Research Methods

Probability and Null Hypothesis Testing (incl. *z*-Test and *t*-Test)

Laurens Rook (Delft University of Technology)

Delft University of Technology

Previous lectures

- Descriptive statistics
- Experiments: building blocks & research designs
- Today: Probability, null hypothesis significance testing (Ch. 4) and inferential statistics (Ch. 5-6)
  - illustrated for *z*-test & various *t*-tests

Learning goals

- Understanding probability and its relation with the normal distribution
- Differentiate null and alternative hypothesis
- Understanding the relation between Type 1 and Type 2 errors and hypothesis testing
- Capable of explaining what statistical significance means (for *z*-test & various *t*-test)

Probability

1.

Basic concepts

- Probability = the number of ways a particular outcome (event) can occur divided by the total number of outcomes (events)

- Proportion = to express probabilities as varying between 0.0 (certainly no occurrence) and 1.0 (certain occurrence of event)

- Uncertainty = the range in event occurrence likelihood

Examples

Flipping a coin once --> probability to get a "head"

Number of ways to get "head" / number of possible outcomes = ½ = .50

Rolling "2" with a die once --> probability to get a "2"

Number of ways to get "2" / number of possible outcomes = 1/6 = .167

- After many trials, we can accurately predict what proportion of an event ("head", "2") will take place

TU Delft

Two probability rules

- Multiplication rule (the “AND” rule)
  - the probability of a series of outcomes occurring on successive trials is the product of their individual probabilities, when the sequence of outcomes is independent
  - E.g., getting “heads” twice in two tosses:

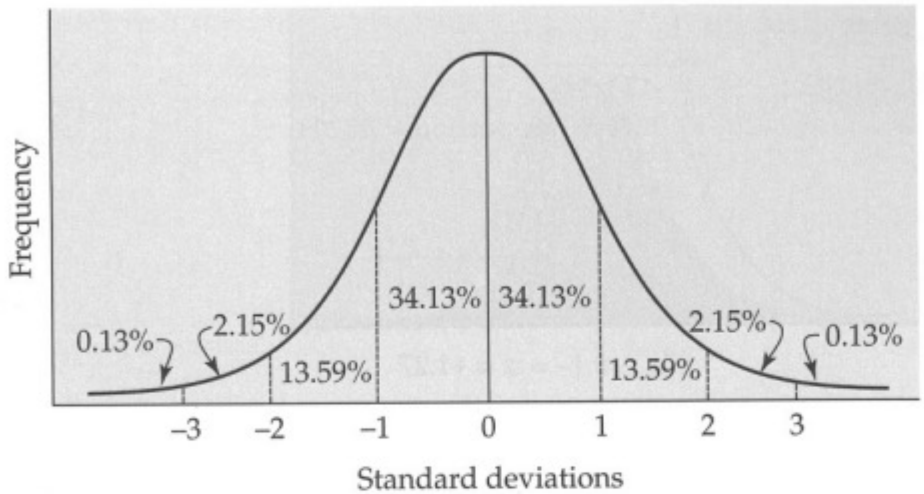
$0.5 * 0.5 = 0.25$

Two probability rules

- Addition rule (the “OR” rule)
  - the probability of one outcome or another outcome occurring on a particular trial is the sum of their individual probabilities, when the outcomes are mutually exclusive
  - E.g., drawing either “clubs” or “hearts” from a deck of cards:

$$0.25 + 0.25 = 0.5$$

The link between probability and the standard normal curve



- We can use the areas under the standard curve to determine the probability that an observation *falls within a certain area under the curve*

How? With *z*-scores

$$Z = \frac{x - \mu}{\sigma}$$

- Where: x = a person's score
  μ = the population mean
  σ = the population standard deviation

- We convert a person's test score into a *z*-score using the formula, and reasd the associated proportion (the area under the curve) from a table (book, Table A1, Appendix A)

TU Delft

Example 1

- We collected intelligence test scores that are normally distributed with a mean = 100 and SD = 15. What is the probability that we select from the general population a person with a test score of 119 or higher?

$$Z = \frac{X-\mu}{\sigma} = \frac{119-100}{15} = \frac{19}{15} = 1.27$$

--> Table A1 from Appendix A: $p\,(X \geq 119) = .10203$ (10.2%)

TU Delft

Example 2

- We collected intelligence test scores that are normally distributed with a mean = 100 and SD = 15. What is the probability that we select from the general population a person with a test score of 70 or lower?

$$Z = \frac{X-\mu}{\sigma} = \frac{70-100}{15} = \frac{-30}{15} = -2.0$$

--> Table A1 from Appendix A: $p\,(X \leq 70) = .02275$ (2.3%)

TU Delft

Example 3 (multiplication)

- We collected intelligence test scores that are normally distributed with a mean = 100 and SD = 15. What is the probability that we select from the general population a person with a test score of 80 or lower AND a person with a test score of 125 or higher?

$Z_1 = \frac{X-\mu}{\sigma} = \frac{80-100}{15} = \frac{-20}{15} = -1.33$ --> $p(X \leq 80) = .09175$

$Z_2 = \frac{X-\mu}{\sigma} = \frac{125-100}{15} = \frac{25}{15} = 1.67$ --> $p(X \geq 125) = .04745$

______________ *

.00435 (0.4%)

Why relevant?

- We can use the individual score --> z-score conversion principle for hypothesis testing research when:
  - We are interested in a sample mean (z-test)
  - We have set critical values (areas under the normal curve that we consider regions of rejection)
  - We have formulated a null hypothesis and an alternative hypothesis

Hypothesis testing

2.

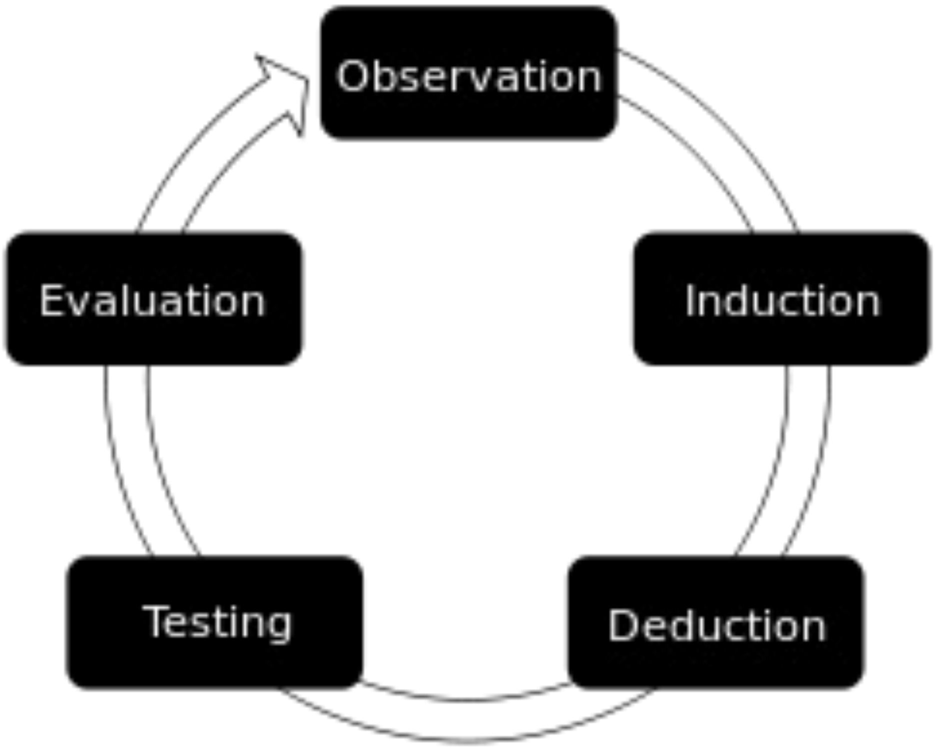
Null hypothesis significance testing

- The aim of research usually is to provide an answer to a research question, which is formulated as a statement that can be accepted / rejected
- The process of discovering if the statement is supported by your data (or not) -->Null hypothesis significance testing (NHST)

Steps in hypothesis development

1. State the null (H_0) and alternative (H_a) hypotheses
2. Determine your significance level (p < .05 or else)
3. Choose the appropriate statistical test based on the type of scales used (nominal, ordinal, interval, ratio)
4. Check your statistical output to see if your null hypothesis is accepted / rejected, or if – instead – the alternative hypothesis is accepted / rejected

How science works (the empirical cycle)



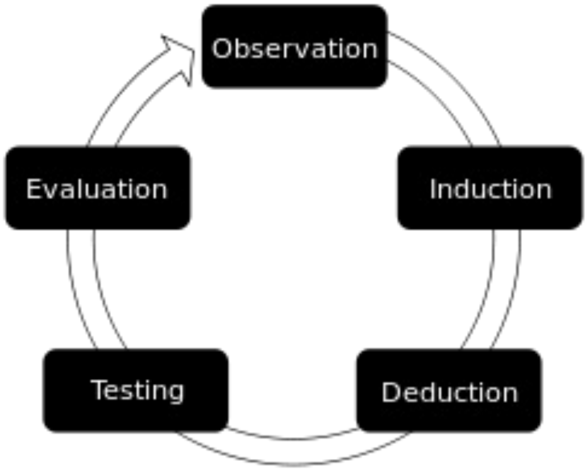
A research dilemma

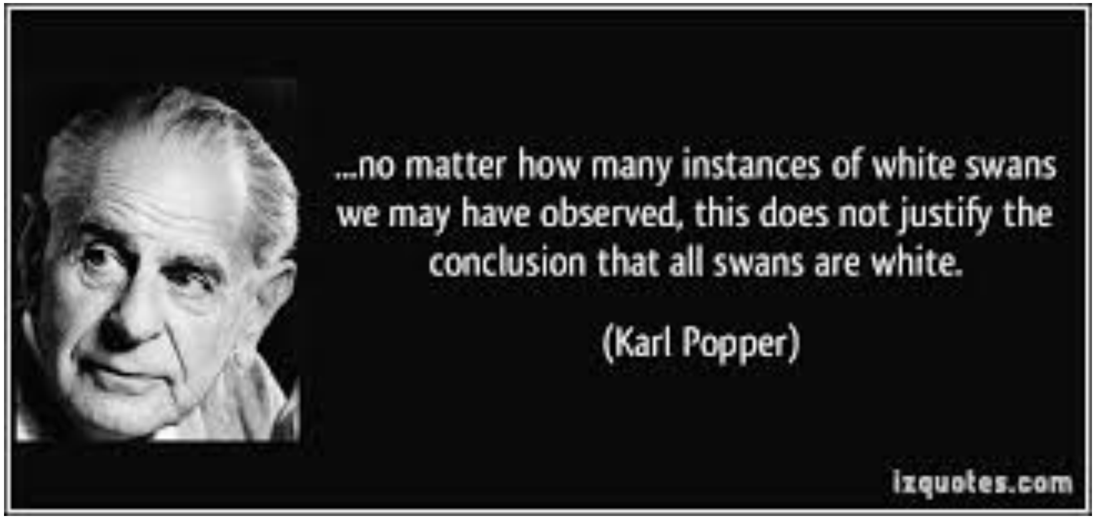
- You cannot statistically demonstrate the **“truth”** of a research statement

- Statistics are better capable of showing that something is not true

  - --> they are designed for **“falsification”** purposes

Karl Popper





TU Delft

The way out of this...

- To propose exactly the opposite of what you wish to demonstrate to be true in a null hypothesis
- Then falsify that opposite statement (or null hypothesis)
- And, what is left (your initial or alternative hypothesis) must then be true

Null hypothesis significance testing (approach)

- We start with two competing hypotheses:
- H_0 (the null hypothesis): the effect does <u>not</u> exist
- H_a (the alternative hypothesis): an effect (a difference between two groups) exists, and is significant
    - this H_a is your research hypothesis, the statement you wish to support --> *but you test it via the null hypothesis!*

H_0 and H_a: two forms

- **One-tailed**: Your H_a is formulated in terms of “higher / lower” (thus directional); same follows for the H_0
- **Two-tailed**: Your H_a is formulated in terms of “differences that exist between groups” (nondirectional); the H_0 then states that no differences exist

Probability and statistical significance

- **Statistical significance** = an observed difference between two descriptive statistics (such as the means), which is unlikely to have occurred by chance

- **Probability value** = in social and management science, researchers usually work with a *p*-value of .05

  - --> they take a 5% risk of making a Type 1 error

NHST is prone to Type-1 and Type-2 errors

TABLE 8.1 The four possible outcomes in statistical decision making

	THE TRUTH (UNKNOWN TO THE RESEARCHER)	
THE RESEARCHER'S DECISION	H_0 is true	H_0 is false
Reject H_0 (say it is false)	Type I error	Correct decision
Fail to reject H_0 (say it is true)	Correct decision	Type II error

Example

Toward Linguistic Recognition of Generalized Anxiety Disorder

Laurens Rook, Maria Chiara Mazza, Iulia Lefter and Frances Brazier*

Faculty of Technology, Policy and Management, Delft University of Technology, Delft, Netherlands

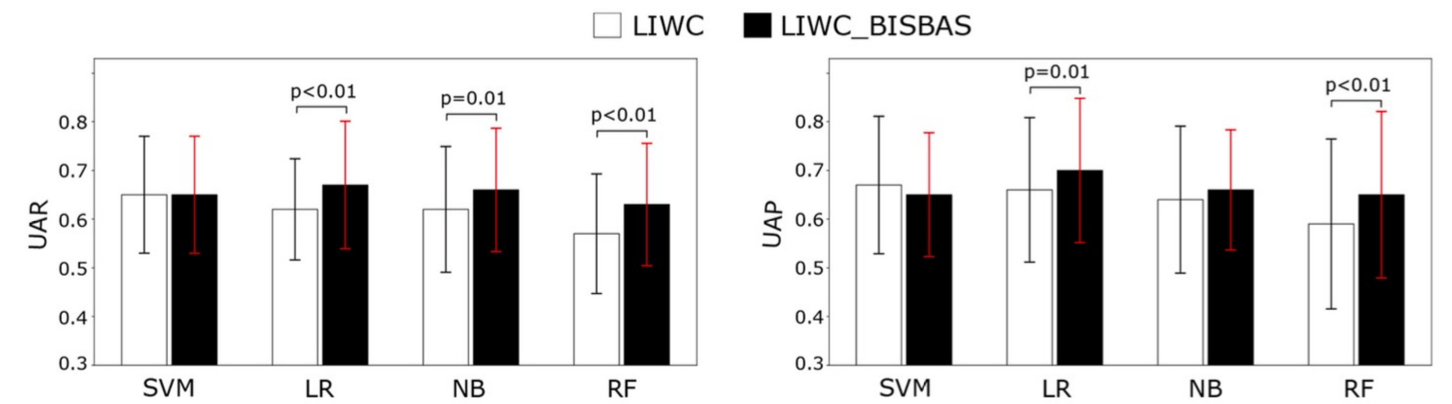


FIGURE 1 | Unweighted average recall (UAR) and unweighted average precision (UAP) for the four classifiers given Language Inventory Word Count (LIWC) features, and for a concatenation of LIWC features, Behavioral Inhibition System and Behavioral Approach System personality features (LIWC_BISBAS), respectively.

GAD ≥ 10

Example

Toward Linguistic Recognition of Generalized Anxiety Disorder

Laurens Rook, Maria Chiara Mazza, Iulia Lefter and Frances Brazier*

Faculty of Technology, Policy and Management, Delft University of Technology, Delft, Netherlands

- False positives --> when text was emotionally negative for a particular episode, but positive on the overall study journey (Type I)

- False negatives --> idem when text had been very short (2-3 condensed sentences; Type II)

In sum: Essence of null hypothesis significance testing

- H_0 (the null hypothesis): the effect does not exist
- H_a (the alternative hypothesis): an effect exists

- We conduct a statistic test that represents H_0. We calculate the probability that we get a value big enough to accept $H_{0.}$ Thus we check the p-value of H_0

- If too small ($p < 0.05$), we reject the idea of H_0 that we have no effect, and accept our H_a instead!

TU Delft

Inferential statistics (*z*-test)

3.

Inferential statistics

- NHST allows us to select a sample, compare it with the population at large, and analyze data collected
- **Inferential statistics** = procedures for drawing conclusions about a (wider) population, based on data collected from a (smaller) sample
- **Parametric tests** --> a (*z* or *t*) test that involves making assumptions about estimates of population characteristics (mean, sd)

Today (Ch. 5-6)

- The single-sample *z*-test
- The single-sample *t*-test
- The *t*-test for related groups
- The *t*-test for independent groups

The (single sample) *z*-test

- **A (single sample) *z*-test** = a parametric inferential statistical test of the null hypothesis for a single sample, *where the population variance is known*

The *z*-score vs. *z*-test

- **A *z*-score** = a single data point (such as a single participant's score) --> you could compare this score to the mean score of the wider population

- **A *z*-test** = a sample mean (from all participant scores in your study) --> you compare this sample mean to the population mean

The *z*-score vs. *z*-test

z-score (single data point):

$Z = \frac{X-\mu}{\sigma}$

z-test (sample mean):

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$

$\sigma_{\bar{x}}$ = standard deviation of the sampling distribution (SE of the mean)
σ = standard deviation
N = distribution of sample means for sample size *N*

$Z = \frac{\bar{X}-\mu}{\sigma_{\bar{x}}}$

$\bar{X}$ = sample mean

μ = mean of sampling distribution

(single sample) *z*-test: when?

- **A single-group design** = A research study in which there is only one group of participants
  - Example: the one-shot case study design from the lectures on (quasi) experiments
  - when the population variance is known

Posttest only (one-shot case study) design*

	[Time]	
	Treatment	Test score
Experimental group	X	O

Treatment effect = O

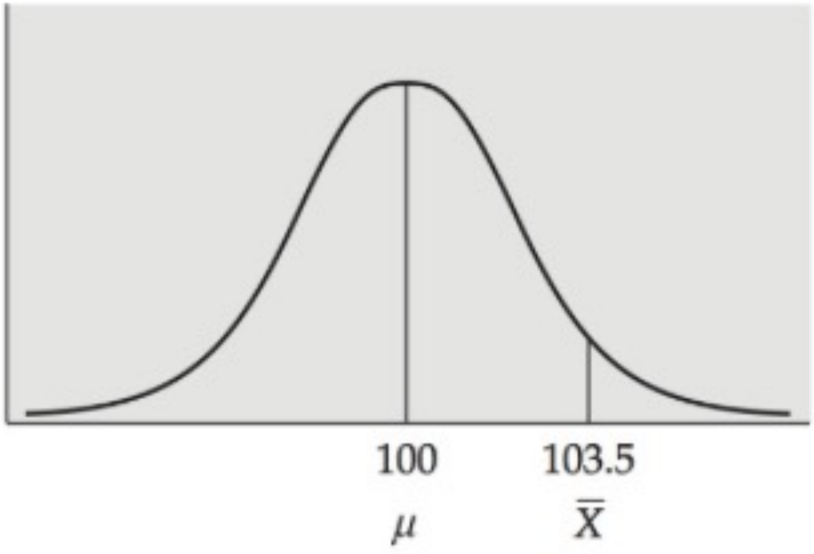
*Quasi

Form of a *z*-test (one-tailed)

- Given are:
    - a normal distribution on scores
    - a population mean
    - a sample mean
    - one-tailed test --> alpha level of .05 or less
- Question to answer: Is the sample mean statistically bigger (or smaller) than the population mean?

Example

FIGURE 9.1
The obtained mean in relation to the population mean



Population mean Sample mean

Interpretation of a *z*-test (one-tailed)

- We convert the z-score to the associated proportion (Table A1, Appendix A) to get the ***z* value obtained**

- We compare the ***z* critical value** (that marks the edge of the region of rejection in a sampling distribution) with the ***z* value obtained**

- **Critical value =** the edge of the region of rejection in a sampling distribution. Values equal or beyond it fall in the region of rejection for the H_0

Interpretation: Example

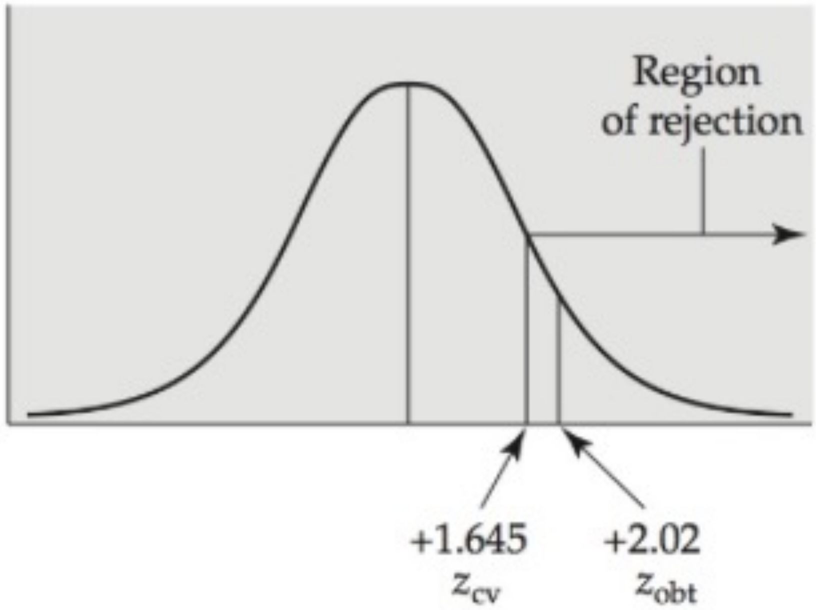


FIGURE 9.2
The *z* critical value and the *z* obtained for the *z* test example

The ***z* value obtained** falls in the region of rejection for the H_0.

Interpretation: the sample mean was statistically different from the population mean at *p* < .05 (one-tailed)

Form of a *z*-test (one-tailed)

- Given are:
  - a normal distribution
  - a population mean
  - a sample mean
  - one-tailed test --> alpha level of .05 or less
- Question to answer: Is the sample mean statistically bigger (or smaller) than the population mean?

TU Delft

Interpretation of a *z*-test (two-tailed)

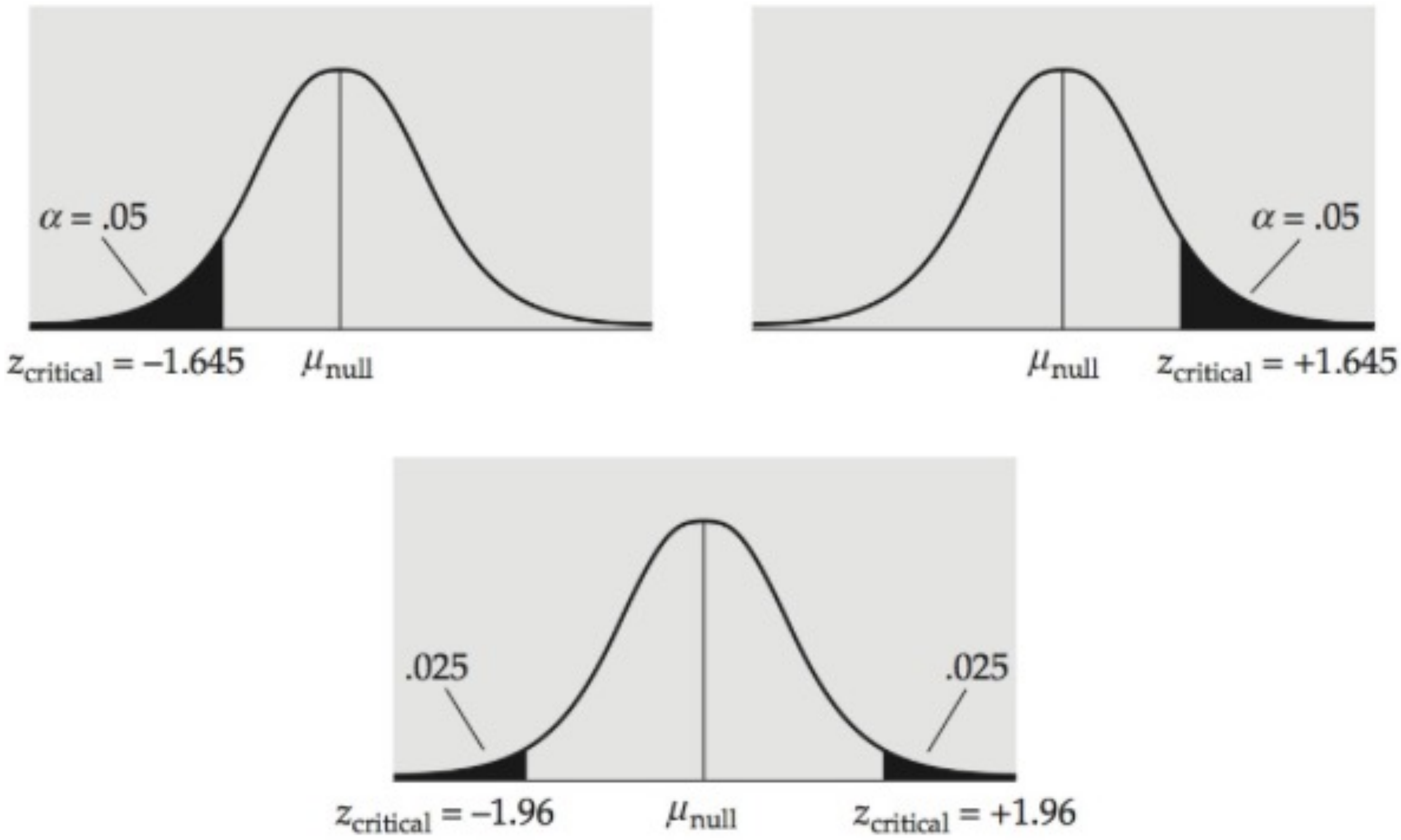
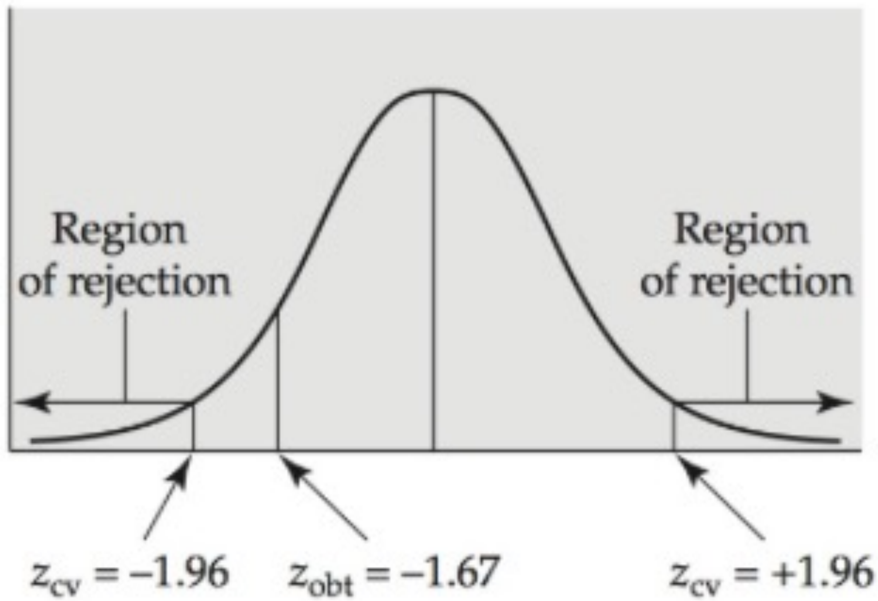


FIGURE 9.3 Regions of rejection and critical values for one-tailed versus two-tailed tests

Interpretation: Example

FIGURE 9.4
The *z* critical value and the *z* obtained for the two-tailed *z* test example



The ***z* value obtained** does NOT fall in the region of rejection for the H_0.
Interpretation: the sample mean was NOT statistically different from the population mean at *p* < .05 (two-tailed)

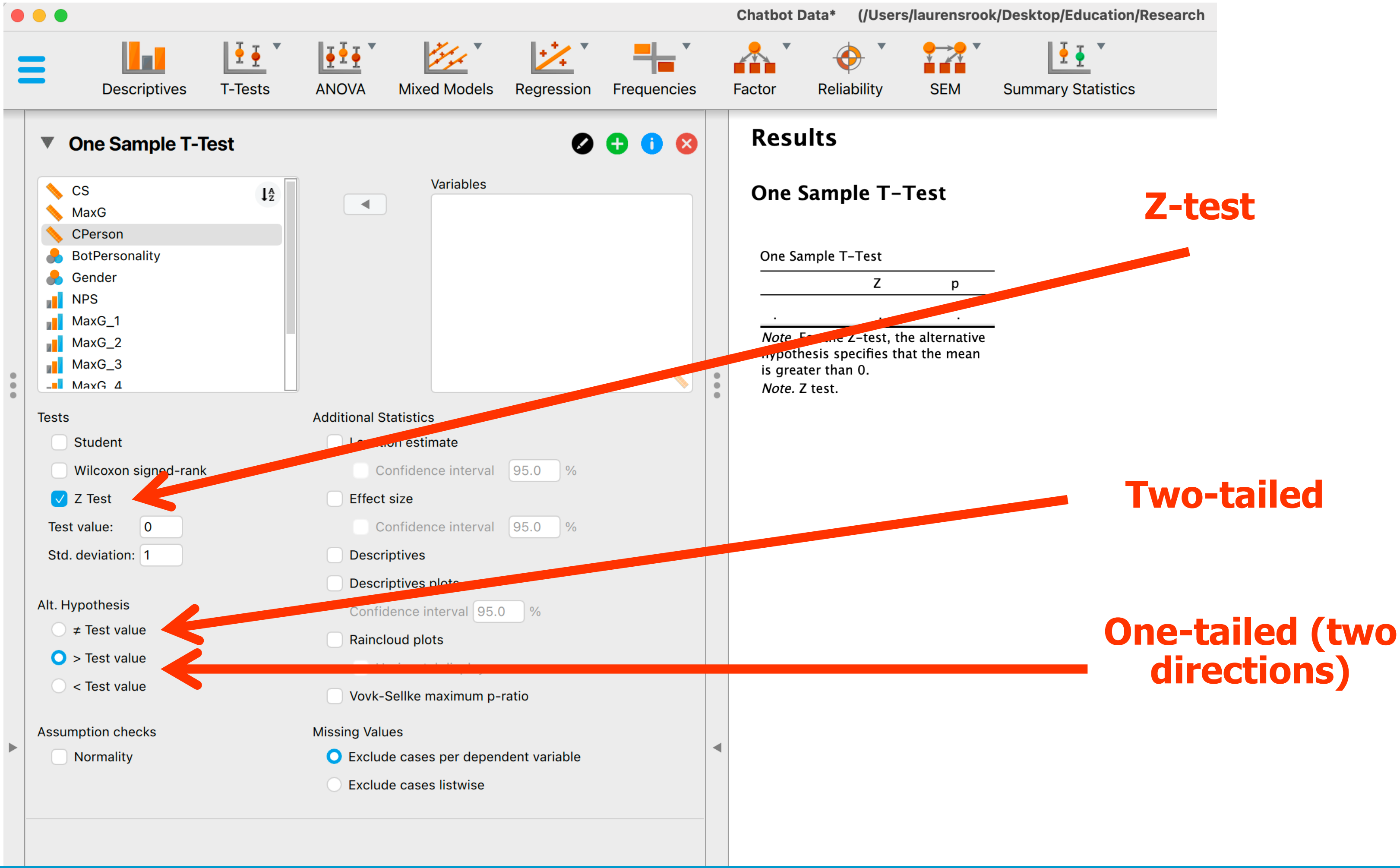
Statistical power & assumptions

- **Statistical power** = the probability that you correctly reject a false H_0

- This is higher with a one-tailed test, given that the *z* critical value does not need to be so large to get significantly different from the population mean

    - A one-tailed test is more statistically powerful than a two-sided test (it increases the chance to find a $p < .05$, and to correctly reject H_0)
    - Another way to achieve this is to increase your sample size (which you usually do for a two-sided test)

In JASP



TU Delft

Inferential statistics (*t*-test)

4.

(a) The (single sample) *t*-test

- **A (single sample) *t*-test** = a parametric inferential statistical test of the null hypothesis for a single sample, where the population variance is **NOT** known

- Unlike the z-test, the (Student's) *t*-test works with *t* distributions that are NOT normally distributed (but have a bell-shaped, symmetrical form)

- Notation includes **degrees of freedom (*df*)** = *N* - 1

The number of scores that are free to vary (sample size - 1)

Example

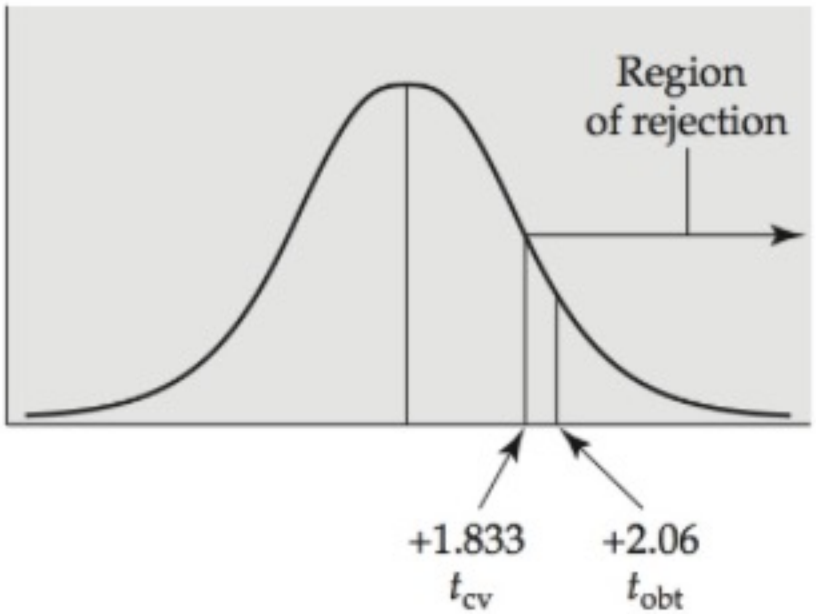


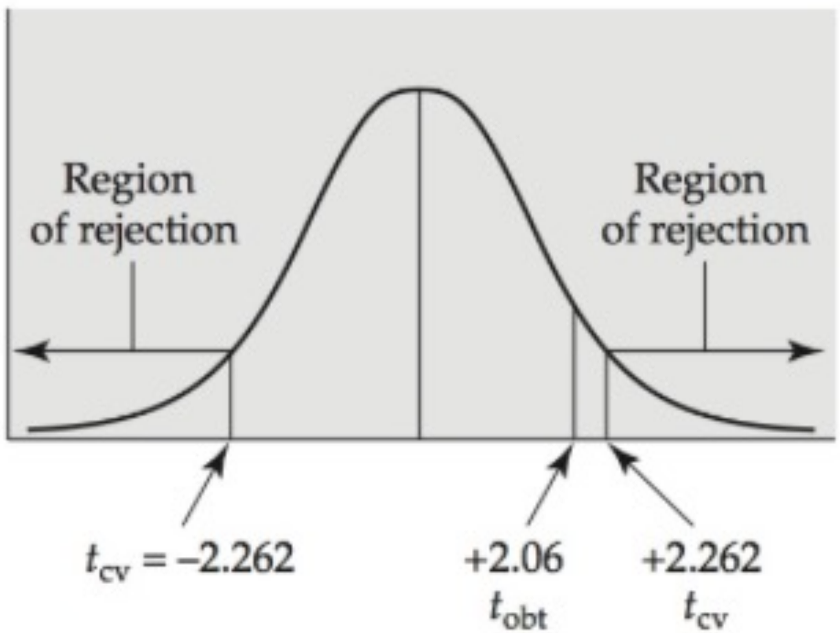
FIGURE 10.1
The *t* critical value and the *t* obtained for the single-sample one-tailed *t* test example

- $t(9) = 2.06$, $p < .05$ (one-tailed)

TU Delft

Example (two-tailed)

FIGURE 10.2
The t critical value and the t obtained for the single-sample two-tailed test example



- $t(9) = 2.06$, *ns.* (two-tailed)

Statistical power & assumptions

- The (single sample) t-test should be used only if:
    - the data are interval / ratio in scale
    - the population distribution of scores is symmetrical
- If those assumptions aren't met --> nonparametric tests should be used

One sample *z*-test and *t*-test in JASP

One Sample T–Test

One Sample T–Test ▼

	Test	Statistic	df	p
CPerson	Student	37.51	119	< .001
	Z	36.83		< .001

(b) Testing hypotheses about two related means

- Paired samples *t*-test = to examine the differences in the same group before and after a treatment
  - H_0 = there is no difference between the pretest and posttest
  - H_a= there exists a difference between pretest and posttest

(One group) pretest-posttest design*

[Time]

	Pretest score	Treatment	Posttest score
Experimental group	O_1	X	O_2

Treatment effect = $(O_2 - O_1)$

***Quasi**

TU Delft

Paired samples *t*-test in JASP

Paired Samples T-Test

Paired Samples T-Test

Measure 1		Measure 2	t	df	p
Howdoyoufeelrightnow_A	-	Howdoyoufeelrightnow	−0.211	141	0.833

Note. Student's t-test.

(c) Testing hypotheses about *unrelated* means

- Independent samples *t*-test = when we are interested whether two groups are different from each other on a particular interval / ratio-scaled factor

- This applies to experimental (treatment) vs. control group designs

Posttest with control group design*

[Time]

	Treatment	Test score
Experimental group	X	O_1
Control group		O_2

Treatment effect = $(O_1 - O_2)$

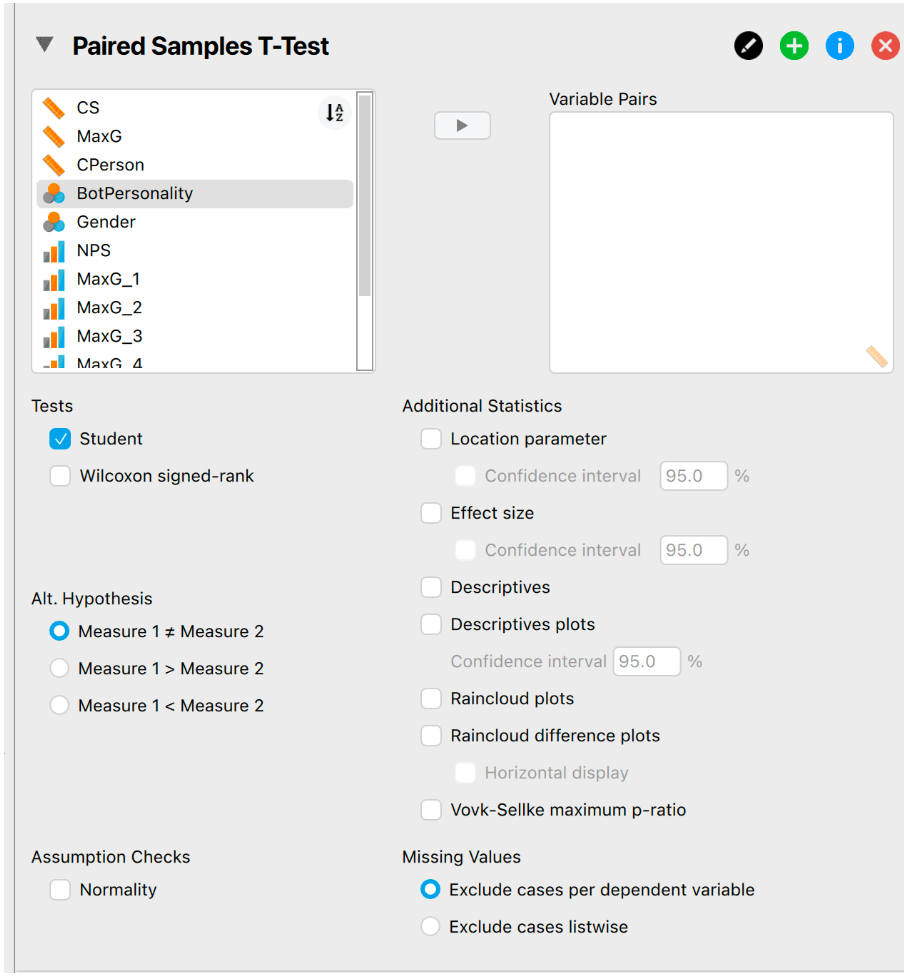
***Quasi**

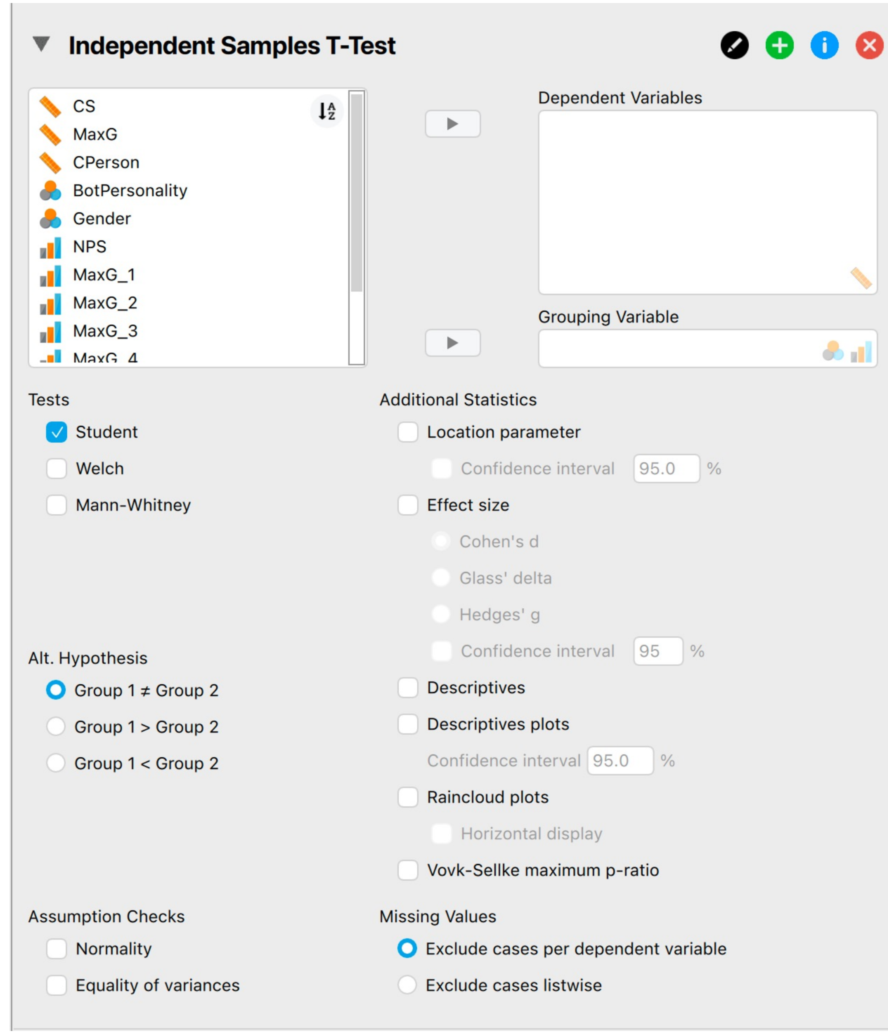
T-tests are about comparing means between two independent groups

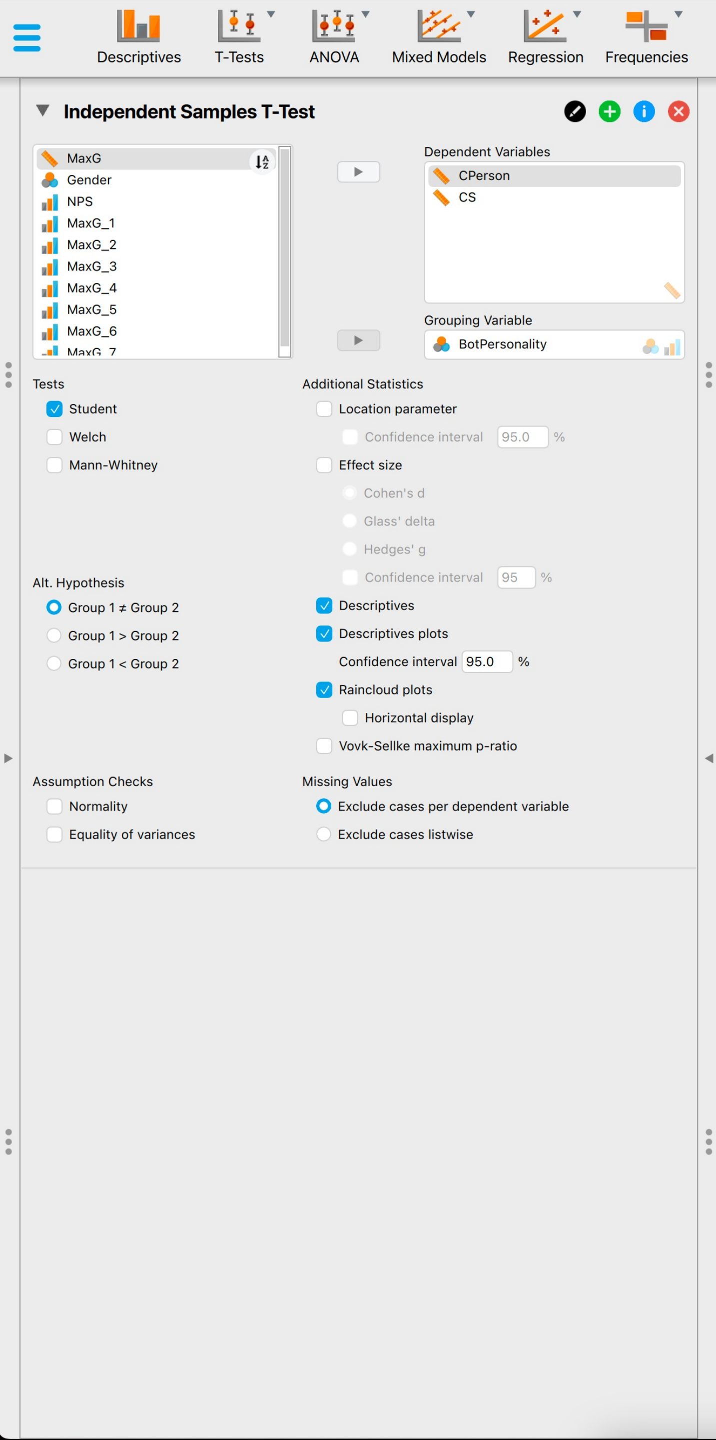
3	4
2	4
3	5
1	4
2	5

CREATIVITY

TU Delft


Paired Samples T-Test
CS
MaxG
CPerson
BotPersonality
Gender
NPS
MaxG_1
MaxG_2
MaxG_3
Variable Pairs
Tests
Student
Wilcoxon signed-rank
Alt. Hypothesis
Measure 1 ≠ Measure 2
Measure 1 > Measure 2
Measure 1 < Measure 2
Assumption Checks
Normality
Additional Statistics
Location parameter
Confidence interval 95.0 %
Effect size
Confidence interval 95.0 %
Descriptives
Descriptives plots
Confidence interval 95.0 %
Raincloud plots
Raincloud difference plots
Horizontal display
Vovk-Sellke maximum p-ratio
Missing Values
Exclude cases per dependent variable
Exclude cases listwise


Independent Samples T-Test
CS
MaxG
CPerson
BotPersonality
Gender
NPS
MaxG_1
MaxG_2
MaxG_3
Dependent Variables
Grouping Variable
Tests
Student
Welch
Mann-Whitney
Alt. Hypothesis
Group 1 ≠ Group 2
Group 1 > Group 2
Group 1 < Group 2
Assumption Checks
Normality
Equality of variances
Additional Statistics
Location parameter
Confidence interval 95.0 %
Effect size
Cohen's d
Glass' delta
Hedges' g
Confidence interval 95 %
Descriptives
Descriptives plots
Confidence interval 95.0 %
Raincloud plots
Horizontal display
Vovk-Sellke maximum p-ratio
Missing Values
Exclude cases per dependent variable
Exclude cases listwise



Results

Independent Samples T-Test

Independent Samples T-Test

	t	df	p
CPerson	−11.219	118	< .001
CS	−0.899	118	0.371

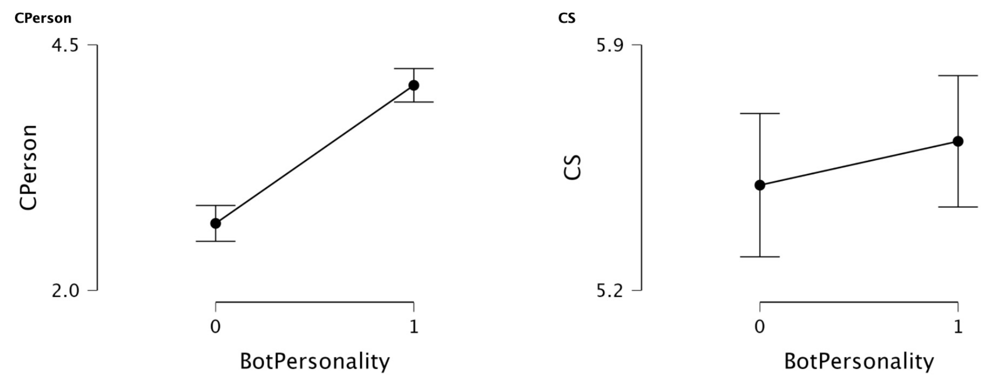
Note. Student's t-test.

Descriptives

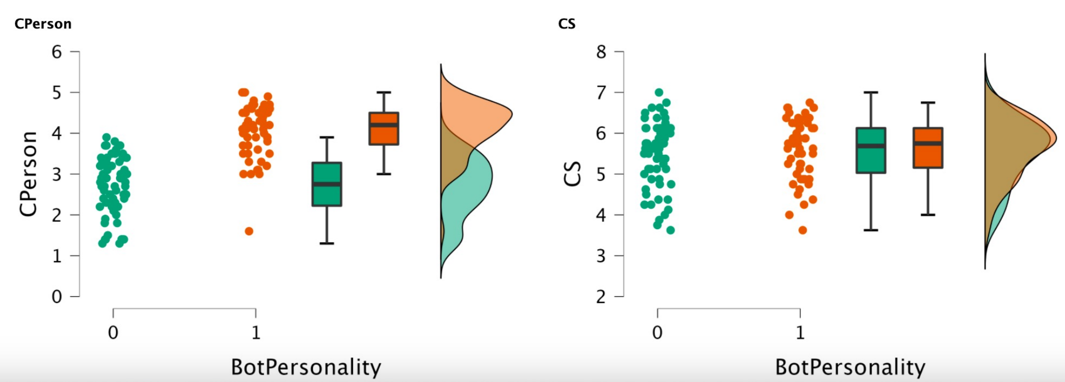
Group Descriptives

	Group	N	Mean	SD	SE
CPerson	0	62	2.682	0.719	0.091
	1	58	4.088	0.648	0.085
CS	0	62	5.500	0.804	0.102
	1	58	5.625	0.712	0.094

Descriptives Plots



Raincloud Plots



JASP output presentation in exam

Independent Samples T–Test

Independent Samples T–Test

	t	df	p
CPerson	−11.219	118	< .001
CS	−0.899	118	0.371

Note. Student's t–test.

In sum

Learning goals (Ch. 4)

- Understanding probability and its relation with the normal distribution
- Differentiate null and alternative hypothesis
- Understanding the relation between Type 1 and Type 2 errors and hypothesis testing
- Capable of explaining what statistical significance means (for *z*-test & various *t*-test)

Learning goals (Ch. 5 - 6)

- Explain what a z-score / z-test / t-test is and how it is computed
- Explain what statistical power is and how to make statistical tests more powerful
- List the assumptions of the z-test / t-test
- Capable of interpreting a z-test / t-test result (one-tailed and two-tailed)