Defensie Ondersteuningscommando
Ministerie van Defensie

Statistiek: college 11

Verschiltoetsen

DOSCO

Dr. ir. Danny Blom
Nederlandse Defensie Academie
Faculteit Militaire Wetenschappen

Mei 2025

Recap

- Punt- en intervalschatters
- Betrouwbaarheids- en voorspellingsintervallen
- De methode van hypothesetoetsen
  - Conclusies trekken op basis van het kritieke gebied of de p-waarde
- Toetsen op verbanden tussen twee variabelen: chikwadraat

Leerdoelen

Aan het eind van dit college kunnen studenten:

- Uitleggen waarom en wanneer verschiltoetsen worden toegepast in kwantitatief onderzoek
- De verschillende situaties kunnen benoemen waarin verschiltoetsen een rol spelen.
- De juiste verschiltoets kiezen op basis van de onderzoeksvraag die ze voor zich hebben
- De uitkomsten van een verschiltoets interpreteren en duiden in de context van een onderzoeksvraag

Recap: verschil tussen twee kansvariabelen

Binnen DCBRNC (Defensie CBRN Centrum) wordt gekeken of de operationele inzet te versnellen is met een nieuw decontaminatieprotocol in een biologisch verontreinigd inzetgebied. Om de effectiviteit hiervan te beoordelen worden twee groepen CBRN-teams vergeleken:

- De eerste groep werkt met het oude protocol (**controlegroep**)
- De tweede groep werkt met het nieuwe protocol (**experimentele groep**)

X: responstijd onder het oude protocol (in min) van "alarmering" tot "operationeel gereed"
Y: responstijd onder het nieuwe protocol (in min) van "alarmering" tot "operationeel gereed"

Stel nu dat we weten dat $X \sim N(\mu_X = ?; \sigma_X = 3)$ en $Y \sim N(\mu_Y = ?; \sigma_Y = 1)$.

Hoe kunnen we toetsen of het nieuwe protocol gemiddeld tot kortere responstijden leidt (in andere woorden: is $\boldsymbol{\mu_X \geq \mu_Y}$)?

Recap: verschil tussen twee kansvariabelen

Stel nu dat we weten dat $X \sim N(\mu_X = ?\,; \sigma_X = 3)$ en $Y \sim N(\mu_Y = ?\,; \sigma_Y = 1)$
Verder nemen we steekproeven van respectievelijk $n = 15$ en $m = 10$ CBRN-teams.

In dit geval kunnen we kijken naar de verschilvariabele

$$V = \overline{X} - \overline{Y} \sim N\left(\mu_V = \mu_X - \mu_Y; \sigma_V = \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)$$

- $\mu_V < 0$: het nieuwe protocol leidt tot gemiddeld **langere** responstijden.
- $\mu_V = 0$: de responstijden zijn gemiddeld **even lang** voor beide protocollen.
- $\mu_V > 0$: het nieuwe protocol leidt tot gemiddeld **kortere** responstijden

Recap: verschil tussen twee kansvariabelen

In dit geval kunnen we kijken naar de **verschilvariabele**

$$V = \overline{X} - \overline{Y} \sim N\left(\mu = \mu_X - \mu_Y; \sigma = \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)$$

Merk op dat we in dit geval twee steekproeven kunnen combineren tot één variabele.
Om te toetsen of de gemiddelde responstijd lager is bij het nieuwe protocol, gebruiken we

$$H_0: \mu \geq 0 \text{ versus } H_1: \mu < 0$$

De toetsingsmethode hiervoor hebben we in het college over hypothesetoetsen behandeld!

Verschiltoetsen

In de praktijk weten we echter vaak helemaal niet de precieze waardes van σ_1 en σ_2.

https://www.defensie.nl/onderwerpen/cbrn-centrum/nationaal-trainingscentrum-cbrn

Stel nu dat we weten dat $X \sim N(\mu_X = ?\,; \sigma_X = ?)$ en $Y \sim N(\mu_Y = ?\,; \sigma_Y = ?)$

Hoe kunnen we toetsen of het nieuwe protocol gemiddeld tot kortere responstijden leidt (in andere woorden: is $\mu_X \geq \mu_Y$)?

Rekenvoorbeeld:

Stel dat we een hypothesetoets willen uitvoeren met de volgende data:

X: responstijd onder oude protocol (in min) van “alarmering” tot “operationeel gereed”
Y: responstijd onder nieuwe protocol (in min) van “alarmering” tot “operationeel gereed”

Verder nemen we steekproeven met de volgende steekproefresultaten:

- Gemiddelde $\overline{x} = 18{,}2$ min, standaardafwijking $s_x = 2{,}5$ min en steekproefgrootte $n = 15$
- Gemiddelde $\overline{y} = 15{,}7$ min, standaardafwijking $s_Y = 2{,}2$ min en steekproefgrootte $m = 10$

Verder kiezen we een significantieniveau van $\alpha = 0{,}02$.

Hoe kunnen we nu toetsen of de gemiddelde responstijd gelijk is voor beide protocollen?

Stap 1: definieer de nul- en alternatieve hypothese

In deze hypothesetoets gaan we voor de **nulhypothese** er van uit dat het nieuwe protocol **niet** tot gemiddeld kortere responstijden leidt, oftewel

$$H_0: \mu_X \leq \mu_Y$$

Daartegenover staat de **alternatieve hypothese** dat het nieuwe protocol **wel** tot gemiddeld kortere responstijden leidt, oftewel:

$$H_1: \mu_X > \mu_Y$$

Stap 2: bepaal het significantieniveau α

In deze hypothesetoets gaan we uit van een significantieniveau van $\alpha = 0{,}02$:

Wat is de betekenis van een type-I en type-II fout in dit geval?

- **Type-I fout (onterecht verwerpen van H_0):** er wordt onterecht aangenomen dat het nieuwe CBRN-protocol de gemiddelde responstijden verlaagt, terwijl dit niet zo is.
- **Type-II fout (onterecht accepteren van H_0):** er wordt onterecht aangenomen dat het nieuwe CBRN-protocol de gemiddelde responstijden **niet** verlaagt, terwijl dit wel zo is.

Welke fout is erger?

Stap 3: verzamelen van data

In deze hypothesetoets gaan we uit van de volgende gegevens:

	Oude protocol	**Nieuwe protocol**
Steekproefgemiddelde	$\overline{x} = 18{,}2$ minuten	$\overline{y} = 15{,}7$ min
Steekproefstandaardafwijking	$s_X = 2{,}5$ minuten	$s_Y = 2{,}2$ minuten
Steekproefgrootte	$n = 15$	$m = 10$

NB: op het tentamen krijg je mogelijk twee datasets aangereikt en moet je zelf nog deze statistieken bepalen.

Stap 4: bepaal de toetsingsgrootheid

Opnieuw kunnen we kijken naar de **verschilvariabele**

$$V = \overline{X} - \overline{Y} \sim N\left(\mu = \mu_X - \mu_Y; \sigma = \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)$$

De geobserveerde toetsingsgrootheid is $v = \overline{x} - \overline{y} = 18{,}2 - 15{,}7 = 2{,}5$ minuten.
Het probleem is echter dat we nu zowel μ als σ niet weten.

Wat zouden we kunnen doen?

Stap 4: bepaal de toetsingsgrootheid

Het meeste intuïtieve idee is zowel σ_X als σ_Y schatten met s_X en s_Y en daarna met een t-verdeling werken. Dit geeft de schatting s_V van de standaardafwijking σ_V:

$$s_V = \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$$

Onder de nulhypothese H_0 ($\mu_X = \mu_Y$) levert dit een ***t*-score** op van:

$$t = \frac{v - \mu_V}{s_V} = \frac{v - (\mu_X - \mu_Y)}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}} = \frac{2{,}5 - 0}{\sqrt{\frac{2{,}5^2}{15} + \frac{2{,}2^2}{10}}} \approx 2{,}6343$$

Dit werkt ook in principe, maar we kunnen wat beters doen als $\sigma_X = \sigma_Y$!

Stap 4: bepaal de toetsingsgrootheid

Als $\sigma = \sigma_X = \sigma_Y$, dan zijn s_X en s_Y beide een schatter van dezelfde onbekende σ. Om een betere schatting te krijgen, mogen we in dat geval werken met een gewogen gemiddelde hiervan (**pooled variance**):

$$s_P^2 = \frac{(n-1)\cdot s_X^2 + (m-1)\cdot s_Y^2}{n-1+m-1} = \frac{(15-1)\cdot 2{,}5^2 + (10-1)\cdot 2{,}2^2}{(15-1+10-1)} \approx 5{,}6983$$

Onder de nulhypothese H_0 ($\mu_X = \mu_Y$) levert dit een ***t*-score** op van:

$$t = \frac{v-\mu_V}{s_V} = \frac{v-(\mu_X-\mu_Y)}{\sqrt{\frac{s_P^2}{n}+\frac{s_P^2}{m}}} = \frac{2{,}5-0}{\sqrt{\frac{5{,}6983}{15}+\frac{5{,}6983}{10}}} \approx 2{,}5653$$

Stap 5: geef een conclusie en formuleer deze in de originele context

We berekenen het **kritieke gebied** $[g; \infty)$ en kijken of de berekende t-score daarin valt:

Geval 1: $\boldsymbol{\sigma_X \neq \sigma_Y}$

- **Berekende $\boldsymbol{t}$-score:** $t \approx 2{,}6343$
- **Grens kritiek gebied (rechtszijdige toets!):**

$$g = \mathrm{InvT}(\mathrm{opp} = 1 - \alpha; \mathrm{df} = \min(n-1, m-1)) = \mathrm{InvT}(0{,}98; 9) \approx 2{,}3984$$

Omdat $t > g$, ligt de t-score in het **kritieke gebied** → verwerpen van H_0

Er is voldoende bewijs om aan te nemen dat het nieuwe protocol inderdaad tot gemiddelde kortere responstijden leidt dan het oude protocol!

Stap 5: geef een conclusie en formuleer deze in de originele context

We berekenen het **kritieke gebied** $[g; \infty)$ en kijken of de berekende t-score daarin valt:

Geval 2: $\sigma_X = \sigma_Y$

- **Berekende t-score:** $t \approx 2{,}5653$
- **Grens kritiek gebied (rechtszijdige toets!)**:

$$g = \text{InvT}(\text{opp} = 1 - \alpha; \text{df} = n + m - 2) = \text{InvT}(0{,}98; 23) \approx 2{,}1770$$

Omdat $t > g$, ligt de t-score in het **kritieke gebied** → verwerpen van H_0

Er is voldoende bewijs om aan te nemen dat het nieuwe protocol inderdaad tot gemiddelde kortere responstijden leidt dan het oude protocol!

F-toets voor gelijke varianties

Om te bepalen of $\boldsymbol{\sigma_X = \sigma_Y}$ of $\boldsymbol{\sigma_X \neq \sigma_Y}$ geldt, moeten we een $\boldsymbol{F}$-toets uitvoeren.

Stap 1: definieer de nulhypothese $\boldsymbol{H_0}$ en de alternatieve hypothese $\boldsymbol{H_1}$

$H_0: \sigma_X = \sigma_Y$ versus $H_1: \sigma_X \neq \sigma_Y$ (**tweezijdig!**)

Stap 2: bepaal het significantieniveau $\boldsymbol{\alpha}$

We mogen werken met $\alpha = 0{,}02$.

Stap 3: verzamelen van data

We hadden al bepaald dat $s_X = 2{,}5$ minuten en $s_Y = 2{,}2$ minuten zijn gegeven.

F-toets voor gelijke varianties

Om te bepalen of $\boldsymbol{\sigma_X = \sigma_Y}$ of $\boldsymbol{\sigma_X \neq \sigma_Y}$ geldt, moeten we een $\boldsymbol{F}$-toets uitvoeren.

Stap 1: bepaal de toetsingsgrootheid.

In dit geval is de toetsingsgrootheid gelijk aan

Hierbij zijn S_A^2 en S_B^2 de puntschatters voor de variantie van A en B.

De geobserveerde toetsingsgrootheid is dus gelijk aan $f = \frac{s_A^2}{s_B^2} = \frac{2{,}5^2}{2{,}2^2} \approx 1{,}2913$

F-toets voor gelijke varianties

De toetsingsgrootheid F volgt de zogenaamde F-verdeling met respectievelijk $n-1$ en $m-1$ vrijheidsgraden.

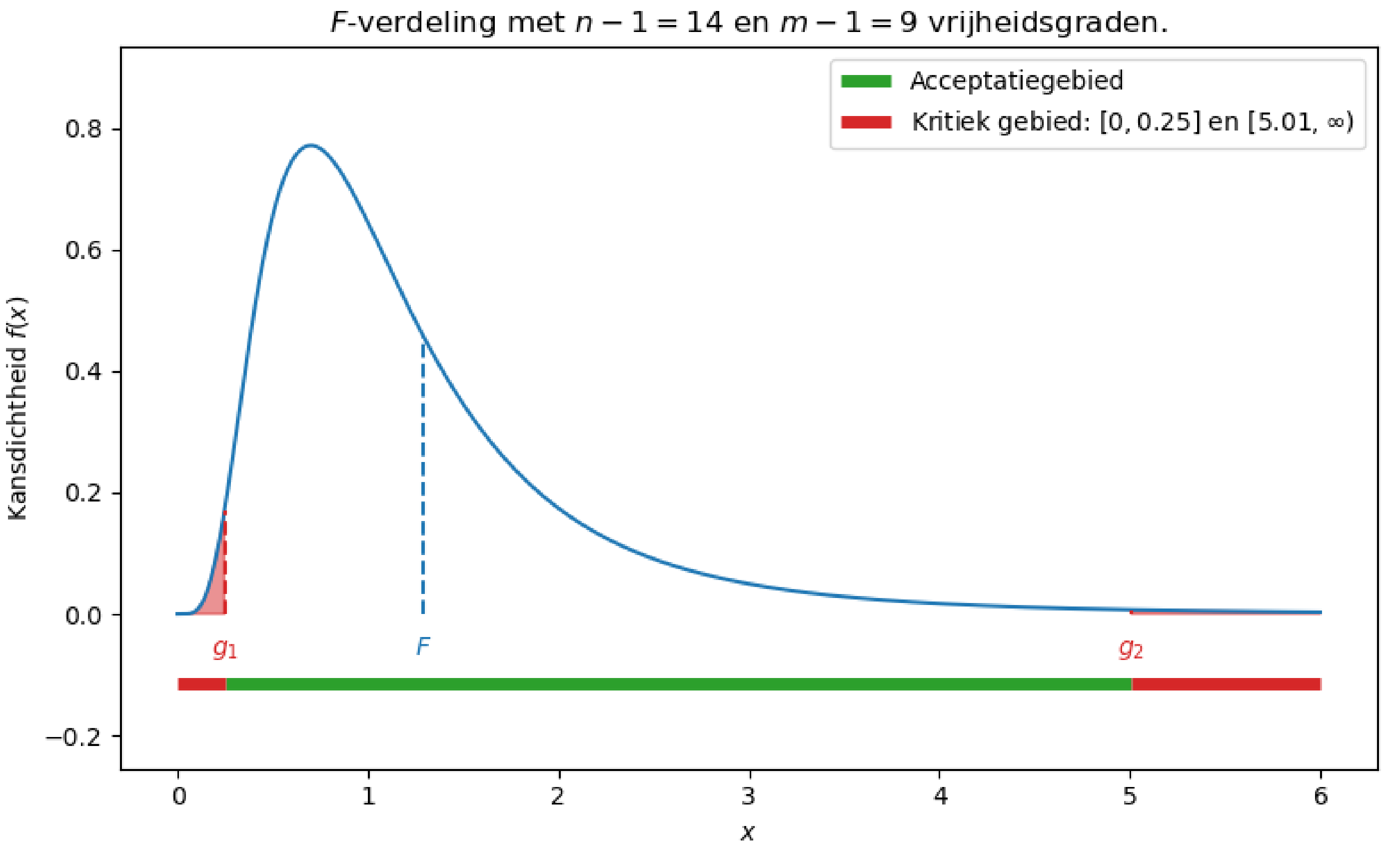
Het **kritieke gebied** is van de vorm $(-\infty; g_1]$ en $[g_2, \infty)$, waarbij:

- $\mathrm{Fcdf}(\mathrm{lower} = 0; \mathrm{upper} = g_1; \mathrm{df}_1 = n-1, \mathrm{df}_2 = m-1) = \frac{\alpha}{2} \rightarrow g_1 = 0{,}2482$
- $\mathrm{Fcdf}(\mathrm{lower} = g_2; \mathrm{upper} = 10^{99}; \mathrm{df}_1 = n-1, \mathrm{df}_2 = m-1) = \frac{\alpha}{2} \rightarrow g_2 = 5{,}0052$

$\rightarrow$ de geobserveerde toetsingsgrootheid $f = 1{,}2913$ ligt niet in het kritieke gebied, dus de nulhypothese wordt niet verworpen.

Er is onvoldoende reden om niet aan te nemen dat de standaardafwijkingen $\sigma_X = \sigma_Y$!

F-toets voor gelijke varianties



Merk op:

- als $\sigma_X = \sigma_Y$, dan is het te verwachten dat schattingen s_X en s_Y dichtbij elkaar liggen
- Oftewel $f = \frac{s_X^2}{s_Y^2}$ ligt dichtbij 1
- Fracties zijn asymmetrisch
- ($\frac{a}{b}$ kan verder van 1 te liggen dan $\frac{b}{a}$).

Schema: toetsen voor gemiddeldes bij twee onafhankelijke populaties

Zijn σ_X en σ_Y bekend?

ja

Voer een z-toets uit op de **verschilvariabele**

$$V = \overline{X} - \overline{Y} \sim N(\mu_X - \mu_Y; \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}})$$

nee

Voer een F-toets voor gelijke varianties uit:

$H_0: \sigma_X = \sigma_Y$ versus $H_1: \sigma_X \neq \sigma_Y$

Wordt H_0 verworpen?

ja

Voer een t-toets uit met toetsingsgrootheid

$$t = \frac{(\overline{x} - \overline{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$$

$$(df = \min(n-1, m-1))$$

nee

Voer een t-toets uit met toetsingsgrootheid

$$t = \frac{(\overline{x} - \overline{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{s_P^2}{n} + \frac{s_P^2}{m}}}$$

$$(df = n + m - 2)$$

Opdracht 11.20

Een transportbedrijf verzorgt de dagelijkse bevoorrading van supermarkten. Tussen het centrale magazijn in Woerden en een aantal supermarkten in Arnhem moet dagelijks een traject worden afgelegd van 80 km. (we noemen dit traject A). Helaas zijn er dikwijls, en op onvoorspelbare tijdstippen, files op dit traject. Hierdoor is de reisduur onzeker, en kunnen vanwege het tijdverlies de kosten voor het transportbedrijf soms onbeheersbaar zijn. Men overweegt daarom een alternatieve route (traject B) te kiezen. Deze is een aantal kilometers groter, maar misschien is er veel minder risico op files. Om een dergelijke keuze te onderbouwen is besloten op een aantal willekeurig gekozen dagen de route via traject A dan wel traject B te rijden. Dit leverde de volgende reistijden (in minuten):

Traject A: 72, 85, 104, 55, 90, 77, 68, 50, 36, 63

Traject B: 78, 72, 81, 74, 85, 71, 69, 84, 66, 76, 65, 79

a) Toets of de variantie van de reistijden voor traject A en traject B aan elkaar gelijk kunnen zijn (kies $\alpha = 0{,}05$).

b) Toets of de gemiddelde reistijd voor traject A en traject B hetzelfde kan zijn (kies $\alpha = 0{,}05$).

Opdracht 11.20

Toets of de variantie van de reistijden voor traject A en traject B aan elkaar gelijk kunnen zijn (kies $\boldsymbol{\alpha = 0,05}$).

Om te toetsen of de variantie van de reistijden voor traject A en traject B aan elkaar gelijk zijn, voeren we een F-toets uit.

Stap 1: definieer de nulhypothese $\boldsymbol{H_0}$ en alternatieve hypothese $\boldsymbol{H_1}$

$$H_0 : \sigma_A^2 = \sigma_B^2 \text{ versus } H_1 : \sigma_A^2 \neq \sigma_B^2$$

Stap 2: bepaal het significantieniveau $\boldsymbol{\alpha}$ van de hypothesetoets

In de vraag is gegeven dat we kunnen werken met $\alpha = 0{,}05$.

Opdracht 11.20a: toets voor gelijke variantie

Stap 3: verzamelen van data

In de opdracht werken we met de volgende gegevens

Traject A: 72, 85, 104, 55, 90, 77, 68, 50, 36, 63

Traject B: 78, 72, 81, 74, 85, 71, 69, 84, 66, 76, 65, 79

We berekenen allereerst de steekproefgemiddeldes $\overline{x}_A$ en $\overline{x}_B$ voor trajecten A en B:

- $\overline{x}_A = \frac{72+85+\cdots+63}{10} = 70$ en $\overline{x}_B = \frac{78+72+\cdots+79}{12} = 75$

Vervolgens berekenen we de steekproefvarianties s_A^2 en s_B^2 voor trajecten A en B:

- $s_A^2 = \frac{(72-70)^2+(85-70)^2+\cdots+(63-70)^2}{10-1} \approx 407{,}5555$
- $s_B^2 = \frac{(78-75)^2+(72-75)^2+\cdots+(79-75)^2}{12-1} \approx 44{,}1818$

Opdracht 11.20a: toets voor gelijke variantie

Stap 4: bereken de toetsingsgrootheid

Op de vorige slide hebben we steekproefgemiddeldes en steekproefvarianties berekend:

- $\overline{x}_A = 70,\ s_A^2 \approx 407{,}5555,\ \overline{x}_B = 75,\ s_B^2 \approx 44{,}1818$

De toetsingsgrootheid van een F-toets is gelijk aan

$$F = \frac{S_A^2}{S_B^2}$$

De geobserveerde toetsingsgrootheid is gelijk aan

$$f = \frac{s_A^2}{s_B^2} = \frac{407{,}5555}{44{,}1818} \approx 9{,}2245$$

Opdracht 11.20a: toets voor gelijke variantie

Stap 5: geef een conclusie en formuleer deze in de originele context

Methode 1 (kritiek gebied): de toetsingsgrootheid F volgt onder H_0 de $F(9, 11)$-verdeling.
Omdat we tweezijdig toetsen, is het kritieke gebied gegeven door $(-\infty; \boldsymbol{g_1}]$ **en** $[\boldsymbol{g_2}, \infty)$.

- $\text{Fcdf}(\text{lower} = 0; \text{upper} = g_1; \text{df1} = 9; \text{df2} = 11) = \frac{\alpha}{2} = 0{,}025 \rightarrow g_1 \approx 0{,}2556$
- $\text{Fcdf}(\text{lower} = g_2; \text{upper} = 10^{99}; \text{df1} = 9; \text{df2} = 11) = \frac{\alpha}{2} = 0{,}025 \rightarrow g_2 \approx 3{,}5879$

De toetsingsgrootheid $f \approx 9{,}2245$ ligt in het kritieke gebied, dus verwerp H_0.

Er is voldoende reden om aan te nemen dat de varianties NIET gelijk zijn.

Opdracht 11.20a: toets voor gelijke variantie

Stap 5: geef een conclusie en formuleer deze in de originele context

Methode 2 (p-waarde): de toetsingsgrootheid $f \approx 9{,}2245$ is onder H_0 een trekking uit de $F(9, 11)$-verdeling.
Omdat we tweezijdig toetsen, is de p-waarde (overschrijdingskans) gegeven door

$$p = \mathrm{Fcdf}(\mathrm{lower} = 9{,}2245; \mathrm{upper} = 10^{99}; \mathrm{df1} = 9; \mathrm{df2} = 11) \approx 5{,}6486 \cdot 10^{-4}$$

De p-waarde van de geobserveerde toetsingsgrootheid f is (veel) kleiner dan het significantieniveau $\alpha = 0{,}05$, dus verwerp H_0.

Er is voldoende reden om aan te nemen dat de varianties NIET gelijk zijn.

Opdracht 11.20b: toets voor gelijke gemiddeldes

b) Toets of de gemiddelde reistijd voor traject A en traject B hetzelfde kan zijn (kies $\boldsymbol{\alpha = 0,05}$).

Uit de F-toets voor gelijke variantie is gebleken dat het onwaarschijnlijk is dat traject A en traject B dezelfde varianties hebben op de reistijd.

We voeren een onafhankelijke t-toets uit met ongelijke varianties.

Stap 1: definieer de nulhypothese $\boldsymbol{H_0}$ en de alternatieve hypothese $\boldsymbol{H_1}$

$$H_0: \mu_A = \mu_B \text{ versus } H_1: \mu_A \neq \mu_B$$

Stap 2: bepaal het significantieniveau $\boldsymbol{\alpha}$ $\rightarrow$ we gaan nog steeds uit van $\alpha = 0{,}05$.

Opdracht 11.20b: toets voor gelijke gemiddeldes

Stap 3: verzamelen van data

We hebben nog steeds te maken met steekproeven voor traject A en traject B met:

$$\overline{x_A} = 70, \qquad s_A^2 \approx 407{,}5555, \qquad \overline{x_B} = 75, \qquad s_B^2 \approx 44{,}1818$$

Stap 4: bepaal de toetsingsgrootheid

Bij een onafhankelijke t-toets met ongelijke varianties is de toetsingsgrootheid gelijk aan:

$$t = \frac{(\overline{x_A} - \overline{x_B}) - (\mu_A - \mu_B)}{\sqrt{\frac{s_A^2}{n} + \frac{s_B^2}{m}}} = \frac{(70 - 75) - 0}{\sqrt{\frac{407{,}5555}{10} + \frac{44{,}1818}{12}}} \approx -0{,}7501$$

Opdracht 11.20b: toets voor gelijke gemiddeldes

Stap 5: geef een conclusie en formuleer deze in de originele context

Toetsingsgrootheid: t-verdeling met $\mathrm{df} = \min(n-1, m-1) = 9$ vrijheidsgraden.

Methode 1 (kritiek gebied): omdat we tweezijdig toetsen, is het kritieke gebied gegeven door $(-\infty; \boldsymbol{g_1}]$ **en** $[\boldsymbol{g_2}, \infty)$.

$$g_1 = \mathrm{InvT}(\mathrm{opp} = \alpha/2; \mathrm{df} = \min(n-1, m-1)) = \mathrm{InvT}(0{,}025; 9) \approx -2{,}2622$$

$$g_2 = \mathrm{InvT}(\mathrm{opp} = 1 - \alpha/2; \mathrm{df} = \min(n-1, m-1)) = \mathrm{InvT}(0{,}975; 9) \approx 2{,}2622$$

Omdat $g_1 < t < g_2$, ligt de t-score NIET in het **kritieke gebied** $\rightarrow H_0$ niet verwerpen.

Er is onvoldoende reden om aan te nemen dat de gemiddelde reistijden verschillend zouden zijn.

Opdracht 11.20b: toets voor gelijke gemiddeldes

Stap 5: geef een conclusie en formuleer deze in de originele context

Toetsingsgrootheid: t-verdeling met $\text{df} = \min(n-1, m-1) = 9$ vrijheidsgraden.

Methode 2 (p-waarde): we bekijken de linkeroverschrijdingskans van de toetsingsgrootheid $t \approx -0{,}7501$ (die is kleiner dan de rechteroverschrijdingskans):

$$p = \text{tcdf}(\text{lower} = -10^{99}; \text{upper} = -0{,}7501;\ \text{df} = 9) \approx 0{,}2362$$

Omdat de p-waarde groter is dan het significantieniveau α, wordt H_0 niet verworpen.

Er is onvoldoende reden om aan te nemen dat de gemiddelde reistijden verschillend zouden zijn.

Samenvatting

- Verschiltoetsen
  - Gemiddeldes van twee onafhankelijke populaties
  - F-verdeling

Huiswerk:

- Lezen van A. Buijs: hoofdstuk 11.1 (blz. 339), 11.2 (blz. 339-348), 11.5 (blz. 354-357)

- Opdrachten:
  - Hoofdstuk 11: m1, m3, m6, 11.1, 11.3, 11.5, 11.7, 11.12, 11.13, 11.14

Volgende les: correlatie en regressie