



STATISTIEK & EXCEL
LES 7: T-TOETSEN

HERHALING LES 6

Een machine die flesjes vult zou afgesteld moeten staan op een eindvolume van 500 mL. We doen een steekproef ($N = 28$) om te bepalen of dit inderdaad het geval is. We vinden een gemiddelde van 495 mL met een standaarddeviatie van 10 mL.

Bevatten de flesjes significant minder dan 500 mL?

- a) Nee, want de 500 mL ligt buiten het 95% BI
- b) Nee, want de 500 mL ligt binnen het 95% BI
- c) Ja, want de 500 mL ligt buiten het 95% BI
- d) Ja, want de 500 mL ligt binnen het 95% BI



HERHALING LES 6

Je wilt onderzoeken of er een verschil is in de chocoladeconsumptie tussen Brabanders en Limburgers.

Stel een nulhypothese (H_0) en een alternatieve hypothese (H_1) op.



HERHALING LES 6



Een onderzoeker wil weten of er een verschil is in de diastolische bloeddruk tussen mensen in Groningen en mensen in Utrecht. Voor de mensen in Groningen voert hij een steekproef uit (data is normaal verdeeld). Het populatie-gemiddelde voor de mensen in Utrecht is bekend. De onderzoeker voert een one-sample t-toets uit en vindt een p-waarde van 0,07. Er geldt $\alpha = 0,05$.

De onderzoeker concludeert dat er een verschil is in de bloeddruk. Heeft de onderzoeker gelijk?

- a) Ja
- b) Nee
- c) Kun je niet zeggen, want je mag geen t-toets uitvoeren op deze data

INHOUD LES 7

- One-sample versus two-sample t-toets
- Two-sample t-toets in Excel
- Gepaard versus ongepaard toetsen
- Ongepaarde toetsen: gelijke of ongelijke SD?

ONE-SAMPLE VERSUS TWO-SAMPLE T-TOETS

ONE-SAMPLE T-TOETS

We willen onderzoeken of er een verschil is tussen **een steekproef (eigen metingen)** en **een bekend populatiegemiddelde (geen eigen meting)**.

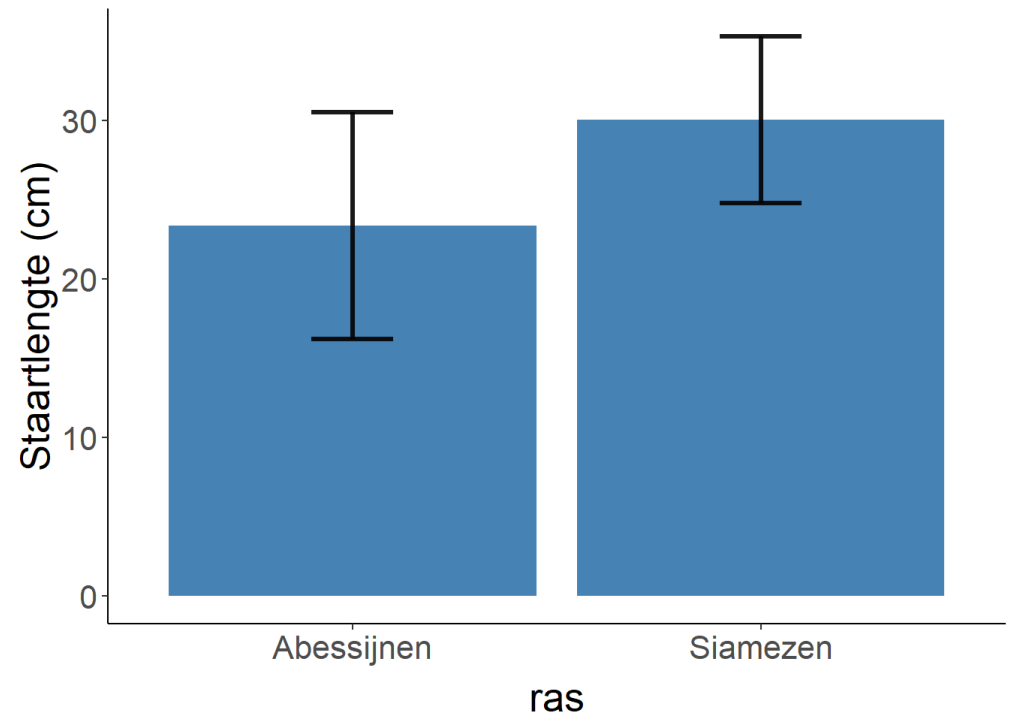
TWO-SAMPLE T-TOETS

We willen onderzoeken of er een verschil is tussen **twee steekproeven (beide steekproeven zijn eigen metingen)**.

ONE-SAMPLE VERSUS TWO-SAMPLE T-TOETS

VOORBEELD TWO-SAMPLE T-TOETS

- We meten de staarlengte van aantal Siamezen (katten) en een aantal Abessijnen (andere katten).
- H_0 : er is geen verschil in staartlengte tussen Siamezen en Abessijnen.
- Omdat de richting van het verschil onbekend is, is er sprake van een **tweezijdige two-sample t-toets**.



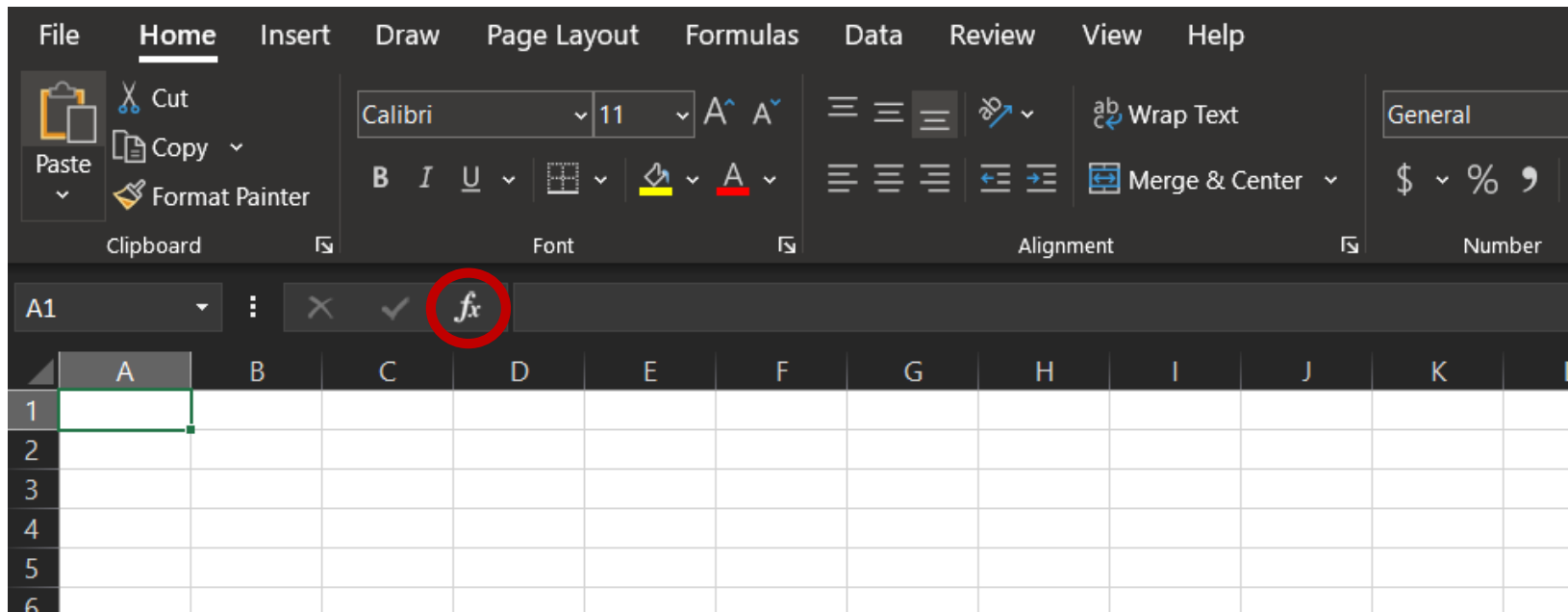
ONE-SAMPLE VERSUS TWO-SAMPLE T-TOETS

- De two-sample t-toets kun je uitvoeren in Excel (zie slides hierna). Het getal dat uit de Excelformule komt is de **p-waarde**. Als de p-waarde erg klein is, dan is het erg onwaarschijnlijk dat de nulhypothese waar is.
- Maar hoe klein is erg klein? Vaak verwerpen we de nulhypothese als **p-waarde < 0.05** is. Die grenswaarde van 0,05 noemen we ook wel **alfa: $\alpha = 0,05$** .

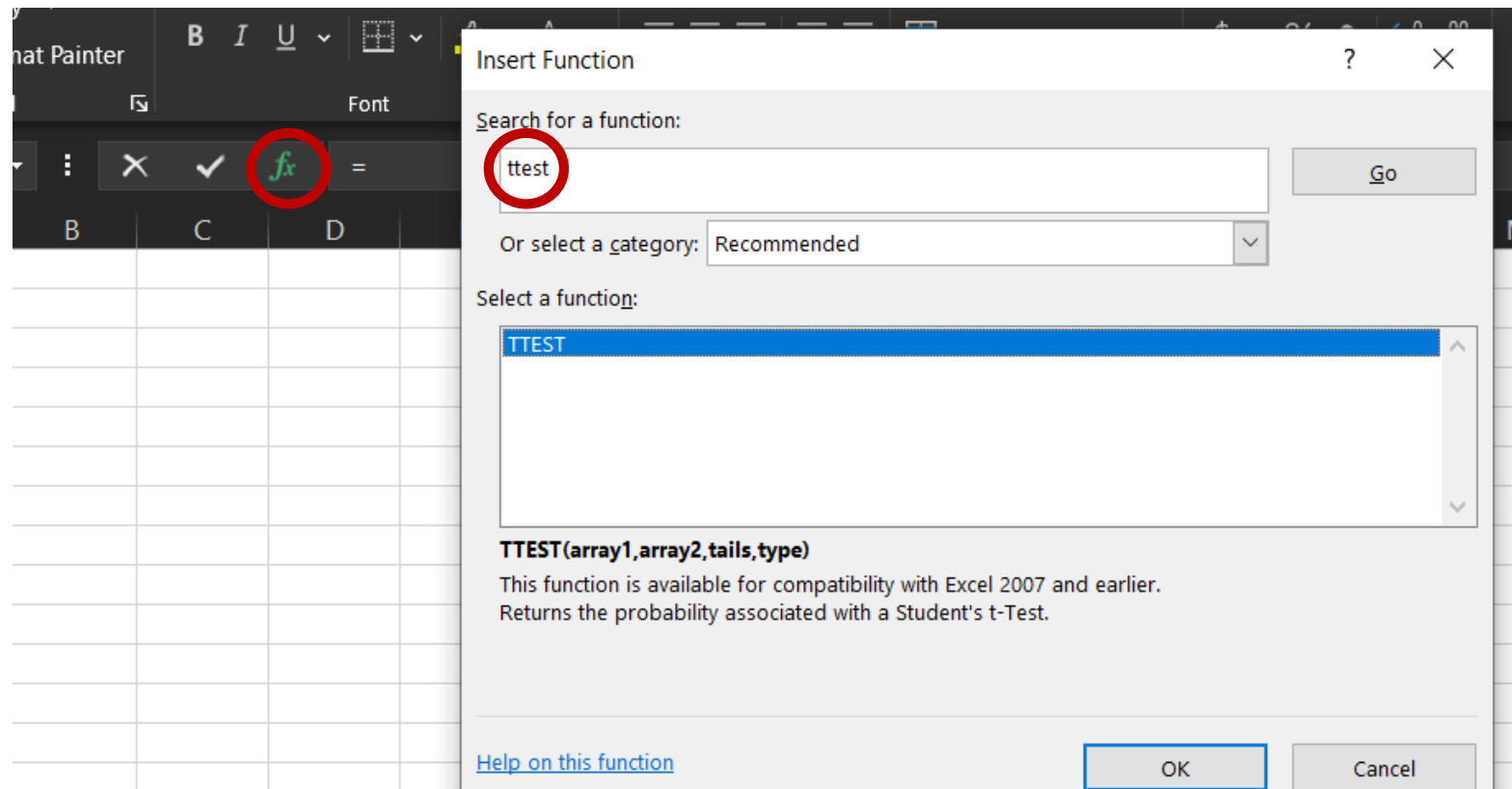
Uitkomst t-toets	Wat doen we met H0?	Wat doen we met H1?	Conclusie
p-waarde < 0,05	H0 wordt verworpen	H1 wordt aangenomen	Er is een statistisch significant verschil ...
p-waarde \geq 0,05	H0 wordt aangenomen	H1 wordt verworpen	Er is geen statistisch significant verschil ...

- **LET OP!** Je mag de t-toets alleen gebruiken als er sprake is van (ongeveer) normaal verdeelde data en van kwantitatieve data (geen telwaarden of data verdeeld in categorieën).

TWO-SAMPLE T-TESTS IN EXCEL



TWO-SAMPLE T-TESTS IN EXCEL



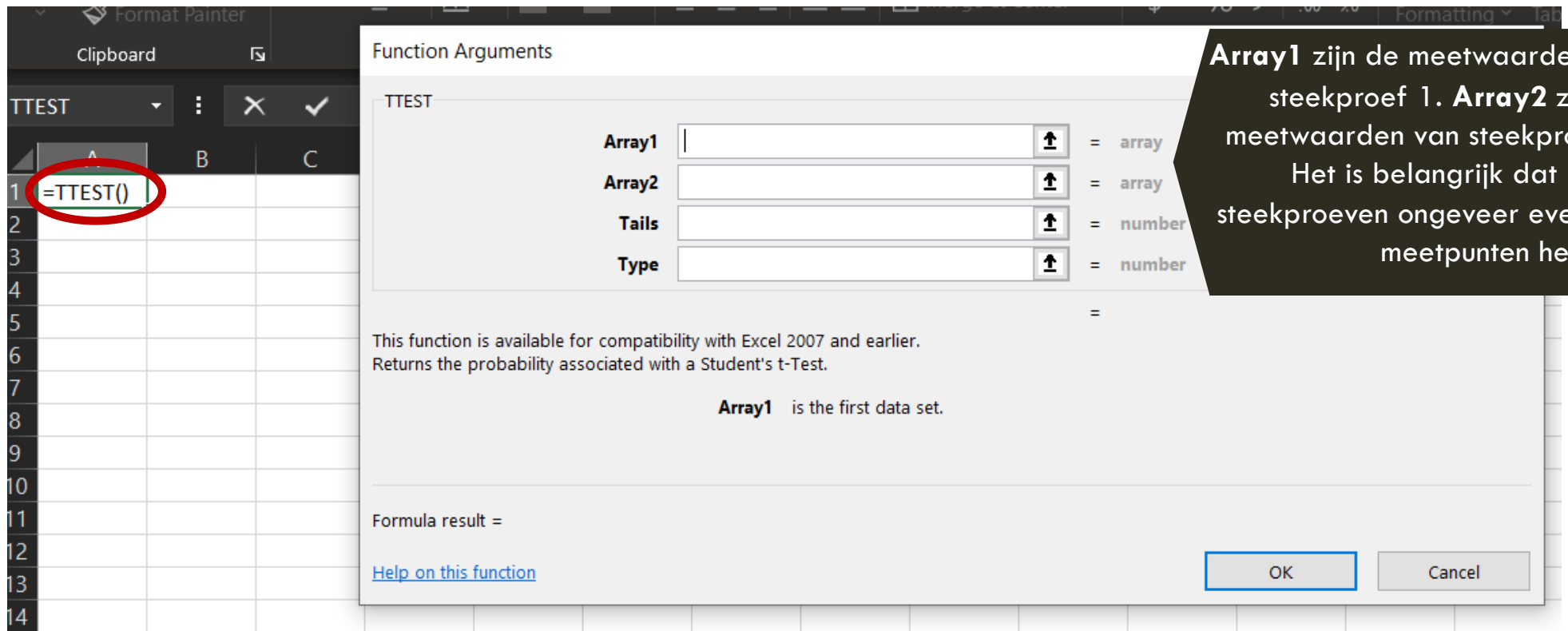
TWO-SAMPLE T-TESTS IN EXCEL

The image shows an Excel spreadsheet with the formula `=TTEST()` entered in cell A1, highlighted with a red circle. The Function Arguments dialog box is open, showing the following fields:

Argument	Value	Default
Array1		array
Array2		array
Tails		number
Type		number

Below the fields, the text reads: "This function is available for compatibility with Excel 2007 and earlier. Returns the probability associated with a Student's t-Test." Below that, it states: "Array1 is the first data set." At the bottom of the dialog, there is a "Formula result =" field, a "Help on this function" link, and "OK" and "Cancel" buttons.

TWO-SAMPLE T-TOETS IN EXCEL



The image shows an Excel spreadsheet with the formula `=TTEST()` entered in cell A1, circled in red. The 'Function Arguments' dialog box is open, showing the following fields:

Argument	Type
Array1	array
Array2	array
Tails	number
Type	number

Below the fields, the text reads: "This function is available for compatibility with Excel 2007 and earlier. Returns the probability associated with a Student's t-Test." and "Array1 is the first data set." At the bottom, there are "OK" and "Cancel" buttons.

Array1 zijn de meetwaarden van steekproef 1. **Array2** zijn de meetwaarden van steekproef 2. Het is belangrijk dat beide steekproeven ongeveer evenveel meetpunten hebben.

TWO-SAMPLE T-TOETS IN EXCEL

The image shows an Excel spreadsheet with the formula `=TTEST()` entered in cell A1, highlighted with a red circle. A 'Function Arguments' dialog box is open, showing the following fields:

- Array1**: = array
- Array2**: = array
- Tails**: = number
- Type**: = number

Below the fields, the text reads: "This function is available for compatibility with Excel 2007 and earlier. Returns the probability associated with a Student's t-Test. **Array1** is the first data set."

A callout box on the right side of the dialog box contains the text: "Tails geeft aan of de toets eenzijdig (1) of tweezijdig (2) is."

At the bottom of the dialog box, there is a "Formula result =" field, a "Help on this function" link, and "OK" and "Cancel" buttons.

TWO-SAMPLE T-TOETS IN EXCEL

The image shows an Excel spreadsheet with the formula `=TTEST()` entered in cell A1, circled in red. A 'Function Arguments' dialog box is open, showing the following fields:

- Array1**: = array
- Array2**: = array
- Tails**: = num
- Type**: =

Below the fields, the text reads: "This function is available for compatibility with Excel 2007 and earlier. Returns the probability associated with a Student's t-Test." and "Array1 is the first data set." At the bottom, there is a "Formula result =" field, a "Help on this function" link, and "OK" and "Cancel" buttons.

Type: gepaard (1), ongepaard met gelijke SD (2), ongepaard met ongelijke SD (3). Zie uitleg hierna.

GEPAAARD VERSUS ONGEPAAARD TOETSEN

- In het geval van **ongepaard toetsen** is er sprake van **twee groepen** die worden vergeleken.

Voorbeeld

We willen onderzoeken of er een verschil in staartlengte is tussen honden (groep 1) en katten (groep 2).

- In het geval van **gepaard toetsen** is er sprake van **één groep en twee metingen**.
NB: het voordeel van gepaard toetsen is dat je minder last hebt van biologische variatie.

Voorbeeld

We willen onderzoeken eten zorgt voor een hogere bloedsuikerspiegel. We meten hiervoor in één groep mensen de bloedsuikerspiegel zowel vóór het eten (meting 1) als na het eten (meting 2).

ONGEPAARD TOETSEN: GELIJKE OF ONGELIJKE SD?

- Bij een ongepaarde toets zijn er twee groepen. Die twee groepen kunnen (ongeveer) dezelfde standaarddeviatie of een verschillende standaarddeviatie hebben.
- Om te bepalen of er sprake is van een gelijke of ongelijke standaarddeviatie, moet je eerst voor elke groep de standaarddeviatie uitrekenen. Daarna gebruik je de volgende vuistregel:

$$\frac{S_{grootste}}{S_{kleinste}} < 3 \rightarrow \text{Gelijke SD}$$

$$\frac{S_{grootste}}{S_{kleinste}} > 3 \rightarrow \text{Ongelijke SD}$$

VRAAG 1



Conditie 1: ex-coronapatiënten die 1 week vrij zijn van klachten

Conditie 2: ex-coronapatiënten die 3 maanden vrij zijn van klachten

We willen weten of er een verschil is in antilichaamtiter tussen deze twee condities.
Hoe kunnen we een gepaard experiment opzetten?

- a) We meten voor elke ex-coronapatiënt de antilichaamtiter na 1 week en na 3 maanden.
- b) We meten voor een aantal ex-coronapatiënt de antilichaamtiter na 1 week en we meten voor andere ex-coronapatiënten de antilichaamtiter na 3 maanden.
- c) Je kunt hier geen gepaard experiment voor opzetten.

VRAAG 2



We willen weten of de staartlengtes van honden en katten verschillen. We meten daarvoor de staartlengtes van 30 honden en van 29 katten. We vinden $s_{kat} = 5 \text{ cm}$ en $s_{hond} = 9 \text{ cm}$. De data is ongeveer normaal verdeeld.

Welke argumenten gebruik je in Excel voor de t-toets functie?

- a) two-tailed, gepaard.
- b) two-tailed, ongepaard met gelijke SD.
- c) two-tailed, ongepaard met ongelijke SD.
- d) Je mag hier geen t-toets doen, want het aantal waarnemingen per groep is te verschillend.

VRAAG 3

Je wilt onderzoeken of er een verschil is in reistijd naar de Hogeschool tussen Life Sciences studenten en Chemie studenten. Je gaat dit statistisch toetsen met een two-sample, ongepaarde t-toets. De t-toets levert een p-waarde op van 0,008. Er geldt $\alpha = 0,05$.

Is er sprake van een statistisch significant verschil in reistijd tussen Life Sciences en Chemie studenten?

- a) nee, want p-waarde is kleiner dan 0,05.
- b) nee, want p-waarde is groter dan 0,05.
- c) ja, want de p-waarde is kleiner dan 0,05.
- d) ja, want de p-waarde is groter dan 0,05.



VRAAG 4

Je wilt onderzoeken of er een verschil is in reistijd naar de Hogeschool tussen Life Sciences studenten en Chemie studenten. Je gaat dit statistisch toetsen met een t-toets. De t-toets levert een p-waarde op van 0,008. Er geldt $\alpha = 0,05$. Welke conclusie kan je trekken?

- a) Er is een statistisch significant verschil in reistijd tussen Life Sciences en Chemie studenten. Life Sciences studenten hebben een langere reistijd dan Chemie studenten.
- b) Er is een statistisch significant verschil in reistijd tussen Life Sciences en Chemie studenten. Op basis van deze gegevens kan je geen uitspraak doen welke studenten de langste reistijd hebben.



“IK HEB EEN T-TOETS GEDAAN” 😞

- eenzijdig of tweezijdig? (tails, ligt aan je vraag)
- one sample t-toets of two sample t-toets? (1 of 2 steekproeven)
- als two-sample t-toets:
 - Gepaard of ongepaard?
 - Gelijke of ongelijke sd?

“Ik heb een eenzijdige, gepaarde two-sample t-toets gedaan met ongelijke varianties.”

AAN DE SLAG MET EXCEL!

In het werkcollege leer je:

- wanneer je een gepaarde toets en wanneer je een ongepaarde toets gebruikt;
- en hoe je een two-sample t-toets uitvoert in Excel.

HEEL VEEL PLEZIER!

Ga aan de slag met de opgaves in het werkcollege.

Stel vragen aan elkaar of aan de docent.