Empirical Research Methods 1

T-Test Independent samples & ANOVA Test

Instructor: Albulene Grajcevci

Tutors: Shravani & Piumal

Slides adopted from Miguel Rejon

Session 7: December 12th, 2023

edutech
SAARLAND UNIVERSITY

Agenda

- Normality testing
  - But why?
  - In SPSS
- Degrees of Freedom (quick explanation)
- T-test family of tests
  - T-test family of tests
  - Student's t-test
- Independent samples t-test in SPSS
  - Assumptions in detail
  - In SPSS
- Self-paced activities

NORMALITY TESTING

But why? ● In SPSS

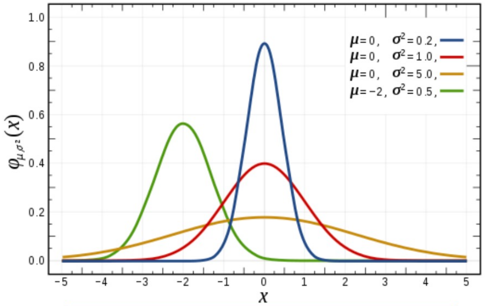
Normality testing. But why?

- Normality testing refers to checking that the data from our variable produces a normal distribution

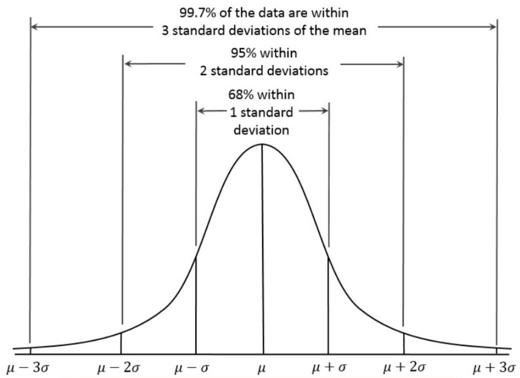
Reminder of what a normal distribution is. Slides from topic 6, slide 8.

Normal distribution

- A.k.a. Gaussian or Gauss or Laplace–Gauss distribution or Bell curve
- Data that produces (or fits) a normal distribution can be called “normal data” (e.g. intelligence scores)
- **It is a probability distribution that is symmetrical, unimodal, and the probabilities for values further away from the mean taper off <u>equally</u> in both directions. Extreme values in both tails of the distribution are similarly unlikely.**
- Normality is a basic assumption/ requirement for parametric tests. Therefore, we want normal data because we want parametric tests!



Different types of normal distributions



Standard normal distributions

8 https://en.wikipedia.org/wiki/Normal_distribution; https://statisticsbyjim.com/basics/normal-distribution

caltech

Normality testing. But why?

- And why we want to ensure that our variable data produces a normal distribution? So we can use such data/variable on a parametric test

- Which tests are parametric? So far: Pearson's correlation (seen on topic 7) and Independent-samples t-test (to be seen on this topic)

Reminder of parametric tests assumptions. Slides from topic 6, slide 10.

Parametric tests

We mentioned that we want normal data because we want (to use) parametric tests. But what are parametric tests?

- A parametric test is a statistical test which makes certain assumptions about the distribution of the unknown parameter of interest and thus the test statistic is valid under these assumptions.

- Parametric tests need to ensure four basic assumptions for the test to be accurate:

1. A normally sampling distribution
2. Homogeneity of variance
3. Interval or ratio data
4. Independence

-Homogeneity of variance: the assumption that the variance of one variable is stable (i.e. relatively similar) at all levels of another variable

-Independence: the assumption that one data point does not influence another. When data come from people, it basically means that the behaviour of one person does not influence the behavior of another. Sometimes also refered as "independence of observations"

https://www.analytics-toolkit.com/glossary/parametric-test/
Field, A. P. (2009). *Discovering statistics using SPSS* (3rd ed.). Sage.

Normality testing. But why?

- So how do we check for normality? Using SPSS, we do it by:

1. Creating an histogram of the variable and inspecting its shape

2. Doing a Shapiro-Wilk (S-W) test, hoping to obtain a non-significant result

Why do we want a non-sig. result on the S-W test? Shouldn't we always strive to get significant results?
In the case of the S-W test, its H0 states that the data comes from a normal distribution whereas its HA predicts that the data is non-normal.

That's why in this case we would want to accept the H0 of this test.

Normality testing in SPSS

- So how do we check for normality? Using SPSS, we do it by:

1. Creating an histogram of the variable and inspecting its shape

2. Doing a Shapiro-Wilk (S-W) test, hoping to obtain a non-significant result

Why do we want a non-sig. result on the S-W test? Shouldn't we always strive to get significant results?
In the case of the S-W test, its H0 states that the data comes from a normal distribution whereas its HA predicts that the data is non-normal.

That's why in this case we would want to accept the H0 of this test.

Normality testing in SPSS

Example:

We have the variable “negativ”, which measures the amount of negative adjectives remembered by our participants across all our 3 experimental conditions. We want to do a parametric test with this variable but before that we need to check if the variable “negativ” is normal.

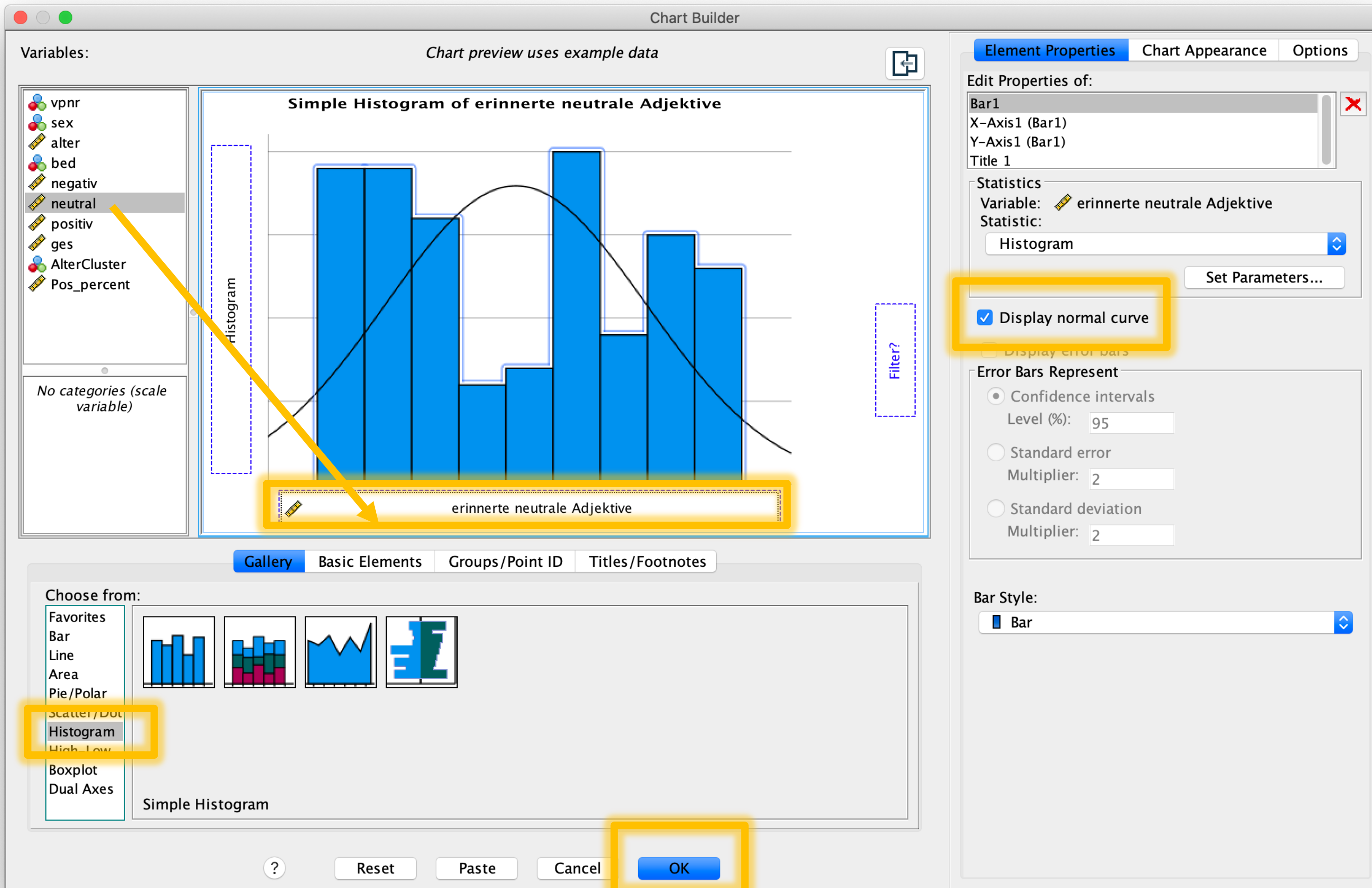
First, create a histogram of the “negativ” variable to visually inspect if the distribution of the variable resembles a normal distribution.

Second, we will do the Shapiro-Wilk test to statistically check for normality.

Normality testing in SPSS

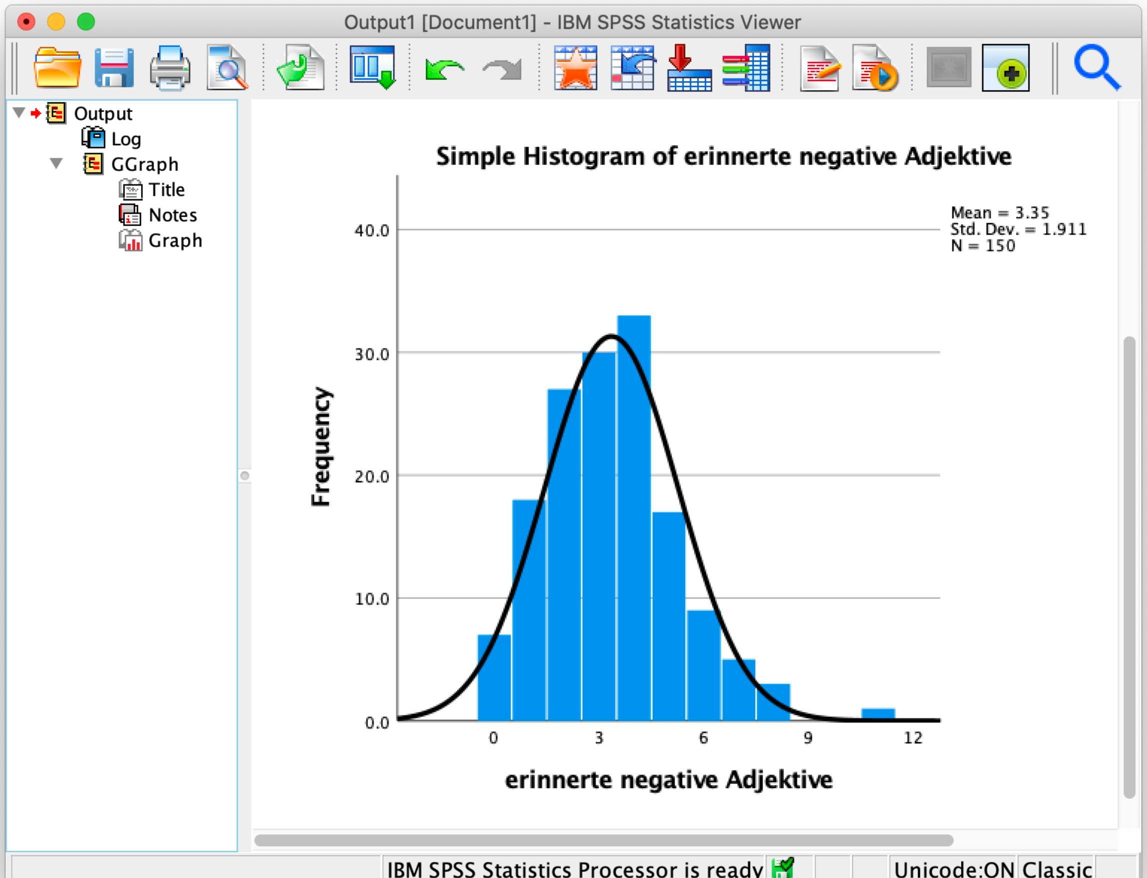
1. Visual inspection of histogram:

- Graphs > Chart Builder > Histogram



Normality testing in SPSS

1. Visual inspection of histogram

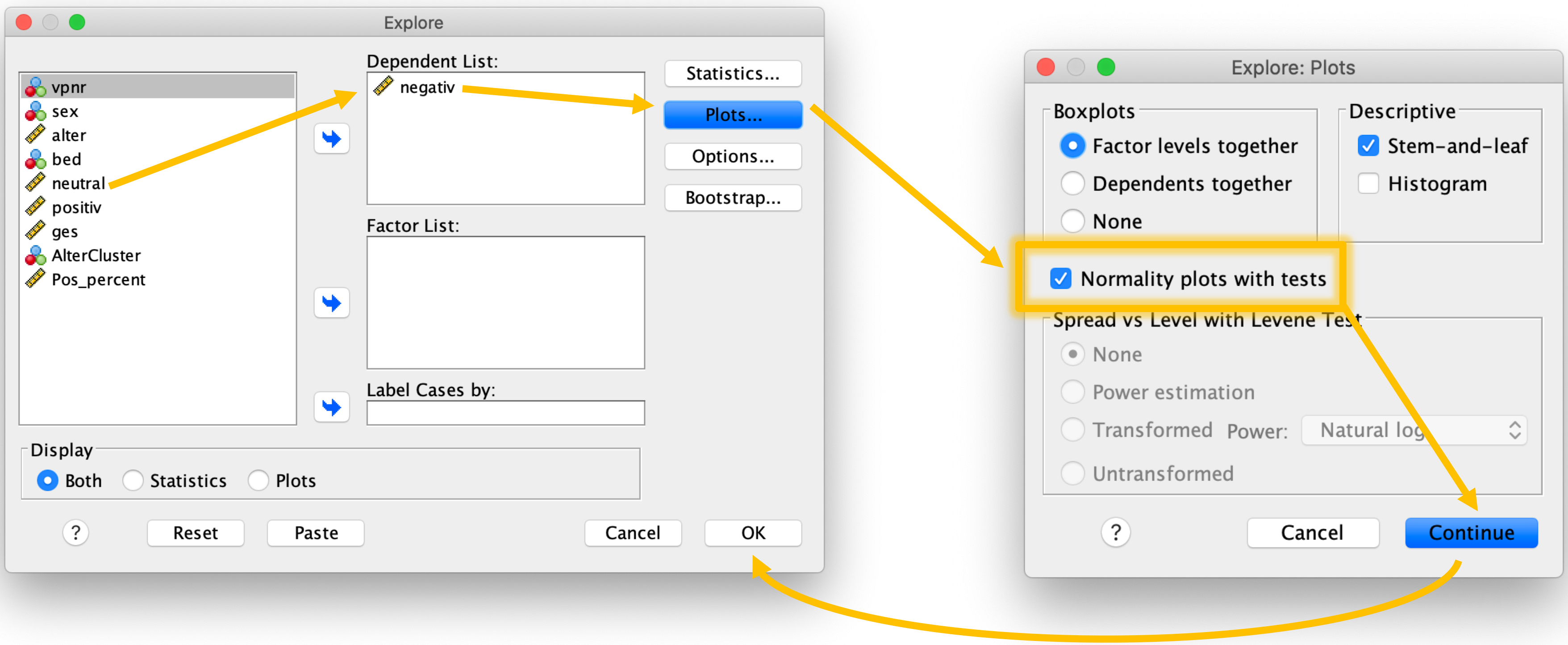


This is the histogram from the “negativ” variable.
If we compare the blue bars (i.e. the data from the variable) with the black line (the “display normal curve” option we selected), we can see that the data more or less fits a normal distribtuion (i.e. the black line).
To gain more certainty about the normality of our data, we will do the Shapiro-Wilk test.

Normality testing in SPSS

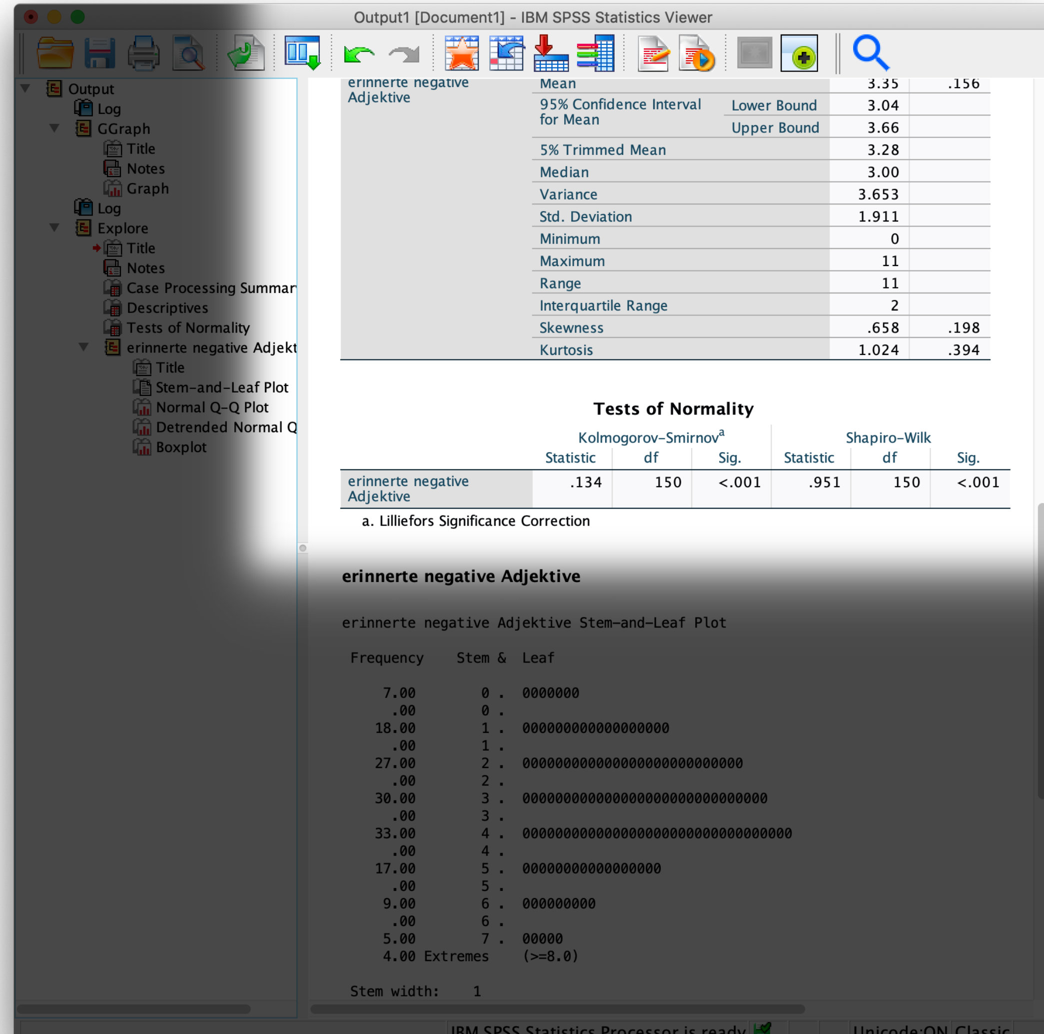
2. Shapiro-Wilk test of normality:

- Analyze > Descriptive Statistics > Explore



Normality testing in SPSS

2. Shapiro-Wilk test of normality



From all the outputs we got, we will only focus on the "Tests of Normality" table.

Normality testing in SPSS

2. Shapiro-Wilk test of normality

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
negativ (erinnerte negative Adjektive)	.134	150	<.001	.951	150	<.001

a. Lilliefors Significance Correction

Here, the Shapiro-Wilk (S-W) test got a Sig. value of <0.001.
The H0 of the Shapiro-Wilk test states that the data is normally distributed, therefore with a p-value of <0.001, we determine that the test is statistically significant and therefore we reject the H0.
Therefore, we conclude that our data is NOT normally distributed

Remember: 1) This is a normality test, 2) We don't always want to reject the H0

Normality testing in SPSS

According to the S-W test, the data from our “negativ” variable is not normally distributed, what now?

- Normality tests often show a violation, and often we use our stat. (parametric) tests anyway (even though normality is violated), because t-test etc. are robust against these violations

- Violating the normality assumption is often not a problem, as statistical tests such as the the t-test are considered to be “robust”, especially when the sample (n) is large enough (> 30; some also say if <10; need to look that up and make a decision yourself). If your n is very small and a normal distribution cannot be assumed, you should switch to a non-parametric test. However, a t-test usually is still the best choice, and “robust” enough.

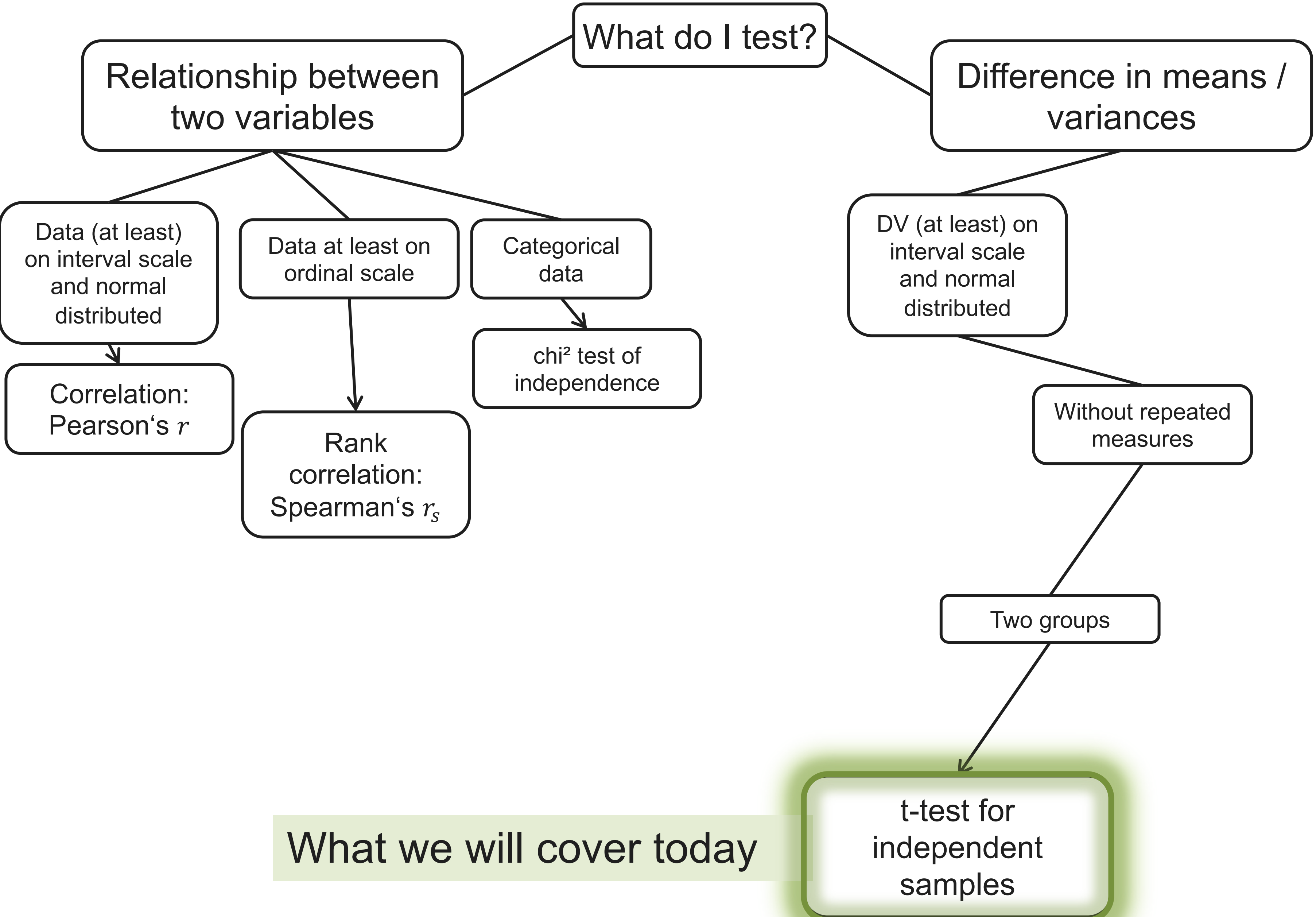
Bortz, J. (2005). Statistik, 6. *Aufl., Heidelberg*.
Pospeschill, M. (2006). Statistische Methoden. *Strukturen, Grundlagen, Anwendungen in Psychologie*.

Normality testing in SPSS

According to the S-W test, the data from our “negativ” variable is not normally distributed, what now?

- If $n \geq 30$: carry on
- If $n < 30$: do a data transformation (beware!) or use a non-parametric test (specially when other parametric assumptions are also violated)

All in all: T and F families of tests are robust enough to handle non-normal distributions. Ideally, try to always have a $n \geq 30$


What do I test?
Relationship between two variables
Difference in means / variances
Data (at least) on interval scale and normal distributed
Data at least on ordinal scale
Categorical data
DV (at least) on interval scale and normal distributed
Correlation: Pearson's r
Rank correlation: Spearman's r_s
chi² test of independence
Without repeated measures
Two groups
t-test for independent samples
What we will cover today

DEGREES OF FREEDOM

Quick explanation

Degrees of Freedom

- df = DF = degrees of freedom = How many values can I vary?

This is a quick mention of what DF are. Below, there are 3 definitions that I copied textually which I think explain the concept well. The point here is for you to have a basic idea of what DF are but is not necessary to have a deep understanding about the concept because that's out of the scope of the course.

- "Degrees of freedom (df) of an estimate is the number of **independent pieces of information that went into calculating the estimate**. It's not quite the same as the number of items in the sample. In order to get the df for the estimate, you have to subtract 1 from the number of items." (Glen, n.d., para 1)

- "In statistics, the degrees of freedom (DF) indicate the **number of independent values that can vary** in an analysis without breaking any constraints." (Frost, 2020, para. 1)

- "Degrees of freedom are often broadly defined as the **number of "observations"** (pieces of information) in the data that are **free to vary when estimating statistical parameters**." (Minitab Blog Editor, 2016, para. 9)

T-TEST FAMILY OF TESTS

T-test family of tests ● Student's t-test

Student‘s t?

- Related to concepts such as: Student’s t statistic, Student’s t distribution, Student’s t-test
- The t distribution is named after William Gosset who used the pseudonym “Student” for publishing
- Gosset worked for the Guiness brewery and was interested in applying statistics to develop better beers
- The t-test we use nowadays builds on his (and others’) work onto
- There is more than one type of t-test; these slightly different t-tests belong to the T-test family

T-test family of tests

- T-test family of tests: When you want to find out if 2, and only 2, **means** are statistically significantly different
- This is a family of parametric tests!
- Three members of the t-test family: independent-samples, repeated-samples, and one-sample
- Regardless if it's an independent-samples, repeated-samples, or one-sample, any t-test we do:

a) consists of comparing the difference between 2 means, accounting for its variability, and

b) the output of such comparison will produce a **t-value** that can be charted on a **t-distribution** and from this we will obtain a p-value to determine if our test was significant or not

The focus of this topic - independent-samples t-test, which we will see in a moment, but before that we will learn about t-values and t-distributions.

t-value. What's it?

- It is the result of comparing 2 means relative to their variability. Therefore, it could be said that the t-value formula is a ratio

- It is the result from conducting a t-test which ”provides” a p-value which is what we use to determine statistical significance

In conclusion: a t-value is the statistic from a t-test. Such value will allow us to determine if, by comparing their means (relative to their variability), **<u>2 groups</u>** are different.

BTW, “group” in this case refers to the 2 levels or conditions from our IV

edutech

The t-distribution

- Similar shape as a normal distribution
- Can we used when comparing 2 means but we don't know a) the population standard deviation (σ) and/ or b) have a sample ≤ 30. (Either of these which happen very often!)
- The t-distribution has thicker tails for accommodating more variation
- Required: mean, variance and degrees of freedom
- An obtained t-value can be located onto a t-distribution

t-distribution and the effect of degrees of freedom

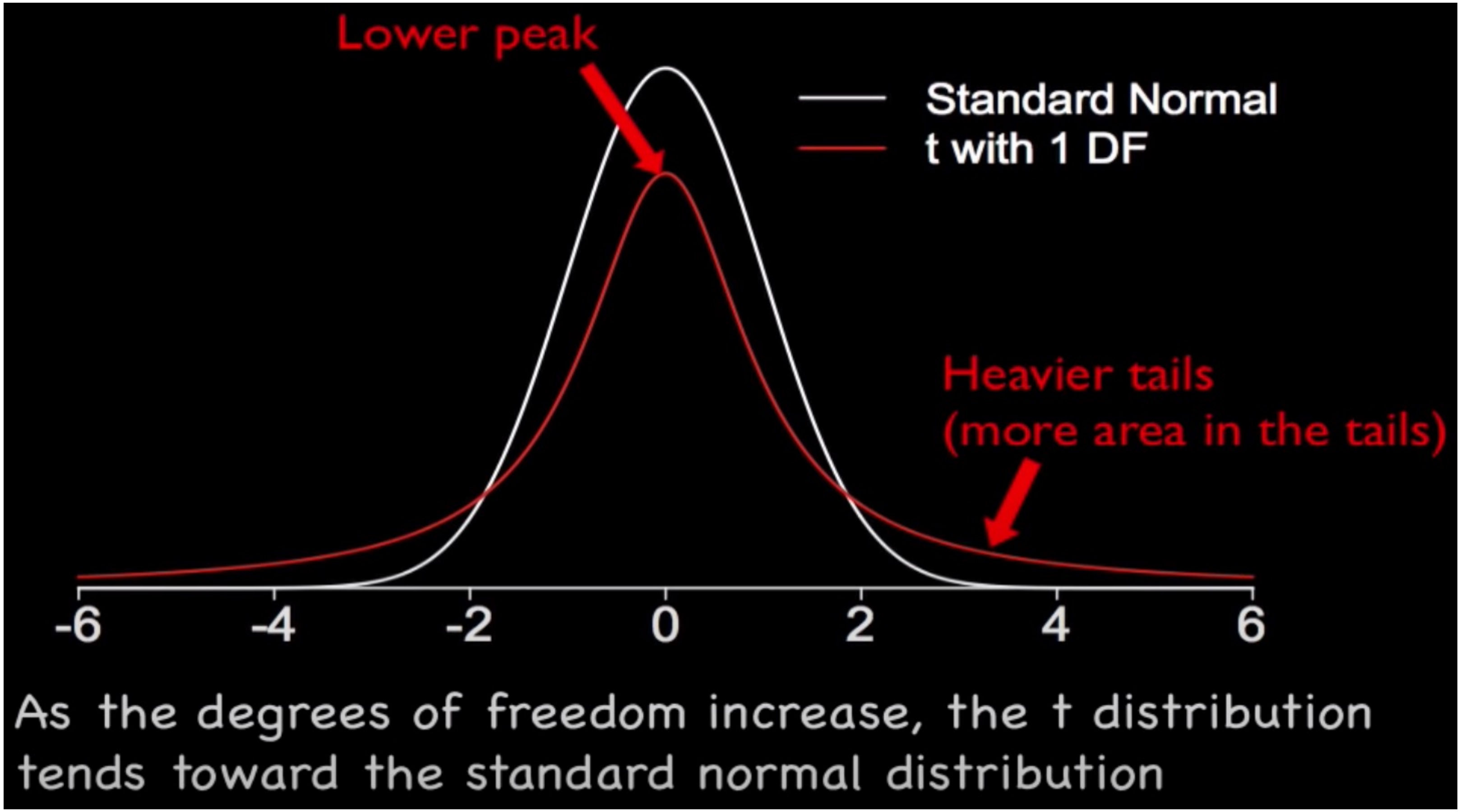


Image source: https://www.youtube.com/watch?v=Uv6nGIgZMVw

Student's t, t, t: Distribution, value, test. Recap

- t-test: parametric test that compares the means of 2 groups to determine if the **groups** are different
- t-value: result of a t-test which will allow us to obtain a p-value and determine statistically significant results
- t-distribution: Similar as a standard normal distribution but with thicker tails to accommodate dispersion (a.k.a. uncertainty)

INDEPENDENT-SAMPLES T-TEST IN SPSS

Assumptions in detail ● In SPSS

Independent-samples t-test. Assumptions in detail

Assumptions			
Number	Name	Description	How to check for that?
1	Categorical independent variable	Independent variable is or nominal/categorical scale level.	Ensure the data from your independent variable corresponds to a nominal/categorical scale level
2	Continuous dependent variable	Dependent variable is either ratio or interval level.	Ensure the data from your dependent variable corresponds to a ratio or interval level
3	Independence of observations	The measurement of each observation was independent and didn't affect the measurement of the other observations.	When collecting the data, ensure each collected/measured datapoint is independent from each other
4	No outliers	Removal or correction of outliers (i.e any datapoint that does not follow the usual pattern all other datapoints follow).	Visual inspection of the boxplot of the dependent variable, ensure there are no indvidual outlier datapoints. If there are, exclude them from the analysis.
5	Normality	Data from the dependent variable should follow a normal distribution curve.	Shapiro-Wilk test for the dependent variable. Test should be not-significant!
6	Homogeneity of variances	The variances of the 2 compared groups are statistically the same, if not very similar.	Levene's test for equality/homogeneity of variances of the 2 levels of the dependent variable. Test should be not-significant! Test is part of the output of the indep. t-test analysis.

Independent-samples t-test in SPSS

Example:

- Research Context: While researching about the impact multimedia content and its influence on helping memorize adjectives, we want to ensure that males and females from all 3 conditions are similar in terms of age.
- H0: There is **no** difference between the average age of females and the average age of males.
- Sample and procedure: We recruit 150 people who are randomly allocated onto 3 conditions. As part of the demographic items, we ask them for their age.
- Variables and levels: Not necessary for this analysis but the experiment had 1 factor with 3 levels (“textual”, “pictorial”, and ”emotional”)
- Before doing any analyses, we decide to follow the statistical conventions of our scientific field and we set our alpha to 5%

Note: dataset taken and adapted from http://quantitative-methoden.de/Downloads_A3.htm

Independent-samples t-test in SPSS. Checking assumptions

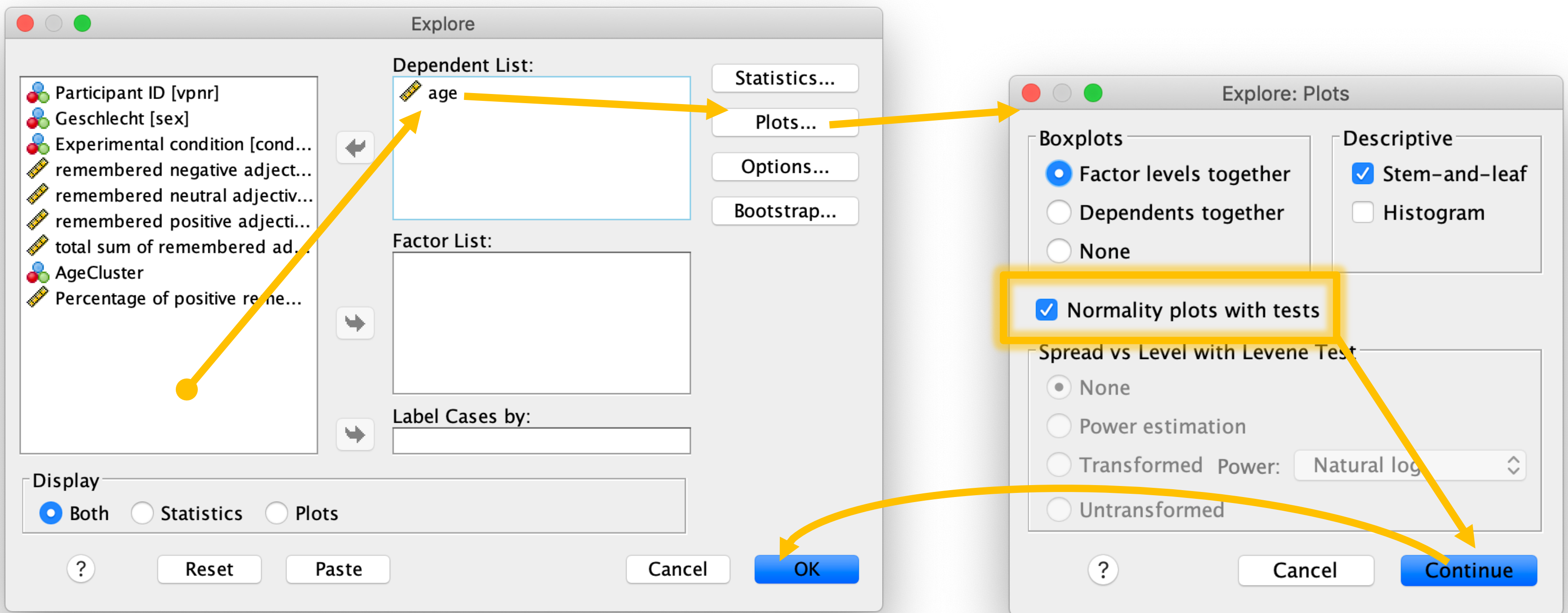
Assumptions			
Number	Name	How to check for that?	Assumtion met?
1	Categorical independent variable	Ensure the data from your independent variable corresponds to a nominal/categorical scale level	Yes
2	Continuous dependent variable	Ensure the data from your dependent variable corresponds to a ratio or interval level	Yes
3	Independence of observations	When collecting the data, ensure each collected/measured datapoint is independent from each other	Yes
4	No outliers	Visual inspection of the boxplot of the dependent variable, ensure there are no indvidual outlier datapoints. If there are, exclude them from the analysis.	Yes
5	Normality	Shapiro-Wilk test for the dependent variable. Test should be not-significant!	pending
6	Homogeneity of variances	Levene's test for equality/homogeneity of variances of the 2 levels of the dependent variable. Test should be not-significant! Test is part of the output of the indep. t-test analysis.	To be checked along with the t-test output

The used data meets all the assumptions but for brevity purposes, contents about how to check for assumptions 2 to 4 won't be shown here.

Indep.-samples t-test in SPSS. Assumption 5 – Normality

Shapiro-Wilk test of normality

- Analyze > Descriptive Statistics > Explore



Indep.-samples t-test in SPSS. Assumption 5 – Normality

Shapiro-Wilk test of normality

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
age	150	100.0%	0	0.0%	150	100.0%

Tests of Normality

	Kolmogorov–Smirnov[a]			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
age	.183	150	<.001	.771	150	<.001

a. Lilliefors Significance Correction

The p-value of the Shapiro-Wilk (S-W) is <0.001, which means that the result is statistically significant and therefore that the data from the “age” variable is NOT normally distributed. Even if the results of the S-W test indicate that the data is not normal, following the advice from slide 15, we will still use a parametric test (i.e. t-test) because our sample is big enough (N = 150).

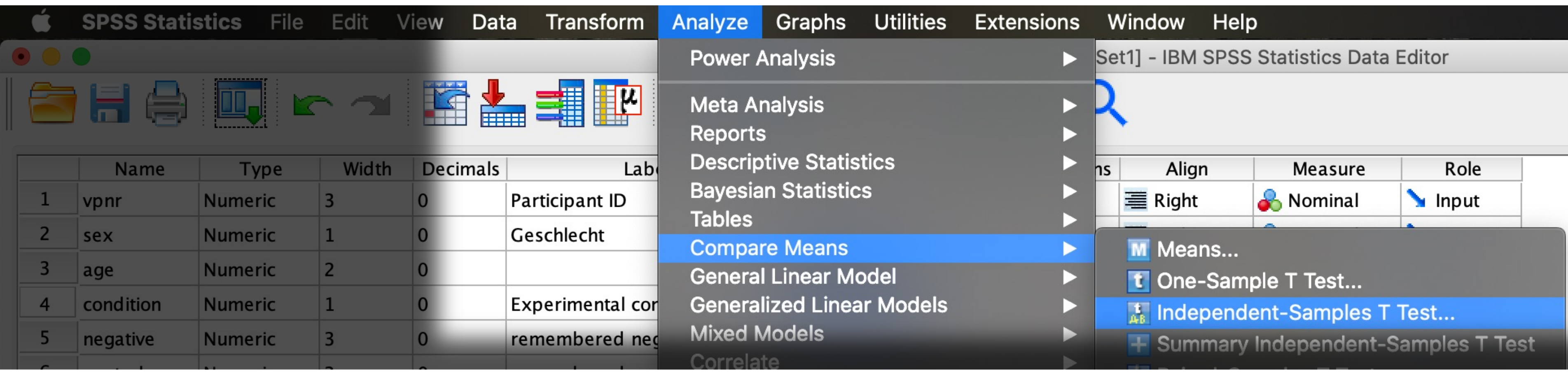
Indep.-samples t-test in SPSS

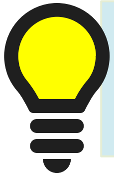
Now that assumptions 1 to 5 are covered, we can proceed with the actual t-test analysis.

Assumption 6 (Homogeneity of variances) will be checked while inspecting the output of the t-test.

Indep.-samples t-test in SPSS

Analyze > Compare Means > Independent-Samples T Test…

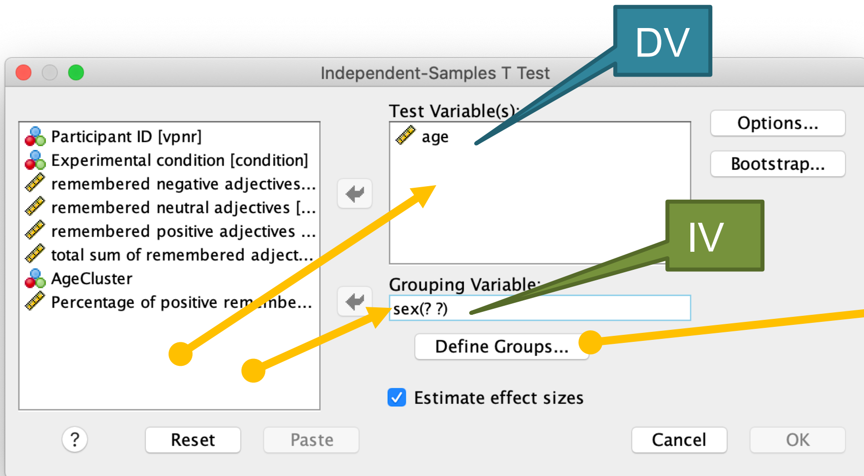


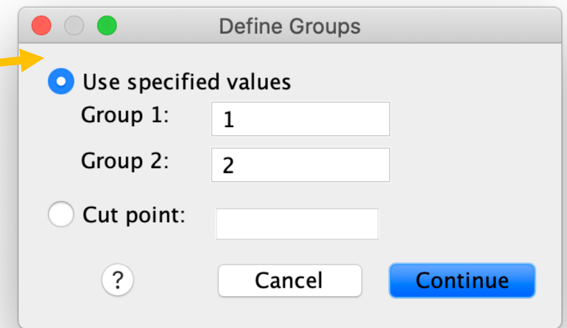


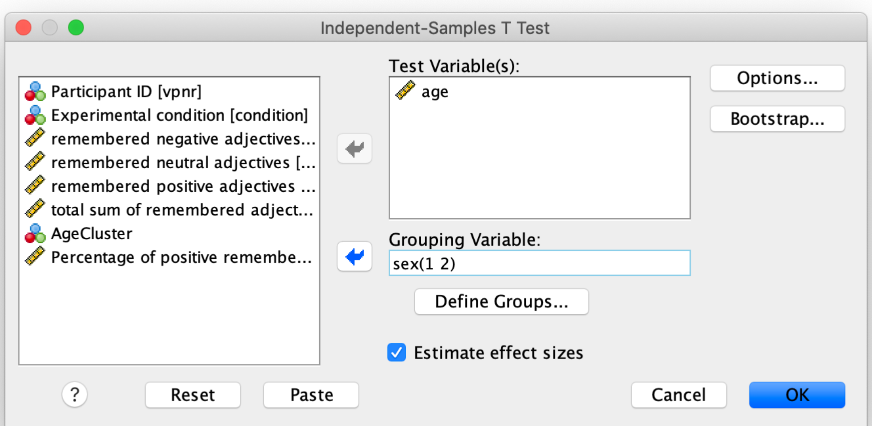
Want to try it out? Use the file "topic-8.indep-t-test.sav" provided in the moodle

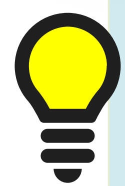
Indep.-samples t-test in SPSS

Analyze > Compare Means > Independent-Samples T Test…





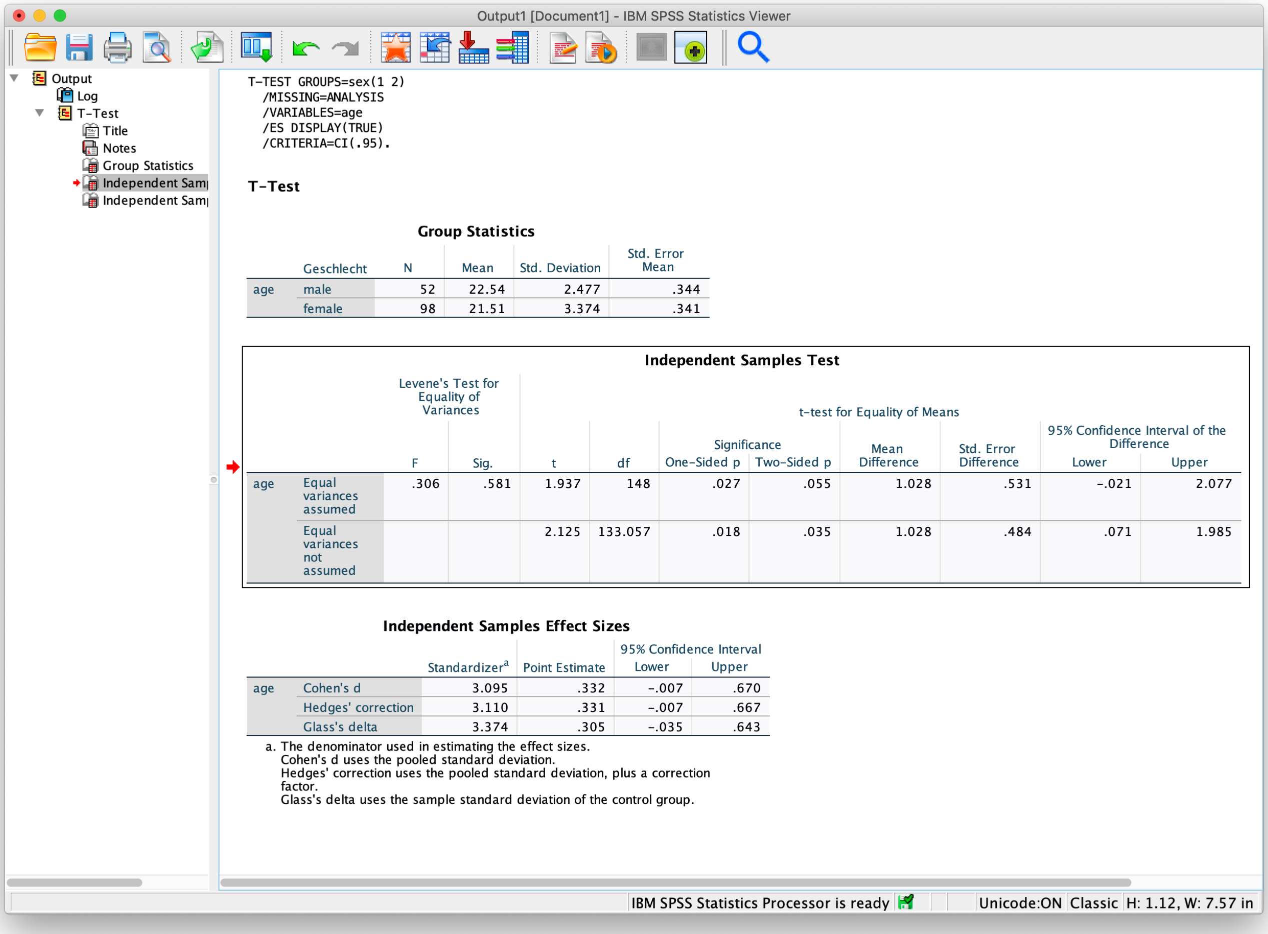




In this case, the specified values for Group 1 and Group 2 were 1 and 2. This was because those are the values we used for coding the different levels in that variable. You can still do a t-test if your IV has more than 2 levels, however, you would only be able to compare 2 levels at once

Indep.-samples t-test in SPSS. Output

The output of this tests consists of 3 tables. We will review them one by one in order of appearance.




```
T-TEST GROUPS=sex(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=age
  /ES DISPLAY(TRUE)
  /CRITERIA=CI(.95).
```

T-Test

Group Statistics

	Geschlecht	N	Mean	Std. Deviation	Std. Error Mean
age	male	52	22.54	2.477	.344
	female	98	21.51	3.374	.341

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
age	Equal variances assumed	.306	.581	1.937	148	.027	.055	1.028	.531	-.021	2.077
	Equal variances not assumed			2.125	133.057	.018	.035	1.028	.484	.071	1.985

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
age	Cohen's d	3.095	.332	-.007	.670
	Hedges' correction	3.110	.331	-.007	.667
	Glass's delta	3.374	.305	-.035	.643

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Indep.-samples t-test in SPSS. Output

DV

IV

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
age	male	52	22.54	2.477	.344
	female	98	21.51	3.374	.341

This first table contains descriptive information about our 2 variables.
Here we can see that there were 52 males and their mean age was of 22.54 years old. For the case of the females, there were 98 of them and their mean age was 21.51 years old.
From these results, we can already say that:
a) the mean ages of males and females are different, and
b) that females were on average younger than males.
However, what we still don't know, is if this difference is statistically significant. For that, we need to look at the next table ;-)

Indep.-samples t-test in SPSS. Output

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
age	Equal variances assumed	.306	.581	1.937	148	.027	.055	1.028	.531	-.021	2.077
	Equal variances not assumed			2.125	133.057	.018	.035	1.028	.484	.071	1.985

This is the second table from our output. Here we have the core statistical values (t-value, p-values, etc.) from our t-test.
First, we need to check assumption 6 (slide 27 and 29), homogeneity of variances. This is done with the Levene's test for Equality of Variances which is automatically computed every time we do a t-test.

Under the "Sig." column of the Levene's test section (green rectangle), is the p-value of the Levene's test. The H0 of Levene's test states that the variances of the 2 levels of the IV (sex, in this case) regarding the DV (age) are the same. Given that we obtained a p-value of 0.581 **for the Levene's test**, we can accept the H0 and therefore determine that both IV levels have non-significantly different variances.

With this, we meet assumption 6 of the t-test and for the next results statistics (t-value, p-value of the t-test, etc.), we will read the results from the "Equal variances assumed" row (blue rectangle).

IMPORTANT: If the Levene's test is significant, the homogeneity is not given and we need to read the results from the row "equal variances not assumed"

Indep.-samples t-test in SPSS. Output

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
age	Equal variances assumed	.306	.581	1.937	148	.027	.055	1.028	.531	-.021	2.077
	Equal variances not assumed			2.125	133.057	.018	.035	1.028	.484	.071	1.985

Still on the second table, we will now have a look at the other statistics from our t-test, focusing only on the first row because we have already determined that there is homogeneity of variances.
Under the “t” column, we will find the t-value (i.e. 1.937) of our t-test. Next to it we find the degrees of freedom; these 2 values would be needed in case we would want to plot our t-distribution. We don’t need to plot a t-distribution.

Under the “Significance” section, we find two p-values. In this case we will take the one under the “Two-Sided p” column. This p-value us of 0.055, which is barely above our alpha of 0.05. Therefore we can conclude that our independent samples t-test was **not significant**.

Indep.-samples t-test in SPSS. Output

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	Upper
age	Cohen's d	3.095	.332	-.007	.670
	Hedges' correction	3.110	.331	-.007	.667
	Glass's delta	3.374	.305	-.035	.643

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

This is the third and last table of our output, which contains the effect sizes of our t-test.

The most often reported effect size value for t-tests is the Cohen's d. Its value can be found under the "Point Estimate" column. Even if our t-test was not significant, we will report its effect size.

On another topic, we will learn about effect sizes. On topic 7 (chi^2) we briefly covered it.

Indep.-samples t-test in SPSS

We‘re done analyzing the results. Before moving to the next and final step, reporting, let‘s have a recap of what we did:

-As part of an experiment with a sample size of 150 participants allocated onto 3 conditions, we want to check that the ages of our male and female participants are statistically the same

-Because we want to compare the averages (of age) of two independent and different groups (i.e. males and females), we decided to do an independent-samples t-test.
-To ensure that the results of the independent-samples t-test analysis would be valid, we checked for the 6 assumptions of this parametric test. Five of these assumptions were checked before doing the t-test, the last one (homogeneity of variances) was done as part of the t-test analysis.

-The main result of our independent-samples t-test, the t-value, was 1.937.
-The corresponding p-value of such t-value was 0.055 (slide 38). **This means that there is a .055 probability that our t-test result was caused by random chance**, which is above our preset alpha of .05. This means that our independent samples t-test is NOT statistically significant and therefore we keep our H0

Is all of this clear? Is it clear from where we got all these numbers? If not, please take note of your questions or doubts and please share them on Monday when we cover this topic.

Now let’s move on and see how to properly report all of this.

Indep.-samples t-test in SPSS

=Context reminder
Research Context: While researching about the impact multimedia content and its influence on helping memorize adjectives, we want to ensure that males and females from all 3 conditions are similar in terms of age.

H0: There is no difference between the average age of females and the average age of males.

=Actual final report
We performed an independent-samples t-test to control for age between male and female participants. There was homogeneity of variances, as assessed by Levene's test for equality of variances (p = .581). There was no statistically significant difference in age between males and females, $t(148) = 1.94$, p = .055 , $d = 0.332$. Therefore we can conclude that there are no significant differences in terms age between male and female participants.

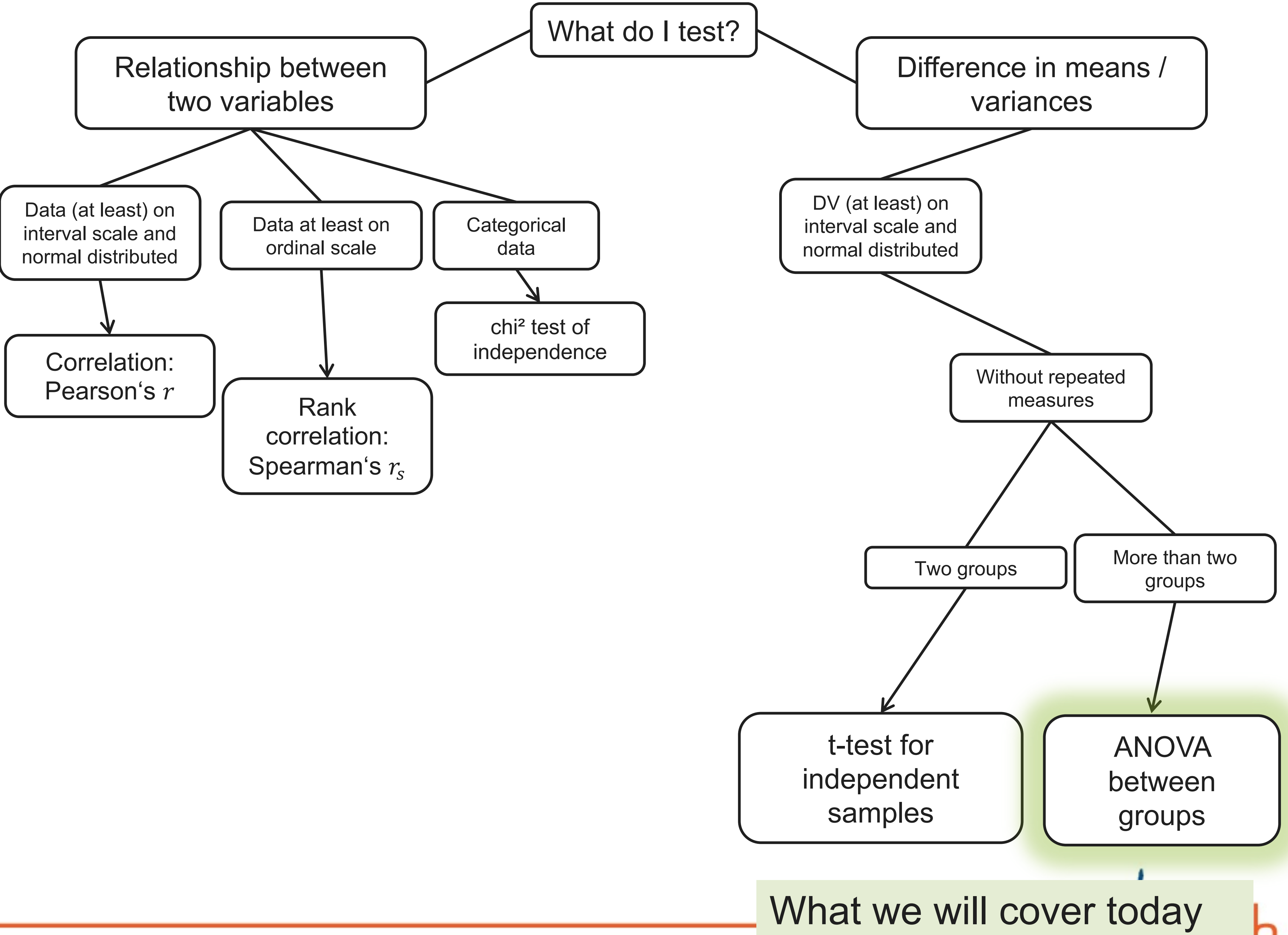
Indep.-samples t-test in SPSS

- Template for reporting the main values from the test:

Template	*t*(df) = [t-value_statistic], *p* = [p-value], *d* = [Cohen's d effect_size_value]
Example 1:	$t(38) = 2.37$, $p = .023$, $d = 0.75$
Example 2:	$t(148) = 1.94$, $p = .055$, $d = 0.332$

Empirical Research Methods 1

ANOVA Test


What do I test?
Relationship between two variables
Difference in means / variances
Data (at least) on interval scale and normal distributed
Data at least on ordinal scale
Categorical data
DV (at least) on interval scale and normal distributed
Correlation: Pearson's r
Rank correlation: Spearman's r_s
chi² test of independence
Without repeated measures
Two groups
More than two groups
t-test for independent samples
ANOVA between groups
What we will cover today

QUICK RECAP

So far on ERM1…

- Independent Variable: a.k.a. IV, <u>factor</u>, input variable, experimental variable, and predictor (variable). Is the variable that is manipulated in our experiment. We expect the IV to have an effect on the dependent variable
- Factors = IV
- Levels: the possible values within a factor. If a factor is a category, a level is a sub-category
- Dependent Variable: a.k.a. DV, outcome variable, response variable. Is the variable that we measure <u>*within our experiment*</u>. We expect the DV to be affected, or depend of, the independent variable
- Categorical Variable: Nominal scale variable with labels used for distinguishing the levels of a factor. <u>The labels can be numerical but won't be treated as such!</u>
- Null hypothesis: Hypothesis that states no difference or change between factors and/or levels
- Alternative hypothesis: Hypothesis that states the existence of a difference or change between factors and/or levels. Can either be directional or non-directional.
- Significance level: a.k.a. p-value, measure of the probability that an observed difference could have occurred just by random chance. That's why we want p-values to be as small as possible.

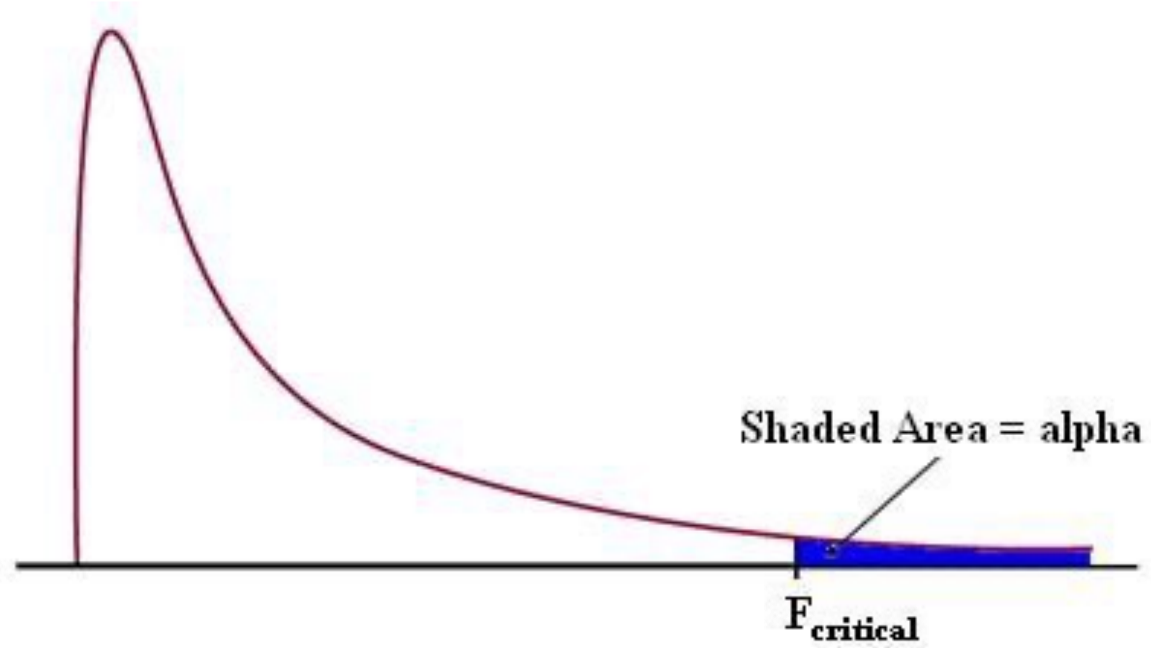
ANOVA EVERYTHING

ANOVA? ● The F distribution ● Different types of ANOVAs ● One-way ANOVA vs t-test ● Two-way ANOVA (factorial design; main and interaction effects) ● ANOVA is an Omnibus test ● A priori and post-hoc ANOVA tests ● Reporting ANOVA ● ANOVA recap

ANOVA?

- ANOVA stands for **AN**alysis **O**f **VA**riance
- ANOVA is part of the family of F-tests which compares means relative to variation within groups
- Similar as the t-test, the ANOVA is a parametric test.
- Somehow similar as the t-test, the ANOVA outputs an F-value which can be located in a F-distribution to determine significance.
- Unlike a t-test, with an ANOVA we can compare the means of >2 levels at once:
  - One-way ANOVA: used when you have **only 1** factor/IV with >2 levels and **only 1** DV
  - Two-way ANOVA: a.k.a. “Factorial ANOVA”, used when you have >1 factor (with multiple levels each) and **only 1** DV

F distribution: The probability distribution for ANOVAs



This is the F distribution, also known as “Fisher distribution”. It is a probability distribution, just like the normal distribution or the t distribution for the t-test.

Works the same: an obtained F-value can be located onto the F distribution. Based on where it is, we can obtain a p-value from the F-value result.

All in all, it can be said that the ANOVA is a generalization/extension of the t-test.

http://www.statisticshowto.com/wp-content/uploads/2013/09/f-table.jpg

ANOVAs

		Independent Variables	
		1	>2
Dependent variables	1	One-way ANOVA	Factorial ANOVA
	>2	Multiple ANOVAS	MANOVA

Underlined: ANOVA tests we will cover on this topic.

Remember that “Factorial ANOVA” is just another way to refer to the “2-way ANOVA”.

Inspired from the table found at: https://www.statisticssolutions.com/conduct-interpret-factorial-anova/

One-way ANOVA vs t-test

Comparing two levels (within the same factor): both are applicable and produce the same p-value, but t-test is more common

Independent Samples Test

		Levene's Test for Equality of Variances				
		F	Sig.	t	df	Sig. (2-tailed)
alter	Equal variances assumed	,306	,581	1,937	148	$\boxed{,055}$
	Equal variances not assumed			2,125	133,057	,035

f Between-Subjects Effects

Dependent Variable: alter

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	35,920[a]	1	35,920	3,751	,055
Intercept	65917,947	1	65917,947	6882,861	,000
sex	35,920	1	35,920	3,751	$\boxed{,055}$
Error	1417,413	148	9,577		
Total	73176,000	150			
Corrected Total	1453,333	149			

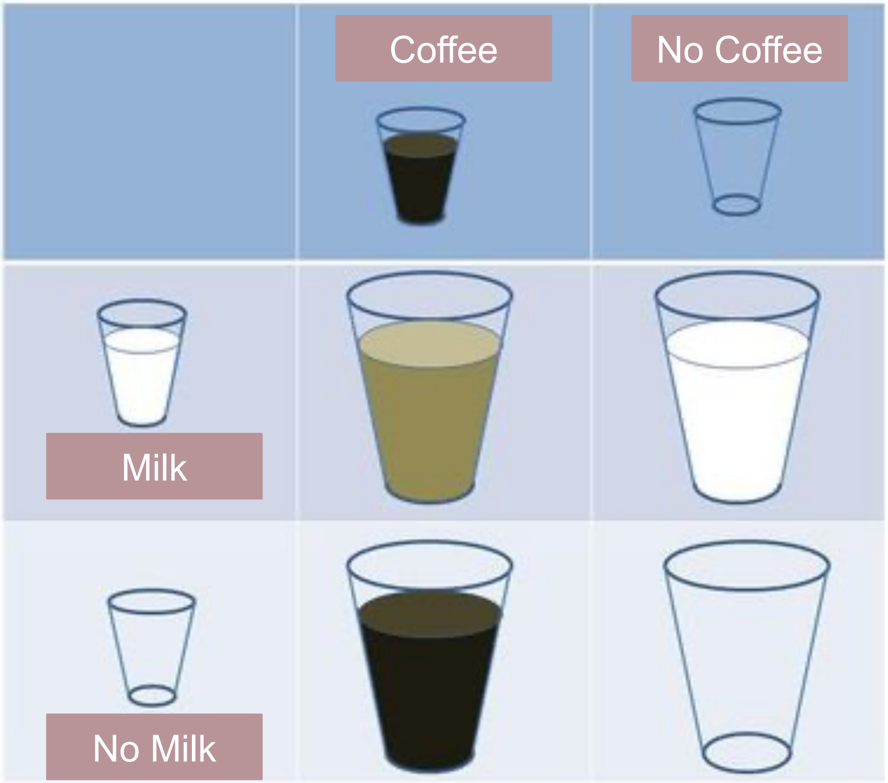
a. R Squared = ,025 (Adjusted R Squared = ,018)

One-way ANOVA vs t-test

- Comparing more than two levels (within the same factor):
  - Alpha inflation: The probability of making an alpha error is increasing when you run many t-tests on the same variables.
  - Therefore, to avoid alpha inflation, the ANOVA is the preferred procedure to compare several groups because the test accounts for multiple paired comparisons
- Comparing more than one factor:
  - → ANOVA: ability to test multi-factorial designs
    - Main effects
    - Interaction effects

We will learn about these 2 in a moment

Two-way ANOVA. Multi-factorial design



For this factorial design, a 2-way ANOVA calculates:

1. Main effect for coffee
2. Main effect for milk
3. Interaction between coffee and milk

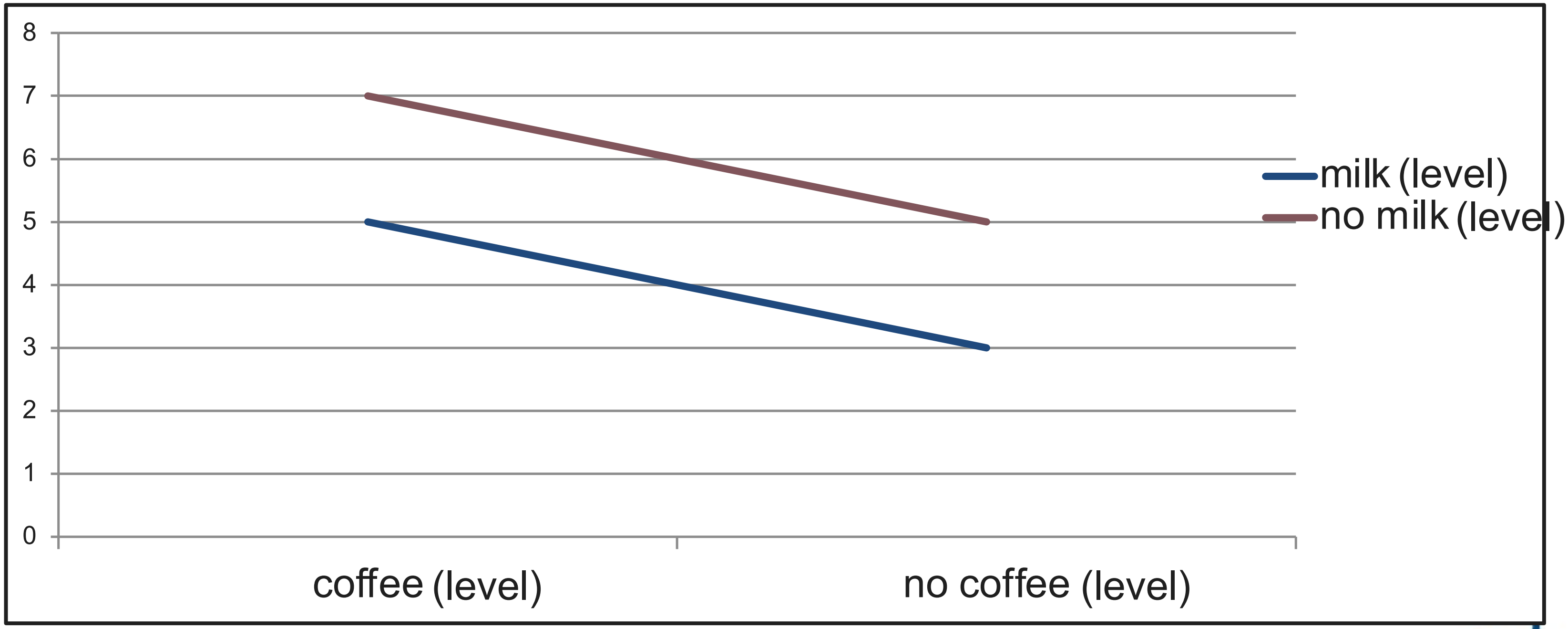
1- From the above-shown factorial design, which are the factors and which are the levels?
2- How would a 1-way ANOVA factorial design look like?

http://versuch.file2.wcms.tu-dresden.de/w/images/thumb/9/98/Mehrfaktoriell_Beispiel_Glas.jpg/391px-Mehrfaktoriell_Beispiel_Glas.jpg

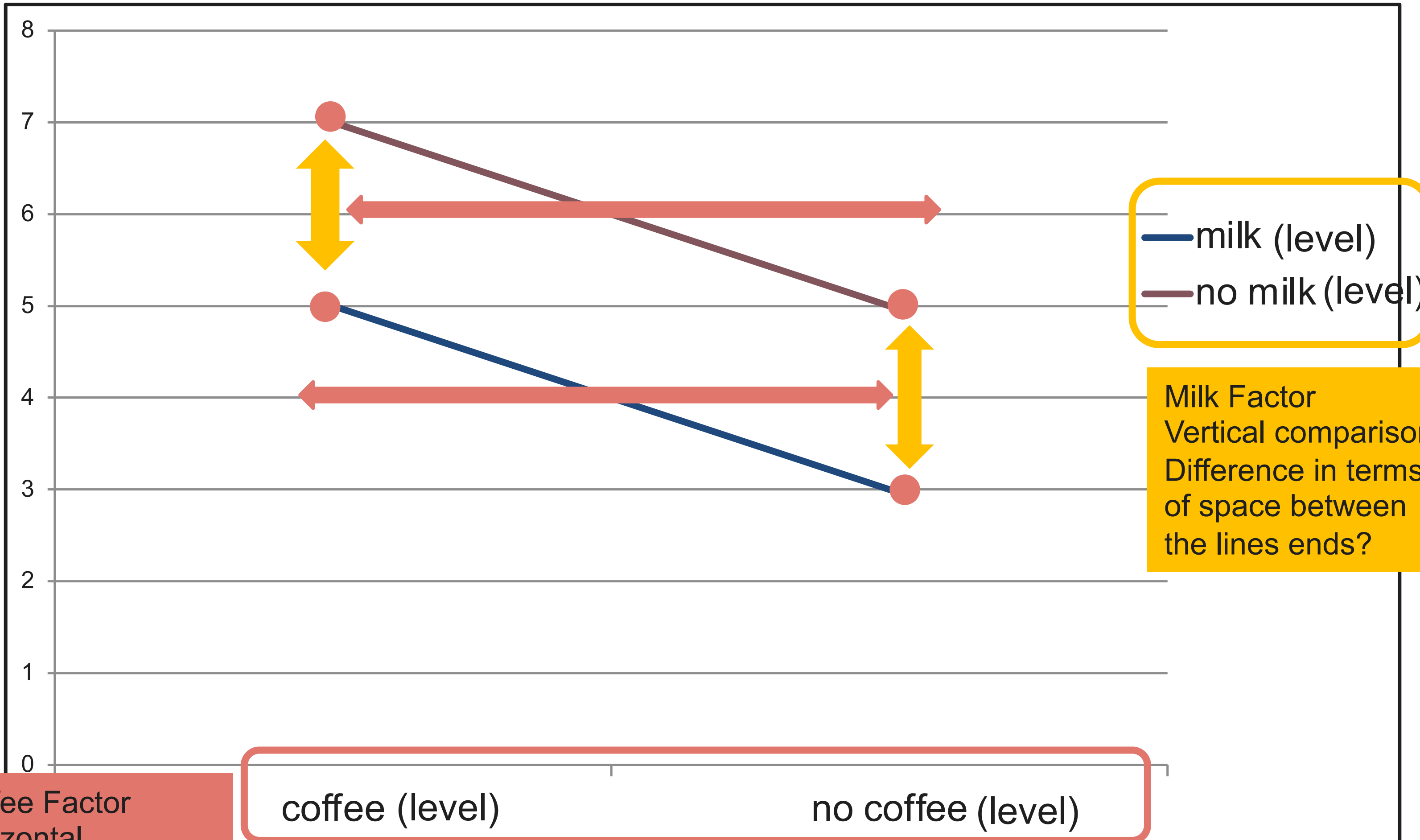
Two-way ANOVA. Main and interaction effects

What you see here below is a plot showing the main an interaction effects from a two-way ANOVA. This plot is part of our SPSS output and it serves as a quick way to estimate if there are main effects and/or an interaction effect between our factors.

The following slides will show you how to understand and interpret these plots



Two-way ANOVA. Main and interaction effects



Milk Factor
Vertical comparison:
Difference in terms of space between the lines ends?

Coffee Factor
Horizontal comparison:
Difference in terms of inclination of the lines?

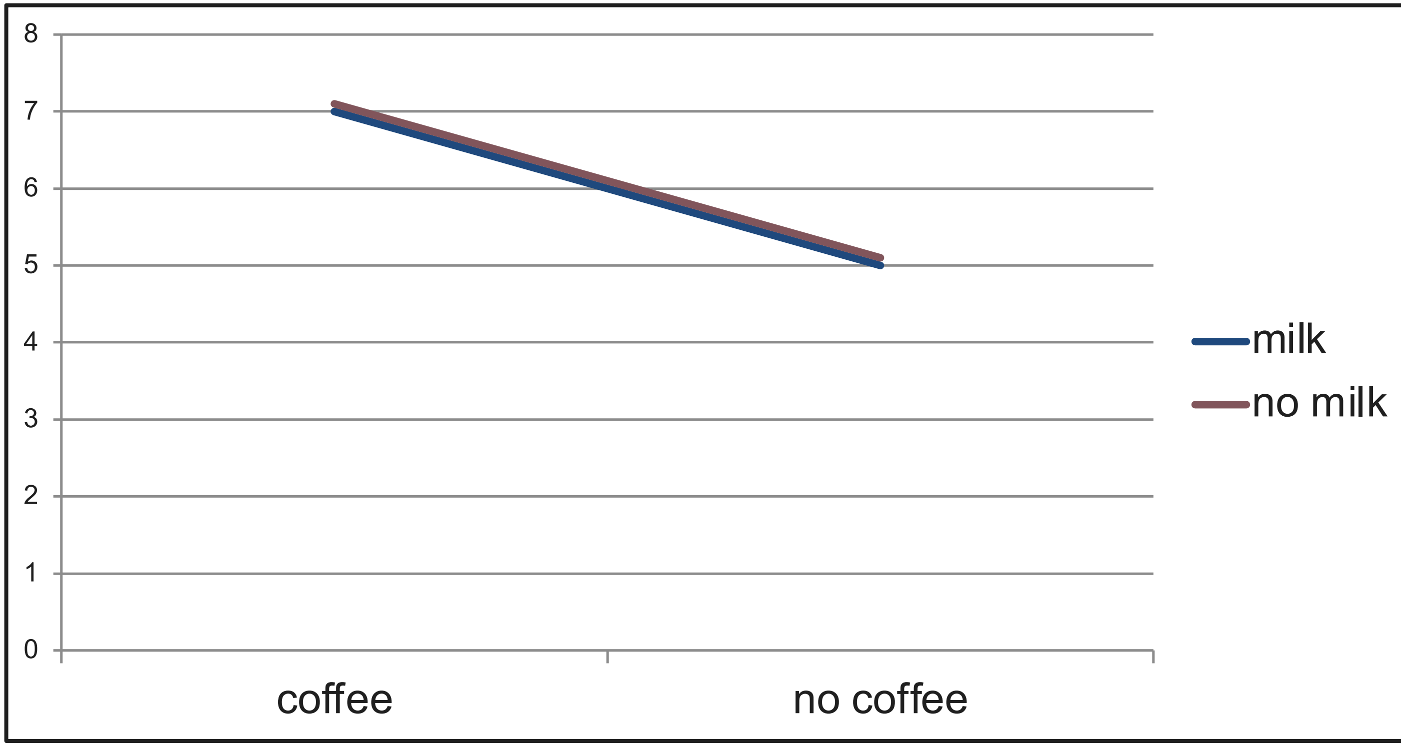
INTERPRETATION:

-Main effect Factor milk.

-Main effect Factor coffee.

-No interaction effect between factors milk-coffee (because the lines don‘t cross each other)

Two-way ANOVA. Main and interaction effects

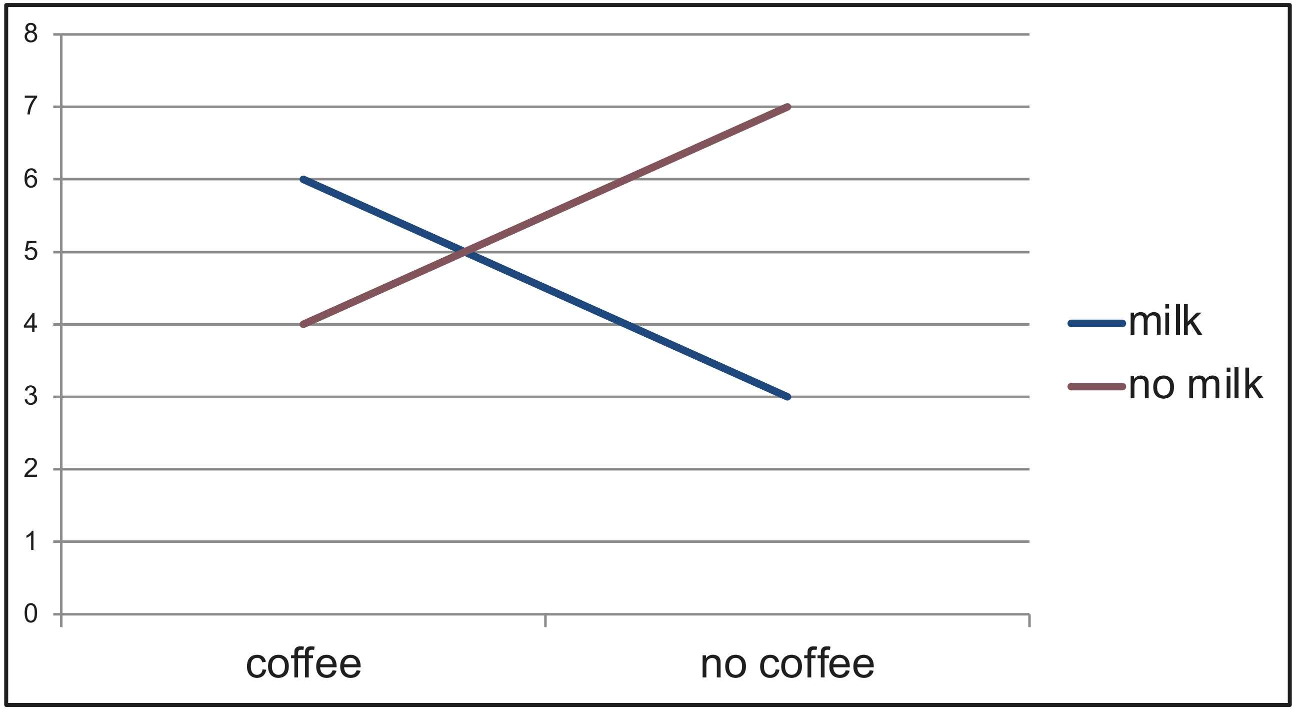


Milk factor: are the lines ends apart from each other? NO → No main effect Factor milk.

Coffee factor: are the lines inclined indicating a diff between the coffee factor levels? YES → Main effect Factor coffee.

Interaction effect: do the lines cross each other or are about to? NO
→ No interaction effect between factors milk-coffee

Two-way ANOVA. Main and interaction effects

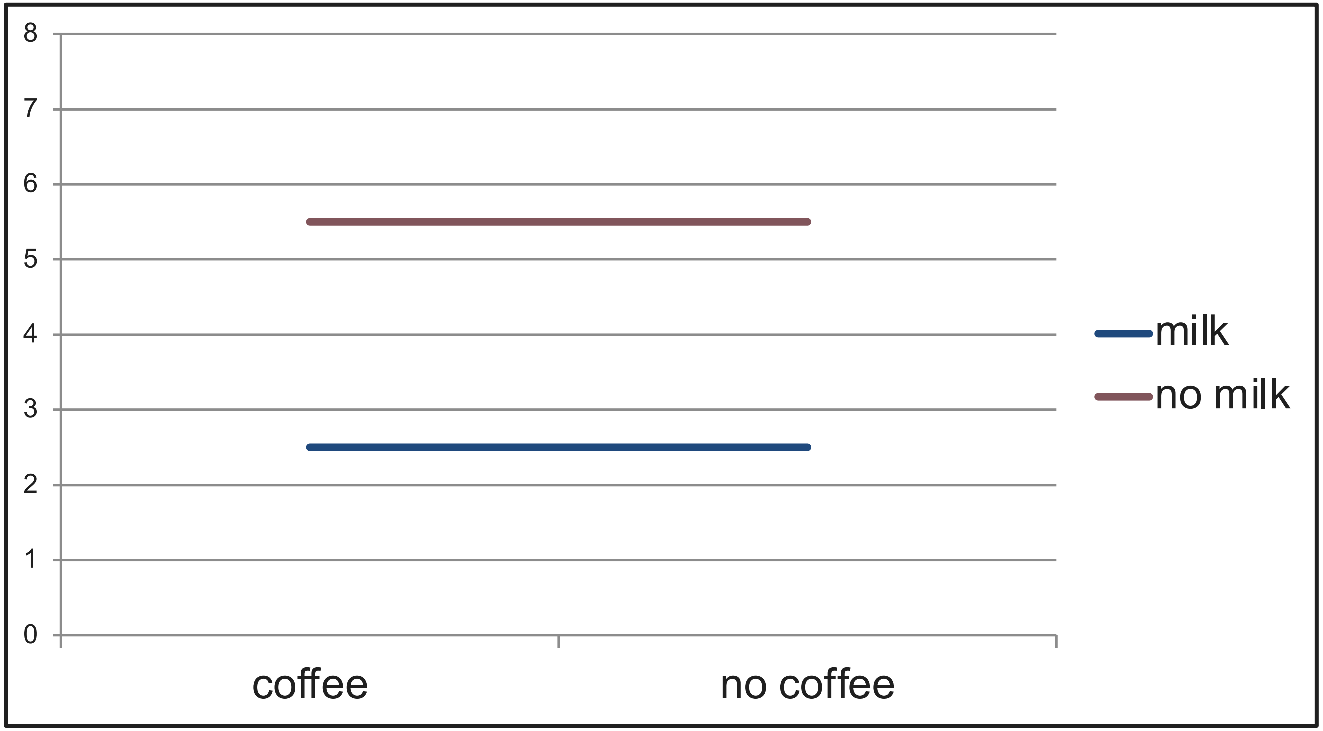


Milk factor: are the lines ends apart from each other? YES → Main effect Factor milk.

Coffee factor: are the lines inclined indicating a diff between the coffee factor levels? YES → Main effect Factor coffee.

Interaction effect: do the lines cross each other or are about to? YES
→ Interaction effect between factors milk-coffee

Two-way ANOVA. Main and interaction effects

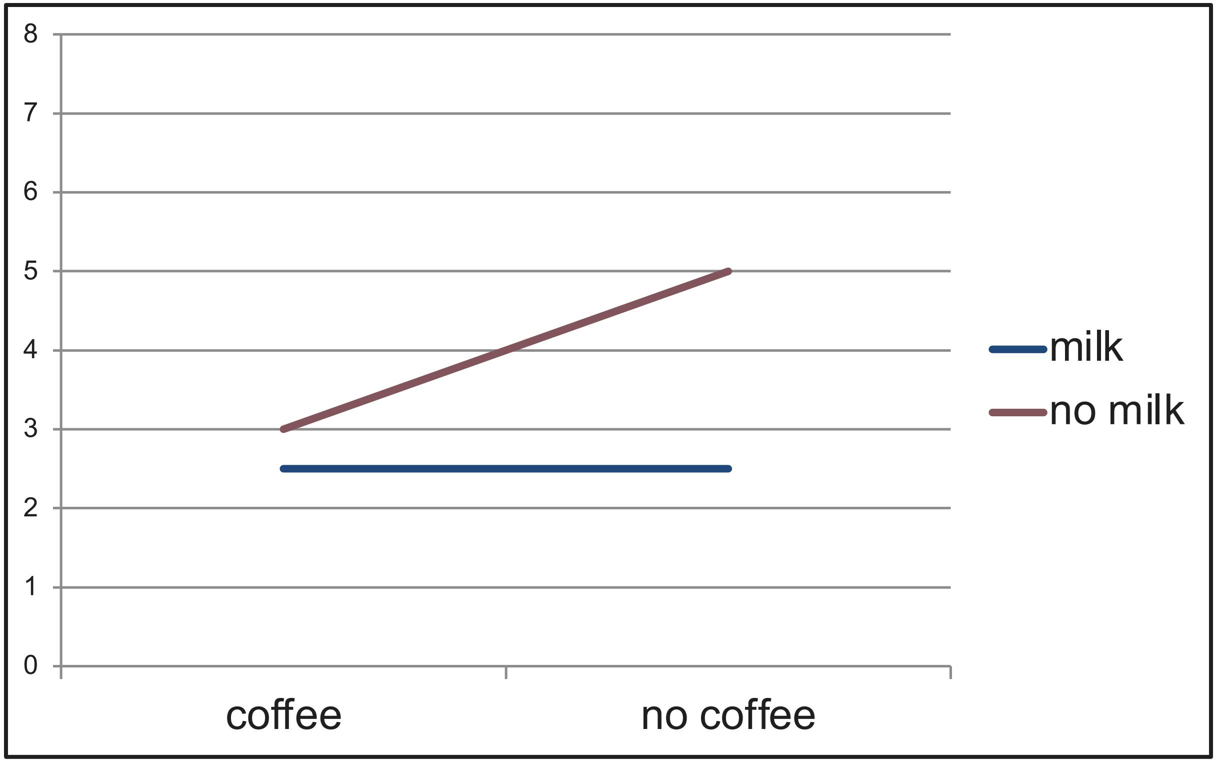


Milk factor: are the lines ends apart from each other? YES → Main effect Factor milk.

Coffee factor: are the lines inclined indicating a diff between the coffee factor levels? NO → No main effect Factor coffee.

Interaction effect: do the lines cross each other or are about to? NO → No interaction effect between factors milk-coffee

Two-way ANOVA. Main and interaction effects



Milk factor: are the lines apart from each other? Kinda → Posible main effect Factor milk.

Coffee factor: are the lines inclined indicating a diff between the coffee factor levels? Kinda → Possible main effect Factor coffee.

Interaction effect: do the lines cross each other or are about to? Almost→ Very likely interaction effect between factors milk-coffee
...that‘s why we always check the results table

Two-way ANOVA. Main and interaction effects

Ok, all these plots are nice and all but what if we have a plot like the last one where it's harder to tell if there may be an interaction effect? Or what if at least 1 of our factors has 3 or more levels?
For that and in general, we also rely on the statistics that SPSS will output after doing this test. These statistics will give us p-value certainty about the existence of main and interaction effects.
We will learn about these main and interaction effects statistics later in this topic.

As mentioned earlier, these plots serve to have a quick reference about the presence of such effects, and also they look cool in your publication / final report 😎. They can also be useful to detect trends or if any of our non-significant effects may "approach significance".

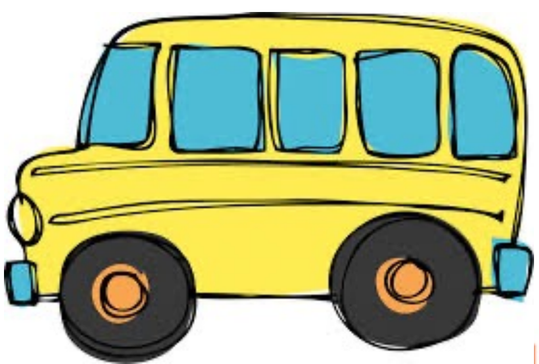
The ANOVA is an "Omnibus" test

- This means that in an ANOVA, a "significant" result indicates that there is a statistically significant difference… somewhere. At least two levels/groups differ significantly from each other.

- What remains unclear is: Where exactly is the difference? Which groups differ significantly from each other? **For that, we need to do a post-hoc analysis.**

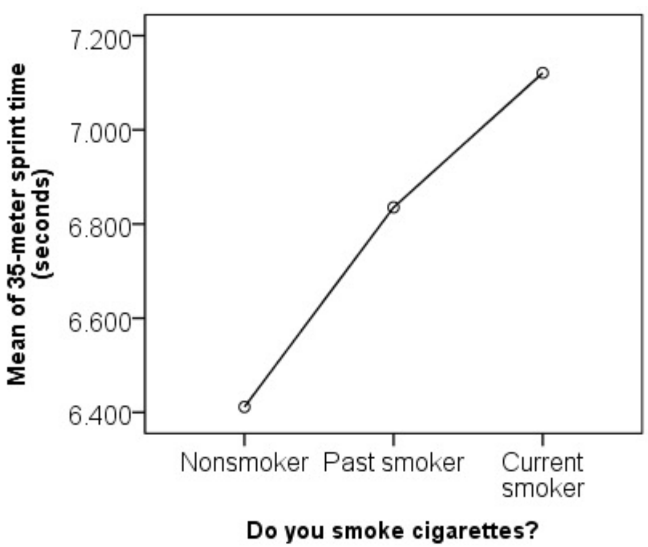
Which other test that we have seen previously could also fit into the definition of an omnibus test?

If the ANOVA is not significant, a post-hoc test is not needed.
Similarly, if you only have two factor levels, a post-hoc test may not be needed because the difference can only be between these two levels (→ you can see it from the graph or the descriptive data)



A priori and post-hoc ANOVA tests

- A priori / planned contrasts
  - If you have specific hypotheses, you can check them with contrasts that consist of specific comparisons between pairs of groups/levels. These comparisons are specified before doing the analyses
- Post-Hoc
  - Used when we don’t have any hypotheses regarding differences between specific pairs of groups/levels. Also used when we find something interesting in your data and want to analyze it in more detail afterwards



For the exam, you don‘t need to select and apply a priori or post-hoc test, but you need to be aware that they exist and when you would need one.)

Reporting ANOVA

"Comparing post-test means, we found a small negative significant main effect of individual preparation on knowledge outcomes, **F(1;124) = 5.121; p = .025; η_p^2 = .04**."

- F = 5.121 → F-value, result from the ANOVA
- (1;124) → df (n of groups minus 1; n of observations minus n of groups)
- p = .025 → result from the significance test
- η_p^2 = .04 → Effect size (next week)

Tsovaltzi, D., Judele, R., Puhl, T., Weinberger, A. (2015). Individual preparation and argumentation scripts in social networking sites. In O. Lindwall, P. Häkkinen, T. Koschman, P. Tchounikine & S. Ludvigsen (Eds.). *Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015, Volume 2.* Gothenburg, Sweden: The International Society of the Learning Sciences, pp. 667-668.

ANOVA recap

- ANOVAs: belong to the F family of tests, therefore they produce a F-value. They are parametric tests normally used for determining if, based on their means, >2 samples are statistically significantly different
- One-way ANOVA: used when you have **only 1 factor/IV** with >2 levels and **only 1 DV**
- Two-way ANOVA: a.k.a. “Factorial ANOVA”, used when you have **>1 factor** (with multiple levels each) and **only 1 DV**

TWO-WAY ANOVA IN SPSS

Assumptions in detail ● In SPSS

Two-way ANOVA. Assumptions in detail

Assumptions			
Num.	**Name**	**Description**	**How to check for that?**
1	Categorical independent variables	Two or more IVs which are on a nominal/categorical scale level.	Ensure the data from your independent variables corresponds to a nominal/categorical scale level
2	Continuous dependent variable	(Single) Dependent variable is either ratio or interval level.	Ensure the data from your dependent variable corresponds to a ratio or interval level
3	Independence of observations	The measurement of each observation was independent and didn't affect the measurement of the other observations.	When collecting the data, ensure each collected/measured datapoint is independent from each other
4	No outliers	No outliers on any of the levels of the IVs regarding the DV. Removal or correction of outliers.	Create boxplots for each level of your IVs measuring your DV. Visual inspection of the boxplots, ensure there are no indvidual outlier datapoints. If there are, exclude them from the analysis.
5	Normality	Data from the dependent variable should follow a normal distribution curve.	Shapiro-Wilk test for the dependent variable. Test should be not-significant!
6	Homogeneity of variances	The variances of the 2 compared groups are statistically the same, if not very similar.	Levene's test for equality/homogeneity of variances. Test should be not-significant! Test is part of the output of the ANOVA analysis.

1-The 1-way ANOVA has the same assumptions except on assumption 1 where only one IV (with 2 or more levels) can be used.
2-Regarding assumption 5, the ANOVA test is fairly robust so as long as the data resembles or "approximates" to a normal distribution, this assumption can be met.

Two-way ANOVA in SPSS

Example:

- Research Context: While researching about the impact multimedia content and its influence on helping memorize adjectives, we want test if the sex of the participants and the condition they were assigned to have an effect on the amount of remembered negative adjectives

a) HA1: There is an interaction effect between gender (sex) and condition on the number of remembered negative adjectives (negative).

b) HA2: There is a main effect of gender (sex) on the number of remembered negative adjectives (negative).

c) HA3: There is a main effect of condition (condition) on the number of remembered negative adjectives (negative).

*Remember that for each of these HAs, there is a corresponding H0

- Sample and procedure: We recruit 150 people who are randomly allocated onto 3 conditions. As part of the demographic items, we ask them for their sex.

- Variables and levels: Factor 1 “condition” with 3 levels (“structural”, “pictorial”, and ”emotional”). Factor 2 “sex” with 2 levels (“male”, and “female”)

- Before doing any analyses, we decide to follow the statistical conventions of our scientific field and we set our alpha to 5%

Note: dataset taken and adapted from http://quantitative-methoden.de/Downloads_A3.htm

Two-way ANOVA in SPSS. Why not one-way?

- We have 3 hypotheses: 2 main effect ones, 1 interaction effect one
  - Therefore, we have 2 IVs (a.k.a. factors)
  - One-way ANOVA is only when you have 1 IV. An one-way ANOVA can't calculate interaction effects

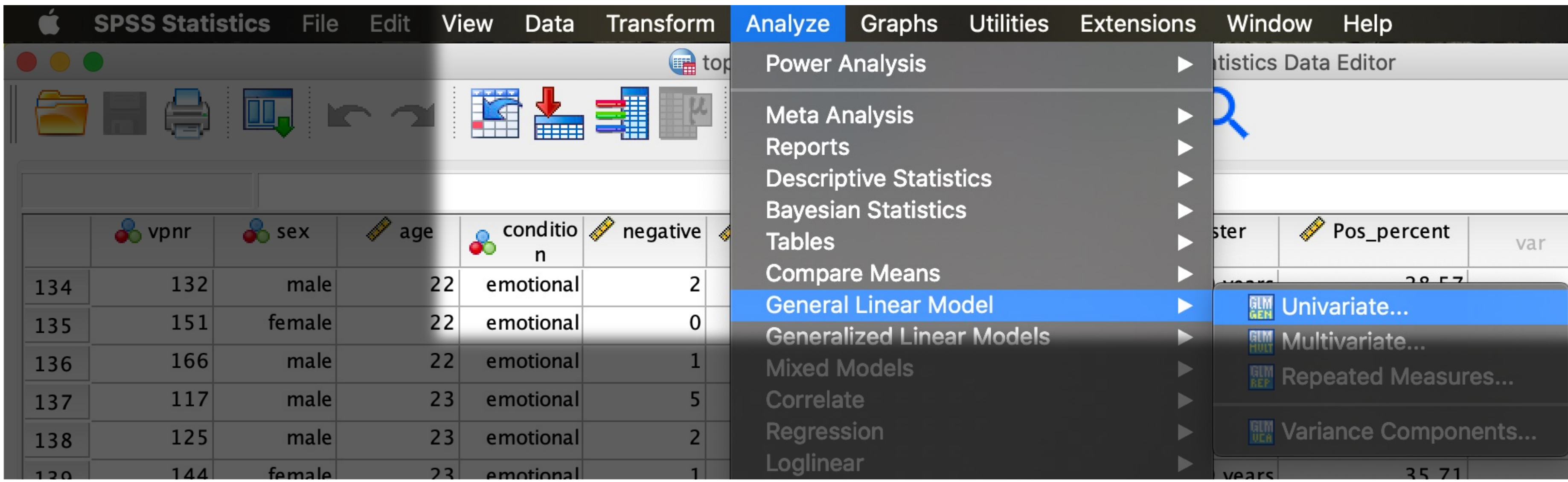
Two-way ANOVA in SPSS. Checking assumptions

Assumptions			
Number	**Name**	**How to check for that?**	**Assumtion met?**
1	Categorical independent variable	Ensure the data from your independent variables corresponds to a nominal/categorical scale level	Yes
2	Continuous dependent variable	Ensure the data from your dependent variable corresponds to a ratio or interval level	Yes
3	Independence of observations	When collecting the data, ensure each collected/measured datapoint is independent from each other	Yes
4	No outliers	Visual inspection of the boxplots, ensure there are no indvidual outlier datapoints. If there are, exclude them from the analysis.	Yes
5	Normality	Shapiro-Wilk test for the dependent variable. Test should be not-significant!	Yes
6	Homogeneity of variances	Levene's test for equality/homogeneity of variances of the 2 levels of the dependent variable. Test should be not-significant! Test is part of the output of the indep. t-test analysis.	To be checked along with the ANOVA output

The used data meets all the assumptions but for brevity purposes, contents about how to check for assumptions 1 to 5 won't be shown here.

Two-way ANOVA in SPSS

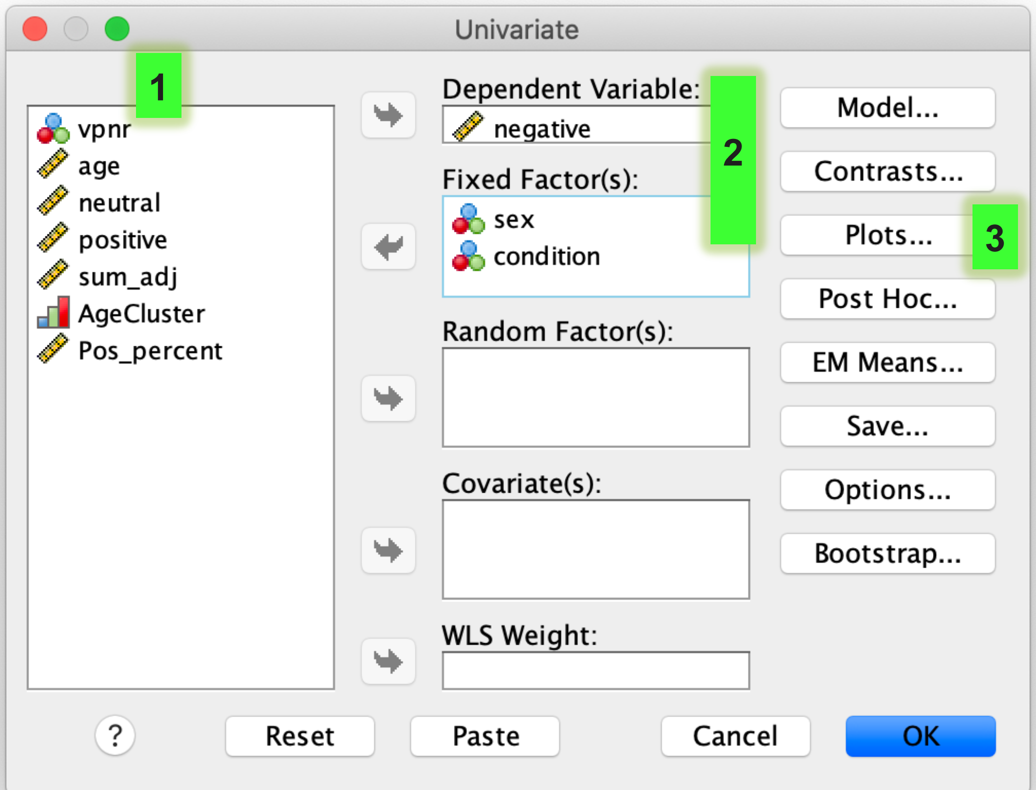
Analyze > General Linear Model > Univariate



Want to try it out? Use the file "topic-9.ANOVA.sav" provided in the moodle

Two-way ANOVA in SPSS

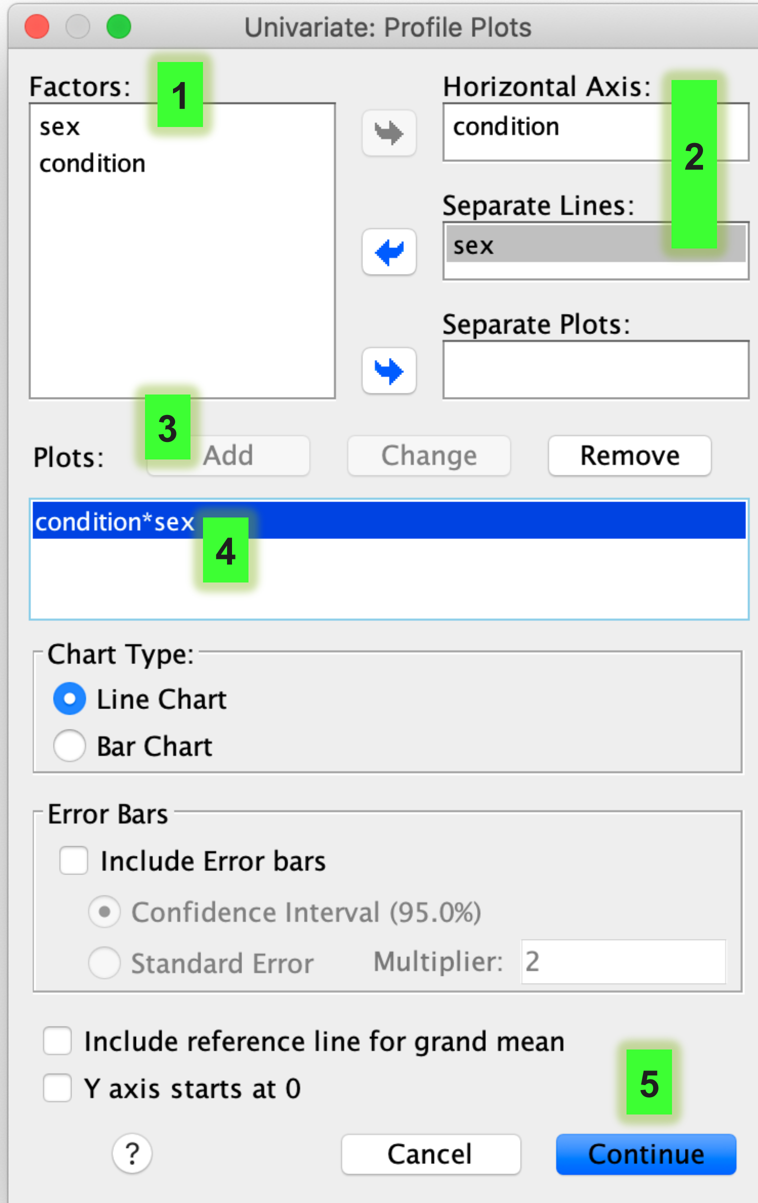
In the “Univariate” window, allocate the necessary variables from the left box [1] to the corresponding box in the right side [2]. Once completed, click on the “Plots” button [3]



Two-way ANOVA in SPSS

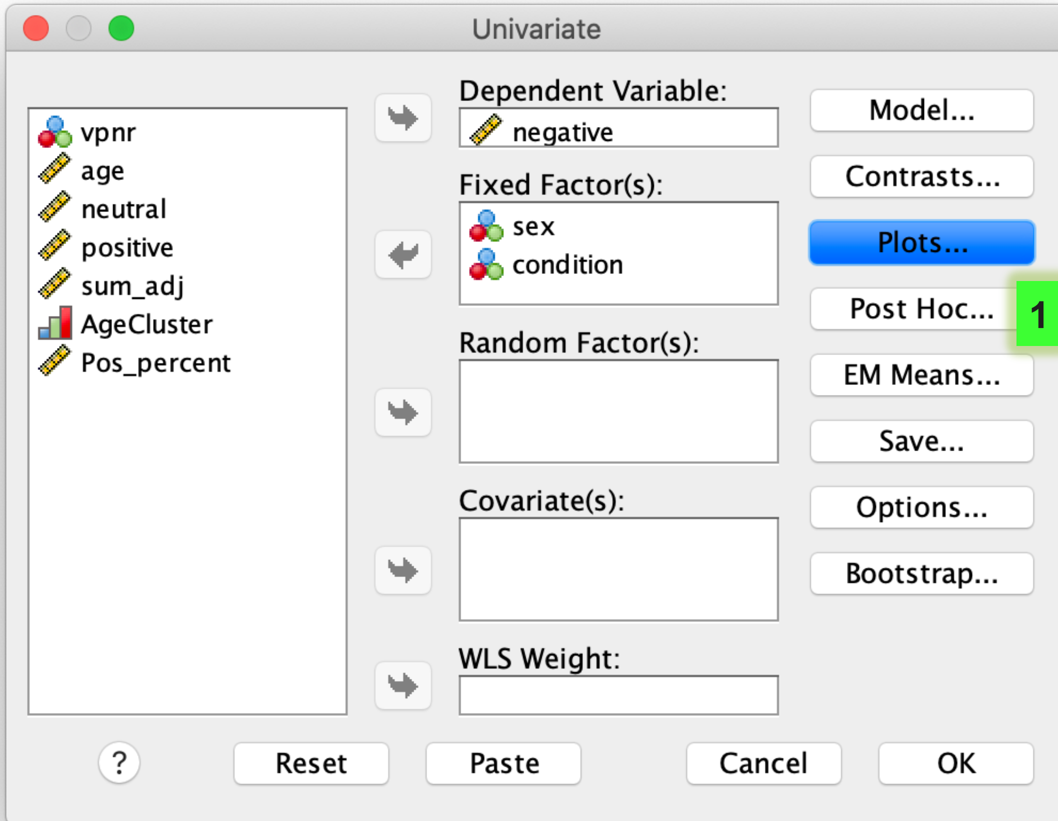
In the “Univariate: Profile plots” window, allocate the necessary variables from the left box [1] to the corresponding box in the right side [2]. Once completed, click on the “Add” button [3] which will show the to-be-created plot in the box below [4]. Ensure the other options are defined as shown.
Click on the “Continue” button [5] to close the window and go back to the main “Univariate” window.

Note: for extra fun, you can also add a “sex*condition” plot



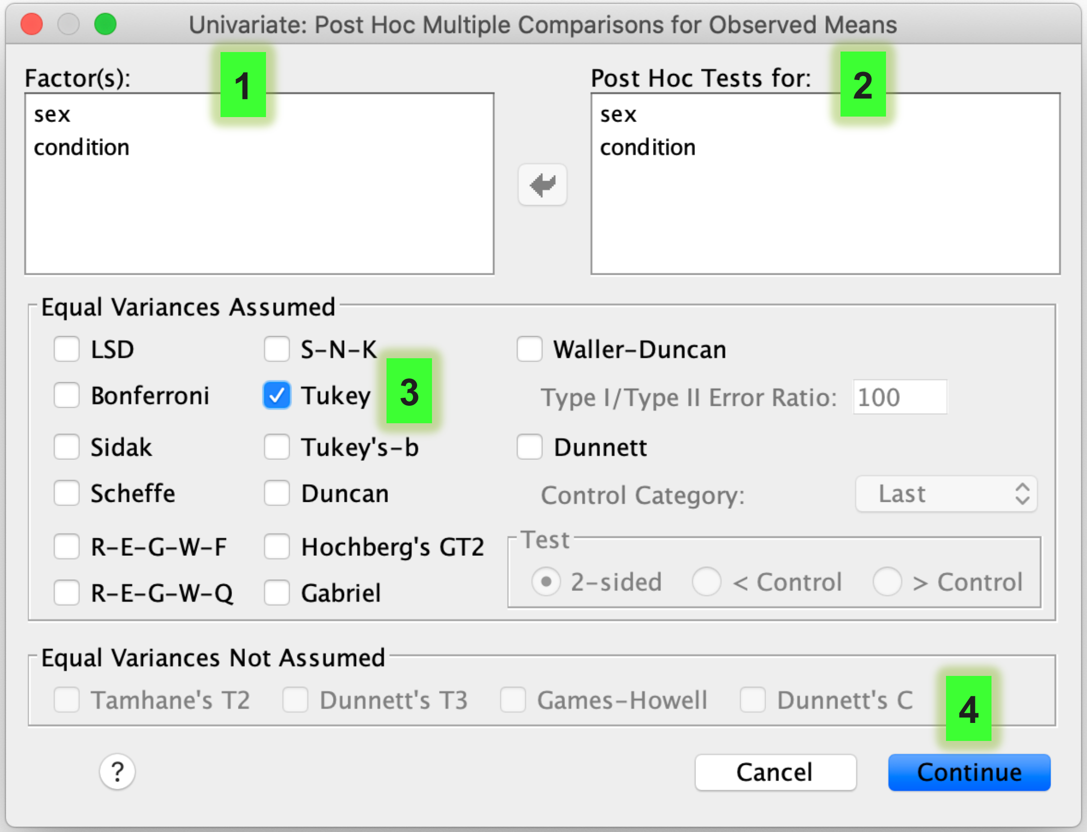
Two-way ANOVA in SPSS

Once back in the main "Univariate" window, click on the "Post-Hoc" button



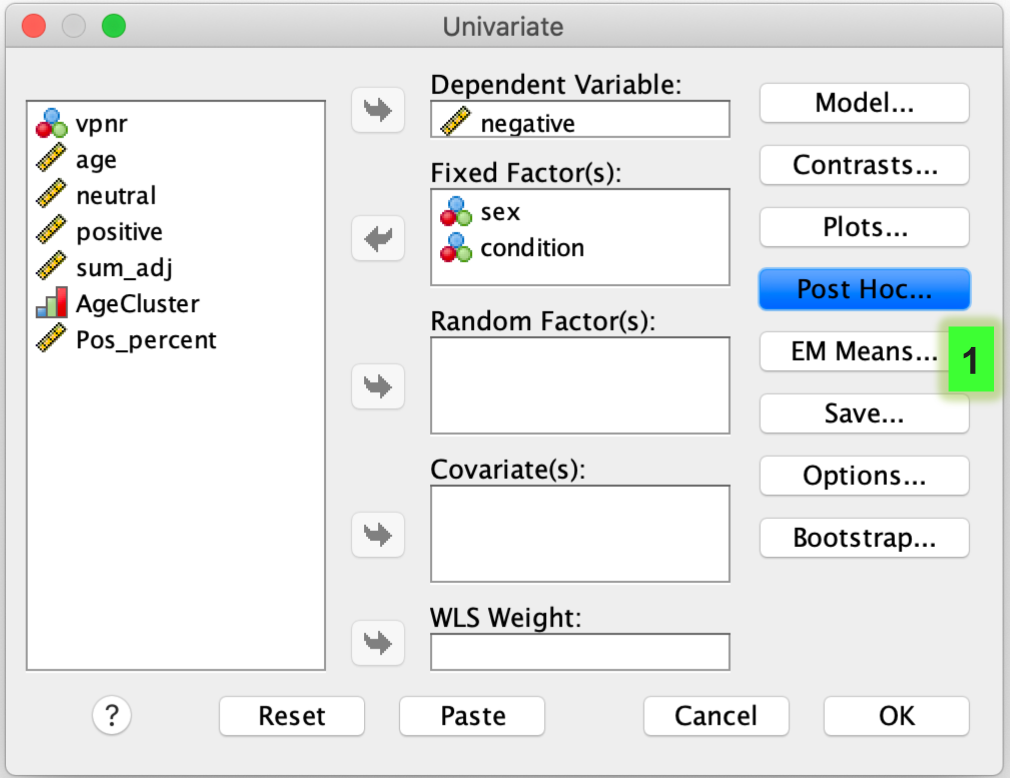
Two-way ANOVA in SPSS

In the “Univariate: Post-Hoc Multiple Comparisons for Observed Means” window, allocate the necessary variables from the left box [1] to the corresponding box in the right side [2]. Once completed, mark the “Tukey” checkbox [3]. Ensure the other options are defined as shown.
Click on the “Continue” button [4] to close the window and go back to the main “Univariate” window.



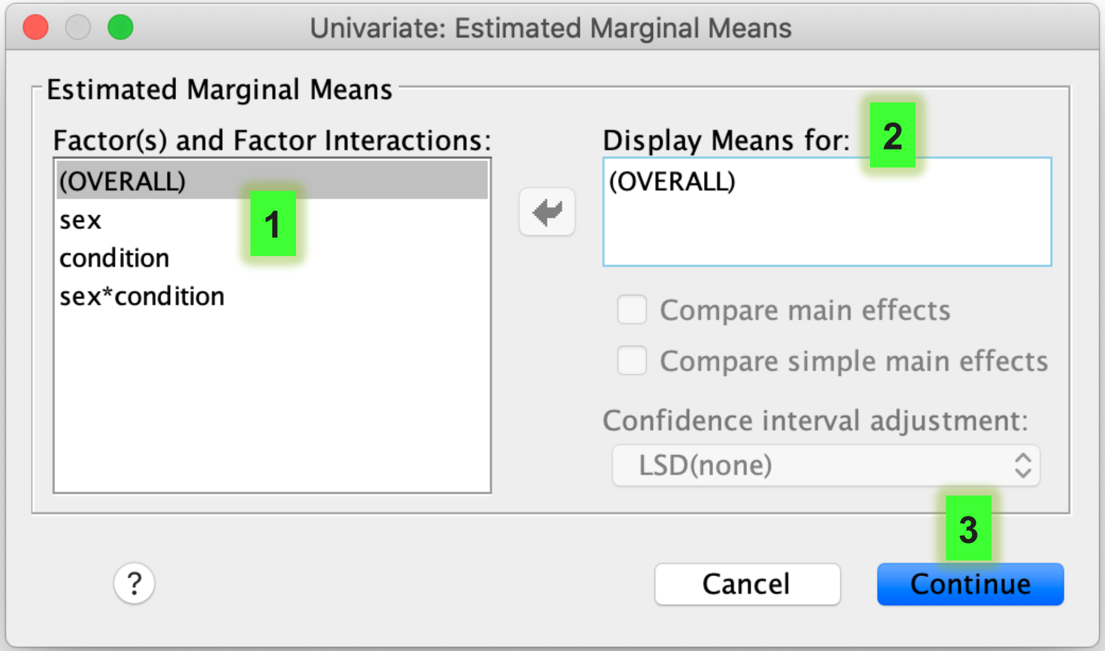
Two-way ANOVA in SPSS

Once back in the main “Univariate” window, click on the “EM Means” button



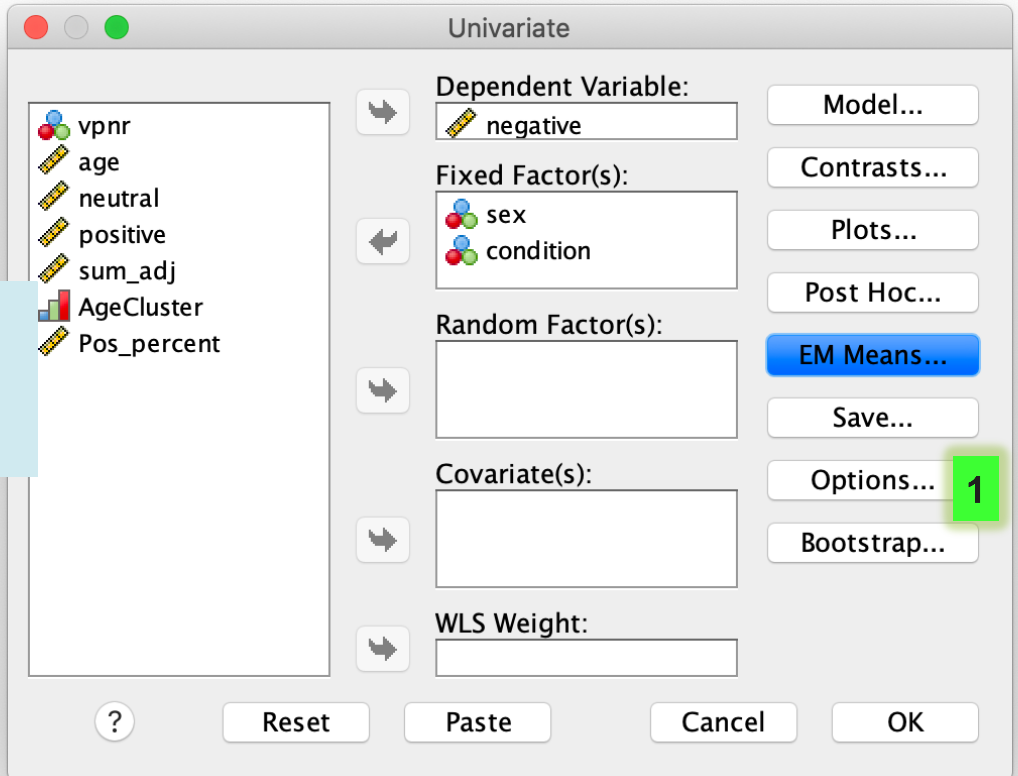
Two-way ANOVA in SPSS

In the “Univariate: Estimated Marginal Means” window, allocate only the “(OVERALL)” factor located in the left box [1] to the box in the right side [2]. Click on the “Continue” button [3] to close the window and go back to the main “Univariate” window.



Two-way ANOVA in SPSS

Once back in the main "Univariate" window, click on the "Options" button

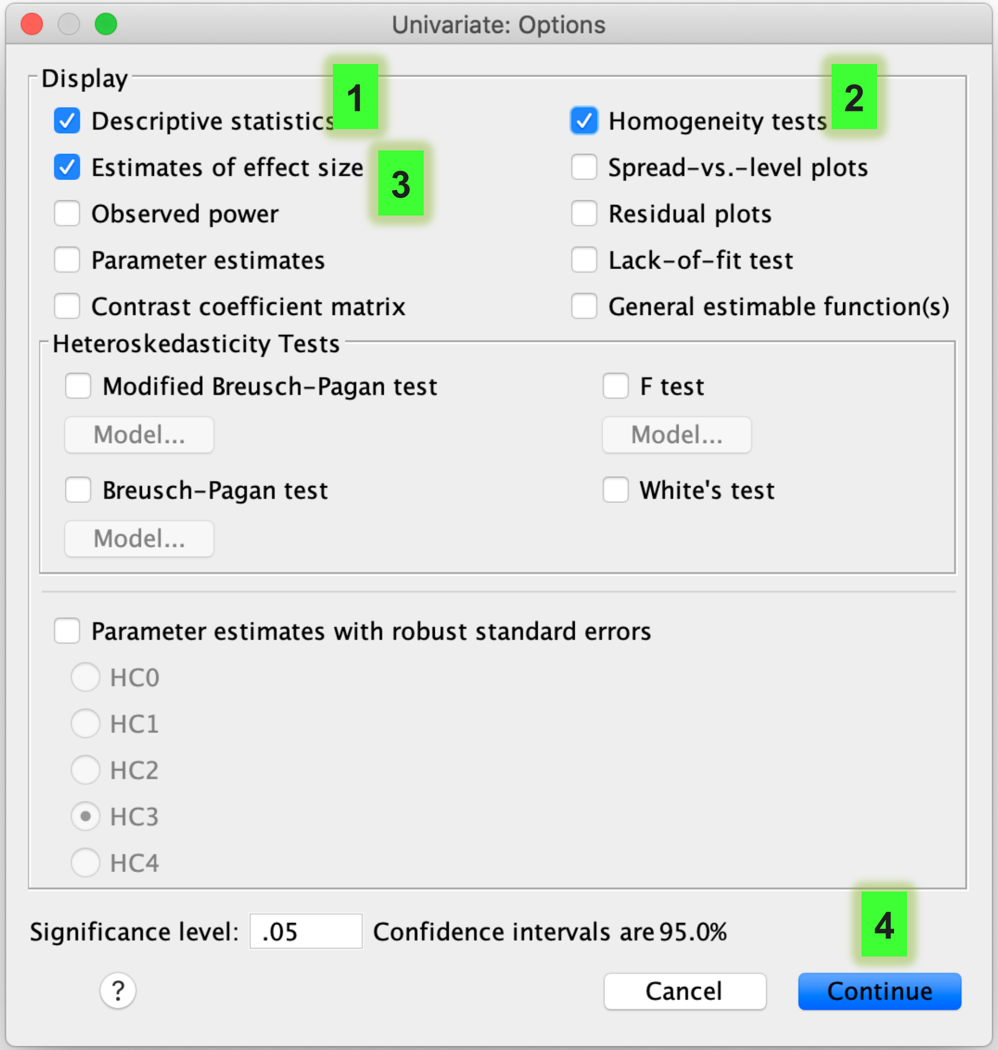


Two-way ANOVA in SPSS

In the "Univariate: Options" window, mark the indicated checkboxes [1], [2], and [3].

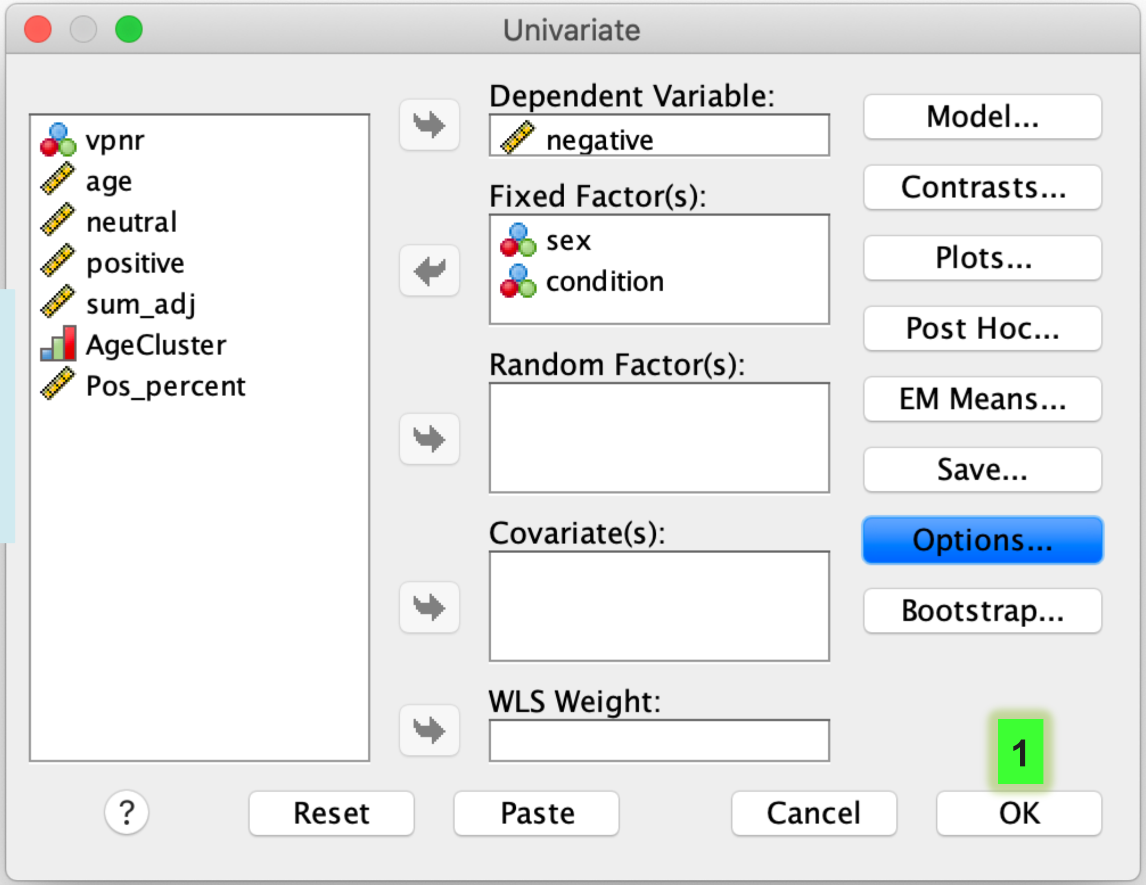
Leave the rest unchecked.

Click on the "Continue" button [4] to close the window and go back to the main "Univariate" window.



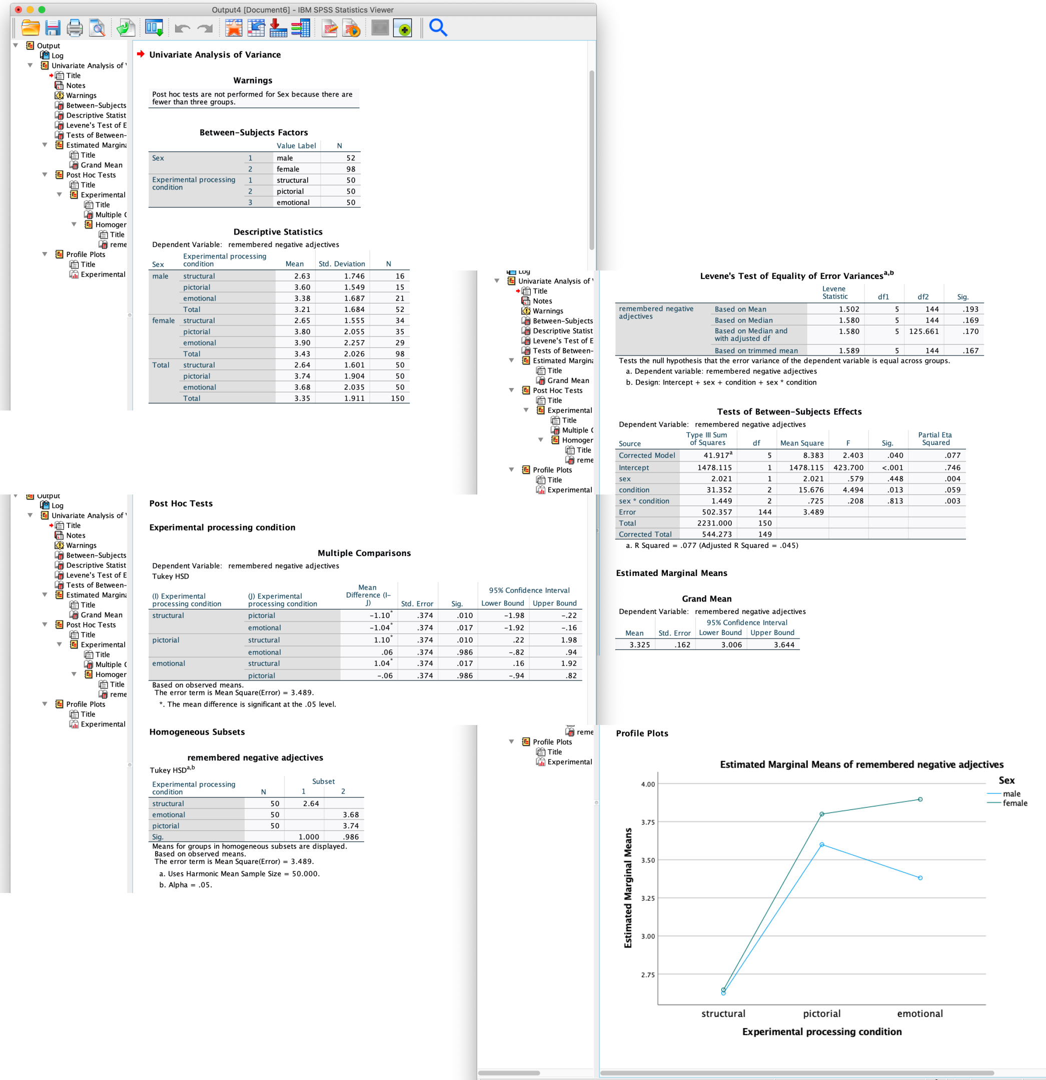
Two-way ANOVA in SPSS

Once finally back in the main "Univariate" window, click on the "OK" button to start the analysis.



Two-way ANOVA in SPSS. Output

The output of this tests consists of a lot of tables and 1 graph. We will review them one by one in order of appearance.



Output4 [Document6] - IBM SPSS Statistics Viewer

Univariate Analysis of Variance

Warnings

Post hoc tests are not performed for Sex because there are fewer than three groups.

Between-Subjects Factors

		Value Label	N
Sex	1	male	52
	2	female	98
Experimental processing condition	1	structural	50
	2	pictorial	50
	3	emotional	50

Descriptive Statistics

Dependent Variable: remembered negative adjectives

Sex	Experimental processing condition	Mean	Std. Deviation	N
male	structural	2.63	1.746	16
	pictorial	3.60	1.549	15
	emotional	3.38	1.687	21
	Total	3.21	1.684	52
female	structural	2.65	1.555	34
	pictorial	3.80	2.055	35
	emotional	3.90	2.257	29
	Total	3.43	2.026	98
Total	structural	2.64	1.601	50
	pictorial	3.74	1.904	50
	emotional	3.68	2.035	50
	Total	3.35	1.911	150

Levene's Test of Equality of Error Variances[a,b]

		Levene Statistic	df1	df2	Sig.
remembered negative adjectives	Based on Mean	1.502	5	144	.193
	Based on Median	1.580	5	144	.169
	Based on Median and with adjusted df	1.580	5	125.661	.170
	Based on trimmed mean	1.589	5	144	.167

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: remembered negative adjectives

b. Design: Intercept + sex + condition + sex * condition

Tests of Between-Subjects Effects

Dependent Variable: remembered negative adjectives

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	41.917[a]	5	8.383	2.403	.040	.077
Intercept	1478.115	1	1478.115	423.700	<.001	.746
sex	2.021	1	2.021	.579	.448	.004
condition	31.352	2	15.676	4.494	.013	.059
sex * condition	1.449	2	.725	.208	.813	.003
Error	502.357	144	3.489			
Total	2231.000	150				
Corrected Total	544.273	149				

a. R Squared = .077 (Adjusted R Squared = .045)

Estimated Marginal Means

Grand Mean

Dependent Variable: remembered negative adjectives

Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
3.325	.162	3.006	3.644

Post Hoc Tests

Experimental processing condition

Multiple Comparisons

Dependent Variable: remembered negative adjectives

Tukey HSD

(I) Experimental processing condition	(J) Experimental processing condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
structural	pictorial	-1.10*	.374	.010	-1.98	-.22
	emotional	-1.04*	.374	.017	-1.92	-.16
pictorial	structural	1.10*	.374	.010	.22	1.98
	emotional	.06	.374	.986	-.82	.94
emotional	structural	1.04*	.374	.017	.16	1.92
	pictorial	-.06	.374	.986	-.94	.82

Based on observed means.
The error term is Mean Square(Error) = 3.489.

*. The mean difference is significant at the .05 level.

Homogeneous Subsets

remembered negative adjectives

Tukey HSD[a,b]

Experimental processing condition	N	Subset 1	Subset 2
structural	50	2.64	
emotional	50		3.68
pictorial	50		3.74
Sig.		1.000	.986

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = 3.489.

a. Uses Harmonic Mean Sample Size = 50.000.

b. Alpha = .05.

Profile Plots

Two-way ANOVA in SPSS. Output

Warnings

Post hoc tests are not performed for Sex because there are fewer than three groups.

We can start with this friendly warning note from SPSS. The warning indicates that because the variable ”sex” has less than 3 groups (a.k.a. levels), we won’t find a post hoc test for this variable.

At this point, we don’t know if there would be a main effect for the variable “sex”, if there would be one, then we can simply assume it’s between the 2 only levels of this factor, which are “male” and “female”.

Two-way ANOVA in SPSS. Output

Between-Subjects Factors

		Value Label	N
Sex	1	male	52
	2	female	98
Experimental processing condition	1	structural	50
	2	pictorial	50
	3	emotional	50

Descriptive Statistics

Dependent Variable: remembered negative adjectives

Sex	Experimental processing condition	Mean	Std. Deviation	N
male	structural	2.63	1.746	16
	pictorial	3.60	1.549	15
	emotional	3.38	1.687	21
	Total	3.21	1.684	52
female	structural	2.65	1.555	34
	pictorial	3.80	2.055	35
	emotional	3.90	2.257	29
	Total	3.43	2.026	98
Total	structural	2.64	1.601	50
	pictorial	3.74	1.904	50
	emotional	3.68	2.035	50
	Total	3.35	1.911	150

The first 2 tables contain descriptive information (freq. counts, means, etc.) of each of the levels of our 2 factors (left table) and of the combination between factor levels (right table).

Two-way ANOVA in SPSS. Output

Levene's Test of Equality of Error Variances[a,b]

		Levene Statistic	df1	df2	Sig.
remembered negative adjectives	Based on Mean	1.502	5	144	.193
	Based on Median	1.580	5	144	.169
	Based on Median and with adjusted df	1.580	5	125.661	.170
	Based on trimmed mean	1.589	5	144	.167

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: remembered negative adjectives

b. Design: Intercept + sex + condition + sex * condition

This table contains the results of our Levene's test which is required for assumption 6 (homogeneity of variances).

Given that the p-value of the test is >.05, we can keep the H0 of the Levene's test and determine that there are equal variances.

Two-way ANOVA in SPSS. Output

Tests of Between-Subjects Effects

Dependent Variable: remembered negative adjectives

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	41.917[a]	5	8.383	2.403	.040	.077
Intercept	1478.115	1	1478.115	423.700	<.001	.746
sex	2.021	1	2.021	.579	.448	.004
condition	31.352	2	15.676	4.494	.013	.059
sex * condition	1.449	2	.725	.208	.813	.003
Error	502.357	144	3.489			
Total	2231.000	150				
Corrected Total	544.273	149				

a. R Squared = .077 (Adjusted R Squared = .045)

This table contains the core results of our Two-way ANOVA. This table will show us if there is:

1) A main effect on our “sex” independent variable
2) A main effect on our “condition” independent variable
3) An interaction effect between our “sex” and “condition” independent variables.

All these effects are in relation to our single DV “negative”.

Next slide will show which values are required to determine significance of the above-mentioned effects.

Two-way ANOVA in SPSS. Output and (partial) reporting

Tests of Between-Subjects Effects

Dependent Variable: remembered negative adjectives

	Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
	Corrected Model	41.917[a]	5	8.383	2.403	.040	.077
	Intercept	1478.115	1	1478.115	423.700	<.001	.746
A	sex	2.021	1	2.021	.579	.448	.004
B	condition	31.352	2	15.676	4.494	.013	.059
A * B	sex * condition	1.449	2	.725	.208	.813	.003
	Error	502.357	144	3.489			
	Total	2231.000	150				
	Corrected Total	544.273	149				

a. R Squared = .077 (Adjusted R Squared = .045)

A

No sign. main effect of gender (sex): $F(1;144) = .579$, $p = .448$, $\eta_p^2 = .004$

B

Sign. main effect of condition (condition): $F(2;144) = 4.494$, $p = .013$, $\eta_p^2 = .059$

A * B

No sign. interaction between sex and condition: $F(2;144) = .208$, $p = .813$, $\eta_p^2 = .003$

…on the DV “negative”

Two-way ANOVA in SPSS. Output and (partial) reporting

Interaction effect:

No sig. interaction between sex (sex) and condition (condition): $F(2;144) = .208$, $p = .813$, $\eta_p^2 = .003$

Main effects:

No sig. main effect of sex (sex): $F(1;144) = .579$, $p = .448$, $\eta_p^2 = .004$

Sig. main effect of condition (condition): $F(2;144) = 4.494$, $p = .013$, $\eta_p^2 = .059$

This means that within the factors of the “condition” variable, 1 or more of its levels should be sig. different to each other.
To see this more in detail, we need to check the post-hoc results.

Two-way ANOVA in SPSS. Output

Estimated Marginal Means

Grand Mean

Dependent Variable: remembered negative adjectives

Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
3.325	.162	3.006	3.644

Before the post hocs table, we have this table. This table contains the mean that is expected by our (ANOVA) model and it’s constraints.

We won’t report any of these values so we can ignore this table

Two-way ANOVA in SPSS. Output

Post Hoc Tests

Experimental processing condition

Multiple Comparisons

Dependent Variable: remembered negative adjectives
Tukey HSD

(I) Experimental processing condition	(J) Experimental processing condition	Mean Difference (I–J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
structural	pictorial	−1.10*	.374	.010	−1.98	−.22
	emotional	−1.04*	.374	.017	−1.92	−.16
pictorial	structural	1.10*	.374	.010	.22	1.98
	emotional	.06	.374	.986	−.82	.94
emotional	structural	1.04*	.374	.017	.16	1.92
	pictorial	−.06	.374	.986	−.94	.82

Based on observed means.
The error term is Mean Square(Error) = 3.489.

*. The mean difference is significant at the .05 level.

This is the post hoc comparisons table from our “condition” factor which has 3 levels: structural, pictorial, and emotional. Remember that we won’t have a similar post hoc table for the “sex” factor (see slide 40). Even if we would have a post hoc table for the variable “sex”, we would not revise it because no significant effect was found for that factor (see slide 44).
The mean of each level is compared to each other level mean to look for significance, hence the 6 comparisons. Half of the comparisons are the same, therefore we focus only on 3 of them.

Two-way ANOVA in SPSS. Output

Post Hoc Tests

Experimental processing condition

Multiple Comparisons

Dependent Variable: remembered negative adjectives
Tukey HSD

(I) Experimental processing condition	(J) Experimental processing condition	Mean Difference (I–J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
structural	pictorial	−1.10*	.374	.010	−1.98	−.22
	emotional	−1.04*	.374	.017	−1.92	−.16
pictorial	structural	1.10*	.374	.010	.22	1.98
	emotional	.06	.374	.986	−.82	.94
emotional	structural	1.04*	.374	.017	.16	1.92
	pictorial	−.06	.374	.986	−.94	.82

Based on observed means.
The error term is Mean Square(Error) = 3.489.
*. The mean difference is significant at the .05 level.

From these post-hoc results we can conclude that, from the IV "condition" (shown here by its label "Experimental processing condition"), the means of the levels:

- structural and pictorial are significantly different (p = .010)
- pictorial and emotional are **not** significantly different (p = .986)
- emotional and structural are significantly different (p = .017)

…in terms of the DV (remembered negative adjectives)

Two-way ANOVA in SPSS. Output

Homogeneous Subsets

remembered negative adjectives

Tukey HSD[a,b]

Experimental processing condition	N	Subset 1	Subset 2
structural	50	2.64	
emotional	50		3.68
pictorial	50		3.74
Sig.		1.000	.986

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = 3.489.

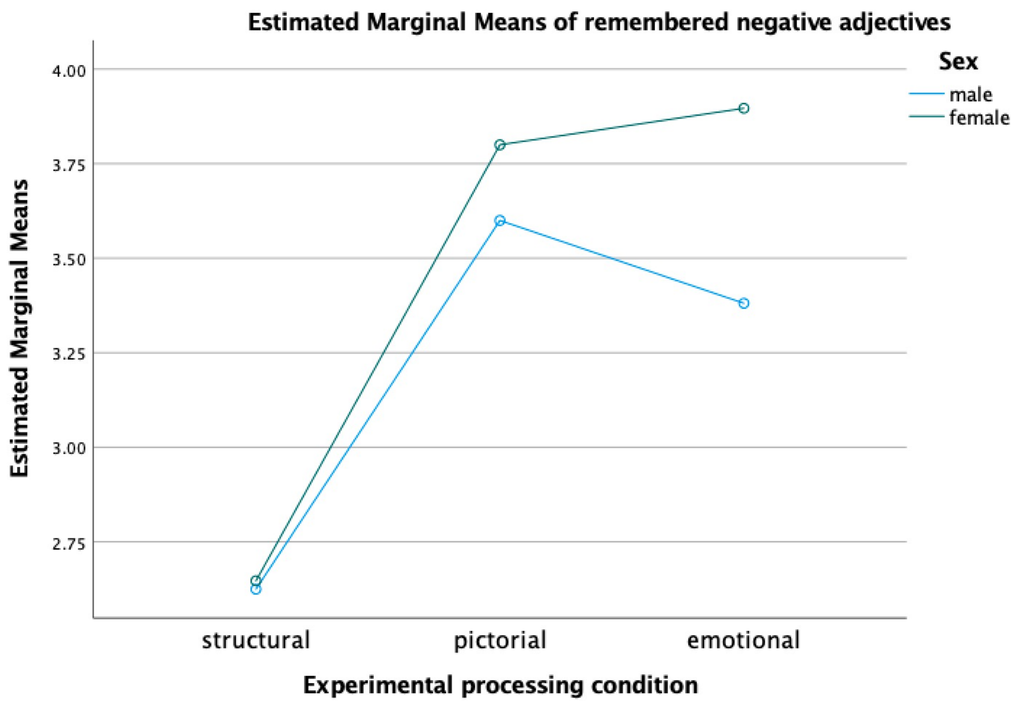
a. Uses Harmonic Mean Sample Size = 50.000.

b. Alpha = .05.

Then we have this other table, which we will ignore…

Two-way ANOVA in SPSS. Output

Profile Plots



Lastly, we have this nice plot which is a graphical representation of the main and interaction effects analyses we saw on the previous tables.
It is slightly harder to interpret compared to the effects plots we saw on slides 13 to 18 but it shows again what we have found:

- That there's no interaction effect between our 2 factors.
- No main effect on the factor "sex".
- There is a main effect on the factor "condition". From it, the level pairs "structural and pictorial" and "emotional and structural" are significantly different

Two-way ANOVA in SPSS. Output

We‘re done analyzing the results. Before moving to the next and final step, reporting, let‘s have a recap of what we did:

-As part of an experiment with a sample size of 150 participants allocated onto 3 conditions, we wanted test if the sex of the participants and the condition they were assigned to had an effect on the amount of remembered negative adjectives.

-Because we want to compare the averages (of remembered negative adjectives) of two factors, one of them having 3 levels (structural, pictorial, emotional), and the other having 2 levels (male, female), we decided to do a two-way ANOVA.
-To ensure that the results of the two-way ANOVA would be valid, we checked for the 6 assumptions of this parametric test. Five of these assumptions were checked before doing the ANOVA, the last one (homogeneity of variances) was done as part of the ANOVA results.

-We got three main results of our two-way ANOVA: (1) an interaction effect that checked if our two IVs interacted with each other, (2) a main effect for our factor “sex”, and (3) a main effect for our factor “condition”. From these 3 main results, only the main effect of our IV “condition” was shown to be significant, $F(2;144) = 4.494$, $p = .013$, $\eta_p^2 = .059$. **This means that there is a .013 probability that our “condition” main effect ANOVA result was caused by random chance**, which is above our preset alpha of .05. This means that for the three alternative hypotheses we had (slide 26), we will only be able to reject the the H0 that corresponds to HA2.

Because ANOVAs are an omnibus test and the factor “condition” has more than 2 levels, we did a post-hoc test to see which levels were statistically significant to each other. Such post-hoc test (slide 48) indicated a significance difference in the means of the level pairs “structural and pictorial” ($p = .010$) and “emotional and structural” ($p = .017$).

Is all of this clear? Is it clear from where we got all these numbers? If not, please take note of your questions or doubts and please share them on Monday when we cover this topic.

Now let’s move on and see how to properly report all of this.

Two-way ANOVA in SPSS. Output

Before moving to the final reporting, and in relation to our alternative hypotheses, now that we have analyzed our data we can reach the following decisions:

a) HA_1: There is an interaction effect between gender (sex) and condition on the number of remembered negative adjectives (negative). → Fail to reject $H0_1$

b) HA_2: There is a main effect of gender (sex) on the number of remembered negative adjectives (negative). → Fail to reject $H0_2$

c) HA_3: There is a main effect of condition (condition) on the number of remembered negative adjectives (negative). → Reject $H0_3$, accept HA_3

   a) Structural and Pictographic conditions are sig. diff. ($p = .010$)

   b) Structural and Emotional are significantly different ($p = .017$)

BTW: what would have happened if we would have had instead only 1 hypothesis for the main effects, in other words, having HA_2 and HA_3 together?

Two-way ANOVA in SPSS. Output

=Context reminder
Research Context: While researching about the impact multimedia content and its influence on helping memorize adjectives, we want test if the sex of the participants and the condition they were assigned to have an effect on the amount of remembered negative adjectives

=Actual final report
We performed a two-way ANOVA to test for a possible interaction effect between the factors “sex” and the experimental conditions (“condition”). No significant interaction between sex and condition was found, $F(2;144) = .208$, $p = .813$, $\eta_p^2 = .003$. Similarly, no significant main effect was found for the factor sex, $F(1;144) = .579$, $p = .448$, $\eta_p^2 = .004$.
However, a significant main effect was found for the factor corresponding to the experimental condition levels, $F(2;144) = 4.494$, $p = .013$, $\eta_p^2 = .059$.
As a post hoc analysis, Tukey’s HSD Test for multiple comparisons revealed that the mean value of the amount of remembered negative adjectives was significantly different between the structural and the pictorial conditions ($p = .010$, 95% C.I. = [-1.98, -0.22]). There was also a significant difference between the emotional and structural conditions ($p = .017$, 95% C.I. = [0.16, 1.92]). Lastly, there was no statistically significant difference in the mean value of the amount of remembered negative adjectives between the pictorial and emotional conditions ($p = .986$).