

UNIVERSITY OF BERGEN

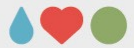


# Refreshing uni- / bivariate statistics: Stat. analyses

Sebastian Jentschke

UNIVERSITY OF BERGEN





# Overview

## Categorical predictor and outcome variables (Ch. 10)

- $\chi^2$ -test of association: step-by-step and in jamovi
- $\chi^2$  goodness-of-fit in jamovi
- McNemar test for Paired Samples in jamovi

## Categorical predictor and continuous outcome variables (Ch. 11)

- z-test and One Sample t-test: step-by-step and in jamovi
- Independent Samples t-test
- Paired Samples t-test
- Assumptions and non-parametric alternatives

## Continuous predictor and continuous outcome variables (Ch. 12)

- Correlations: step-by step and in jamovi
- Where correlations fail: Anscombe's quartet



During this lecture, we are dealing with three large categories of analyses. The first one, which are analyses where we have a categorical predictor and a categorical outcome, may not have gotten very much coverage in your B.Sc. statistics course. This is one reason, why they get relatively strong weight in our crash course.

Another reason for giving them much weight is that they are mathematically relatively simple. That is, they are suited to introduce concepts behind statistics that might be more difficult to teach for the more complex methods (even though I still try do this to some degree for t-test and correlation).



# Overview

## Categorical predictor and outcome variables (Ch. 10)

- $\chi^2$ -test of association: step-by-step and in jamovi
- $\chi^2$  goodness-of-fit in jamovi
- McNemar test for Paired Samples in jamovi

## Categorical predictor and continuous outcome variables (Ch. 11)

- z-test and One Sample t-test: step-by-step and in jamovi
- Independent Samples t-test
- Paired Samples t-test
- Assumptions and non-parametric alternatives

## Continuous predictor and continuous outcome variables (Ch. 12)

- Correlations: step-by step and in jamovi
- Where correlations fail: Anscombe's quartet



The second class of analyses is possibly the one that most of you used in your B.Sc. thesis. Here, we have methods where we have a categorical predictor and a continuous outcome variable (e.g., to explore whether certain test scores differ with gender or the experimental group the participant was in). We cover three different kinds of t-tests and the z-test as an introduction. I also briefly mention non-parametric alternatives to these tests that you use if assumptions to conduct a t-test are not met (e.g., if your data are not normally distributed).



# Overview

## Categorical predictor and outcome variables (Ch. 10)

- $\chi^2$ -test of association: step-by-step and in jamovi
- $\chi^2$  goodness-of-fit in jamovi
- McNemar test for Paired Samples in jamovi

## Categorical predictor and continuous outcome variables (Ch. 11)

- z-test and One Sample t-test: step-by-step and in jamovi
- Independent Samples t-test
- Paired Samples t-test
- Assumptions and non-parametric alternatives

## Continuous predictor and continuous outcome variables (Ch. 12)

- Correlations: step-by step and in jamovi
- Where correlations fail: Anscombe's quartet



Finally, the third part is about exploring the relations between continuous predictors and continuous outcomes.

The word “relation” already indicates that using the terms predictor and outcome are a bit misleading. With correlations, we assess the relation between two variables. We are not capable to say which variable 1 “caused” a change in variable 2, whether it was the other way around or whether there was maybe a third outside variable that “caused” the changes in variable 1 and 2.

I will also speak about a special example of variable pairs – Anscombe's quartet – that are correlated even though this is mainly based on that assumptions for running correlation analyses are violated.



# Categorical predictor and outcome variables (Ch. 10)

Let's turn to refreshing some basic statistical procedures (or to briefly introduce them if you did not have them before). And first, we will turn to statistical procedures that are used to explore relationships between a categorical predictor (= independent) variable and a categorical outcome (= dependent) variable.



# Introduction

when exploring frequency data, we often are interested whether:

- whether the frequencies in one variable are influenced by another (e.g., do more female finish university) →  $\chi^2$ -test of association
- several categories occur equally often (e.g., all numbers on a dice) →  $\chi^2$ -goodness of fit
- whether a change in frequency occurs (e.g., in health vs. sick as consequence of a treatment) → McNemar test



The class of statistical procedures that I am going to introduce now are used to explore relationships between a categorical independent (= predictor) variable and a categorical dependent (= outcome) variable.

Most statistical tests exploring such relationships are based upon a  $\chi^2$ -distribution (chi-squared). Basically, all tests based upon this distribution compare whether an empirical distribution of frequencies or occurrences is (significantly) different from a theoretical expected distribution.



## Introduction

when exploring frequency data, we often are interested whether:

- whether the frequencies in one variable are influenced by another (e.g., do more female finish university) →  $\chi^2$ -test of association
- several categories occur equally often (e.g., all numbers on a dice) →  $\chi^2$ -goodness of fit
- whether a change in frequency occurs (e.g., in health vs. sick as consequence of a treatment) → McNemar test



When exploring such frequency data, we often are interested whether: (1) whether the frequencies in one variable are influenced by another (e.g., do more females finish university); (2) several categories occur equally often (e.g., all numbers on a dice); or (3) whether a change in frequency occurs (e.g., a change in health vs. sick as consequence of a treatment over time, considering whether you were already healthy or ill).

We use the  $\chi^2$ -test of association for questions of type (1), the  $\chi^2$ -goodness of fit to explore questions belonging to (2), and the McNemar test for questions of type (3).



## Introduction

when exploring frequency data, we often are interested whether:

- whether the frequencies in one variable are influenced by another (e.g., do more female finish university) →  $\chi^2$ -test of association
- several categories occur equally often (e.g., all numbers on a dice) →  $\chi^2$ -goodness of fit
- whether a change in frequency occurs (e.g., in health vs. sick as consequence of a treatment) → McNemar test



To make it easier to understand which test to use lets see with which type of data we are dealing for each test: For (1), we have two columns with combinations of data, e.g., for assessing whether participants are cured or not after a treatment (vs. a placebo). Data would look like (treatment – cured, treatment – cured, placebo – not cured, placebo – cured, treatment – cured, treatment – not cured, ... [with treatment / placebo being in the first column and cured / not cured in the second]).

For (2), we have one column where different categories occur with certain frequencies (e.g., heart, diamond, club, heart, spade, diamond, ...).

For (3) the data assess the same characteristics (e.g., an opinion or a health state: healthy or sick) at two points in time (e.g., before and after treatment). Data would look like (if the characteristic is sick: yes – yes, yes – no, yes – no, no – no, yes – no, no – yes, ... [also arranged in two columns]).





## $\chi^2$ -test of association: step-by-step

- $\chi^2$ -distribution: compare an **observed** distribution of frequencies with a theoretically **expected** distribution
- example: new drug (compared to a placebo) × cured vs. not cured

<i>observed</i>	<i>cured</i>			<i>expected</i>	<i>cured</i>		
<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>	<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>
<i>drug</i>	30	10	40	<i>drug</i>	20	20	40
<i>placebo</i>	10	30	40	<i>placebo</i>	20	20	40
<i>total</i>	40	40	80	<i>total</i>	40	40	80

REFRESHER: STATISTICAL ANALYSES

SEBASTIAN.JENTSCHKE@UIB.NO

SLIDE 9



Let's try to do this in a step-by-step fashion, using method (1): the  $\chi^2$ -test of association. For example, let's assume that you tried out a new drug (compared to a placebo) and then assess whether this new drug leads to a higher proportion of people becoming healthy as a consequence. So, let's say: In the group that received the drug, 30 participants got healthy while 10 weren't cured. In the placebo group, 10 participants got healthy and 30 weren't cured. This distribution is shown in the left table. In addition, totals were calculated for all rows and columns.



## $\chi^2$ -test of association: step-by-step

- **observed**: row sums, drug-recipients (cured + not cur.)  
col. sums, cured (drug + placebo)
- **expected**:  $40 / 80 = 0.5$  (rows)  $\cdot$   $0.5$  (cols)  $\cdot$   $80 = 20$

<i>observed</i>	<i>cured</i>		
<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>
<i>drug</i>	30	10	40
<i>placebo</i>	10	30	40
<i>total</i>	40	40	80

<i>expected</i>	<i>cured</i>		
<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>
<i>drug</i>	20	20	40
<i>placebo</i>	20	20	40
<i>total</i>	40	40	80



Based upon these totals, in the right table is shown what theoretically would have been expected (given these totals) if all conditions were evenly distributed (which is 20 for each condition).

If you wonder, how these 20 came about, think about the totals in the rows: the probability for having got treatment is  $40 / 80 = 0.5$  (the probability for the placebo-condition is the same). Now, we turn to the columns: The probability for being cured is  $40 / 80 = 0.5$  (as it is for not cured). That is for the combination drug – cured, the expected occurrence is  $80 \cdot 0.5 \cdot 0.5 = 20$ .



## $\chi^2$ -test of association: step-by-step

- for each cell:  
 $(\text{observed} - \text{expected})^2 / \text{expected}$   
 $(30 - 20)^2 / 20 = 10^2 / 20 = 100 / 20 = 5$   
 $(10 - 20)^2 / 20 = -10^2 / 20 = 100 / 20 = 5$

<i>observed</i>	<i>cured</i>			<i>expected</i>	<i>cured</i>		
<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>	<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>
<i>drug</i>	30	10	40	<i>drug</i>	20	20	40
<i>placebo</i>	10	30	40	<i>placebo</i>	20	20	40
<i>total</i>	40	40	80	<i>total</i>	40	40	80



That means, under our alternative hypothesis, the treatment has an effect and more people are cured, we expect an distribution like in the left table. Under the null hypothesis, there is no effect and the expected frequency of occurrences would be similar distributed as in the right table. Now, a  $\chi^2$ -value is calculated for each cell according to the formula  $(\text{observed} - \text{expected})^2 / \text{expected}$  (and afterwards summed up).



## $\chi^2$ -test of association: step-by-step

- summed over all four cells / conditions:  
 $\chi^2 = 5 + 5 + 5 + 5 = 20$  – four conditions  $\rightarrow$   $df = 4 - 1$
- compared to a critical  $\chi^2$  for  $df = 3$ ,  $\alpha = 0.05$ :  
 $20 > 7.815 \rightarrow$  reject  $H_0 \rightarrow$  *drug or placebo affect if cured*

<i>observed</i>	<i>cured</i>			<i>expected</i>	<i>cured</i>		
<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>	<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>
<i>drug</i>	30	10	40	<i>drug</i>	20	20	40
<i>placebo</i>	10	30	40	<i>placebo</i>	20	20	40
<i>total</i>	40	40	80	<i>total</i>	40	40	80



That is, for the first cell (drug, cured) this is  $(30 - 20)^2 / 20 = 100 / 20 = 5$ . For the second cell (placebo, cured) the value is  $(10 - 20)^2 / 20 = 100 / 20 = 5$ . The third (drug, not cured) and the fourth cell (placebo, not cured) the results are the same as for first cell and the second cell (how conveniently I chose that example...).



## $\chi^2$ -test of association: step-by-step

- summed over all four cells / conditions:  
 $\chi^2 = 5 + 5 + 5 + 5 = 20$  – four conditions  $\rightarrow$   $df = 4 - 1$
- compared to a critical  $\chi^2$  for  $df = 3$ ,  $\alpha = 0.05$ :  
 $20 > 7.815 \rightarrow$  reject  $H_0 \rightarrow$  *drug or placebo affect if cured*

<i>observed</i>	<i>cured</i>			<i>expected</i>	<i>cured</i>		
<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>	<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>
<i>drug</i>	30	10	40	<i>drug</i>	20	20	40
<i>placebo</i>	10	30	40	<i>placebo</i>	20	20	40
<i>total</i>	40	40	80	<i>total</i>	40	40	80



The values for all cells are summed up, i.e.,  $\chi^2 = 5 + 5 + 5 + 5 = 20$ . We had four cells / categories (denoted  $k$ ), therefore the degrees of freedom ( $df$ ) used for evaluating this  $\chi^2$ -value are  $df = k - 1$ .

Now, we only have to determine if that empirical  $\chi^2$  is above the critical threshold were we can reject the  $H_0$ . The critical threshold for  $\alpha = 0.05$  and  $df = 3$  is 7.815, meaning that the empirical  $\chi^2 = 20$  is far above that critical value. We can therefore reject the null hypothesis and conclude that our drug made an effect on how many people were cured.

When look at the table in the jamovi-book where the critical  $\chi^2$ -values are given (Figure 10.3, p. 220; Navarro & Foxcroft, 2022) we can see that our empirical  $\chi^2 = 20$  would even have been significant at  $\alpha = 0.001$  (critical  $\chi^2 = 16.266 >$  empirical  $\chi^2 = 20$ ).



## $\chi^2$ -test of association: step-by-step

- repetition with a (little) more complicated example
- rows:  $44 / 80 = 0.55$   
columns:  $50 / 80 = 0.625$   
expected:  $0.55 \cdot 0.625 \cdot 80 (= 50 \cdot 44 / 80) = 27.5$

<i>observed</i>	<i>cured</i>			<i>expected</i>	<i>cured</i>		
<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>	<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>
<i>drug</i>	40	4	44	<i>drug</i>	27.5	16.5	44
<i>placebo</i>	10	26	36	<i>placebo</i>	22.5	13.5	36
<i>total</i>	50	30	80	<i>total</i>	50	30	80



Now, let's do the same with an example where the values are little less evenly distributed. The probability of occurrence for the rows are: drugs  $\rightarrow 44 / 80 = 0.55$ ; placebo  $\rightarrow 36 / 80 = 0.45$ .

The probability of occurrence for the columns are: cured  $\rightarrow 50 / 80 = 0.625$ ; not cured  $\rightarrow 30 / 80 = 0.375$ .

From these probabilities we can calculate what is expected in the cells:

drugs – cured  $\rightarrow 0.55 \cdot 0.625 \cdot 80 = 27.5$   
 placebo – cured  $\rightarrow 0.45 \cdot 0.625 \cdot 80 = 22.5$   
 drug – not cured  $\rightarrow 0.55 \cdot 0.375 \cdot 80 = 16.5$   
 placebo – not cured  $\rightarrow 0.45 \cdot 0.375 \cdot 80 = 13.5$



## $\chi^2$ -test of association: step-by-step

- $\chi^2$ :
  - $(40 - 27.5)^2 / 27.5 = 12.5^2 / 27.5 = 5.682$
  - $(4 - 16.5)^2 / 16.5 = -12.5^2 / 16.5 = 9.470$
  - $(10 - 22.5)^2 / 22.5 = -12.5^2 / 22.5 = 6.944$
  - $(26 - 13.5)^2 / 13.5 = 12.5^2 / 13.5 = 11.574$

<i>observed</i>	<i>cured</i>			<i>expected</i>	<i>cured</i>		
<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>	<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>
<i>drug</i>	40	4	44	<i>drug</i>	27.5	16.5	44
<i>placebo</i>	10	26	36	<i>placebo</i>	22.5	13.5	36
<i>total</i>	50	30	80	<i>total</i>	50	30	80



The  $\chi^2$  with the cells is calculated as shown on the slide. Maybe, take the chance to follow the calculations on paper, with a calculator or in Excel so that you can get a feel for the calculations.



## $\chi^2$ -test of association: step-by-step

- $\chi^2 = 5.682 + 9.470 + 6.944 + 11.574 = 33.670$
- $\chi^2 = 33.670 > 7.815$  (crit.  $\chi^2$  for  $\alpha = 0.05$ ,  $df = 3$ )

<i>observed</i>	<i>cured</i>			<i>expected</i>	<i>cured</i>		
<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>	<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>
<i>drug</i>	40	4	44	<i>drug</i>	27.5	16.5	44
<i>placebo</i>	10	26	36	<i>placebo</i>	22.5	13.5	36
<i>total</i>	50	30	80	<i>total</i>	50	30	80



Now, we sum up the values we calculated on the previous slide:  $5.682 + 9.470 + 6.944 + 11.574 = 33.670$ .

From before do we know the critical  $\chi^2$  is 7.815 (for  $\alpha = 0.05$ ) and 16.266 (for  $\alpha = 0.001$ ). The empirical  $\chi^2$  is above either of these critical values and the effect of our treatment (in comparison with the placebo) on how many people were cured is statistically highly significant.

That means, given that low p-value, we can be relatively certain that we don't make an error rejecting the  $H_0$ . The  $H_0$  assumes no relation between our treatment and whether people are cured and would result in a distribution like in the right table (expected).





## $\chi^2$ -test of association: step-by-step

- $\chi^2 = 5.682 + 9.470 + 6.944 + 11.574 = 33.670$
- $\chi^2 = 33.670 > 7.815$  (crit.  $\chi^2$  for  $\alpha = 0.05$ ,  $df = 3$ )

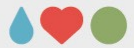
<i>observed</i>				<i>expected</i>			
	<i>cured</i>			<i>cured</i>			
<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>	<i>treatment</i>	<i>yes</i>	<i>no</i>	<i>total</i>
<i>drug</i>	40	4	44	<i>drug</i>	27.5	16.5	44
<i>placebo</i>	10	26	36	<i>placebo</i>	22.5	13.5	36
<i>total</i>	50	30	80	<i>total</i>	50	30	80



This demonstrates a very basic principle underlying statistical analyses. We compare one case, where we assume the distribution could have occurred “naturally” (the right “expected”-table) to our measured (empirical) distribution of values (the left “observed”-table). If the difference is large enough, that we can assume that they did not happen “by chance”.

That is we – in a way – compare the differences we find in our data to what could have happened randomly (plus a certain margin of error).

Let’s briefly summarize again what the  $\chi^2$ -test is about: It compares an observed to an expected frequency distribution (i.e. how often a certain event is occurring vs. it is expected to occur if it were “by chance”).



## $\chi^2$ -test of association in jamovi

- download ChiSquared\_DrugPlacebo.omv from MittUiB → Modules → Refreshing univariate...
- open it in jamovi: ☰ → Open → This PC
- check the first example analysis: Frequencies → Contingency Tables → Independent Samples assign “**condition**” → Rows, “**state**” → Columns
- check the second example analysis: Frequencies → Contingency Tables → Independent Samples assign “**condition2**” → Rows, “**state2**” → Columns



Now, let's try to “replicate” the two analyses that I demonstrated on the slides. To do this, please download the data set from MittUiB → Modules → Refreshing univariate... → ChiSquared\_DrugPlacebo.csv.

Open jamovi and select ☰ (located in the top-left corner of the jamovi-window). Choose “Open” → “This PC” and go to the place where you stored the downloaded file.

In the data spreadsheet, you will find four columns: condition and state are the data from the first example, condition2 and state2 those from the second.



# $\chi^2$ -test of association in jamovi

Contingency Tables

condition2

state2

Q

Rows

→ condition

Columns

→ state

Counts (optional)

→

Layers

→

> | Statistics

> | Cells

Contingency Tables

condition	state		Total
	cured	not cured	
drug	30	10	40
placebo	10	30	40
<b>Total</b>	<b>40</b>	<b>40</b>	<b>80</b>

$\chi^2$ Tests			
	Value	df	p
$\chi^2$	20.000	1	< .0001
Fisher's exact test	8.704		< .0001
N	80		

REFRESHER: STATISTICAL ANALYSES

SEBASTIAN.JENTSCHKE@UIB.NO

SLIDE 19



For replicating the first example analysis above, chose Frequencies → Contingency Tables →  $\chi^2$ -test of association. If you assign the variable named “condition” to Rows and “state” to Columns you will get a Contingency Table looking pretty much like the left table in the first example on the slides above.

The  $\chi^2$ -value that was calculated is exactly the one we got in the first example: 20 and the p-value is shown to be < .0001.

Please note, that your value might show as < .001 (i.e., only having three decimals). This is a setting, not a mistake. You can change that setting in the options ⋮ (the three small dots in the top-right corner of the jamovi-window). My settings are: Number format: 3 dp (decimal places) and p-value format: 4 dp.



# $\chi^2$ -test of association in jamovi

Contingency Tables

Rows: condition2

Columns: state2

Counts (optional):

Layers:

> | Statistics

> | Cells

Contingency Tables

condition2	state2		Total
	cured	not cured	
drug	40	4	44
placebo	10	26	36
Total	50	30	80

$\chi^2$  Tests

	Value	df	p
$\chi^2$	33.670	1	< .0001
N	80		



The second example analysis is replicated using Frequencies → Contingency Tables →  $\chi^2$ -test of association and assigning “condition2” to Rows and “state2” to Columns. Again, the  $\chi^2$ -value is the same as we calculated on the slides. Try one last thing. Revert the assignment of “condition2” and “state2”, i.e., assign “state2” to Rows and “condition2” to Columns. You will see that the  $\chi^2$ -value stays the same. That means it doesn't matter which variable you assign to rows or columns (and it shouldn't matter).

You can watch Barton Poulson's videos if you want to have an explanation from a different angle: go to <https://datalab.cc/jamovi/>, click on ☰ (top-right in the video) and select video 45 for a general introduction into all method in the “Frequencies” menu and video 48 for an introduction into the  $\chi^2$ -test of association we just conducted.



## $\chi^2$ goodness-of-fit in jamovi

### main difference to the $\chi^2$ -test of association:

- the  $\chi^2$ -test of association uses two variables and determines the expected frequency based on how often each step (e.g., cured vs. not cured) occurs (i.e., what is expected is based upon the totals in rows / columns)
- the  $\chi^2$  goodness-of-fit uses one variable with n trials; it assumes that all possible events occur with the same probability / frequency



Compared to the  $\chi^2$ -test of association, the  $\chi^2$  goodness-of-fit test is a bit simpler but follows the same principles.

For the  $\chi^2$ -test of association we conducted before, we had cross-tabulated the different steps within two conditions (placebo vs. drugs; cured vs. not cured) and derived the expected values from the totals in the rows and columns.

For the  $\chi^2$  goodness-of-fit, the expected occurrences are based on the number of trials and the number of categories in the data. Taking our deck of cards and assuming 200 trials, we expect that each suit (clubs, diamonds, hearts, spades) should be drawn about 50 times (assuming that it is a “normal” deck that wasn’t faked).



## $\chi^2$ goodness-of-fit in jamovi

- $\equiv$  → Open → Data Library → learning statistics with jamovi” (lsj-data) → Randomness  
*can't find it?* install lsj-data (cf. Introduction lecture)
- use the R-code (details on the next slide)
- Frequencies → One Sample Proportion Test → N outcomes ( $\chi^2$  Goodness of fit)  
assign “**choice\_1**” to Variable
- Frequencies → One Sample Proportion Test → N outcomes ( $\chi^2$  Goodness of fit)  
assign “**choice\_2**” to Variable



Now we turn to an example for the  $\chi^2$  goodness-of-fit test that is described in the jamovi-book (Navarro & Foxcroft, 2022; Chapter 10.1.1 to 10.1.7, p. 212 – 221).

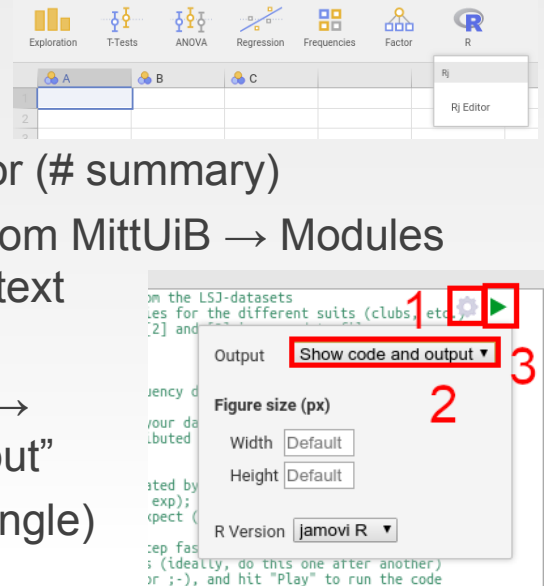
For conducting that analysis, we open the Randomness-dataset. Choose  $\equiv$  → Open → Data Library and enter the folder “learning statistics with jamovi” (lsj-data). If you can’t find it, did you possibly forget to install the lsj-data module (see the Introduction lecture for how to do that)?

The Randomness-data-set contains three columns (of which the first “id” isn’t of particular interest to us). The other two columns (“choice\_1”, “choice\_2”) contain data about which suit of cards (diamonds, hearts, clubs, spades) was drawn from a deck by different participants. We would like to determine whether each suit (clubs, diamonds, hearts, spades) occurs with equal frequency.



## $\chi^2$ goodness-of-fit in jamovi

- “R”-icon → Rj editor  
*can't find it? install the Rj module*
- remove the code that is in the editor (# summary)
- download Syntax\_ChiSquared.R from MittUiB → Modules → Refreshing..., copy & paste the text from this file to the editor
- Cogwheel (next to “Play”-symbol) → set Output to “Show code and output”
- run the code with “Play” (green triangle)



REFRESHER: STATISTICAL ANALYSES

It is also possible to run a chi-squared statistics manually, using R-syntax (you will be surprised, it is much easier than you might expect).

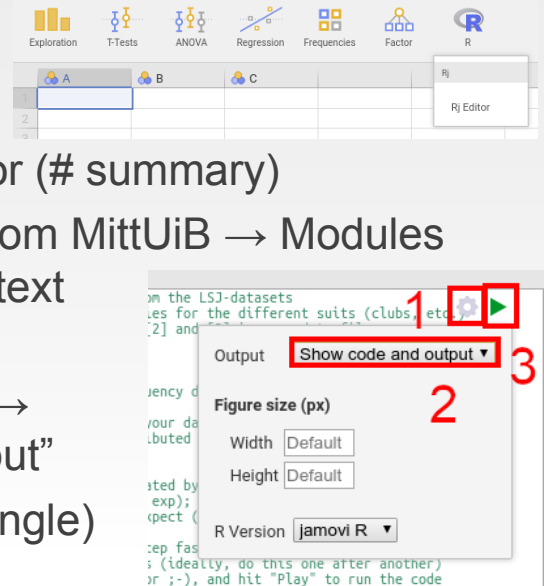
Download ChiSquare.R from MittUiB → Modules → Refreshing... → Syntax\_ChiSquared.R and open it in a text editor. Open Rj you click on the R-symbol in the menu bar and choose Rj editor (if you can't find it, ensure that you installed the module). There is some R-code written (# summary...); please remove that. Instead copy and paste the content of Syntax\_ChiSquared.R into it.

Click on the cogwheel (left of the green triangle) in the Rj-window and select Output – “Show code and output” (it would also work with “Show output” but it is easier if you have the R-commands, my comments and the results in the output instead of the results alone). You now run the syntax by pressing the green triangle (“Play”).



## $\chi^2$ goodness-of-fit in jamovi

- “R”-icon → Rj editor  
*can't find it? install the Rj module*
- remove the code that is in the editor (# summary)
- download Syntax\_ChiSquared.R from MittUiB → Modules  
→ Refreshing..., copy & paste the text from this file to the editor
- Cogwheel (next to “Play”-symbol) → set Output to “Show code and output”
- run the code with “Play” (green triangle)



REFRESHER: STATISTICAL ANALYSES

You will see that the  $\chi^2$  and p-values are the same as when running the analyses through the graphical user interface in jamovi.

There is another thing that you can try. If you go to l. 19 of the script:  `#(obs_1 - exp)` and remove the `#`-sign from the begin of the line, you can stepwise follow the analysis in chapter 10.1.3. of the jamovi-book (p. 214 – 215). Please note that the l. 19 is in the table on p. 214. Now put the `#` back at the begin of l. 19, remove the `#` from l. 20 and hit play again. Continue until you reached l. 22

One thing you might realize is that R-code (hopefully) appears less frightening than you possibly anticipated.

Among Barton Poulson’s videos ([datalab.cc/jamovi](https://datalab.cc/jamovi)), video 47 provides an introduction into the  $\chi^2$ -goodness-of-fit test we just conducted.





# $\chi^2$ goodness-of-fit in jamovi

Proportion Test (N Outcomes)

choice\_2

id

Variable

choice\_1

Counts (optional)

Expected counts

> | Expected Proportions

Proportion Test (N Outcomes)

Proportions **choice\_1**

Level	Count	Proportion
clubs	35	0.175
diamonds	51	0.255
hearts	64	0.320
spades	50	0.250

$\chi^2$  Goodness of Fit

$\chi^2$	df	p
8.440	3	0.0377

Proportion Test (N Outcomes)

Proportions **choice\_2**

Level	Count	Proportion
clubs	68	0.340
diamonds	44	0.220
hearts	41	0.205
spades	47	0.235

$\chi^2$  Goodness of Fit

$\chi^2$	df	p
9.000	3	0.0293



To do this in jamovi, we use Frequencies → N outcomes ( $\chi^2$  Goodness of fit), and assign choice\_1 to Variable. Create a new analysis by repeating that but this time assign choice\_2 to Variable. Now, you should see two outputs.

In the first, there are two tables, one saying “Proportions – choice\_1” and another saying “ $\chi^2$  Goodness of Fit”. The other output contains the tables “Proportions – choice\_2” and another with the  $\chi^2$ -test.

The  $\chi^2$ -value are 8.440 for choice\_1 (the same result as in the jamovi-book, p. 215) and 9.000 for choice\_2. Both values are significant (0.0377 for choice\_1, 0.0293 for choice\_2).



## McNemar test in jamovi

- used when we have repeated measurements that are dependent of each other (e.g., how many people were cured after an intervention or how many people changed their opinions)
- key to distinguish McNemar from other  $\chi^2$ -tests is the notion of dependence: a person already has an opinion or is healthy vs. sick  $\leftrightarrow$  when we draw a card, the card is independent of the card before



In some cases, we have repeated measurements that are dependent of each other (e.g., because we want to assess how many people were cured after an intervention or how opinions changed over time). In such case, we can use the McNemar-test to assess that.

When comparing the McNemar-test to the  $\chi^2$ -test of association we conducted before, the key distinction is the notion of independence: With the Randomness-example and the different suits of cards, the first card that is drawn is independent of the second one.

In contrast, with the AGPP-example, a person already has a political conviction. Therefore, the second poll is dependent on what opinion or conviction the person had in the beginning.



## McNemar test in jamovi

- ≡ → Open → Data Library → learning statistics with jamovi” (lsj-data) → AGPP (containing opinions / claims whether to vote for the party before and after an advertisement)
- Frequencies → Contingency Tables → Paired Samples (McNemar test)  
assign “**response.before**” to rows and “**response.after**” to columns



To conduct this analysis in jamovi, we open the AGPP-dataset. That dataset contains three columns, two of which we are using in our analysis – response.before and response.after – saying how opinions of people (“How likely are you going to vote for our party?”) changed in connection with seeing advertisement for the party.

We click on “Frequencies” → “Paired Samples (McNemar test)” and assign response.before to rows and response.after to columns.



# McNemar test in jamovi

Paired Samples Contingency Tables

id

Rows  
response\_before

Columns  
response\_after

Counts (optional)

$\chi^2$

$\chi^2$  continuity correction

Log odds ratio exact

Percentages

Row

Column

## Paired Samples Contingency Tables

### Contingency Tables

response_before	response_after		Total
	no	yes	
no	65	5	70
yes	25	5	30
Total	90	10	100

### McNemar Test

	Value	df	p
$\chi^2$	13.333	1	0.0003
N	100		

Why is this not the outcome the party anticipated?

REFRESHER: STATISTICAL ANALYSES

SEBASTIAN.JENTSCHKE@UIB.NO

SLIDE 28



We got a highly significant result:  $\chi^2 = 13.33$ ,  $p < 0.001$ .

However, although we can say that opinions changed, possibly not in the direction the party intended. Whereas 30 people would have voted for the party before seeing the advertisement, only 10 would do so afterwards. Whereas the party lost 25 of its original voters, they got only 5 new ones.

Among Barton Poulson's videos, video 49 provides an introduction into the McNemar test we just conducted.

There are another two helpful section in Chapter 10 of the jamovi-book (10.1.9 and 10.1.10). These sections describe how to report results like the one from the  $\chi^2$ -test in a manuscript and make some comments on statistical notation.



# Categorical predictor and continuous outcome variables (Ch. 11)



## Introduction

- t-tests explore whether a **categorical predictor** (with two steps) affects values in a **continuous variable**
- **three types:**
  1. **One Sample t-test:** does the mean in a sample differ from a test value?
  2. **Independent Samples t-test:** are the values in two groups different from another?
  3. **Paired Samples t-test:** do the values change between a first and a second measurement?



When we assess the influence of a categorical predictor with two steps (e.g., male-female) onto a continuous variable, we typically use a t-test. The two categories can also be two measurement-time-points (e.g., before vs. after treatment).

For variables that differ between participants (a so-called between-subjects-factor) we use the Independent samples t-tests (since the measurements are collected e.g., from one male participant which is independent from another female participant and yet from another male participant).

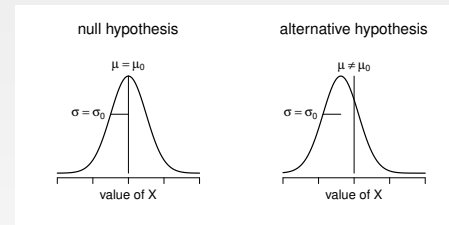
For variables that differ within a participant (a so-called within-subject-factor), we use a Paired Samples t-test (since we assess the difference in a pair of variables, e.g., of a measurement before vs. after an intervention).

However, we will first turn to what is called a One Sample t-test.



## z-test: step-by-step

- “model” for the principle of the One Sample t-test
- we know mean and std. deviation for the characteristic in the population
- we assess whether our sample mean is different from that in the population
- idea and procedure are based upon the **Central Limit Theorem**: the distribution of values follows a normal distribution; the larger the sample, the better estimates the sample mean the population mean



The basic principle behind the One Sample t-test is more easily explained using the z-test.

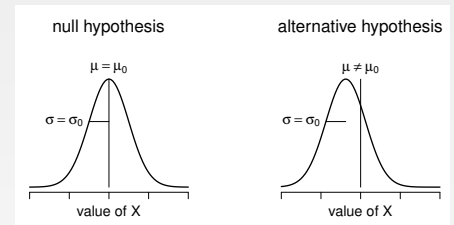
The basis for the z-test is that we know mean and standard deviation for a certain variable in the population. This condition is only fulfilled on very few occasions, one being standardized tests (e.g., IQ tests) which are normed (typically to have a population mean of 100 and a standard deviation of 15 points).

Given that these conditions are rarely fulfilled, the z-test is almost never used in reality, but – given that it follows the same logic as the One Sample t-test – is often used to introduce the concept and to serve didactic purposes.



## z-test: step-by-step

- “model” for the principle of the One Sample t-test
- we know mean and std. deviation for the characteristic in the population
- we assess whether our sample mean is different from that in the population



- idea and procedure are based upon the **Central Limit Theorem**: the distribution of values follows a normal distribution; the larger the sample, the better estimates the sample mean the population mean



The general idea and procedure of the z-test are based upon the Central Limit Theorem.

It assumes that:

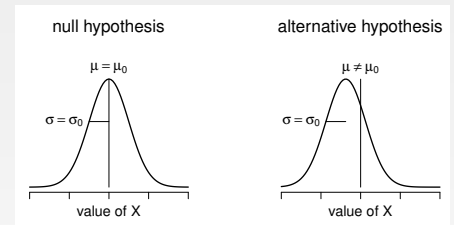
- 1) The distribution of measurements recorded from a sample corresponds approximately to the normal distribution.
- 2) The mean (average) of almost any set of independent and randomly generated variables ( $\bar{X}$ ) converges to the mean in the underlying population ( $\mu$ ) from which the sample is taken.
- 3) The standard deviation in the sample ( $s$ ) converges to the standard deviation in the population ( $\sigma$ ). However, the smaller the sample, the larger the standard deviation (as we have a higher chance to take measurements that fall within the more extreme ends [tails] of the distribution).





## z-test: step-by-step

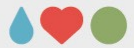
- “model” for the principle of the One Sample t-test
- we know mean and std. deviation for the characteristic in the population
- we assess whether our sample mean is different from that in the population
- idea and procedure are based upon the **Central Limit Theorem**: the distribution of values follows a normal distribution; the larger the sample, the better estimates the sample mean the population mean



Therefore, we have to “correct” the standard deviation in the population ( $\sigma$ ) with this formula 
$$\sigma = s \cdot \frac{(n - 1)}{n}$$
 The formula says that we have to multiply the standard deviation by  $(n - 1)$  and divide this by  $n$ . This makes it smaller. The fewer participants the more  $\sigma$  is corrected down. This is based upon the assumption that extreme values (leading to large standard deviation in the sample) are more likely when the sample is small (assumption 3 of the Central Limit Theorem on the previous slide).

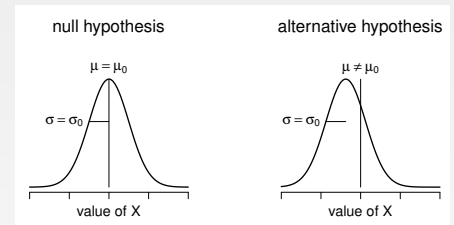
The central formula for the z-test is: 
$$z = \frac{\mu - \bar{x}}{\sigma / \sqrt{n}}$$

Above the fraction bar, the mean the “real” mean of a characteristic in the population ( $\mu$ ) and mean of the characteristic measured in the sample ( $\bar{x}$ ) are subtracted. This is set in relation to the error we likely make when taking these measurements (the part below the fraction bar). It is the standard deviation in the population ( $\sigma$ ) divided by the square root of the sample size ( $n$ ).



## z-test: step-by-step

- “model” for the principle of the One Sample t-test
- we know mean and std. deviation for the characteristic in the population
- we assess whether our sample mean is different from that in the population
- idea and procedure are based upon the **Central Limit Theorem**: the distribution of values follows a normal distribution; the larger the sample, the better estimates the sample mean the population mean



So, that was a lot of rather complex sentences and some formulas.

Let's try to take this into praxis. Using the formula

$$z = \frac{\mu - \bar{x}}{\sigma / \sqrt{n}}$$

is basically all we need.

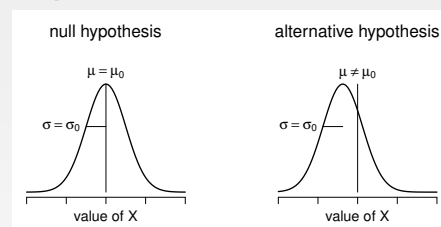
We subtract the mean in the sample ( $\bar{x}$ ) from the known population mean ( $\mu$ ) and set this into relation to the standard error that we are likely going to make when comparing the mean in sample and population ( $\sigma / \text{square root of } n$ ).

The rationale is as follows: We know that when we measure the mean in our sample it always can only be an estimate or an approximation of the “true” value in the population.



## z-test: step-by-step

- “model” for the principle of the One Sample t-test
- we know mean and std. deviation for the characteristic in the population
- we assess whether our sample mean is different from that in the population



- idea and procedure are based upon the **Central Limit Theorem**: the distribution of values follows a normal distribution; the larger the sample, the better estimates the sample mean the population mean



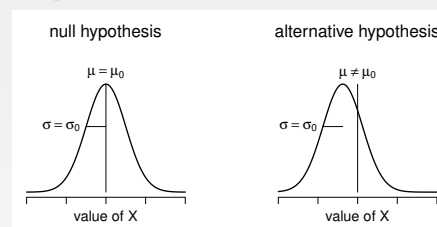
We also know that the standard deviation in the population makes the mean that we get for our sample “jitter” a little around the “true” mean in the population. The larger our sample, the better estimates the mean we measured in our sample the mean in the population, hence the lower is the error we make with this estimation.

As a consequence we set the standard deviation in the population in relation to (i.e., divide it by) the square root of the sample size.



## z-test: step-by-step

- “model” for the principle of the One Sample t-test
- we know mean and std. deviation for the characteristic in the population
- we assess whether our sample mean is different from that in the population
- idea and procedure are based upon the **Central Limit Theorem**: the distribution of values follows a normal distribution; the larger the sample, the better estimates the sample mean the population mean



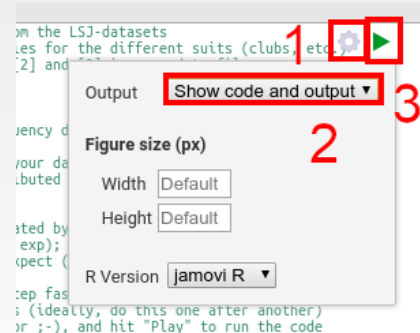
The standard deviation in the population is so to say a natural “source” of variation and for the error we make when estimating the mean in the population (because the sample we take will have values around the mean and vary according to the standard deviation in the population).

We now compare this error that occurred “naturally” or “randomly” with the difference we obtained between the mean in our sample and the mean in the population. If the latter difference is much larger than the “natural” variation, we can assume that the difference is “real” and has (statistical) significance.



## z-test: step-by-step

- open zeppo – Example analysis.omv under MittUiB  
→ Modules → Refreshing univariate...
- go to the part  
click the cogwheel (next to “Play”), set Output to “Show code and output” and hit “Play”, check output / comments



REFRESHER: STATISTICAL ANALYSES



For a demonstration of the basic principles, I prepared a R-script. Download zeppo – Example analysis.omv from MittUiB → Modules → Refreshing univariate...

This example already contains the R-code. Go to the part that contains the R output (under the heading “z-test: step-by-step” and click on the code underneath. This should open the Rj Editor and you can use the green triangle symbol (“Play”) to run the code. Ensure that you changed the Output to “Show code and output” using the cogwheel in Rj. The logic of the script follow the z-test that is described on p. 242 – 246 of the jamovi-book (Navarro & Foxcroft, 2022).



## z-test: step-by-step

- for a visualization of the Central Limit theorem  
<https://garhtarr.shinyapps.io/statstar/>  
→ Inference → Central Limit Theorem
- change the slider “Sample size” and check how the green distribution changes – narrower, wider? what happens to mean and std. deviation? can you say why?



For those, who want to play a little around with this concepts and the mathematics beyond it in a visual way, use the demonstration on Garth Tarr’s webpage <https://garhtarr.shinyapps.io/statstar/> → Inference → Central Limit Theorem.

This visualization simulates what will happen if you take a sample of size  $n$  (this can be set with the slider under “Sample size:”) and repeat collecting data from that sample several times (this can be set with Number of samples:”).



## z-test: step-by-step

- for a visualization of the Central Limit theorem  
<https://garhtarr.shinyapps.io/statstar/>  
→ Inference → Central Limit Theorem
- change the slider “Sample size” and check how the green distribution changes – narrower, wider? what happens to mean and std. deviation? can you say why?



Play a little around with these values.

It will show you two things (I used Sample size: 20 and Number of samples: 200):

- (1) the mean of  $\bar{x}$  in the green figure (showing the distribution of means in the sample you took) is always quite close to the mean in the population – typically you will get differences in the second decimal (e.g., -0.08 or 0.02, the most extreme value I got was 0.30 which is still quite small given that the standard deviation in the population is 10).
- (2) The sd of  $\bar{x}$  (SE – denoting the standard error) is around between 2.2 and 2.3 (i.e., close to  $10 / \sqrt{20} = 2.24$ ). You can “repeat” that with moving the “Number of samples”-slider a little up and down, the values for “mean of  $\bar{x}$ ” and “sd of  $\bar{x}$ ” will vary a bit, but not substantially.



## z-test: step-by-step

- for a visualization of the Central Limit theorem  
<https://garhtarr.shinyapps.io/statstar/>  
→ Inference → Central Limit Theorem
- change the slider “Sample size” and check how the green distribution changes – narrower, wider?  
what happens to mean and std. deviation?  
can you say why?



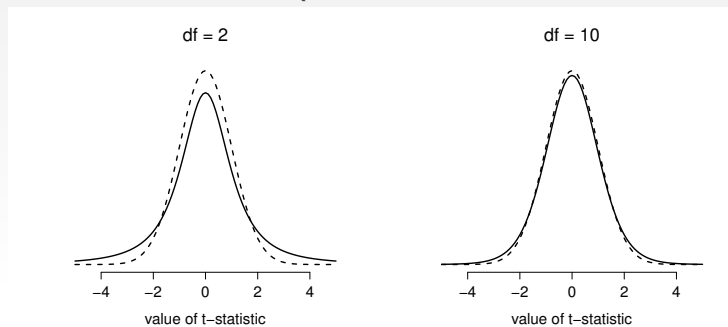
Now, change the slider “Sample size” from 20 to 40. You will see that “mean of x” will vary a bit (but still be around 0) whereas “sd of x” goes down from between 2.2 and 2.3 (as we had for  $n=20$ ) to between 1.5 and 1.6 (which is close to  $10 / \sqrt{40} = 1.58$ ).





## One Sample t-test: step-by-step

- assume we don't know mean and std. deviation in the population, but estimate it from the sample
  - uncertainty we have to correct for
  - t-distribution (instead of z-distribution)



SLIDE 41

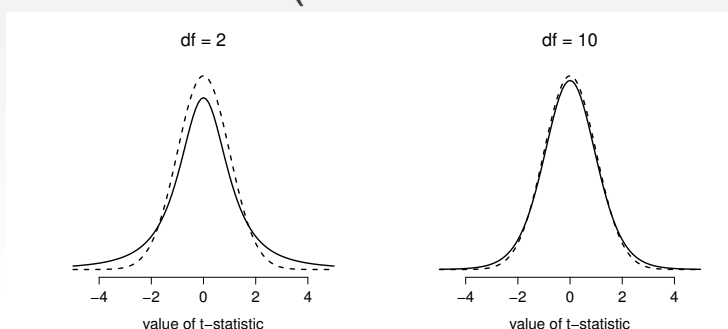


Now, let's continue with a One Sample t-test. A essential difference to the z-test is that we DON'T assume this time that we knew the "true" value of the standard deviation in the population. What we do instead is that we estimate the standard deviation in the population using the standard deviation in our sample. Such estimation introduces some degree of uncertainty (we estimate, we don't know for sure) and we therefore have to apply some form of correction.



## One Sample t-test: step-by-step

- assume we don't know mean and std. deviation in the population, but estimate it from the sample
  - uncertainty we have to correct for
  - t-distribution (instead of z-distribution)



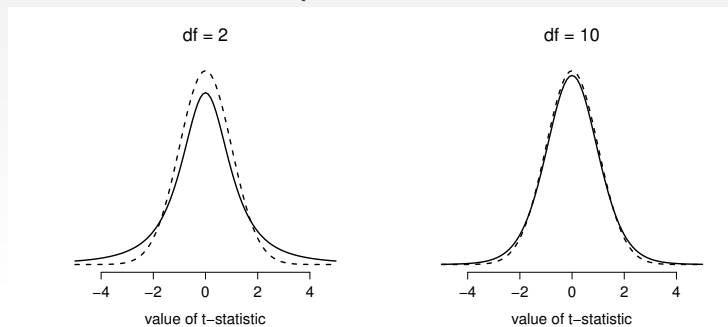
One way to do that is to use a (slightly different) distribution the so-called t-distribution. At the first glance, this distribution very much looks like a normal distribution but it is a little lower in the middle and a little higher on the tails. It also introduces degrees of freedom (df) and changes its shape depending on how many degrees of freedom we have (whereas the z-distribution always is the same). The degrees of freedom are dependent on the size of your sample ( $df = n - 1$ ).

The larger the degrees of freedom get, the more the t-distribution looks like a z-distribution (i.e., a standard normal distribution). For  $df > 30$  it is (more or less) impossible to see a difference to a standard normal distribution.



## One Sample t-test: step-by-step

- assume we don't know mean and std. deviation in the population, but estimate it from the sample
  - uncertainty we have to correct for
  - t-distribution (instead of z-distribution)



SLIDE 43



If we now compare the formula for the t-test is based upon (left) with that from the z-test (right), it is easy to see that they pretty much look the same:

$$t = \frac{\mu - \bar{x}}{\hat{\sigma} / \sqrt{n}} \qquad z = \frac{\mu - \bar{x}}{\sigma / \sqrt{n}}$$

The only exception is that for the t-test, we are using an estimate for the standard deviation in the population (which we derive from the standard deviation in the sample) – a  $\sigma$  with a little “hat” on it – whereas in the z-test, we know the “true” value  $\sigma$ .



## One Sample t-test: step-by-step

- ☰ → Open → Data Library → learning statistics with jamovi (lsj-data) → Zeppo
- T-Tests → One Sample t-test  
assign “x” to “Dependent variable”  
enter 67.5 in “Test value”  
tick “Shapiro-Wilk” and check for normality  
tick “Mean difference”, “Confidence interval”,  
“Effect size” and “Descriptives”



To conduct a One Sample t-test in jamovi, we first have to open a data file. This can be found via ☰ → Open → “Data Library” → “learning statistics with jamovi” (lsj-data) → Zeppo (hint, type ze in the search field above the file list to find the file easier). Then, click on the “T-Tests”-icon and choose “One Sample t-test”.

One Sample T-Test

Dependent Variables: x

Tests:  Student's t,  Bayes factor,  Wilcoxon rank

Hypothesis: Test value 67.5,  ≠ Test value,  > Test value,  < Test value

Additional Statistics:  Mean difference,  Effect size,  Confidence interval,  Descriptives

Assumption Checks:  Normality (Shapiro-Wilk),  Normality (Q-Q plot)

One Sample T-Test

	statistic	df	p	Mean difference	95% Confidence Interval		Cohen's d	
					Lower	Upper		
x	Student's t	2.255	19.000	0.0361	4.800	0.344	9.256	0.504

Note. H<sub>a</sub> population mean ≠ 67.5

Test of Normality (Shapiro-Wilk)

	W	p
x	0.962	0.5856

Note. A low p-value suggests a violation of the assumption of normality

Descriptives

	N	Mean	Median	SD	SE
x	20	72.300	75.000	9.521	2.129



Assign “x” to “Dependent variable”. Now, we have to enter the population mean – 67.5 – under “Test value” (left column, middle, under “Hypothesis”). Let’s briefly have a look on the output we got a  $t(19) = 2.255$  (meaning a t value that was based on a sample with 19 degrees of freedom),  $p = 0.036$  (which is significant using  $\alpha = 0.05$ ).

**One Sample T-Test**

One Sample T-Test

	statistic	df	p	Mean difference	95% Confidence Interval		Cohen's d	
					Lower	Upper		
x	Student's t	2.255	19.000	0.0361	4.800	0.344	9.256	0.504

Note. H<sub>a</sub> population mean ≠ 67.5

Test of Normality (Shapiro-Wilk)

	W	p
x	0.962	0.5856

Note. A low p-value suggests a violation of the assumption of normality

Descriptives

	N	Mean	Median	SD	SE
x	20	72.300	75.000	9.521	2.129



The t-test has the assumption that the data we are using follow a normal distribution. That can be checked with ticking “Normality” (right column, middle, under “Assumption Tests”). The output has a quite high p-value (0.5856) which is not significant. This indicates that the assumption of normality is **not violated**: For Shapiro-Wilk, a low p-value would have indicated that the distribution in our sample [significantly] differs from a standard normal distribution. If the p-value is high, it indicates that our sample distribution is rather in accordance with a normal distribution. This doesn't give us reason for concern about normality.

**One Sample T-Test**

One Sample T-Test

	statistic	df	p	Mean difference	95% Confidence Interval		Cohen's d	
					Lower	Upper		
x	Student's t	2.255	19.000	0.0361	4.800	0.344	9.256	0.504

Note. H<sub>a</sub> population mean = 67.5

Test of Normality (Shapiro-Wilk)

	W	p
x	0.962	0.5856

Note. A low p-value suggests a violation of the assumption of normality

Descriptives

	N	Mean	Median	SD	SE
x	20	72.300	75.000	9.521	2.129



We can now ask for additional output by ticking “Mean difference”, “Confidence interval”, “Effect size” and “Descriptives”.

This gives us a mean difference of 4.8, a confidence interval of [0.344 9.256] and a Cohen’s d of 0.504. This means that we have a mean difference of 4.8 points, which with 95% probability falls within a confidence interval between 0.344 and 9.256.

That is, based upon what can be expected given the data in the current sample, we expect that (if we took the measurement again and again) in 95% the mean score of psychology students would be at least  $67.5 + 0.344 = 67.844$  and at maximum  $67.5 + 9.256 = 76.756$ .

Please note that these values are relative to the test score we entered, whereas it is common to report the confidence interval in square brackets, like  $CI_{95} = [0.344 \ 9.256]$ . Finally, the Cohen’s d (0.504) indicates a moderate effect size.

## One Sample T-Test

One Sample T-Test

Dependent Variables: x

Tests

- Student's
- Bayes factor
- Prior: 0.707
- Wilcoxon rank

Hypothesis

Test value: 67.5

- ≠ Test value
- > Test value
- < Test value

Missing values

- Exclude cases analysis by analysis
- Exclude cases listwise

Additional Statistics

- Mean difference
- Effect size
- Confidence interval
- Interval: 95 %
- Descriptives
- Descriptives plots

Assumption Checks

- Normality (Shapiro-Wilk)
- Normality (Q-Q plot)

## One Sample T-Test

One Sample T-Test

	statistic	df	p	Mean difference	95% Confidence Interval		Cohen's d	
					Lower	Upper		
x	Student's t	2.255	19.000	0.0361	4.800	0.344	9.256	0.504

Note. H<sub>a</sub> population mean ≠ 67.5

Test of Normality (Shapiro-Wilk)

	W	p
x	0.962	0.5856

Note. A low p-value suggests a violation of the assumption of normality

Descriptives

	N	Mean	Median	SD	SE
x	20	72.300	75.000	9.521	2.129



In a results section we could report this like:

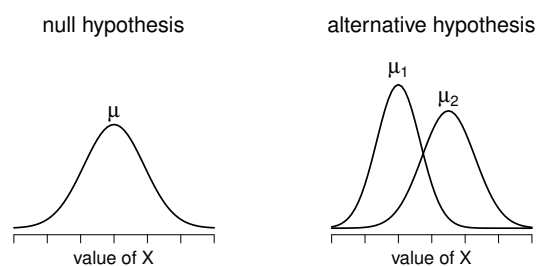
“With a mean grade of 72.3, the psychology students scored on average 4.80 points higher than the average grade of 67.5 ( $t(19) = 2.25$ ,  $p = 0.036$ ,  $CI_{95} = [0.344 \ 9.256]$ , Cohen's  $d = 0.504$ ).”





## Independent Samples t-test in jamovi

- most commonly used form of the t-test
- used to compare two groups of observations
- mathematically more complex (than One-Sample t-test)
- have to consider (pool) variances in two groups
- additional assumption: homogeneity of variances



The possibly most common example of the t-test is the Independent Samples t-test. This test is used to compare two samples / groups of observations. To say two sentences regarding the mathematics behind the t-test: Different from what we did in the One Sample t-test (where we had to deal only with one standard deviation, i.e., the square root of the variance,  $\sigma$ ). Now, comparing two groups, we have to deal with the variances from both groups which we have to combine (pool). The formulas are a little complex and can be found at the bottom of p. 255 and in the box in the middle of p. 257.



## Independent Samples t-test in jamovi

- ☰ → Open → Data Library → learning statistics with jamovi (lsj-data) → Harpo
- T-Tests → Independent Samples t-test  
assign “grade” to “Dependent variable” and  
”tutor” to “Grouping variable”  
tick “Normality (Shapiro-Wilk)” and “Equality of variances”  
tick “Welch’s”  
tick “Mean difference”, “Effect size”, “Confidence interval”  
and “Descriptives”



The data file used in the jamovi-book (Navarro & Foxcroft, 2022; p. 252 – 263) reports 33 test scores from a class with two different tutors. We now would like to know, who of them is the better tutor. In order to do this, open the harpo file. Use ☰ → Open → Data Library and choose harpo (hint: type ha in the search field).

Then you click on “T-tests” → “Independent Samples T-Test” and assign “grade” to “Dependent variable” and “tutor” to “Grouping variable”.

**Independent Samples T-Test**

Dependent Variables: grade

Grouping Variable: tutor

**Tests**

- Student's
- Bayes factor
- Prior: 0.707
- Welch's**
- Mann-Whitney U

**Additional Statistics**

- Mean difference
- Confidence interval 95 %
- Effect size
- Confidence interval 95 %
- Descriptives
- Descriptives plots

**Assumption Checks**

- Homogeneity test
- Normality test
- Q-Q plot

**Hypothesis**

- Group 1 ≠ Group 2
- Group 1 > Group 2
- Group 1 < Group 2

**Missing values**

- Exclude cases analysis by analysis
- Exclude cases listwise

**Independent Samples T-Test**

	statistic	df	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d	
						Lower	Upper		
grade	Student's t	2.115	31.000	0.0425	5.478	2.589	0.197	10.759	0.740
	Welch's t	2.034	23.025	0.0536	5.478	2.693	-0.092	11.048	0.740

**Assumptions**

Test of Normality (Shapiro-Wilk)

	W	p
grade	0.981	0.8266

Note. A low p-value suggests a violation of the assumption of normality

Test of Equality of Variances (Levene's)

	F	df	df2	p
grade	2.485	1	31	0.1251

Note. A low p-value suggests a violation of the assumption of equal variances

Group Descriptives

	Group	N	Mean	Median	SD	SE
grade	Anastasia	15	74.533	76.000	8.999	2.324
	Bernadette	18	69.056	69.000	5.775	1.361

We can see that there seems to be a difference between the two tutors: we obtain a t-value of 2.115 and a p of 0.0425 (just significant).

There are **three assumptions** for the Independent Samples t-test: **Normality**, **Independence**, and **Homogeneity of variance**.

The first can be checked by ticking “**Normality test**” under “Assumption Checks”. This reveals a p-value of 0.8266, indicating that there is no deviation from a normal distribution.

For assessing the second, **Independence**, we need to know a little more about the sample. We should be fine, as long as the measurements are collected from different participants and it is ensured that participants were not measured twice (e.g., by having to change the tutor in the middle of the semester).

**Independent Samples T-Test**

Dependent Variables: grade

Grouping Variable: tutor

**Tests**

- Student's
- Bayes factor
- Prior: 0.707
- Welch's**
- Mann-Whitney U

**Additional Statistics**

- Mean difference
- Confidence interval 95 %
- Effect size
- Confidence interval 95 %
- Descriptives
- Descriptives plots

**Assumption Checks**

- Homogeneity test
- Normality test
- Q-Q plot

**Hypothesis**

- Group 1 ≠ Group 2
- Group 1 > Group 2
- Group 1 < Group 2

**Missing values**

- Exclude cases analysis by analysis
- Exclude cases listwise

**Independent Samples T-Test**

		statistic	df	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d
							Lower	Upper	
grade	Student's t	2.115	31.000	0.0425	5.478	2.589	0.197	10.759	0.740
	Welch's t	2.034	23.025	0.0536	5.478	2.693	-0.092	11.048	0.740

**Assumptions**

Test of Normality (Shapiro-Wilk)

	W	p
grade	0.981	0.8266

Note. A low p-value suggests a violation of the assumption of normality

Test of Equality of Variances (Levene's)

	F	df	df2	p
grade	2.485	1	31	0.1251

Note. A low p-value suggests a violation of the assumption of equal variances

Group Descriptives

	Group	N	Mean	Median	SD	SE
grade	Anastasia	15	74.533	76.000	8.999	2.324
	Bernadette	18	69.056	69.000	5.775	1.361

For the third **Homogeneity of variances**, the fact that we have to deal with variances from two different groups also means that we have to ensure that the variances in the two groups are as equal as possible.

In order to check that, we tick “Homogeneity test” under “Assumption Checks”. A new table appears showing the results from Levene's Test of Equality of Variances. The p-value is not significant ( $p = 0.1251$ ) indicating that the variances and the two groups are not significantly different.

**Independent Samples T-Test**

Dependent Variables: grade

Grouping Variable: tutor

**Tests**

- Student's
- Bayes factor
- Prior: 0.707
- Welch's**
- Mann-Whitney U

**Additional Statistics**

- Mean difference
- Confidence interval: 95 %
- Effect size
- Confidence interval: 95 %
- Descriptives
- Descriptives plots

**Assumption Checks**

- Homogeneity test
- Normality test
- Q-Q plot

**Independent Samples T-Test**

	Statistic	df	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d	
						Lower	Upper		
grade	Student's t	2.115	31.000	0.0425	5.478	2.589	0.197	10.759	0.740
	Welch's t	2.034	23.025	0.0536	5.478	2.693	-0.092	11.048	0.740

**Assumptions**

Test of Normality (Shapiro-Wilk)

	W	p
grade	0.981	0.8266

Note. A low p-value suggests a violation of the assumption of normality.

Test of Equality of Variances (Levene's)

	F	df	df2	p
grade	2.485	1	31	0.1251

Note. A low p-value suggests a violation of the assumption of equal variances.

Group Descriptives

	Group	N	Mean	Median	SD	SE
grade	Anastasia	15	74.533	76.000	8.999	2.324
	Bernadette	18	69.056	69.000	5.775	1.361

However, that p-value is relatively low. To be on the safe side, we may decide to use Welch's instead of (or in addition to Student's). Just tick the box with "Welch's" (under "Tests") to switch it on. Welch's is another type of t-test that applies a correction for Inhomogeneity of variances. In R, using Welch's is even the default.

After applying the correction, the difference between the tutors isn't significant any more. This also nicely illustrates how arbitrary the threshold  $\alpha = 0.05$  is. And the threshold is not missed by a large margin, it is just a little above ( $p = 0.0536$ ).

We are now in a bit of a dilemma. The Homogeneity of variances is given ( $p = 0.1251$  is clearly above the threshold of  $\alpha = 0.05$ ), and the p-value for the group difference based on Student's t is just significant. Yet, "Welch's" is the more conservative (i.e. safer) measure.

**Independent Samples T-Test**

Dependent Variables: grade

Grouping Variable: tutor

**Tests**

- Student's
- Bayes factor
- Prior: 0.707
- Welch's
- Mann-Whitney U

**Additional Statistics**

- Mean difference
- Confidence interval: 95 %
- Effect size
- Confidence interval: 95 %
- Descriptives
- Descriptives plots

**Assumption Checks**

- Homogeneity test
- Normality test
- Q-Q plot

**Hypothesis**

- Group 1 ≠ Group 2
- Group 1 > Group 2
- Group 1 < Group 2

**Missing values**

- Exclude cases analysis by analysis
- Exclude cases listwise

**Independent Samples T-Test**

	Statistic	df	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d	
						Lower	Upper		
grade	Student's t	2.115	31.000	0.0425	5.478	2.589	0.197	10.759	0.740
	Welch's t	2.034	23.025	0.0536	5.478	2.693	-0.092	11.048	0.740

**Assumptions**

Test of Normality (Shapiro-Wilk)

	W	p
grade	0.981	0.8266

Note. A low p-value suggests a violation of the assumption of normality

Test of Equality of Variances (Levene's)

	F	df	df2	p
grade	2.485	1	31	0.1251

Note. A low p-value suggests a violation of the assumption of equal variances

Group Descriptives

	Group	N	Mean	Median	SD	SE
grade	Anastasia	15	74.533	76.000	8.999	2.324
	Bernadette	18	69.056	69.000	5.775	1.361

In the hope for making a more informed decision, we can continue to ask for further information, ticking “Mean difference” (and its “Confidence interval”), “Effect size”, and “Descriptives” (all under “Additional statistics”).

Turning to the table “Descriptive Statistics” to better understand how the two groups of students differ from each other, we see that students in Anastasia’s group got higher test scores compared to the students in Bernadette’s group. The standard deviations in Anastasia’s group (8.999) are about 1.6-times as high as those in Bernadette’s group (5.775). Since the variance is the standard deviation squared, the difference in variances between groups is about 2.5. That contributes to a better understanding of the results we got when correcting for unequal variances using Welch’s.

**Independent Samples T-Test**

Dependent Variables: grade

Grouping Variable: tutor

**Tests**

- Student's
- Bayes factor
- Prior: 0.707
- Welch's
- Mann-Whitney U

**Additional Statistics**

- Mean difference
- Confidence interval: 95 %
- Effect size
- Confidence interval: 95 %
- Descriptives
- Descriptives plots

**Assumption Checks**

- Homogeneity test
- Normality test
- Q-Q plot

**Independent Samples T-Test**

	Statistic	df	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d	
						Lower	Upper		
grade	Student's t	2.115	31.000	0.0425	5.478	2.589	0.197	10.759	0.740
	Welch's t	2.034	23.025	0.0536	5.478	2.693	-0.092	11.048	0.740

**Assumptions**

Test of Normality (Shapiro-Wilk)

	W	p
grade	0.981	0.8266

Note. A low p-value suggests a violation of the assumption of normality.

Test of Equality of Variances (Levene's)

	F	df	df2	p
grade	2.485	1	31	0.1251

Note. A low p-value suggests a violation of the assumption of equal variances.

Group Descriptives

	Group	N	Mean	Median	SD	SE
grade	Anastasia	15	74.533	76.000	8.999	2.324
	Bernadette	18	69.056	69.000	5.775	1.361

We could now report something like “The students in Anastasia’s group had a group mean of 74.533 in their test scores which was higher than the test scores in Bernadette’s group where students got on average 69.056 points. Standard deviations in the two groups differed considerably and were about 1.6-times higher in Anastasia’s group (8.999) than in Bernadette’s group (5.775). However, the Levene’s test for Equality of Variances was not significant:  $F_{(1,31)} = 2.485$ ,  $p = 0.1251$ . We therefore chose Student’s t to compare the difference between the two groups:  $t_{(31)} = 2.115$ ,  $p = 0.0425$ ,  $CI_{95} = [0.197 \ 10.759]$ , Cohen’s  $d = 0.740$ .

It possibly might be warranted to mention in addition that if Welch’s t would have been used to account for inhomogeneous variances, the difference would have missed significance by a little margin ( $p = 0.0536$ ).



# Independent Samples t-test in jamovi

for those who are brave at heart:

- download Syntax\_t-test\_Independent.R from MittUiB → Files → Data sets → Examples4jamovi
- click “R” → “Rj editor”, remove the syntax in it (# summary) and copy-and-paste the downloaded file  
click the cogwheel, set “Output” to “show code and output” and press “Play”
- check syntax and output, esp. check the comments and try to compare the formulas with those in the book (page numbers are given in the comments)



For those of you who are in for a bit of self-tormenting, I created also an R-script to calculate the t-values and the degrees of freedom using Student's and Welch's approach.

However, I am afraid, it is a bit complex and maybe not too didactic. To some degree, it served me to understand the maths behind the tests better.

Those, who are brave enough can find the file under:  
MittUiB → Files → Data sets → Examples4jamovi  
→ Syntax\_t-test\_Independent.R





## Paired Samples t-test in jamovi

- two measurement within the same person, e.g., a measurement before and after the intervention
- mathematically more like the One Sample t-test (the difference between the variables is used)



There is one last category of t-tests, called Paired Samples t-test. It is used if we have two measurement within the same person, e.g., a measurement before and after the intervention.

This is called a repeated measures design and we can use a Paired Samples t-test for exploring changes between these two measurements over time. The test compares these two measurements (within one person) by assessing whether the difference ( $D_i = X_{i1} - X_{i2}$ ) is significant.

Given that we now only have to deal with the variation within one group, the calculations (shown on p. 265 and 266 in the jamovi-book; Navarro & Foxcroft, 2019) are very much like those for the One Sample t-test (i.e., relatively easy again, much less complicated as those for the Independent Samples t-test where we had to pool variances).



## Paired Samples t-test in jamovi

- ☰ → Open → Data Library → learning statistics with jamovi (lsj-data) → Chico
- T-Tests → Paired Samples t-test  
mark “grade\_test1” and “grade\_test2” (Shift + mouse click) and assign them to “Dependent variable” and “tutor” to “Grouping variable”  
tick “Normality (Shapiro-Wilk)” and “Equality of variances”  
tick “Welch’s”  
tick “Mean difference”, “Effect size”, “Confidence interval” and “Descriptives”



The data set that is used to demonstrate the test in the jamovi-book (p. 264) is about a teacher, Dr Chico, where the designation “evil lecturer” (that you might remember from chapter 1 in Andy Fields book: “Why is my evil lecturer forcing me to learn statistics?”) is rather appropriate. He runs a very hard class. What he does is to have two exams, one at the beginning, one at the end. The first test is supposed to be a bit of a “wake up call” for students. The idea is that after having got unsatisfactory results in the first test, students will put in effort and better prepare for the second test. We open the dataset using ☰ → “Open” → “Data Library” → “learning statistics with jamovi” (lsj-data) → “chico” (please note that there is also a chico2-dataset; don’t choose that).



Paired Samples T-Test

Paired Variables: grade\_test1, grade\_test2

1

Additional Statistics: Mean difference, Effect size, Confidence interval, Descriptives

3

Assumption Checks: Normality (Shapiro-Wilk)

2

Paired Samples T-Test

		statistic	df	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d
grade_test1	grade_test2	Student's t					Lower	Upper	
		-6.475	19.000	< .0001	-1.405	0.217	-1.859	-0.951	-1.448

Test of Normality (Shapiro-Wilk)

		W	p
grade_test1	grade_test2	0.966	0.6778

Note. A low p-value suggests a violation of the assumption of normality.

Descriptives

	N	Mean	Median	SD	SE
grade_test1	20	56.980	57.700	6.616	1.479
grade_test2	20	58.385	59.700	6.406	1.432



To see whether Dr. Chico's rationale (the first test is a wake-up call and the students prepare better for the second) is working and the increase from the first to the second test is statistically significant, we click on "T-Tests" → "Paired Samples T-Test" select both (click on both while holding the Shift or the Ctrl-key) and assign them to "Paired Variables". The difference is highly significant:  $t(19) = -6.475$ ,  $p < 0.001$ .

Strictly speaking, we are dealing with a directed hypothesis here: Dr Chico expected an increase in grade from the first to the second test, so we could change "Measure 1  $\neq$  Measure 2" into "Measure 1  $<$  Measure 2" (this, however, doesn't change anything since the p-value already is very low).



Paired Samples T-Test

Paired Variables: grade\_test1, grade\_test2

1

Tests:  Student's,  Bayes factor,  Wilcoxon rank

Hypothesis:  Measure 1 = Measure 2,  Measure 1 > Measure 2,  Measure 1 < Measure 2

Missing values:  Exclude cases analysis by analysis,  Exclude cases listwise

Additional Statistics:  Mean difference,  Effect size,  Confidence interval,  Descriptives

Assumption Checks:  Normality (Shapiro-Wilk),  Normality (Q-Q plot)

2

3

### Paired Samples T-Test

Paired Samples T-Test

		statistic	df	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d
grade_test1	grade_test2	Student's t					Lower	Upper	
		-6.475	19.000	< .0001	-1.405	0.217	-1.859	-0.951	-1.448

Test of Normality (Shapiro-Wilk)

		W	p
grade_test1	- grade_test2	0.966	0.6778

Note. A low p-value suggests a violation of the assumption of normality

Descriptives

	N	Mean	Median	SD	SE
grade_test1	20	56.980	57.700	6.616	1.479
grade_test2	20	58.385	59.700	6.406	1.432



Now, we tick “Normality (Shapiro-Wilk)” (under “Assumption Checks”). The p-value for Shapiro-Wilk is 0.6778 and we therefore have no reason to doubt that the data are normally distributed. We then tick “Mean difference”, “Effect size” and “Confidence interval” (under “Additional Statistics”). When looking at the “Mean difference”, we see that it is quite small, even though the test is so highly statistically significant and even though the effect size is also considerable (1.448, when a Cohen’s d of 0.8 is the threshold for a “large” effect size). Briefly checking the Standard Deviation for the two variables in the table “Descriptive statistics” confirms that there is no reason for concern either.



## Paired Samples T-Test

Paired Variables

grade\_test1 grade\_test2

1

**Tests**

Student's

Bayes factor

Prior: 0.707

Wilcoxon rank

**Hypothesis**

Measure 1 = Measure 2

Measure 1 > Measure 2

Measure 1 < Measure 2

**Missing values**

Exclude cases analysis by analysis

Exclude cases listwise

**Additional Statistics**

Mean difference

Effect size

Confidence interval

Interval: 95 %

Descriptives

Descriptives plots

**Assumption Checks**

Normality (Shapiro-Wilk)

Normality (Q-Q plot)

2

3

## Paired Samples T-Test

Paired Samples T-Test

		statistic	df	p	Mean difference	SE difference	95% Confidence Interval		Cohen's d
grade_test1	grade_test2	Student's t					Lower	Upper	
		-6.475	19.000	< .0001	-1.405	0.217	-1.859	-0.951	-1.448

## Test of Normality (Shapiro-Wilk)

		W	p
grade_test1	- grade_test2	0.966	0.6778

Note. A low p-value suggests a violation of the assumption of normality

## Descriptives

	N	Mean	Median	SD	SE
grade_test1	20	56.980	57.700	6.616	1.479
grade_test2	20	58.385	59.700	6.406	1.432



We could report these results with a sentence like:  
 “Students increased their performance from the first (M = 56.980, SD = 6.62) to the second test (M = 58.39, SD = 6.41). The difference between the tests (M = -1.405) was highly significant:  $t(19) = -6.475$ ,  $p < 0.001$ ,  $CI_{95} = [-1.86 -0.95]$ , Cohen’s  $d = -1.45$ .”



## Assumptions, non-parametric altern.

- t-tests have assumptions: all require normality, the Independent Samples t-test Equivalence of Variances
- easily check whether these assumptions are met with the tests under “Assumption Checks”
- if assumptions are not met, you can use non-parametric methods, e.g., Wilcoxon or Mann-Whitney U; these are the last entry under “Tests”

REFRESHER: STATISTICAL ANALYSES

SEBASTIAN...

Independent Samples T-Test

Dependent Variables: grade

Grouping Variable: tutor

**Tests**

Student's  
 Bayes factor  
 Prior: 0.707  
 Welch's  
 Mann-Whitney U

**Hypothesis**

Group 1 = Group 2  
 Group 1 > Group 2  
 Group 1 < Group 2

**Missing values**

Exclude cases analysis by analysis  
 Exclude cases listwise

**Additional Statistics**

Mean difference  
 Effect size  
 Confidence interval  
 Interval: 95 %  
 Descriptives  
 Descriptives plots

**Assumption Checks**

Normality (Shapiro-Wilk)  
 Normality (Q-Q plot)  
 Equality of variances

Finally, a couple of words regarding the assumptions of the different classes of t-tests and what we could do if one of these assumptions is violated.

Generally, and this is also one reason why I find jamovi so logic and didactic, has each of the tests options to check the necessary assumptions easily from within the user interface of the test itself:

There is always a header (or for the more complex tests such as ANOVAs it is a drop-down-menu) where you have all necessary assumption tests in one place.



## Assumptions, non-parametric altern.

- t-tests have assumptions: all require normality, the Independent Samples t-test Equivalence of Variances
- easily check whether these assumptions are met with the tests under “Assumption Checks”
- if assumptions are not met, you can use non-parametric methods, e.g., Wilcoxon or Mann-Whitney U; these are the last entry under “Tests”

REFRESHER: STATISTICAL ANALYSES

SEBASTIAN.

Independent Samples T-Test

Dependent Variables: grade

Grouping Variable: tutor

Tests:

- Student's
- Bayes factor (Prior: 0.707)
- Welch's
- Mann-Whitney U

Hypothesis:

- Group 1 = Group 2
- Group 1 > Group 2
- Group 1 < Group 2

Missing values:

- Exclude cases analysis by analysis
- Exclude cases listwise

Additional Statistics:

- Mean difference
- Effect size
- Confidence interval (Interval: 95 %)
- Descriptives
- Descriptives plots

Assumption Checks:

- Normality (Shapiro-Wilk)
- Normality (Q-Q plot)
- Equality of variances

For One Sample and Paired Sample t-tests, the two assumptions we have to assess are: (1) Independence and (2) Normality.

As I mentioned before, “Independence” asks whether we can assume that each observation is a completely random sample from the population that you’re interested in. That assumption can’t be tested but that you have to consider it when designing your experiments and collecting your data. For the datasets we are dealing with (where you haven’t been involved in design and data collected), we just have to assume that the assumption isn’t violated.



## Assumptions, non-parametric altern.

- t-tests have assumptions: all require normality, the Independent Samples t-test Equivalence of Variances
- easily check whether these assumptions are met with the tests under “Assumption Checks”
- if assumptions are not met, you can use non-parametric methods, e.g., Wilcoxon or Mann-Whitney U; these are the last entry under “Tests”

REFRESHER: STATISTICAL ANALYSES

SEBASTIAN.

Independent Samples T-Test

Dependent Variables: grade

Grouping Variable: tutor

Tests:

- Student's
- Bayes factor (Prior: 0.707)
- Welch's
- Mann-Whitney U

Hypothesis:

- Group 1 = Group 2
- Group 1 > Group 2
- Group 1 < Group 2

Missing values:

- Exclude cases analysis by analysis
- Exclude cases listwise

Additional Statistics:

- Mean difference
- Effect size
- Confidence interval (Interval: 95 %)
- Descriptives
- Descriptives plots

Assumption Checks:

- Normality (Shapiro-Wilk)
- Normality (Q-Q plot)
- Equality of variances

Normality can easily be assessed by ticking “Normality (Shapiro-Wilk)” and assessing the p-values in the “Test of Normality”-table. If those are significant or close to significance ( $p < 0.1$ ), it is possible safer to consider non-parametric analyses (described on the next slide).

For the Independent Samples t-test, we have an additional assumption “Homogeneity of Variances”. As I said, when demonstrating the test before, you assess that by ticking “Equality of Variances” (also under “Assumption Checks”) and depending on whether the p-value in the “Test of Equality of Variances (Levene’s)”-table is significant (or close to;  $p < 0.1$ ), decide to untick “Student’s t” and to tick “Welch’s t” (under “Tests”) which is correcting for unequal variances.





## Assumptions, non-parametric altern.

- t-tests have assumptions: all require normality, the Independent Samples t-test Equivalence of Variances
- easily check whether these assumptions are met with the tests under “Assumption Checks”
- if assumptions are not met, you can use non-parametric methods, e.g., Wilcoxon or Mann-Whitney U; these are the last entry under “Tests”

REFRESHER: STATISTICAL ANALYSES

SEBASTIAN...

Independent Samples T-Test

Dependent Variables: grade

Grouping Variable: tutor

Tests

Student's t

Bayes factor

Prior: 0.707

Welch's t

Mann-Whitney U

Hypothesis

Group 1 = Group 2

Group 1 > Group 2

Group 1 < Group 2

Missing values

Exclude cases analysis by analysis

Exclude cases listwise

Additional Statistics

Mean difference

Effect size

Confidence interval

Interval: 95 %

Descriptives

Descriptives plots

Assumption Checks

Normality (Shapiro-Wilk)

Normality (Q-Q plot)

Equality of variances

Especially, if the assumption of normality is violated, it is recommended to use so-called non-parametric methods or analyses.

For the One Sample and the Paired Samples t-test you would have to tick “Wilcoxon rank” (under “Tests”), for the Independent Samples t-test you have to tick “Mann-Whitney U”. In that case, it is possibly wise to untick the parametric analyses (“Student’s t” and “Welch’s t”) to keep the output clean and to be less confusing.



## Assumptions, non-parametric altern.

- t-tests have assumptions: all require normality, the Independent Samples t-test Equivalence of Variances
- easily check whether these assumptions are met with the tests under “Assumption Checks”
- if assumptions are not met, you can use non-parametric methods, e.g., Wilcoxon or Mann-Whitney U; these are the last entry under “Tests”

REFRESHER: STATISTICAL ANALYSES

SEBASTIAN...

Independent Samples T-Test

Dependent Variables: grade

Grouping Variable: tutor

Tests

Student's

Bayes factor

Prior: 0.707

Welch's

Mann-Whitney U

Hypothesis

Group 1 = Group 2

Group 1 > Group 2

Group 1 < Group 2

Missing values

Exclude cases analysis by analysis

Exclude cases listwise

Additional Statistics

Mean difference

Effect size

Confidence interval

Interval: 95 %

Descriptives

Descriptives plots

Assumption Checks

Normality (Shapiro-Wilk)

Normality (Q-Q plot)

Equality of variances

One advantage of non-parametric methods is that you can use them without having to consider whether certain assumptions are met.

That is, you can always use non-parametric methods (Wilcoxon, Mann-Whitney, etc.) to analyse data that in principle also would have allowed to use parametric methods (t-tests, etc.).

The reason, why we not just use non-parametric methods for all analyses are: (1) they usually are more conservative and less powerful than their parametric equivalents (it is more difficult to get significant results using non-parametric methods); and (2) you typically can explore more complex statistical models using parametric methods (e.g., it is not possible to explore two or more factors using non-parametric methods, whereas you can easily do this using an ANOVA).



# Continuous predictor and continuous outcome variables (Ch. 12)

If we are interested in exploring relationships between continuous variables, the easiest way of doing this is calculating a correlation between these variables. There is another way, which is visualizing the two variables in a scatter plot. You can select this visualization (under “Plot”) when using the “Correlation Matrix”-analysis in jamovi.



## Correlations: step-by step

- correlations can be used to explore relationships between continuous variables
- correlation coefficients ( $r$ ) can range from -1 to 1
  - $0 < r < 1$  → positive correlation
  - $-1 < r < 0$  → negative correlation
- calculated by multiplying the position of a person (relative to the mean) in the first variable and the position (relative to the mean) in the second variable



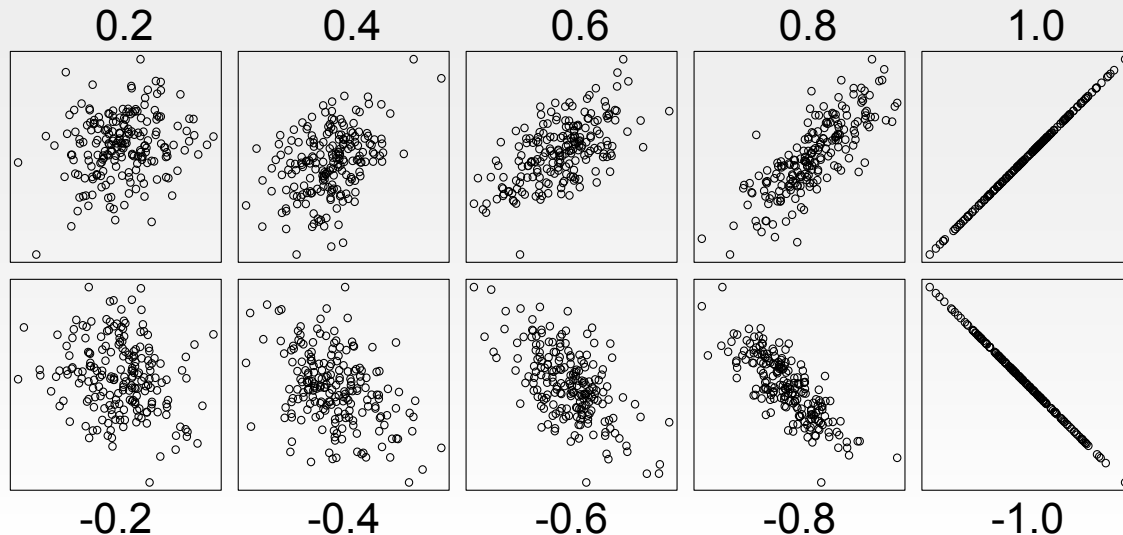
A correlation coefficient (i.e., the statistics we used to describe the strength of such a relation between variables) can take values between -1 and 1.

If we have a correlation coefficient between 0 and 1 we speak of a positive correlation. In that case, the values in the two variables change in the same direction, i.e., one person who has a variable value larger than the mean in the first variable is expected to have a variable value above the mean in the second variable too.

If we have a correlation coefficient between 0 and -1 we speak of a negative correlation. In that case, the values in the two variables change in opposite direction, i.e., one person who has a variable value larger than the mean in the first variable is expected to have a variable value below the mean in the second variable.



## Correlations: step-by step



REFRESHER: STATISTICAL ANALYSES

SEBASTIAN.JENTSCHKE@UIB.NO

SLIDE 69



We can use this logic to calculate the correlation by multiplying the position of a person relative to the mean in the first variable and the position relative to the mean in the second variable. The product of this multiplication is then “standardized” by (i.e., divided) the standard deviations for the first ( $\sigma_X$ ) and the second variable ( $\sigma_Y$ ) which are multiplied.

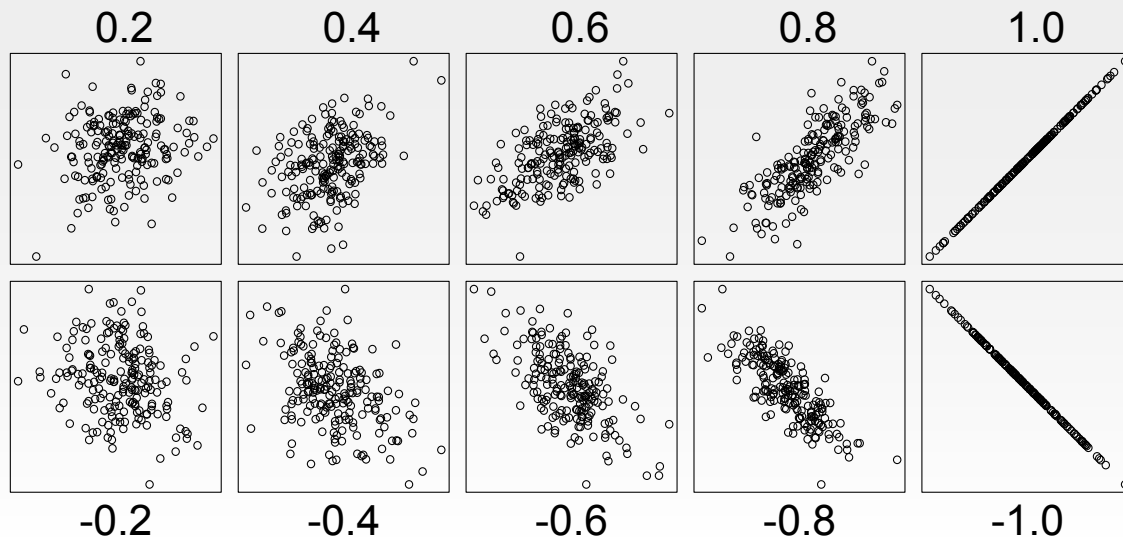
The whole formula looks like this: 
$$r_{XY} = \frac{\text{Cov}(X, Y)}{\hat{\sigma}_X \cdot \hat{\sigma}_Y}$$

The part above the fraction bar is called covariance (as the name indicates, it expresses common variation in the two variables). It is calculated with this formula:

$$\text{Cov}(X, Y) = \frac{1}{N-1} \cdot \sum_{i=1}^N (X_i - \bar{X}) \cdot (Y_i - \bar{Y})$$



## Correlations: step-by step



REFRESHER: STATISTICAL ANALYSES

SEBASTIAN.JENTSCHKE@UIB.NO

SLIDE 70



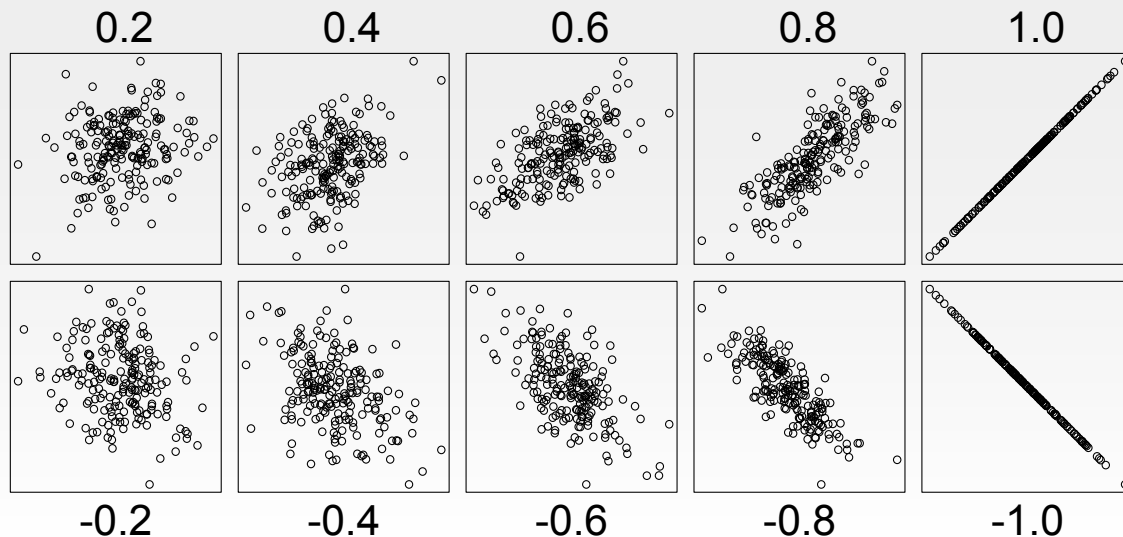
$$\text{Cov}(X, Y) = \frac{1}{N-1} \cdot \sum_{i=1}^N (X_i - \bar{X}) \cdot (Y_i - \bar{Y}) \quad (\text{same formula as on the previous slide})$$

If a person has a value above the mean in both variables the product of  $(X_i - \bar{X}) \cdot (Y_i - \bar{Y})$  is positive, the same happens if a person has a value below the mean in both variables (if we multiply two negative numbers, the product is positive). If not, i.e., if a person has a value above the mean in one and below the mean in the other variable, the product is negative. These products are summed up per individuum.

If all (or the majority of) products are positive and point in the same direction we will get a large positive correlation (top row in the figure). If all (or the majority of) products are negative and point in opposite directions we will get a large negative correlation (bottom row in the figure). If some products are positive and some are negative, the correlation will be relatively weak (left in the figure).



## Correlations: step-by step



REFRESHER: STATISTICAL ANALYSES

SEBASTIAN.JENTSCHKE@UIB.NO

SLIDE 71



$$r_{XY} = \frac{\text{Cov}(X, Y)}{\hat{\sigma}_X \cdot \hat{\sigma}_Y} \quad (\text{same formula as two slides before})$$

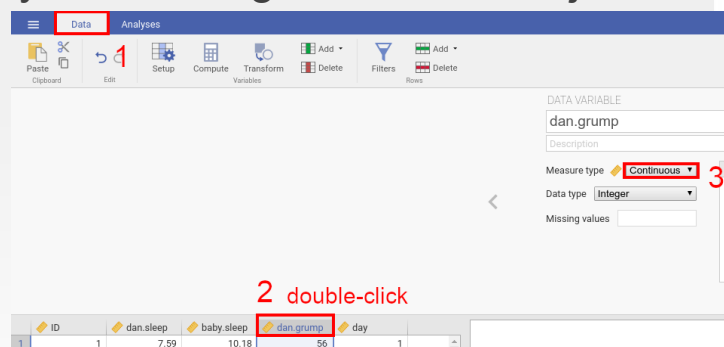
As I mentioned is the covariance – Cov(X, Y) expressing to what degree a group of participants had the same tendency in their values on two variables). It is divided by product of the standard deviation within the first ( $\sigma_X$ ) and the second variable ( $\sigma_Y$ ). This product represents the “natural” variation within each of the variables.

By doing so, i.e. by dividing the covariance and the product of the standard deviations, the correlation always falls in the range of -1 to 1.



## Correlations in jamovi

- ☰ → Open → Data Library → learning statistics with jamovi (lsj-data) → Parenthood
- change measurement type for dan.grump to “Continuous” (and ID to ID)
- Regression → Correlation Matrix  
assign all variables except “ID” to the variable box



The dataset that we will use now to demonstrate correlations is called **Parenthood** and can be opened using ☰ → Open → “Data Library” → “learning statistics with jamovi” (lsj-data).

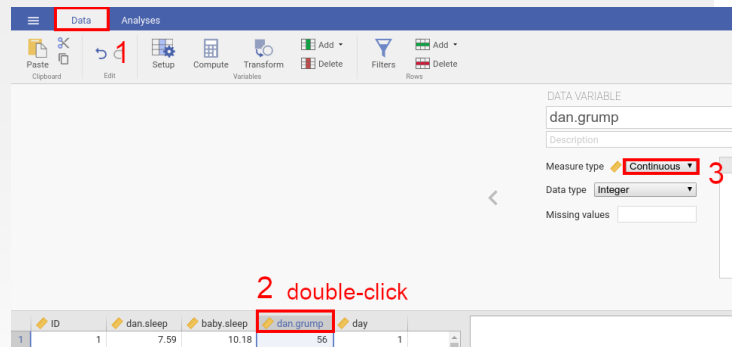
The dataset contains five columns of which we will use four: dan.sleep, baby.sleep, dan.grump, and day. Dan is parent to a baby and registered for every day (day), how long the baby slept (baby.sleep), how long Dan slept (dan.sleep) and how grumpy Dan was on that day (dan.grump). We are now interested in how these variables are related. Is it, e.g., reasonable to assume that the older the baby gets, the more Dan is going to get sleep? Or is it reasonable to assume that Dan’s grumpiness increases as the sleeps Dan gets decreases?





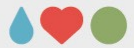
## Correlations in jamovi

- ☰ → Open → Data Library → learning statistics with jamovi (lsj-data) → Parenthood
- change measurement type for dan.grump to “Continuous” (and ID to ID)
- Regression → Correlation Matrix  
assign all variables except “ID” to the variable box



When we look at the variables, then we see that jamovi has assigned the measurement level “Ordinal” to dan.grump that we should change into “Continuous”. To do this, we click on dan.grump in the top row with the variable names in the spreadsheet and then change the “Measurement type”. If we don’t do that, the calculations are still correct. However, the variables are re-arranged (with the originally ordinal one dan.grump as the last row / column) which may lead to confusion when comparing your outputs with the ones on the slides.

In order to explore these relations using jamovi, we use the icon “Regression” → “Correlation Matrix”. We assign dan.sleep, baby.sleep, dan.grump, and day to the box for the variables.



# Correlations in jamovi

Correlation Matrix

ID

- dan.sleep
- baby.sleep
- dan.grump
- day

**Correlation Coefficients**

Pearson  
 Spearman  
 Kendall's tau-b

**Additional Options**

Report significance  
 Flag significant correlations  
 N  
 Confidence intervals  
 Interval: 95 %

**Hypothesis**

Correlated  
 Correlated positively  
 Correlated negatively

**Plot**

Correlation matrix  
 Densities for variables  
 Statistics

## Correlation Matrix

		dan.sleep	baby.sleep	dan.grump	day
dan.sleep	Pearson's r	–			
	p-value	–			
baby.sleep	Pearson's r	0.628	–		
	p-value	< .0001	–		
dan.grump	Pearson's r	-0.903	-0.566	–	
	p-value	< .0001	< .0001	–	
day	Pearson's r	-0.098	-0.010	0.076	–
	p-value	0.3300	0.9179	0.4495	–

SEBASTIAN.JENTSCHKE@UIB.NO

SLIDE 74



Contrary to the expectation, none of the correlations with age of the baby (day) is significant (grey box). If we look at the three other combinations (green box), then we see that all of them are highly significant. The correlation between how much sleep Dan got and Dan's grumpiness is the one with the highest absolute value for the correlation coefficient ( $r = -0.903$ ,  $p < 0.001$ ), i.e. the strongest negative correlation (very close to 1 and hence almost perfect). The second highest correlation is the one between how much the baby and Dan slept ( $r = 0.628$ ,  $p < 0.001$ ). In this case we got a positive correlation coefficient, indicating that the more the baby slept, the more slept Dan. Be careful, even though it seems quite logical (based on our everyday knowledge) that how much the baby slept caused how much Dan slept, but strictly speaking we can't make such claims based upon a correlation.



# Correlations in jamovi

Correlation Matrix

ID

- dan.sleep
- baby.sleep
- dan.grump
- day

**Correlation Coefficients**

Pearson  
 Spearman  
 Kendall's tau-b

**Additional Options**

Report significance  
 Flag significant correlations  
 N  
 Confidence intervals  
 Interval: 95 %

**Hypothesis**

Correlated  
 Correlated positively  
 Correlated negatively

**Plot**

Correlation matrix  
 Densities for variables  
 Statistics

## Correlation Matrix

Correlation Matrix

		dan.sleep	baby.sleep	dan.grump	day
dan.sleep	Pearson's r	–			
	p-value	–			
baby.sleep	Pearson's r	0.628	–		
	p-value	< .0001	–		
dan.grump	Pearson's r	-0.903	-0.566	–	
	p-value	< .0001	< .0001	–	
day	Pearson's r	-0.098	-0.010	0.076	–
	p-value	0.3300	0.9179	0.4495	–

SEBASTIAN.JENTSCHKE@UIB.NO

SLIDE 75



Finally, we have the negative correlation between how much the baby slept and Dan's grumpiness ( $r = -0.566$ ,  $p < 0.001$ ). Given that size and direction of this correlation is about that of the product of the two others ( $r = -0.903 \cdot 0.628 = -0.567 \sim -0.566$ ) we might be very tempted to speculate that the direction of effect goes: babies sleep  $\rightarrow$  Dan's sleep  $\rightarrow$  Dan's grumpiness. However, again the warning that correlations don't allow to assess (and even less to claim) certain causal relations.



# Non-parametric correl. in jamovi

Correlation Matrix

→
 

- dan.sleep
- baby.sleep
- dan.grump
- day

**Correlation Coefficients**

Pearson

Spearman

Kendall's tau-b

**Additional Options**

Report significance

Flag significant correlations

N

Confidence intervals

Interval  %

**Hypothesis**

Correlated

Correlated positively

Correlated negatively

**Plot**

Correlation matrix

Densities for variables

Statistics

Correlation Matrix

		dan.sleep	baby.sleep	dan.grump	day
baby.sleep	Pearson's r	0.628	–		
	p-value	< .0001	–		
	Spearman's rho	0.645	–		
dan.grump	p-value	< .0001	–		
	Kendall's Tau B	0.455	–		
	p-value	< .0001	–		
dan.grump	Pearson's r	-0.903	-0.566	–	
	p-value	< .0001	< .0001	–	
	Spearman's rho	-0.887	-0.587	–	
day	p-value	< .0001	< .0001	–	
	Kendall's Tau B	-0.732	-0.418	–	
	p-value	< .0001	< .0001	–	
day	Pearson's r	-0.098	-0.010	0.076	–
	p-value	0.3300	0.9179	0.4495	–
	Spearman's rho	-0.089	0.003	0.056	–
	p-value	0.3787	0.9770	0.5823	–
	Kendall's Tau B	-0.061	-0.005	0.042	–
p-value	0.3653	0.9383	0.5371	–	

In addition to Pearson's  $r$ , jamovi allows us also to calculate two non-parametric correlations: Spearman does something very similar to Pearson's correlation coefficient. However, whereas the parametric correlation coefficient (Pearson's) uses the distance of individuals from the mean of each variable in the correlation, the non-parametric correlation (Spearman) compares the ranks each person has within the two variables.



# Non-parametric correl. in jamovi

Correlation Matrix

→
 

- dan.sleep
- baby.sleep
- dan.grump
- day

**Correlation Coefficients**

Pearson

Spearman

Kendall's tau-b

**Additional Options**

Report significance

Flag significant correlations

N

Confidence intervals

Interval  %

**Hypothesis**

Correlated

Correlated positively

Correlated negatively

**Plot**

Correlation matrix

Densities for variables

Statistics

Correlation Matrix

		dan.sleep	baby.sleep	dan.grump	day
baby.sleep	Pearson's r	0.628	–		
	p-value	< .0001	–		
	Spearman's rho	0.645	–		
dan.grump	Pearson's r	-0.903	-0.566	–	
	p-value	< .0001	< .0001	–	
	Spearman's rho	-0.887	-0.587	–	
day	Pearson's r	-0.098	-0.010	0.076	–
	p-value	0.3300	0.9179	0.4495	–
	Spearman's rho	-0.089	0.003	0.056	–
	p-value	0.3787	0.9770	0.5823	–
	Kendall's Tau B	-0.061	-0.005	0.042	–
	p-value	0.3653	0.9383	0.5371	–

This is done by sorting each of the two variables and then assigning a rank depending on which position in the list that individual achieved. Afterwards, the same formula as for the Pearson correlation coefficient above is applied (only this time not calculating the statistics for the variable values but for the ranks these values represent).

If all individuals have the same (or a very similar) rank on both variables, the correlation coefficient is positive.

If they ranks are opposite (i.e., being among the first on one variable and rather at the back on the other), the correlation is negative.

If the rank in one variable doesn't really bear any predictive power for saying something about the rank in the other variable, the correlation is low (close to zero).



# Non-parametric correl. in jamovi

Correlation Matrix

→
 

- dan.sleep
- baby.sleep
- dan.grump
- day

**Correlation Coefficients**

Pearson

Spearman

Kendall's tau-b

**Additional Options**

Report significance

Flag significant correlations

N

Confidence intervals

Interval  %

**Hypothesis**

Correlated

Correlated positively

Correlated negatively

**Plot**

Correlation matrix

Densities for variables

Statistics

Correlation Matrix

		dan.sleep	baby.sleep	dan.grump	day
baby.sleep	Pearson's r	0.628	–		
	p-value	< .0001	–		
	Spearman's rho	0.645	–		
dan.grump	Pearson's r	-0.903	-0.566	–	
	p-value	< .0001	< .0001	–	
	Spearman's rho	-0.887	-0.587	–	
day	Pearson's r	-0.098	-0.010	0.076	–
	p-value	0.3300	0.9179	0.4495	–
	Spearman's rho	-0.089	0.003	0.056	–
	p-value	0.3787	0.9770	0.5823	–
	Kendall's Tau B	-0.061	-0.005	0.042	–
	p-value	0.3653	0.9383	0.5371	–

When taking the parenthood dataset, we observe the interesting fact that the Spearman's rho correlation coefficients have a very similar magnitude to the parametric Pearson's coefficients. Sometimes the Spearman coefficients are even numerically larger: dan.sleep – baby.sleep and dan.grump – baby.sleep. However, in most cases, non-parametric measures tend to lower coefficients and lower p-values.



# Non-parametric correl. in jamovi

Correlation Matrix

ID  →

- dan.sleep
- baby.sleep
- dan.grump
- day

**Correlation Coefficients**

Pearson

Spearman

Kendall's tau-b

**Additional Options**

Report significance

Flag significant correlations

N

Confidence intervals

Interval  %

**Hypothesis**

Correlated

Correlated positively

Correlated negatively

**Plot**

Correlation matrix

Densities for variables

Statistics

Correlation Matrix

		dan.sleep	baby.sleep	dan.grump	day
baby.sleep	Pearson's r	0.628	–		
	p-value	< .0001	–		
	Spearman's rho	0.645	–		
dan.grump	Pearson's r	-0.903	-0.566	–	
	p-value	< .0001	< .0001	–	
	Spearman's rho	-0.887	-0.587	–	
day	Pearson's r	-0.098	-0.010	0.076	–
	p-value	0.3300	0.9179	0.4495	–
	Spearman's rho	-0.089	0.003	0.056	–
	p-value	0.3787	0.9770	0.5823	–
	Kendall's Tau B	-0.061	-0.005	0.042	–
	p-value	0.3653	0.9383	0.5371	–

Finally, we have a measure called Kendall's tau b that also operates on ranks but compares how many individuals got the same rank on both variables and for how many the rank was different. Using Kendall's tau b, we typically get much lower correlation coefficients (even though all correlations are still statistically highly significant;  $p < 0.001$  for all).



## Where corr. fail: Anscombe's quartet

- open the file using ☰ → Open → Data library (→ learning statistics with jamovi) → Anscombe
- select “Exploration” → “Descriptives”  
assign all variables (X1 to X4, Y1 to Y4) to “Variables”  
tick “Shapiro-Wilk”



For another word of caution regarding correlations and for demonstrating what effect outliers in your data can cause, we have to turn to a famous dataset called Anscombe's quartet.

You can open the file using ☰ → Open → Data library (→ learning statistics with jamovi) → Anscombe.





## Where corr. fail: Anscombe's quartet

- Anscombe's quartet are four variable pairs (X1–Y1, ... X4–Y4) with nearly identical descriptive statistics (incl. correlation coeff.), yet very different distributions

Descriptives								
	X1	X2	X3	X4	Y1	Y2	Y3	Y4
N	11	11	11	11	11	11	11	11
Mean	9.000	9.000	9.000	9.000	7.501	7.501	7.500	7.501
Median	9	9	9	8	7.580	8.140	7.110	7.040
Standard deviation	3.317	3.317	3.317	3.317	2.032	2.032	2.030	2.031
Minimum	4	4	4	8	4.260	3.100	5.390	5.250
Maximum	14	14	14	19	10.840	9.260	12.740	12.500
Shapiro-Wilk W	0.968	0.968	0.968	0.345	0.977	0.828	0.834	0.875
Shapiro-Wilk p	0.8698	0.8698	0.8698	< .0001	0.9467	0.0222	0.0260	0.0908



First, let's have a look at the Descriptive statistics.

Use the icon "Exploration" → Descriptives and assign X1 to X4, and Y1 to Y4 to "Variables". Then open the Drop-down-menu "Statistics", select "Std. deviation" (under "Dispersion") and "Shapiro-Wilk" (under Normality).

You will see a table as the one shown above. As you can see, the Mean is the same within all X-variables (9.000) and within all Y-variables (7.501). The same applies for the Standard deviation which is also identical within X- (3.317) and Y-variables (2.032). Within the X variables, most ( $\frac{3}{4}$ ) also are in accordance with a normal distribution.



## Where corr. fail: Anscombe's quartet

- Regression → Correlation Matrix  
assign “X1” and “Y1” to the variable box (remember, the variables come in pairs, so X1 → Y1, X2 → Y2, etc.)  
tick “Correlation matrix” (under “Plot”)
- right-click on the analysis in the output window (e.g., the table with the correlations), select Analysis → Duplicate, repeat this three times (you need four analyses altogether)
- go to the second and replace X1 and Y1 with X2 and Y2, and repeat this for the third (X3, Y3) and fourth (X4, Y4)



Select “Regression” → “Correlation Matrix” and start with assigning the first pair (X1, Y1) to the variable box.

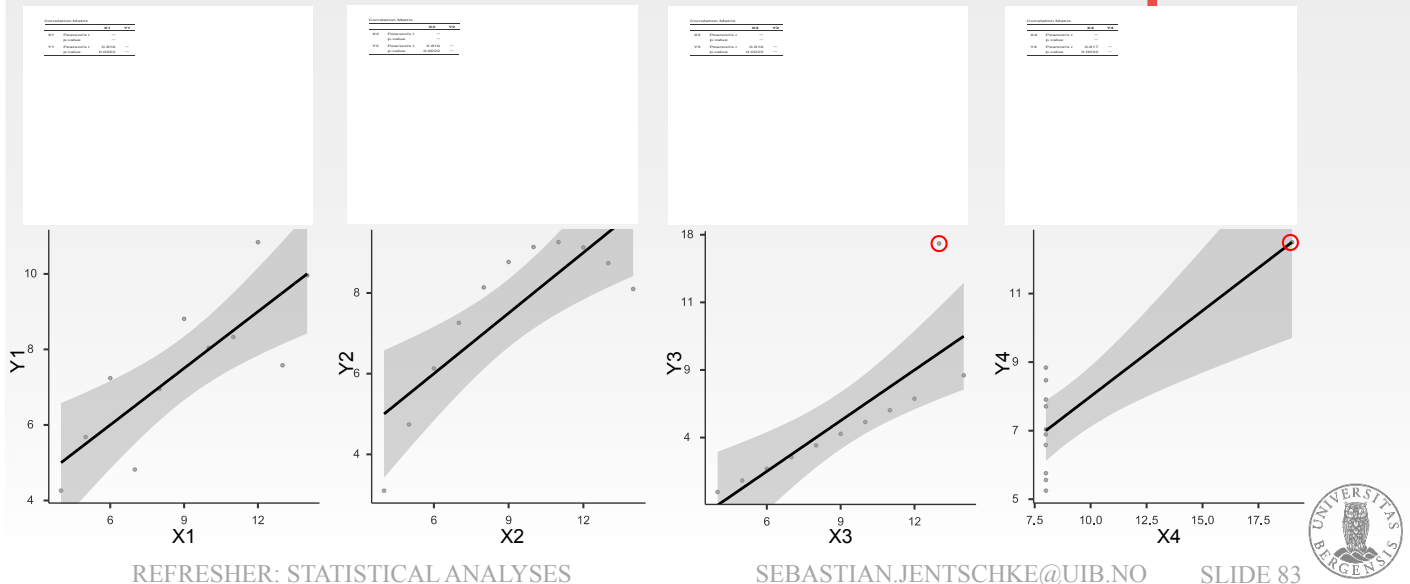
Ensure that X1 comes first in the variable list (before Y1; this doesn't make a difference for the correlation coefficient but for how the plots look like).

You obtain a correlation coefficient of  $r = 0.816$  which is significant  $p = 0.002$ . Now, switch on “Correlation matrix” (under “Plots”).

Right-click somewhere in that output, select “Analysis” → “Duplicate” and repeat that another two times (so that you have four copies altogether). Now, you click onto the second analysis and replace X1, Y1 with X2, Y2. For the third analysis you replace X1, Y1 with X3, Y3 and for the fourth X1, Y1 with X4, Y4.



## Where corr. fail: Anscombe's quartet



Now, let's have a look at all of them in comparison: Most importantly, we see that all correlation coefficients and p-values are identical ( $r = 0.816$ ,  $p = 0.002$ ). What we also see is that the plots differ quite a bit.

For the X1 – Y1 combination, the dots are relatively close to the regression line and how much they are away from the regression line seems random. This is pretty much what we typically would expect for data points following a normal distribution.

The X2 – Y2 combination has a curvilinear relationship. Since the mathematics underlying our statistics are (in principle) only suited to describe linear relationships, such a curvilinear relationship falls a bit outside the scope of these methods. There are methods to describe curvilinear relationships, but those are quite advanced and not introduced in the course.



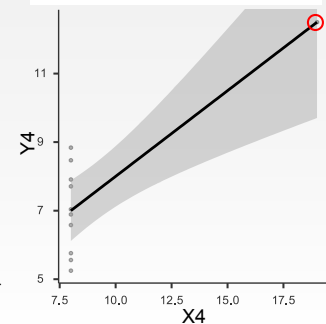
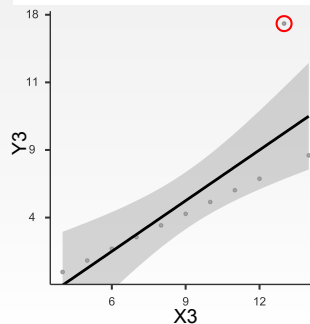
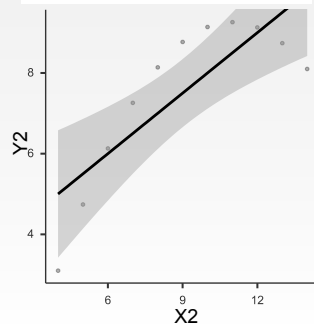
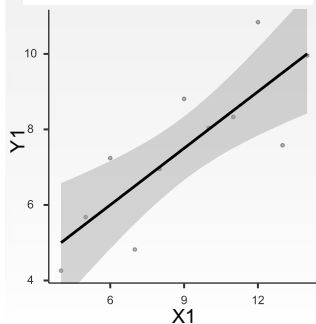
## Where corr. fail: Anscombe's quartet

Correlation Matrix		
	X1	Y1
X1	Pearson's r	—
	p-value	—
Y1	Pearson's r	0.816
	p-value	0.0022

Correlation Matrix		
	X2	Y2
X2	Pearson's r	—
	p-value	—
Y2	Pearson's r	0.816
	p-value	0.0022

Correlation Matrix		
	X3	Y3
X3	Pearson's r	—
	p-value	—
Y3	Pearson's r	0.816
	p-value	0.0022

Correlation Matrix		
	X4	Y4
X4	Pearson's r	—
	p-value	—
Y4	Pearson's r	0.817
	p-value	0.0022



REFRESHER: STATISTICAL ANALYSES

SEBASTIAN.JENTSCHKE@UIB.NO

SLIDE 84



For the X3 – Y3 relationship we see a line of dots that is much steeper than the regression line plus a single dot at the very right end that is much underneath the regression line. This example perfectly illustrates the effect an outlier can have. All our statistical procedures are based upon minimizing the squared deviation of all points included in the calculation of our correlation.

You can now try something: Change to the “Data”-tab so that you have access to the spreadsheet, move the cursor to line 3 in the column Y3 and remove the value. See what happens to the scatterplot and the correlation – we now have a perfect (positive) correlation. Just from removing one outlier. Hit Ctrl-Z to revert that. We could have been made aware of that outlier by running “Descriptives” and checking for Normality using the Shapiro-Wilk-test (which is significant for Y3, indicating that this variable deviates from a normal distribution).



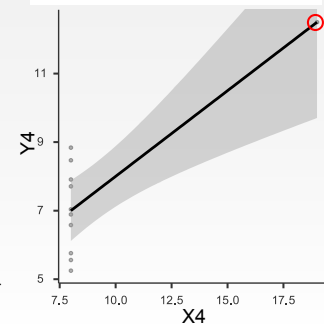
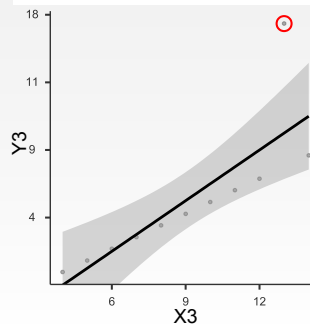
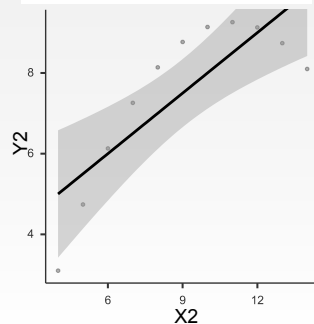
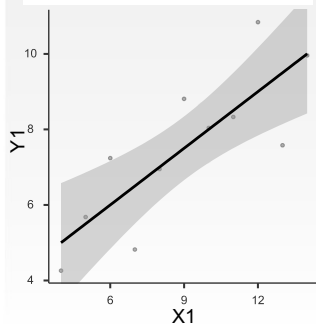
## Where corr. fail: Anscombe's quartet

Correlation Matrix		
	X1	Y1
X1	Pearson's r	—
	p-value	—
Y1	Pearson's r	0.816
	p-value	0.0022

Correlation Matrix		
	X2	Y2
X2	Pearson's r	—
	p-value	—
Y2	Pearson's r	0.816
	p-value	0.0022

Correlation Matrix		
	X3	Y3
X3	Pearson's r	—
	p-value	—
Y3	Pearson's r	0.816
	p-value	0.0022

Correlation Matrix		
	X4	Y4
X4	Pearson's r	—
	p-value	—
Y4	Pearson's r	0.817
	p-value	0.0022



REFRESHER: STATISTICAL ANALYSES

SEBASTIAN.JENTSCHKE@UIB.NO

SLIDE 85



The combination of X4 – Y4 is even more extreme.

All values in X4 except from one outlier (case 19) have the same value: 8. In the plot the outlier is a bit difficult to spot (it is lying on the regression line).

There is also something interesting happening when you remove that outlier. Go to “X4” and remove the 19 in line 8. Suddenly, the correlation got a “NaN” (meaning NaN) because now all values in X4 are 8 and we do not have any variation in that variable anymore. We can therefore not calculate a correlation: If all variables have the same value, the difference from the mean that we use to calculate our correlation gets 0 (since all variables are now equal to the mean). Hit Ctrl-Z do revert that removal. The Shapiro-Wilk-test in Descriptives, assessing normality, should already have worked as a warning: the variable X4 got a highly significant p-value, indicating a clear deviation from a normal distribution.



# Summary

## Categorical predictor and outcome variables (Ch. 10)

- $\chi^2$ -test of association: step-by-step and in jamovi
- $\chi^2$  goodness-of-fit in jamovi
- McNemar test for Paired Samples in jamovi

## Categorical predictor and continuous outcome variables (Ch. 11)

- z-test and One Sample t-test: step-by-step and in jamovi
- Independent Samples t-test
- Paired Samples t-test
- Assumptions and non-parametric alternatives

## Continuous predictor and continuous outcome variables (Ch. 12)

- Correlations: step-by step and in jamovi
- Where correlations fail: Anscombe's quartet



Let's briefly recapture what we learned in this lecture: We introduced three classes of analyses for different types of data: (1) categorical predictors and categorical outcomes, based upon the  $\chi^2$ -distribution (and did even some calculations by hand). (2) We introduced different types of t-tests that we can use to assess the influence of a categorical predictor on a continuous outcome variable. (3) Finally, we looked at correlations used to explore relations between continuous variables.

There is hopefully another thing you were made aware of, e.g., by the example with Anscombe's quartet. Statistical measures are one side of thing. Applying them sensibly and responsibly another. Typically, this is not achieved by just following "recepies" in a cookbook. You should rather aim to understand what you data "mean" and how you can interpret them. Don't be afraid to try things out and see what happens. Only that way, you learn and get better.



**Finally, you are done!  
Thank you for  
bearing with me!**



---

UNIVERSITY OF BERGEN

