met α = 1%.

Stap 2

Aan de hand van een hulptabel bereken je de chikwadraatwaarde.

O	**E**	**O – E**	$(O-E)^2$	$\frac{(O-E)^2}{E}$
110	120	-10	100	0,833
90	80	10	100	1,250
130	120	10	100	0,833
70	80	-10	100	1,250
400	400	0		4,166

Je vindt de toetsingsgrootheid

$$X^2 = \sum \frac{(O-E)^2}{E} = \mathbf{4{,}166}.$$

Stap 3

Je vindt de grens door de tabel van de chikwadraatverdeling uit bijlage 3 er bij te nemen:

- de eerste ingang is α = 1% = 0,01;
- de tweede ingang is het aantal vrijheidsgraden (aantal kolommen minus 1) * (aantal rijen minus 1) = (2 – 1) * (2 – 1) = 1.

Je vindt de tabelwaarde **6,635**.

Stap 4

De conclusie is dat **4,166 < 6,635**, dus verwerp de nulhypothese niet. Je kunt niet met 99% betrouwbaarheid beweren dat mannen en vrouwen een afwijkende mening hebben over de nieuwe chocolade.

Opgave 2

a. Je moet de kolommen op 100% zetten.

b.

Stap 1

H_0 : De drie leeftijdsgroepen hebben dezelfde voorkeur voor productattribuut.

H_1 : Minstens één leeftijdsgroep heeft een afwijkende voorkeur voor productattribuut.

Met α = 5%.

Stap 2

Aan de hand van een hulptabel bereken je de chikwadraatwaarde.

O	**E**	**O – E**	$(O-E)^2$	$\frac{(O-E)^2}{E}$
40	43,333	-3,333	11,109	0,256
40	43,333	-3,333	11,109	0,256
50	43,333	6,667	44,449	1,026
80	66,667	13,333	177,769	2,667
70	66,667	3,333	11,109	0,167
50	66,667	-16,667	277,789	4,167
80	90	-10	100	1,111
90	90	0	0	0
100	90	10	100	1,111
600	600	0		10,761

Je vindt de toetsingsgrootheid

$$X^2 = \sum \frac{(O-E)^2}{E} = \mathbf{10{,}761}.$$

Stap 3

Je vindt de grens door de tabel van de chikwadraatverdeling uit bijlage 3 er bij te nemen:

- de eerste ingang is α = 5% = 0,05;
- de tweede ingang is het aantal vrijheidsgraden (aantal kolommen minus 1) * (aantal rijen minus 1) = (3 – 1) * (3 – 1) = 1.

Je vindt de tabelwaarde **9,488**.

Stap 4

De conclusie is dat **10,761 > 9,488**, dus verwerp de nulhypothese. Met 95% betrouwbaarheid kun je beweren dat minstens één leeftijdsgroep een afwijkende voorkeur heeft voor productattribuut.

c. De verwachte frequenties zijn **43,333, 66,667** en **90**. Deze zijn allemaal groter of gelijk aan 5. Er is voldaan aan de voorwaarde van toepassing van de chikwadraat toets.

Opgave 3

a. Je moet de rijen op 100% zetten.
b. **Stap 1**

H_0 : De drie leeftijdsgroepen hebben gelijke interesse voor beleggen.
H_1 : Minstens één leeftijdsgroep heeft een afwijkende interesse voor beleggen.

Met α = 5%.

Stap 2

Aan de hand van een hulptabel bereken je de chikwadraatwaarde.

O	**E**	**O – E**	$(O-E)^2$	$\frac{(O-E)^2}{E}$
80	78,333	1,667	2,779	0,035
120	121,667	-1,667	2,779	0,023
70	78,333	-8,333	69,439	0,886
130	121,667	8,333	69,439	0,571
85	78,333	6,667	44,449	0,567
115	121,667	-6,667	44,449	0,365
600	600	0		2,447

Je vindt de toetsingsgrootheid

$$X^2 = \sum \frac{(O-E)^2}{E} = \mathbf{2{,}447}.$$

Stap 3

Je vindt de grens door de tabel van de chikwadraat verdeling uit bijlage 3 er bij te nemen:

- de eerste ingang is α = 5% = 0,05;
- de tweede ingang is het aantal vrijheidsgraden (aantal kolommen minus 1) * (aantal rijen minus 1) = (3 – 1) * (2 – 1) = 2.

Je vindt de tabelwaarde **5,991**.
Stap 4

De conclusie is dat **2,447** < **5,991**, dus verwerp de nulhypothese niet. Met 95% betrouwbaarheid kun je niet beweren dat minstens één leeftijdsgroep een afwijkende interesse heeft voor beleggen.
c. De verwachte frequenties zijn **78,333** en **121,667**. Deze zijn groter of gelijk aan 5. Er is voldaan aan de voorwaarde van toepassing van de chikwadraat toets.

Competentieprikkel

a. Je moet de rijen op 100% zetten.
b. De verwachte frequenties zijn **72, 128, 108** en **192**. Deze vind je door het kolomtotaal te vermenigvuldigen met het rijtotaal en dan te delen door de steekproefgrootte.
c. **Stap 1**

H_0 : De vier leeftijdsgroepen zijn gelijk spontaan bekend met Arke.
H_1 : Minstens één leeftijdsgroep heeft een afwijkende spontane naamsbekendheid met Arke.

Met α = 5%.

Stap 2

Aan de hand van een hulptabel bereken je de chikwadraatwaarde.

O	**E**	**O – E**	$(O-E)^2$	$\frac{(O-E)^2}{E}$
40	72	-32	1.024	14,222
160	128	32	1.024	8,000
70	72	-2	4	0,056
130	128	2	4	0,031
110	108	2	4	0,037
190	192	-2	4	0,021
140	108	32	1.024	9,481
160	192	-32	1.024	5,333
1.000	1.000	0		37,179

Je vindt de toetsingsgrootheid

$$X^2 = \sum \frac{(O-E)^2}{E} = \mathbf{37{,}179}.$$

Stap 3

Je vindt de grens door de tabel van de chikwadraatverdeling uit bijlage 3 er bij te nemen:

- de eerste ingang is α = 5% = 0,05;
- de tweede ingang is het aantal vrijheidsgraden (aantal kolommen minus 1) * (aantal rijen minus 1) = (4 – 1) * (2 – 1) = 3.

Je vindt de tabelwaarde **7,815**.

Stap 4

De conclusie is dat **37,179 > 7,815**, dus verwerp de nulhypothese. Met 95% betrouwbaarheid kun je beweren dat minstens één leeftijdsgroep een afwijkende spontane naamsbekendheid heeft met Arke.

d. De p-waarde is de rechteroppervlakte van 37,179. Inspectie in bijlage 3 geeft dat de rechteroppervlakte van 7,815 gelijk is aan 0,05. De rechteroppervlakte is daarom heel klein, in ieder geval kleiner dan 0,05. Als je de p-waarde bepaalt met de grafische rekenmachine, dan blijkt deze in drie decimalen **0%** te zijn.

Uitwerkingen hoofdstuk 9

Opdrachten hoofdstuk 9

Opdracht 1

Bij de eerste tabel is de steekproefstandaarddeviatie van de mannen 2,55 euro en die van de vrouwen 3,24 euro. Bij de tweede tabel is de steekproefstandaarddeviatie van de mannen 16,02 euro en die van de vrouwen 10,05 euro. Je ziet dat de standaarddeviaties van de tweede tabel veel groter zijn dan die van de eerste tabel. Dit in overeenstemming met de opmerking dat de spreiding in de tweede tabel veel groter is dan de spreiding in de eerste tabel.

Opdracht 2

Je ziet dat **3,02 < 8,84**, dus je verwerpt de nulhypothese niet. Je kunt niet beweren met 95% betrouwbaarheid dat minimaal één verpakking afwijkende gemiddelde verkoopresultaten per week heeft.

Opdracht 3

Als je bij de variantieanalyse geen significantie hebt gevonden, dan betekent dit dat je geen significante verschillen hebt gevonden tussen de steekproefgemiddelden. Het is dan niet meer zinvol om de steekproefgemiddelden twee aan twee te gaan vergelijken op een significant verschil, zoals je doet bij de Tukey toets.

Kennisopgaven hoofdstuk 9

Opgave 1

Normaliteit van de populatiegegevens en de populatiestandaarddeviaties dienen gelijk te zijn.

Opgave 2

Bij variantieanalyse vergelijk je de tussenvariantie met de binnenvariantie. Als de tussenvariantie groot is in verhouding tot de binnenvariantie, dan vind je een significant verschil. Als de tussenvariantie klein is in verhouding tot de binnenvariantie, dan vind je geen significant verschil.

Opgave 3

De tussenvariantie is het kwadraat van de standaarddeviatie van de steekproefgemiddelden, de binnenvariantie is het kwadraat van de standaarddeviatie in iedere groep.

Opgave 4

Bij variantieanalyse maak je gebruik van een rechtseenzijdige toets.

Opgave 5

Door het twee aan twee vergelijken van de steekproefgemiddelden met elkaar krijg je zicht op de onderlinge verhouding tussen de steekproefgemiddelden ten aanzien van significantie.

Toegepaste opgaven hoofdstuk 9

Opgave 1

Stap 1

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
H_1: Minstens één μ wijkt af

Met α = 5%.

Stap 2

48 49 49	54 58 60	46 47 44	40 42 38	
$T_1 = 145$	$T_2 = 172$	$T_3 = 137$	$T_4 = 120$	T =574

$$SS_{totaal} = \sum (x_{ij} - \overline{x})^2 = \sum x_{ij}^{\ 2} - \frac{T^2}{n} = 48^2 + 49^2 + ... + 38^2 - \frac{574^2}{12} = 512{,}917$$

$$SS_{prijs} = \sum (\overline{x}_i - \overline{x})^2 = \sum \frac{T_i^{\ 2}}{n_i} - \frac{T^2}{n} = \frac{145^2}{3} + \frac{172^2}{3} + \frac{137^2}{3} + \frac{120^2}{3} - \frac{574^2}{12} = 470{,}917$$

De laatste kwadratensom SS_{fout} vind je door de kwadratensom van het totaal te verminderen met de kwadratensom van de verpakking, dus

$$SS_{fout} = SS_{totaal} - SS_{prijs} = 512{,}917 - 470{,}917 = 32$$

Bron	**Vrijheidsgraden**	**Kwadratensom**	**Gemiddelde Kwadratensom**	**F-waarde**
Prijs Fout	k -1 = 4 – 1 = 3 n – k = 12 – 4 = 8	470,917 32	156,972 4	39,24
Totaal	n – 1 = 12 – 1 = 11	512,917		

Stap 3

De bovengrens vind je als volgt:

- De onbetrouwbaarheid α = 5%.
- De vrijheidsgraden van de teller zijn gelijk aan 3.
- De vrijheidsgraden van de noemer zijn gelijk aan 8.

Je vindt de tabelwaarde **4,07**.

Stap 4

Je ziet dat de berekende F-waarde F = **39,24** groter is dan de tabelwaarde **4,07**. De conclusie luidt dat er minimaal één populatiegemiddelde significant afwijkt.

Opgave 2

Stap 1

$H_0 : \mu_1 = \mu_2 = \mu_3$
H_1: Minstens één μ wijkt af

Met α = 5%.

Stap 2

20 21 19 17 20	16 17 15 18 17	15 18 14 12 16	
$T_1 = 97$	$T_2 = 83$	$T_3 = 75$	T = 255

$$SS_{totaal} = \sum (x_{ij} - \overline{x})^2 = \sum x_{ij}^{\ 2} - \frac{T^2}{n} = 20^2 + 21^2 + ... + 16^2 - \frac{255^2}{15} = 84{,}000$$

$$SS_{verpakking} = \sum (\overline{x}_i - \overline{x})^2 = \sum \frac{T_i^{\ 2}}{n_i} - \frac{T^2}{n} = \frac{97^2}{5} + \frac{83^2}{5} + \frac{75^2}{5} - \frac{255^2}{15} = 49{,}600$$

De laatste kwadratensom SS_{fout} vind je door de kwadratensom van het totaal te verminderen met de kwadratensom van de verpakking, dus

$$SS_{fout} = SS_{totaal} - SS_{verpakking} = 84{,}000 - 49{,}600 = 34{,}400$$

Bron	**Vrijheidsgraden**	**Kwadratensom**	**Gemiddelde Kwadratensom**	**F-waarde**
Verpakking Fout	k -1 = 3 – 1 = 2 n – k = 15 – 3 = 12	49,600 34,400	24,800 2,867	8,65
Totaal	n – 1 = 15 – 1 = 14	84,000		

Stap 3

De bovengrens vind je als volgt:

- De onbetrouwbaarheid α = 5%.
- De vrijheidsgraden van de teller zijn gelijk aan 2.
- De vrijheidsgraden van de noemer zijn gelijk aan 12.

Je vindt de tabelwaarde **3,89**.

Stap 4

Je ziet dat de berekende F-waarde F = **8,65** groter is dan de tabelwaarde **3,89**. De conclusie luidt dat er minimaal één populatiegemiddelde significant afwijkt.

Opgave 3

a. De gegevens dienen afkomstig te zijn uit de normale verdeling en de populatiestandaarddeviaties dienen gelijk te zijn.
b.

Stap 1

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
H_1: Minstens één μ wijkt af

Met α = 5%.

Stap 2

110 106 102 107	87 98 95 93	102 106 87 103	100 92 98 95	
$T_1 = 425$	$T_2 = 373$	$T_3 = 398$	$T_4 = 385$	T = 1.581

$$SS_{totaal} = \sum (x_{ij} - \overline{x})^2 = \sum x_{ij}^{\ 2} - \frac{T^2}{n} = 110^2 + 106^2 + ... + 95^2 - \frac{1.581^2}{16} = 724{,}438$$

$$SS_{verkoper} = \sum (\overline{x}_i - \overline{x})^2 = \sum \frac{T_i^{\ 2}}{n_i} - \frac{T^2}{n} = \frac{425^2}{4} + \frac{373^2}{4} + \frac{398^2}{4} + \frac{385^2}{4} - \frac{1.581^2}{16} = 373{,}188$$

De laatste kwadratensom SS_{fout} vind je door de kwadratensom van het totaal te verminderen met de kwadratensom van de verpakking, dus

$$SS_{fout} = SS_{totaal} - SS_{verkoper} = 724{,}438 - 373{,}188 = 351{,}250$$

Bron	**Vrijheidsgraden**	**Kwadratensom**	**Gemiddelde Kwadratensom**	**F-waarde**
Verkoper Fout	k -1 = 4 – 1 = 3 n – k = 16 – 4 = 12	373,188 351,250	124,396 29,271	4,25
Totaal	n – 1 = 16 – 1 = 15	724,438		

Stap 3

De bovengrens vind je als volgt:

- De onbetrouwbaarheid α = 5%.
- De vrijheidsgraden van de teller zijn gelijk aan 3.
- De vrijheidsgraden van de noemer zijn gelijk aan 12.

Je vindt de tabelwaarde **3,49**.

Stap 4

Je ziet dat de berekende F-waarde F = **4,25** groter is dan de tabelwaarde **3,49**. De conclusie luidt dat er minimaal één populatiegemiddelde significant afwijkt.

c. De Tukey toets geeft als resultaat dat de eerste verkoper significant gemiddeld meer verkoopt per week dan de andere verkopers.

Competentieprikkel

a. De gegevens dienen afkomstig te zijn uit de normale verdeling en de populatiestandaarddeviaties dienen gelijk te zijn.
b.

Stap 1

$H_0 : \mu_1 = \mu_2 = \mu_3$
H_1: Minstens één μ wijkt af

Met α = 5%.

Stap 2

2.500 3.600 2.400 3.000	2.900 4.300 3.900 3.600	4.100 4.200 3.800 4.500	
$T_1 = 11.500$	$T_2 = 14.700$	$T_3 = 16.600$	T = 42.800

$$SS_{totaal} = \sum (x_{ij} - \overline{x})^2 = \sum x_{ij}^{\ 2} - \frac{T^2}{n} = 2.500^2 + 3.600^2 + ... + 4.500^2 - \frac{42.800^2}{12} = 5.526.666,667$$

$$SS_{haven} = \sum (\overline{x}_i - \overline{x})^2 = \sum \frac{T_i^{\ 2}}{n_i} - \frac{T^2}{n} = \frac{11.500^2}{4} + \frac{14.700^2}{4} + \frac{16.600^2}{4} - \frac{42.800^2}{12} = 3.321.666,667$$

De laatste kwadratensom SS_{fout} vind je door de kwadratensom van het totaal te verminderen met de kwadratensom van de verpakking, dus

$$SS_{fout} = SS_{totaal} - SS_{haven} = 5.526.666,667 - 3.321.666,667 = 2.205.000$$

Bron	**Vrijheidsgraden**	**Kwadratensom**	**Gemiddelde Kwadratensom**	**F-waarde**
Haven Fout	k -1 = 3 – 1 = 2 n – k = 12 – 3 = 9	3.321.666,667 2.205.000	1.660.833,333 245.000	6,78
Totaal	n – 1 = 12 – 1 = 11	5.526.666,667		

Stap 3

De bovengrens vind je als volgt:

- De onbetrouwbaarheid α = 5%.
- De vrijheidsgraden van de teller zijn gelijk aan 2.
- De vrijheidsgraden van de noemer zijn gelijk aan 9.

Je vindt de tabelwaarde **4,26**.

Stap 4

Je ziet dat de berekende F-waarde F = **6,78** groter is dan de tabelwaarde **4,26**. De conclusie luidt dat er minimaal één populatiegemiddelde significant afwijkt.

c. De Tukey toets geeft als resultaat dat naar Hamburg significant gemiddeld minder containers vervoerd worden dan naar Antwerpen en New York.

Uitwerkingen hoofdstuk 10

Opdrachten hoofdstuk 10

Opdracht 1

De voorspelling vind je door 60 in te vullen in de regressievergelijking. Je vindt dat 370 + 3,20 * 60 = 562. Dit betekent een voorspelde maandomzet van **562.000** euro.

Opdracht 2

De voorspelling bij 60.000 euro aan advertentiekosten is 562.000 euro. Verder is de tabelwaarde onbekend:

- de rechteroppervlakte voor de t-waarde is 10%/2 = 5% = 0,05,
- de vrijheidsgraden zijn n – 2 = 4 – 2 = 2.

Dit levert de tabelwaarde 2,920 op.

Het voorspellingsinterval is:

$$\hat{Y} \pm t_{\alpha/2;n-2}.s.\sqrt{1+\frac{1}{n}+\frac{(X-\overline{X})^2}{\sum(X-\overline{X})^2}} = 562 \pm 2{,}920 * 6{,}325 * \sqrt{1+\frac{1}{4}+\frac{(60-25)^2}{500}} = 562 \pm 35{,}526$$

Dat betekent dat het 90% voorspellingsinterval loopt van **526.474** euro tot **597.526** euro.

Opdracht 3

Als r = 0,90, dan is $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}90\sqrt{4-2}}{\sqrt{1-0{,}90^2}} = 2{,}920$. Dit is bij eenzijdig toetsen precies de grenswaarde voor significantie. Dat betekent dat bij r = 0,90 en n = 4 de conclusie luidt dat er een significant positief lineair verband aanwezig is tussen advertentiekosten per maand en maandomzet. Voor een hogere waarde van r, zoals r = 0,96, vind je een hogere t-waarde en dus ook een significante correlatie.

Kennisopgaven hoofdstuk 10

Opgave 1

Het model is:

$$Y = \alpha + \beta X + \varepsilon$$

Hierin is

α	de populatieconstante
β	de populatierichtingscoëfficiënt
ε	storingsterm

Opgave 2

De regressievergelijking is $\hat{Y} = a + bX$. Met deze vergelijking kun je een voorspelling geven door een X-waarde in te vullen.

Opgave 3

- De betrouwbaarheid.
- De steekproefgrootte.
- De steekproefstandaarddeviatie.
- De waarde voor X die ingevuld wordt om de voorspelling te geven
- Het gemiddelde van X.
- De som van de kwadratische afwijkingen van X.

Opgave 4

Alle waarden vanaf -1 tot en met 1 kan r aannemen.

r = -1	perfect dalend lineair verband
r = 0	geen lineair verband
r = 1	perfect stijgend lineair verband

Opgave 5

Zowel r als n hebben invloed op de toets op correlatie $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$:

- als r gelijk aan nul is, dan is de bijbehorende t-waarde ook nul;
- als r bijna gelijk aan 1 is, dan wordt de bijbehorende t-waarde een hele grote waarde;
- als r bijna gelijk aan -1 is, dan wordt de bijbehorende t-waarde een hele kleine waarde;
- hoe groter het steekproefaantal n is, hoe groter de t-waarde.

Opgave 6

De voorwaarden om de toets op correlatie uit te voeren zijn

- de afhankelijke variabele, de omzet dus in de case, moet normaal verdeeld zijn,
- de standaarddeviatie rond de lijn moet voor iedere X-waarde even groot zijn.

Toegepaste opgaven hoofdstuk 10

Opgave 1

a. Zelf.
b. Allereerst maak je een hulptabel:

X	Y	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
10	50	-3	-16,375	49,125	9	268,141
12	62	-1	-4,375	4,375	1	19,141
13	67	0	0,625	0	0	0,391
14	73	1	6,625	6,625	1	43,891
16	80	3	13,625	40,875	9	185,641
12	65	-1	-1,375	1,375	1	1,891
13	66	0	-0,375	0	0	0,141
14	68	1	1,625	1,625	1	2,641
104	531	0	0	104	22	521,878

Hiermee bereken je b = 104/22 = 4,727. Als je b berekend hebt, dan kun je a bepalen:

$\overline{X} = 104/8 = 13$,

$\overline{Y} = 531/8 = 66{,}375$,

$a = \overline{Y} - b\,\overline{X} = 66{,}375 - 4{,}727 * 13 = 4{,}924$.

De regressievergelijking luidt: $\hat{Y} = 4{,}924 + 4{,}727X$.

c. Het voorspellingsinterval is:

$$\hat{Y} \pm t_{\alpha/2;n-2}.s.\sqrt{1+\frac{1}{n}+\frac{(X-\bar{X})^2}{\sum(X-\bar{X})^2}} = 99{,}46 \pm 1{,}943 * 2{,}245 * \sqrt{1+\frac{1}{8}+\frac{(20-13)^2}{22}} = 99{,}46 \pm 7{,}99$$

Het voorspellingsinterval loopt van **914.700** euro tot **1.074.500** euro.

d.

Stap 1

$H_0 : \rho = 0$
$H_1 : \rho > 0$

Je toetst met α = 1%.

Stap 2
Allereerst bereken je de correlatiecoëfficiënt. Deze is:

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2.\sum(Y-\bar{Y})^2}} = \frac{104}{\sqrt{22*521{,}878}} = 0{,}971$$

Daarna bereken je de t-waarde:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}971\sqrt{8-2}}{\sqrt{1-0{,}971^2}} = 9{,}948$$

Stap 3

Je vindt de tabelwaarde met de ingangen:

- 8 – 2 = 6 vrijheidsgraden
- α = 0,01

Je vindt in bijlage 2 de tabelwaarde **3,143**.

Stap 4

Omdat de berekende t-waarde t = **9,948** veel groter is dan de tabelwaarde **3,143** is er sprake van een significant positief lineair verband tussen het aantal gereden kilometers per week en de omzet per week.

Opgave 2

a. Zelf.
b. Allereerst maak je een hulptabel:

X	Y	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
34	20	-4,583	-3,500	16,041	21,004	12,250
42	25	3,417	1,500	5,126	11,676	2,250
35	21	-3,583	-2,500	8,958	12,838	6,250
50	35	11,417	11,500	131,300	130,348	132,250
44	27	5,417	3,500	18,960	29,344	12,250
38	22	-0,583	-1,500	0,875	0,340	2,250
36	21	-2,583	-2,500	6,458	6,672	6,250
40	25	1,417	1,500	2,126	2,008	2,250
43	26	4,417	2,500	11,043	19,510	6,250
37	22	-1,583	-1,500	2,375	2,506	2,250
31	18	-7,583	-5,500	41,707	57,502	30,250
33	20	-5,583	-3,500	19,541	31,170	12,250
463	282	0	0	264,510	324,918	227

Hiermee bereken je b = 264,510/324,918 = 0,814. Als je b berekend hebt, dan kun je a bepalen:

$\overline{X} = 463/12 = 38{,}583$,

$\overline{Y} = 282/12 = 23{,}500$,

$a = \overline{Y} - b\,\overline{X} = 23{,}500 - 0{,}814 * 38{,}583 = -7{,}907$

De regressievergelijking luidt: $\hat{Y} = -7{,}907 + 0{,}814X$. De voorspelling is -7,914 + 0,814 * 50 = 32,79, hetgeen **3.279** euro betekent.

c. Het voorspellingsinterval is:

$$\hat{Y} \pm t_{\alpha/2;n-2}.s.\sqrt{1+\frac{1}{n}+\frac{(X-\overline{X})^2}{\sum(X-\overline{X})^2}} = 32{,}79 \pm 2{,}228 * 1{,}081 * \sqrt{1+\frac{1}{12}+\frac{(50-38{,}583)^2}{324{,}918}} = 32{,}79 \pm 2{,}93$$

Het voorspellingsinterval loopt van **2.986** euro tot **3.572** euro.

d. De omzet van 3.500 euro zit inderdaad in het voorspellingsinterval, zodat de accountant niet meer twijfelt aan de correctheid van deze waarde.

e.

Stap 1

$H_0 : \rho = 0$

$H_1 : \rho > 0$

Je toetst met α = 5%.

Stap 2

Allereerst bereken je de correlatiecoëfficiënt. Deze is:

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2.\sum(Y-\bar{Y})^2}} = \frac{264{,}510}{\sqrt{324{,}918 * 227}} = 0{,}974$$

Daarna bereken je de t-waarde:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}974\sqrt{12-2}}{\sqrt{1-0{,}974^2}} = 13{,}596$$

Stap 3

Je vindt de tabelwaarde met de ingangen:

- 12 – 2 = 10 vrijheidsgraden
- α = 0,05

Je vindt in bijlage 2 de tabelwaarde **1,812**.

Stap 4

Omdat de berekende t-waarde t = **13,596** veel groter is dan de tabelwaarde **1,812** is er sprake van een significant positief lineair verband tussen het aantal verhuurde kamers en de omzet per maand.

Opgave 3

a. Allereerst maak je een hulptabel:

X	Y	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$(X-\overline{X})^2$	$(Y-\overline{Y})^2$
12	16	-3,200	-3,800	12,160	10,240	14,440
15	20	-0,200	0,200	-0,040	0,040	0,040
13	17	-2,200	-2,800	6,160	4,800	7,840
16	21	0,800	1,200	0,960	0,640	1,440
20	25	4,800	5,200	24,960	23,040	27,040
76	99	0	0	44,200	38,760	50,800

Hiermee bereken je b = 44,200/38,760 = 1,140. Als je b berekend hebt, dan kun je a bepalen:

$\overline{X} = 76/5 = 15{,}200$,

$\overline{Y} = 99/5 = 19{,}800$,

$a = \overline{Y} - b\,\overline{X} = 19{,}800 - 1{,}140 * 15{,}200 = 2{,}472$.

De regressievergelijking luidt: $\hat{Y} = 2{,}472 + 1{,}140X$.

a. De constante 2,472 euro is het vast bedrag dat standaard bij de inkoopprijs wordt opgeteld, de richtingscoëfficiënt 1,14 betekent dat bovenop de 2,47 euro nog eens 14% extra van de inkoopprijs geteld wordt.

b. **Stap 1**

$H_0 : \rho = 0$

$H_1 : \rho > 0$

Je toetst met α = 10%.

Stap 2

Allereerst bereken je de correlatiecoëfficiënt. Deze is:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 . \sum (Y - \bar{Y})^2}} = \frac{44{,}200}{\sqrt{38{,}760 * 50{,}800}} = 0{,}996$$

Daarna bereken je de t-waarde:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}996\sqrt{5-2}}{\sqrt{1-0{,}996^2}} = 19{,}307$$

Stap 3

Je vindt de tabelwaarde met de ingangen:

- 5 – 2 = 3 vrijheidsgraden
- α = 0,10

Je vindt in bijlage 2 de tabelwaarde **1,638**.

Stap 4

Omdat de berekende t-waarde t = **19,307** veel groter is dan de tabelwaarde **1,638** is er sprake van een significant positief lineair verband tussen de inkoopprijs en de verkoopprijs.

Competentieprikkel

a. Allereerst maak je een hulptabel:

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$
5	20	-10	-9,600	96	100	92,160
10	26	-5	-3,600	18	25	12,960
15	32	0	2,400	0	0	5,760
20	33	5	3,400	17	25	11,560
25	37	10	7,400	74	100	59,290
75	148	0	0	205	250	181,730

Hiermee bereken je b = 205/250 = 0,820. Als je b berekend hebt, dan kun je a bepalen:

$\bar{X} = 75/5 = 15$,

$\bar{Y} = 148/5 = 29{,}600$,

$a = \bar{Y} - b\,\bar{X} = 29{,}600 - 0{,}820 * 15 = 17{,}300$.

De regressievergelijking luidt: $\hat{Y} = 17{,}300 + 0{,}820X$.

b. De voorspelling luidt: 17,300 + 0,820 * 30 = 41,900, dus **41.900** euro.
c. Het voorspellingsinterval is:

$$\hat{Y} \pm t_{\alpha/2;n-2}.s.\sqrt{1+\frac{1}{n}+\frac{(X-\bar{X})^2}{\sum(X-\bar{X})^2}} = 41{,}900 \pm 5{,}841*1{,}742*\sqrt{1+\frac{1}{5}+\frac{(35-15)^2}{250}} = 41{,}900 \pm 17{,}026$$

Het interval loopt van **24.874** euro tot **58.926** euro.

c.

Stap 1

$H_0 : \rho = 0$
$H_1 : \rho > 0$

Je toetst met α = 1%.

Stap 2

Allereerst bereken je de correlatiecoëfficiënt. Deze is:

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2.\sum(Y-\bar{Y})^2}} = \frac{205}{\sqrt{250*181{,}730}} = 0{,}962$$

Daarna bereken je de t-waarde:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}962\sqrt{5-2}}{\sqrt{1-0{,}962^2}} = 6{,}102$$

Stap 3

Je vindt de tabelwaarde met de ingangen:

- 5 – 2 = 3 vrijheidsgraden
- α = 0,01

Je vindt in bijlage 2 de tabelwaarde **4,541**.

Stap 4

Omdat de berekende t-waarde t = **6,102** groter is dan de tabelwaarde **4,541** is er sprake van een significant positief lineair verband tussen het percentage klanten dat op rekening koopt per maand en de omzet per maand.

Uitwerkingen hoofdstuk 11

Opdrachten

Opdracht 1

De p-waarde is P($\underline{k} \leq 1$ of $\underline{k} \geq 5$) = 0,1094 + 0,1094 = 0,2188 = **21,88%.**

Opdracht 2

Als n = 10, dan

P($\underline{k} \leq 0$) = 0,0010 < 0,025
P($\underline{k} \leq 1$) = 0,0107 < 0,025
P($\underline{k} \leq 2$) = 0,0547 > 0,025.

De ondergrens is 1 en de bovengrens is 9. Het kritieke gebied is **{0, 1, 9, 10}**.

Opdracht 3

De z-waarde is $z = \frac{k - np}{\sqrt{np(1-p)}} = \frac{270 - 250}{\sqrt{500 * 50\% * 50\%\%}} = 1{,}79$. De grenswaarden zijn -1,645 en 1,645. Omdat de berekende z-waarde z = **1,79 > 1,645**, is de conclusie dat er een significante voorkeur is voor ontwerp 1.

Opdracht 4

Je berekent $\mu = \frac{n(n+1)}{4} = \frac{61.62}{4} = 945{,}5$ en $\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{61.62.123}{24}} = 139{,}22$.
Hiermee bepaal je $z = \frac{T - \mu}{\sigma} = \frac{745 - 945{,}5}{139{,}22} = -1{,}44$.

Opdracht 5

De z-waarde is $z = \frac{T - \mu}{\sigma} = \frac{720 - 945{,}5}{139{,}22} = -1{,}62$, dit vergeleken met de grenzen -1,645 en 1,645 geeft geen significante voorkeur!

Opdracht 6

8 is groter dan 7, 1 maal groter,
8 is gelijk aan 8, 0,5 maal groter,
8 is groter dan 6, 1 maal groter,
8 is groter dan 5, 1 maal groter,
8 is gelijk aan 8, 0,5 maal groter.

In totaal is 8 1 + 0,5 + 1 + 1 + 0,5 = **4** maal groter dan de getallen uit groep 2.

Opdracht 7

48 is groter dan 46 en 44,
47 is groter dan 46 en 44,
45 is groter dan 4,
51 is groter dan 46, 50, 44 en 49.

In totaal zijn de getallen van groep 1 **9** maal groter dan die van groep 2.

Opdracht 8

De z-waarde is $z = \frac{9-66}{\sqrt{264}} = -3{,}51$.

Opdracht 9

R_1 = 7 + 7 + 2 + 10,5 + 1 = **27,5**
R_2 = 10,5 + 4 + 4 + 10,5 + 14 = **43**
R_3 = 10,5 + 14 + 7 + 4 + 14 = **49,5**.

Opdracht 10

Met α = 0,05 en 2 vrijheidsgraden is de tabelwaarde **5,991**.

Opdracht 11

Als n = 10, dan is $z = r_s.\sqrt{n-1} = 0{,}40.\sqrt{10-1} = 1{,}20$;
als n = 100, dan is $z = r_s.\sqrt{n-1} = 0{,}40.\sqrt{100-1} = 3{,}98$.

Vergelijking van deze waarden met 1,96 geeft bij de eerste z-waarde geen significante correlatie, maar bij de tweede z-waarde wel.

Kennisopgaven hoofdstuk 11

Opgave 1

Je gebruikt de binomiale verdeling.

Opgave 2

De Wilcoxontoets is een gepaarde toets.

Opgave 3

De Wilcoxontoets heeft een groter onderscheidend vermogen omdat de rangordes gebruikt worden, bij de tekentoets is er slechts sprake van wel of geen voorkeur. Bij de Wilcoxontoets wordt een ordinaal meetniveau gebruikt, bij de tekentoets een nominaal meetniveau.

Opgave 4

Je telt hoe vaak de getallen uit groep 1 groter zijn dan de getallen uit groep 2, daarna tel je hoe vaak de getallen uit groep 2 groter zijn dan de getallen van groep 1. Hiervan pak je het minimum en dat noem je U.

Opgave 5

Bij de Mann-Whitneytoets vergelijk je twee groepen met elkaar, bij de Kruskal-Wallistoets vergelijk je meer dan twee groepen met elkaar.

Opgave 6

Bij de Pearson correlatiecoëfficiënt is minimaal een interval meetniveau vereist, de Spearman correlatiecoëfficiënt gebruik je bij een ordinaal meetniveau.

Toegepaste Opgaven hoofdstuk 11

Opgave 1

a.
Stap 1

$H_0 : p = 50\%$
$H_1 : p \neq 50\%$

Met α = 10%.

Stap 2

Laat <u>k</u>: aantal malen voorkeur voor Shrek, dan is <u>k</u> = 2.

Stap 3

- Herhaling n = 9;
- Twee uitkomstmogelijkheden: voorkeur voor Flippo's of plaatjes van Shrek;
- Succeskans p = 50%, zeer grote doelgroep.

Voor de grenswaarden heb je de binomiale verdeling nodig in bijlage 5:

$P(\underline{k} \leq 0) = 0{,}0020 < 0{,}05$
$P(\underline{k} \leq 1) = 0{,}0195 < 0{,}05$
$P(\underline{k} \leq 2) = 0{,}0898 > 0{,}05$.

De ondergrens is 1 en de bovengrens is 8, het kritieke gebied is {0, 1, 8, 9}.

Stap 4

Conclusie: 2 > 1, er is geen aantoonbare significante voorkeur.

b. Analoog, maar nu met $z = \frac{k - np}{\sqrt{np(1-p)}} = \frac{145 - 125}{\sqrt{250 * 50\% * 50\%}} = 2{,}53$. Controle voor z-benadering is np = 250 * 50% = 125 ≥ 5en n(1 – p) = 250 * 50% = 125 ≥ 5. Omdat **2,53 > 1,645**, is er een significante voorkeur voor plaatjes van Shrek aangetoond.

Opgave 2

Stap 1

H_0 : De twee locaties worden gelijk beoordeeld
H_1 : De twee locaties worden niet gelijk beoordeeld

Met α = 5%.

Stap 2

Locatie A	Locatie B	Verschil	Rangorde
9	6	3	7
8	5	3	7
9	4	5	10
8	5	3	7
6	7	-1	1
7	4	3	7
8	5	3	7
7	5	2	3
8	6	2	3
7	5	2	3

T^+ = 7 + 7 + 10 + 7 + 7 + 7 + 3 + 3 + 3 = 54;

$T^- = 1$.

Het minimum van T^+ en T^- is **T = 1**.

Stap 3

- α = 5% = 0,05;
- n = 10

De tabelwaarde is 8, het kritieke gebied is **{0, 1, 2, 3, 4, 5, 6, 7, 8}**.

Stap 4

Omdat **1 ≤ 8**, is de conclusie dat er een significant verschil is in de beoordeling van locatie A en locatie B.

b. Analoog, maar nu met $z = \dfrac{T - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}} = \dfrac{400 - \dfrac{50.51}{4}}{\sqrt{\dfrac{50.51.101}{24}}} = -2{,}29$. Als je deze waarde vergelijkt met de ondergrens -1,96 en de bovengrens 1,96, is de conclusie dat er een significante voorkeur is.

Opgave 3

a.
Stap 1

H_0 : De bruto maandinkomens van bakkers en slagers zijn gelijk
H_1 : De bruto maandinkomens van bakkers en slagers zijn niet gelijk

Met α = 5%.

Stap 2

Door twee aan twee vergelijken vind je dat **U = 8**.

Stap 3

- α = 5%;
- $n_1 = 6$;
- $n_2 = 5$.

De tabelwaarde in bijlage 9 is 3. Het kritieke gebied is **{0, 1, 2, 3}**.

Stap 4

De conclusie is dat **8 > 5**, er is geen significant verschil in de bruto maandinkomens van bakkers en slagers.

b. Analoog als a, maar nu met $z = \dfrac{U - \dfrac{n_1 n_2}{2}}{\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} = \dfrac{1.200 - \dfrac{50.50}{2}}{\sqrt{\dfrac{50.50.101}{12}}} = -0{,}34$. Vergelijking van deze waarde met de bekende grenzen -1,96 en 1,96 levert geen significant verschil op tussen de bruto maandinkomens van bakkers en slagers.

Opgave 4

Stap 1

H_0 : De drie producten verkopen wekelijks even veel

H_1 : Minstens één product wijkt af met de wekelijkse verkoopaantallen

Met α = 5%.

Stap 2

Zookies		Chocopain		Tjokidz	
120	**7,5**	80	**4**	400	**18**
180	**10**	100	**6**	300	**17**
230	**14**	70	**3**	250	**15**
200	**12**	50	**1**	275	**16**
150	**9**	120	**7,5**	190	**11**
90	**5**	60	**2**	210	**13**
	R_1 = **57,5**		R_2 = **23,5**		R_3 = **90**

In iedere tweede kolom zie je de rangordes staan van de gegeven data. Nu bereken je de chikwadraatwaarde:

$$\chi^2 = \frac{12}{n(n+1)} \sum \frac{R_i^{\,2}}{n_i} - 3(n+1) = \frac{12}{18.19}\{\frac{57{,}5^2}{6} + \frac{23{,}5^2}{6} + \frac{90^2}{6}\} - 3.19 = \mathbf{12{,}933.}$$

Stap 3

- α = 5% = 0,05;
- het aantal vrijheidsgraden is 3 – 1 = 2.

De tabelwaarde is **5,991**.

Stap 4

Omdat **12,933 > 5,991**, luidt de conclusie dat er minimaal één kokje significant afwijkt met de wekelijkse verkoopaantallen.

Opgave 5

a. Allereerst maak je de volgende hulptabel:

Jaren werkervaring	Rangorde	Bruto jaarinkomen	Rangorde	Verschil rangordes	Kwadraat verschil
2	1	30.000	1	0	0
5	2	35.000	3,5	-1,5	2,25
15	6	60.000	6	0	0
10	5	50.000	5	0	0
8	4	34.000	2	2	4
6	3	35.000	3,5	-0,5	0,25
20	7	65.000	7	0	0
					6,50

Nu bereken je Spearman's rangcorrelatiecoëfficiënt:

$$r_s = 1 - \frac{6\sum d_i^{\,2}}{n(n^2-1)} = 1 - \frac{6.6{,}50}{7.(7^2-1)} = 0{,}88.$$

b.
Stap 1

H_0 : Geen rangcorrelatie tussen jaren werkervaring en het bruto jaarinkomen
H_1 : Wel rangcorrelatie tussen jaren werkervaring en het bruto jaarinkomen

Met α = 10%.

Stap 2

Je berekent de z-waarde $z = r_s.\sqrt{n-1} = 0{,}88.\sqrt{7-1} = 2{,}16$.

Stap 3

De ondergrens is **-1,645** en de bovengrens is **1,645**.

Stap 4

De conclusie is dat **2,16 > 1,645**, dus is er een significante positieve rangcorrelatie tussen aantal jaren werkervaring en het bruto jaarinkomen.

Competentieprikkel

a.

Stap 1

H_0 : De klanten tussen 30 en 45 jaar besteden evenveel als de klanten van 45 tot 60 jaar

H_1 : De klanten tussen 30 en 45 jaar besteden niet evenveel als de klanten van 45 tot 60 jaar

Met α = 5%.

Stap 2

30 tot 45 jaar	**45 tot 60 jaar**
23.000	40.000
19.000	50.000
14.000	29.000
18.000	24.000
21.000	22.000
20.000	25.000
19.000	45.000

Je kijkt hoe vaak de getallen van de eerste groep groter zijn dan de getallen van de tweede groep. Alleen 23.000 is groter dan 22.000. Dat betekent dat **U = 1**.

Stap 3

- α = 5% = 0,05;
- $n_1 = 7$;
- $n_2 = 7$.

De tabelwaarde is 8, het kritieke gebied is **{0, 1, 2, 3, 4, 5, 6, 7, 8}.**

Stap 4

Conclusie: **1 ≤ 8**, je ziet een significant verschil tussen de twee leeftijdsgroepen ten aanzien van de gemiddelde besteding aan een Mercedes.

b.

Stap 1

H_0 : De drie leeftijdsgroepen geven een gelijk algemeen oordeel

H_1 : Minstens één leeftijdsgroep heeft een afwijkend algemeen oordeel

Met α = 10%.

Stap 2

Allereerst construeer je de tabel met gegevens en berekent de rangordes met totalen.

30 tot 45 jaar		**45 tot 60 jaar**		**60 jaar en ouder**	
8	**13,5**	7	**4**	8	**13,5**
9	**22**	7	**4**	7	**4**
8	**13,5**	8	**13,5**	8	**13,5**
8	**13,5**	9	**22**	9	**22**
7	**4**	8	**13,5**	8	**13,5**
7	**4**	9	**22**	8	**13,5**
9	**22**	7	**4**	8	**13,5**
7	**4**			8	**13,5**
				8	**13,5**
	R_1 = **96,5**		R_2 = **83**		R_3 = **120,5**

Hierna bereken je de chikwadraatwaarde:

$$\chi^2 = \frac{12}{n(n+1)}\sum\frac{R_i^{\,2}}{n_i} - 3(n+1) = \frac{12}{24.25}\{\frac{96{,}5^2}{8} + \frac{83^2}{7} + \frac{120{,}5^2}{9}\} - 3.24 = 0{,}231$$

Stap 3

- α = 10% = 0,10;
- de vrijheidsgraden zijn 3 – 1 = 2.

Tot je verbazing zie je dat de tabelwaarde ontbreekt. Je bent genoodzaakt om of een andere tabel te gaan raadplegen, of om met je GRM de p-waarde te bepalen. Deze blijkt gelijk te zijn aan 0,89 = **89%**.

Stap 4

Conclusie**: 89% > 10%**, je kunt niet zeggen dat er minstens één leeftijdsgroep significant afwijkt ten aanzien van het algemeen oordeel over de Mercedesgarage.

c.
Deze opgave is met de hand al bijna niet meer te doen. Het gebruik van de pc lijkt hier handiger.

Stap 1

H_0 : Geen rangcorrelatie tussen vraag 10 en vraag 8
H_1 : Wel rangcorrelatie tussen vraag 10 en vraag 8

Met α = 5%.

Stap 2

Je construeert de tabel met gegevens, de rangorde, de verschillen in de rangordes en de kwadraten hiervan. Dit tel je tot slotte bij elkaar op.

Vraag 8	**Rangorde**	**Vraag 10**	**Rangorde**	**Verschil rangordes**	**Kwadraat verschil**
2	**15,5**	7	**4**	11,5	132,25
2	**15,5**	7	**4**	11,5	132,25
2	**15,5**	8	**13,5**	2,0	4,00
1	**3,5**	8	**13,5**	-10,0	100,00
2	**15,5**	9	**22**	-6,5	42,25
2	**15,5**	8	**13,5**	2,0	4,00
1	**3,5**	7	**4**	-0,5	0,25
2	**15,5**	8	**13,5**	2,0	4,00
2	**15,5**	9	**22**	-6,5	42,25
2	**15,5**	9	**22**	-6,5	42,25
2	**15,5**	8	**13,5**	2,0	4,00
2	**15,5**	8	**13,5**	2,0	4,00
2	**15,5**	8	**13,5**	2,0	4,00
1	**3,5**	8	**13,5**	-10,0	100,00
2	**15,5**	8	**13,5**	2,0	4,00
2	**15,5**	9	**22**	-6,5	42,25
1	**3,5**	8	**13,5**	-10,0	100,0
2	**15,5**	7	**4**	11,5	132,25
2	**15,5**	7	**4**	11,5	132,25
2	**15,5**	7	**4**	11,5	132,25
1	**3,5**	9	**22**	-18,5	342,25
2	**15,5**	8	**13,5**	2,0	4,00
2	**15,5**	8	**13,5**	2,0	4,00
1	**3,5**	7	**4**	-0,5	0,25
					1.509

Daarna bereken je de Spearman correlatiecoëfficiënt:

$$r_s = 1 - \frac{6\sum d_i^{\,2}}{n(n^2-1)} = 1 - \frac{6*1.509}{24.(24^2-1)} = 0{,}3439\,.$$

Hiermee bepaal je de z-waarde:

$$z = r_s.\sqrt{n-1} = 0{,}3439.\sqrt{24-1} = 1{,}649\,.$$

Stap 3

De ondergrens is **-1,96** en de bovengrens is **1,96**.

Stap 4

De conclusie is dat **1,649< 1,96**, dus is er een geen significante rangcorrelatie aantoonbaar tussen vraag 8 en vraag 10.

Uitwerkingen hoofdstuk 12

Toegepaste opgaven hoofdstuk 12

Opgave 1

a. Trefwoorden: Percentage – Betrouwbaarheidsinterval – 22% * 2.000 = 440 ≥ 5, 78% * 2.000 = 1.560 ≥ 5, dus z-benadering. Je gebruikt de formule van paragraaf 6.2.

$$\hat{p} \pm z_{\alpha/2} . \sqrt{\frac{\hat{p}.(1-\hat{p})}{n}} = 22\% \pm 1{,}645 . \sqrt{\frac{22\%.78\%}{2.000}} = 22\% \pm 1{,}52\%$$

Het 90% betrouwbaarheidsinterval loopt van **20,48%** tot **23,52%.**

b. Trefwoorden: Percentage – Toetsen op één percentage – np = 2.000 * 20% = 400 ≥ 5 en n(1 – p) = 2.000 * 80% = 1.600 ≥ 5, je komt uit op de z-benadering van paragraaf 7.2.

Stap 1

H_0: p = 20%

H_1: p > 20%

Met α = 1%.

Stap 2

Invullen van alle gegevens levert op:

$$z = \frac{k - np}{\sqrt{np(1-p)}} = \frac{440 - 400}{\sqrt{2.000 * 20\% * 80\%}} = 2{,}24$$

Stap 3

Er is zoals eerder aangetoond voldaan aan de voorwaarden van de benadering met de normale verdeling. De grenswaarde is bij een α van 1% gelijk aan **2,33**.

Stap 4

De conclusie vind je door de gevonden z-waarde van 2,24 te vergelijken met de grenswaarde 2,33. Omdat **2,24 kleiner is dan 2,33**, verwerp je de nulhypothese niet. Je kunt niet beweren dat het percentage mensen dat vindt dat de klussector het minst positieve imago heeft, significant groter is dan 20%.

c. Trefwoorden: - Percentage – Minimale steekproefgrootte – Zeer grote doelgroep. Je komt uit in paragraaf 6.4. De berekening is:

$n \geq \frac{z_{\alpha/2}{}^2.p.(1-p)}{F^2} = \frac{1{,}96^2.50\%.50\%}{2{,}5\%^2} = 1.536{,}64$; de minimale steekproefgrootte is n = **1.537** respondenten.

Opgave 2

a. **Stap 1**

H_0: μ = 1,70
H_1: μ > 1,70

Met α = 1%.

b. Trefwoorden: - Gemiddelde – Toetsen op één gemiddelde – Grote n = 400 ≥ 200 – Zeer grote N. De keuze valt op de toets uit hoofdstuk 2.

Stap 2

Bereken de toetsingsgrootheid:

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1{,}80 - 1{,}70}{0{,}40/\sqrt{400}} = 5{,}00$$

Stap 3

De bovengrens is **2,33**.

Stap 4

Conclusie: **5,00 > 2,33**, er is sprake van een significante toename van de gemiddelde prijs voor een witbrood.

Opgave 3

Trefwoorden – niet-parametrische toets (gegevens niet normaal verdeeld) – twee onafhankelijke steekproeven – Mann-Whitney toets – paragraaf 11.4.

Stap 1

H_0: De bruto maandinkomens van ZZP'ers en ondernemers met max. vijf personeelsleden zijn gelijk
H_1: De bruto maandinkomens van ZZP'ers en ondernemers met max. vijf personeelsleden zijn ongelijk

Met α = 5%.

Stap 2

ZZP'ers	Ondernemers
3.500	4.900
3.200	5.500
5.000	4.200
4.400	6.500
5.200	6.900
4.200	5.700

Je telt hoe vaak de getallen van groep 1 groter zijn dan de getallen van groep 2:

5.000 is groter dan 4.900,
5.000 is groter dan 4.200,
4.400 is groter dan 4.200,
5.200 is groter dan 4.900,
5.200 is groter dan 4.200,
4.200 is gelijk aan 4.200.

Dit levert op dat **U = 5,5**.

Stap 3

Neem bijlage 9 bij de hand met de ingangen

- $\alpha = 0{,}05$;
- $n_1 = n_2 = 6$.

Je vindt de tabelwaarde 5, het kritieke gebied is **{0, 1, 2, 3, 4, 5}.**

Stap 4

Omdat **5,5 groter is dan 5** is er geen significant verschil aantoonbaar voor de bruto inkomens van ZZP'ers en ondernemers met maximaal vijf personeelsleden.

Opgave 4

a. In deze kruistabel zet je de mannen en vrouwen op 100% om ze procentueel met elkaar te kunnen vergelijken. Je kijkt naar hoofdstuk 8: de analyse van een kruistabel.

Stap 1

H_0 : Mannen en vrouwen hebben gelijke voorkeur voor Pepsi en Coca Cola.
H_1 : Mannen en vrouwen hebben geen gelijke voorkeur voor Pepsi en Coca Cola.

Met $\alpha = 5\%$.

Stap 2

	Coca Cola	**Pepsi**	**Totaal**
Man	120	80	200
Vrouw	170	230	400
Totaal	290	310	600

Bereken de verwachte frequenties met het algoritme (rijtotaal maal kolomtotaal)/absoluut totaal. Zo vind je de volgende tabel voor de verwachte frequenties:

	Coca Cola	**Pepsi**	**Totaal**
Man	96,667	103,333	200
Vrouw	193,333	206,667	400
Totaal	290	310	600

Nu bereken je de chikwadraatwaarde:

O	**E**	**O – E**	$(O-E)^2$	$\frac{(O-E)^2}{E}$
120	96,667	23,333	544,429	5,632
80	103,333	-23,333	544,429	5,269
170	193,333	-23,333	544,429	2,816
230	206,667	23,333	544,429	2,634
600	600	0		16,351

Je vindt

$$X^2 = \sum \frac{(O-E)^2}{E} = \mathbf{16{,}351}.$$

Stap 3

Neem de tabel van de chikwadraatverdeling bij de hand met de ingangen:

- $\alpha = 5\%$;
- (aantal rijen min één) maal (aantal kolommen min één) = (2 – 1) * (2 – 1) = 1 vrijheisgraad.

De tabelwaarde is **3,841**.

Stap 4

Conclusie: **16,351 is groter dan 3,841**, dus er is een significant verschil in voorkeur voor Coca Cola en Pepsi tussen mannen en vrouwen.

b. Alle verwachte frequenties zijn groter of gelijk 5, dus er is hier voldaan aan de voorwaarde voor toepassing van de chikwadraattoets.

Opgave 5

a. Zelf.
b. Allereerst maak je een hulptabel:

X (in 100 euro)	Y (in 1.000 euro)	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
43	120	-11,333	-5	56,665	128,437	25
43	130	-11,333	5	-56.665	128,437	25
54	125	-0,333	0	0	0,111	0
58	110	3,667	-15	-55,005	13,447	225
66	140	11,667	15	175,005	136,119	25
62	125	7,667	0	0	58,783	0
326	750	0	0	120	465,334	500

Hierin is

$$\bar{X} = 326/6 = 54{,}333 \text{ en}$$

$$\bar{Y} = 750/6 = 125.$$

Dan bereken je a en b:

$$b = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sum (X-\bar{X})^2} = \frac{120}{465{,}334} = 0{,}258.$$

$$a = \bar{Y} - b\,\bar{X} = 125 - 0{,}258 * 54{,}333 = 110{,}982.$$

Hieruit volgt dat $\hat{Y} = a + bX = 110{,}982 + 0{,}258X$.

c. € 5.000,- betekent X = 50. Als je dit invult inde regressievergelijking, dan vind je $\hat{Y} = 10{,}982 + 0{,}258 * 50 = 123{,}882$. Vermenigvuldiging met € 1.000,- levert een voorspelling op van **€ 123.883**.
d. Trefwoord is correlatie – toetsen op correlatie – paragraaf 10.5.

Stap 1

$H_0 : \rho = 0$
$H_1 : \rho > 0$

Je toetst met α = 5%.

Stap 2

Allereerst bereken je de correlatiecoëfficiënt. Deze is:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 . \sum (Y - \bar{Y})^2}} = \frac{120}{\sqrt{465{,}334 * 500}} = 0{,}249$$

Daarna bereken je de t-waarde:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}249\sqrt{6-2}}{\sqrt{1-0{,}249^2}} = 0{,}514$$

Stap 3

Je vindt de tabelwaarde met de ingangen:

- 6 – 2 = 4 vrijheidsgraden
- α = 0,05

Je vindt in bijlage 2 de tabelwaarde **2,132**.

Stap 4

Omdat de berekende t-waarde t = **0,514** kleiner is dan de tabelwaarde **2,132** is er niet sprake van een significant positief lineair verband tussen advertentiekosten per maand en omzet per maand.

Opgave 6

Trefwoorden: percentage – toetsen – toetsen op twee percentages – toetsen op gelijkheid van twee percentages – paragraaf 7.4.

Controleer voor je de toets toepast of aan de voorwaarden voldaan is:

90 ≥ 5, 660 ≥5, 120 ≥ 5 en 630 ≥ 5.

Stap 1

$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$

Bij deze toets is α = 10%.

Stap 2

Allereerst bereken je het gepoolde percentage (90 + 120)/(750 + 750) = 14%. Daarna bereken je

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}.(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{12{,}0\% - 16\%}{\sqrt{14\% * 86\% * (\frac{1}{750}+\frac{1}{750})}} = -2{,}23$$

Stap 3

De hier geformuleerde toets is tweezijdig. Bij een α van 10% hoort de bovengrens **1,645** en de ondergrens **-1,645**.

Stap 4

Conclusie: **-2,23 < -1,645**, dus er is een significant verschil in de spontane naamsbekendheid van Heineken in Hamburg en Berlijn .

Bij deze toets gaan we er van uit dat de steekproefaantallen minimaal 200 zijn en achterliggende populaties zeer groot. Verder moet voor beide populaties gelden dat np ≥ 5 en n(1 – p) ≥ 5. Dit kun je controleren door naar het aantal successen en het aantal mislukkingen te kijken in beide steekproeven. In het voorbeeld zijn deze 40 ≥ 5, 460 ≥5, 34 ≥ 5 en 466 ≥ 5.

Opgave 7

a/b. Trefwoorden: percentage – toetsen – toetsen op meer dan twee percentages – paragraaf 7.6. Aan de voorwaarde dat alle verwachte frequenties minimaal vijf zijn, is voldaan, omdat alle verwachte frequenties 20 zijn.

Stap 1

$H_0 : p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9 = 10\%$
$H_1 :$ Minstens één percentage is niet gelijk aan 10%

Met α = 5%.

Stap 2
Je berekent de chikwadraatwaarde:

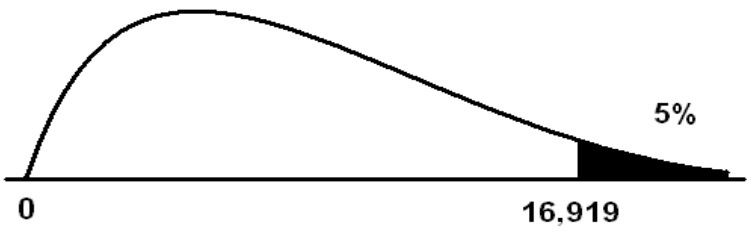
Cijfers	**O**	**E**	**O – E**	$(O-E)^2$	$\frac{(O-E)^2}{E}$
0	16	20	-4	16	0,800
1	17	20	-3	9	0,450
2	22	20	-2	4	0,200
3	23	20	3	9	0,450
4	18	20	-2	4	0,200
5	19	20	-1	1	0,050
6	21	20	1	1	0,050
7	20	20	0	0	0
8	21	20	1	1	0,050
9	23	20	3	9	0,450
Totaal	200	200	0		2,700

Stap 3

Je vindt de tabelwaarde als volgt:

- Neem bijlage 3 bij de hand.
- Kijk bij de eerste tabel met een bovengrens.
- De eerste ingang is in de eerste rij α = 0,05.
- Het aantal vrijheidsgraden is 10 – 1 = 9.
- Je vindt de bovengrens **16,919**.

Stap 4



Je ziet dat de berekende chikwadraatwaarde van **2,700** kleiner is dan de bovengrens **16,919**. De conclusie is dat je niet kunt aantonen dat er minstens één percentage significant afwijkt van 10%.

Opgave 8

a. Trefwoorden: percentage – toetsen – toetsen op één percentage – np = 800 * 0,005 = 4 < 5 – je komt uit op het toetsen met de Poissonbenadering paragraaf 7.3.

Stap 1

H_0: p ≥ 0,5%
H_1: p < 0,5%

Je werkt met α = 10%.

Stap 2

De toetsingsgrootheid is het aantal niet correcte posten in de steekproef.

Uit de gegevens blijkt dat er twee niet correcte posten zijn.

Stap 3

De toetsingsgrootheid is binomiaal verdeeld, want

- je werkt met herhaling, je onderzoekt n = 800 posten;
- er zijn twee uitkomstmogelijkheden, correcte post of een niet correcte post;
- het percentage niet correcte posten is constant omdat we te maken hebben met een zeer grote boekhouding.

Zodoende is het aantal niet correcte posten binomiaal verdeeld met n = 800 en onder de nulhypothese gebruiken we p = 0,5% als uiterste waarde om de toets uit te kunnen voeren.

Omdat n groot is, kijken we of we een benadering toe mogen passen:

- np = 800 * 0,5% = 4 < 5.
- n(1 – p) = 800 * 99,5% = 796 ≥ 5.

Je komt uit op de benadering met de Poissonverdeling.

De linkergrens vind je nu door bijlage 6 bij de hand te nemen:

- Kijk bij μ = np = 4.
- Kijk bij de eerste waarde P($\underline{k}$ ≤ 0) = 0,0183.
- Deze waarde is kleiner dan 10% = 0,10, hetgeen betekent dat 0 tot het kritieke gebied behoort.
- Kijk bij de tweede waarde P($\underline{k}$ ≤ 1). Deze waarde bedraagt 0,0916.
- Deze waarde is ook kleiner dan de onbetrouwbaarheidsdrempel α = 0,10.
- Als je kijkt naar P($\underline{k}$ ≤ 2) = 0,2381, dan is deze kans groter dan de onbetrouwbaarheidsdrempel van α = 0,10.

Het kritieke gebied is **{0,1}**.We kunnen ook zeggen dat er nul niet correcte posten dienen gevonden te worden of één niet correcte post om de conclusie te trekken dat de verantwoording goedgekeurd wordt.

Stap 4

Er zijn daadwerkelijk twee niet correcte posten gevonden in de steekproef. Om de verantwoording goed te keuren had je geen niet correcte posten moeten vinden of één niet correcte post. De nulhypothese wordt daarom niet verworpen en de verantwoording wordt nader onderzocht.

b. Trefwoorden: percentage – minimale steekproefgrootte – accountancycontrole – paragraaf 7.3. Bereken $n \geq \frac{\log \alpha}{\log(1-p)} = \frac{\log 0{,}10}{\log 0{,}995} = 459{,}36$. Je vindt **n = 460 posten**.

c. Als de alternatieve hypothese correct is, dan is μ = np = 800 * 0,25% = 2 niet correcte posten. Berekening van het onderscheidend vermogen:

P(nulhypothese verwerpen| alternatieve hypothese correct) = P($\underline{k}$ ≤ 1| μ = 2) = 0,4060 = **40,60%.**

Opgave 9

Trefwoorden: gemiddelde – toetsen op meer dan twee gemiddelden – aanname van normaliteit van de gegevens – paragraaf 9.2.

Stap 1

$H_0 : \mu_1 = \mu_2 = \mu_3$
H_1: Minstens één μ wijkt af

Met α = 5%.

Stap 2

60 50 22 38	50 73 72 80 81	40 44 63	
$T_1 = 170$	$T_2 = 356$	$T_3 = 147$	T = 673

$$SS_{totaal} = \sum (x_{ij} - \overline{x})^2 = \sum x_{ij}^{\ 2} - \frac{T^2}{n} = 60^2 + 50^2 + ... + 63^2 - \frac{673^2}{12} = 3.762{,}917$$

$$SS_{haven} = \sum (\overline{x}_i - \overline{x})^2 = \sum \frac{T_i^{\ 2}}{n_i} - \frac{T^2}{n} = \frac{170^2}{4} + \frac{356^2}{5} + \frac{147^2}{3} - \frac{673^2}{12} = 2.031{,}117$$

De laatste kwadratensom SS_{fout} vind je door de kwadratensom van het totaal te verminderen met de kwadratensom van de verpakking, dus

$$SS_{fout} = SS_{totaal} - SS_{haven} = 3.762{,}917 - 2.031{,}117 = 1.731{,}800$$

Bron	**Vrijheidsgraden**	**Kwadratensom**	**Gemiddelde Kwadratensom**	**F-waarde**
Haven Fout	k -1 = 3 – 1 = 2 n – k = 12 – 3 = 9	2.031,117 1.731,800	1.015,559 192,422	5,28
Totaal	n – 1 = 12 – 1 = 11	3.762,917		

Stap 3

De bovengrens vind je als volgt:

- De onbetrouwbaarheid α = 5%.
- De vrijheidsgraden van de teller zijn gelijk aan 2.
- De vrijheidsgraden van de noemer zijn gelijk aan 9.

Je vindt de tabelwaarde **4,26**.

Stap 4

Je ziet dat de berekende F-waarde F = **5,28** groter is dan de tabelwaarde **4,26**. De conclusie luidt dat er minimaal één populatiegemiddelde significant afwijkt.

b. Trefwoord: niet-parametrische toets – Kruskal-Wallis toets – paragraaf 11.5.

Stap 1

H_0 : Totaaloordelen van de drie mobiele telefoons zijn gelijk
H_1 : Minstens één mobiele telefoon wijkt af met het totaaloordeel

Met α = 5%.

Stap 2

Zookies		Chocopain		Tjokidz	
60	**7**	50	**5,5**	40	**3**
50	**5,5**	73	**10**	44	**4**
22	**1**	72	**9**	63	**8**
38	**2**	80	**11**		
		81	**12**		
	R_1 = **15,5**		R_2 = **47,5**		R_3 = **15**

In iedere tweede kolom zie je de rangordes staan van de gegeven data. Nu bereken je de chikwadraatwaarde:

$$\chi^2 = \frac{12}{n(n+1)}\sum\frac{R_i^{\,2}}{n_i} - 3(n+1) = \frac{12}{12.13}\{\frac{15{,}5^2}{4} + \frac{47{,}5^2}{5} + \frac{15^2}{3}\} - 3.13 = \mathbf{6{,}101.}$$

Stap 3

- α = 5% = 0,05;
- het aantal vrijheidsgraden is 3 – 1 = 2.

De tabelwaarde is **5,991**.

Stap 4

Omdat **6,101 > 5,991**, luidt de conclusie dat er minimaal één mobiele telefoon significant afwijkt met het totaaloordeel.

Opgave 10

a. Trefwoorden: Gemiddelde – betrouwbaarheidsinterval – kleine n – paragraaf 4.2.

$$\bar{x}_1 \pm t_{\alpha/2;n_1-1} \cdot \frac{s_1}{\sqrt{n_1}} = 10.000 \pm 2{,}776 * \frac{2.549{,}51}{\sqrt{5}} = 10.000 \pm 3.165{,}13$$

Het eerste betrouwbaarheidsinterval loopt van **6.834,87** euro tot **13.165,13** euro.

$$\bar{x}_2 \pm t_{\alpha/2;n_2-1} \cdot \frac{s_2}{\sqrt{n_2}} = 12.000 \pm 2{,}776 * \frac{1.581{,}14}{\sqrt{5}} = 12.000 \pm 1.962{,}93$$

Het tweede betrouwbaarheidsinterval loopt van **10.037,07** euro tot **13.962,93** euro.

b. Trefwoorden: Standaarddeviatie – toetsen op twee standaarddeviaties – paragraaf 5.3.

Stap 1

$H_0 : \sigma_1 = \sigma_2$
$H_1 : \sigma_1 \neq \sigma_2$

Hierbij is α = 10%.

Stap 2

Je berekent de F-waarde met de grootste standaarddeviatie in de teller:

$$F = \frac{s_1^{\,2}}{s_2^{\,2}} = \frac{2.549{,}51^2}{1.581{,}14^2} = 2{,}60$$

Stap 3

De ingangen voor de tabel zijn:

- α/2 = 10%/2 = 5% = 0,05;
- $n_1 - 1$ = 5 – 1 = 4 is het aantal vrijheidsgraden van de teller.
- $n_2 - 1$ = 5 – 1 = 4 is het aantal vrijheidsgraden van de noemer.

De tabelwaarde is **6,39**.

Stap 4

Conclusie: de berekende F-waarde **2,60** is kleiner dan de bovengrens **6,39**, hetgeen betekent dat de populatiestandaarddeviaties niet significant verschillen.

c. Trefwoorden: Gemiddelde – toetsen – toetsen op twee gemiddelden – kleine n – onafhankelijke steekproeven – paragraaf 4.3.

Stap 1

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

Met α = 5%.

Stap 2

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{10.000 - 12.000}{2.121{,}32\sqrt{\frac{1}{5} + \frac{1}{5}}} = 1{,}491 .$$

Hierbij is $s^2 = \frac{(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2}}{n_1 + n_2 - 2} = \frac{(5-1) * 2.549{,}51^2 + (5-1) * 1.581{,}14^2}{5 + 5 - 2}$, je vindt s = 2.121,32 euro.

Stap 3

- α/2 = 0,025;
- het aantal vrijheidsgraden is 5 + 5 – 2 = 8;
- de tabelwaarde is 2,306.

De ondergrens is **-2,306** en de bovengrens is **2,306**.

Stap 4

De conclusie: **-2,306 < 1,491 < 2,306**, je kunt niet beweren met 95% betrouwbaarheid dat er een verschil is in de gemiddelde dagomzetten van beide transporteurs.