Hoofdstuk 2. Het ABC van de statistiek – Oefenreeks – Oplossingen

Oefening 2.1. Intervalschattingen.

Antwoord: Het 95 % betrouwbaarheidsinterval is 164.93 – 171.07

Stap 1. Je gebruikt hiervoor de onderstaande formule voor een betrouwbaarheidsinterval van het populatiegemiddelde op basis van de z-verdeling:

$$\mu_X = \overline{X} \pm z_{\alpha/2} * \frac{\sigma_X}{\sqrt{n}}$$

Waarbij:	
μ_X = populatiegemiddelde	$\overline{X}$ = steekproefgemiddelde
$\frac{\sigma_X}{\sqrt{n}}$ = standaardfout van het gemiddelde	n = steekproefomvang
$z_{\alpha/2}$ = z-score van gekozen betrouwbaarheidsinterval	

$z_{\alpha/2}$ is bij α = 5 % (100 % – 95 %) gelijk aan 1.96

bij α = 1 % (100 % – 99 %) gelijk aan 2.58

Stap 2. Formule oplossen:

$$\mu_X = 168 \pm 1.96 * \frac{35}{\sqrt{500}} = 168 \pm 3.07$$

Oefening 2.2. Intervalschattingen.

Antwoord: Het 99 % betrouwbaarheidsinterval is 163.96 – 172.04

Stap 1. Je gebruikt hiervoor de onderstaande formule voor een betrouwbaarheidsinterval van het populatiegemiddelde op basis van de z-verdeling:

$$\mu_X = \overline{X} \pm z_{\alpha/2} * \frac{\sigma_X}{\sqrt{n}}$$

Waarbij:	
μ_X = populatiegemiddelde	$\overline{X}$ = steekproefgemiddelde
$\frac{\sigma_X}{\sqrt{n}}$ = standaardfout van het gemiddelde	n = steekproefomvang
$z_{\alpha/2}$ = z-score van gekozen betrouwbaarheidsinterval	

$z_{\alpha/2}$ is bij α = 5 % (100 % – 95 %) gelijk aan 1.96

bij α = 1 % (100 % – 99 %) gelijk aan 2.58

Stap 2. Formule oplossen:

$$\mu_X = 168 \pm 2.58 * \frac{35}{\sqrt{500}} = 168 \pm 4.04$$

Oefening 2.3. Intervalschattingen.

Antwoord: Het 95 % betrouwbaarheidsinterval is 41.73 % – 48.27 %

Stap 1. Je gebruikt hiervoor de onderstaande formule voor een betrouwbaarheidsinterval van de populatieproportie op basis van de z-verdeling:

$$\pi_X = p \pm z_{\alpha/2} * \sqrt{\frac{(p)(1-p)}{n}}$$

Waarbij:

π_X = populatieproportie

p = steekproefproportie

$\sqrt{\frac{(p)(1-p)}{n}}$ = standaardfout van de proportie

n = steekproefomvang

$z_{\alpha/2}$ = z-score van gekozen betrouwbaarheidsinterval

$z_{\alpha/2}$ is bij α = 5 % (100 % – 95 %) gelijk aan 1.96

bij α = 1 % (100 % – 99 %) gelijk aan 2.58

Stap 2. Formule oplossen:

$$\mu_X = 0.45 \pm 1.96 * \sqrt{\frac{(0.45)(1-0.45)}{888}} = 0.45 \pm 0.0327$$

Oefening 2.4 Hypothesen toetsen.

Antwoord: Significant verschil: JA

De toetsingsgrootheid is **t(1099) = -2.88** met **p = .005 (tweezijdig toetsen). Indien eenzijdig toetsen: p = .0025**

Stap 1. Stel de nulhypothese (H_0) en de alternatieve hypothese (H_a) op uit de bovenstaande opgave:

H_0: *De gemiddelde score op levenstevredenheid in de steekproef van Vlamingen is gelijk aan het gemiddelde in de populatie België. Het steekproefgemiddelde is gelijk aan 7.3.*

H_a: *De gemiddelde score op levenstevredenheid in de steekproef van Vlamingen is niet gelijk aan het gemiddelde in de populatie België. Het steekproefgemiddelde is significant lager of hoger dan 7.3.*

Stap 2. Om na te gaan of de nulhypothese aanvaard of verworpen wordt, gebruik je de onderstaande formule voor een toetsingsgrootheid t, op basis van de t-verdeling met vrijheidsgraden df = n – 1:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Waarbij:

μ_0 = een bepaald populatiegemiddelde

$\bar{X}$ = steekproefgemiddelde

s = steekproefstandaardafwijking

n = steekproefomvang

Stap 3. Formule oplossen:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{7.1 - 7.3}{2.3/\sqrt{1100}} = -2.88$$

Stap 4. Bepalen van p-waarde op basis van t-tabel:

Tabel 2 in appendix 1 geeft de p-waarden mee voor de verschillende mogelijke t-verdelingen. Om te bepalen welke t-verdeling we moeten gebruiken, bepaal je eerst de vrijheidsgraden (df = n – 1). Hier is df gelijk aan 1099. Omdat de rij van de t-verdeling met df = 1000 hier het dichtst bij ligt, gebruiken we die rij in de tabel. De absolute waarde van onze berekende t-waarde is gelijk aan 2.88, wat tussen 2.813 en 3.098 ligt. Je neemt hier de laagste van deze twee en leest bovenaan de tabel af welke overschrijdingskans voor een rechts eenzijdige t-toets daarmee overeenkomt: p = .0025. Aangezien H_a hier tweezijdig opgesteld werd, moeten we die p-waarde verdubbelen: p = .005.

Stap 5. Vergelijken van p-waarde met minimaal significantieniveau α = .05

Onze berekende p-waarde (.005) is kleiner dan α (.05), waardoor je de nulhypothese kunt verwerpen en er een significant verschil is tussen het steekproefgemiddelde en het vooropgestelde populatiegemiddelde.

Oefening 2.5. Hypothesen toetsen.

Antwoord: Significant verschil: NEE

De toetsingsgrootheid is **t(1798) = - 1.62** met **p = .20 (tweezijdig toetsen). Indien eenzijdig toetsen: p = .10**

Stap 1. Stel de nulhypothese (H_0) en de alternatieve hypothese (H_a) op uit de bovenstaande opgave:

> H_0: *Er is geen verband tussen herkomst en levenstevredenheid: Vlamingen en Walen verschillen niet van elkaar op het vlak van levenstevredenheid.*
>
> H_a: *Er is een verband tussen herkomst en levenstevredenheid: een van beide groepen (Vlamingen of Walen) heeft gemiddeld gezien een betere levenstevredenheid dan de andere groep.*

Stap 2. Om na te gaan of de nulhypothese aanvaard of verworpen wordt, gebruik je de onderstaande formule voor een toetsingsgrootheid t, op basis van de t-verdeling met vrijheidsgraden df = $n_1 + n_2 - 2$:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Waarbij:

$$s_p = \frac{(n_1-1)\,s_1^2 + (n_2-1)\,s_2^2}{(n_1-1) + (n_2-1)}$$

Stap 3. Formule oplossen:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{7.1 - 7.5}{5.12\sqrt{\frac{1}{1100} + \frac{1}{700}}} = -1.62$$

Stap 4. Bepalen van p-waarde op basis van t-tabel:

Tabel 2 in appendix 1 geeft de p-waarden mee van enkele specifieke t-waarden, per t-verdeling. Om te bepalen welke t-verdeling we moeten gebruiken, bepaal je eerst de vrijheidsgraden (df = $n_1 + n_2 - 2$). Hier is df gelijk aan 1798. Omdat de rij van de t-verdeling met df = 1000 hier het dichtst bij ligt, gebruiken we die rij in de tabel. De absolute waarde van onze berekende t-waarde is gelijk aan 1.62, wat tussen 1.282 en 1.646 ligt. Je neemt hier de laagste van deze twee en leest bovenaan de tabel af welke overschrijdingskans voor een rechts eenzijdige t-toets daarmee overeenkomt: p = .10. Aangezien H_a hier tweezijdig opgesteld werd, moeten we die p-waarde verdubbelen: p = .20.

Stap 5. Vergelijken van p-waarde met minimaal significantieniveau α = .05

Onze berekende p-waarde (.20) is groter dan α (.05), waardoor je de nulhypothese moet aanvaarden. Er is geen significant verschil tussen de twee steekproefgemiddelden.

Oefening 2.6. Hypothesen toetsen.

Antwoord: Significant verschil: JA

De toetsingsgrootheid is **t(80) = -3.21** met **p = .001 (eenzijdig toetsen)**

Stap 1. Stel de nulhypothese (H_0) en de alternatieve hypothese (H_a) op uit de bovenstaande opgave:

> H_0: *De gemiddelde score op intelligentie in de steekproef van studenten sociale wetenschappen is gelijk aan het gemiddelde in de populatie geneeskundestudenten. Het steekproefgemiddelde is gelijk aan 121.*
>
> H_a: *De gemiddelde score op intelligentie in de steekproef van studenten sociale wetenschappen is niet gelijk aan het gemiddelde in de populatie België. Het steekproefgemiddelde is significant lager dan 121.*

Stap 2. Om na te gaan of de nulhypothese aanvaard of verworpen wordt, gebruik je de onderstaande formule voor een toetsingsgrootheid t, op basis van de t-verdeling met vrijheidsgraden df = n – 1:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Waarbij:	
μ_0 = een bepaald populatiegemiddelde	$\bar{X}$ = steekproefgemiddelde
s = steekproefstandaardafwijking	n = steekproefomvang

Stap 3. Formule oplossen:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{119 - 121}{5.6/\sqrt{81}} = -3.21$$

Stap 4. Bepalen van p-waarde op basis van t-tabel:

Tabel 2 in appendix 1 geeft de p-waarden mee van enkele specifieke t-waarden, per t-verdeling. Om te bepalen welke t-verdeling we moeten gebruiken, bepaal je eerst de vrijheidsgraden (df = n – 1). Hier is df gelijk aan 80. Neem hier dus de rij van de t-verdeling met df = 80. De absolute waarde van onze berekende t-waarde is gelijk aan 3.21, wat tussen 3.195 en 3.416 ligt. Je neemt hier de laagste van deze twee en leest bovenaan de tabel af welke overschrijdingskans voor een rechts eenzijdige t-toets daarmee overeenkomt: p = .001. Aangezien H_a hier eenzijdig opgesteld werd, moeten we de p-waarde hier niet verdubbelen.

Stap 5. Vergelijken van p-waarde met minimaal significantieniveau α = .05

Onze berekende p-waarde (.001) is kleiner dan α (.05), waardoor je de nulhypothese kunt verwerpen en er een significant verschil is tussen het steekproefgemiddelde en het vooropgestelde populatiegemiddelde.