Statistic 101 & Dataanalysis

with SPSS

This guide is a cooperation project between:

Wilfried Grassegger, BSc
wilfried.grassegger@uni-graz.at

Maximilian Gerschütz, BSc
maximilian.gerschuetz@uni-graz.at

English translation by:

David Matischek, BSc
david.matischek@uni-graz.at

About the content of this guide

As of 01.05.2024

- This guide aims to take essential concepts of statistics that occur in BSc psychology and discuss them in a sequence that makes sense from our perspective. We focus primarily on those concepts that we believe are particularly important for a comprehensive understanding of the methodology. Please don't be put off by slides with lots of text! ☺

- In addition, the documents and knowledge from
  - „Anwendung statistischer Verfahren am Computer", as well as
  - the VOs „Psychologische Statistik" & „Forschungsmethodik und wissenschaftliches Arbeiten"

  are required.

T **This slide primarily contains theoretical notes on the respective concept**

P **This slide primarily contains information on practical application and implementation with SPSS**

Basic statistical concepts

– that everyone should know and understand

The meaning of variance

- Uniform and consistent method for quantifying scatter→ provides basic information about the **distribution of the data** and the extent to which they vary.
- **Statistical distributions:** Statistical distributions (*e.g. normal distribution)* **are defined in relation to their parameters** (mean μ und variance σ^2).
  - variance: quantifies the spread of a distribution around the mean value → essential for characterizing and understanding these distributions
- **Parameter estimation:** Variances are used to evaluate the accuracy of estimators (e.g. standard error or standard error of the mean)
- **Hypothesis test**: The variance is crucial for the construction of test statistics used in hypothesis tests (*e.g. in the t-test or F-test*)

- **Importance of variance in statistical methods *(example):***
  - **ANOVA**: Analyzes **ratios of variability** within persons and between (groups of) persons
  - **Multiple linear Regression:** R^2 = **explained variability of the dependent variable** → How well does the model fit the data

Variance (and *SD*) can also be considered as „Fit-Measures": They serve as a measure of the extent to which the mean value represents the data (→ the higher the variance, the worse the mean value represents the data)

Variance & standard deviation in the sample

- **Empirical variance:** Measure of the dispersion of the data around the mean value
  - = arithmetic mean of the squared deviation values

the multiplication with $\frac{1}{n}$ corresponds to the division by n: $\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$

$$s^2 = \frac{1}{n} * \sum_{i=1}^{n}(x_i - \bar{x})^2$$

- The values are squared, as the positive and negative deviation values would cancel each other out!
- Although squaring provides a stronger weighting of large deviations **(→ sensitivity to outliers!),** the **favorable mathematical properties of variance** (**additivity**) are a major reason for its importance in statistics

- **empirical standard deviation: (positive) square root of the variance**
  - Drawing the square root serves to transfer to the original metric → easier to interpret

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Note: Even if we usually calculate with sample data in SPSS , **SPSS always estimates the variance of the population automatically** by adapting the formula by $n - 1$ (*see next slide*)

Variance & SD in the Population

- **Sample**variance/standard-deviation is usually given as s^2/s, sometimes also as $s^2{}_{emp}\,/s_{emp.}$
  - **Population** → greek letter σ^2/σ
- We **generally do not know the variance in the population and estimate it on the basis of our sample.**

  To do this, we divide by $n-1$ based on two considerations:

  1. **Correction of the bias in the estimation of σ^2**: The calculation of the sample variance with n as divisor leads to a systematic underestimation of σ^2, especially with small n. The use of $n-1$ corrects this underestimation, as the smaller divisor leads to a higher value for σ^2.
  2. **Consideration of degrees of freedom (df):** When estimating the population variance on the basis of the sample, 1 degree of freedom (due to the already determined sample mean) is lost in the estimation → 1 value can no longer vary freely in order to achieve the sum of the squared deviations.
- **Estimated values** are often reffered to as $\hat{\sigma}^2$ or $\hat{\sigma}$.

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

The reason for the underestimation is that the probability of extreme values (which increase the variance!) occurring in a sample is significantly lower than their occurrence in the population.

Note!

Estimated values are often referred to as σ^^2 or σ^. Sometimes estimated values are referred to as s and s^2, while the sample values are reffered to as s_{emp} and s_{emp}^2.

In the following, $\hat{\sigma}^2$ and $\hat{\sigma}$ as well as s und s^2 <u>always</u> refer to estimated population values (also in SPSS)

Why do we use estimates?

- We want to make statements about population(s), but usually only draw a sample of data from population(s) → Inferential statistical methods make it possible to make statements about a population based on the sample data
- **Inferential statistics includes estimation of population parameters, construction of confidence intervals & hypothesis tests**
- We consider **probabilistic assumptions** that form the basis for the development and application of statistical models and methods for estimating population parameters.
  - By adhering to these assumptions, we can ensure the validity and reliability of our statistical conclusions
  - The assumptions relate, for example, to the underlying distributions of i.i.d. random variables and aspects of quantifying the uncertainty of statistical conclusions

In inferential statistics, most of the values resulting from the analyses are estimates or are based on estimates, as we usually do not know the true values of the population.

Standard error *(SE)*

- = a measure of the precision of an estimate
  - *e.g.: how accurately a sample mean estimates the mean of the population*

Lower Limit

Upper Limit

e.g. CI for μ: $\left[\bar{x} - t_{1-\frac{a}{2}} \cdot \sqrt{\frac{s^2}{n}}, \bar{x} + t_{1-\frac{a}{2}} \cdot \sqrt{\frac{s^2}{n}}\right]$

- = the basis for the calculation of confidence intervals
  - Depending on the type of the CI, the ***SE*** is computed differently, as ist **numerator** contains the **estimated population Variance** (s^2)!
  - The value of the $t_{1-\frac{a}{2}}$ quantile depends on the selected significance level and the sample size or the degrees of freedom of the *t*-distribution

- Influences whether a result is statistically significant in the context of **hypothesis tests**
  - *SE* in the denominator
  - → The smaller *SE*, the greater the value of the test statistic (*t*-value)

e.g. hypothesis test on a single mean value:

$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}}$$

- There are *SE* for the estimation of **various parameters/characteristic values**
  - *e.g. mean value, mean value of the differences between pairs of measured values, regression coefficients*
- The *SE* of a parameter **corresponds to the standard deviation of the sample-parameter-distribution**

Confidence intervals

As we have no other values available

- Sample parameters serve as a "best" estimate for population parameters (e.g. μ, π), but the two values will rarely match exactly
- Problem: *How accurate is the estimate based on our sample?*
- Solution: Confidence intervals for **interpreting the accuracy of an estimate**

General structure of confidence intervals (using the example of the estimation of μ):

$$CI = [U, O] = \left[\underbrace{\boxed{P} - \boxed{t/z_{1-\frac{\alpha}{2}}} * \boxed{\sqrt{\frac{\sigma^2}{n}}}}_{\text{Lower Limit}}, \underbrace{P + t/z_{1-\frac{\alpha}{2}} * \sqrt{\frac{\sigma^2}{n}}}_{\text{Upper Limit}}\right]$$

with

$\boxed{P}$: **Sample parameter** for estimating the population parameter, e.g. $\bar{x}$ (for μ or π)

$\boxed{t/z_{1-\frac{\alpha}{2}}}$: **t- or z-quantil** (depending on the parameter sought and confidence level)

$\boxed{\sqrt{\frac{\sigma^2}{n}}}$: **Standard error** (Calculation depends on the parameter searched for)

Specific confidence intervals

CI for μ

$I(x_1, \dots, x_n) = [u, o] = \left[\boxed{\bar{x}} - \boxed{t_{1-\frac{\alpha}{2}}} \cdot \boxed{\sqrt{\frac{s^2}{n}}}, \bar{x} + t_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s^2}{n}}\right]$ here t, because not standard-normally distributed

CI for π

$I(x_1, \dots, x_n) = [u, o] = \left[\boxed{\bar{x}} - \boxed{z_{1-\frac{\alpha}{2}}} \cdot \boxed{\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}}, \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$ here z, because we assume that the approximate probability distribution of Z^* für $n \to \infty$ converges to the standard normal distribution

CI for parameter differences of independent samples $\mu_1 - \mu_2$

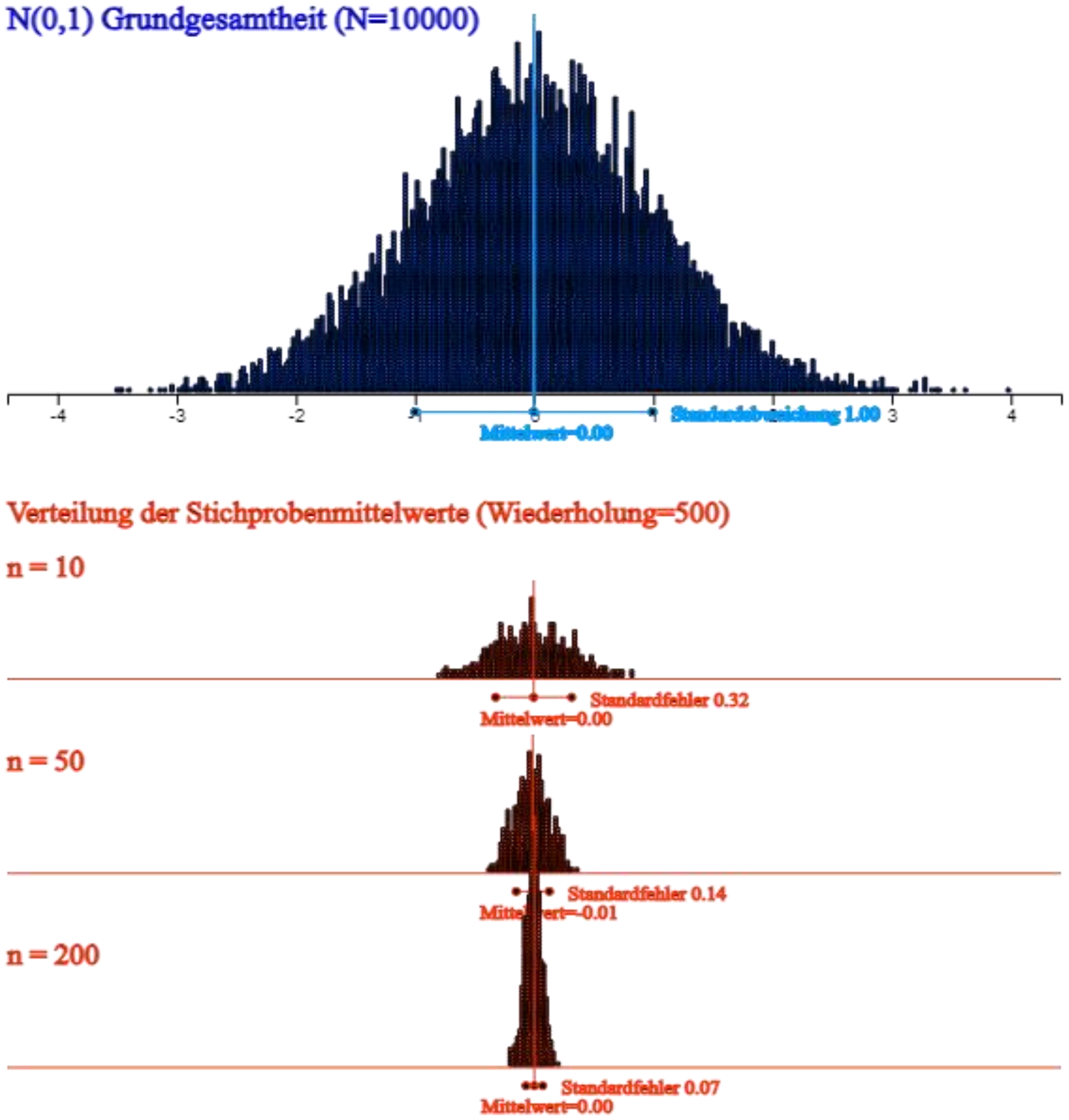
$$I(x_1, \dots, x_n) = \left[\boxed{(\bar{x}_1 - \bar{x}_2)} - \boxed{t_{1-\frac{\alpha}{2}}} \cdot \boxed{\sqrt{\frac{s_{pool}^2}{n_1} + \frac{s_{pool}^2}{n_2}}}, (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_{pool}^2}{n_1} + \frac{s_{pool}^2}{n_2}}\right]$$

CI for parameter differences of dependent samples $\mu_1 - \mu_2$

$$I(x_1, \dots, x_n) = \left[\boxed{(\bar{x}_1 - \bar{x}_2)} - \boxed{t_{1-\frac{\alpha}{2}}} \cdot \boxed{\sqrt{\frac{s_{Diff}^2}{n}}}, (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_{Diff}^2}{n}}\right]$$

Sampling distribution 1/2

- describes how a statistical parameter (e.g. mean value) behaves when it is drawn from many samples from the same population.
  - The **resulting distribution** shows which values the sample parameter can assume and how likely these different values are if **an infinite number of samples of the same size are drawn from the population**.



Upper Graphic: population with N = 10000 with μ = 0 and σ = 1.

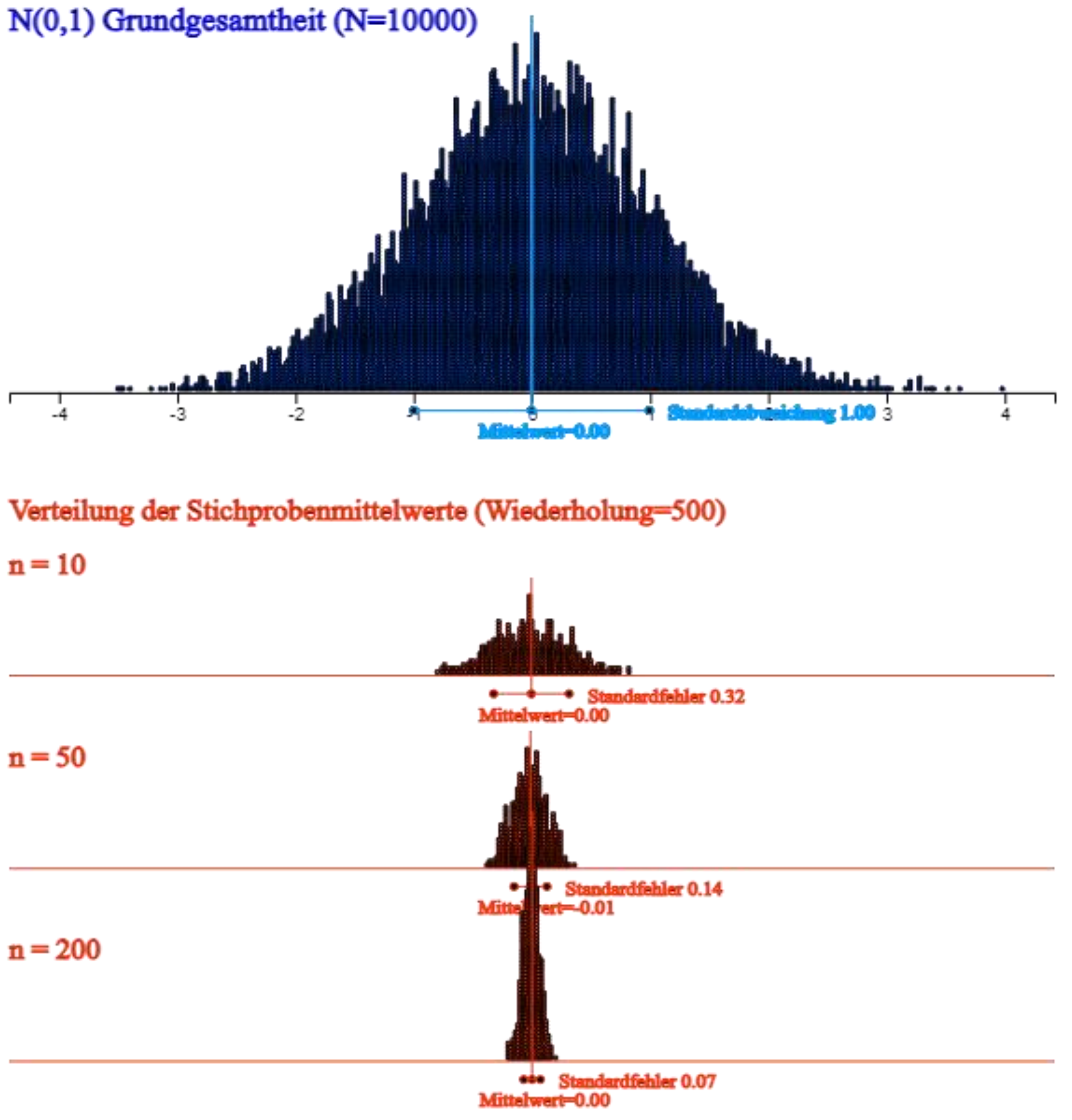
Lower Graphic: Samples of different sizes (n = 10, n = 50, n = 200) are taken 500 times and the **distribution of the mean values resulting from the samples is mapped**

As the sample size increases, the SE decreases → "more information about the population = more accurate estimate"

Graphic created with: http://www.estat.me/estat/eStatU/62SampleEng.htm

Sampling Distribution 2/2

- **Central limit theorem (CLT):** The **distribution of the mean values** of sufficiently large samples **is approximately normally distributed**, regardless of the shape of the underlying population distribution (often: $n \geq 30$)
  - Assumption: i.i.d random variables
  - CLT “justifies” the use of methods based on the normal distribution assumption.
- By understanding the sampling distribution, we can **draw conclusions about the population** and, for example, construct confidence intervals for population parameters or test hypotheses about the population.
- The sampling distribution illustrates that **different samples from the same population** can **lead to different estimates of the characteristic value of interest**.

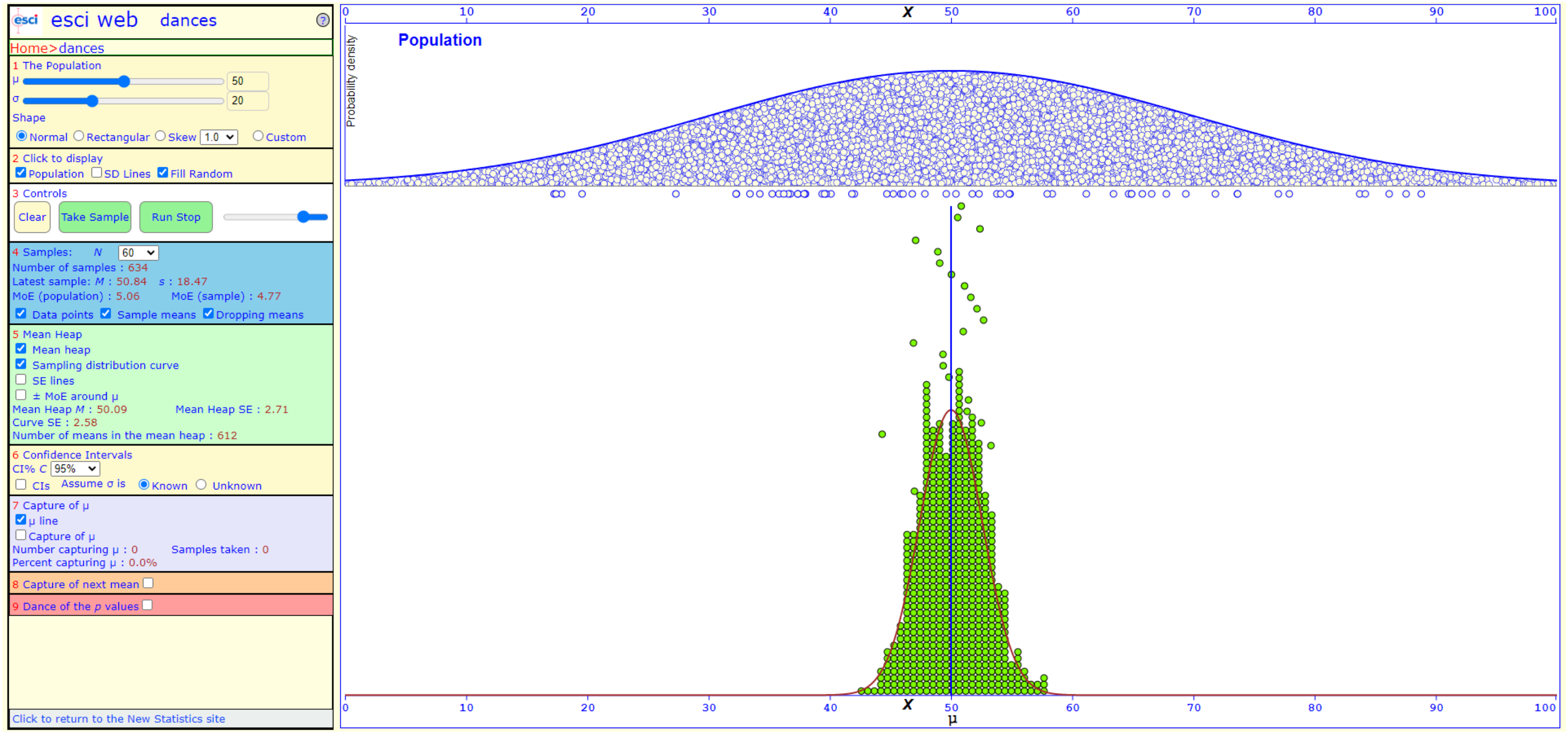


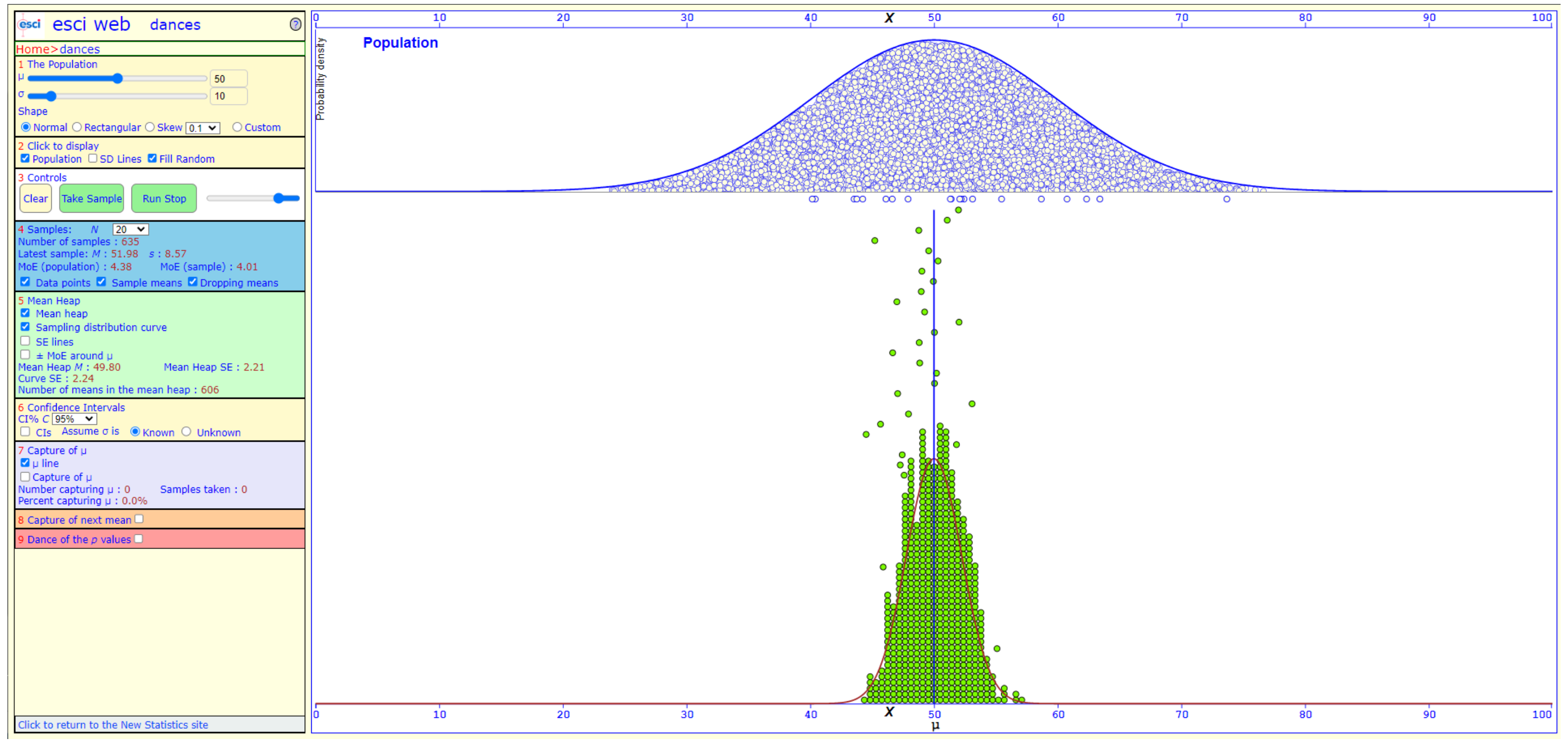
Bootstrapping is based on a similar principle in that the sample is regarded as a population and an “empirical sample distribution” is created by drawing numerous (usually at least 1000) secondary samples, on the basis of which confidence intervals or test decisions can be constructed.

Graphic created with: http://www.estat.me/estat/eStatU/62SampleEng.htm

Visualization for better understanding

- https://www.esci.thenewstatistics.com/esci-dances.html
  - This link will help you to visualize the principle of the sample characteristic value distribution, among other things





Graphic created with: https://www.esci.thenewstatistics.com/esci-dances.html

Interpreation of Confidence intervals

Don't forget: We always want to estimate values in the population!

What does **95% CI** mean?

- Assuming we were to draw 100 samples, 95 of them would include the true population value within their CI.

Valid conclusions:

- We can say that **if our sample is one of the 95 that includes the population value, then it is within the bounds of our CI**.

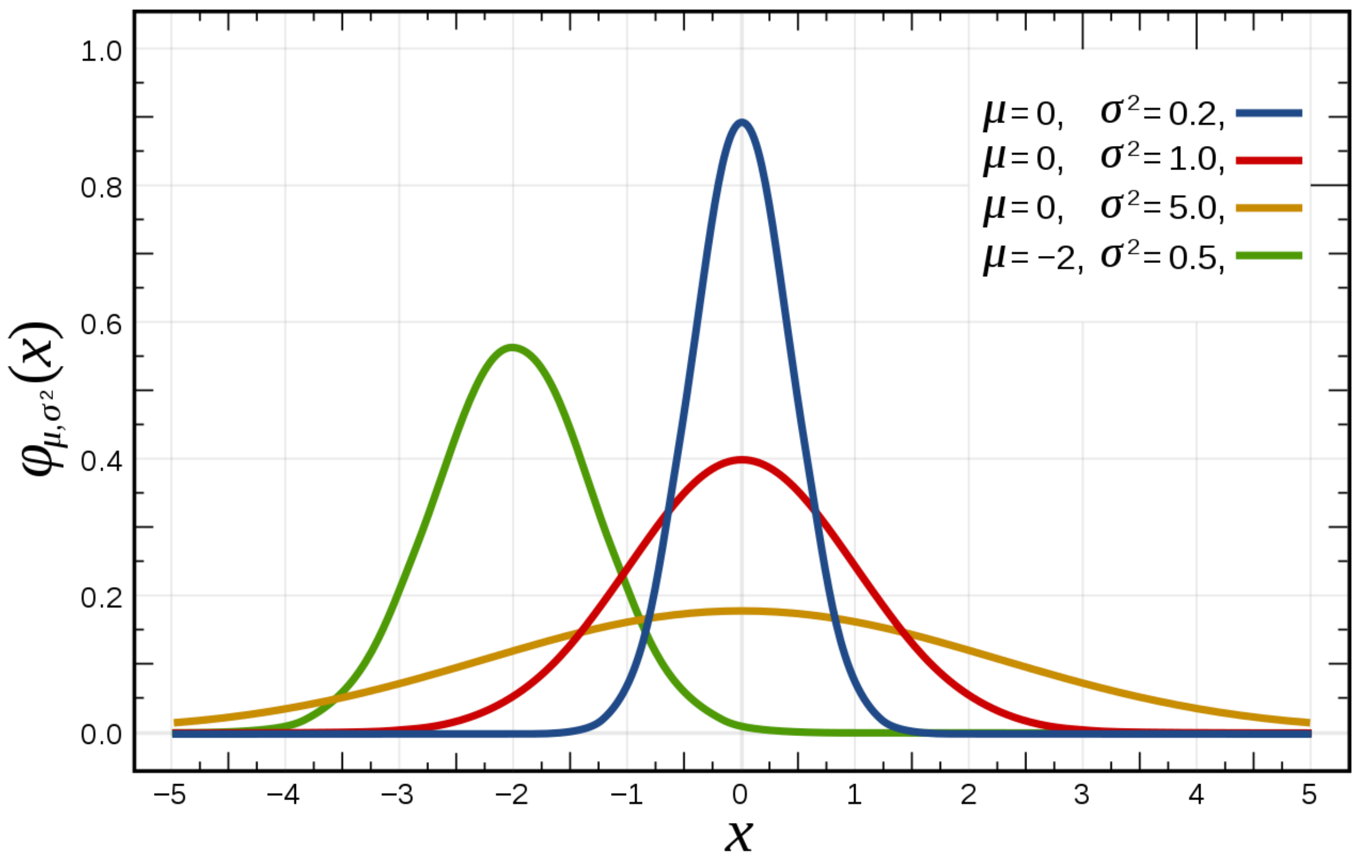
→ but we cannot say whether our CI includes the population value.

(Standard) normal distribution

- = Probability distribution in “bell shape
- Fully describable by 2 parameters:
  - **μ**: Maximum of the density function
  - $\boldsymbol{\sigma^2}$: Width” of the density function (= scattering)

→ Special case: Standard-normal distribution

- Results from z-standardization of any normally distributed variable
- The parameters of the Standard-normal distribution are **always**:
  - $\mu = 0$
  - $\sigma^2 = 1$



The Graphic shows different normal distributions.
In red: standard-normal distribution

Graphic: https://de.wikipedia.org/wiki/Normalverteilung#/media/Datei:Normal_Distribution_PDF.svg

Why is normal distribution important?

- Many statistical methods are based on the assumption of normal distribution. **The mathematical properties of the normal distribution make it possible to calculate probabilities and significances precisely.**
- Many real characteristics follow at least approximately a normal distribution, especially in biology, psychology and the social sciences
- **The central limit theorem** "justifies" the use of methods based on the normal distribution assumption, even if the characteristic to be analyzed is not normally distributed.

- Attention: The ZGS does **NOT** state that a characteristic is normally distributed in a sample of ≥ 30 persons!
- The ZGS is **ALWAYS** about the resulting sample characteristic value distribution, which we use for parameter estimation!

Covariance & Correlation 1/2

- **Covariance** = the average product of the corresponding deviation values
  - Range: $-\infty$ to $+\infty$
  - Can **only** describe **the direction** of the linear relationship, **but not the strength** and it is **unit-dependent**

$$cov_{emp}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Caution: As already noted for the variance, **SPSS estimates the covariance and correlation on the basis of the sample** by entering **n-1** in the denominator!

- **Correlation** = (z-)standardized covariance
  - Formular for z-standardisation $z_x = \frac{x-\bar{x}}{s}$
  - The empirically determined covariance is relativized to the maximum covariance (= the dispersion of the two variables)
  - Range: -1 to+1
  - $r_{xy} = 0$ does not mean that there can be no correlation, but indicates that there is no linear correlation
  - Correlation ≠ Causality

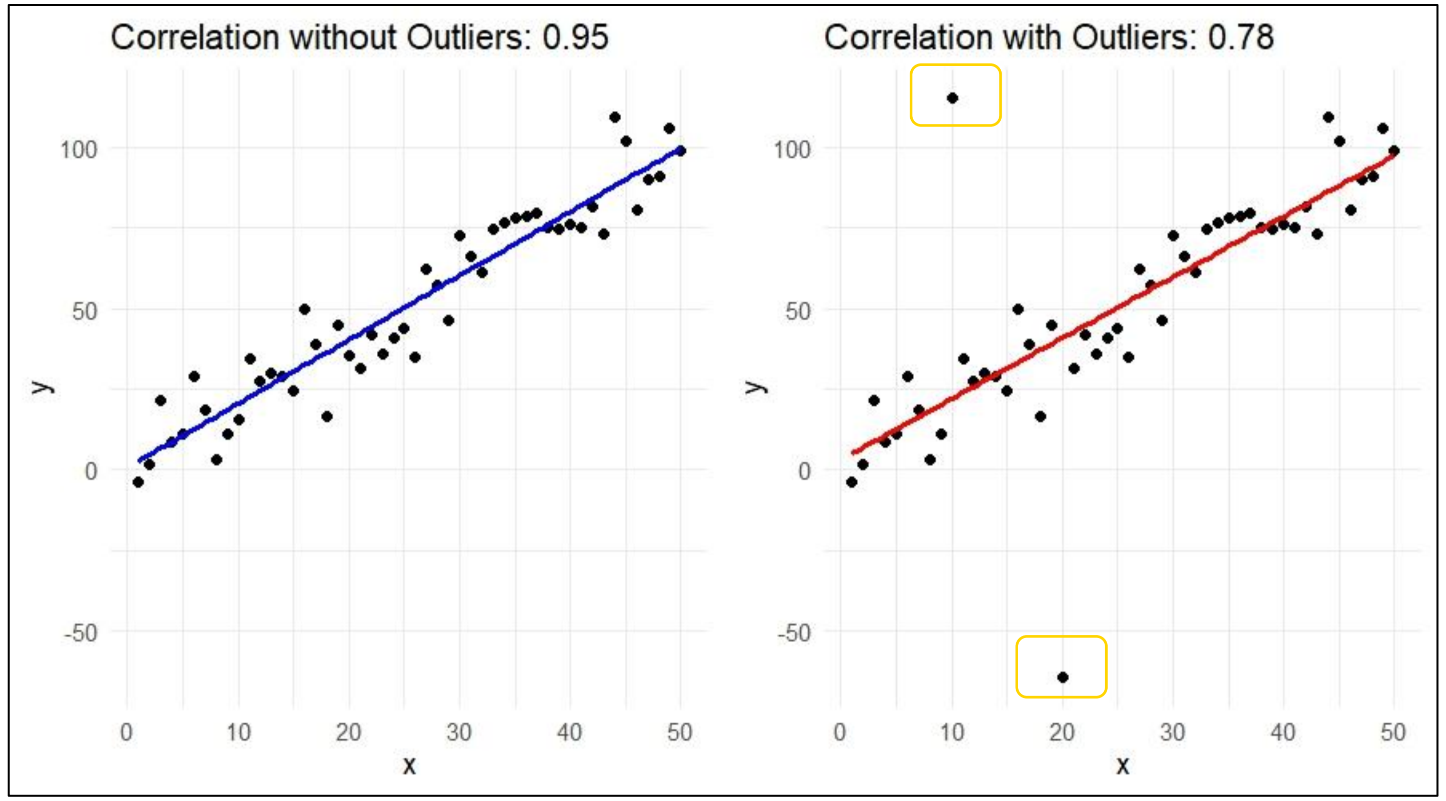
If two variables are perfectly positively linearly correlated (r of +1), this means that their empirical covariance is as large as would be possible while retaining their individual variances.

$$r(x,y) = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x\, s_y}$$

In a simple linear regression, the standardized regression weight β corresponds to the correlation coefficient of the two variables

Covariance & Correlation 2/2

- **Caution**: Both measures are **sensitive to outliers,** which is why graphical inspection (scatterplot) and analysis of the descriptive statistics are indispensable.
- If outliers are to be taken into account in the analyses, the Spearman rank correlation can be used as an alternative.
- The comparison of the correlation coefficient with outliers and without outliers can be just as insightful!
  - Example with 50 data points and 2 massive outliers bzw. 1 Ausreißer und ein potenziell fehlerhafter Wert!
  - → Data inspection is essential!!!!

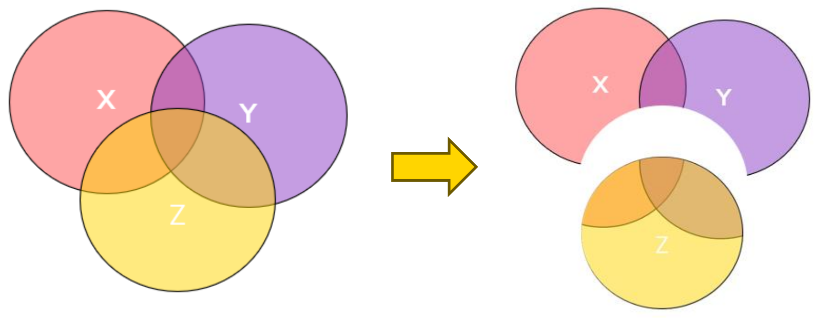


- **Helpful tool for visualizing correlations:** https://rpsychologist.com/correlation/

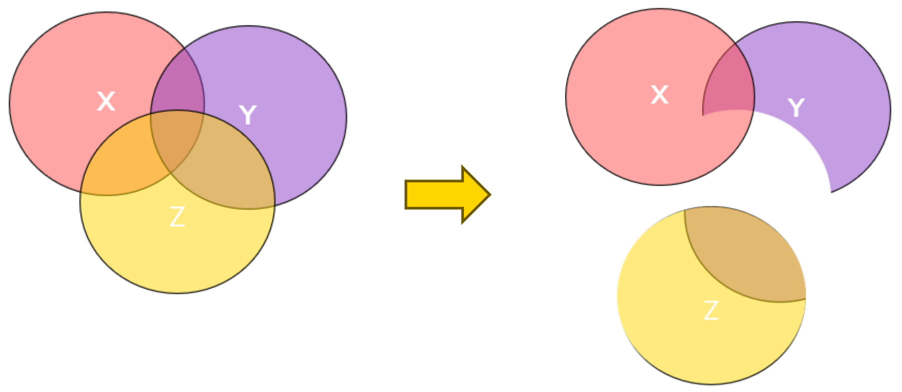
(Semi-)partial-correlation 1/2

- Both are important for understanding multiple regression analysis and structural equation modeling

- **Partial-correlation $r_{xy.z}$:** : The influence of a third variable (Z) is removed from both other variables
  - = the bivariate correlation between 2 residual variables



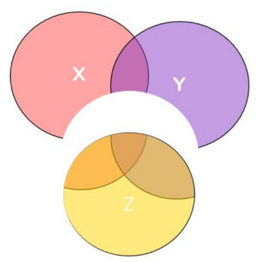
- **Semipartial correlation (SPK) $r_{x(y.z)}$:** The influence of a third variable Z is removed from one of the other variables
  - = Correlation of a variable X with a residual variable Y that has been adjusted for the influence of Z
  - **Underlying concept**: Can a variable Y make an independent contribution to the variance explanation of X (if Z already makes a contribution)? → quantified by squared SPK

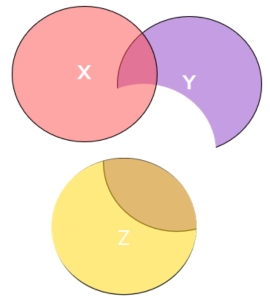


(Semi-)partial-correlation 2/2

- **Significance in multiple linear regression (MLR)**
  - **Partial-correlation**: quantifies the relationship between a UV and the AV, controlling for the influences of all other UVs (on the respective UV and AV). This makes it possible to isolate the specific effect of a UV on the AV.
- **Semipartialkorrelation:** *(SPSS: „**Part**")* quantifies the unique contribution of a UV to the prediction of AV after the influence of the other UVs on AV, **but not on each other,** has been considered = quantifies the additional variance portion of AV explained by each UV **when the effects of the other variables are already in the model → "unique variance elucidation"**
  - This does not remove the joint effect that the other UVs have on each other from the calculation.

Correlations		
Zero-order	Partial	Part
,096	-,130	-,128
-,054	,054	,053
,178	,188	,186





Zero-order = without controlling for the potential influence of other variables → classic Pearson correlation

Statistical Decision

Statistical Decision	Reality	
	H_0 is true; H_1 is false	H_0 is false; H_1 is true
H_0 is rejected; H_1 is accepted	Type I error (α)	Corrct decision ($1 - \beta$)
H_0 is retained; H_1 is rejected	Correct decision ($1 - \alpha$)	Type II error (β)

„Power“

Caution: In contrast to the risk of an α-error, where we incorrectly reject a true H0, it is often not possible to determine the risk of a β-error in advance.
This is because of how we set up hypotheses in statistical tests: While the H0 always makes a specific assertion (e.g. $\mu = 0$), the H1 usually covers all other possibilities and is therefore usually formulated in more general terms (e.g. $\mu \neq 0$).
To determine β in advance, you need information about the effect size in the population (and this is not known a priori.)

p-value 1/2

- „[...] is defined as the probability of finding an empirical result, or a result that is even more strongly, against the null hypothesis, **assuming that the null hypothesis is true.**"[1]
  - or "the maximum probability under H0 that the test statistic is realized in the observed realization or a more extreme realization in the direction of H1."[2]
- Achtung: The *p*-value is NOT the probability that the null hypothesis is true!

- If the **p-value** is **less than/equal to** the previously defined significance level **α**, we opt **for the alternative hypothesis**
- **The calculation of the *p-value*** is based on the distribution of the respective test statistic under the assumption that H0 is true. This distribution describes how likely different possible values of the test statistic are if the H0 is true. The *p-value* is then used to assess how extreme the observed data is compared to this distribution.
  - The actual calculation is usually automated using the statistical software of our choice.

[1]Eid et al., 2017, p.219 [2] VO Psychologische Statistik 1, Foliensatz „Hypothesentests 2", Folie 4. WS23 - Freudenthaler

p-value 2/2: directed/undirected hypotheses

- The *p*-value is calculated in the same way regardless of whether the hypotheses are undirected (two-sided) or directed (one-sided).
  - **The difference lies in how the *p*-value is interpreted and how the critical region for the test is defined - based on the type of hypothesis.**

- **undirecrted (two-sided) Hypotheses:** all "more extreme realizations in the direction of H1 that are greater in magnitude than the observed realization t of the test statistic T" → $H_1: \mu \neq \mu_o$

See the definition on the previous slide!

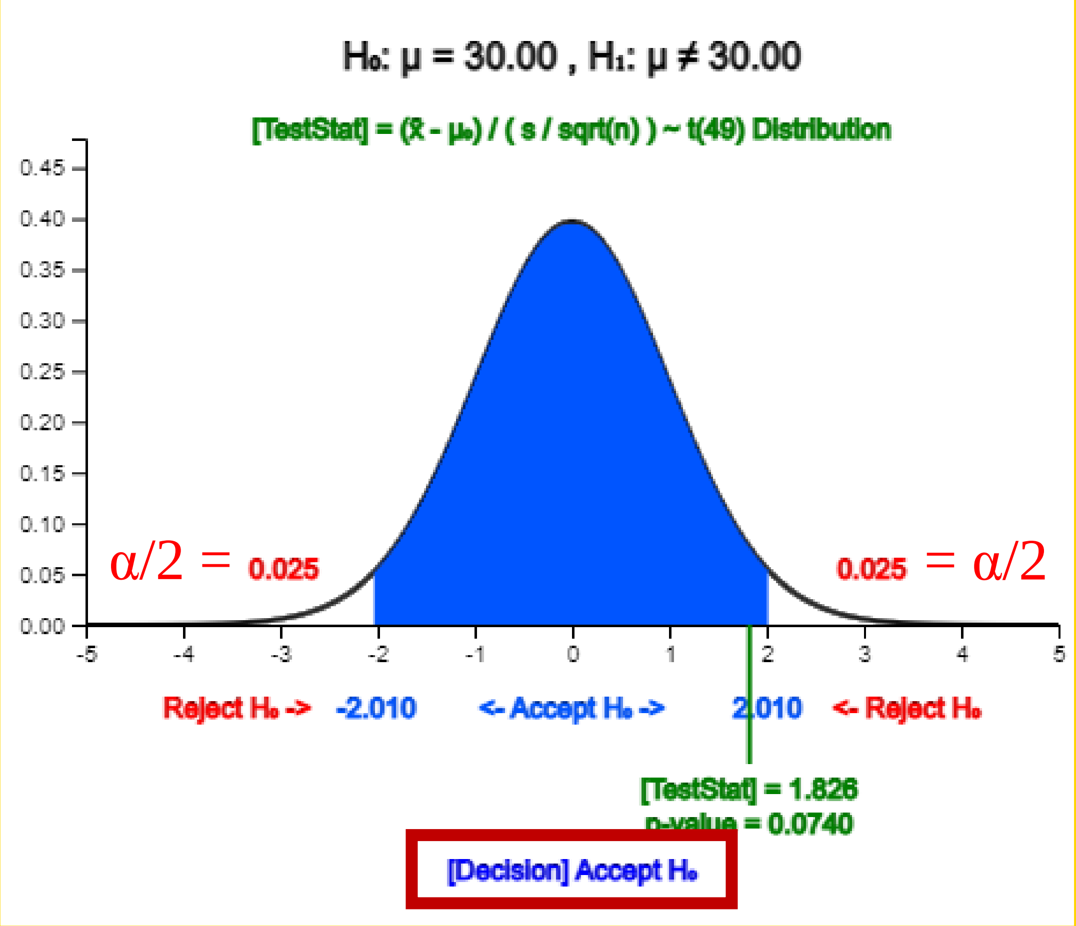
- **directed (one-sided) Hypotheses**
  - **Rightward**: "more extreme realizations in the direction of H1 to the right of the observed realization t of the test statistic T" → $H1 : \mu < \mu_0$
  - **Leftward**: "more extreme realizations in the direction of H1 to the left of the observed realization t of the test statistic T" → $H_1 : \mu > \mu_o$

See example on the next slide!

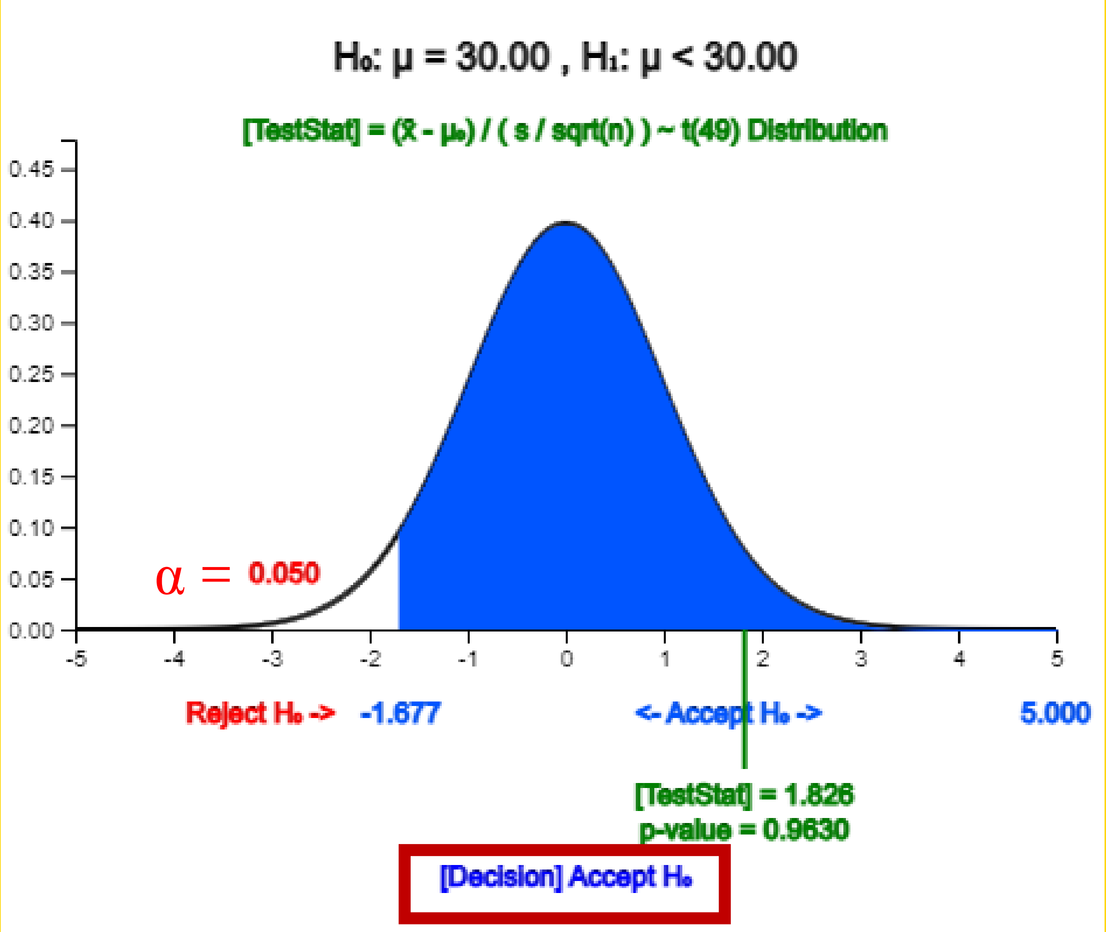
p-values - examples

- H0: $\mu = \mu_o = 30$
- $n = 50$, $\bar{x} = 31$, $s^2 = 15$
- The parameters are always the same here, only the direction of the alternative hypothesis is changed and thus the critical regions that lead to the rejection of H0 *(here only the rightward H1)*
- *Green: Realisation of the test-statistic (empiric t-value) & p-value, Blue: critical t-value*

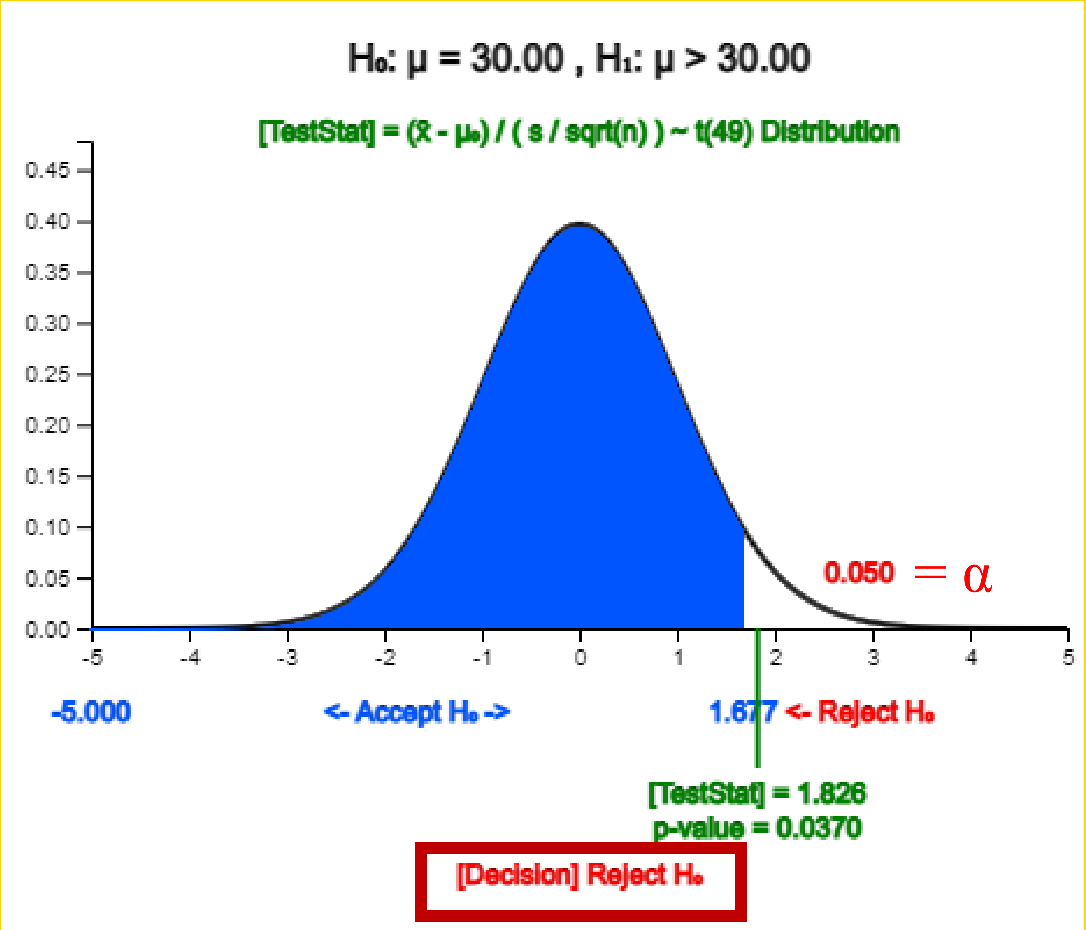
Undirected



Leftward



Rightward



Graphics created with: http://www.estat.me/estat/eStatU/70TestingMu.htm

Effect size

- **Effect size** = unit-independent measure of the size of a difference or the magnitude of a relationship
- **Caution:** There are different measures of effect size! https://www.psychometrica.de/effektstaerke.html

Example Cohen´s δ:

- For indipendent samples: $\delta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma^2_{pool}}}$
- For dependent samples: $\delta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma^2_{Diff}}}$
- Unit: "Standard deviations"

Interpretation:

- |δ| > .20 → small Effect
- |δ| > .50 → middle Effect
- |δ| > .80 → big Effect

Interpretation always in relation to the content-related question!

Power (1-ß)

Konvention für psychologische Untersuchungen:
Power = 1-ß ≥ .80

- **Power** = the probability that, if H_1 is true, we find the precicted effect in our sample

Factors influencing the power

- **Effect size**
- **Significance level α**
- **Sample size n**

Note:
1. if three of these four quantities are known, the fourth quantity can be determined unambiguously
2. Positive relationship between the influencing variables and the power! The greater one of the influencing variables, the greater the power

Application in practice:

1. **Determination of the power for an examination:**
   "What is the chance that a phenomenon that exists will be discovered by the planned investigation?"
2. **Calculation of the sample size:**
   "How large must the sample size be in order to find the expected effect?"

With G*Power

- Download: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower
- Manual: https://www.psychologie.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf

Determining the power using G*Power

Selection of the test family: t-Test, F-Test (=ANOVA),…

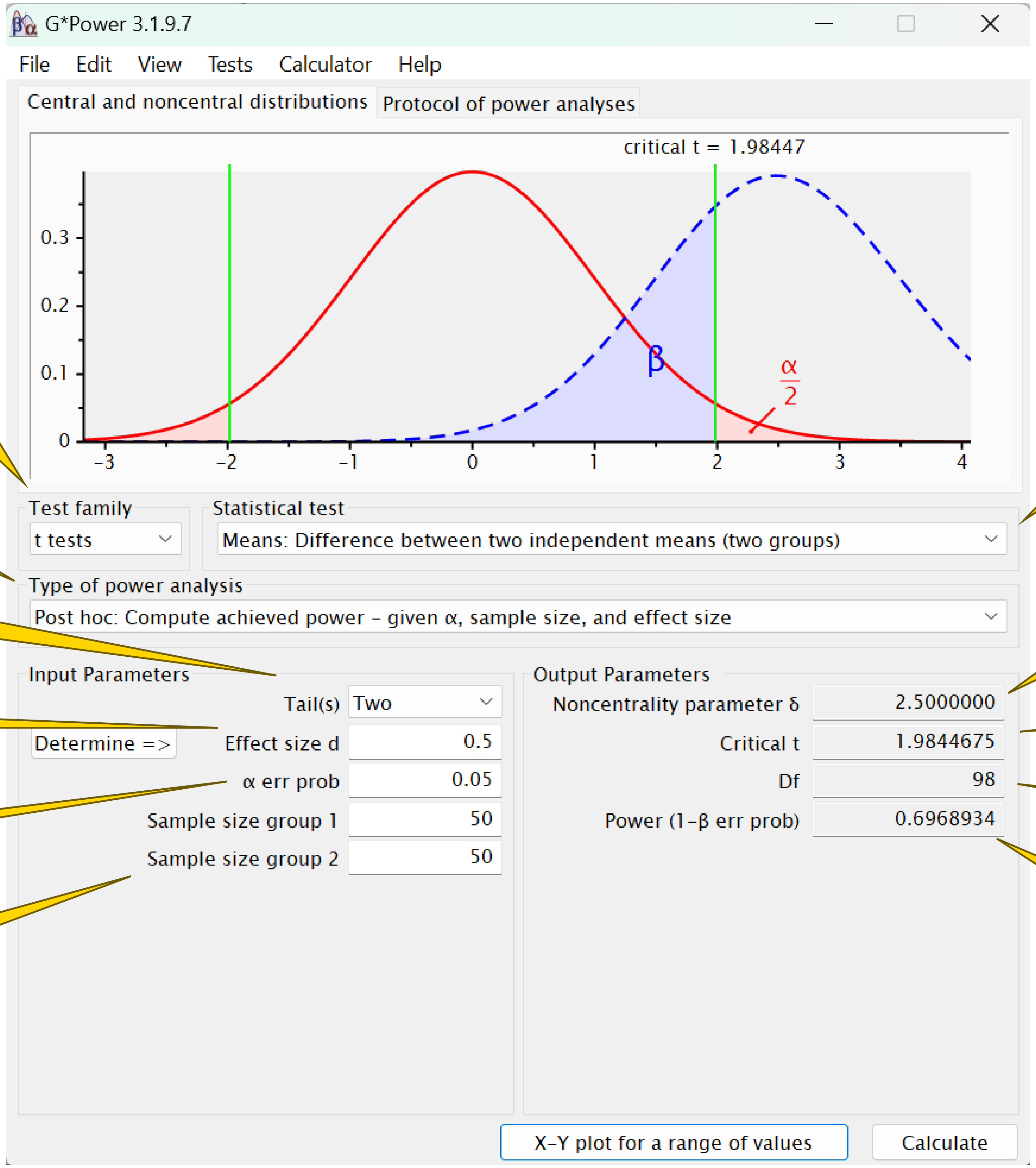
Size to be calculated 1-ß, n,…

directed vs. undirected Hypothesis

Effectsize d / f

Significance level α

Sample size(s)



Specification of the test: e.g. dependent vs. indipendent, Regression,…

Effectsize δ

critical t-value

Degrees of freedom

Power

Sample size planing with G*Power

Selection of the test family: t-Test, F-Test (=ANOVA),…

Size to be calculated 1-ß, n,…

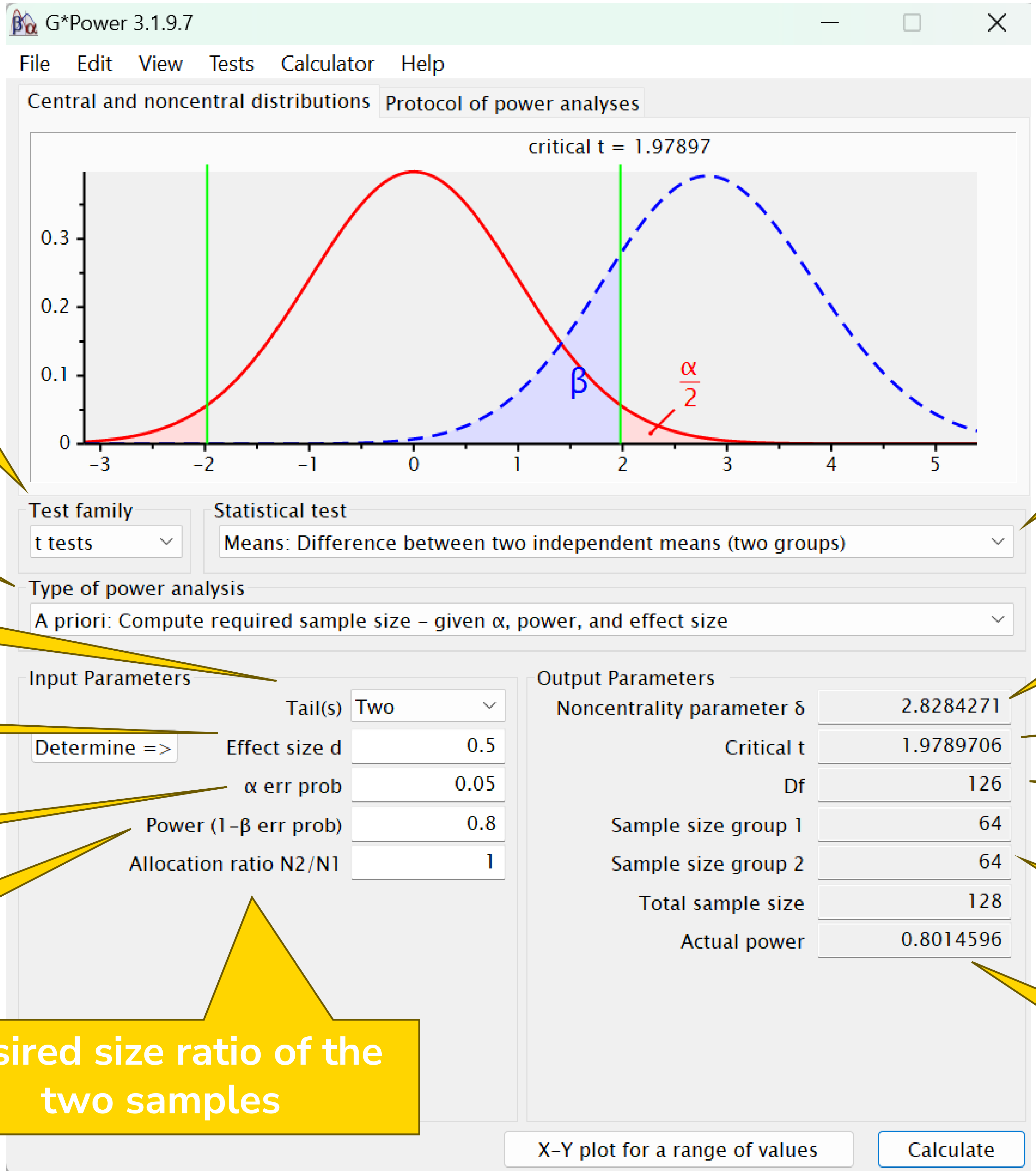
directed vs. undirected Hypothesis

Effect size d / f

Significance level α

Desired power

Desired size ratio of the two samples



Spezifikation des Tests: z. B. Abhängig vs. Unabhängiger, Regression,…

Effect size δ

critical t-value

Degrees of freedom

Needed sample size of the groups and total sample size

Actual predicted power

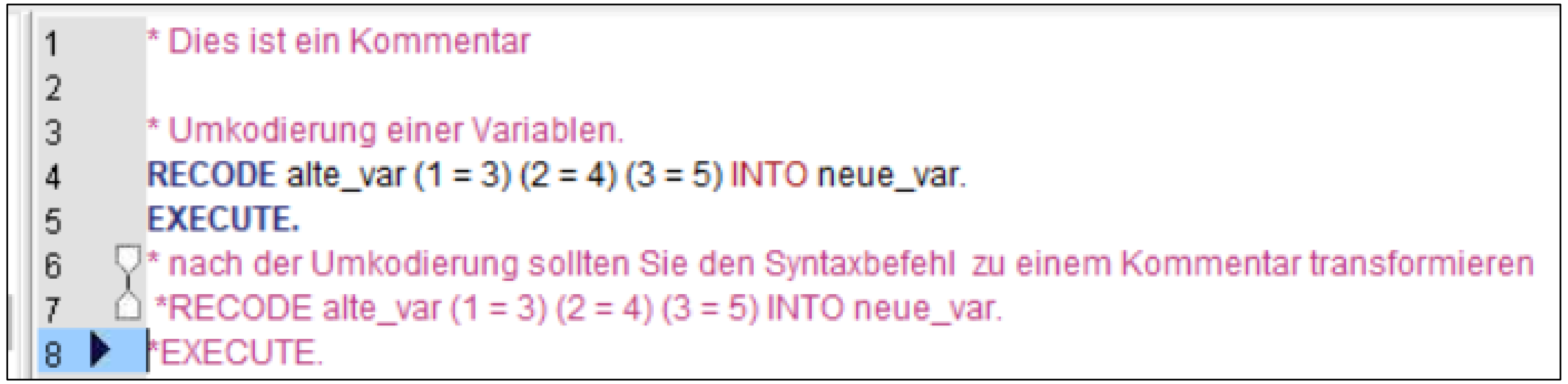
SPSS

Basics

Data preparation and working with SPSS syntax

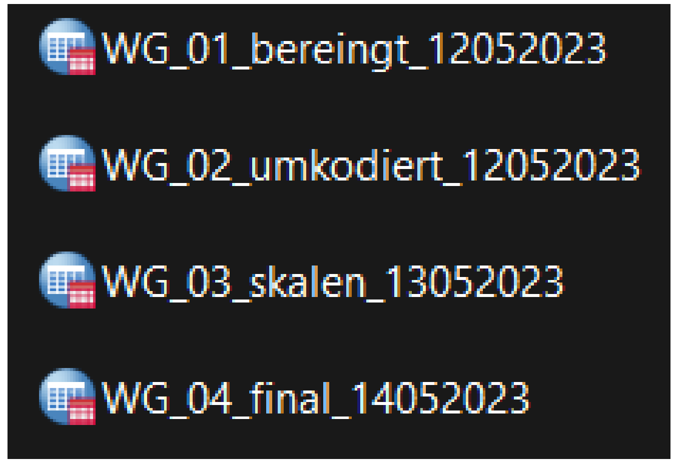
General information 1/2

- Save **raw data** twice and preferably once externally (backup copy)
- Create a separate "**data work file**" (→ if something is accidentally deleted, the original raw data is still available!)
- **Syntax**: Commands can be selected and executed individually
- Functions such as the creation/modification of variables, z-standardization/recoding of variables etc. are **executed each time the syntax is run**!
    - It is best to mark the syntax command as a comment afterwards!
- Use (detailed) **Comments** („*" before your text) in the Syntax!
- There must **be a period** at the end of a command!
- You can use tools like Chat-GBT to create Syntax code ;)



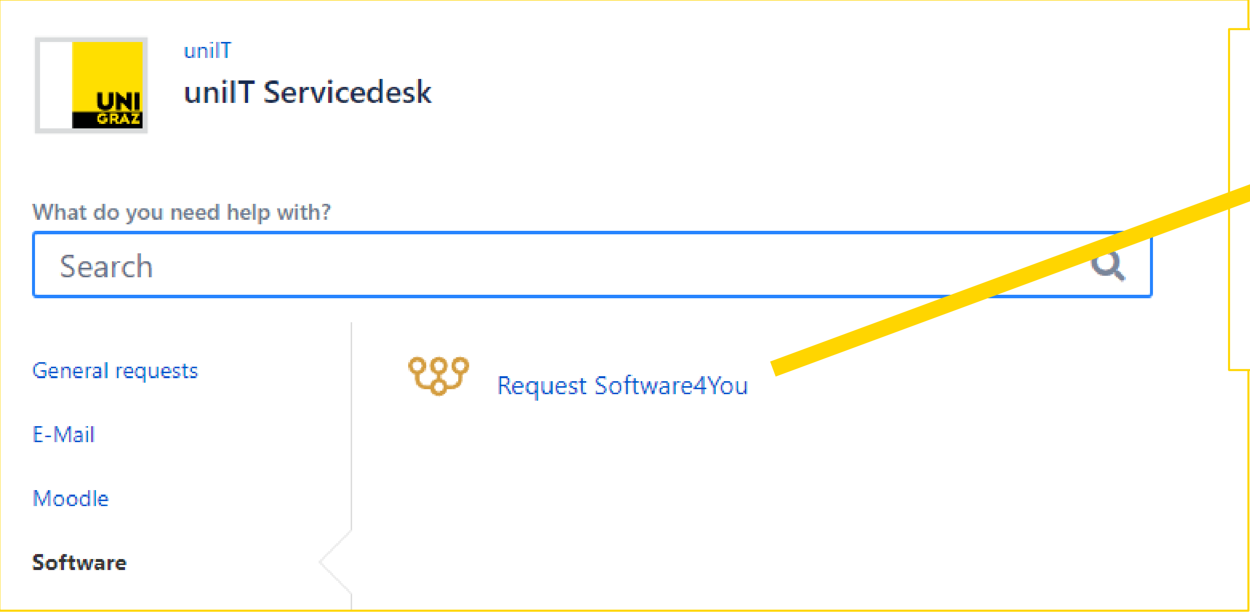
General information 2/2

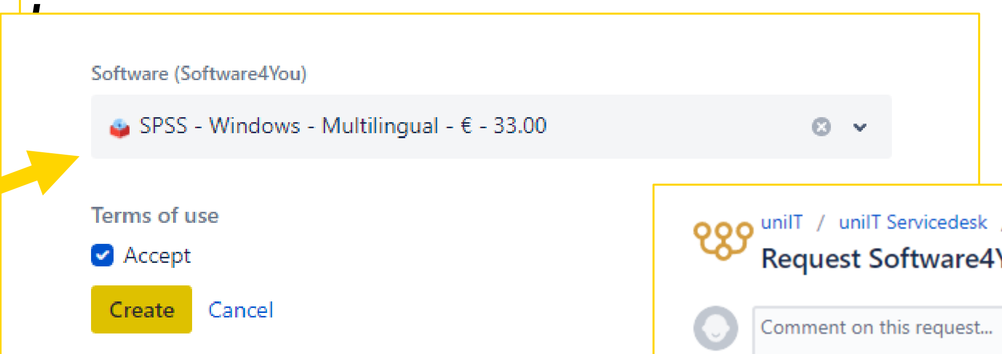
- In Excel or similar: document the number of excluded cases for each step
  - Incomplete*data records, extreme values (→ filter), occasional missing values*
- Always create new variables (do not replace originals!)
- Choose a clear and simple description of your data/syntax
- add missing SPSS labels for every Variable!!!
- Select unique, simple and meaningfull variable names

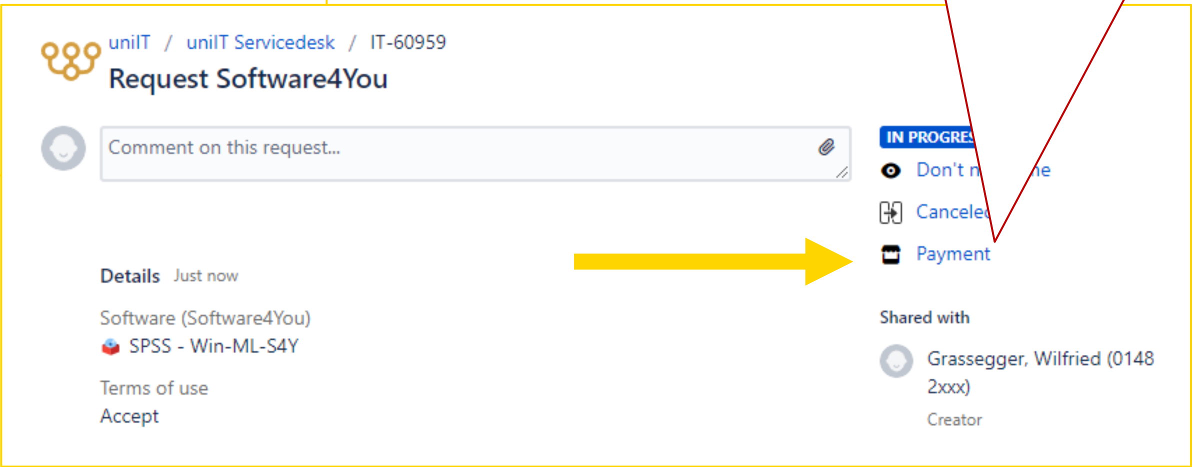


Access to SPSS

- Use a PC at the University (PC-Room at the institut)
- https://it.uni-graz.at/de/it-services/arbeiten/software-fuer-studierende/

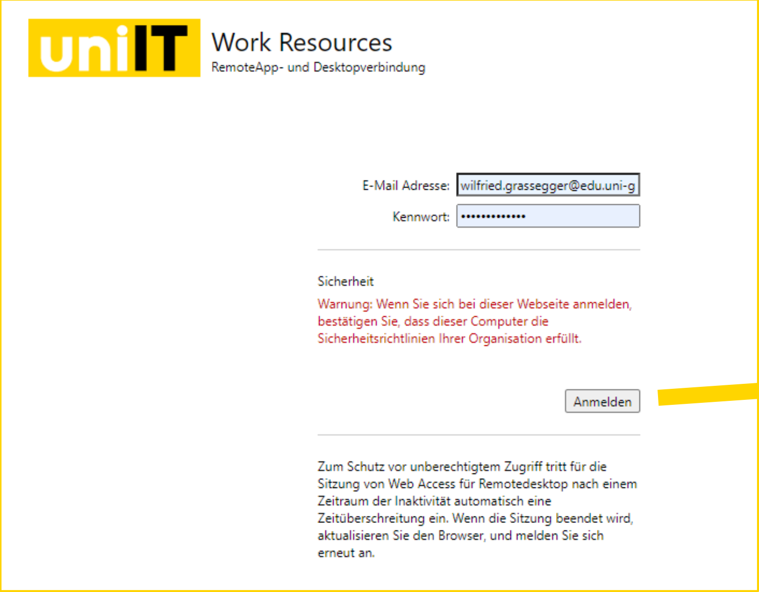






After the request, the amount can be transferred directly, the link to the download + installation instructions will be sent automatically!

B)

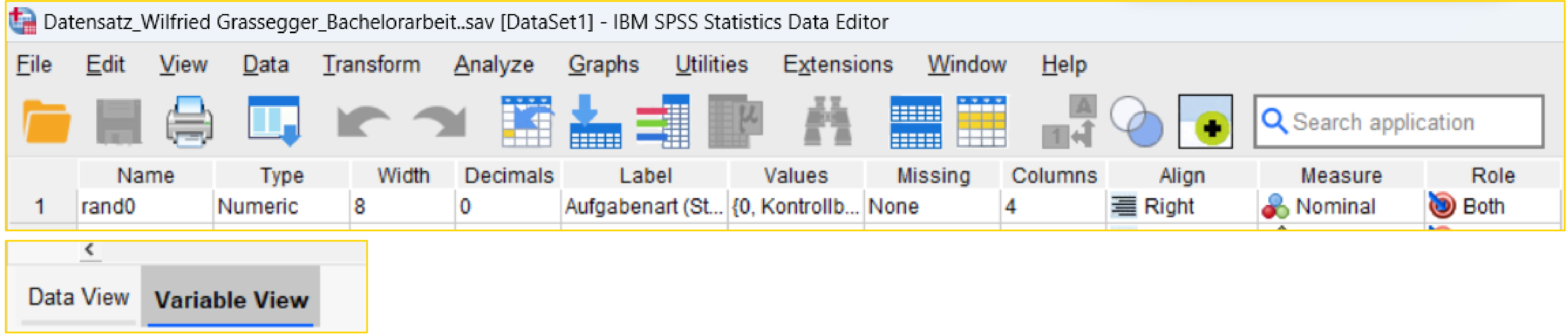


The remote desktop connection is then downloaded:
Open →
Connect →
Enter user data.

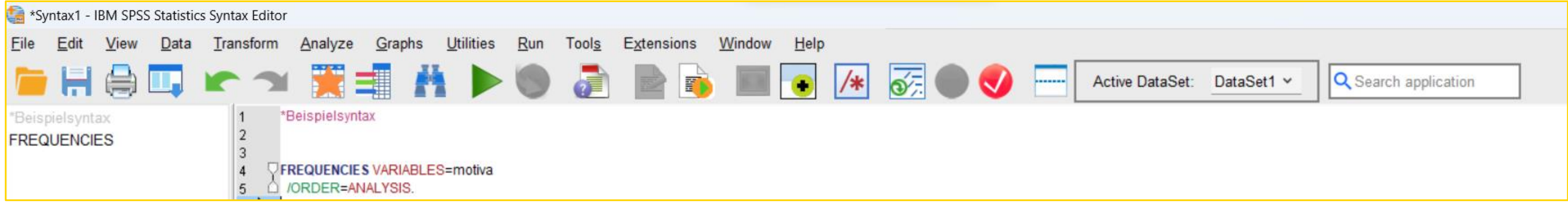
After confirmation, SPSS opens automatically.

Open data & screens

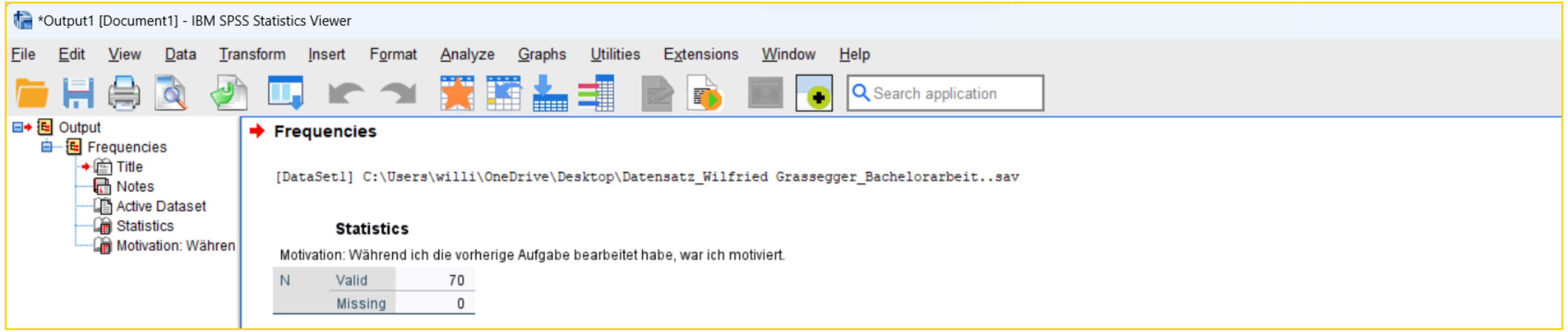
- **SPSS** data files have the extension **.sav**
  - Data View & Variable View



- SPSS-**Syntax** have the extension **.sps**
  - programmable command language
  - Automatisierung, Reproduktion & Dokumentation von Analyseschritten



- SPSS-**Outputs** have the extension **.spv**
  - Result of the analyses & error messages



Tips: Before starting your analysis 1/2

By activating this function, **the respective syntax command is included in the output.**

```
FREQUENCIES VARIABLES=motiva
  /ORDER=ANALYSIS.
```

Frequencies

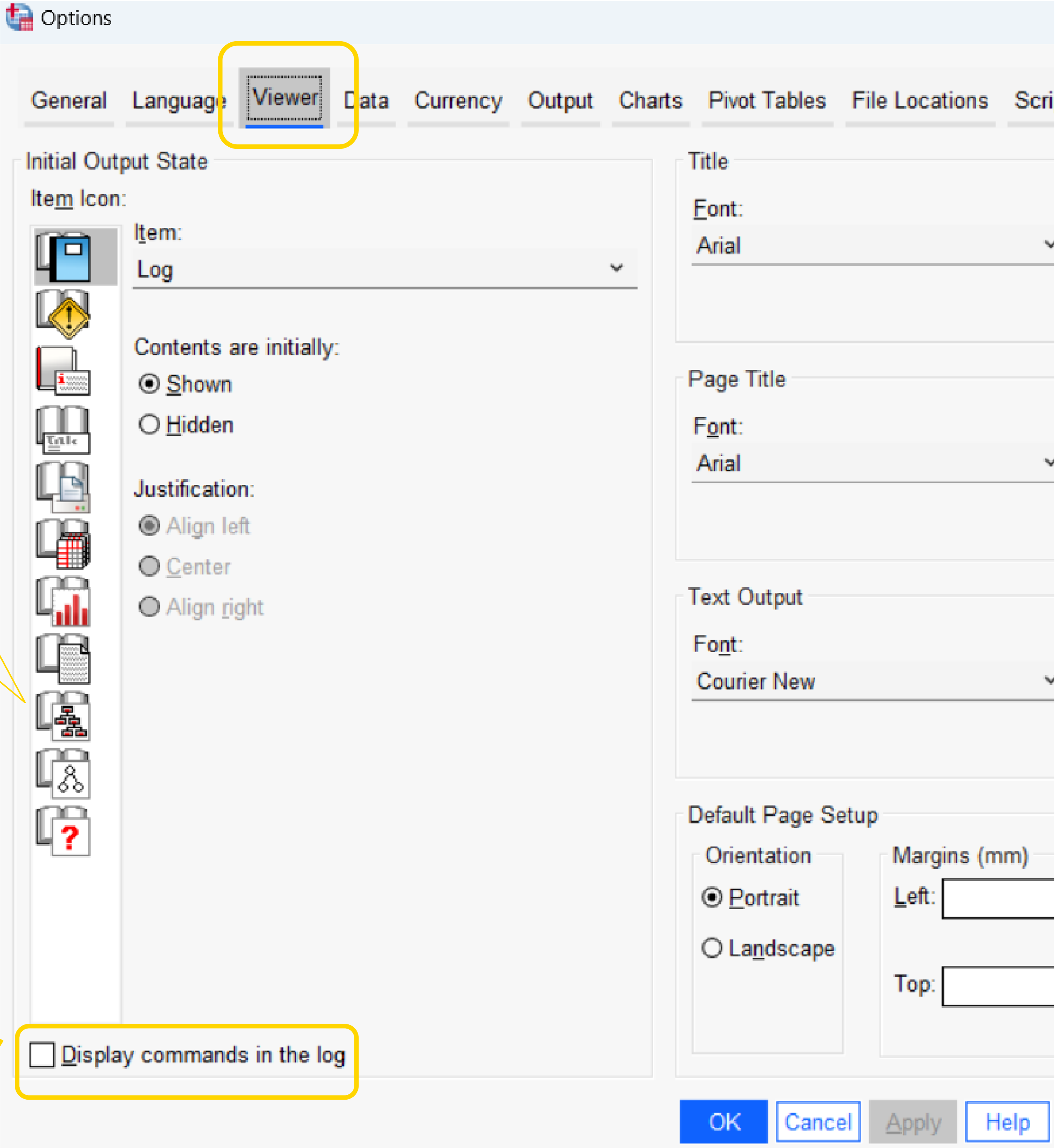
Statistics

Motivation: Während ich die vorherige Aufgabe bearbeitet habe, war ich motiviert.

N	Valid	70
	Missing	0

Motivation: Während ich die vorherige Aufgabe bearbeitet habe, war ich motiviert.

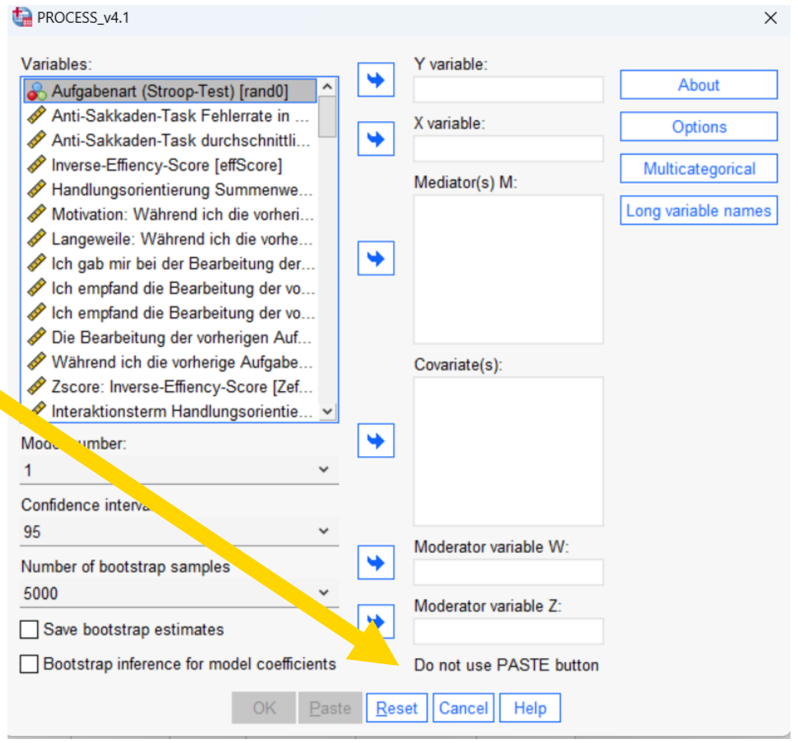
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	trifft überhaupt nicht zu	4	5,7	5,7	5,7
	2	11	15,7	15,7	21,4
	3	12	17,1	17,1	38,6
	4	16	22,9	22,9	61,4
	5	15	21,4	21,4	82,9
	trifft völlig zu	12	17,1	17,1	100,0
	Total	70	100,0	100,0	



Tips: Before starting your analysis 2/2

- Even if you prefer the graphical user interface to working with the syntax, **ALWAYS use the "*PASTE*" command** instead of "*OK*" to document each step in the syntax.
  - This has the great advantage that you can edit and/or copy any command in the syntax.
  - This saves you having to "click through" the graphical user interface if you want to perform the same analysis with other variables, for example.

- Use the "*OK*" button instead of "*PASTE*", only if you are working with the PROCESS macro



Creating new Variables

- Is needed for e.g.:
  - total scores (questionnaires)
  - Means (questionnaires, reaction times)
  - difference value (reaction times)

Exemplary syntax command

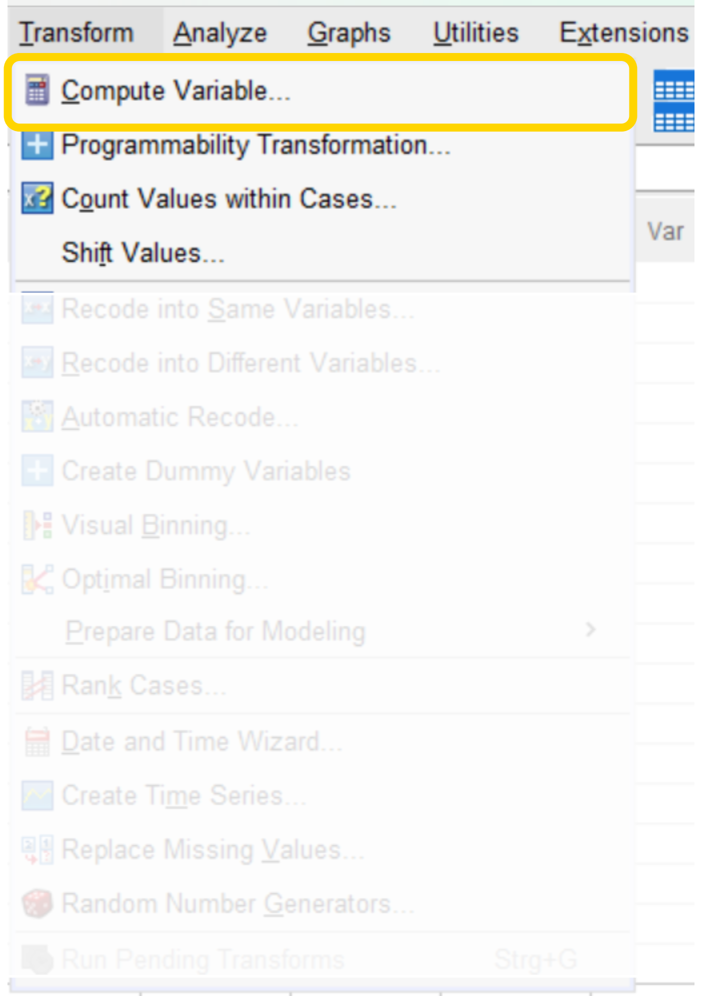
```
Compute random_sum=random1 + random2.
EXECUTE.
```

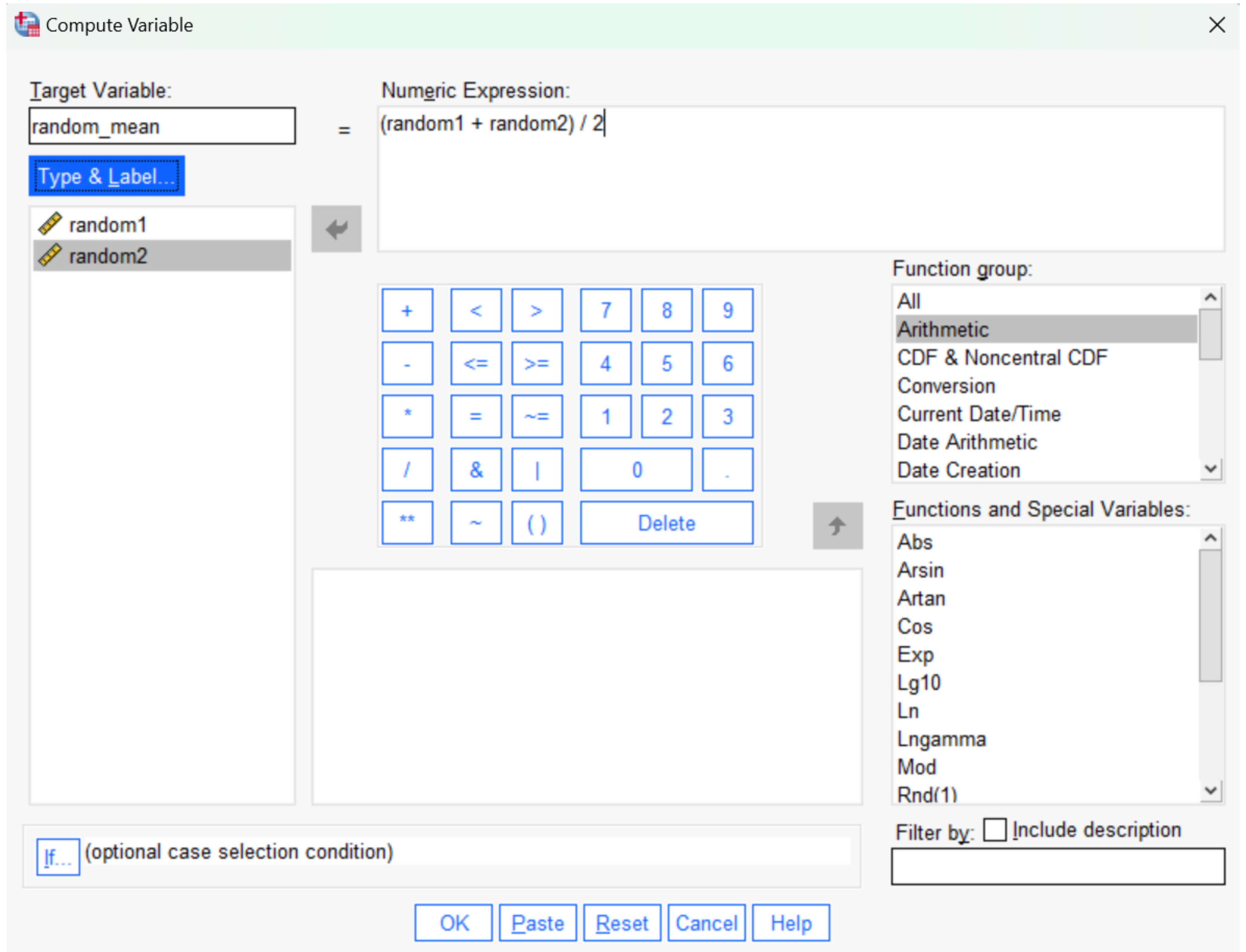
Exemplary syntax command

```
Compute random_mean=(random1 + random2)/2.
EXECUTE.
```

Exemplary syntax command

```
Compute random_diff=random1 - random2.
EXECUTE.
```





Note:
In the case of (e.g.) sum scores from a large number of variables, it is not necessary to enumerate all variables individually. The “to” command is used for this:
Instead:

`x_sum=var1 + var2 + …`

→

`x_sum=var1 to varN`

In this case, however, the variables var1 to varN <u>must</u> follow each other directly in the variable view!

Recoding existing variables

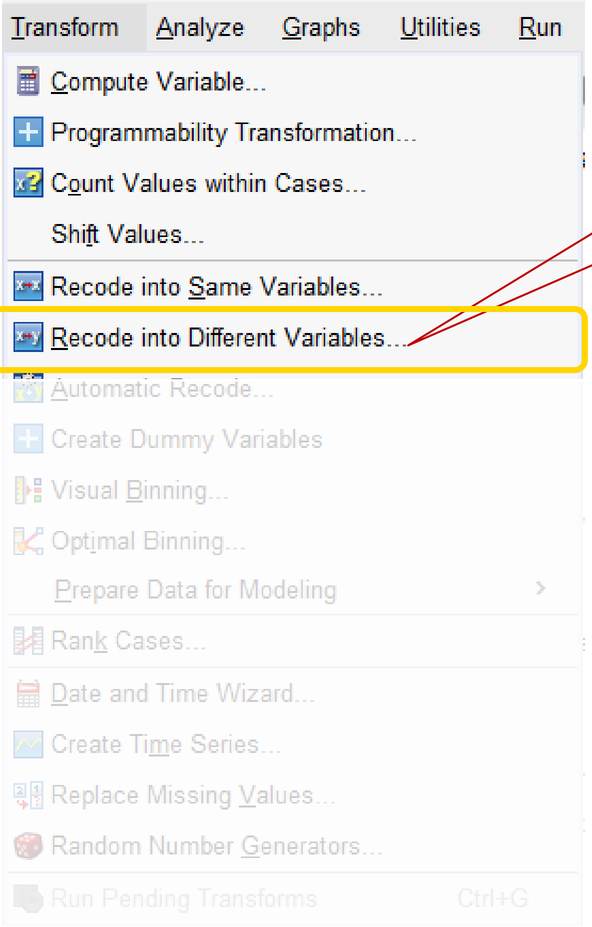
- Is needed for e.g.:
  - Simplification of categorys
  - Inversion of scales
  - Creation of dummy variables
  - Error corrections
  - Mediansplit

Exemplary syntax command

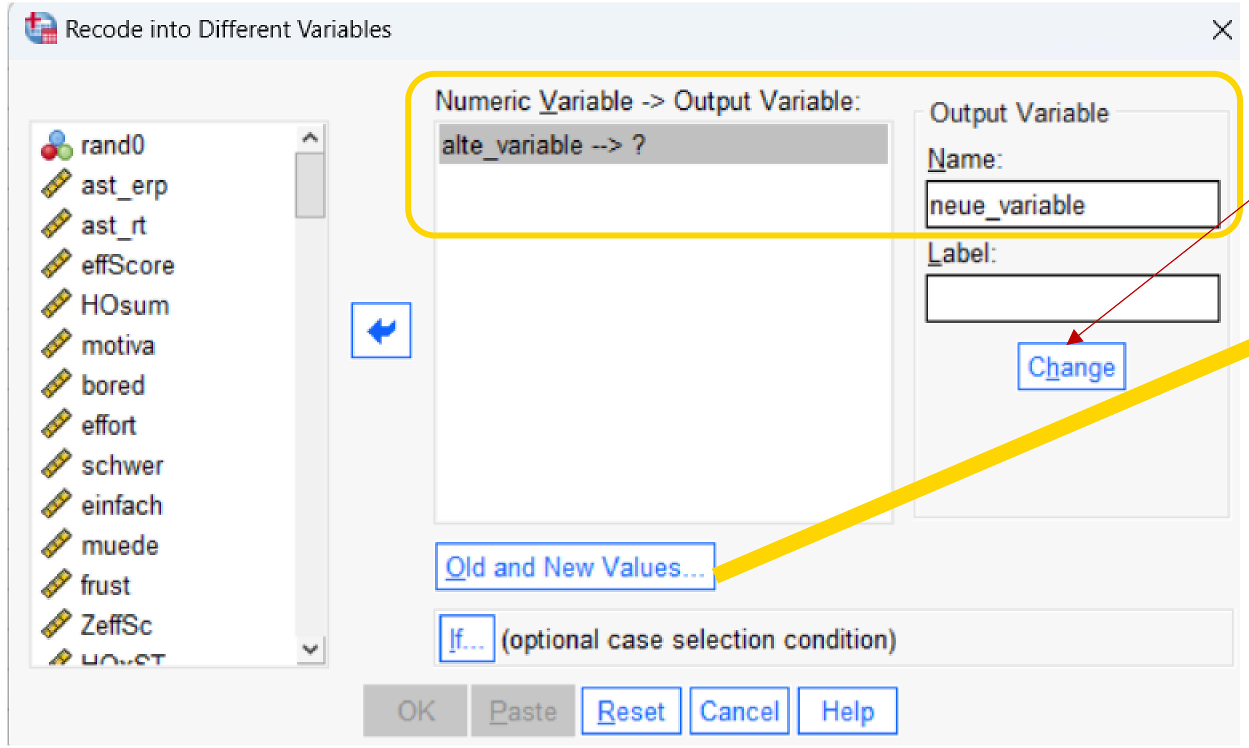
```
RECODE alte_variable (1 THRU 3 = 1) (4 THRU 7 = 2) (8
THRU 10 = 3) INTO neue_variable.
EXECUTE.
```

Exemplary syntax command

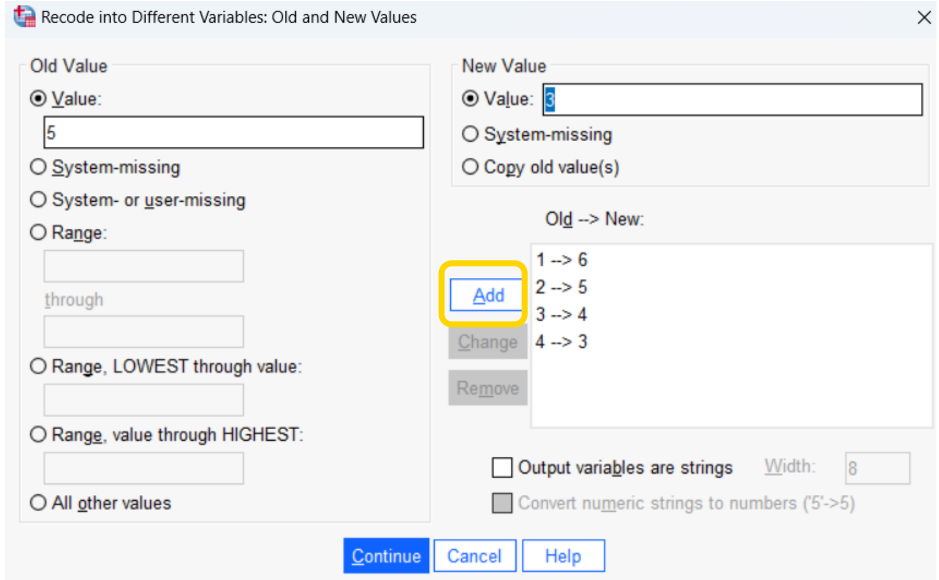
```
RECODE alte_variable (1 = 6) (2 = 5) (3 = 4) (4 = 3)
(5 = 2) (6 = 1) INTO neue_variable.
EXECUTE.
```



ALWAYS create a NEW variable!



- The rules for recoding can also be entered via the graphical user interface instead of via the syntax!
- Click on *Change* to add the newly defined variable as an output variable



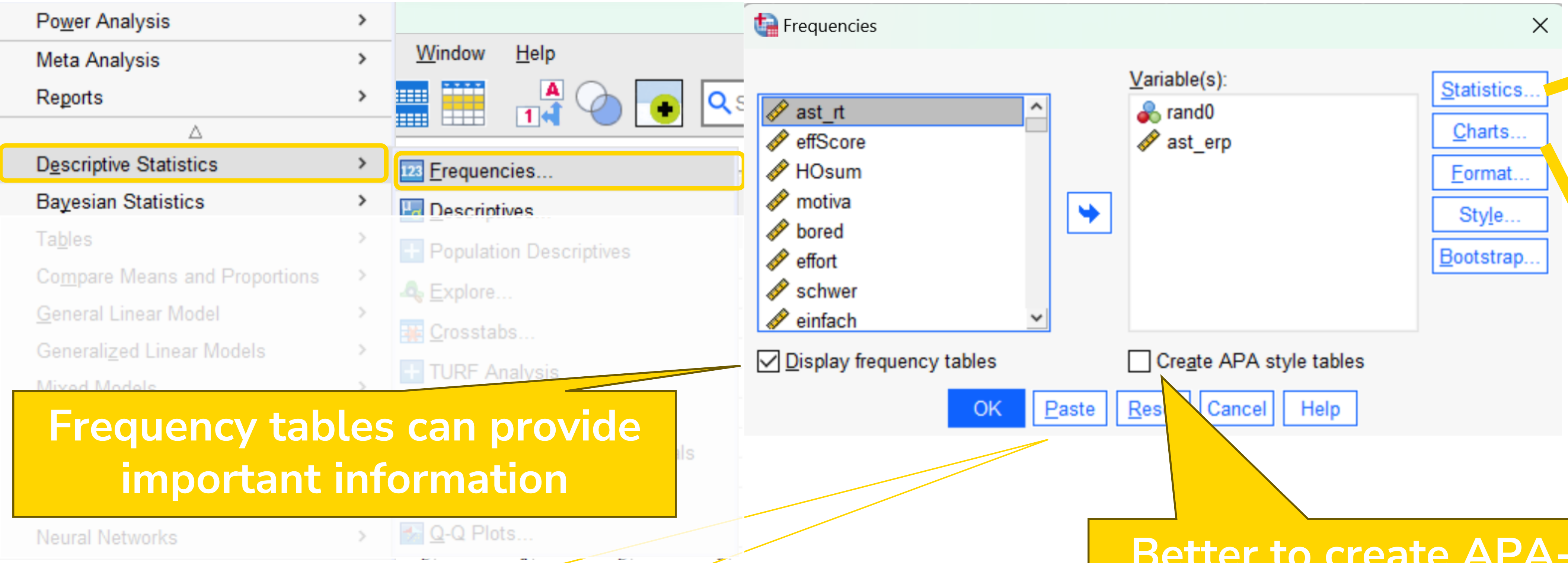
Descriptive statistics

- Characteristic values for **describing data**
  - Composition of the sample
  - “Basic information” about the variables used
  - (Verification of assumptions)

- Important to understand your own data!!!

- Different possibilities:
  - (cumulative) absolute/relative frequencies
  - Measures of central tendency (modal value, median, mean value)
  - Dispersion measures (variance, standard deviation)
  - Graphical forms of presentation
  - Correlation

Descriptive statistics - Frequencies

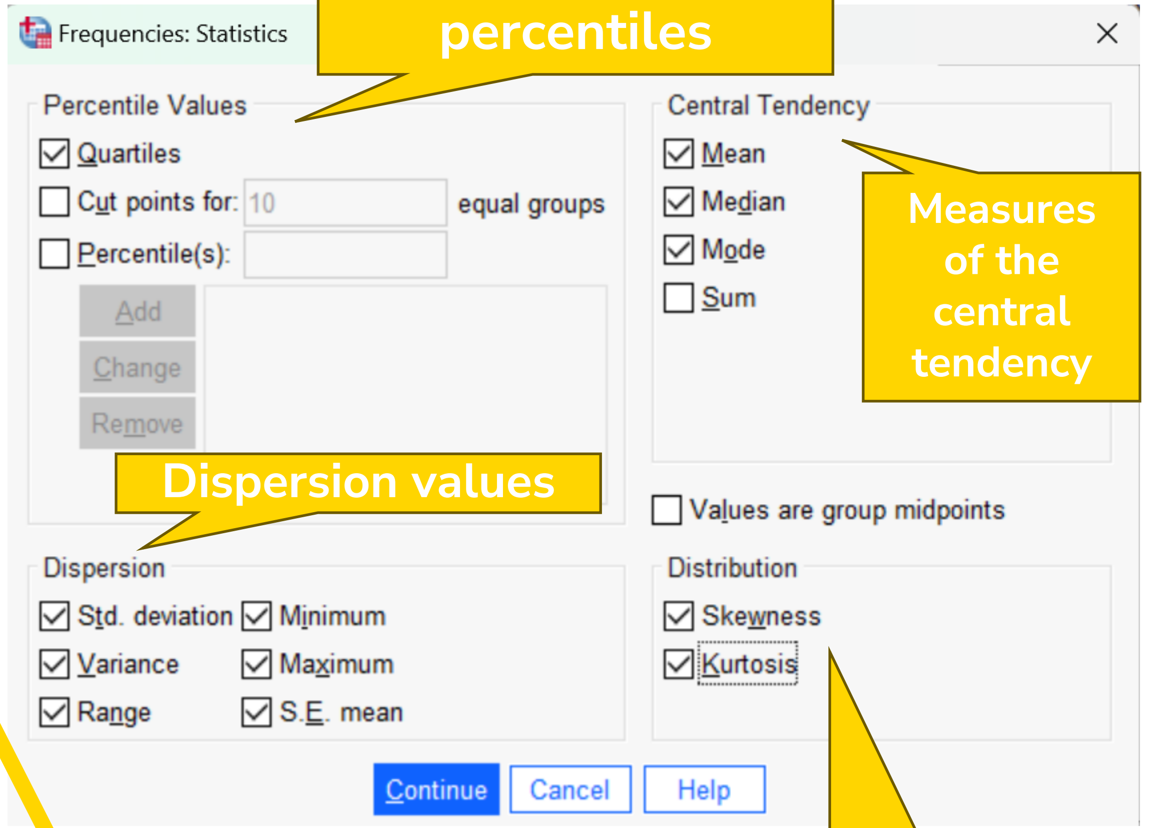


Frequency tables can provide important information

Better to create APA-compliant table yourself

Exemplary syntax command

```
FREQUENCIES VARIABLES=rand0 ast_erp
  /NTILES=4
  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM
MAXIMUM SEMEAN MEAN MEDIAN MODE SKEWNESS SESKEW
    KURTOSIS SEKURT
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.
```

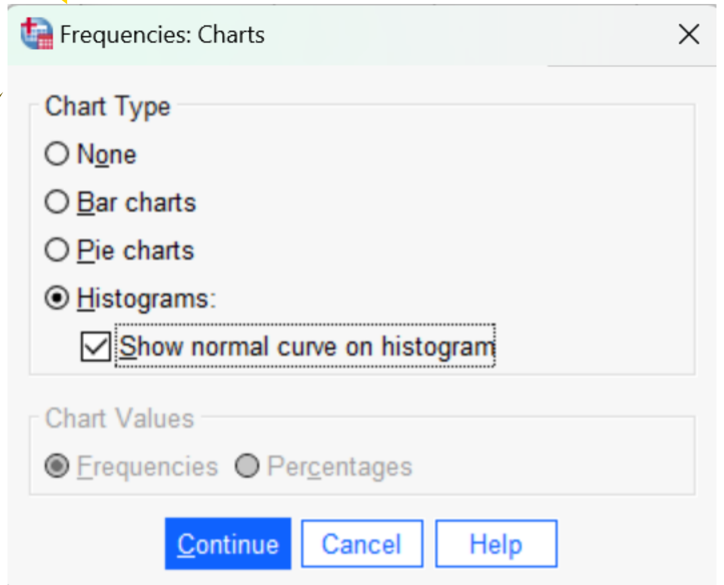


(user-defined) percentiles

Measures of the central tendency

Dispersion values

Measures for assessing whether a variable is normally distributed



Graphics can be helpful for an initial overview

Note:
Depending on the variable (scale level, number of gradations, ...), decide whether more or fewer measures make sense

Descriptive statistics - Frequencies

Selected Variables

Selected Characteristics

Frequencies

Statistics

		Aufgabenart (Stroop-Test)	Anti-Sakkaden-Task Fehlerrate in %
N	Valid	70	70
	Missing	0	0
Mean		,46	24,5714
Std. Error of Mean		,060	1,68038
Median		,00	24,1650
Mode		0	10,00[a]
Std. Deviation		,502	14,05908
Variance		,252	197,658
Skewness		,176	,481
Std. Error of Skewness		,287	,287
Kurtosis		-2,028	-,755
Std. Error of Kurtosis		,566	,566
Range		1	51,67
Minimum		0	3,33
Maximum		1	55,00
Percentiles	25	,00	11,6700
	50	,00	24,1650
	75	1,00	35,0000

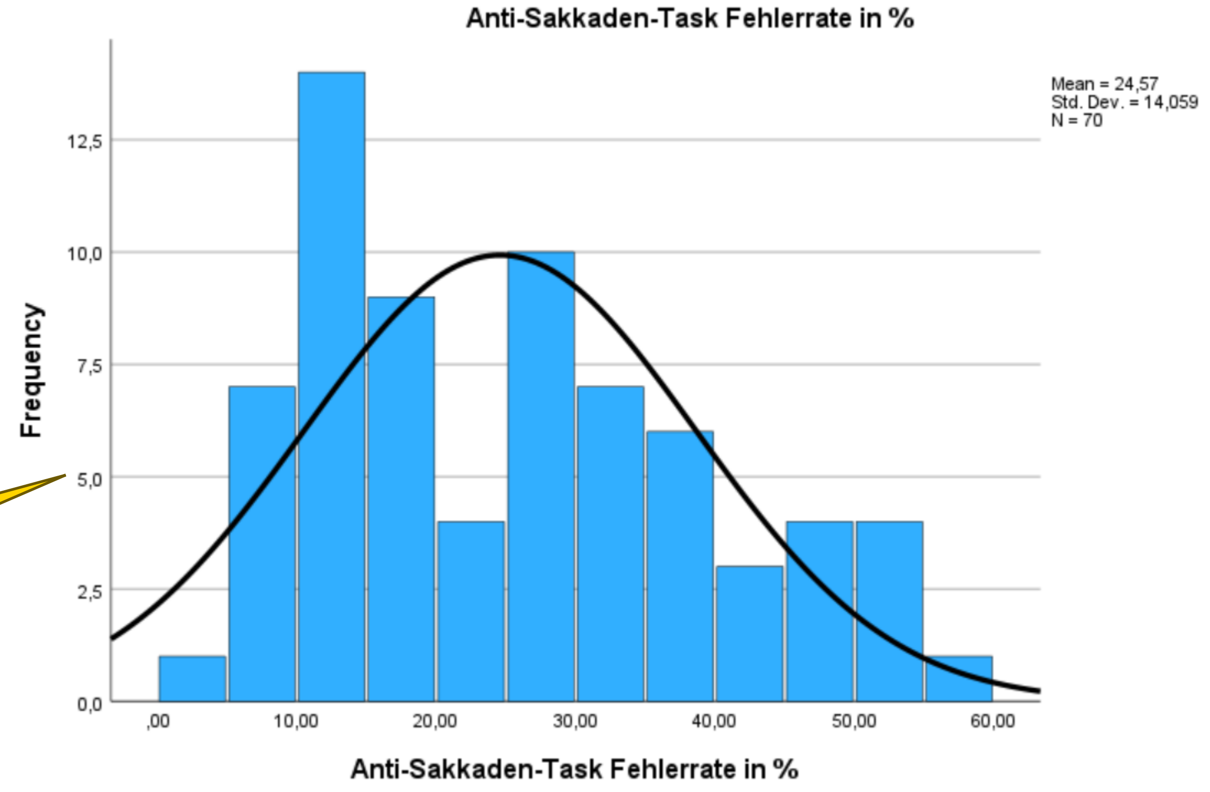
a. Multiple modes exist. The smallest value is shown

Frequency Table

Aufgabenart (Stroop-Test)

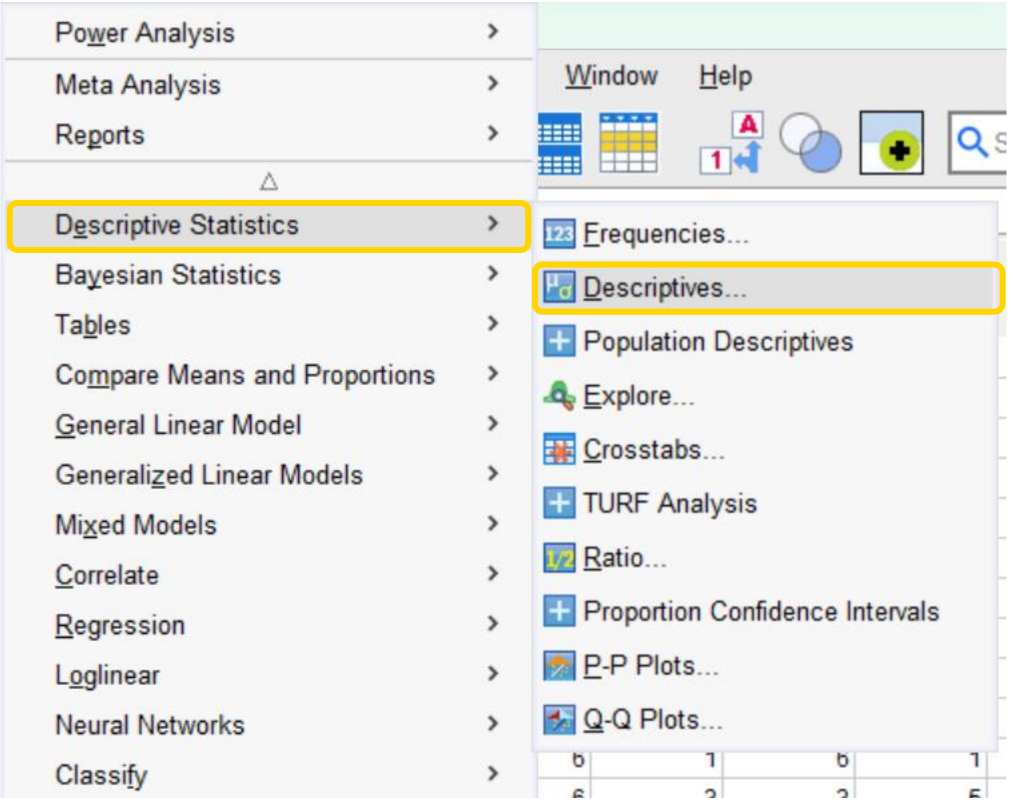
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Kontrollbedingung	38	54,3	54,3	54,3
	Depletion-Bedingung	32	45,7	45,7	100,0
	Total	70	100,0	100,0	

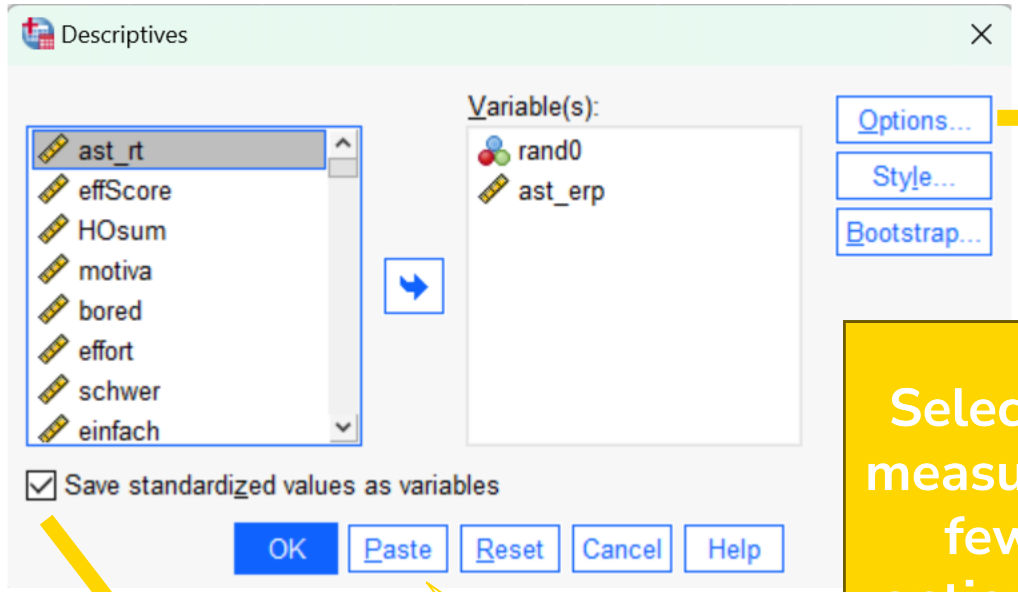
Caution: variables that contain e.g., reaction times or mean scores can result in a very long frequency table

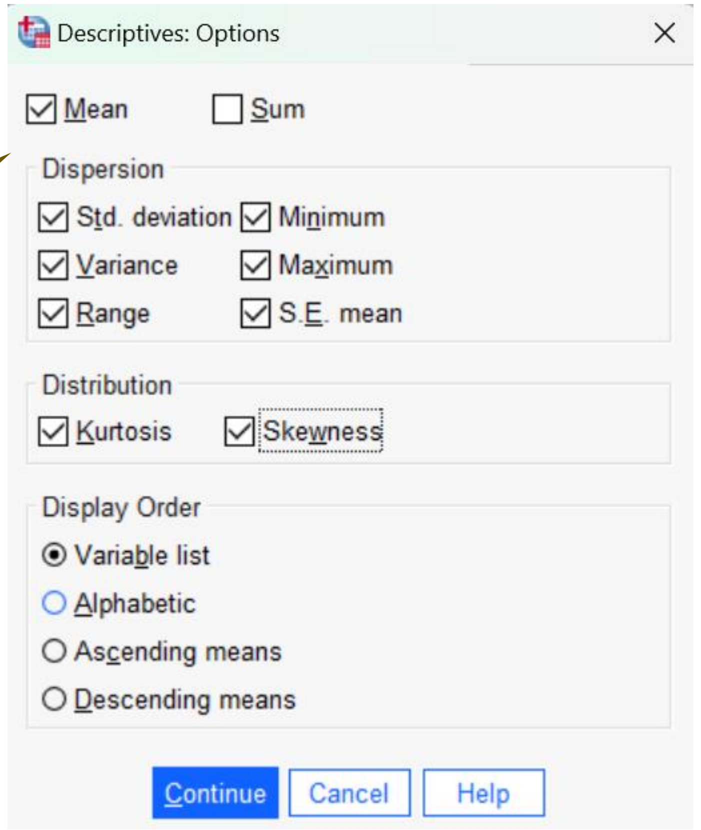


Histogram (with fit-curve)

Descriptive statistics - Descriptives







Selection of desired measures (but overall fewer selection options compared to “Frequencies”)

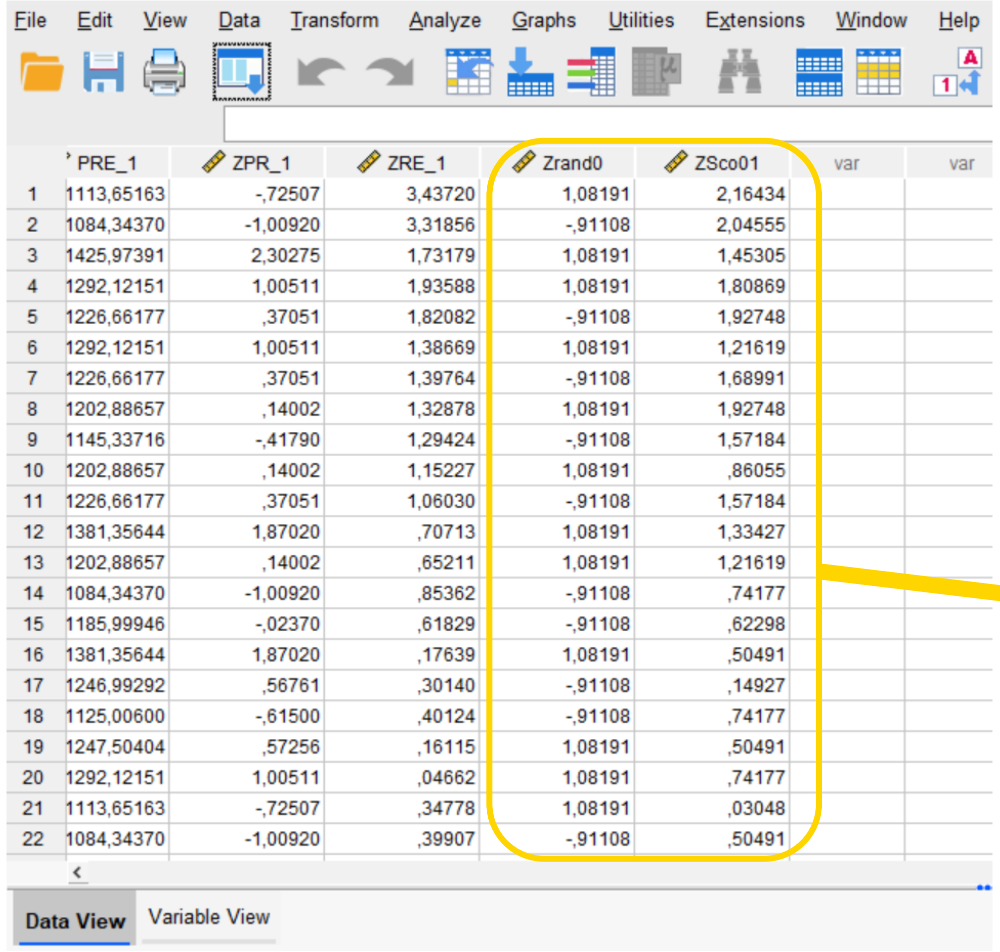
Use to z-standardize your variables ;)

Caution: a new z-standardized variable is created each time the command is executed

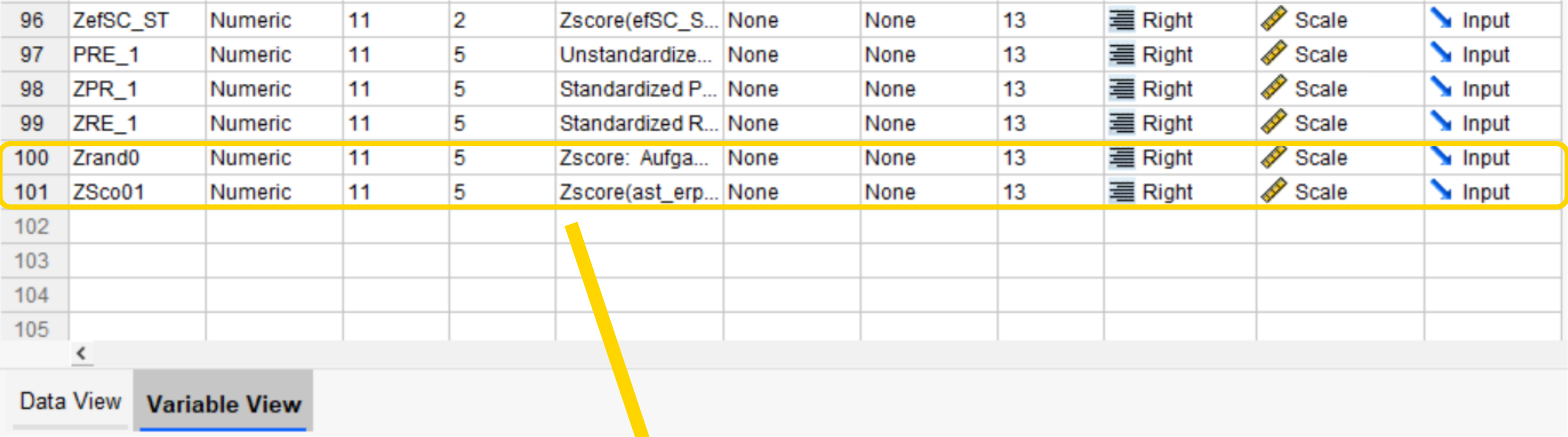
Exemplary syntax command

```
DESCRIPTIVES VARIABLES=rand0 ast_erp
  /SAVE
  /STATISTICS=MEAN STDDEV VARIANCE RANGE MIN MAX SEMEAN KURTOSIS SKEWNESS.
```

Descriptive statistics - Descriptives



	PRE_1	ZPR_1	ZRE_1	Zrand0	ZSco01	var	var
1	1113,65163	-,72507	3,43720	1,08191	2,16434		
2	1084,34370	-1,00920	3,31856	-,91108	2,04555		
3	1425,97391	2,30275	1,73179	1,08191	1,45305		
4	1292,12151	1,00511	1,93588	1,08191	1,80869		
5	1226,66177	,37051	1,82082	-,91108	1,92748		
6	1292,12151	1,00511	1,38669	1,08191	1,21619		
7	1226,66177	,37051	1,39764	-,91108	1,68991		
8	1202,88657	,14002	1,32878	1,08191	1,92748		
9	1145,33716	-,41790	1,29424	-,91108	1,57184		
10	1202,88657	,14002	1,15227	1,08191	,86055		
11	1226,66177	,37051	1,06030	-,91108	1,57184		
12	1381,35644	1,87020	,70713	1,08191	1,33427		
13	1202,88657	,14002	,65211	1,08191	1,21619		
14	1084,34370	-1,00920	,85362	-,91108	,74177		
15	1185,99946	-,02370	,61829	-,91108	,62298		
16	1381,35644	1,87020	,17639	1,08191	,50491		
17	1246,99292	,56761	,30140	-,91108	,14927		
18	1125,00600	-,61500	,40124	-,91108	,74177		
19	1247,50404	,57256	,16115	1,08191	,50491		
20	1292,12151	1,00511	,04662	1,08191	,74177		
21	1113,65163	-,72507	,34778	1,08191	,03048		
22	1084,34370	-1,00920	,39907	-,91108	,50491		



96	ZefSC_ST	Numeric	11	2	Zscore(efSC_S...	None	None	13	Right	Scale	Input
97	PRE_1	Numeric	11	5	Unstandardize...	None	None	13	Right	Scale	Input
98	ZPR_1	Numeric	11	5	Standardized P...	None	None	13	Right	Scale	Input
99	ZRE_1	Numeric	11	5	Standardized R...	None	None	13	Right	Scale	Input
100	Zrand0	Numeric	11	5	Zscore: Aufga...	None	None	13	Right	Scale	Input
101	ZSco01	Numeric	11	5	Zscore(ast_erp...	None	None	13	Right	Scale	Input
102											
103											
104											
105											

z-standardized variables appear in the data set and can be used for analyses

Compared to “Frequencies”:
Same information (but fewer selection options) with a different arrangement

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Anti-Sakkaden-Task Fehlerrate in %	70	51,67	3,33	55,00	24,5714	1,68038	14,05908	197,658	,481	,287	-,755	,566
Valid N (listwise)	70											

z-Standardization

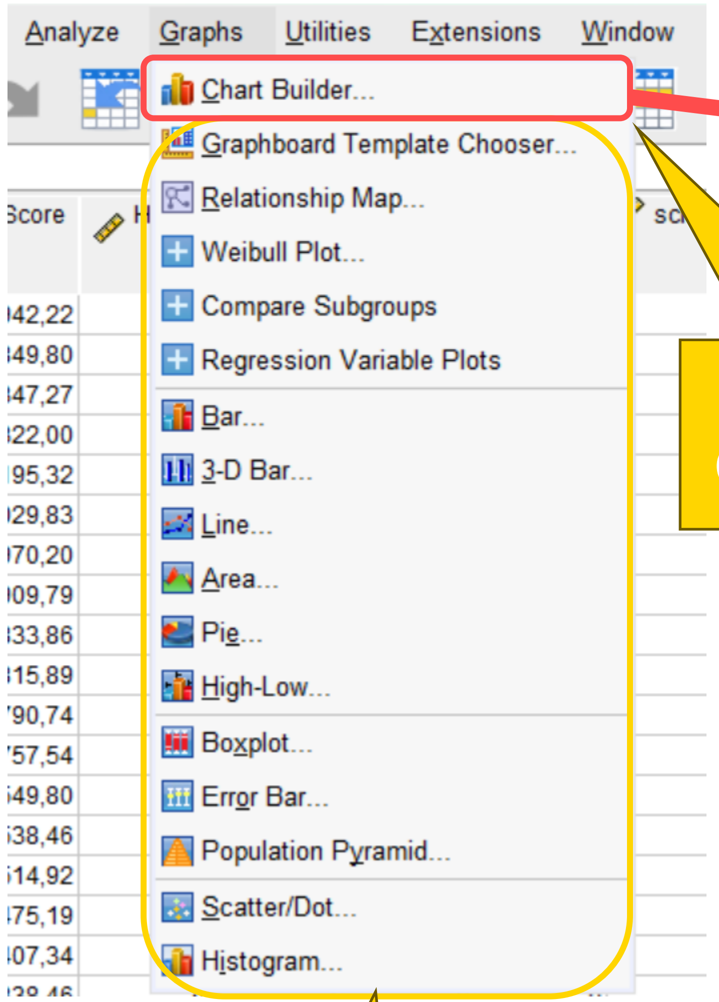
- Standardization is usually used to compare variables with different units by transforming the measured values so that they have a given mean and variance
- In z-standardization, the mean value of the variable $\bar{x}$ is subtracted from each measured value x_i and the resulting difference is divided by the standard deviation s_{emp} :

$$z_i = \frac{x_i - \bar{x}}{s_{emp}}$$

- If this is carried out for all measured values of a variable, the result for the z-standardized measured values is a mean value of 0 and a standard deviation of 1

- Note: z-standardization does not change the direction of the deviations of the individual measured values from the respective mean value!
- Interpretation: A z-value indicates by how many standard deviations a value deviates from the mean value

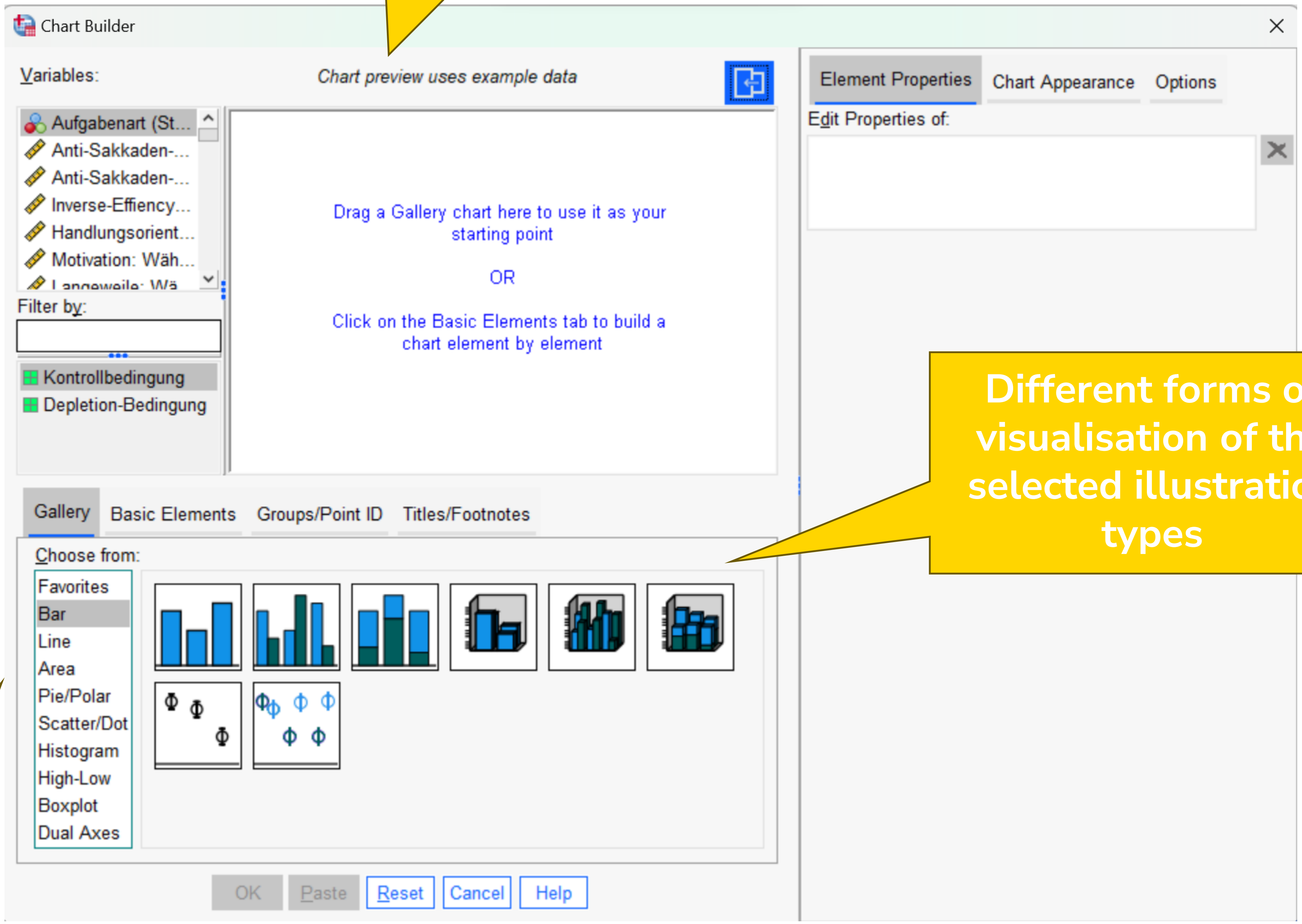
Graphical representations



„general“ Chart Builder

Direct selection of different presentation types

"Preview" of the resulting figure

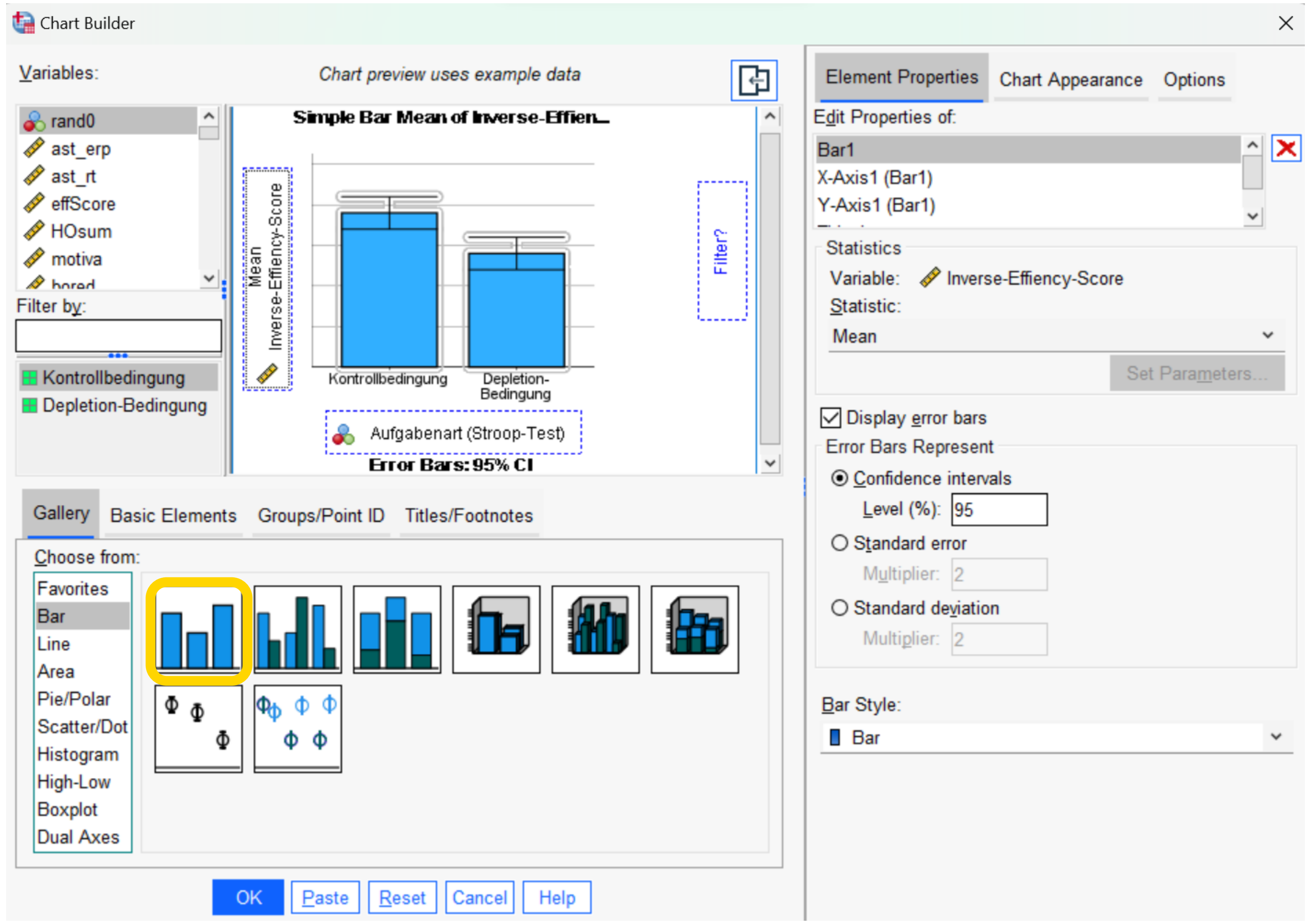


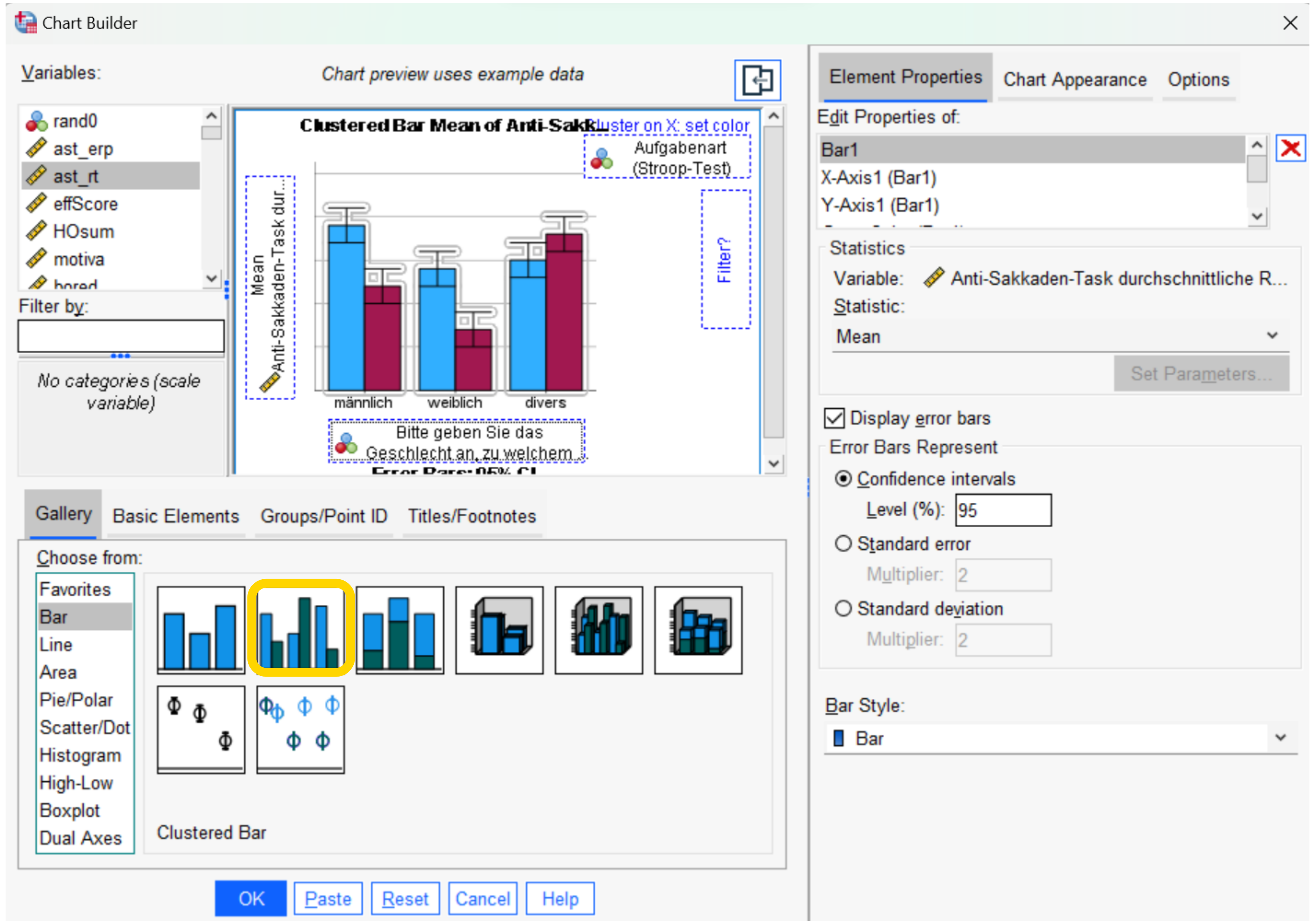
Possible types of figures

Different forms of visualisation of the selected illustration types

Graphical representations
Bar chart

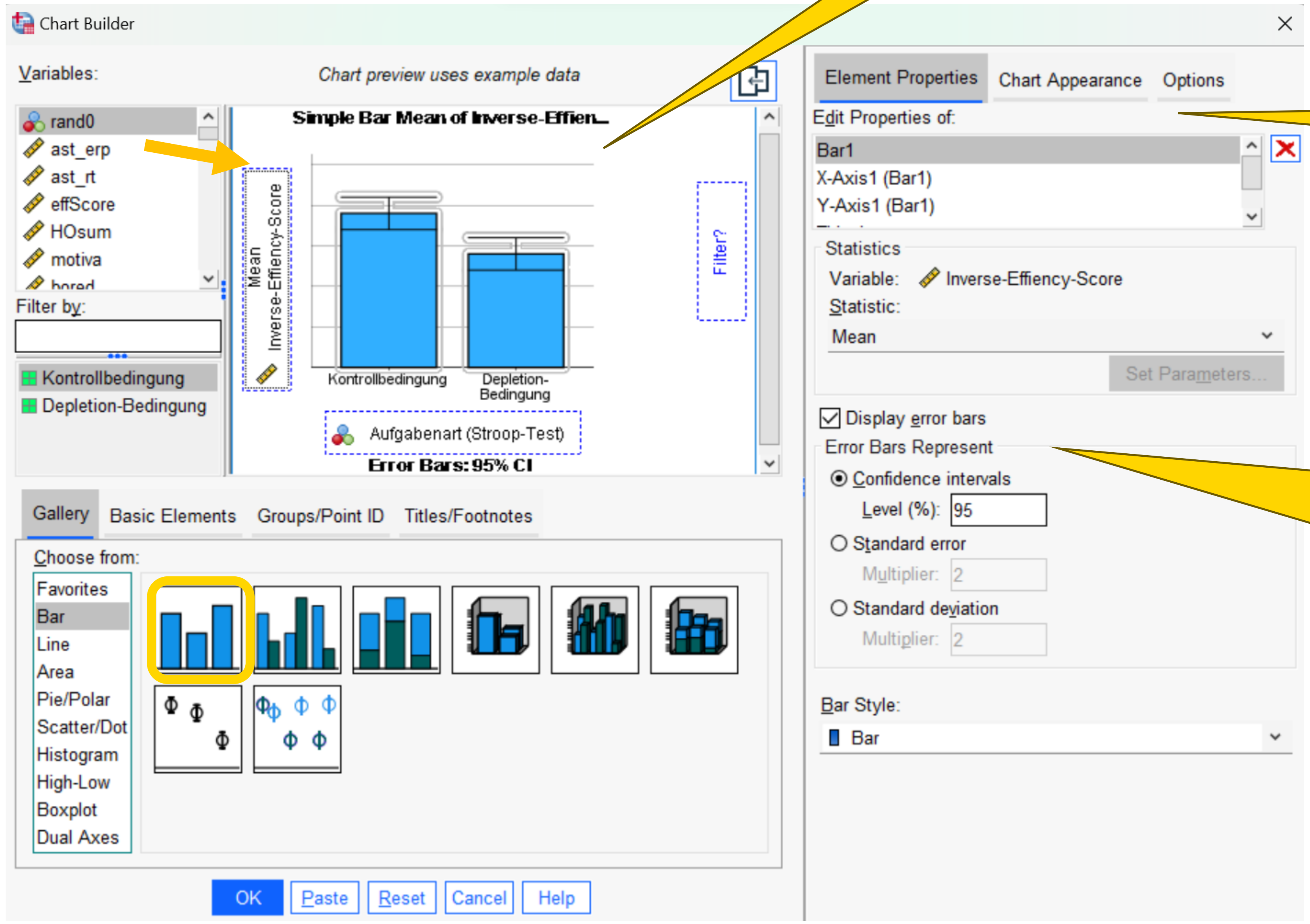
Try it out to get a feel for the data!





Graphical representations
Bar chart

Drag & Drop selection of your Variables



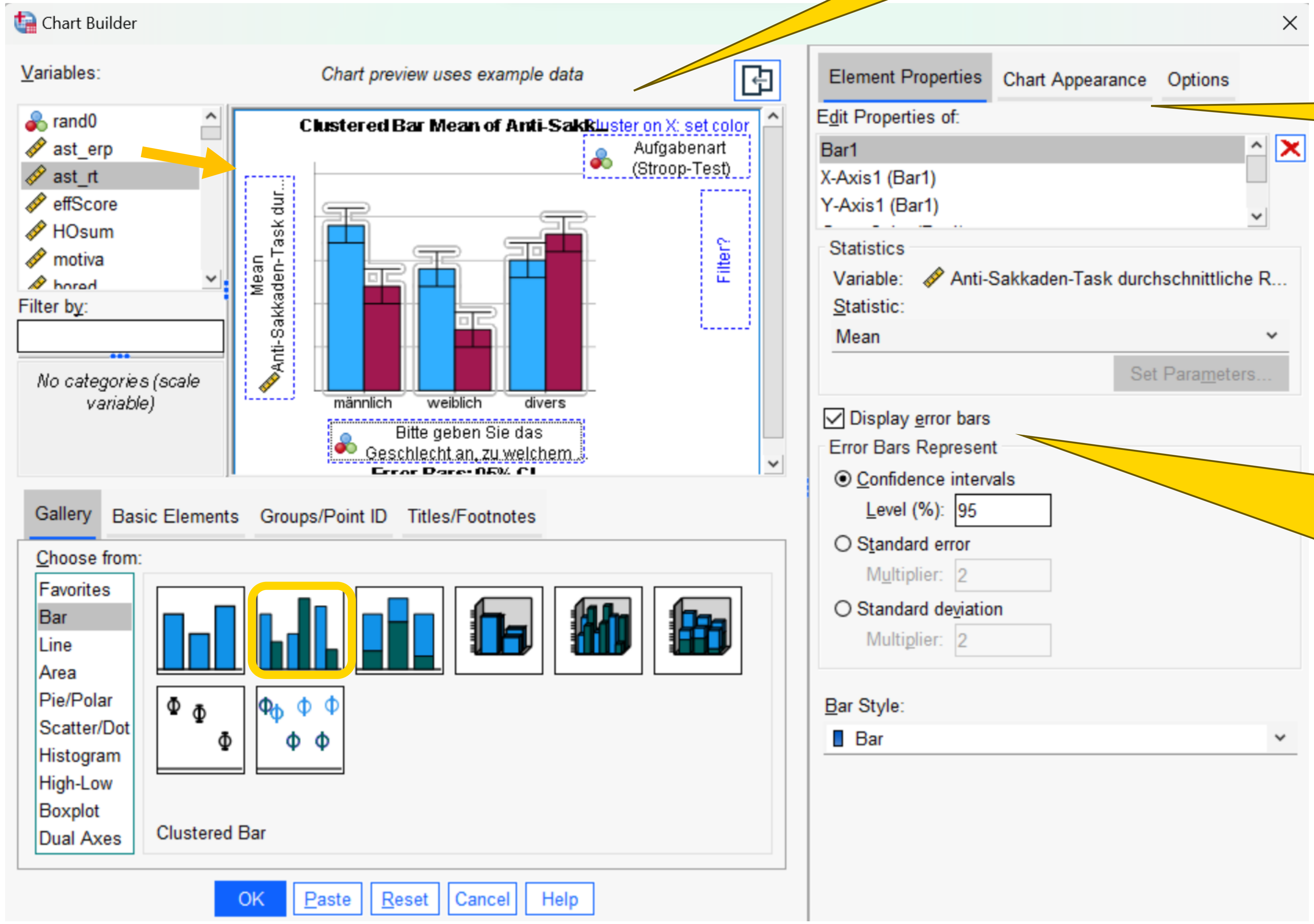
Other fitting methods -> simply try them out and "explore"

Insert error bars in the graphic by selecting "Bar1" under "Edit Properties of"

Note: Always indicate what the error bars represent in illustrations

Graphical representations
Bar chart

Drag & Drop selection of your grouping variable



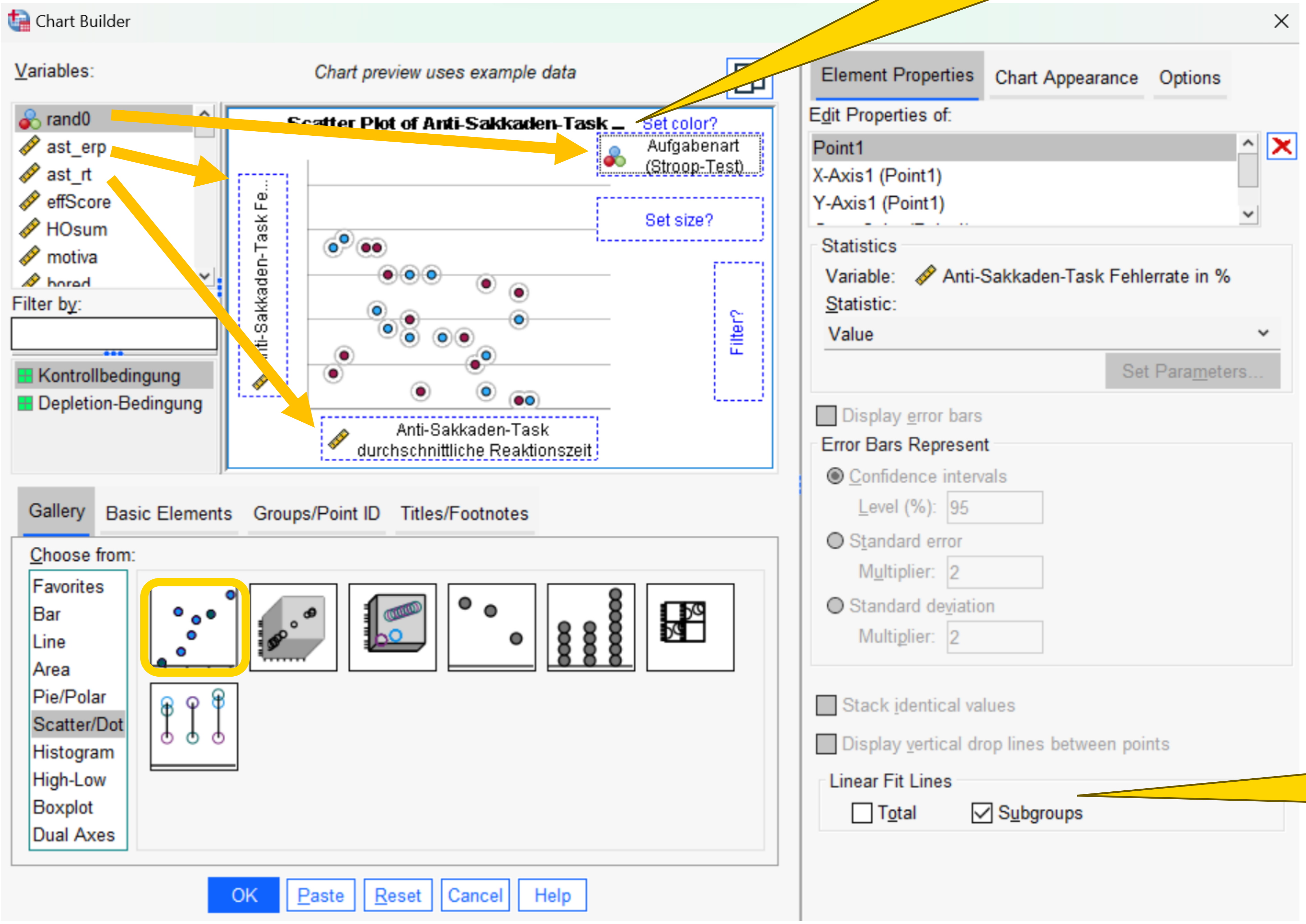
Other fitting methods -> simply try them out and "explore"

Insert error bars in the graphic by selecting "Bar1" under "Edit Properties of"

Note: Always indicate what the error bars represent in illustrations

Graphical representations
Scatterplot

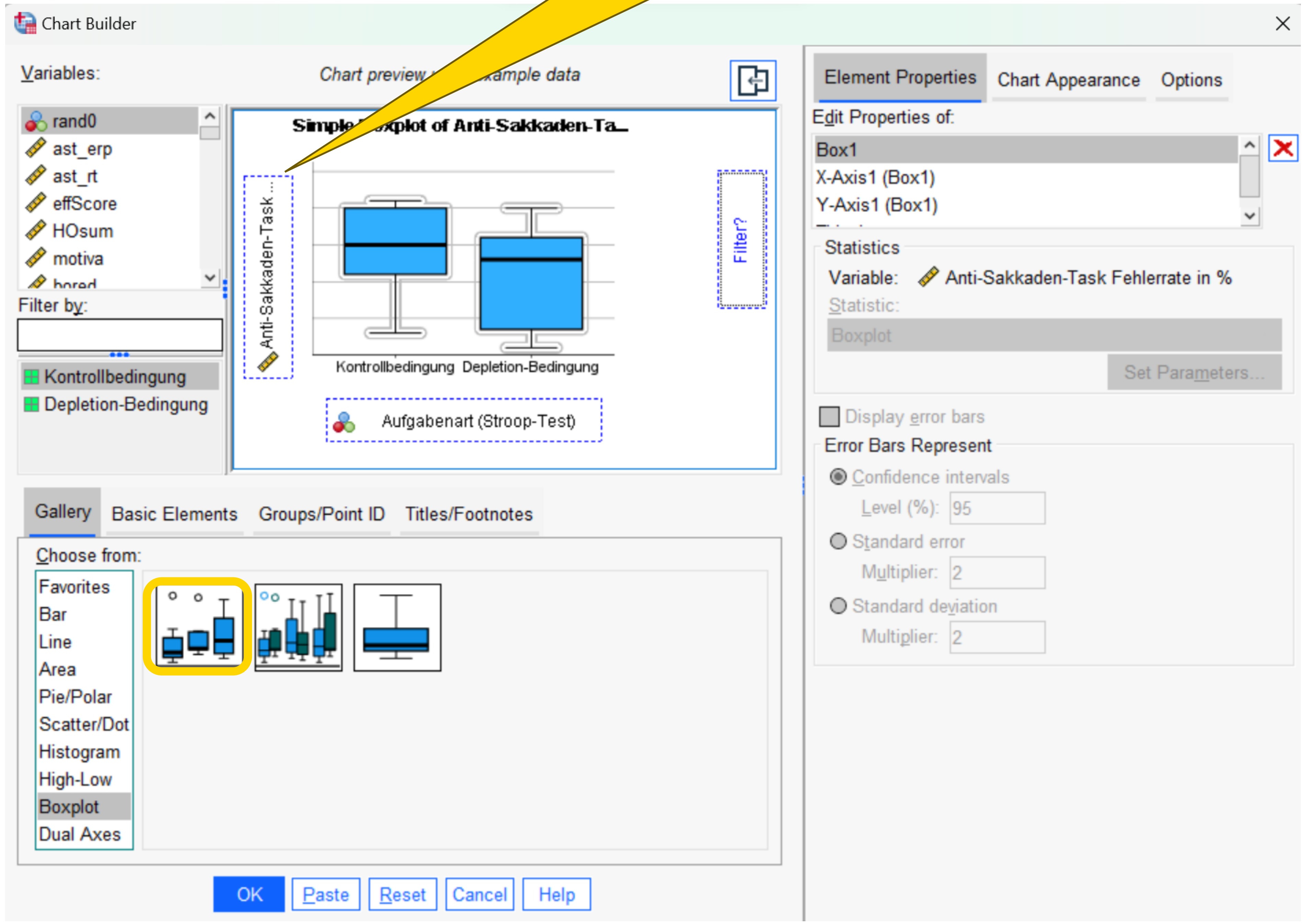
Possibility to display different groups in different colors



Possibility to display Fit-Lines

Graphical representations
Box-Plot

Possibility to define only the y-axis to search for outliers in the selected variable

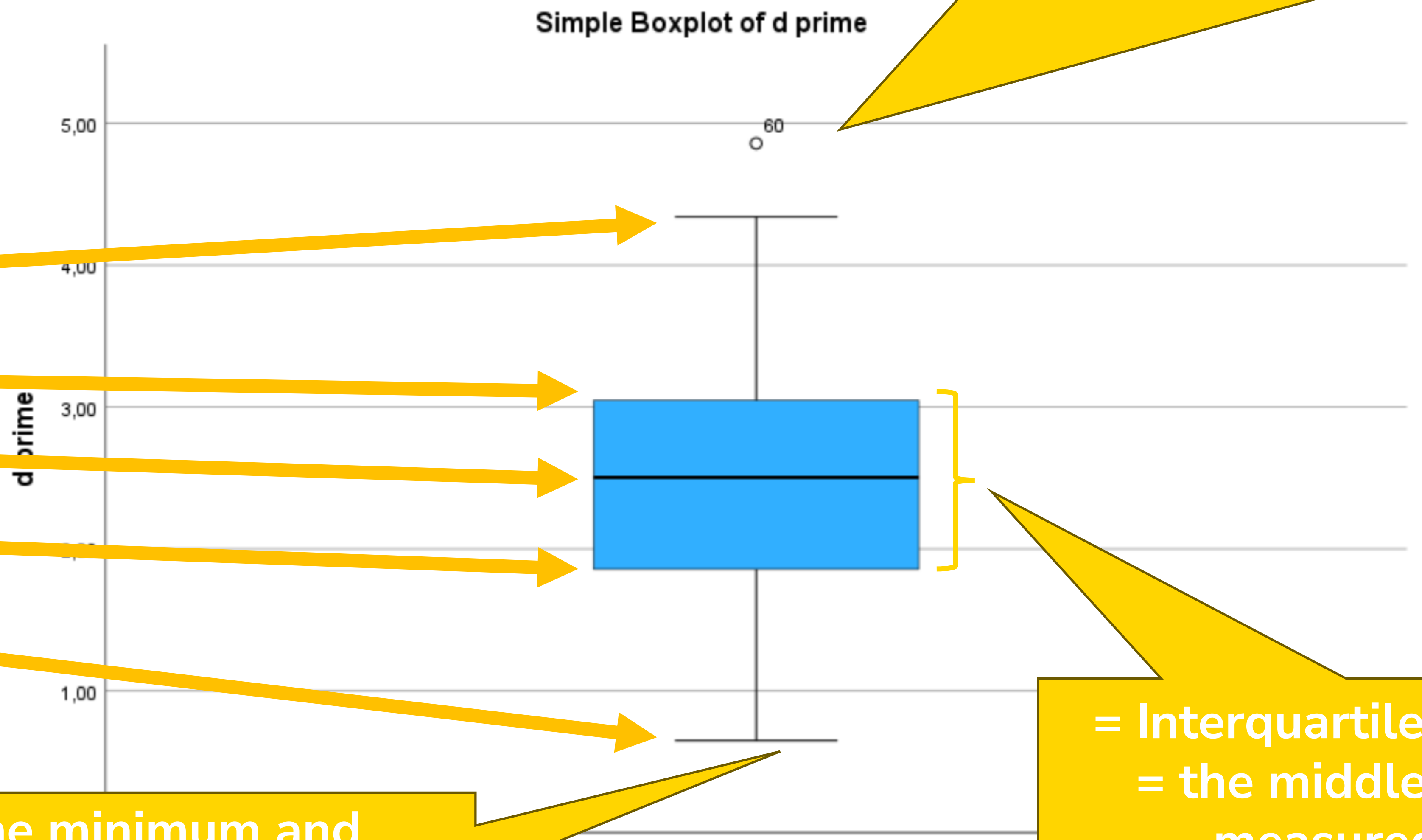


Graphical representations
Box-Plot

Outliers, if present, are directly marked, with the number indicating the row in the data file to which this data set belongs. Outliers that are more than 1.5 times the IQR away from the third or first quartile are marked with a circle, while those at least 2.5 times the IQR (i.e., "extreme outliers") are marked with a star.

- Five relevant values are displayed:
  - Maximal Value
  - Quartil Q3
  - Quartil Q2 (= Median!)
  - Quartil Q1
  - Minimal Value

  (+ Outliers)



= Interquartile Range (IQR) = the middle 50% of all measured values

The horizontal lines indicate the minimum and maximum measurement values, excluding any marked outliers.

Outliers in the data

- Are noticeably higher/lower than the majority of values and do not appear to fit with the remaining data points.
- Often have a disproportionate influence on statistical analyses → leading to skewed results.
- **Identification**
  - Visuel: Scatterplots, Boxplots (see Box-Plot slides)
  - Statistical: z-values & 1.5×IQR-rule

Especially in small samples!
However, in small samples, values may be considered outliers that, in reality, represent entirely plausible values. Excluding them could introduce bias or lead to an underestimation of variance within the population.

- **How to handle outliers**
  - Consider possible causes of extreme values. (real values vs. Data entry errors)
  - Set a threshold: Exclude certain values. (e.g. ± 2.5/3 *SD* assuming a normal distribution)
  - Data transformation (Logarithmize) to reduce the influence
  - Robust Methods (Median insted of Mittelwert, IQA instead of *SD*, robust Regression/Estimations, non-parametric Tests)
  - (Bootstrapping)

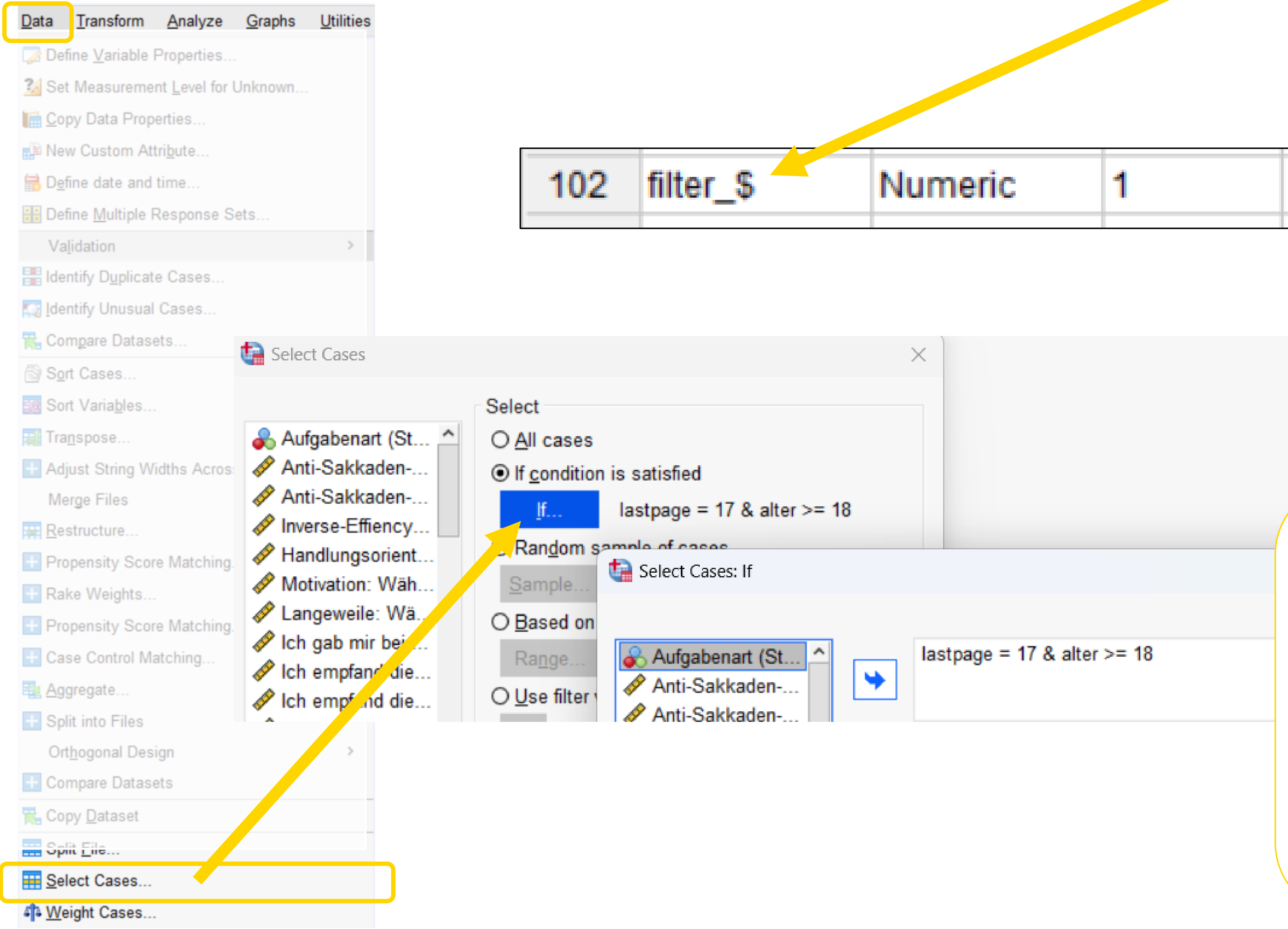
The approach should always be well thought out, logically justified, and documented.
Comparing the results with outliers and without outliers can also be insightful!

Filter 1/2

- Reduction of the dataset to relevant cases for analysis (e.g., exclusion of outliers)
- Targeted analysis of subgroups
- SPSS creates a corresponding filter variable in the dataset!

Exemplary syntax command *(easy)*

```
USE ALL.
COMPUTE filter_$=(lastpage=17 &
alter>=18).
FILTER BY filter_$.
EXECUTE.
```



Exemplary syntax command *(more complex → via „PASTE“)*

```
USE ALL.
COMPUTE filter_$=(lastpage = 17 & alter >= 18 ).
VARIABLE LABELS filter_$ 'lastpage = 17 & alter >= 18 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

→ Enhances the documentation and user-friendliness of the syntax code!

Filter 2/2

- It is possible to implement multiple Filters at once

Syntax command ***to deactivate Filters:***

```
Filter OFF.
USE ALL.
EXECUTE.
```

- **Important** at the end: "Turn off" the filter.
  - Either through the graphical user interface or via syntax.
  - Otherwise, the filter remains active, which can lead to errors in further analyses

Filtering Outliers

Example:

Only individuals whose average reaction time deviates by a maximum of 2.5 standard deviations from the overall sample's average reaction time should be analyzed.

Procedure:

1. Standardize the variable with the average reaction time using the z-score *(via „Descriptives", see slide „Descriptive Statistics – Descriptives")*
2. Define a filter so that individuals with values < -2.5 and > 2.5 in the newly z-standardized variable are excluded.

100	ZSco01	Numeric	11	5	Zscore(ast_rt) ...

USE ALL.
COMPUTE filter_$=(ZSco01 > -2.5 & ZSco01 < 2.5).
VARIABLE LABELS filter_$ 'ZSco01 > -2.5 & ZSco01 < 2.5 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

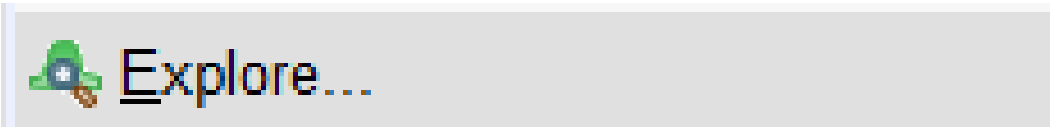
	ZSco01
~~1~~	-2,84115
40	,13652
41	,15209
42	,20919
43	,20919
44	,29224
45	,31819
46	,32338
47	,35453
48	,35972
49	,38567
50	,42200
51	,48429
52	,48429
53	,58292
54	,61925
55	,65040
56	,74902
57	,76978
58	,86322
59	,88398
60	,92550
61	,93589
62	,97741
63	1,05008
64	1,20061
65	1,66258
66	1,71968
67	1,83907
68	2,39447
~~69~~	2,56577
~~70~~	2,59691

Check for normal distribution 1/7

Theoretically, there are three possibilities:

1. **Inferential statistical testing**: Kolmogorov-Smirnov test or Shapiro-Wilk test
2. **Descriptive statistical testing**: Using skewness and kurtosis
3. **Graphical testing**: For example, using a histogram or Q-Q plot

→ via „Explorative data analysis“

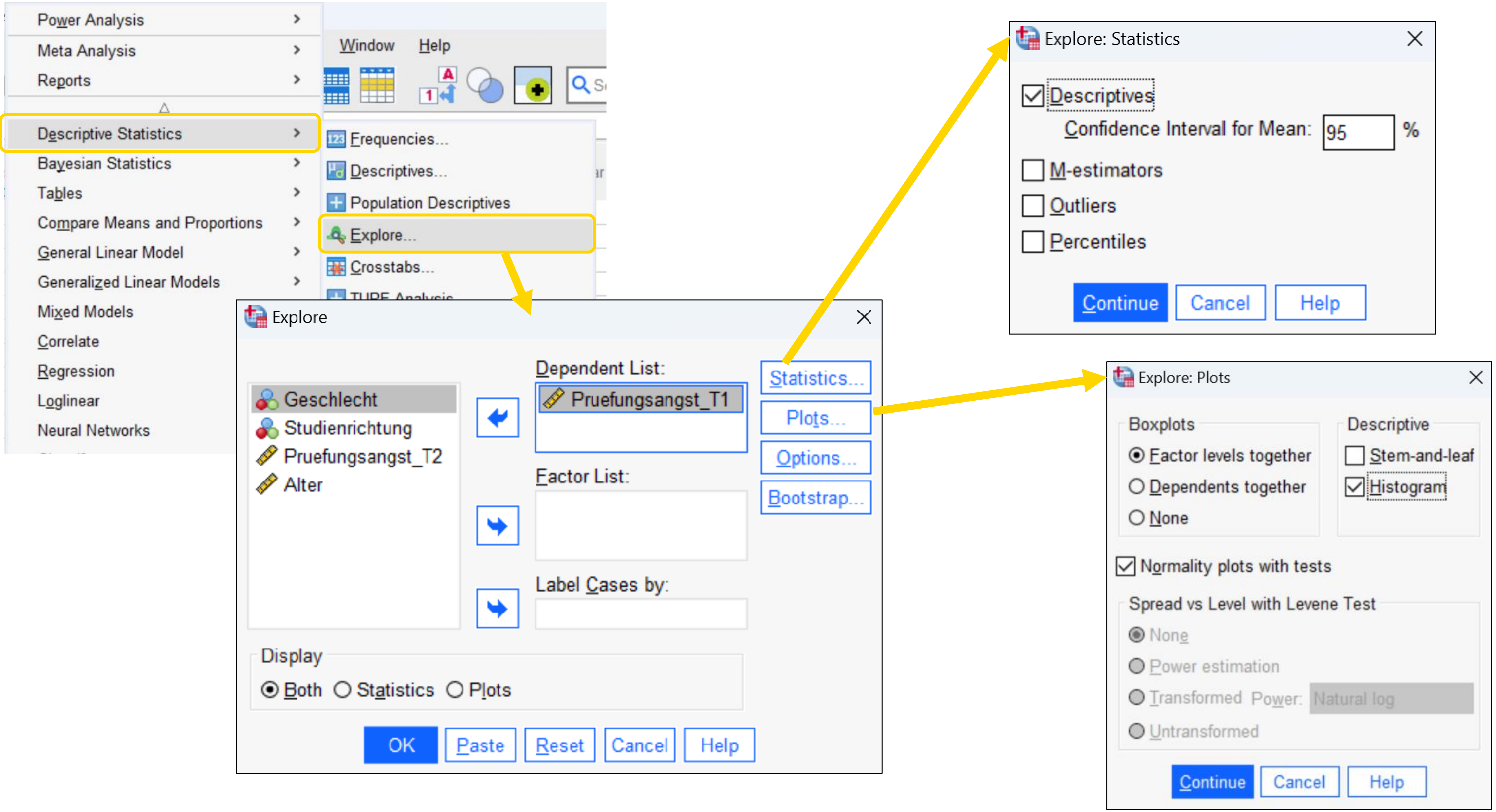


Note:
In some areas, skewness and kurtosis are preferred for assessing the normality of a variable (due to the disadvantages of inferential statistical methods). However, inferential statistical methods can generally still be used.

References to check:

- Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, 22(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18
- Kim H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative dentistry & endodontics*, *38*(1), 52–54. https://doi.org/10.5395/rde.2013.38.1.52

Check for normal distribution 2/7



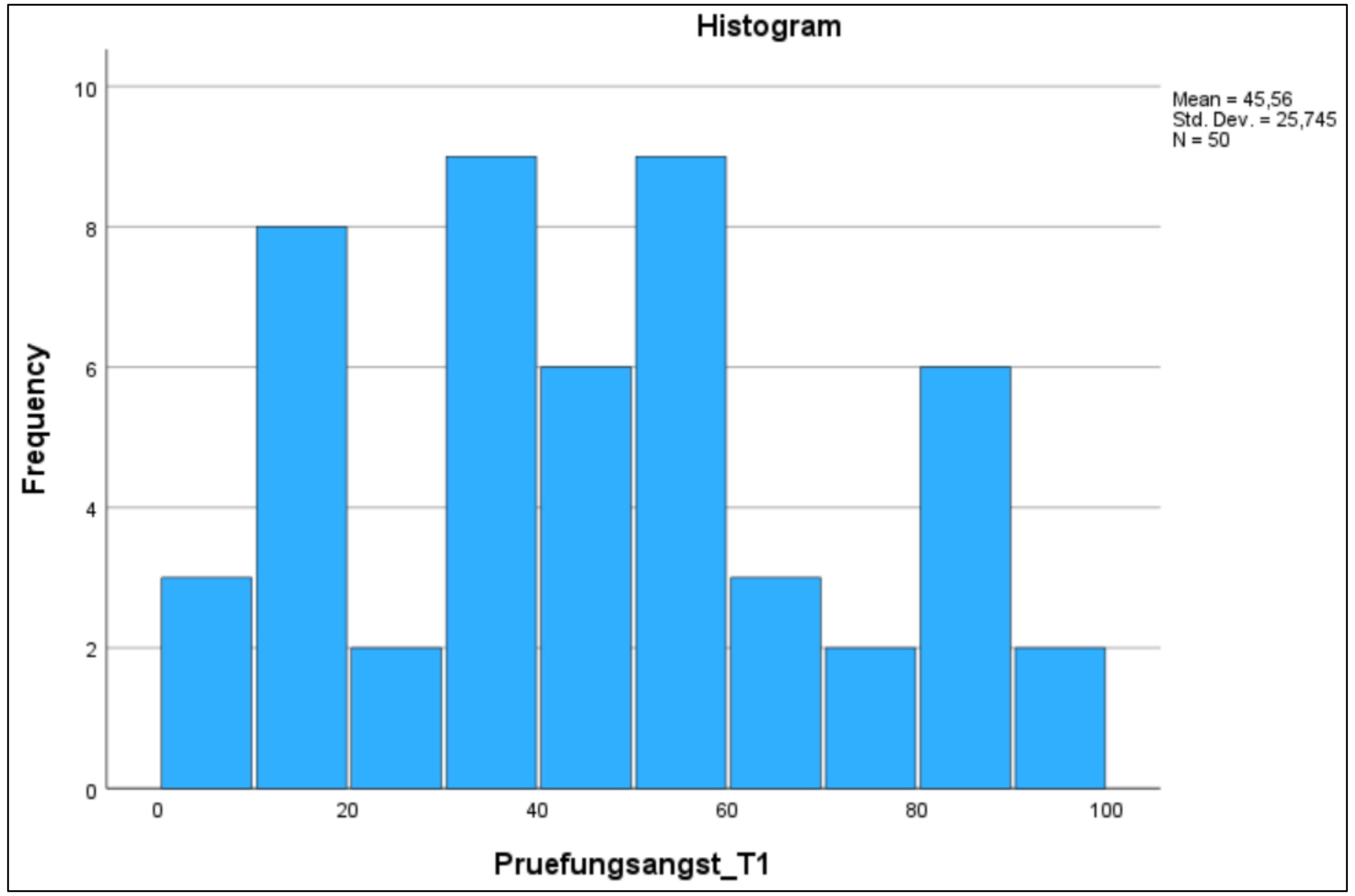
Check for normal distribution 3/7

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Pruefungsangst_T1	50	100,0%	0	0,0%	50	100,0%

Descriptives

			Statistic	Std. Error
Pruefungsangst_T1	Mean		45,56	3,641
	95% Confidence Interval for Mean	Lower Bound	38,24	
		Upper Bound	52,88	
	5% Trimmed Mean		45,28	
	Median		45,50	
	Variance		662,782	
	Std. Deviation		25,745	
	Minimum		0	
	Maximum		98	
	Range		98	
	Interquartile Range		35	
	Skewness		,179	,337
	Kurtosis		-,837	,662



Testing for normality using **skewness** and **kurtosis**:
If both are **less than 1** in **absolute value**, we can assume the variable is normally distributed
If the **skewness is less than 1** in absolute value and the **kurtosis is less than 5** in absolute value, we can still assume an **acceptable fit**

Check for normal distribution 4/7

- Both the Kolmogorov-Smirnov test and the Shapiro-Wilk-test test **the null hypothesis that the measurements of the respective variable are normally distributed**
- A **significant result** indicates that the data are **not normally distributed**

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pruefungsangst_T1	,096	50	,200*	,968	50	,185

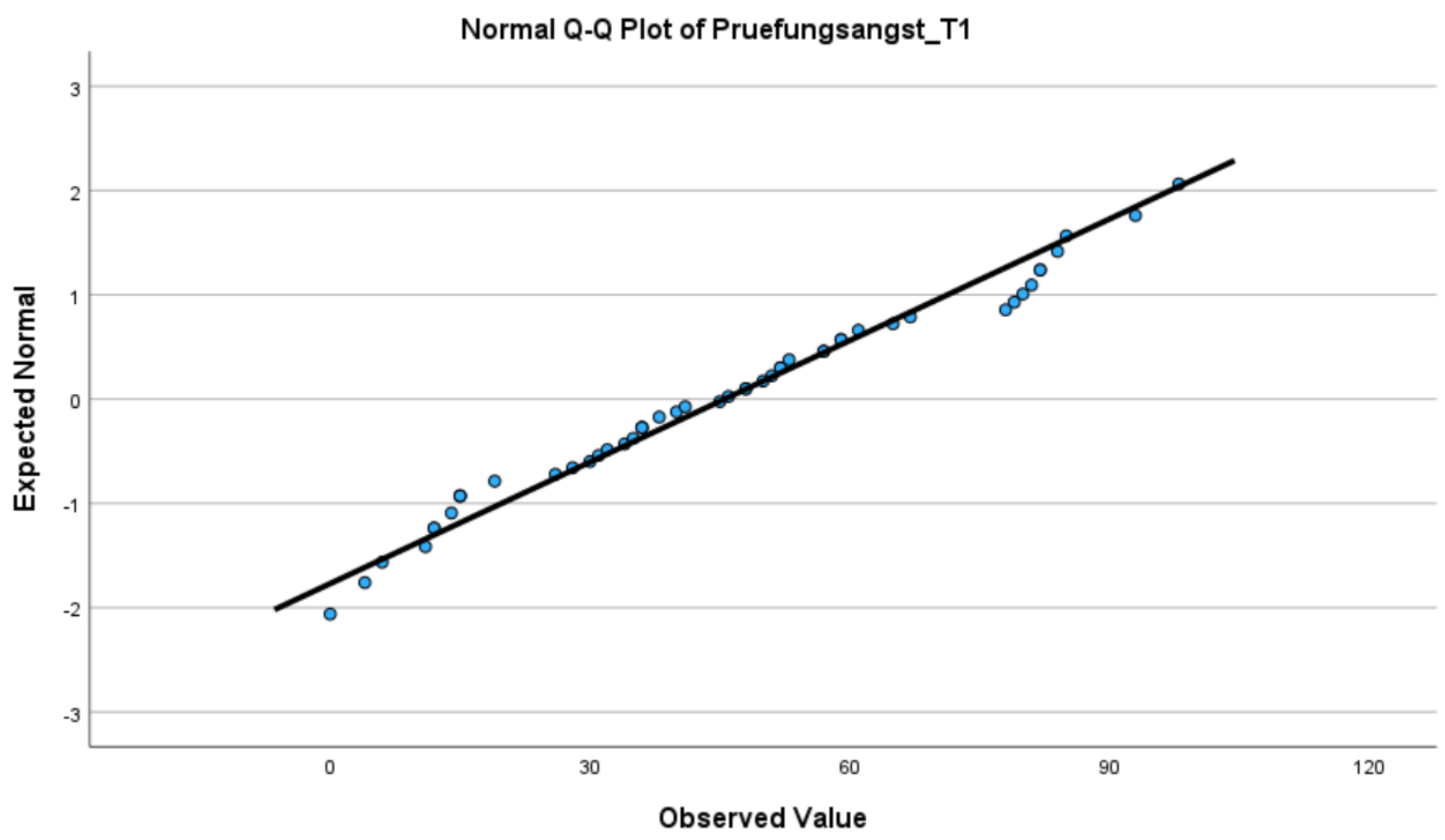
*. This is a lower bound of the true significance.

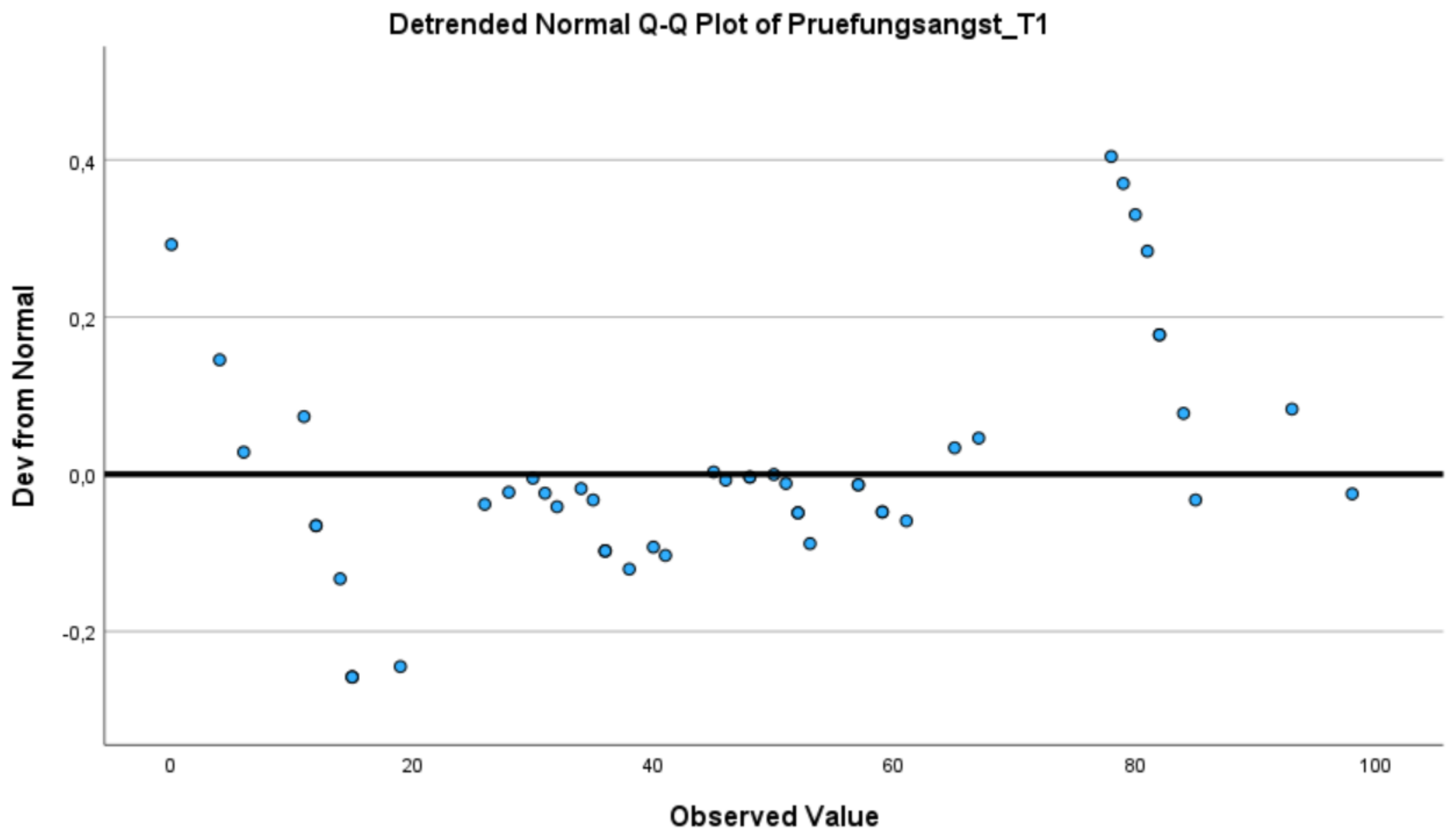
a. Lilliefors Significance Correction

Problem with these tests:

- Too high power in very large samples → significant results even for very small and practically insignificant deviations from "perfect normality“
- Too low power in small samples → "problematic" deviations from normality are not detected, leading to no significant result

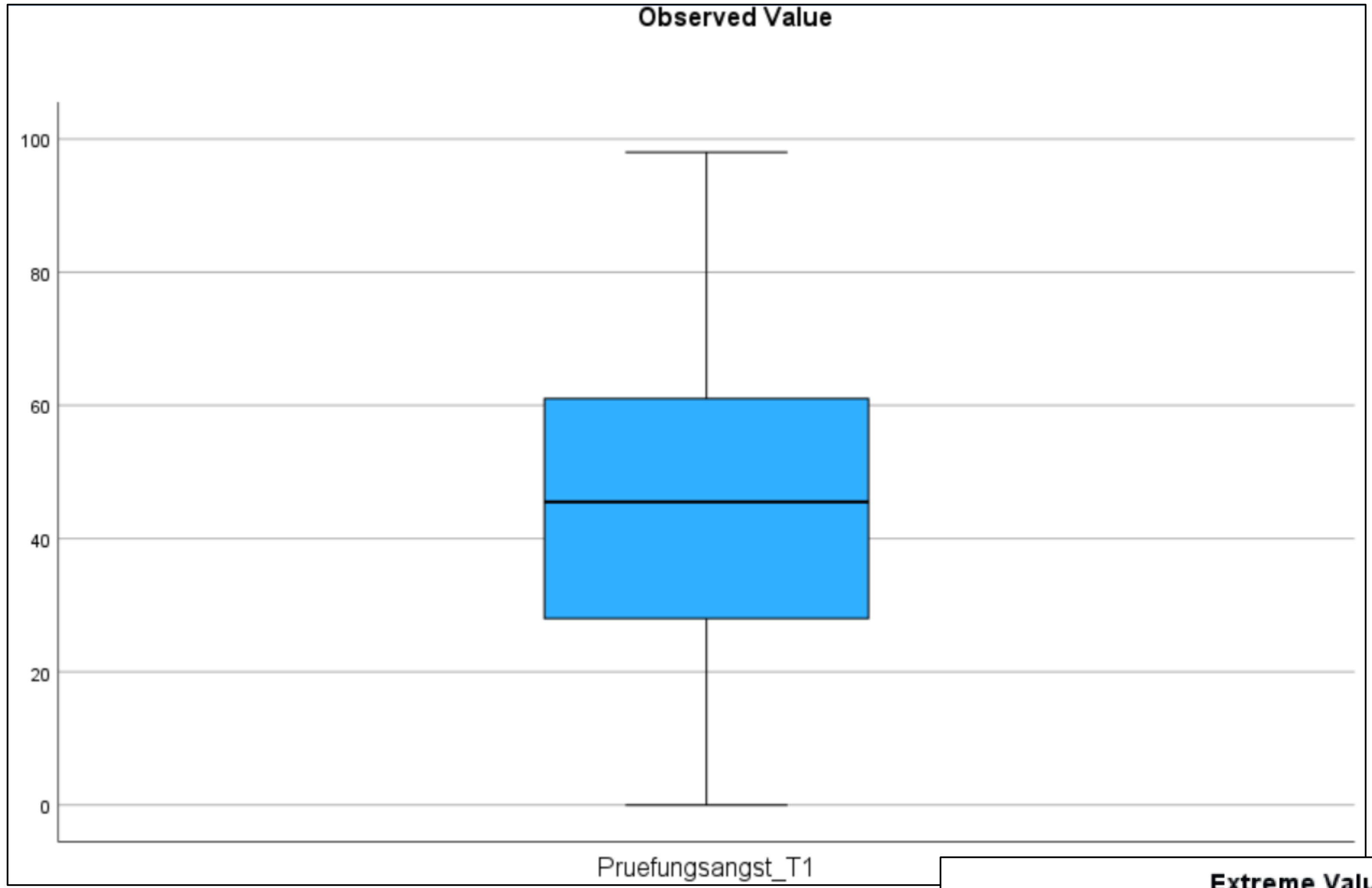
Check for normal distribution 5/7





In a "perfect" normal distribution, all data points in both graphs would lie on the black line. Q-Q plots provide a graphical illustration that can help better understand the data (especially the left plot). However, they should not be used as the sole criterion for deciding whether a variable is normally distributed.

Check for normal distribution 6/7



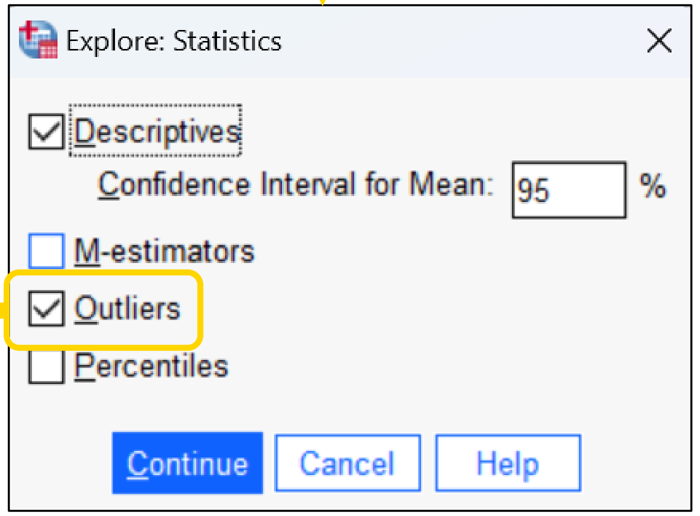
The exploratory data analysis additionally provides boxplots, where potential outliers are directly marked (see slides on boxplots). However, the decision on what to do with outliers should again depend on substantive considerations and the expected impact of the outliers on the planned analysis. There is also the option to output extreme values of the variable, which can help gain a better understanding of the data

Extreme Values

			Case Number	Value
Pruefungsangst_T1	Highest	1	20	98
		2	46	93
		3	45	85
		4	16	84
		5	32	82[a]
	Lowest	1	44	0
		2	40	4
		3	24	6
		4	49	11
		5	50	12[b]

a. Only a partial list of cases with the value 82 are shown in the table of upper extremes.

b. Only a partial list of cases with the value 12 are shown in the table of lower extremes.



Check for normal distribution 7/7

What to do if one or more variables to be analyzed are not normally distributed?

- Transform the data (e.g., logarithmize)
- Use the resampling method of bootstrapping
- Argue using the central limit theorem
- (Use non-parametric methods)

Be transparent and report it!

Reliability analyses/"internal consistency"

- To assess the reliability (measurement accuracy, dependability, stability) of tests and measurements
- Reliability values can be considered as the proportion of the true variance to the total variance of the test.

$$\alpha = \frac{m}{m-1}\left(1 - \frac{\sum_{i=1}^{m} s_i^2}{s_t^2}\right)$$

m = number of parallel measurements (Items)
s_i^2 = Variance of the i-th parallel measurement (items)
s_t^2 = Variance of the test t(sum of all item values)

- **Cronbachs Alpha (α)**
  - Dependent on: 1) number of items, 2) variance of the items, 3) variance of the test scores, 4) Covariance/intercorrelation of the items
  - α = Sample-dependent | the more items the higher α
  - It is a generalization of the split-half method (where each item is considered as an independent test part)
  - A high α Is desirable for homogeneous constructs/tests

A high value for α does NOT AUTOMATICALLY mean that a unidimensional or consistent construct is being measured!

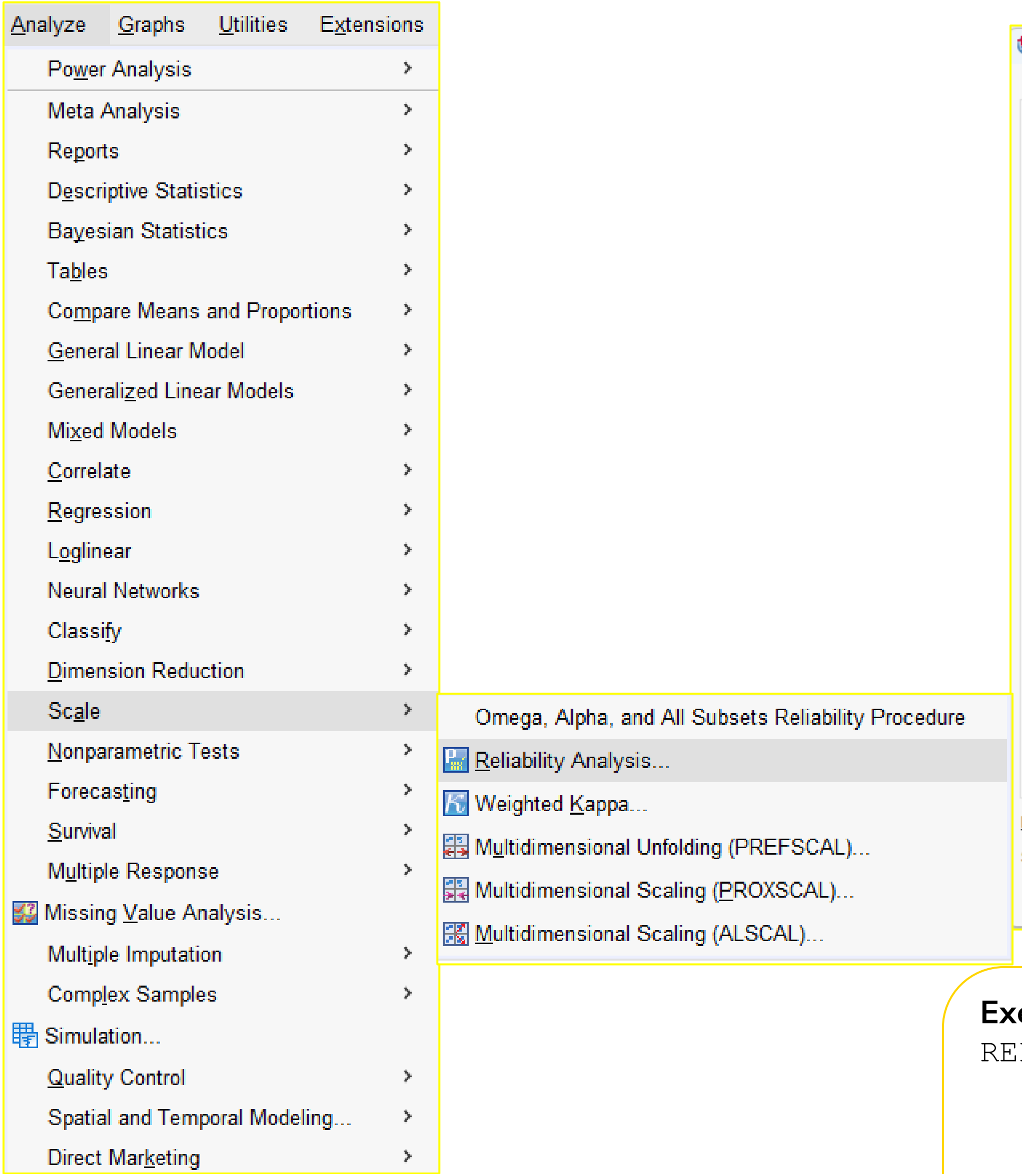
→ **.7** (Heterogeneous constructs, personality tests)
.8 - .9 (Homogeneous constructs) | **>.9** (Performance tests)

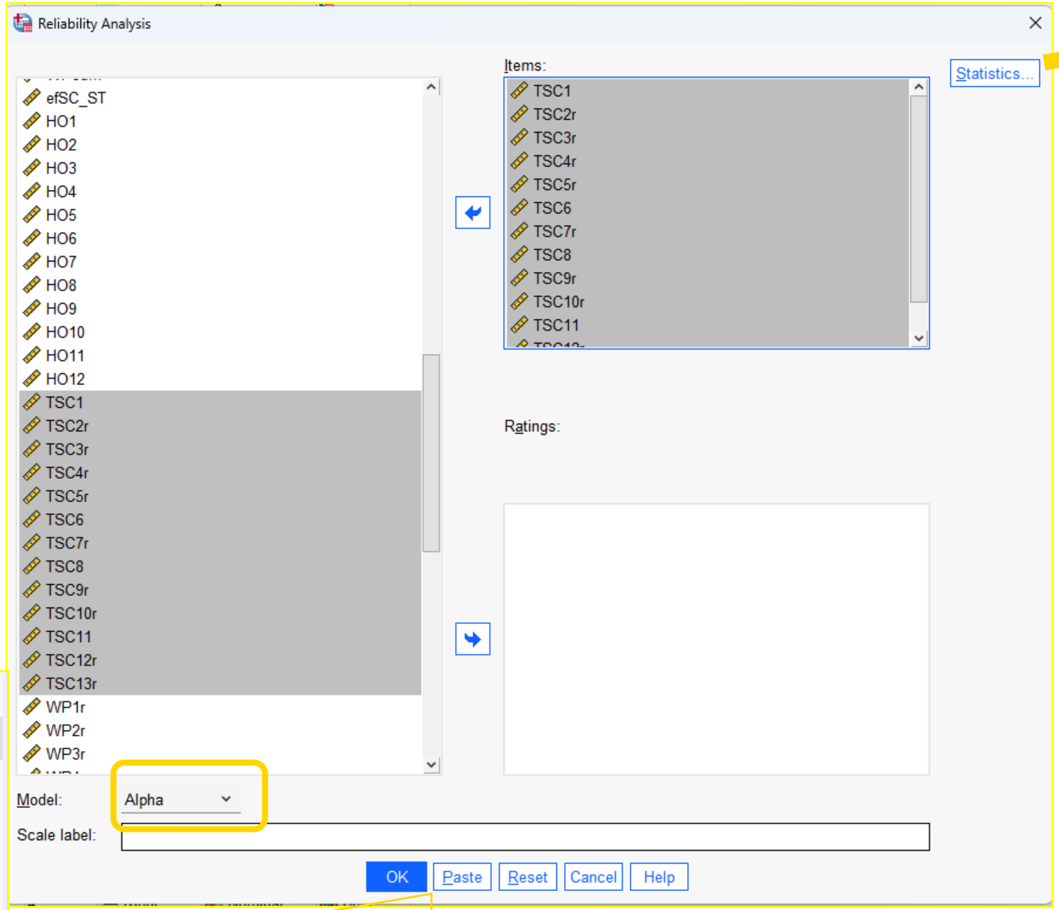
CAUTION: The frequently cited Kuder-Richardson formula (KR20) for **dichotomous items** is not explicitly provided in SPSS, but an adequate approximation can be calculated using Cronbach's alpha.

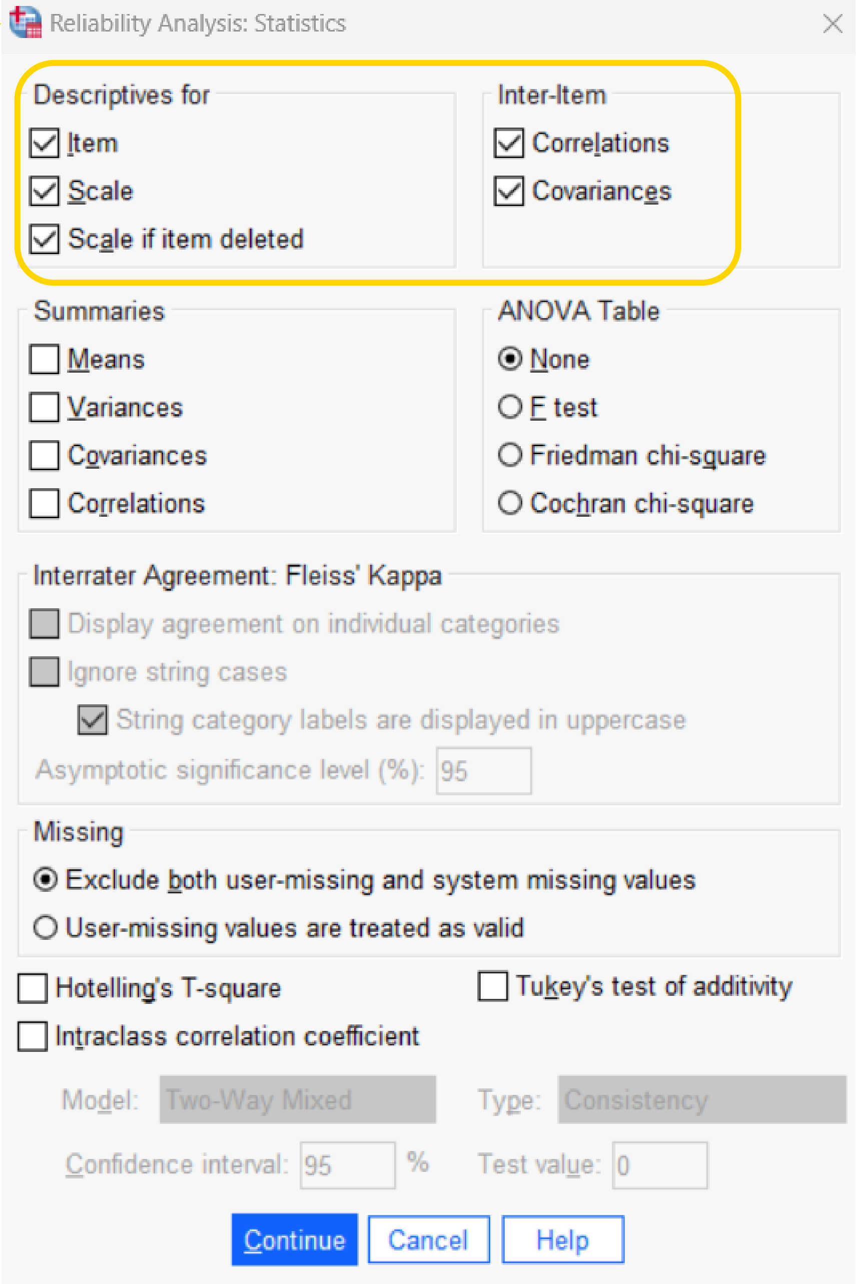
- Alternative: **McDonalds Omega (ω)**
  - Less strict assumptions (~~identical unstandardized loadings and error variances; essential τ-equivalence~~)
  - Recommended when the items correlate with the latent construct to varying degrees.
  - Integrated since version 27 in SPSS! -> you can use it :)

These assumptions are (unfortunately) often violated in empirical research!

Reliability analyses/"internal consistency"







Exemplary syntax command

```
RELIABILITY
  /VARIABLES=TSC1 TSC2r TSC3r TSC4r TSC5r TSC6
             TSC7r TSC8 TSC9r TSC10r TSC11 TSC12r TSC13r
  /SCALE('ALL VARIABLES') ALL
  /MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE SCALE CORR COV
  /SUMMARY=TOTAL CORR.
```

1. Items of a scale are dragged into the right field (be careful with **reverse-coded items** → use them accordingly!)
2. Click on the framed statistics
3. Copy the command into the syntax and execute it

Reliability analyses/"internal consistency"

Case Processing Summary

		N	%
Cases	Valid	70	100,0
	Excluded[a]	0	,0
	Total	70	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,846	,846	13

Uses correlations instead of covariances. This can be used, for example, when the items of a test are measured in different units.

What values should be aimed for?

- **>.70** (Heterogeneous constructs, personality tests)
- **.80 - .90** (Homogeneous constructs)
- **> .90** (Performance tests)

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item Correlations	,296	,040	,615	,576	15,534	,017	13

The mean of the correlations between all item pairs within the measurement instrument. Very high inter-item correlations may indicate redundant items.
.15 - .50 is acceptable

Reliability analyses/"internal consistency"

= **Discriminative power** (correlation of an item with the scale, excluding the respective item)

- Items of a scale should be able – when there is a high reliability – to distinguish between individuals with low and high levels of the trait → requires discriminative items.

>.25 is acceptable; .40 - .70 is good

= Improvement/deterioration of Cronbach's α by removing the respective item from the scale.

- *In the present case, removing individual items would not lead to an improvement in α.*

Item-Total Statistics

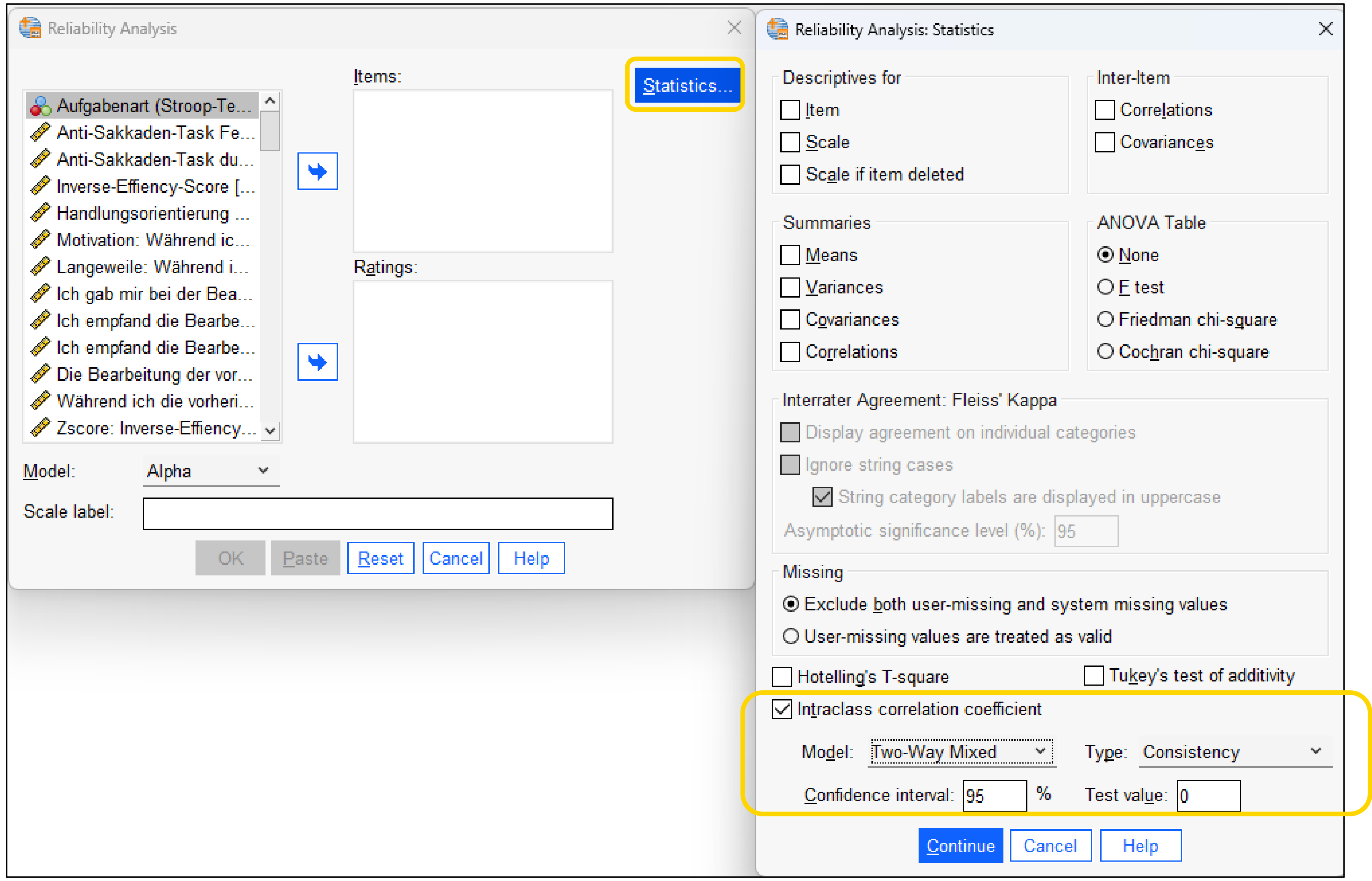
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
TSC1	46,66	91,968	,422	,435	,840
TSC 2 Recode	46,84	88,714	,512	,358	,834
TSC 3 Recode	45,69	88,219	,470	,539	,837
TSC 4 Recode	45,56	87,265	,527	,581	,833
TSC 5 Recode	47,04	88,911	,499	,418	,835
TSC6	46,57	93,205	,393	,319	,841
TSC 7 Recode	47,56	87,729	,548	,487	,831
TSC8	46,93	88,125	,615	,544	,828
TSC 9 Recode	46,83	90,840	,452	,401	,838
TSC 10 Recode	45,86	86,472	,596	,584	,828
TSC11	45,87	93,099	,463	,494	,837
TSC 12 Recode	46,40	85,229	,646	,604	,824
TSC 13 Recode	45,80	93,264	,341	,244	,845

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,846	,846	13

Intraclass correlation coefficient (ICC)

- Is rarely used in the context of bachelor theses, so please refer to this video:
  - https://www.youtube.com/watch?v=qbM3ryAl4rE



t-test

Principle & Execution in SPSS

t-tests

"t-Test", because the test statistic calculated in this procedure follows a t-distribution!!

- checks whether the means of two groups differ significantly from each other or whether a mean differs significantly from a given value
  - **One-sample t-test**
    - *Does the mean of a sample differ significantly from a known or hypothetical population mean?*

$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}}$$

  - **Two-sample t-test for independent samples**
    - *Does the difference in means between two independent samples significantly deviate from a known or hypothetical population value?*
    - "The special case is the normal case": Do two populations differ in their mean?

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_{pool}^2}{n_1} + \frac{s_{pool}^2}{n_2}}}$$

  - **Two-sample t-test for dependent samples**
    - As with the two-sample t-test for independent samples, we are interested in the parameter differences here, but from dependent samples.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_{diff}^2}{n}}}$$

In all cases, the t-value quantifies how far the observed mean/the observed difference is from the expected value (under the null hypothesis) (numerator), measured in units of the standard error (denominator).

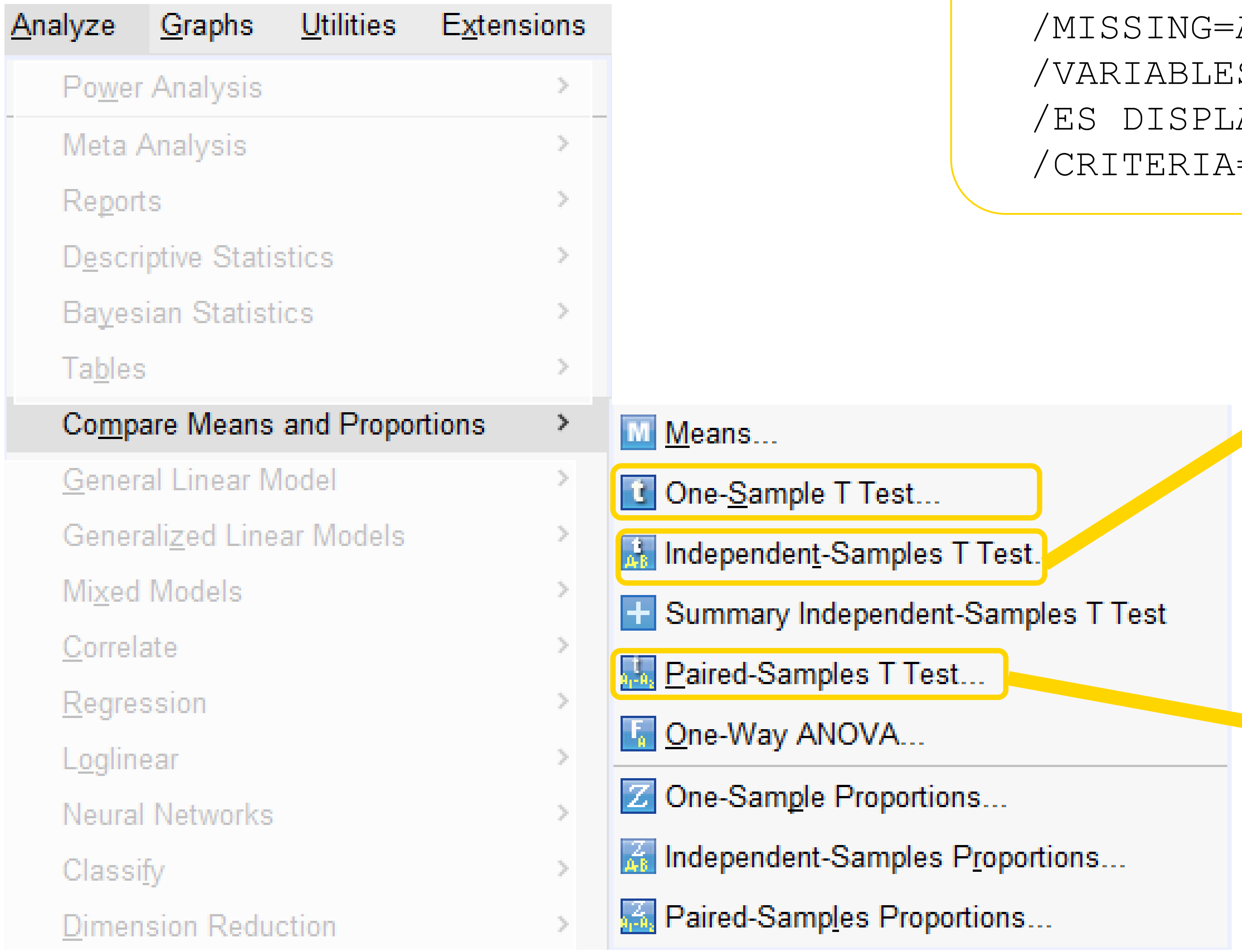
t-test: Assumptions

- **Interval-scaled** dependent variable
- **Normal distributed** dependent variable (*Check: see slides „Check for normal distribution“)*

- For two-sample t-test for independent samples: **Homogeneity of variance**
  - = Assumption that the variance of the dependent variable is equal across all groups of the factor
  - Checked using **Levene's test (H0: estimated population variances are equal)**
  - For a non-significant result, the values under "Equal variances assumed" are used for the t-test
  - For a significant result, the values under "Equal variances not assumed" are used for the t-test

In principle, t-tests are quite robust to violations of assumptions (meaning they still provide stable results even when assumptions are violated).

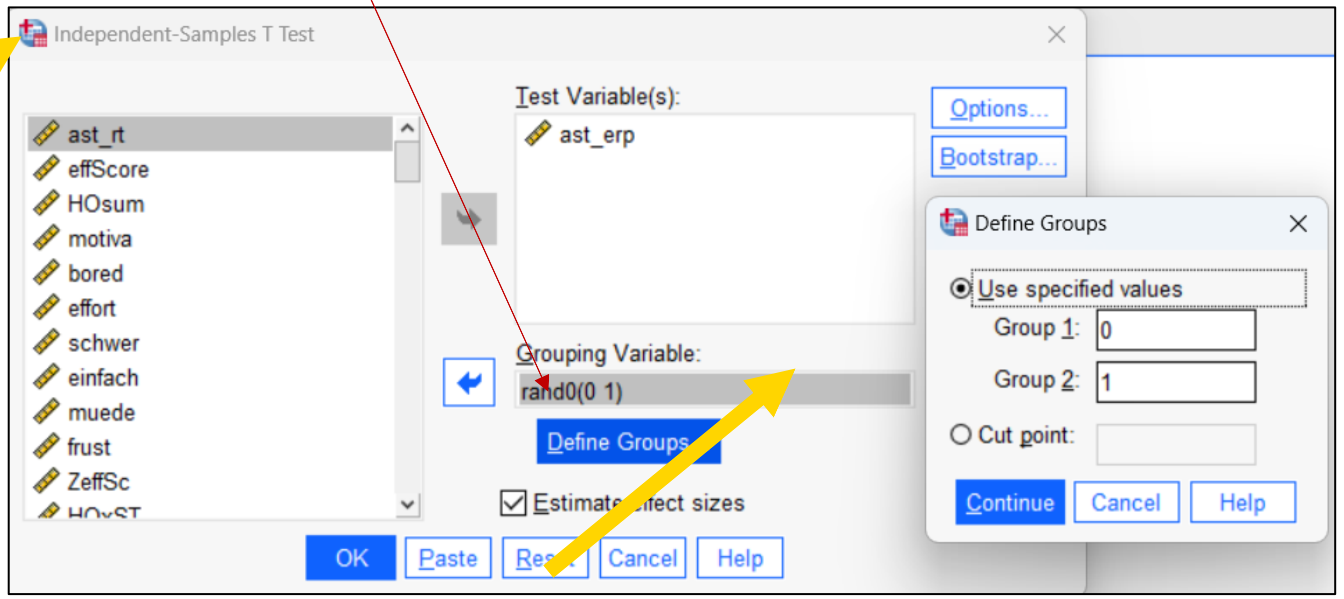
t-tests in SPSS 1/2

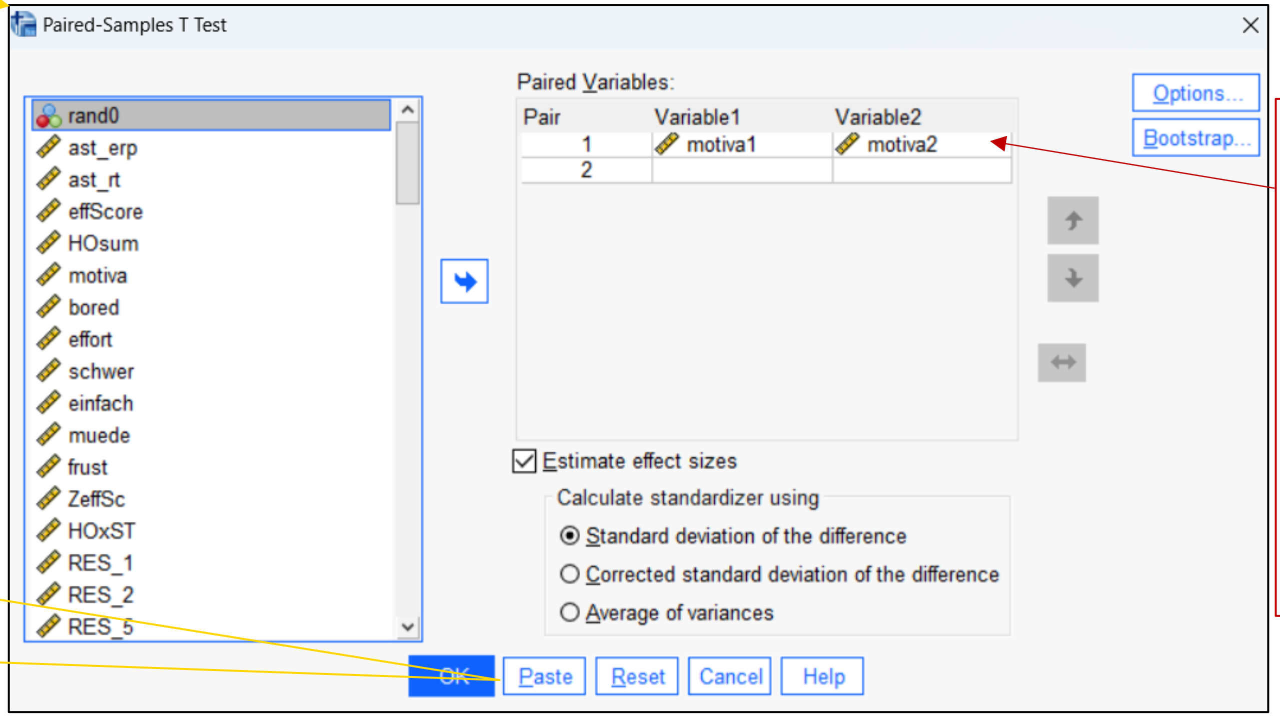


Exemplary syntax command

```
T-TEST GROUPS=rand0(0 1)
  /MISSING=ANALYSIS
  /VARIABLES=ast_erp
  /ES DISPLAY(TRUE)
  /CRITERIA=CI(.95).
```

The two groups are defined according to their numerical coding (see in the data view).





We define which variables are considered as dependent measurements or measurement pairs.

Here: Motivation at t1 and at t2.

Exemplary syntax command

```
T-TEST PAIRS=motiva1 WITH motiva2
(PAIRED)
  /ES DISPLAY(TRUE) STANDARDIZER(SD)
  /CRITERIA=CI(.9500)
  /MISSING=ANALYSIS
```

t-Tests in SPSS 2/2

Levene's test to check the assumption of homogeneity of variance

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
Anti-Sakkaden-Task Fehlerrate in %	Equal variances assumed	,122	,728	-,573	68	,284	,569	-1,94137	3,38972	-8,70545	4,82272
	Equal variances not assumed			-,572	65,818	,285	,569	-1,94137	3,39192	-8,71391	4,83118

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	Upper
Anti-Sakkaden-Task Fehlerrate in %	Cohen's d	14,12804	-,137	-,608	,334
	Hedges' correction	14,28629	-,136	-,601	,330
	Glass's delta	14,18536	-,137	-,607	,336

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.

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	Motivation: Während ich die vorherige Aufgabe bearbeitet habe, war ich motiviert. - Motivation: Während ich die vorherige Aufgabe bearbeitet habe, war ich motiviert.	-,386	1,921	,230	-,844	,072	-1,680	69	,049	,097

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	Upper
Pair 1	Motivation: Während ich die vorherige Aufgabe bearbeitet habe, war ich motiviert. - Motivation: Während ich die vorherige Aufgabe bearbeitet habe, war ich motiviert.	Cohen's d	1,921	-,201	-,437	,037
		Hedges' correction	1,942	-,199	-,432	,036

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

In the context of a t-test for dependent samples, the assumption of homogeneity of variance does not need to be checked. However, interval scaling and normality should still be assessed.

ANOVA

Execution in SPSS

ANOVA against alpha error accumulation

- Starting point: Examine a difference between three groups in a variable
  - Idea: three t-tests *(1 vs. 2, 1 vs. 3, 2 vs. 3*), each with a significance level of 5%
  - Assumption: The three groups do not differ (and none of the three t-tests should yield a significant result).
  - In each t-test, the probability that we (in this case, incorrectly) choose H1 is 5%, and the probability that we retain H0 is 95%.
  - Problem: To retain the "global hypothesis" that the three groups do not differ significantly, none of the three t-tests should yield a significant result.
  - However, the probability of this (since the three t-tests are not independent of each other!) is not 0.95, → instead $0.95^3 = 0.86!$

  → Accordingly, the probability of incorrectly choosing the alternative hypothesis, even though the null hypothesis holds in the population (our alpha error), is not .05, but rather .14!

→ **To avoid this so-called alpha error accumulation, a different analysis method should be used and/or a correction of the alpha error should be applied.**

ANOVA – basic idea

- Generalization of the t-test for comparing more than 2 means
- "**AN**alysis **O**f **VA**riance" because, based on variance (a measure of the variability of values), a judgment is made about a possible effect.
- Analyses of variance are classified, among other ways, by how many and what type of factors are simultaneously examined for their impact on a dependent variable.
  - one-way/multifactorial
  - without/with repeated measures

- **Caution**: The omnibus test ("F-test") in analysis of variance tests exclusively the following:
  - H0: The (group) means in the populations do not differ
  - H1: At least one of the (group) means in the populations differs from another
  - **The test does not indicate which means differ (significantly) from each other!**
    - For this purpose, **a priori planned contrasts** or **post-hoc comparisons** are used – which are more common, but not entirely free from criticism (ideally, hypotheses about mean differences are formed in advance, rather than taking an "exploratory" approach).

ANOVA

Assumptions

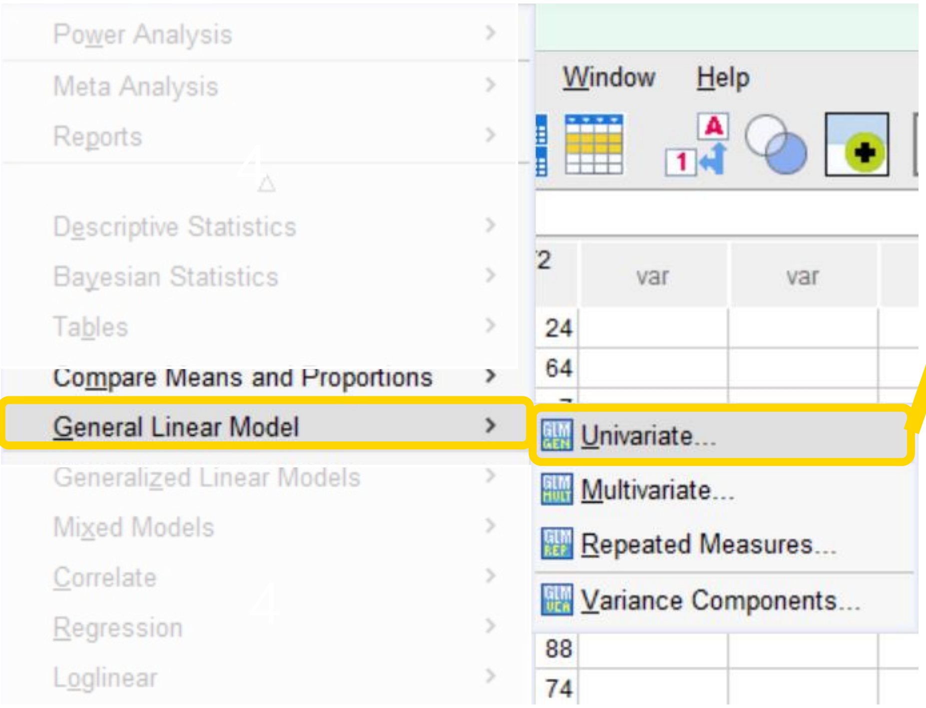
ANOVA: Assumptions - Overview

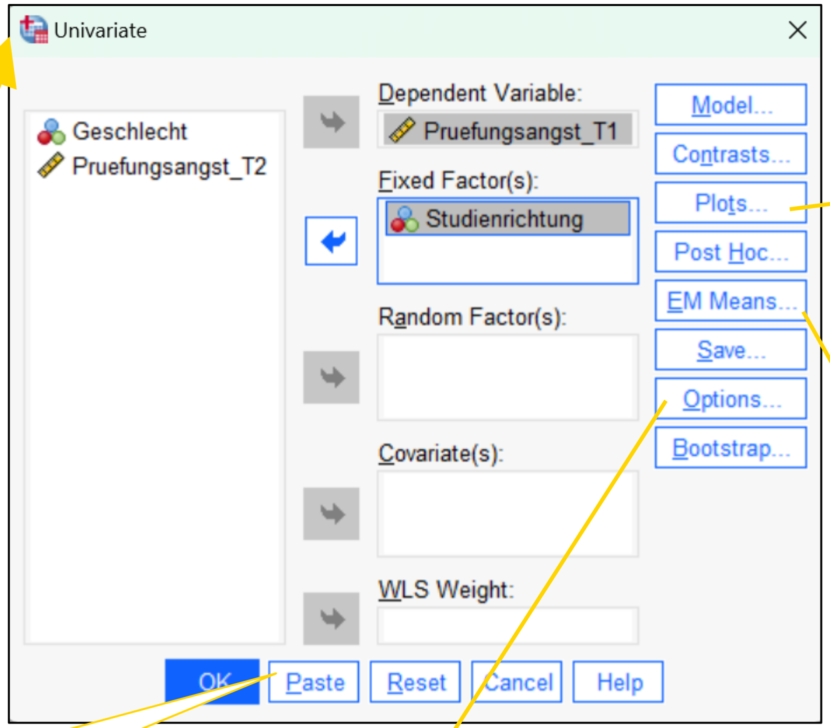
One-way without repeated measures	One-way with repeated measures	two-way without repeated measures	two-way wit repeated measures (at least 1 factor)
DV: interval-scaled			
DV in all IV groups normally distributed → WHEN sample sizes are small.			
IV: nominal-scaled (groups)			
Homogeneity of variance	Sphericity when ≥ 3 levels	Homogeneity of variance for both factors	Sphericity for within-factor when ≥ 3 levels
→ Levene´s-Test	→ Mauchly´s-Test	→ Levene´s-Test	Homogeneity of variance for between-factor
-	-	-	→ Mauchly´s-Test → Box-M Test

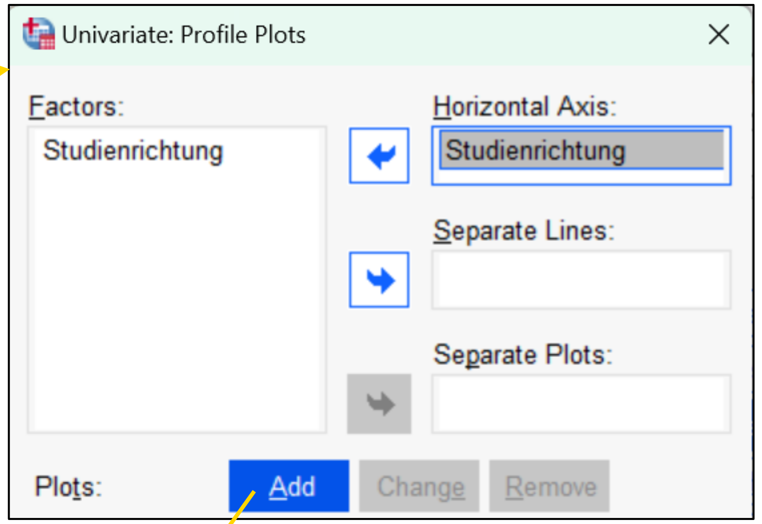
One-way ANOVA (between-subjects factor)

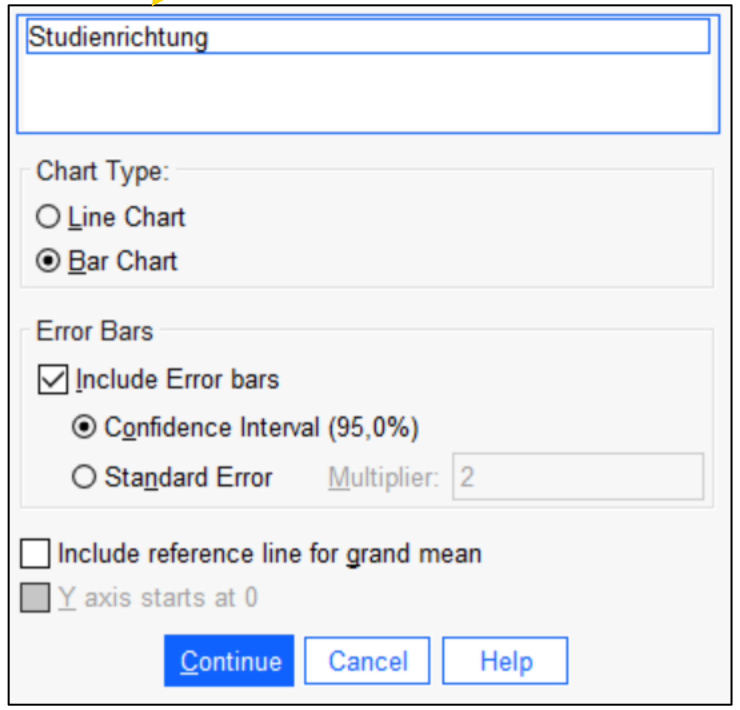
ANOVA – one-way (between-factor)

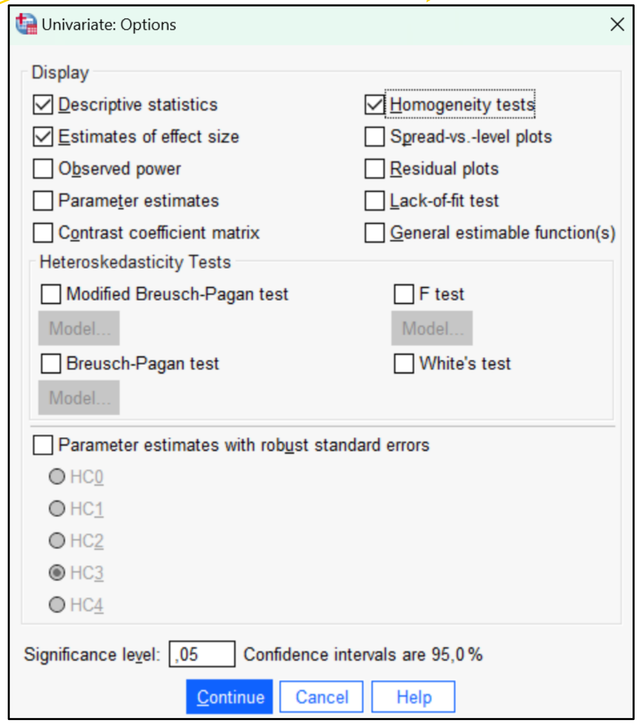
- to examine differences in a variable across more than 2 (independent) groups

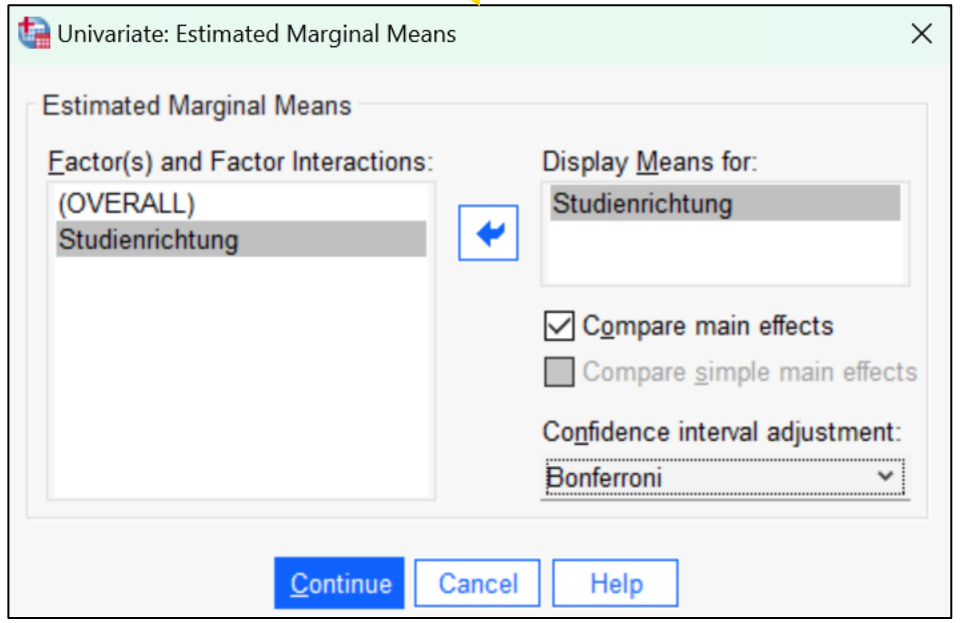












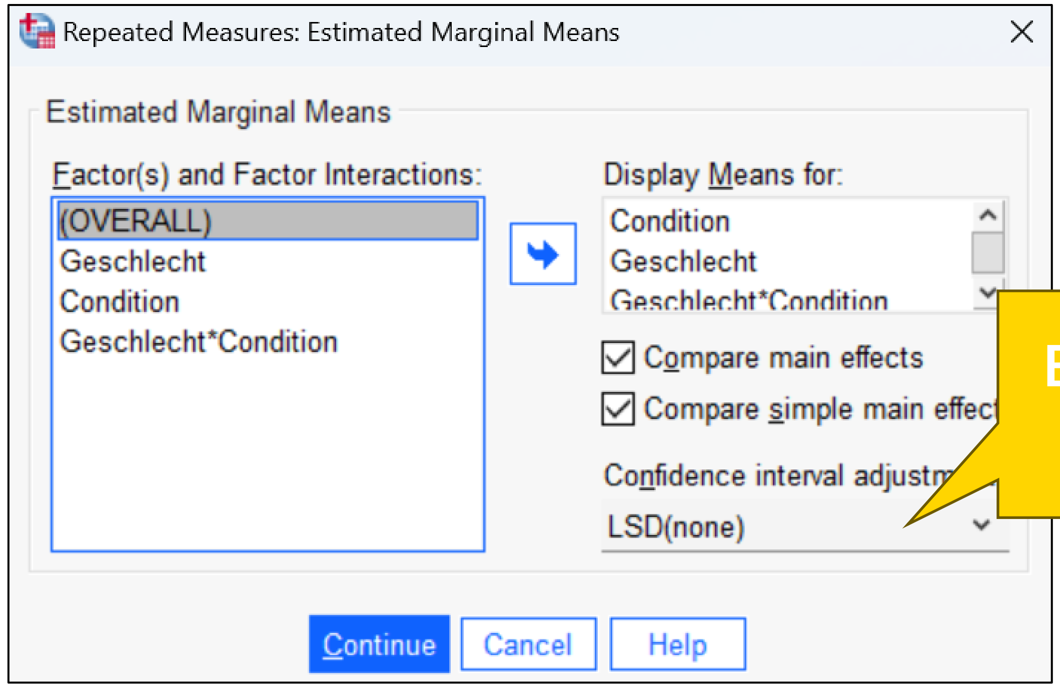
Exemplary syntax command

```
UNIANOVA Pruefungsangst_T1 BY
Studienrichtung
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /PLOT=PROFILE(Studienrichtung) TYPE=BAR
ERRORBAR=CI MEANREFERENCE=NO
  /EMMEANS=TABLES(Studienrichtung) COMPARE
ADJ(BONFERRONI)
  /PRINT ETASQ DESCRIPTIVE HOMOGENEITY
  /CRITERIA=ALPHA(.05)
  /DESIGN=Studienrichtung.
```

Supplement: Estimated Marginal Means in SPSS

- Also referred to as "estimated marginal means" or "least squares means“
- These estimated means represent a form of “standardized means”
- Important for interpreting the effects of categorical predictors within the model, especially **when the sample sizes between the groups vary or when the model includes covariates**
- **Always select for interactions and when an independent variable has more than 2 levels!**

Caution: Post-Hoc-Tests under Post Hoc... **are generally only possible for between-subjects factors. For multifactorial ANOVAs, it is best to conduct these through "EM Means"**



Better to use Bonferroni

ANOVA – one-way (between-factor)

Between-Subjects Factors

		N
Studienrichtung	0	19
	1	13
	2	18

Basic descriptive statistics

Descriptive Statistics

Dependent Variable: Pruefungsangst_T1

Studienrichtung	Mean	Std. Deviation	N
0	32,79	19,156	19
1	68,15	20,330	13
2	42,72	25,474	18
Total	45,56	25,745	50

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

		Levene Statistic	df1	df2	Sig.
Pruefungsangst_T1	Based on Mean	,747	2	47	,479
	Based on Median	,611	2	47	,547
	Based on Median and with adjusted df	,611	2	40,268	,548
	Based on trimmed mean	,740	2	47	,483

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

a. Dependent variable: Pruefungsangst_T1

b. Design: Intercept + Studienrichtung

The Levene's test for homogeneity of variances checks whether the variances in the different groups are equal (= H0).If a significant result is found, it can be "ignored" in the case of a conservative method for correction in pairwise comparisons (e.g., Bonferroni), but it should still be reported!

Global F-/Omnibus test with the null hypothesis that the groups do not differ from each other

Tests of Between-Subjects Effects

Dependent Variable: Pruefungsangst_T1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	9879,859[a]	2	4939,929	10,275	<,001	,304
Intercept	111499,997	1	111499,997	231,917	<,001	,831
Studienrichtung	9879,859	2	4939,929	10,275	<,001	,304
Error	22596,461	47	480,776			
Total	136262,000	50				
Corrected Total	32476,320	49				

a. R Squared = ,304 (Adjusted R Squared =

Report F-test:

basic: F(df_UV, df_Rrror) = F-value, p =/< p-value, η_p^2 = value

Here: $F(2, 47) = 10.28$, $p < .001$, $\eta_p^2 = .30$

ANOVA – one-way (between-factor)

Estimates

Dependent Variable: Pruefungsangst_T1

Studienrichtung	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
0	32,789	5,030	22,670	42,909
1	68,154	6,081	55,920	80,388
2	42,722	5,168	32,325	53,119

Estimated means (of the population) and their standard errors

Pairwise Comparisons

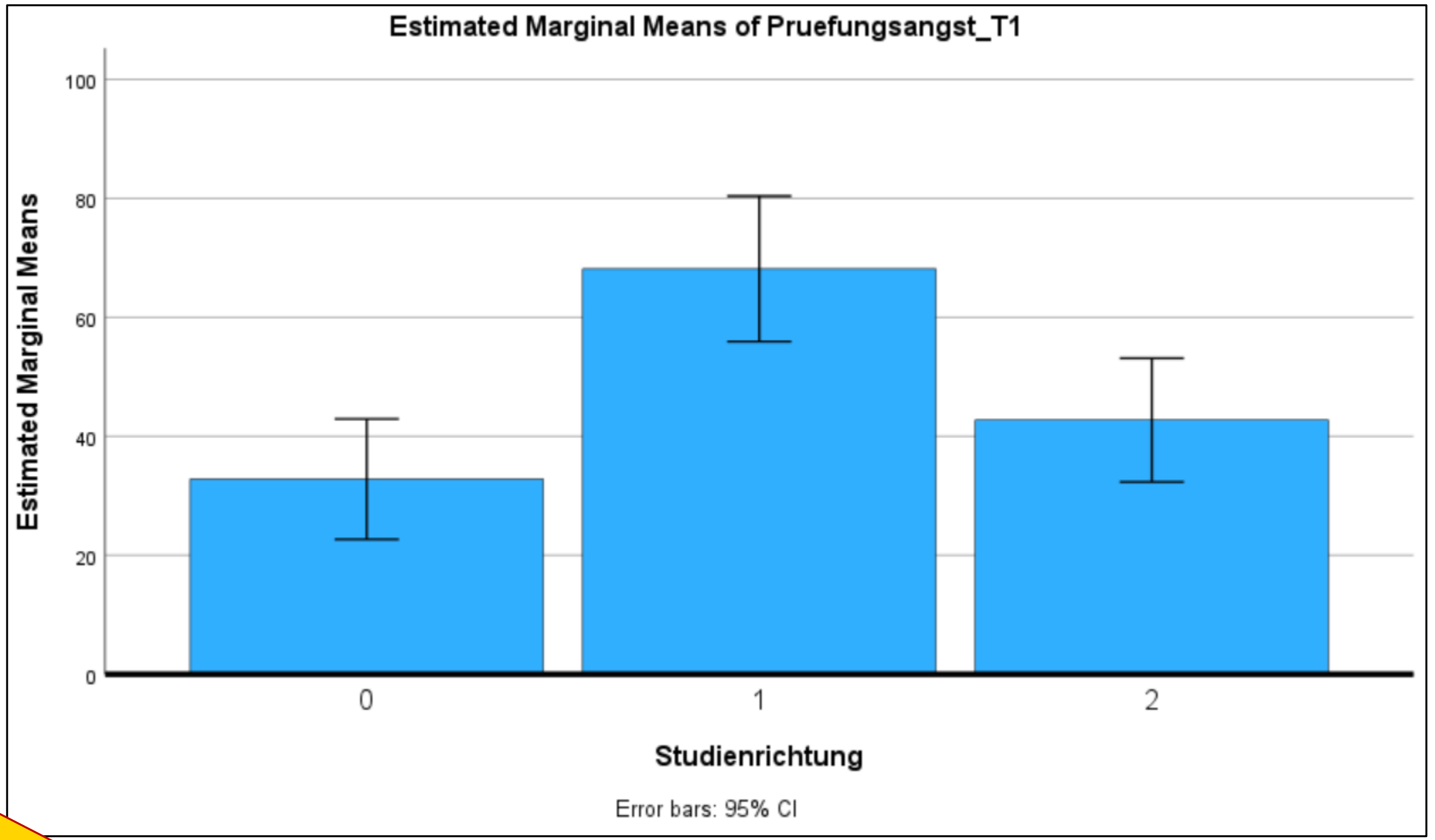
Dependent Variable: Pruefungsangst_T1

(I) Studienrichtung	(J) Studienrichtung	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
0	1	-35,364*	7,892	<,001	-54,958	-15,770
	2	-9,933	7,212	,525	-27,838	7,973
1	0	35,364*	7,892	<,001	15,770	54,958
	2	25,432*	7,981	,008	5,618	45,245
2	0	9,933	7,212	,525	-7,973	27,838
	1	-25,432*	7,981	,008	-45,245	-5,618

Based on estimated marginal means

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

b. Adjustment for multiple comparisons: Bonferroni.



Pairwise comparisons with a previously defined correction method do not provide an effect size measure! If needed, conduct separate t-tests.

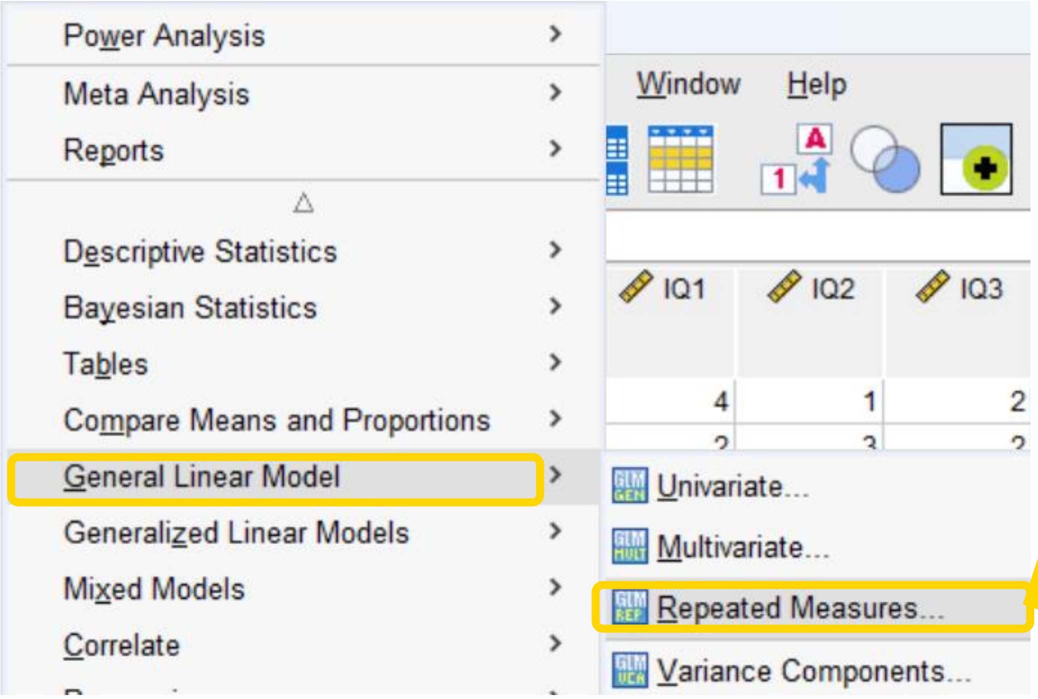
Report pairwise-comparison
basic: Mean_Diff = value, 95% CI [Lower, Upper], *p* =/< p-value

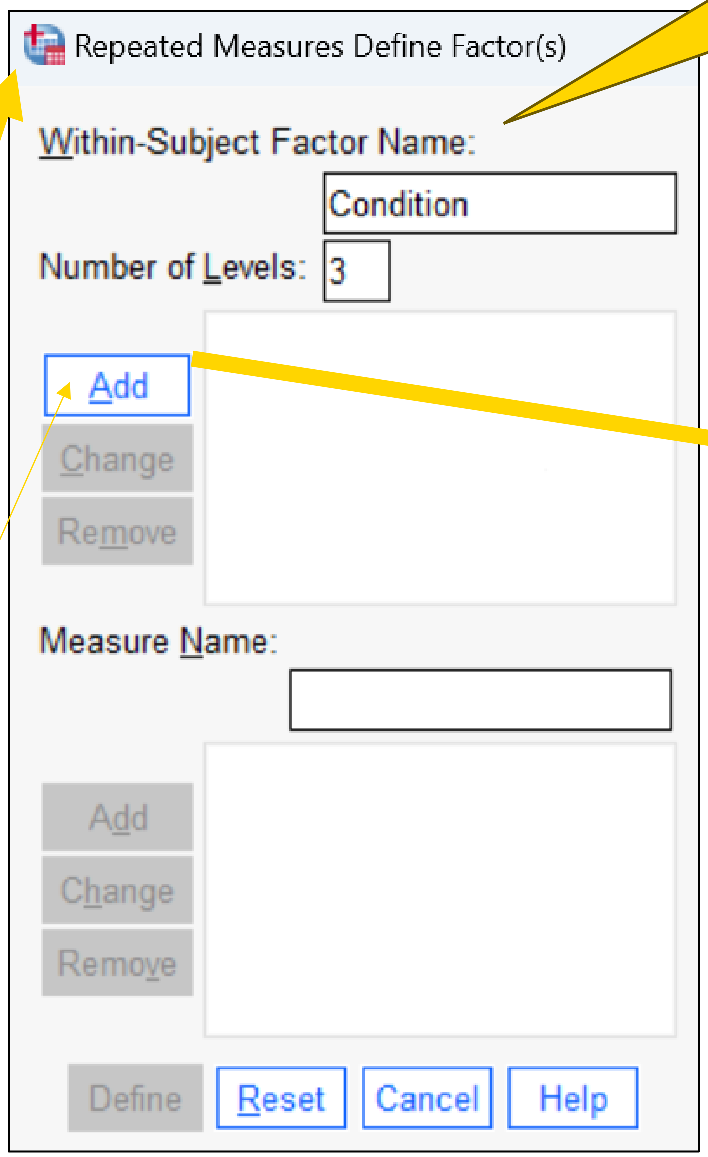
Here: Mean_Diff0/1 = -35.36, 95% CI [-54.96, -15.77], *p* < .001

one-way ANOVA (within-subjects factor)

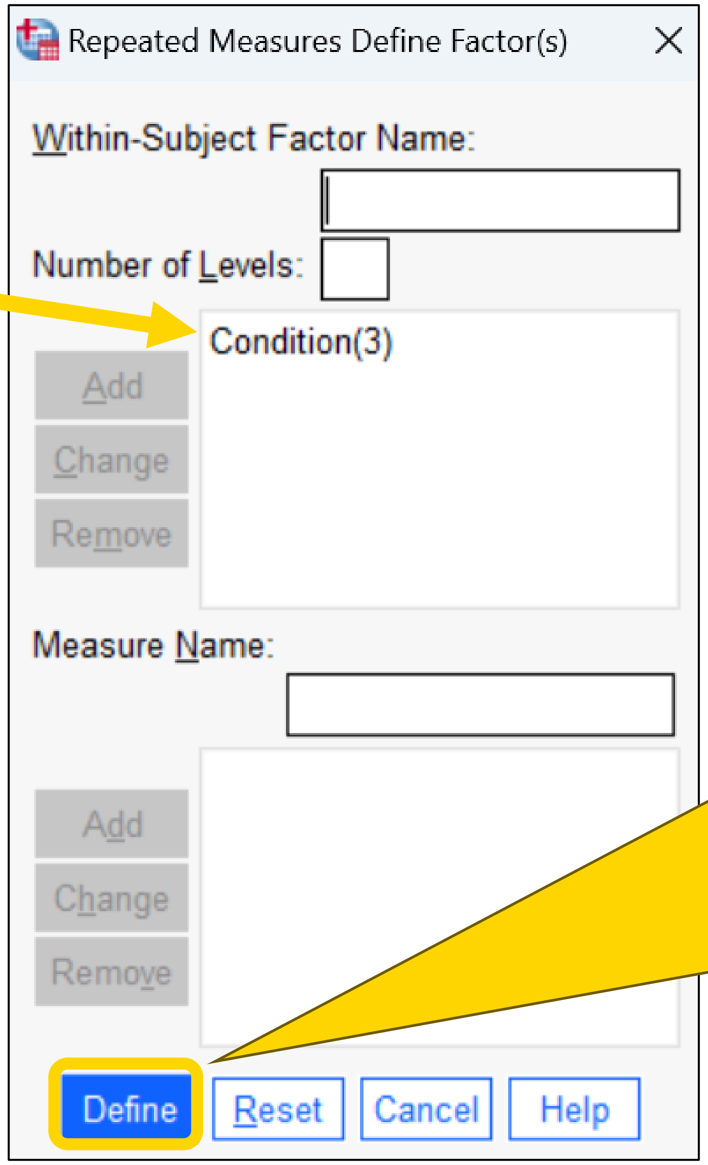
ANOVA – one-way (within-Faktor)

- First, we define a within-subjects factor Analysis





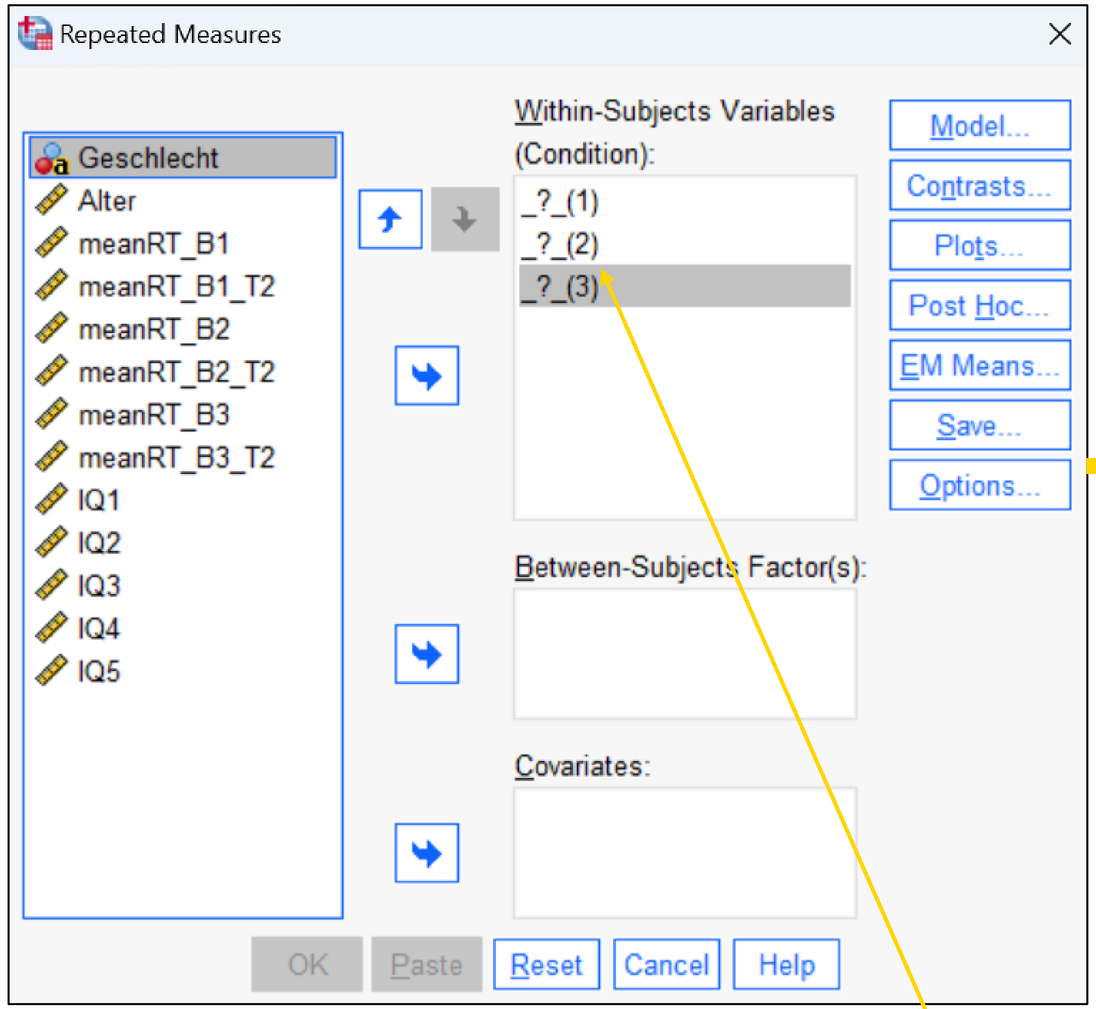
Define the name and number of levels of the within-subjects factor (e.g., number of measurement points)



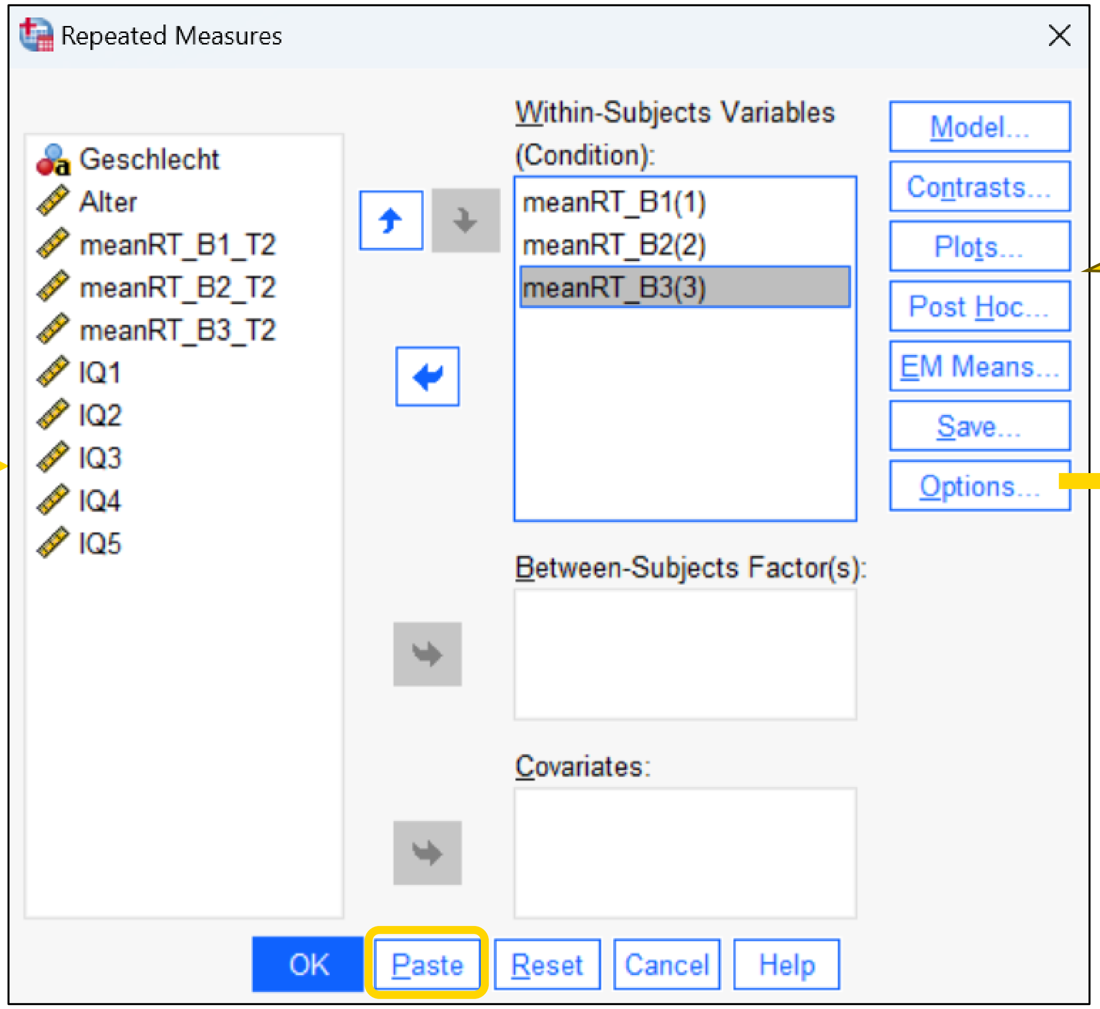
"Add" adds the factor to the list (allows the definition of multiple within-subjects factors)

In the next step, the defined variables are assigned to the levels of the repeated measures factor

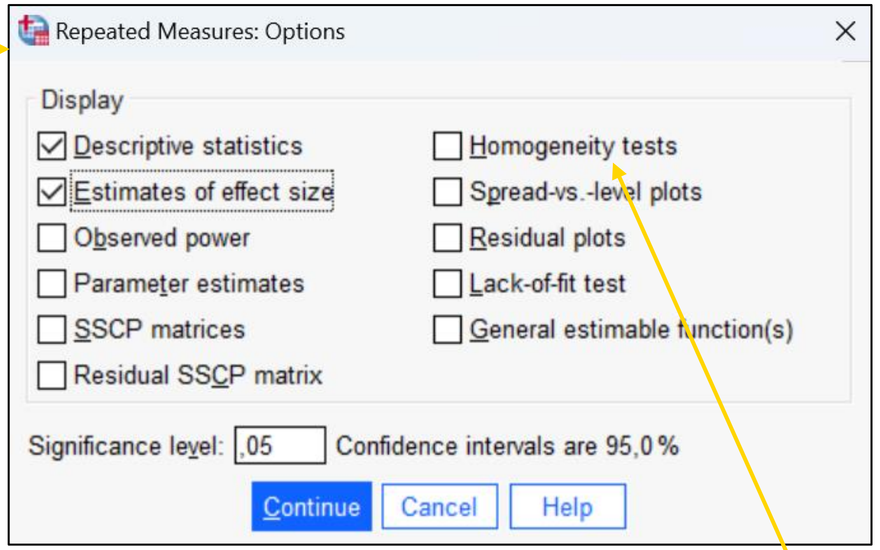
ANOVA – one-way (within-Faktor)



Define the levels of the factor using drag & drop or the "arrow button"



Plots and EM means are equivalent to a one-way ANOVA with a between-subjects factor



Exemplary syntax command

```
GLM meanRT_B1 meanRT_B2 meanRT_B3
  /WSFACTOR=Condition 3 Polynomial
  /METHOD=SSTYPE(3)
  /PLOT=PROFILE(Condition) TYPE=BAR ERRORBAR=CI MEANREFERENCE=NO
  /EMMEANS=TABLES(Condition) COMPARE ADJ(LSD)
  /PRINT=DESCRIPTIVE ETASQ
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=Condition.
```

Homogeneity tests are only necessary for between-subjects factors

ANOVA – one-way (within-Faktor)

Within-Subjects Factors

Measure: MEASURE_1

Condition	Dependent Variable
1	meanRT_B1
2	meanRT_B2
3	meanRT_B3

Information about factor levels and descriptive statistics

Descriptive Statistics

	Mean	Std. Deviation	N
meanRT_B1	702,62	117,955	150
meanRT_B2	682,11	99,675	150
meanRT_B3	679,54	102,824	150

Mauchly's Test of Sphericity (for within-subjects factors with > 2 levels): If the result is not significant, the within-subjects effect can be read under "Sphericity Assumed“
If the result is significant, a correction of the degrees of freedom should be applied, and the corresponding row for the within-subjects effect should be used (→ next slide!)

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

					Epsilon[b]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Condition	,985	2,211	2	,331	,985	,998	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Condition

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Condition	Pillai's Trace	,026	1,955[b]	2,000	148,000	,145	,026
	Wilks' Lambda	,974	1,955[b]	2,000	148,000	,145	,026
	Hotelling's Trace	,026	1,955[b]	2,000	148,000	,145	,026
	Roy's Largest Root	,026	1,955[b]	2,000	148,000	,145	,026

a. Design: Intercept
Within Subjects Design: Condition

b. Exact statistic

The table of multivariate tests is actually only relevant in MANOVA (with more than one dependent variable)

ANOVA – one-way (within-Faktor)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Condition	Sphericity Assumed	48003,551	2	24001,776	2,207	,112	,015
	Greenhouse-Geisser	48003,551	1,971	24357,716	2,207	,113	,015
	Huynh-Feldt	48003,551	1,997	24037,937	2,207	,112	,015
	Lower-bound	48003,551	1,000	48003,551	2,207	,140	,015
Error(Condition)	Sphericity Assumed	3240873,116	298	10875,413			
	Greenhouse-Geisser	3240873,116	293,645	11036,693			
	Huynh-Feldt	3240873,116	297,552	10891,798			
	Lower-bound	3240873,116	149,000	21750,826			

Reporting the F-test is equivalent to a one-way ANOVA with a between-subjects factor,except when degrees of freedom are corrected due to a violation of the sphericity assumption. In this case, the correction method should be reported in the text!

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Condition	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Condition	Linear	39951,480	1	39951,480	3,384	,068	,022
	Quadratic	8052,071	1	8052,071	,810	,370	,005
Error(Condition)	Linear	1758834,520	149	11804,259			
	Quadratic	1482038,596	149	9946,568			

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	213059843,56	1	213059843,56	16815,414	<,001	,991
Error	1887905,778	149	12670,509			

The contrast table (unless a contrast analysis was planned in advance) and the table of between-subjects effects can be ignored

ANOVA – one-way (within-Faktor)

Estimated Marginal Means

Condition

Estimated means (of the population) and their standard errors

Estimates

Measure: MEASURE_1

Condition	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	702,620	9,631	683,589	721,651
2	682,107	8,138	666,025	698,188
3	679,540	8,396	662,950	696,130

Multivariate tests are not relevant in our case, even for pairwise comparisons

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	,026	1,955[a]	2,000	148,000	,145	,026
Wilks' lambda	,974	1,955[a]	2,000	148,000	,145	,026
Hotelling's trace	,026	1,955[a]	2,000	148,000	,145	,026
Roy's largest root	,026	1,955[a]	2,000	148,000	,145	,026

Each F tests the multivariate effect of Condition. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Pairwise Comparisons

Measure: MEASURE_1

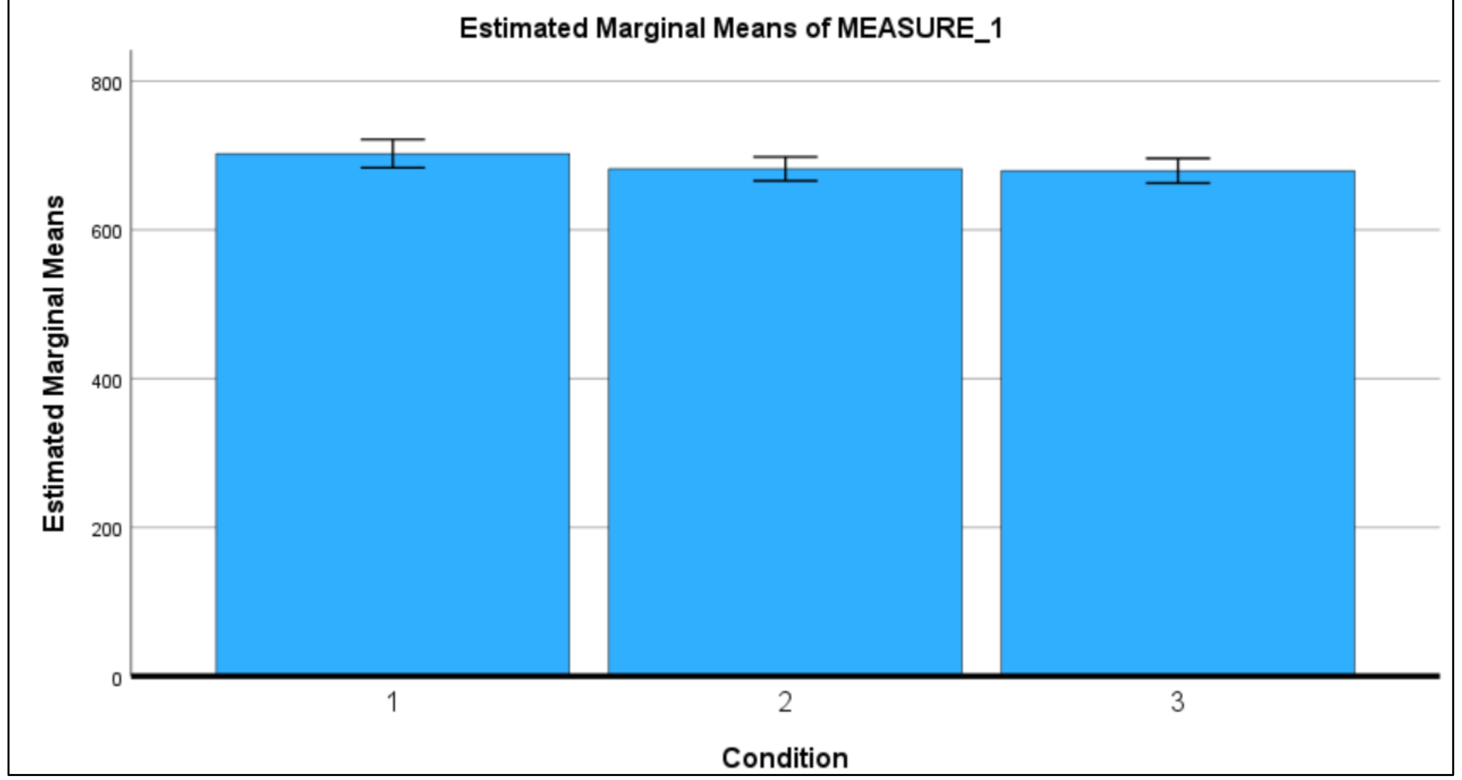
(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	95% Confidence Interval for Difference[a] Upper Bound
1	2	20,513	12,236	,096	-3,665	44,691
	3	23,080	12,546	,068	-1,710	47,870
2	1	-20,513	12,236	,096	-44,691	3,665
	3	2,567	11,310	,821	-19,782	24,915
3	1	-23,080	12,546	,068	-47,870	1,710
	2	-2,567	11,310	,821	-24,915	19,782

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Pairwise comparisons with the previously defined correction procedure are reported equivalently to pairwise comparisons with a between-subjects factor

Profile Plots



One-way ANOVA reporting results

One-way ANOVA – reporting results

- Relevant descriptive statistics (either mentioned in the text or referenced in an **APA-compliant** table)
  - Factor + levels
  - *N, M, SD* of the groups
  - Estimated population means and standard errors for (used) pairwise comparisons
- F-/Omnibus test: *F*(df_UV, df_Error) = F-value, *p* =/< p-value, η_p^2 = value
- Mean difference, standard error, and p-value for all pairwise comparisons
  - Significant pairwise comparisons should be discussed in the text. Non-significant pairwise comparisons can be presented in an **APA-compliant table**, with a reference to the table in the text
- In the case of a significant Mauchly's Test of Sphericity in a one-way ANOVA with a within-subjects factor, the correction method used (e.g., Greenhouse-Geisser) must also be reported.

One-way ANOVA – reporting results

- **Always** prioritize the substantive conclusion, with statistical values serving only to support the statement!

- **Don‘t:** *"The three groups differ significantly from each other.“*

- **Instead e.g.:** „A one-way repeated measures ANOVA was conducted with the repeated measures factor [Factor Name]. The descriptive statistics can be found in Table 1. The results indicated that students from the fields of psychology, history, and business administration significantly differed in perceived exam anxiety (report F-test correctly). Pairwise comparisons further revealed that psychology students had significantly higher levels of exam anxiety than business administration students (report pairwise comparison). No significant differences were found otherwise (see Table X; reference to table with non-significant pairwise comparisons)”

Supplement: Effect sizes η^2 and η^2_p

→ Both are measures of effect size within the context of analysis of variance

η^2 : Proportion of the total variance explained by the factor

- *How much of the variance in the dependent variable can be attributed to the individual factor?*
- Contra: Comparison of effect size with other analyses is not possible, as the effect size depends on which other effects are being tested in the analysis.

$$\hat{\eta}^2 = \frac{QS_{Factor}}{QS_{Total}}$$

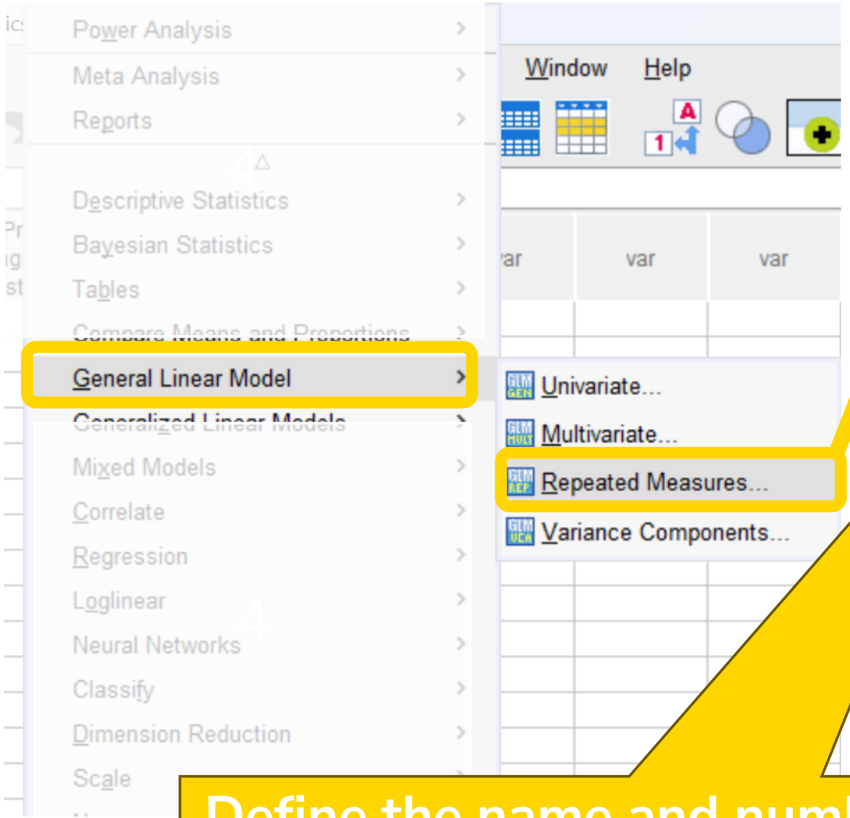
- **η^2_p** : All components except the variance of interest and the error variance are removed from the total variance (in the denominator)
  - *What is the proportion of the dependent variable's variance attributed to the <u>individual factor and the error</u> that can be explained by each respective factor?*
  - Pro: Comparable across different studies
  - (small) Contra: Partial effect size measures within a study cannot be summed to a total effect ⟶ Representation of the relationships between the effects within the analysis is not possible

$$\hat{\eta}^2_p = \frac{QS_{Factor}}{QS_{Factor} + QS_{Error}}$$

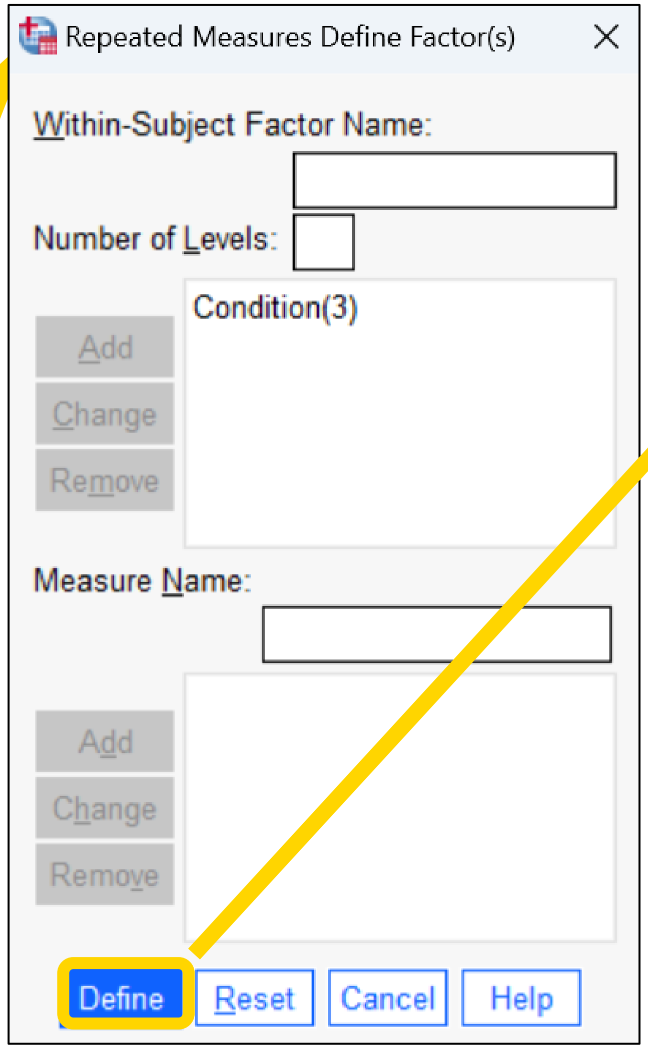
Both effect sizes OVERSTATE the respective population effects due to the use of sample variances, which is why Eid et al. (2017) recommend calculating ω^2. However, this is (currently) not integrated into SPSS.

Mixed-ANOVA (Between + Within)

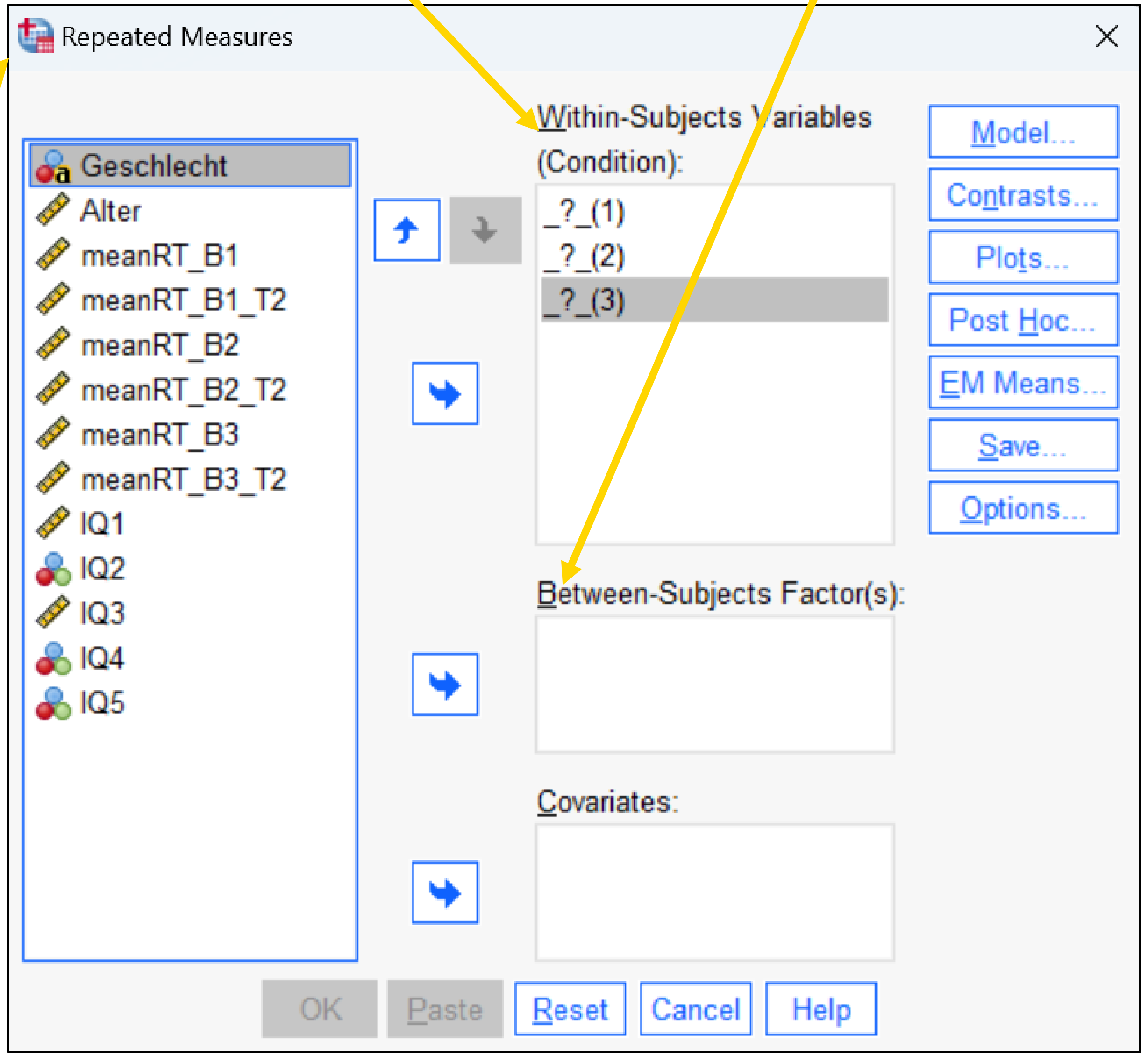
- First, the within-subject factor is defined: Analyze



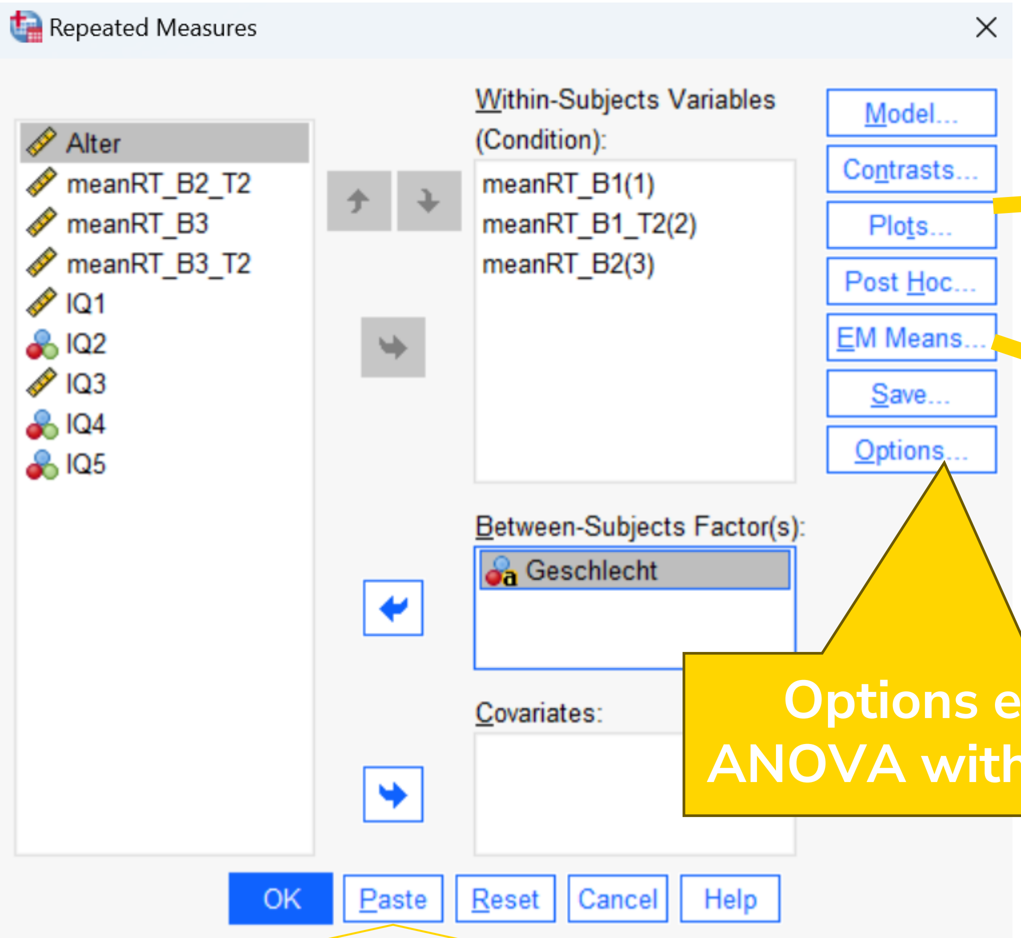
Define the name and number of levels, and add them using "Add" (equivalent to the one-way ANOVA with a within-subject factor)



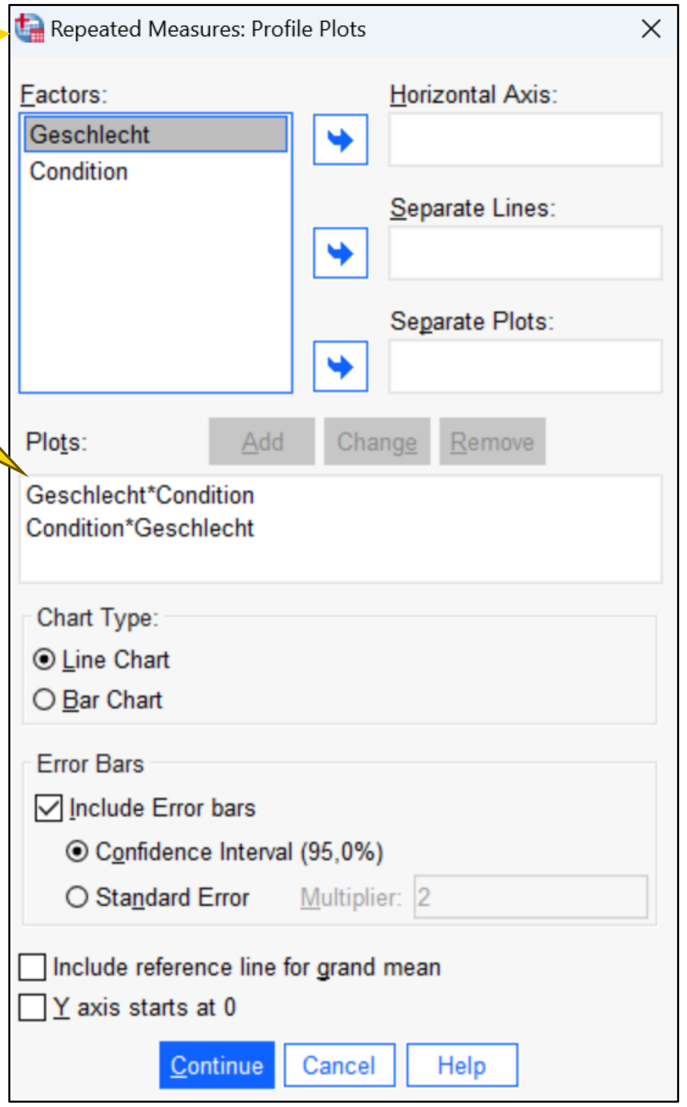
Assignment of variables to the levels is equivalent to a one-way ANOVA with a within-subject factorAssignment of the between-subject factor is equivalent to a one-way ANOVA with a between-subject factor



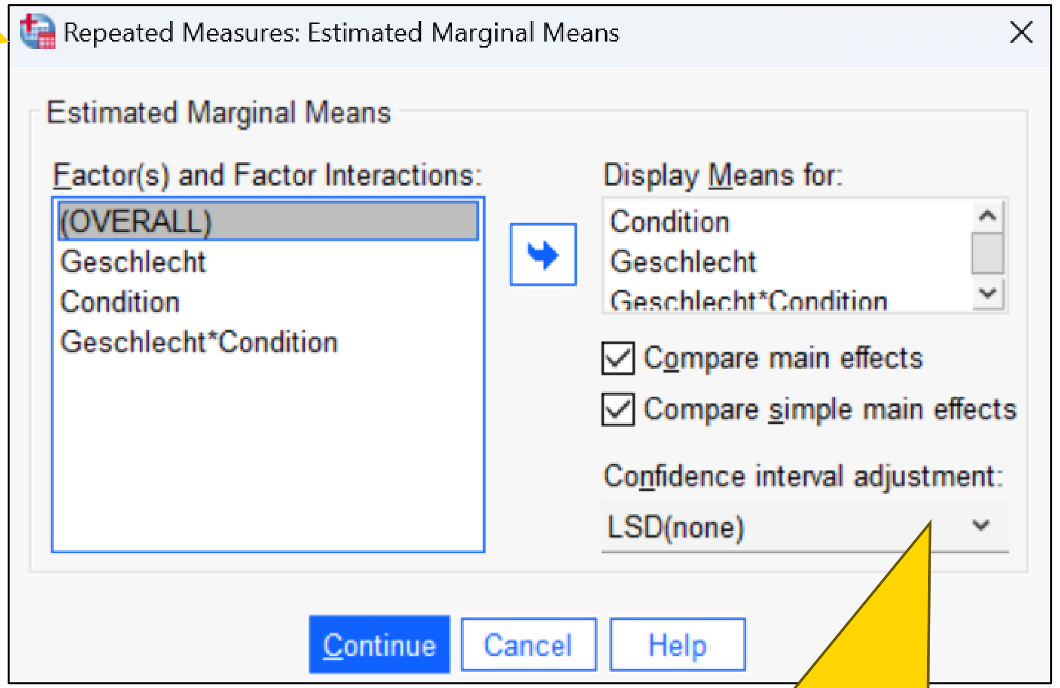
Mixed-ANOVA



Options equivalent to the ANOVA with a Between-Factor



Generate plots for both AxB and BxA!



Request pairwise comparisons for both main effects and the interaction (+ correction method – Bonferroni!)

Exemplary syntax command

```
GLM meanRT_B1 meanRT_B2 meanRT_B3 BY Geschlecht
  /WSFACTOR=Condition 3 Polynomial
  /METHOD=SSTYPE(3)
  /PLOT=PROFILE(Geschlecht*Condition Condition*Geschlecht) TYPE=LINE
ERRORBAR=CI MEANREFERENCE=NO
    YAXIS=AUTO
  /EMMEANS=TABLES(Condition) COMPARE ADJ(LSD)
  /EMMEANS=TABLES(Geschlecht) COMPARE ADJ(LSD)
  /EMMEANS=TABLES(Geschlecht*Condition) COMPARE(Geschlecht) ADJ(LSD)
  /EMMEANS=TABLES(Geschlecht*Condition) COMPARE(Condition) ADJ(LSD)
  /PRINT=DESCRIPTIVE ETASQ HOMOGENEITY
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=Condition
  /DESIGN=Geschlecht.
```

Mixed-ANOVA

Multivariate tests can also be ignored here

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Condition	Pillai's Trace	,035	2,697[b]	2,000	147,000	,071	,035
	Wilks' Lambda	,965	2,697[b]	2,000	147,000	,071	,035
	Hotelling's Trace	,037	2,697[b]	2,000	147,000	,071	,035
	Roy's Largest Root	,037	2,697[b]	2,000	147,000	,071	,035
Condition * Geschlecht	Pillai's Trace	,065	5,088[b]	2,000	147,000	,007	,065
	Wilks' Lambda	,935	5,088[b]	2,000	147,000	,007	,065
	Hotelling's Trace	,069	5,088[b]	2,000	147,000	,007	,065
	Roy's Largest Root	,069	5,088[b]	2,000	147,000	,007	,065

a. Design: Intercept + Geschlecht
Within Subjects Design: Condition

b. Exact statistic

Within-Subjects Factors

Measure: MEASURE_1

Condition	Dependent Variable
1	meanRT_B1
2	meanRT_B2
3	meanRT_B3

Description of the respective factor levels and descriptive statistics

Descriptive Statistics

	Geschlecht	Mean	Std. Deviation	N
meanRT_B1	m	745,97	104,300	68
	w	666,67	117,089	82
	Total	702,62	117,955	150
meanRT_B2	m	687,31	109,730	68
	w	677,79	90,964	82
	Total	682,11	99,675	150
meanRT_B3	m	685,97	117,415	68
	w	674,21	89,340	82
	Total	679,54	102,824	150

Between-Subjects Factors

		N
Geschlecht	m	68
	w	82

The Box-M test checks for the equality of covariance matrices. A commonly used approach interprets the ANOVA as "normal" if the individual groups are approximately the same size, but the Box-M test is still reported, for example: Box M: 14.16, F(6, 144622.370) = 2.307, p = .031. → When in doubt, discuss with the supervising person.
More information

Box's Test of Equality of Covariance Matrices[a]

Box's M	14,158
F	2,307
df1	6
df2	144622,370
Sig.	,031

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Geschlecht
Within Subjects Design: Condition

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Condition	,992	1,234	2	,540	,992	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + Geschlecht
Within Subjects Design: Condition

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Mauchly's Test of Sphericity is equivalent to a one-way ANOVA with a within-subjects factor. A significant result, however, necessitates a correction for both the main effect and the interaction!

Mixed-ANOVA

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Condition	Sphericity Assumed	62410,131	2	31205,065	2,957	,054	,020
	Greenhouse-Geisser	62410,131	1,983	31465,826	2,957	,054	,020
	Huynh-Feldt	62410,131	2,000	31205,065	2,957	,054	,020
	Lower-bound	62410,131	1,000	62410,131	2,957	,088	,020
Condition * Geschlecht	Sphericity Assumed	116922,611	2	58461,305	5,539	,004	,036
	Greenhouse-Geisser	116922,611	1,983	58949,829	5,539	,004	,036
	Huynh-Feldt	116922,611	2,000	58461,305	5,539	,004	,036
	Lower-bound	116922,611	1,000	116922,611	5,539	,020	,036
Error(Condition)	Sphericity Assumed	3123950,505	296	10553,887			
	Greenhouse-Geisser	3123950,505	293,547	10642,079			
	Huynh-Feldt	3123950,505	296,000	10553,887			
	Lower-bound	3123950,505	148,000	21107,774			

Reporting the F-test for the main effect is equivalent to a one-way ANOVA with a Between/Within factor. The same applies to the interaction (here, the error degrees of freedom are the same as for the Within main effect!). In case of a violation of the assumption of sphericity, the correction procedure should be reported!

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Condition	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Condition	Linear	51158,125	1	51158,125	4,523	,035	,030
	Quadratic	11252,006	1	11252,006	1,149	,286	,008
Condition * Geschlecht	Linear	84777,325	1	84777,325	7,495	,007	,048
	Quadratic	32145,286	1	32145,286	3,281	,072	,022
Error(Condition)	Linear	1674057,195	148	11311,197			
	Quadratic	1449893,310	148	9796,576			

The table of contrasts (unless a contrast analysis was planned in advance) can be ignored.

Mixed-ANOVA

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

		Levene Statistic	df1	df2	Sig.
meanRT_B1	Based on Mean	,805	1	148	,371
	Based on Median	,771	1	148	,381
	Based on Median and with adjusted df	,771	1	145,809	,381
	Based on trimmed mean	,803	1	148	,372
meanRT_B2	Based on Mean	1,814	1	148	,180
	Based on Median	1,925	1	148	,167
	Based on Median and with adjusted df	1,925	1	144,190	,167
	Based on trimmed mean	1,833	1	148	,178
meanRT_B3	Based on Mean	6,345	1	148	,013
	Based on Median	6,372	1	148	,013
	Based on Median and with adjusted df	6,372	1	143,189	,013
	Based on trimmed mean	6,384	1	148	,013

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

a. Design: Intercept + Geschlecht
Within Subjects Design: Condition

In essence the Box-M Test split in three Levene Tests, *see Box-M Test*

Report the F-test for the main effect, equivalent to a one-way ANOVA with a Between/Within factor

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	212165411,97	1	212165411,97	17815,320	<,001	,992
Geschlecht	125350,825	1	125350,825	10,526	,001	,066
Error	1762554,953	148	11909,155			

Mixed-ANOVA - Estimated Marginal Means

Multivariate tests can also be ignored here

1. Condition

Estimates

Measure: MEASURE_1

Condition	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
1	706,321	9,142	688,254	724,387
2	682,551	8,192	666,362	698,740
3	680,089	8,447	663,397	696,781

Pairwise Comparisons

Measure: MEASURE_1

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	23,770*	11,993	,049	,071	47,469
	3	26,232*	12,335	,035	1,857	50,606
2	1	-23,770*	11,993	,049	-47,469	-,071
	3	2,462	11,397	,829	-20,061	24,984
3	1	-26,232*	12,335	,035	-50,606	-1,857
	2	-2,462	11,397	,829	-24,984	20,061

Based on estimated marginal means

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

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	,035	2,697[a]	2,000	147,000	,071	,035
Wilks' lambda	,965	2,697[a]	2,000	147,000	,071	,035
Hotelling's trace	,037	2,697[a]	2,000	147,000	,071	,035
Roy's largest root	,037	2,697[a]	2,000	147,000	,071	,035

Each F tests the multivariate effect of Condition. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Pairwise comparisons of the individual factors are equivalent to the corresponding one-way ANOVA

Mixed-ANOVA - Estimated Marginal Means

2. Geschlecht

Estimates

Measure: MEASURE_1

Geschlecht	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
m	706,417	7,641	691,318	721,515
w	672,890	6,958	659,141	686,640

Pairwise Comparisons

Measure: MEASURE_1

(I) Geschlecht	(J) Geschlecht	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
m	w	33,526*	10,334	,001	13,105	53,947
w	m	-33,526*	10,334	,001	-53,947	-13,105

Based on estimated marginal means

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

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Univariate Tests

Measure: MEASURE_1

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	41783,608	1	41783,608	10,526	,001	,066
Error	587518,318	148	3969,718			

The F tests the effect of Geschlecht. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

Pairwise comparisons of the individual factors are equivalent to the corresponding one-way ANOVA

Mixed-ANOVA - Estimated Marginal Means

3. Geschlecht * Condition

Estimates

Measure: MEASURE_1

Geschlecht	Condition	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
m	1	745,971	13,519	719,255	772,686
	2	687,309	12,114	663,369	711,248
	3	685,971	12,491	661,287	710,654
w	1	666,671	12,311	642,343	690,999
	2	677,793	11,032	655,992	699,593
	3	674,207	11,375	651,730	696,685

Univariate Tests

Measure: MEASURE_1

Condition		Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
1	Contrast	233763,289	1	233763,289	18,809	<,001	,113
	Error	1839344,051	148	12428,000			
2	Contrast	3366,303	1	3366,303	,337	,562	,002
	Error	1476963,990	148	9979,486			
3	Contrast	5143,843	1	5143,843	,485	,487	,003
	Error	1570197,417	148	10609,442			

Each F tests the simple effects of Geschlecht within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

Pairwise Comparisons

Measure: MEASURE_1

Condition	(I) Geschlecht	(J) Geschlecht	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
1	m	w	79,300*	18,285	<,001	43,167	115,432
	w	m	-79,300*	18,285	<,001	-115,432	-43,167
2	m	w	9,516	16,385	,562	-22,862	41,894
	w	m	-9,516	16,385	,562	-41,894	22,862
3	m	w	11,763	16,894	,487	-21,621	45,148
	w	m	-11,763	16,894	,487	-45,148	21,621

Based on estimated marginal means

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

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Pairwise comparisons of the interaction are the "actually relevant/interesting" ones! Are there differences between the levels of the other factor (columns 2+3) within a specific level of one factor (leftmost column)?It is very important to always consider both directions!

Mixed-ANOVA – Estimated Marginal Means

4. Geschlecht * Condition

Estimates

Measure: MEASURE_1

Geschlecht	Condition	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
m	1	745,971	13,519	719,255	772,686
	2	687,309	12,114	663,369	711,248
	3	685,971	12,491	661,287	710,654
w	1	666,671	12,311	642,343	690,999
	2	677,793	11,032	655,992	699,593
	3	674,207	11,375	651,730	696,685

Multivariate Tests

Geschlecht		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
m	Pillai's trace	,086	6,922[a]	2,000	147,000	,001	,086
	Wilks' lambda	,914	6,922[a]	2,000	147,000	,001	,086
	Hotelling's trace	,094	6,922[a]	2,000	147,000	,001	,086
	Roy's largest root	,094	6,922[a]	2,000	147,000	,001	,086
w	Pillai's trace	,003	,239[a]	2,000	147,000	,788	,003
	Wilks' lambda	,997	,239[a]	2,000	147,000	,788	,003
	Hotelling's trace	,003	,239[a]	2,000	147,000	,788	,003
	Roy's largest root	,003	,239[a]	2,000	147,000	,788	,003

Each F tests the multivariate simple effects of Condition within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Pairwise Comparisons

Measure: MEASURE_1

Geschlecht	(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
m	1	2	58,662*	17,734	,001	23,617	93,706
		3	60,000*	18,240	,001	23,956	96,044
	2	1	-58,662*	17,734	,001	-93,706	-23,617
		3	1,338	16,854	,937	-31,967	34,643
	3	1	-60,000*	18,240	,001	-96,044	-23,956
		2	-1,338	16,854	,937	-34,643	31,967
w	1	2	-11,122	16,149	,492	-43,035	20,791
		3	-7,537	16,610	,651	-40,359	25,286
	2	1	11,122	16,149	,492	-20,791	43,035
		3	3,585	15,348	,816	-26,744	33,914
	3	1	7,537	16,610	,651	-25,286	40,359
		2	-3,585	15,348	,816	-33,914	26,744

Based on estimated marginal means

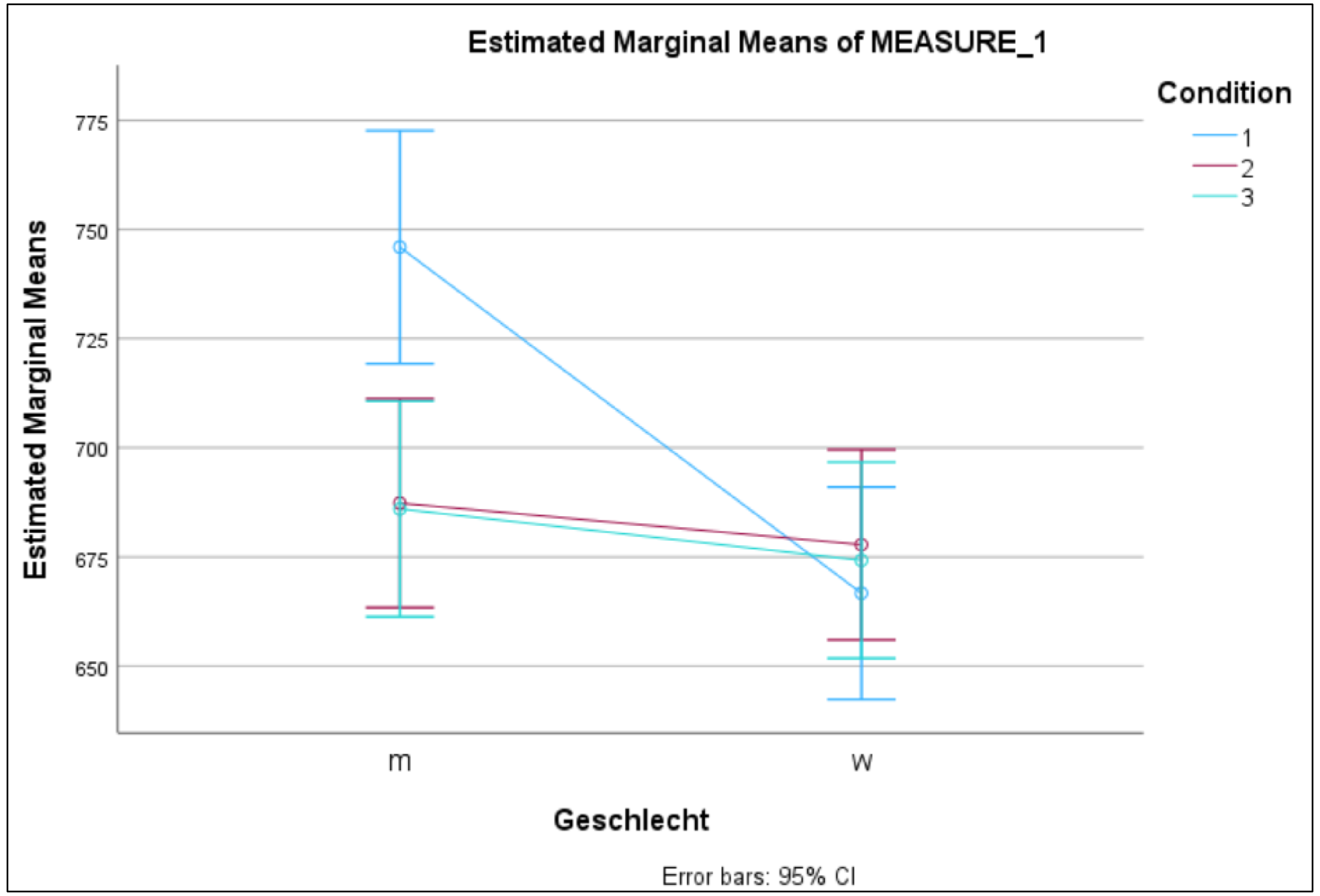
*. The mean difference is significant at the ,05 level.

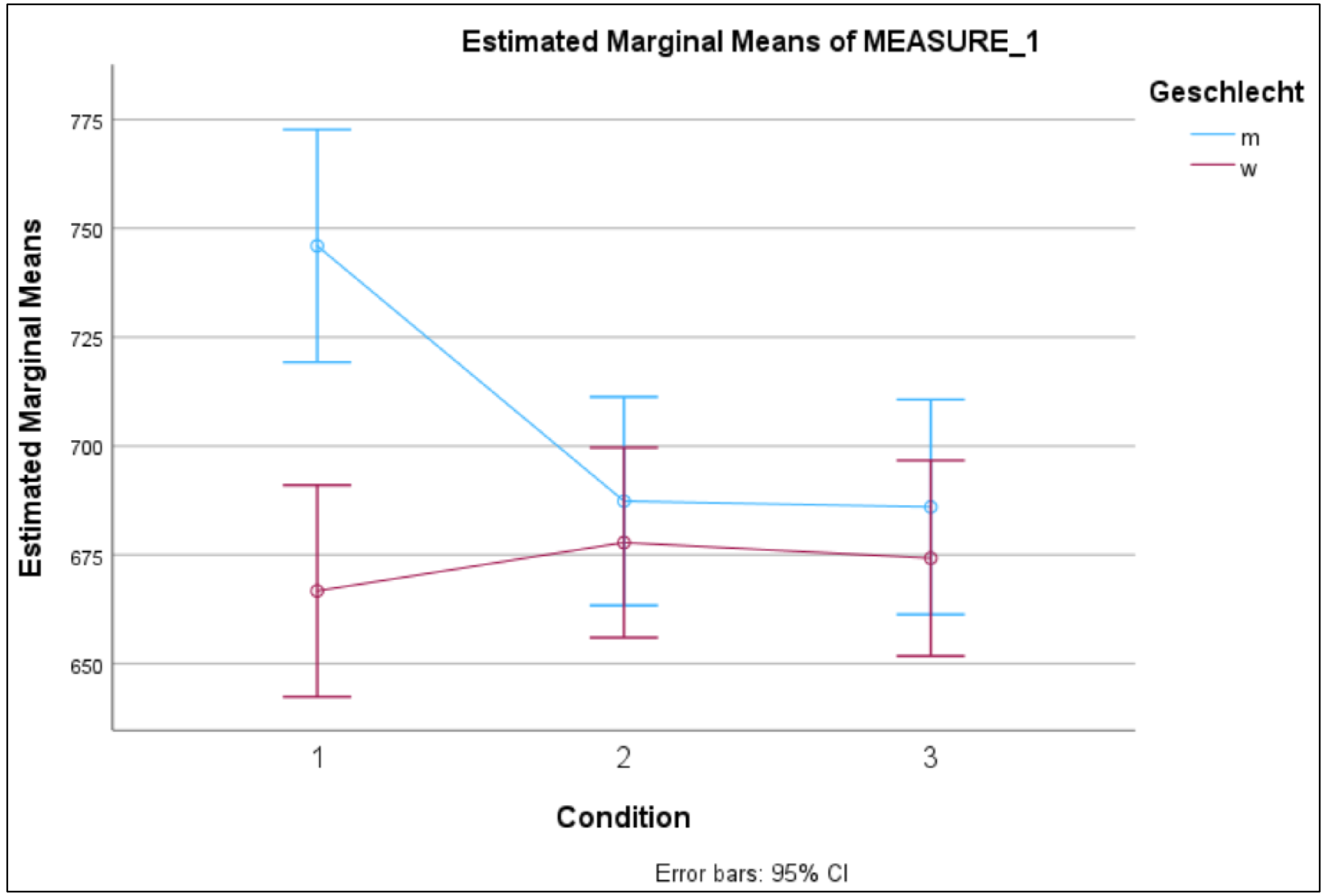
b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Pairwise comparisons of the interaction are the "actually relevant/interesting" ones! Are there differences between the levels of the other factor (columns 2+3) within a specific level of one factor (leftmost column)?It is very important to always consider both directions!

Mixed-Anova

- Always consider both diagrams → Ordinal, hybrid, or disordinal interaction?





Types of interactions

- **Ordinal**: Both main effects are globally interpretable in addition to the interaction
  - The lines do not cross in either diagram
- **Hybrid**: Only one of the main effects is globally interpretable in addition to the interaction
  - The lines cross in only one of the two diagrams
- **Disordinal**: Only the interaction is interpretable
  - *The lines cross in both diagrams*
  - → Reversal of the effect of the respective other factor

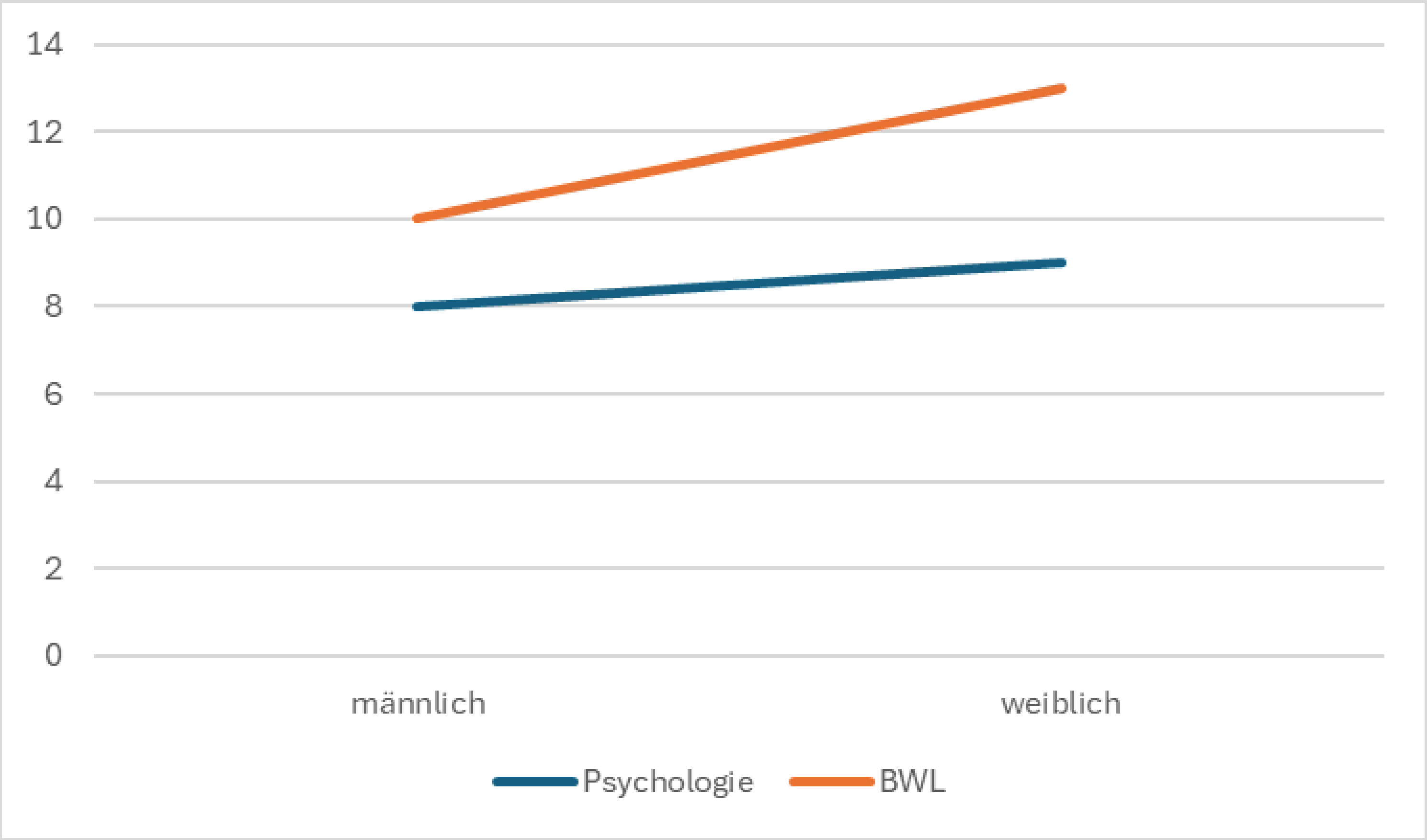
- The site https://statistikgrundlagen.de/ebook/chapter/zweifaktorielle-varianzanalyse_two_way_anova/ visualises the three types of interaction very nice

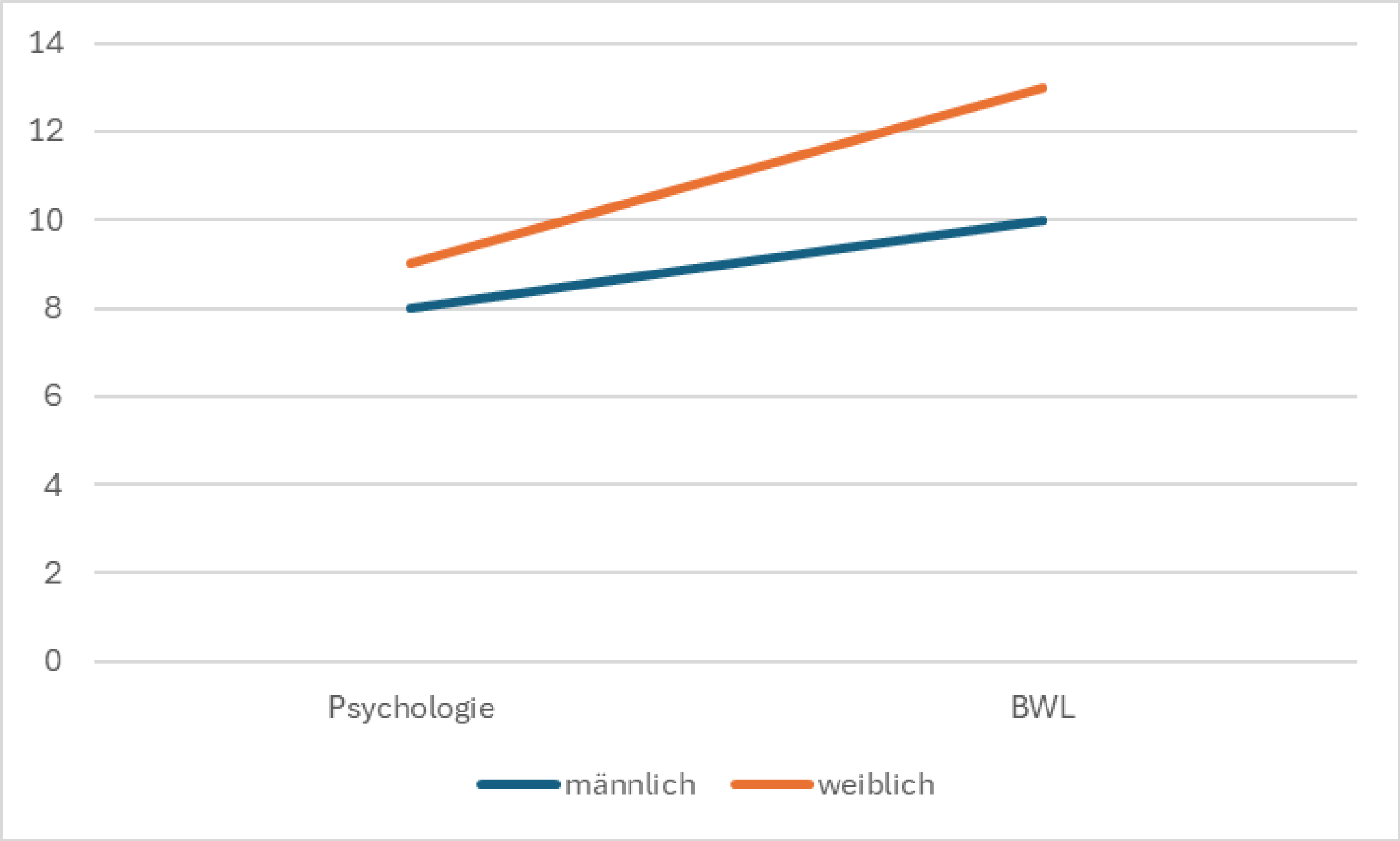
On the following slides, it is assumed that all means with different values represent significant differences!

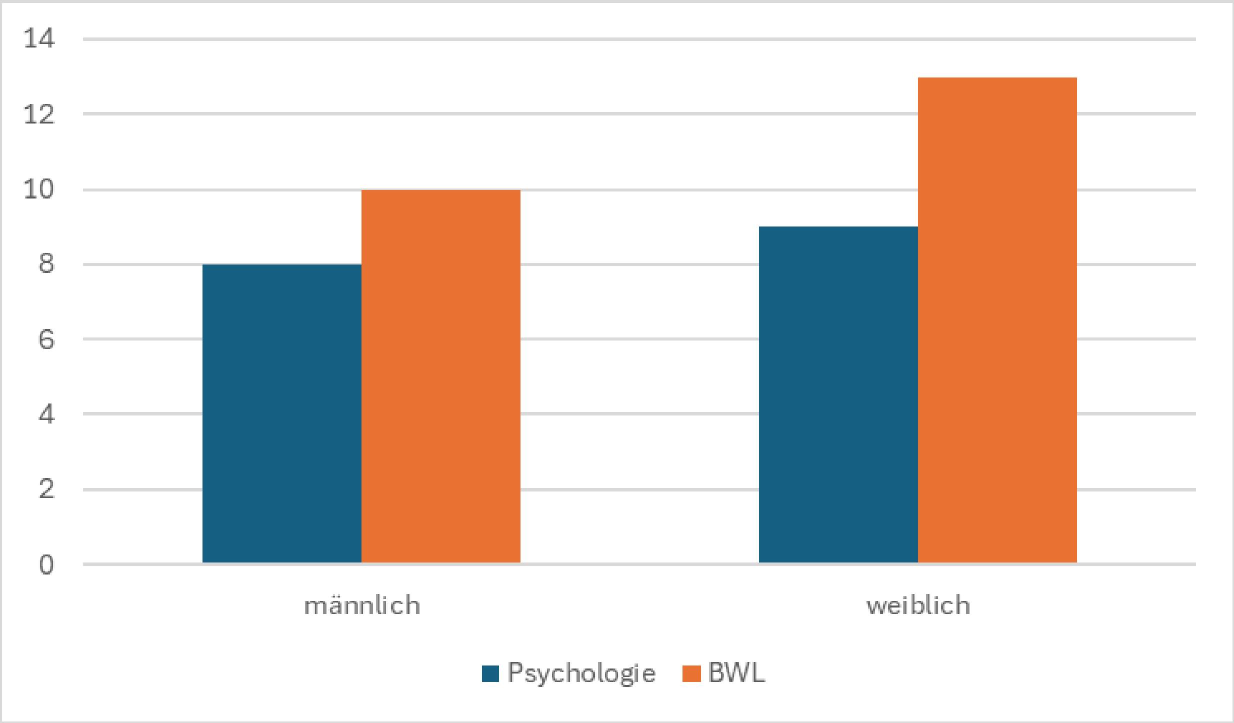
Ordinale interaction

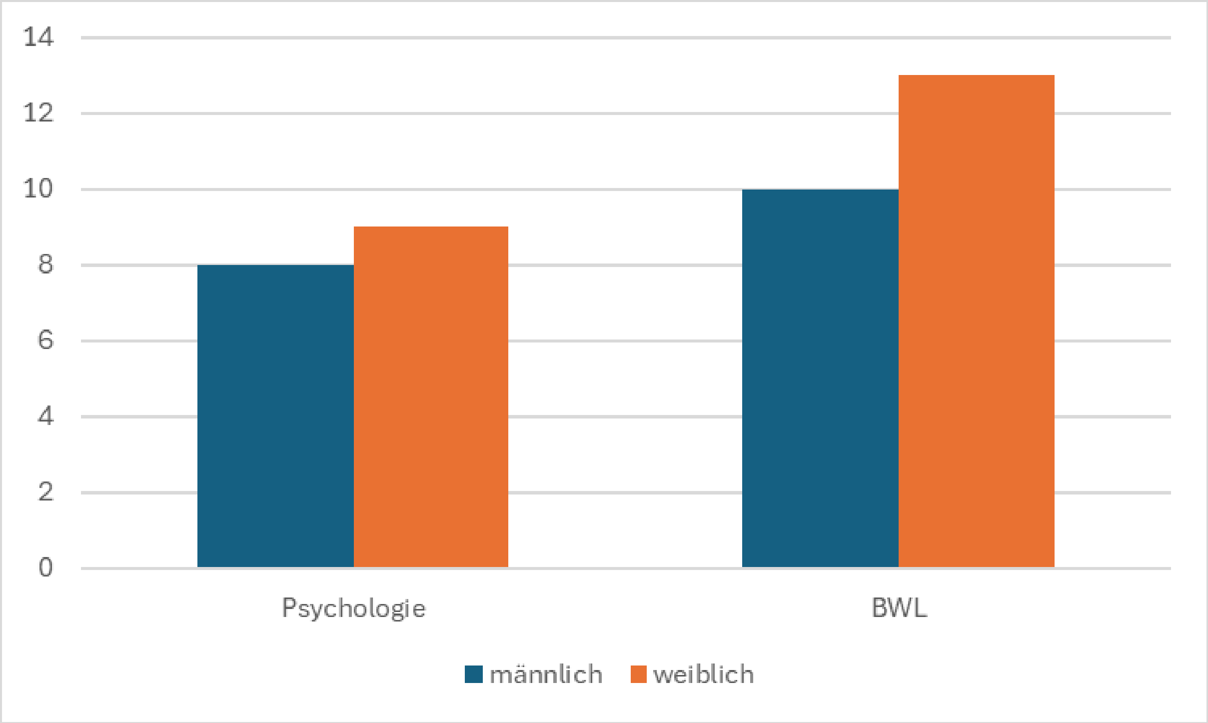
	männlich	weiblich
Psychologie	8	9
BWL	10	13

	Psychologie	BWL
männlich	8	10
weiblich	9	13





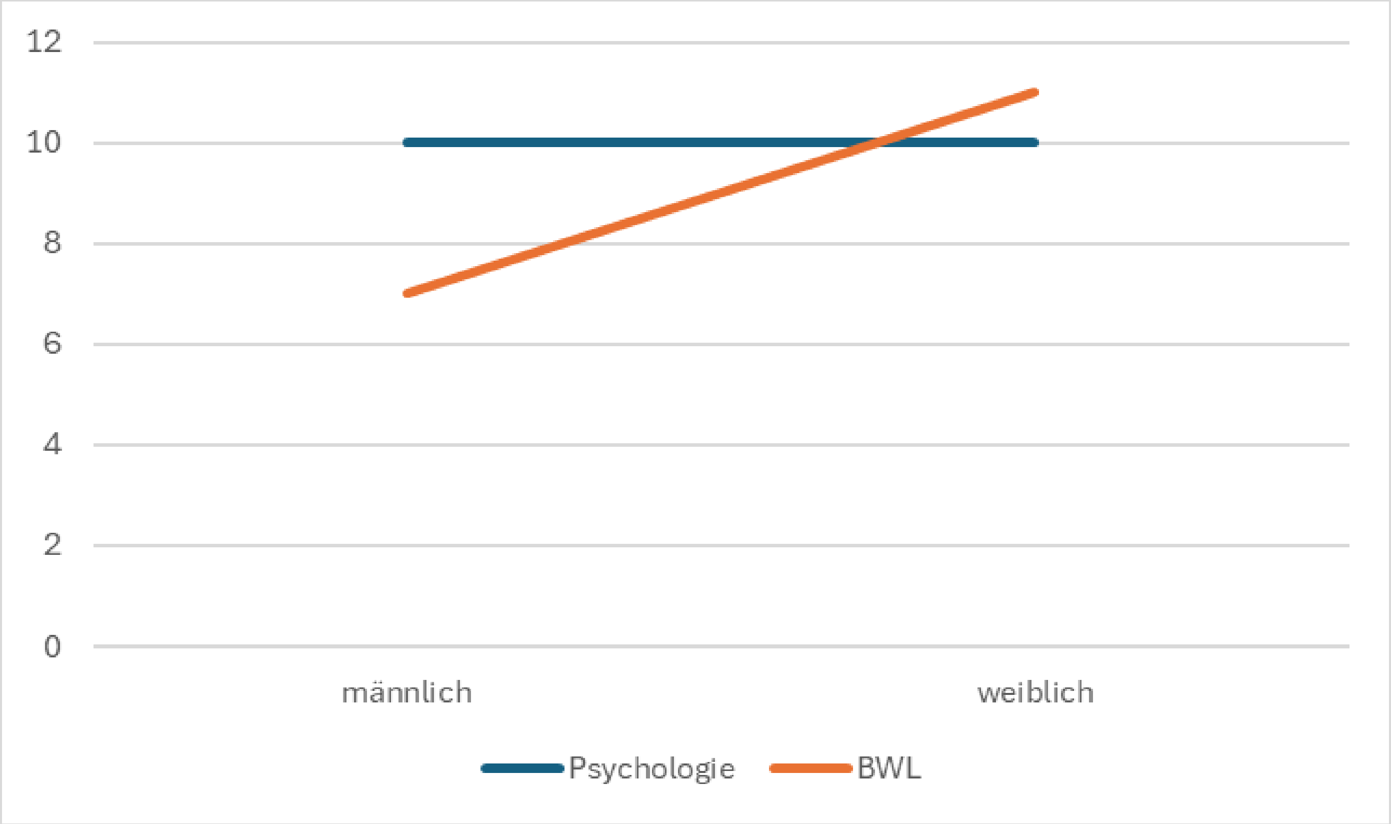


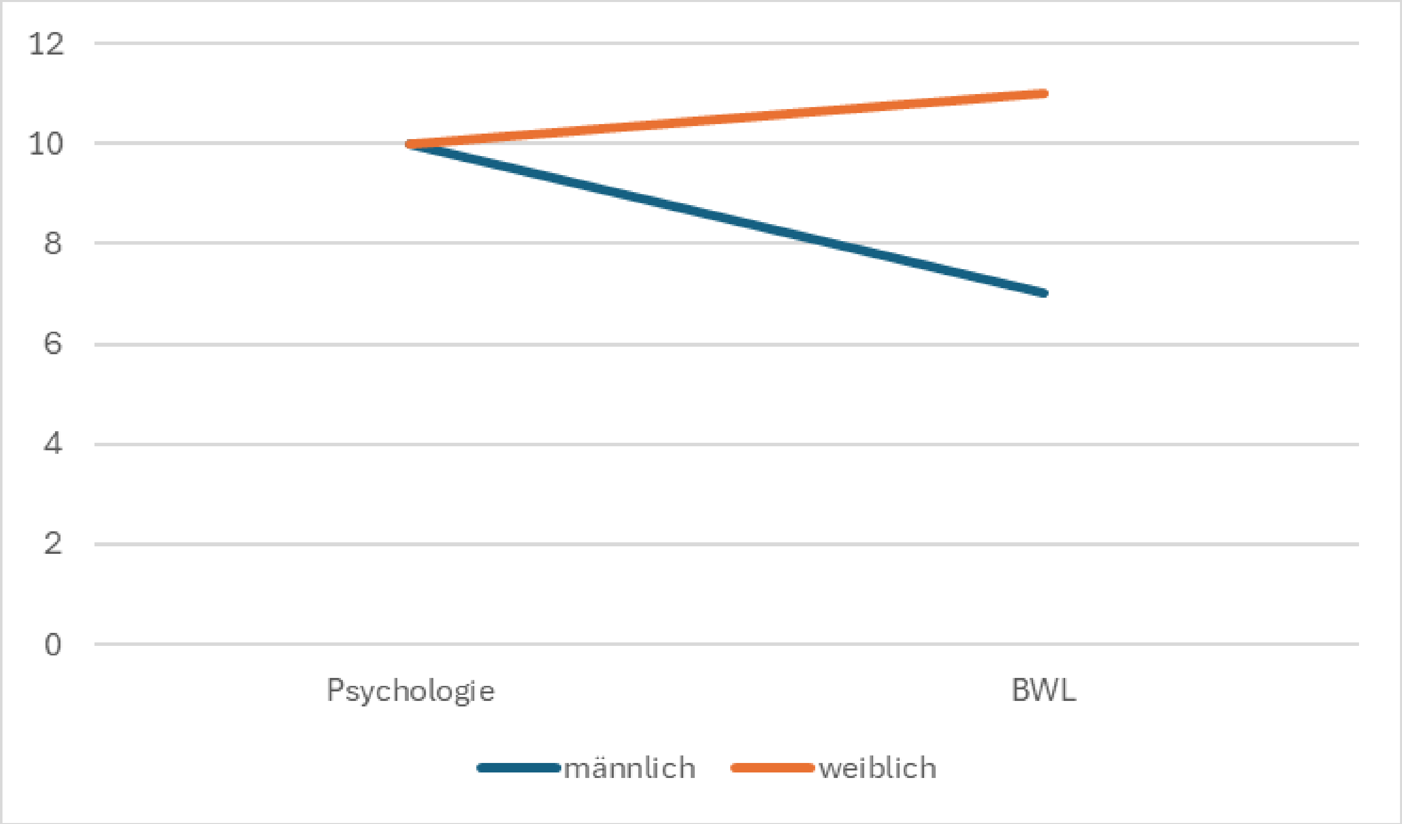


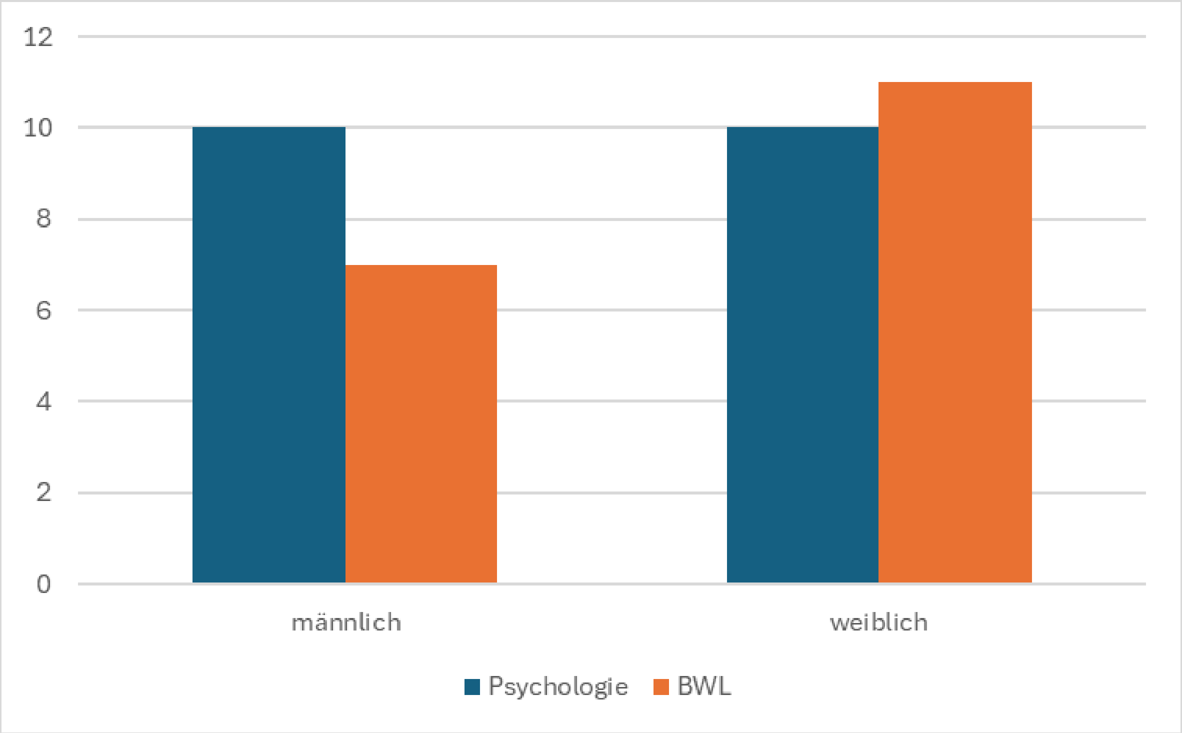
Hybrid interaction

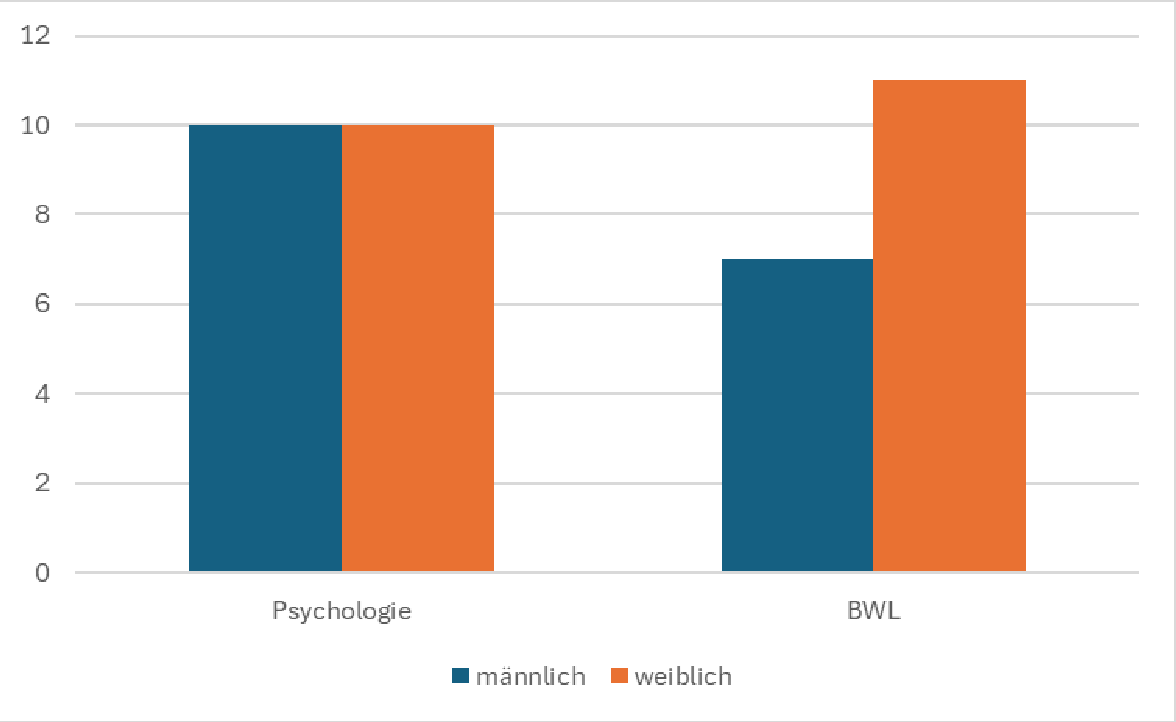
	männlich	weiblich
Psychologie	10	10
BWL	7	11

	Psychologie	BWL
männlich	10	7
weiblich	10	11





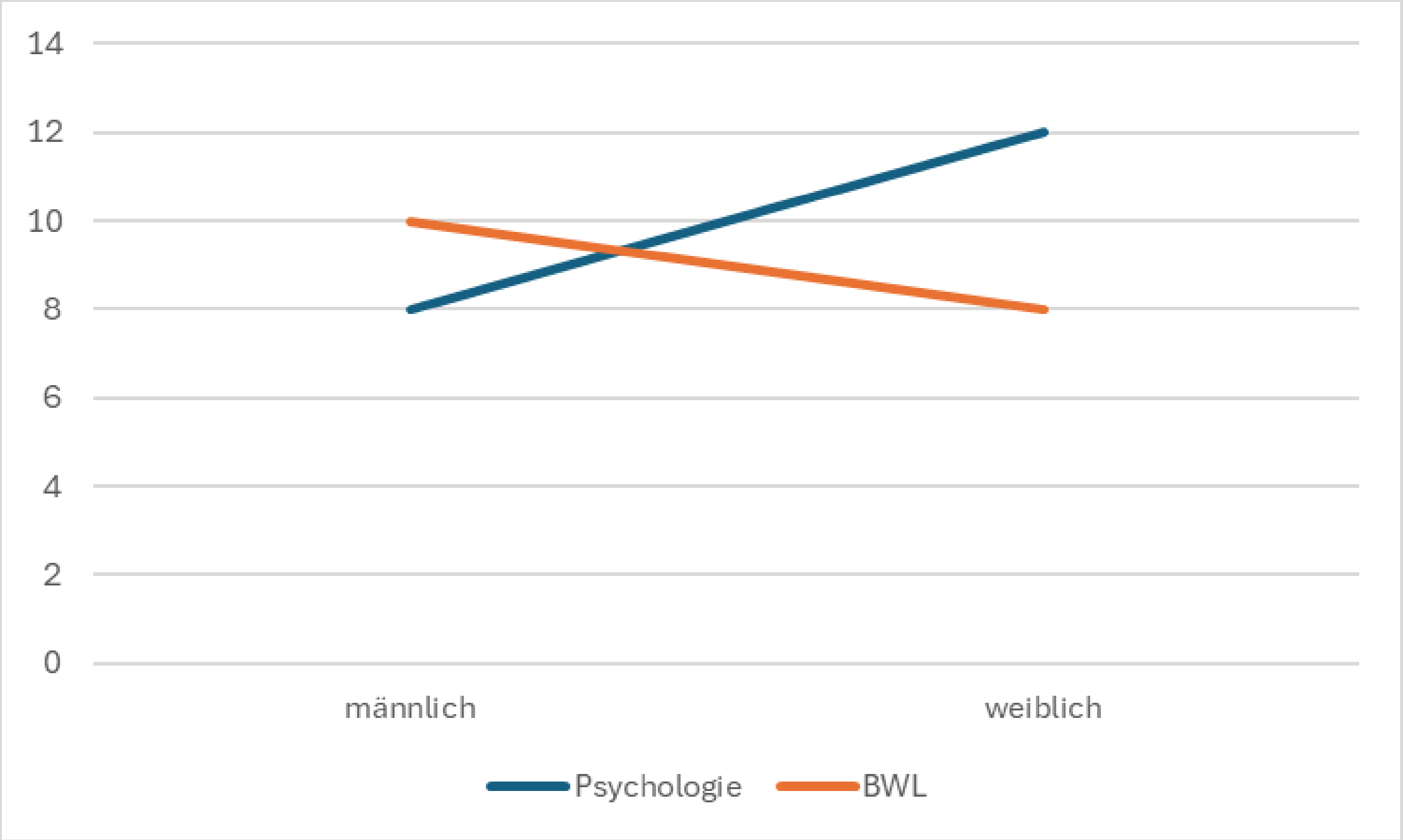


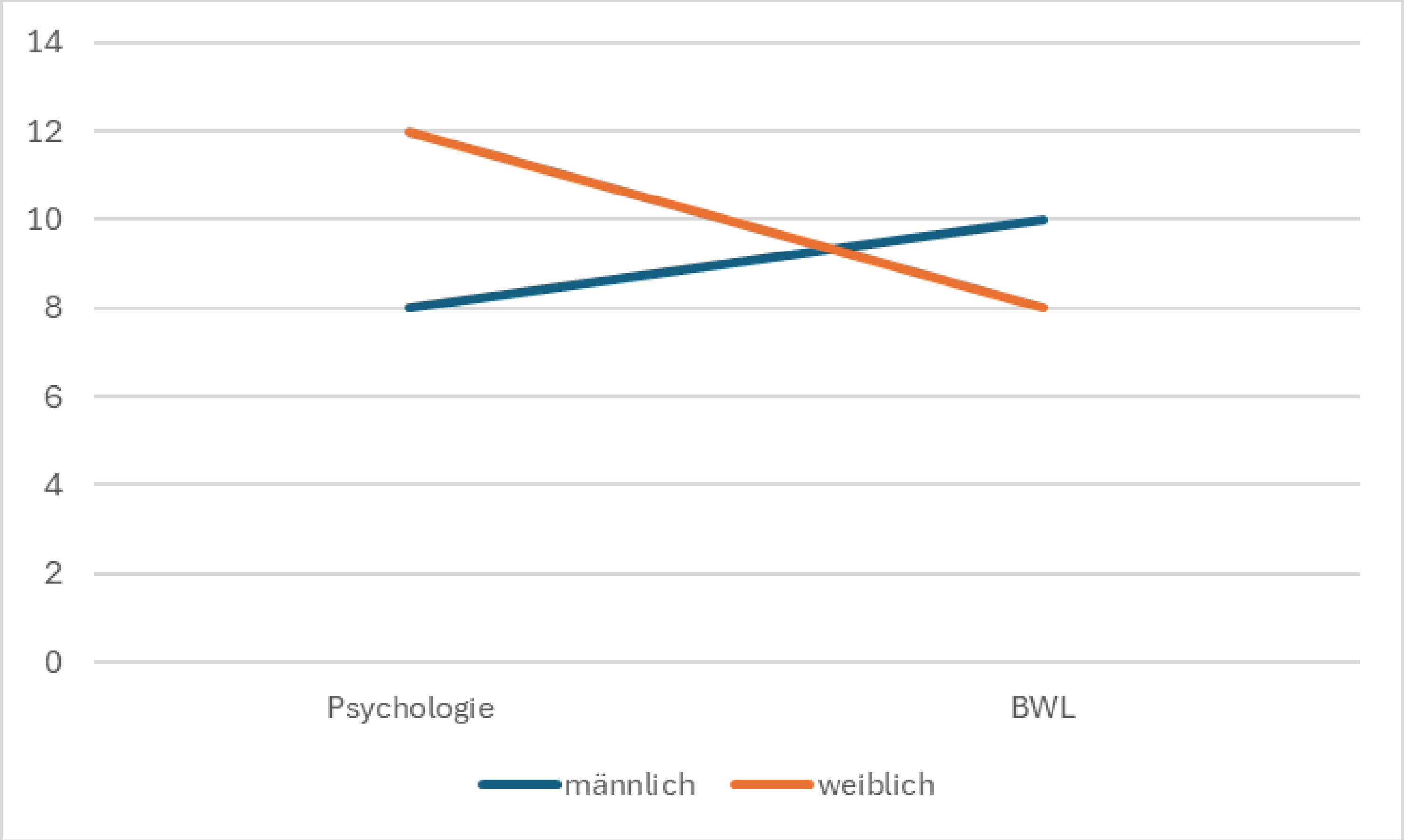


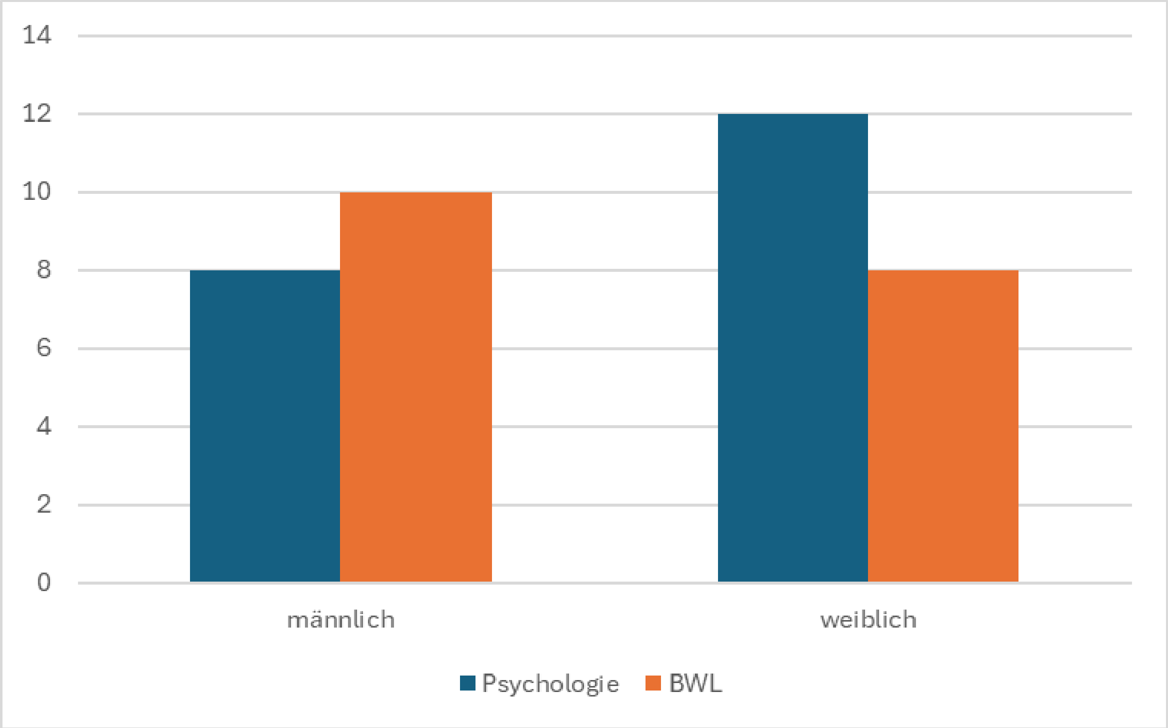
Disordinal interaction

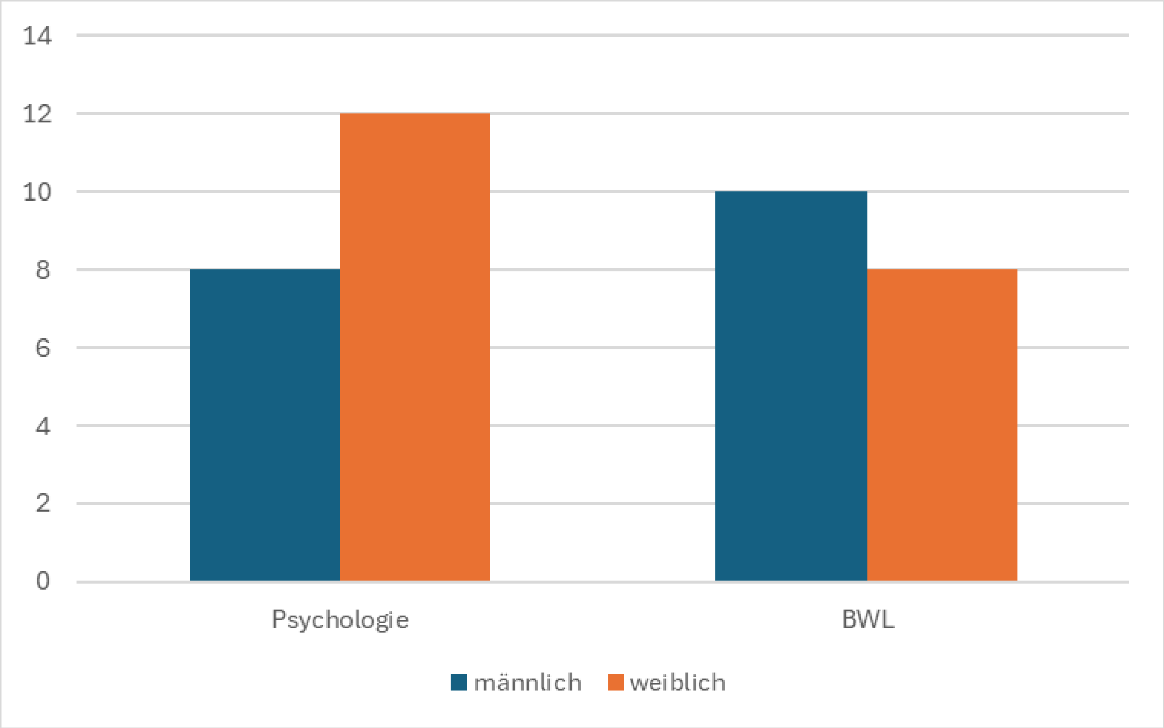
	männlich	weiblich
Psychologie	8	12
BWL	10	8

	Psychologie	BWL
männlich	8	10
weiblich	12	8





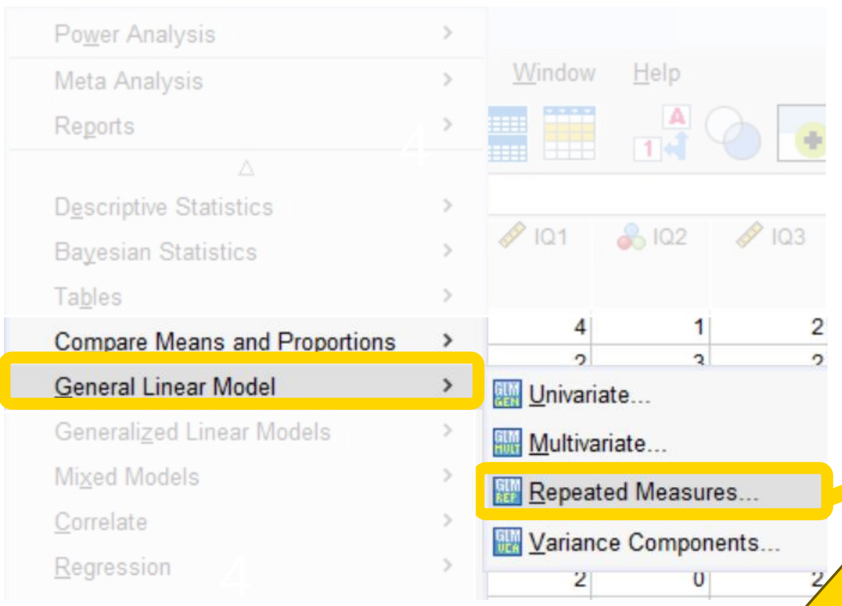


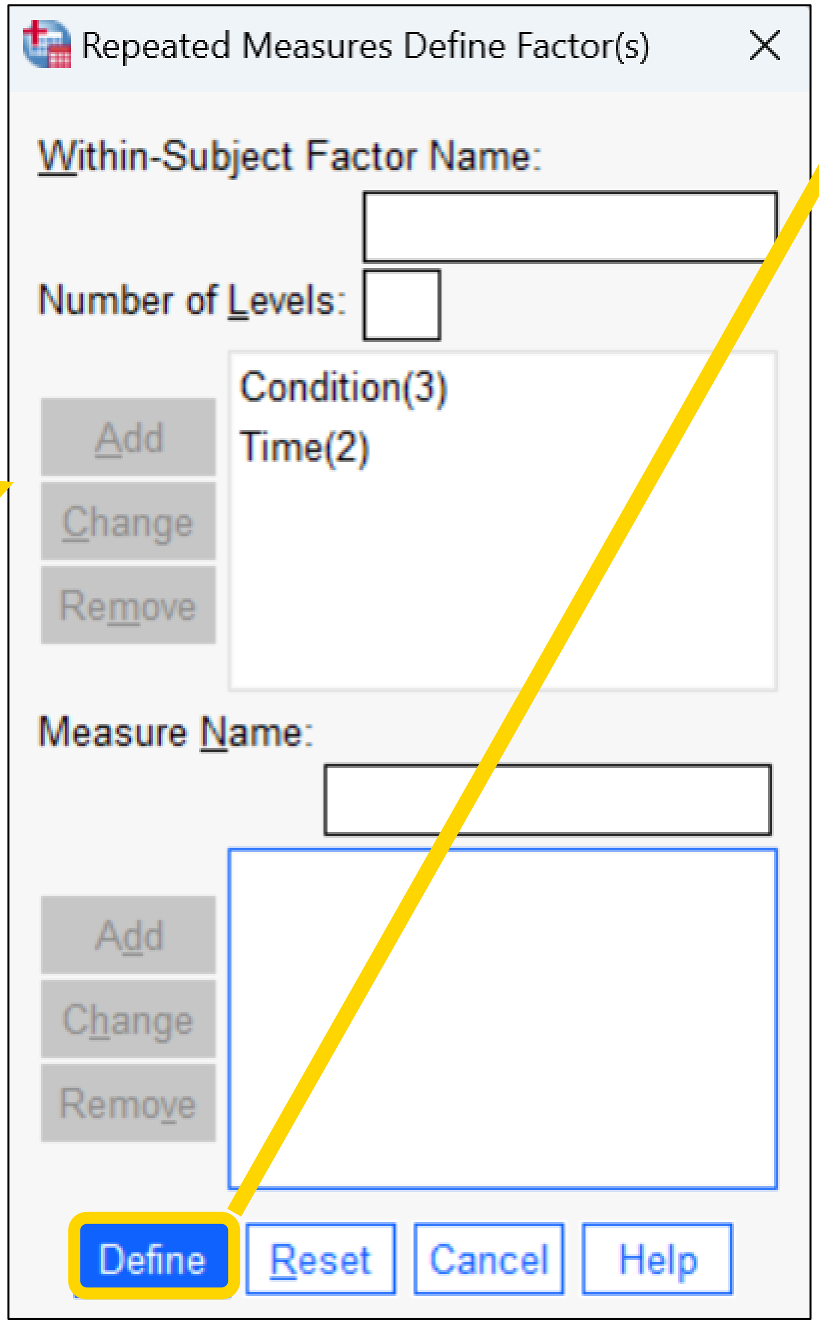


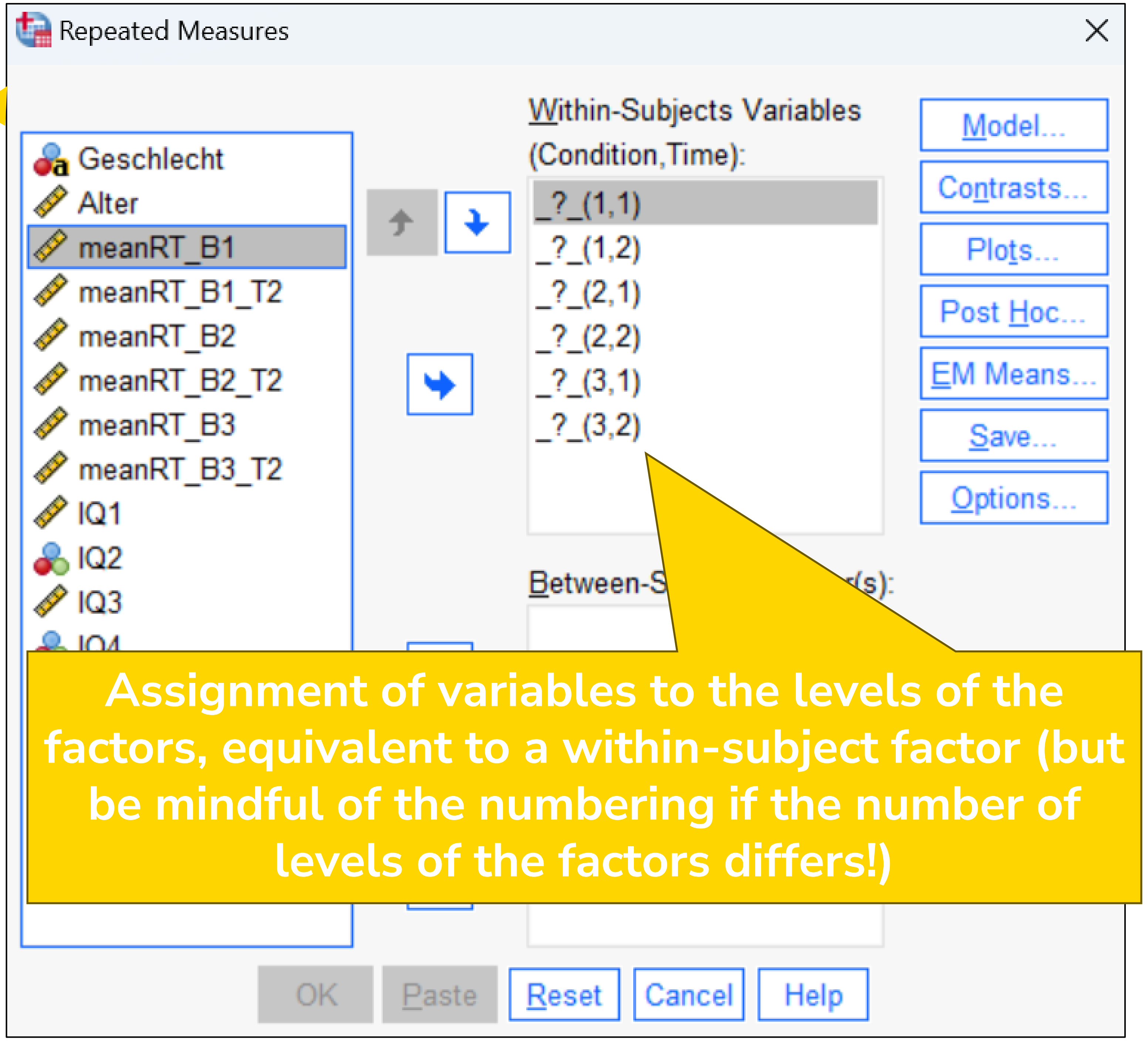
ANOVA with two within-subjects Factors

ANOVA with two within-subjects Factors

Zunächst die beiden within-subjects Faktoren definiert:



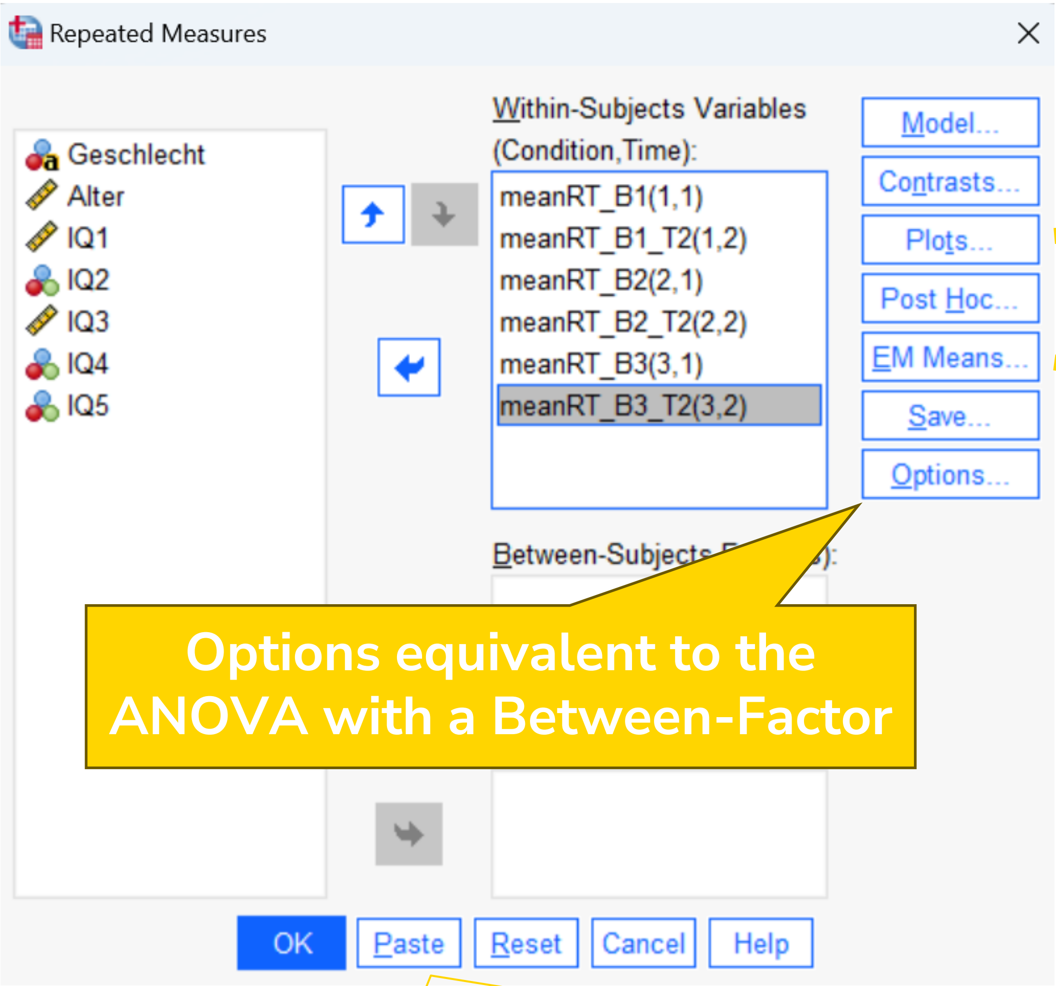




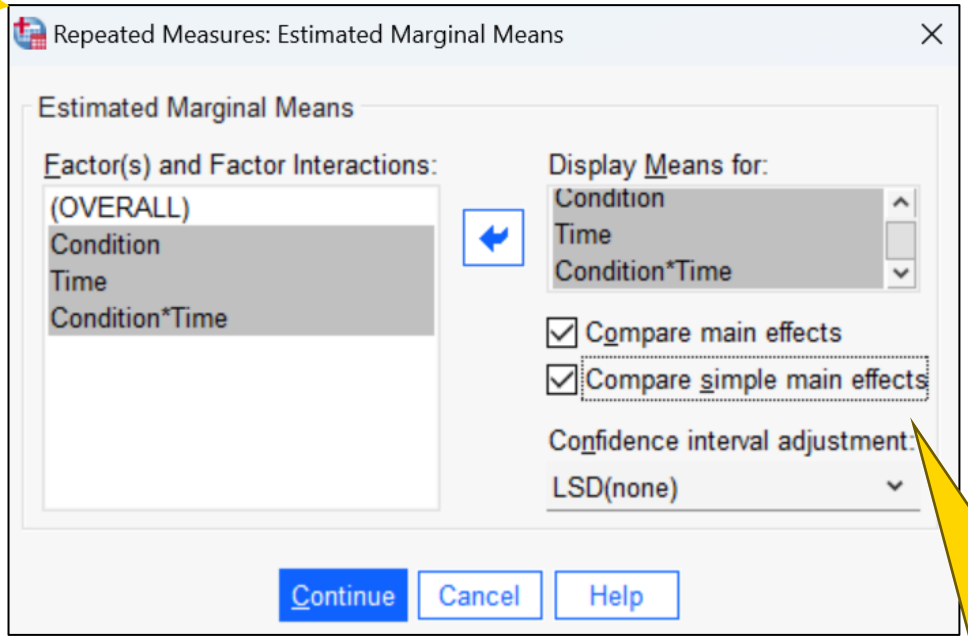
Define the name and number of levels of the within-subject factors and add them with "Add"(see one-way ANOVA with within-subjects factor)

Assignment of variables to the levels of the factors, equivalent to a within-subject factor (but be mindful of the numbering if the number of levels of the factors differs!)

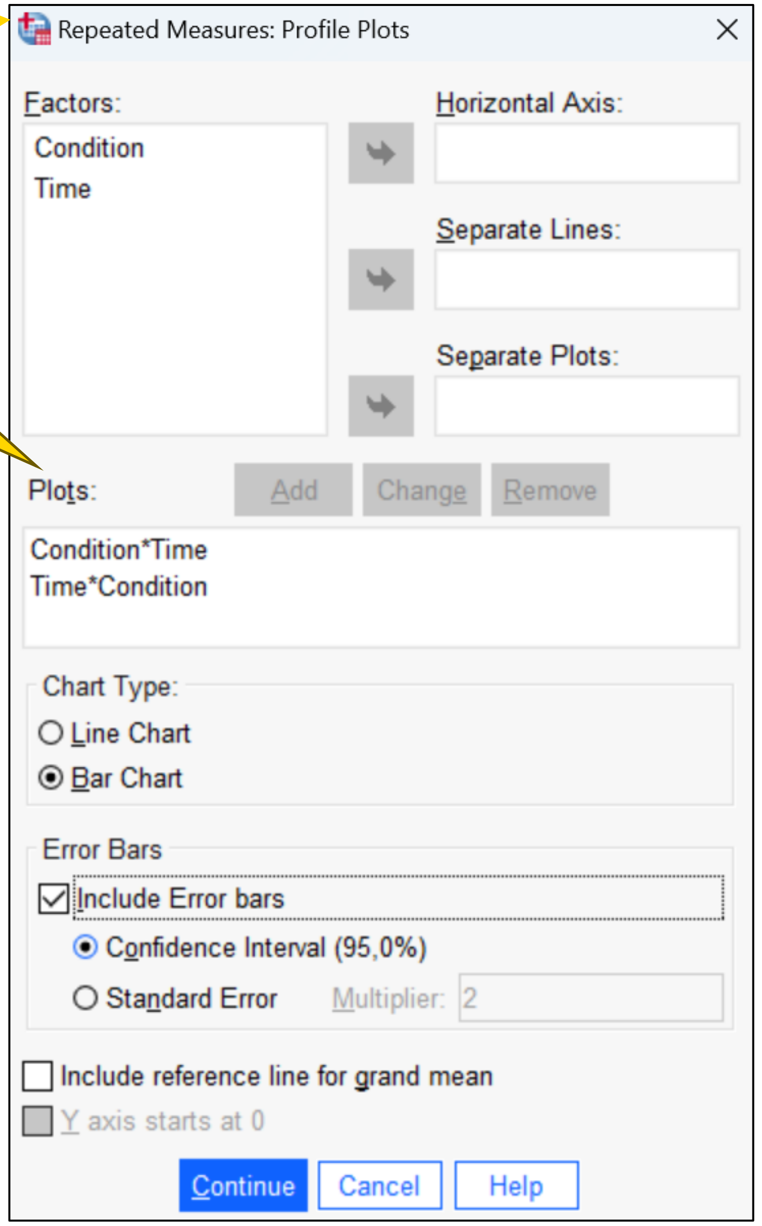
ANOVA with two within-subjects Factors



Options equivalent to the ANOVA with a Between-Factor



Generate plots for both AxB and BxA!



Request pairwise comparisons for both main effects and the interaction (+ correction method – Bonferroni!)

Exemplary syntax command

```
GLM meanRT_B1 meanRT_B1_T2 meanRT_B2 meanRT_B2_T2 meanRT_B3 meanRT_B3_T2
  /WSFACTOR=Condition 3 Polynomial Time 2 Polynomial
  /METHOD=SSTYPE(3)
  /PLOT=PROFILE(Condition*Time Time*Condition) TYPE=BAR ERRORBAR=CI
MEANREFERENCE=NO
  /EMMEANS=TABLES(Condition) COMPARE ADJ(LSD)
  /EMMEANS=TABLES(Time) COMPARE ADJ(LSD)
  /EMMEANS=TABLES(Condition*Time) COMPARE(Condition) ADJ(LSD)
  /EMMEANS=TABLES(Condition*Time) COMPARE(Time) ADJ(LSD)
  /PRINT=DESCRIPTIVE ETASQ
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=Condition Time Condition*Time.
```

ANOVA with two within-subjects Factors

Within-Subjects Factors

Measure: MEASURE_1

Condition	Time	Dependent Variable
1	1	meanRT_B1
	2	meanRT_B1_T2
2	1	meanRT_B2
	2	meanRT_B2_T2
3	1	meanRT_B3
	2	meanRT_B3_T2

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Condition	Pillai's Trace	,012	,889[b]	2,000	148,000	,413	,012
	Wilks' Lambda	,988	,889[b]	2,000	148,000	,413	,012
	Hotelling's Trace	,012	,889[b]	2,000	148,000	,413	,012
	Roy's Largest Root	,012	,889[b]	2,000	148,000	,413	,012
Time	Pillai's Trace	,016	2,405[b]	1,000	149,000	,123	,016
	Wilks' Lambda	,984	2,405[b]	1,000	149,000	,123	,016
	Hotelling's Trace	,016	2,405[b]	1,000	149,000	,123	,016
	Roy's Largest Root	,016	2,405[b]	1,000	149,000	,123	,016
Condition * Time	Pillai's Trace	,053	4,152[b]	2,000	148,000	,018	,053
	Wilks' Lambda	,947	4,152[b]	2,000	148,000	,018	,053
	Hotelling's Trace	,056	4,152[b]	2,000	148,000	,018	,053
	Roy's Largest Root	,056	4,152[b]	2,000	148,000	,018	,053

a. Design: Intercept
Within Subjects Design: Condition + Time + Condition * Time

b. Exact statistic

Multivariate tests can also be ignored here

Mauchly's Test of Sphericity for both factors (irrelevant for "Time" since this factor only has two levels) + interaction → Indicates here as well whether a correction procedure should be applied when examining the effects/interactions or not

Descriptive Statistics

	Mean	Std. Deviation	N
meanRT_B1	702,62	117,955	150
meanRT_B1_T2	641,71	200,903	150
meanRT_B2	682,11	99,675	150
meanRT_B2_T2	693,69	182,375	150
meanRT_B3	679,54	102,824	150
meanRT_B3_T2	680,01	189,360	150

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Condition	,982	2,711	2	,258	,982	,995	,500
Time	1,000	,000	0	.	1,000	1,000	1,000
Condition * Time	,991	1,395	2	,498	,991	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Condition + Time + Condition * Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Description of the factors and descriptive statistics, equivalent to a one-way ANOVA with a within-subjects factor, expanded to include the levels of the other factor

ANOVA with two within-subjects Factors

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Condition	Sphericity Assumed	37143,842	2	18571,921	,812	,445	,005
	Greenhouse-Geisser	37143,842	1,964	18908,987	,812	,443	,005
	Huynh-Feldt	37143,842	1,990	18661,973	,812	,445	,005
	Lower-bound	37143,842	1,000	37143,842	,812	,369	,005
Error(Condition)	Sphericity Assumed	6817171,158	298	22876,413			
	Greenhouse-Geisser	6817171,158	292,688	23291,603			
	Huynh-Feldt	6817171,158	296,562	22987,337			
	Lower-bound	6817171,158	149,000	45752,827			
Time	Sphericity Assumed	59649,921	1	59649,921	2,405	,123	,016
	Greenhouse-Geisser	59649,921	1,000	59649,921	2,405	,123	,016
	Huynh-Feldt	59649,921	1,000	59649,921	2,405	,123	,016
	Lower-bound	59649,921	1,000	59649,921	2,405	,123	,016
Error(Time)	Sphericity Assumed	3695247,912	149	24800,322			
	Greenhouse-Geisser	3695247,912	149,000	24800,322			
	Huynh-Feldt	3695247,912	149,000	24800,322			
	Lower-bound	3695247,912	149,000	24800,322			
Condition * Time	Sphericity Assumed	228657,349	2	114328,674	4,573	,011	,030
	Greenhouse-Geisser	228657,349	1,981	115401,355	4,573	,011	,030
	Huynh-Feldt	228657,349	2,000	114328,674	4,573	,011	,030
	Lower-bound	228657,349	1,000	228657,349	4,573	,034	,030
Error(Condition*Time)	Sphericity Assumed	7450466,318	298	25001,565			
	Greenhouse-Geisser	7450466,318	295,230	25236,140			
	Huynh-Feldt	7450466,318	298,000	25001,565			
	Lower-bound	7450466,318	149,000	50003,130			

Statistics for the effect of the factor „Condition“

Statistics for the effect of the factor "Time"

Statistics for the effect of the interaction „Condition x Time“

Reporting the effects/interactions is equivalent to the one-way ANOVA with within-subjects factor

ANOVA with two within-subjects Factors

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Condition	Time	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Condition	Linear		8686,815	1	8686,815	,335	,564	,002
	Quadratic		28457,027	1	28457,027	1,437	,233	,010
Error(Condition)	Linear		3865776,435	149	25944,808			
	Quadratic		2951394,723	149	19808,018			
Time		Linear	59649,921	1	59649,921	2,405	,123	,016
Error(Time)		Linear	3695247,912	149	24800,322			
Condition * Time	Linear	Linear	141281,415	1	141281,415	5,641	,019	,036
	Quadratic	Linear	87375,934	1	87375,934	3,501	,063	,023
Error(Condition*Time)	Linear	Linear	3731456,835	149	25043,334			
	Quadratic	Linear	3719009,483	149	24959,795			

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	416096082,45	1	416096082,45	17824,069	<,001	,992
Error	3478348,046	149	23344,618			

The table of contrasts (unless a contrast analysis was planned in advance) and the table of Between-Subjects effects can be ignored, as with the one-way ANOVA with a within-subject factor

ANOVA with two within-subjects Factors
Estimated Marginal Means

1. Condition

Estimates

Measure: MEASURE_1

Condition	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	672,167	9,045	654,294	690,039
2	687,900	8,443	671,217	704,583
3	679,777	8,789	662,410	697,143

Pairwise Comparisons

Measure: MEASURE_1

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-15,733	11,858	,187	-39,165	7,699
	3	-7,610	13,152	,564	-33,598	18,378
2	1	15,733	11,858	,187	-7,699	39,165
	3	8,123	11,998	,499	-15,584	31,831
3	1	7,610	13,152	,564	-18,378	33,598
	2	-8,123	11,998	,499	-31,831	15,584

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	,012	,889[a]	2,000	148,000	,413	,012
Wilks' lambda	,988	,889[a]	2,000	148,000	,413	,012
Hotelling's trace	,012	,889[a]	2,000	148,000	,413	,012
Roy's largest root	,012	,889[a]	2,000	148,000	,413	,012

Each F tests the multivariate effect of Condition. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Pairwise comparisons of the individual factors are equivalent to the corresponding one-way ANOVA

ANOVA with two within-subjects Factors
Estimated Marginal Means

2. Time

Estimates

Measure: MEASURE_1

Time	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	688,089	5,306	677,604	698,574
2	671,807	8,879	654,262	689,351

Pairwise Comparisons

Measure: MEASURE_1

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	95% Confidence Interval for Difference[a] Upper Bound
1	2	16,282	10,499	,123	-4,463	37,028
2	1	-16,282	10,499	,123	-37,028	4,463

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	,016	2,405[a]	1,000	149,000	,123	,016
Wilks' lambda	,984	2,405[a]	1,000	149,000	,123	,016
Hotelling's trace	,016	2,405[a]	1,000	149,000	,123	,016
Roy's largest root	,016	2,405[a]	1,000	149,000	,123	,016

Each F tests the multivariate effect of Time. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Pairwise comparisons of the individual factors are equivalent to the corresponding one-way ANOVA

ANOVA with two within-subjects Factors
Estimated Marginal Means

3. Condition * Time

Estimates

Measure: MEASURE_1

Condition	Time	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1	702,620	9,631	683,589	721,651
	2	641,713	16,404	609,300	674,127
2	1	682,107	8,138	666,025	698,188
	2	693,693	14,891	664,269	723,118
3	1	679,540	8,396	662,950	696,130
	2	680,013	15,461	649,462	710,565

Multivariate Tests

Time		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
1	Pillai's trace	,026	1,955[a]	2,000	148,000	,145	,026
	Wilks' lambda	,974	1,955[a]	2,000	148,000	,145	,026
	Hotelling's trace	,026	1,955[a]	2,000	148,000	,145	,026
	Roy's largest root	,026	1,955[a]	2,000	148,000	,145	,026
2	Pillai's trace	,037	2,840[a]	2,000	148,000	,062	,037
	Wilks' lambda	,963	2,840[a]	2,000	148,000	,062	,037
	Hotelling's trace	,038	2,840[a]	2,000	148,000	,062	,037
	Roy's largest root	,038	2,840[a]	2,000	148,000	,062	,037

Each F tests the multivariate simple effects of Condition within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Pairwise Comparisons

Measure: MEASURE_1

Time	(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
1	1	2	20,513	12,236	,096	-3,665	44,691
		3	23,080	12,546	,068	-1,710	47,870
	2	1	-20,513	12,236	,096	-44,691	3,665
		3	2,567	11,310	,821	-19,782	24,915
	3	1	-23,080	12,546	,068	-47,870	1,710
		2	-2,567	11,310	,821	-24,915	19,782
2	1	2	-51,980*	22,194	,020	-95,835	-8,125
		3	-38,300	22,857	,096	-83,466	6,866
	2	1	51,980*	22,194	,020	8,125	95,835
		3	13,680	21,566	,527	-28,935	56,295
	3	1	38,300	22,857	,096	-6,866	83,466
		2	-13,680	21,566	,527	-56,295	28,935

Based on estimated marginal means

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

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Pairwise comparisons of the interaction are the "actually relevant/interesting" ones! Are there differences between the levels of the other factor (columns 2+3) within a specific level of one factor (leftmost column)?It is very important to always consider both directions!

ANOVA with two within-subjects Factors
Estimated Marginal Means

4. Condition * Time

Estimates

Measure: MEASURE_1

Condition	Time	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1	702,620	9,631	683,589	721,651
	2	641,713	16,404	609,300	674,127
2	1	682,107	8,138	666,025	698,188
	2	693,693	14,891	664,269	723,118
3	1	679,540	8,396	662,950	696,130
	2	680,013	15,461	649,462	710,565

Multivariate Tests

Condition		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
1	Pillai's trace	,059	9,357[a]	1,000	149,000	,003	,059
	Wilks' lambda	,941	9,357[a]	1,000	149,000	,003	,059
	Hotelling's trace	,063	9,357[a]	1,000	149,000	,003	,059
	Roy's largest root	,063	9,357[a]	1,000	149,000	,003	,059
2	Pillai's trace	,003	,462[a]	1,000	149,000	,498	,003
	Wilks' lambda	,997	,462[a]	1,000	149,000	,498	,003
	Hotelling's trace	,003	,462[a]	1,000	149,000	,498	,003
	Roy's largest root	,003	,462[a]	1,000	149,000	,498	,003
3	Pillai's trace	,000	,001[a]	1,000	149,000	,979	,000
	Wilks' lambda	1,000	,001[a]	1,000	149,000	,979	,000
	Hotelling's trace	,000	,001[a]	1,000	149,000	,979	,000
	Roy's largest root	,000	,001[a]	1,000	149,000	,979	,000

Each F tests the multivariate simple effects of Time within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Pairwise Comparisons

Measure: MEASURE_1

Condition	(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
1	1	2	60,907*	19,911	,003	21,562	100,251
	2	1	-60,907*	19,911	,003	-100,251	-21,562
2	1	2	-11,587	17,053	,498	-45,284	22,111
	2	1	11,587	17,053	,498	-22,111	45,284
3	1	2	-,473	17,610	,979	-35,271	34,324
	2	1	,473	17,610	,979	-34,324	35,271

Based on estimated marginal means

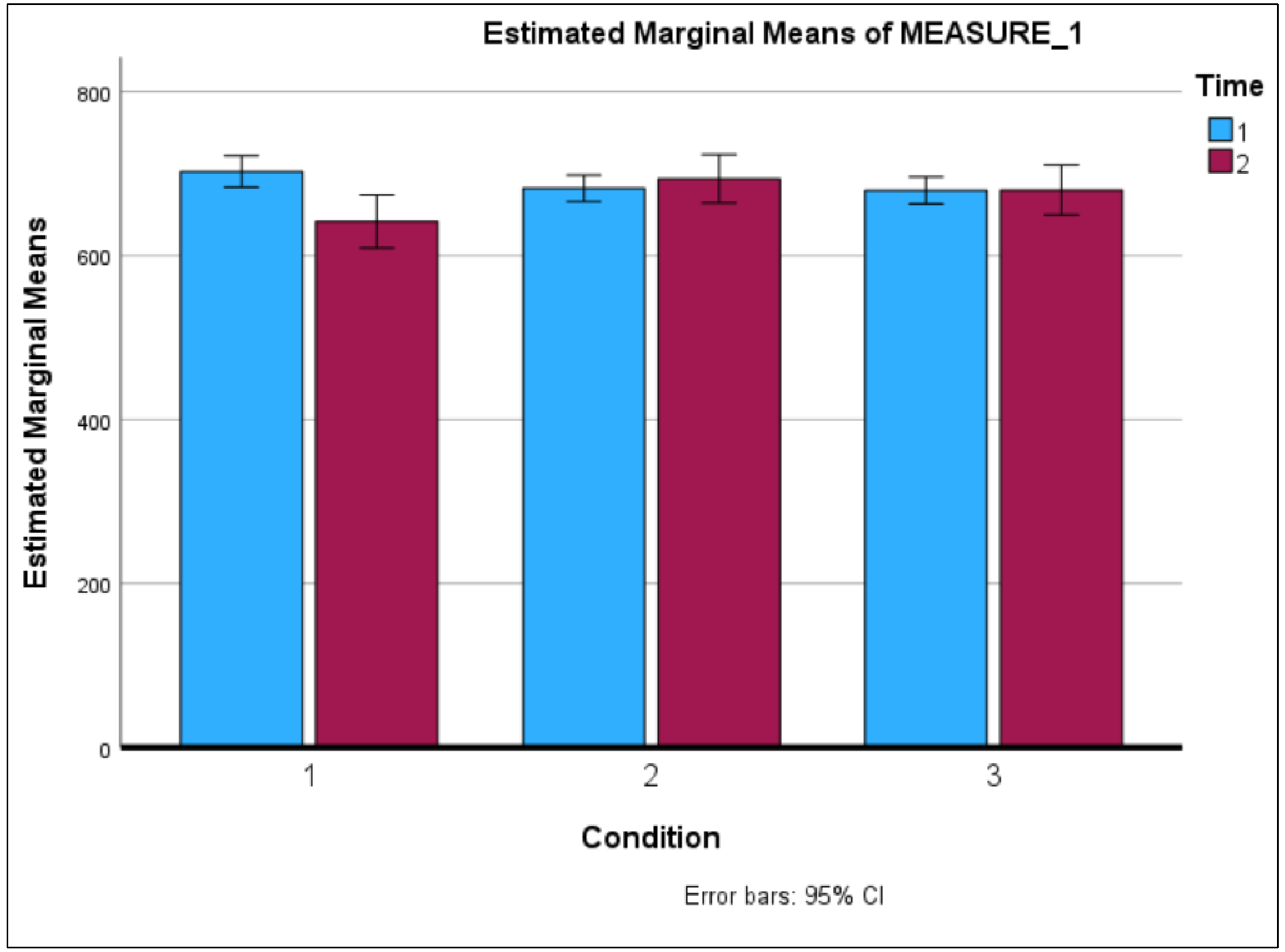
*. The mean difference is significant at the ,05 level.

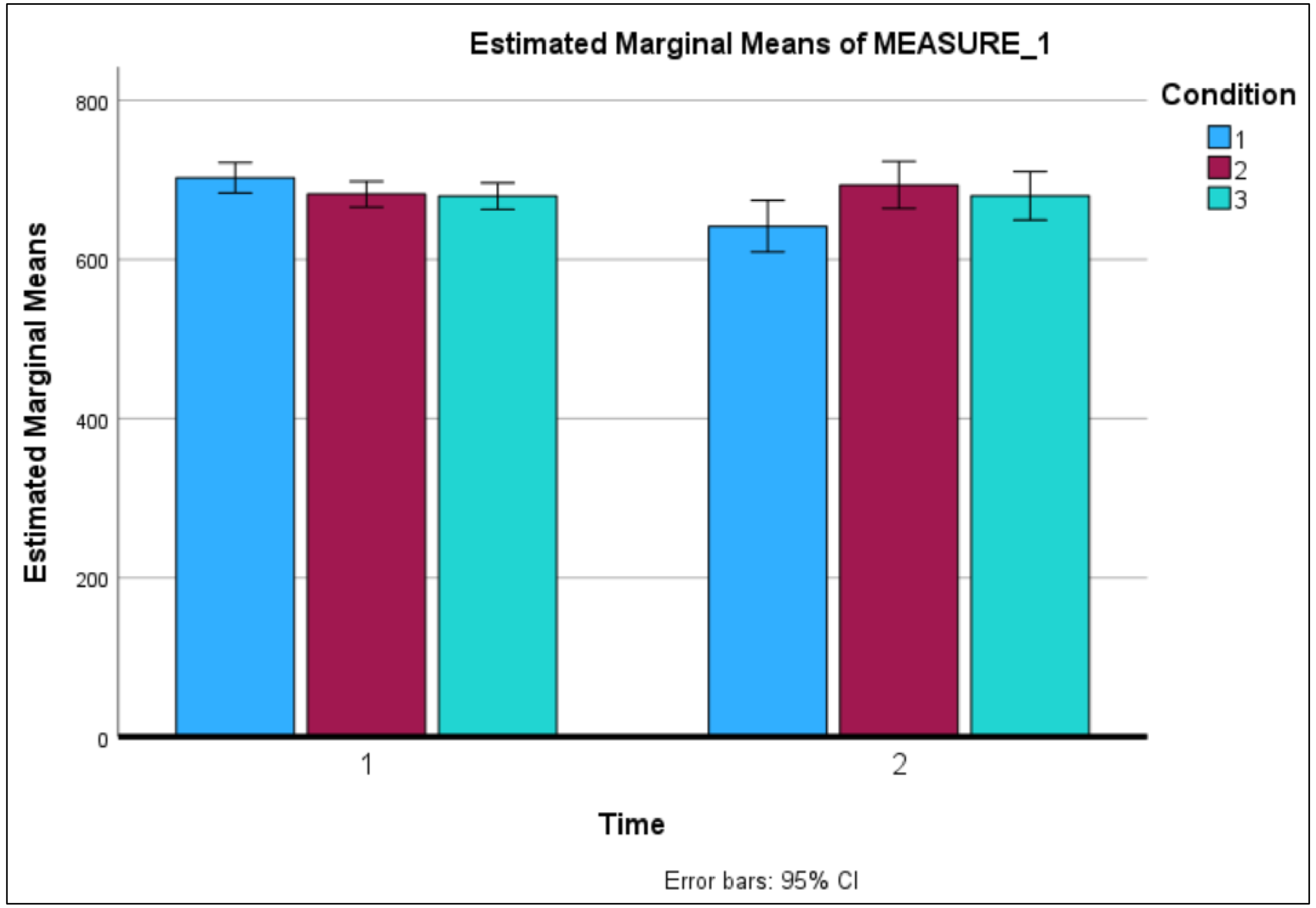
b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Pairwise comparisons of the interaction are the "actually relevant/interesting" ones! Are there differences between the levels of the other factor (columns 2+3) within a specific level of one factor (leftmost column)?It is very important to always consider both directions!

ANOVA with two within-subjects Factors
Estimated Marginal Means

- Always look at both figures!





Reporting an ANOVA with more than one factor

Reporting an ANOVA with more than one factor

- Relevant descriptive statistics. Equivalent to a one-way ANOVA (Note: for pairwise comparisons within the interaction, use the correct descriptive values!)
- F-/Omnibus Test: F(df, df_error) = F_value, *p* =/< p-value, η_p^2 = value
- Ideally, discuss with the supervising person whether main effects and interactions should always be reported. In principle, we begin with the interaction, as it may provide the most interesting information! Depending on the type of interaction, the main effects may no longer be meaningfully interpretable.
- Reporting of pairwise comparisons is equivalent to a one-way ANOVA
- In the case of a significant Mauchly's Test of Sphericity, as with a one-way ANOVA with a within-subjects factor, it is necessary to report which correction method was used (e.g., Greenhouse-Geisser). In the event of a violation of the assumption of homogeneity of variances (Box M test or individual Levene tests), it should be discussed with the supervising person how to proceed (if the groups are of equal size, this is usually not a problem, but still report the test and justify with the group size!).

Reporting an ANOVA with more than one factor

- <u>**Always**</u> prioritize the content of the findings, with <u>statistical measures serving only to support the statement!</u>
- <u>**Instead of saying:**</u> "The interaction between condition and time showed a significant result“

- <u>**Say:**</u> A two-way within/mixed ANOVA was conducted with the factors 'Name of the first factor' (levels of the first factor) and 'Name of the second factor' (levels of the second factor). The descriptive statistics can be found in Table Y. The significant interaction revealed that Factor 1 had a different effect on 'Name of the dependent variable' depending on Factor 2 ($F(2, 298) = 4.57$, $p = .011$, $\eta_p^2 = .03$). In Condition 1, significantly higher values were observed at Time 1 compared to Condition 2 (Mean_Diff = 35.36, 95% CI [15.77, 54.96], p < .001). At Time 2, Condition 2 showed lower values than Condition 3 (Mean_Diff = -28.47, 95% CI [-46.27, -9.65], *p* = .004). Additionally, no differences were found between the conditions. Within a condition, no changes in values were observed across the different measurement times (non-significant pairwise comparisons can be found in Table Z). (Effect size for the pairwise comparisons can be calculated and also reported)

ANOVA with a Three-Way Interaction

2 between and 1 within-Factor

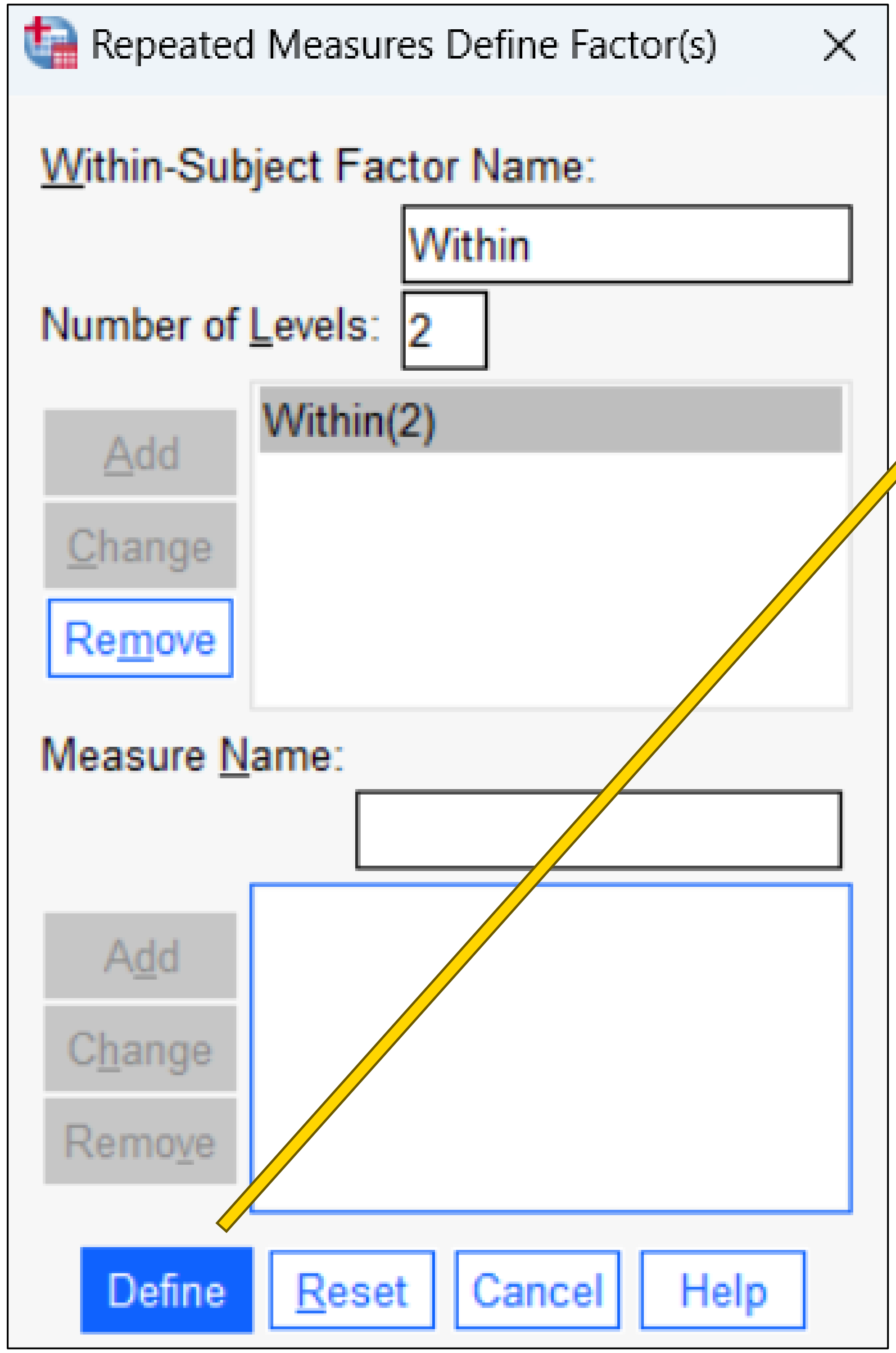
ANOVA – 3-factors (2 betw, 1 with)

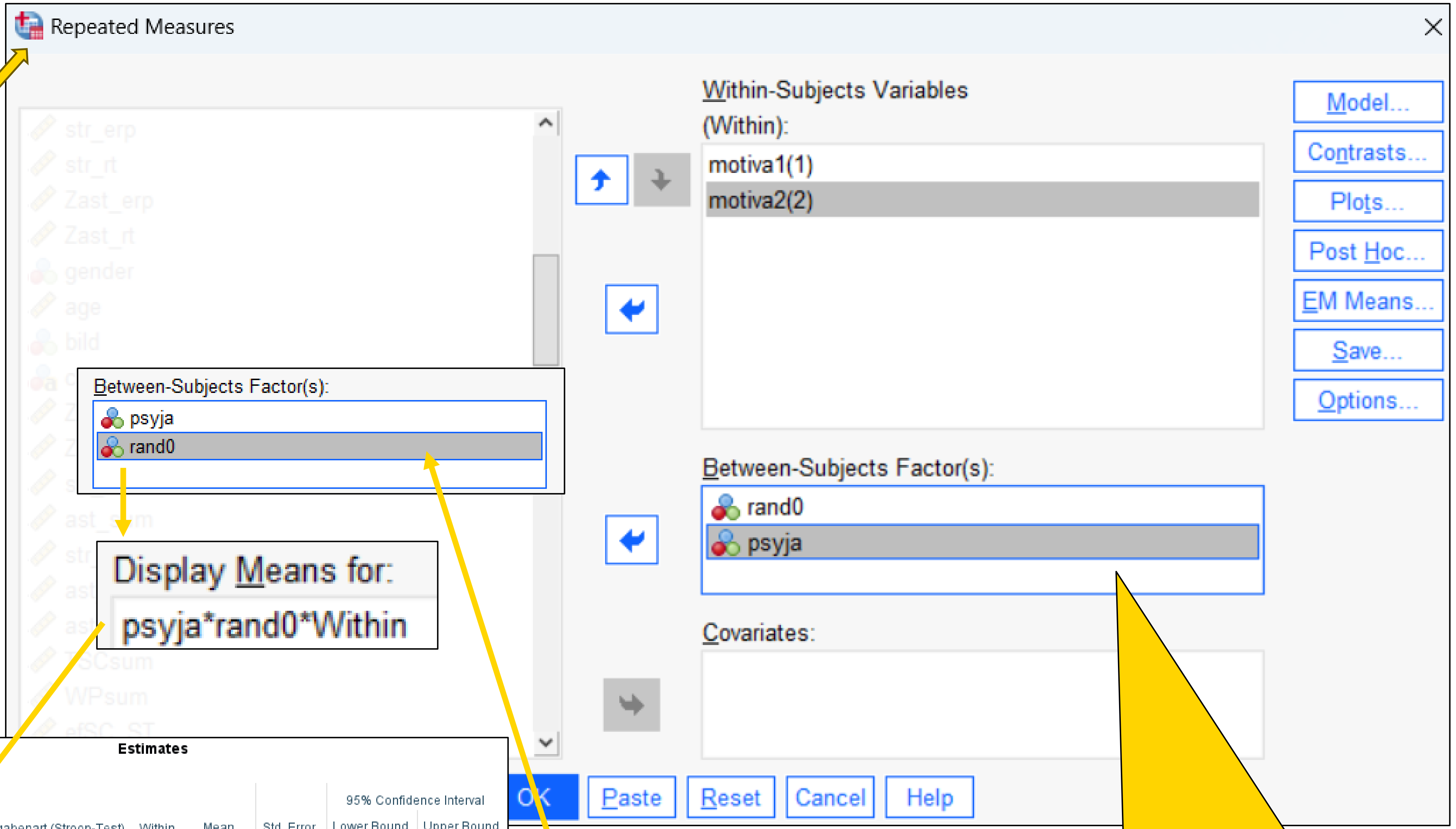
- The procedure in SPSS for a three-way ANOVA is analogous to that of a two-way ANOVA
- With tree IV there are:
  - 3 Main effects (A, B, C)
  - 3 Interaction 1.Order (A*B, A*C, B*C)
  - 1 Interaction 2. Order (A*B*C)

Start interpretation with the highest-order effect

- A three-way interaction means that the effect of one factor on the dependent variable depends on the levels of two other factors, or that the interaction between two factors depends on the levels of a third factor
  - The interpretation of first-order interactions should also be made in the context of higher-order interactions (→ significance)
  - The interpretation of main effects should also only be made in the context of interactions (→ significance)
- ***Here, an exemplary process is roughly outlined, but without a results table***
  → ***Please contact me/us if you have questions about your own analyses***

ANOVA – 3-factors (2 betw, 1 with)





Estimates

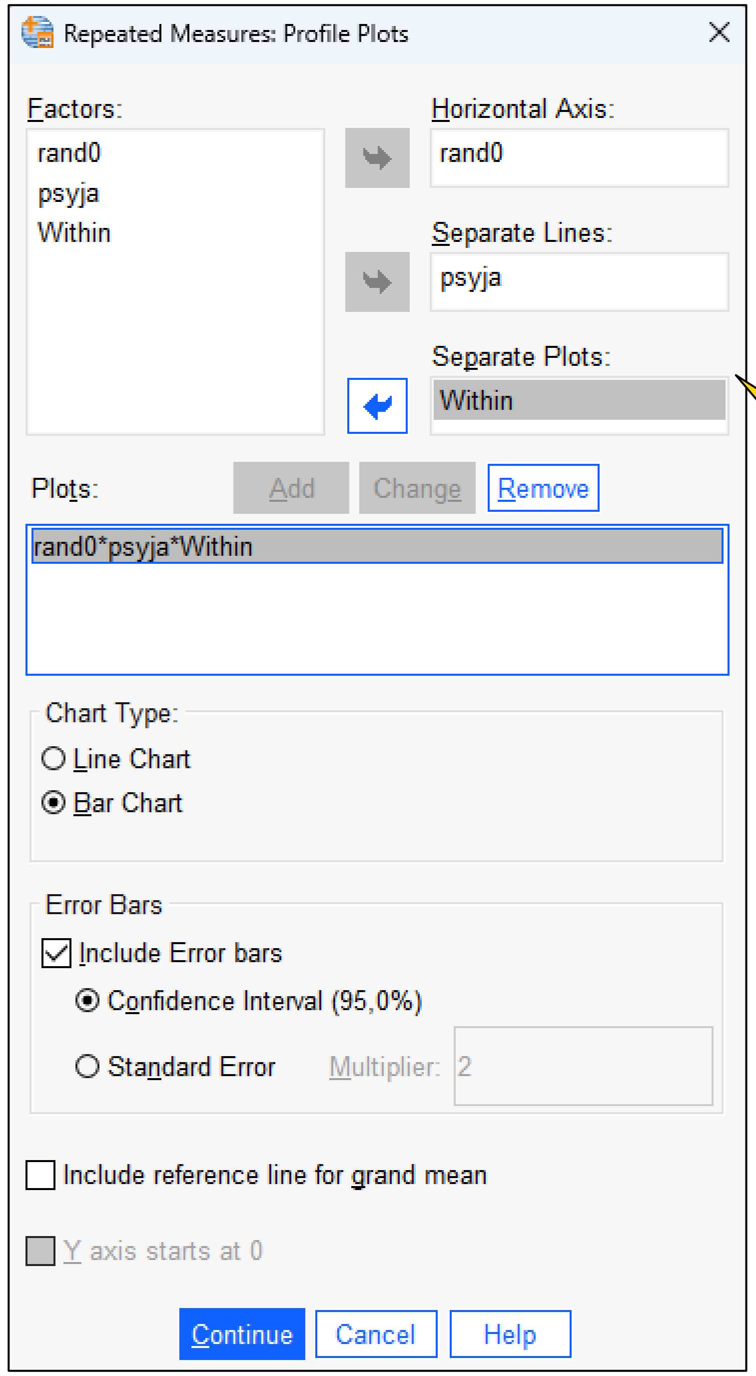
Measure: MEASURE_1

Studieren Sie aktuell Psychologie (Bachelor oder Master)?	Aufgabenart (Stroop-Test)	Within	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Ja	Kontrollbedingung	1	3,316	,344	2,629	4,003
		2	3,895	,249	3,397	4,392
	Depletion-Bedingung	1	3,923	,416	3,093	4,753
		2	4,385	,301	3,783	4,986
Nein	Kontrollbedingung	1	3,632	,344	2,945	4,318
		2	3,842	,249	3,345	4,339
	Depletion-Bedingung	1	4,053	,344	3,366	4,739
		2	4,368	,249	3,871	4,866

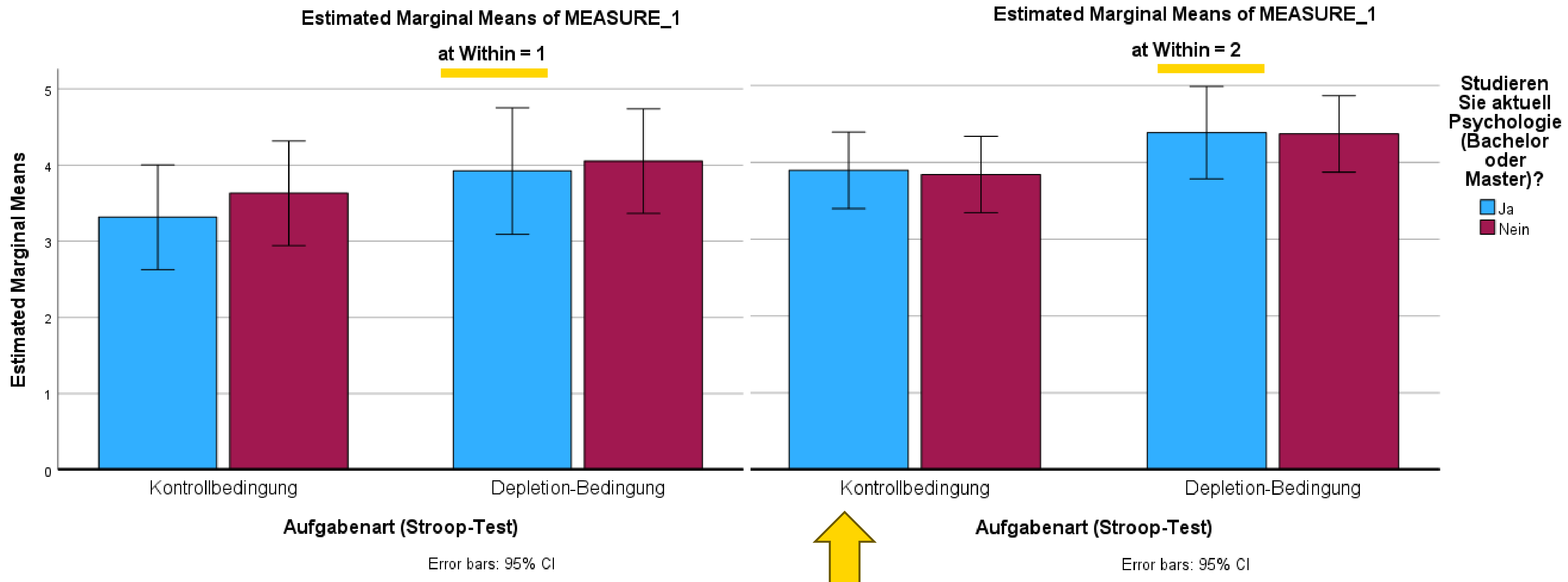
Estimated Marginal Means are displayed in the order in which the variables appear here → See comparison table on the next slide!

ANOVA – 3-factors (2 betw, 1 with)

- For the interpretation of the results, a graphical representation of the three-way interaction is useful (if available!)



Two diagrams for the two levels of the within-subjects factor



Here, the order of the between-subjects factors is reversed compared to the previous slides

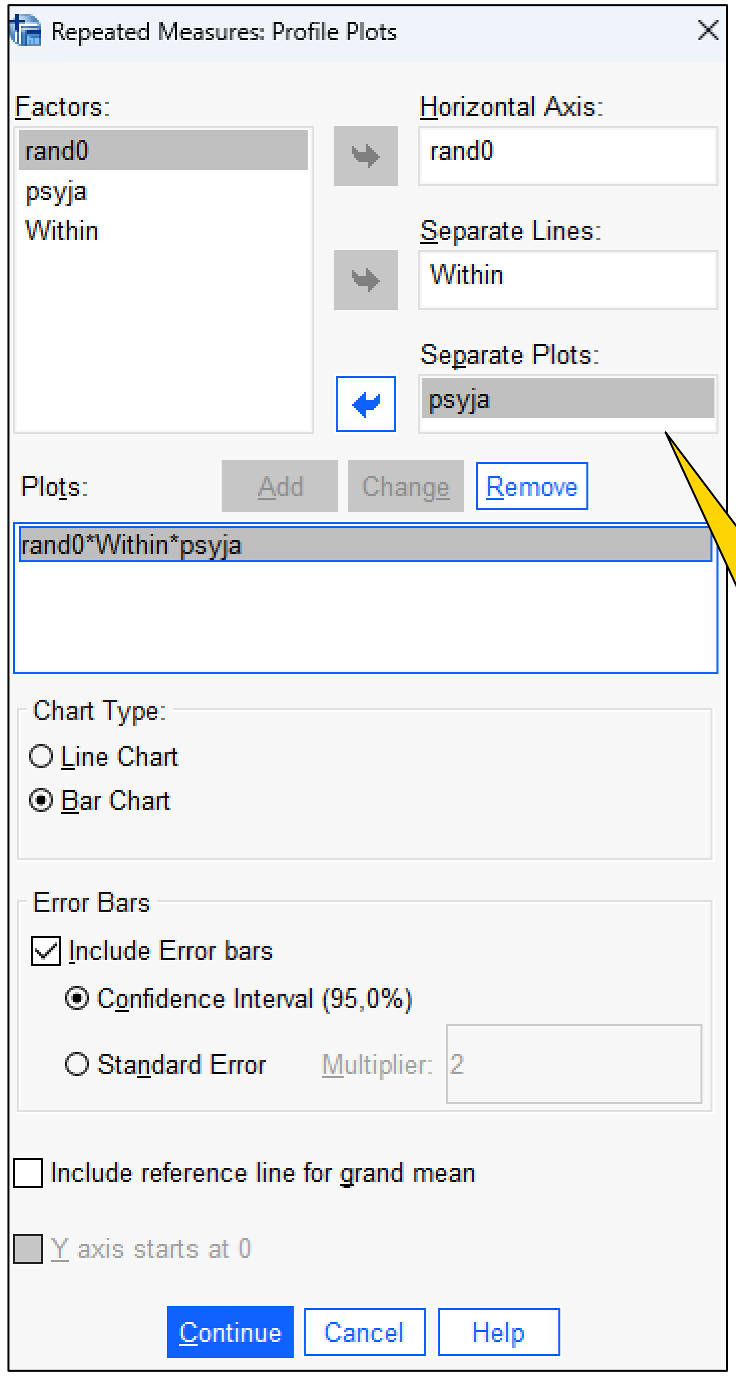
Estimates

Measure: MEASURE_1

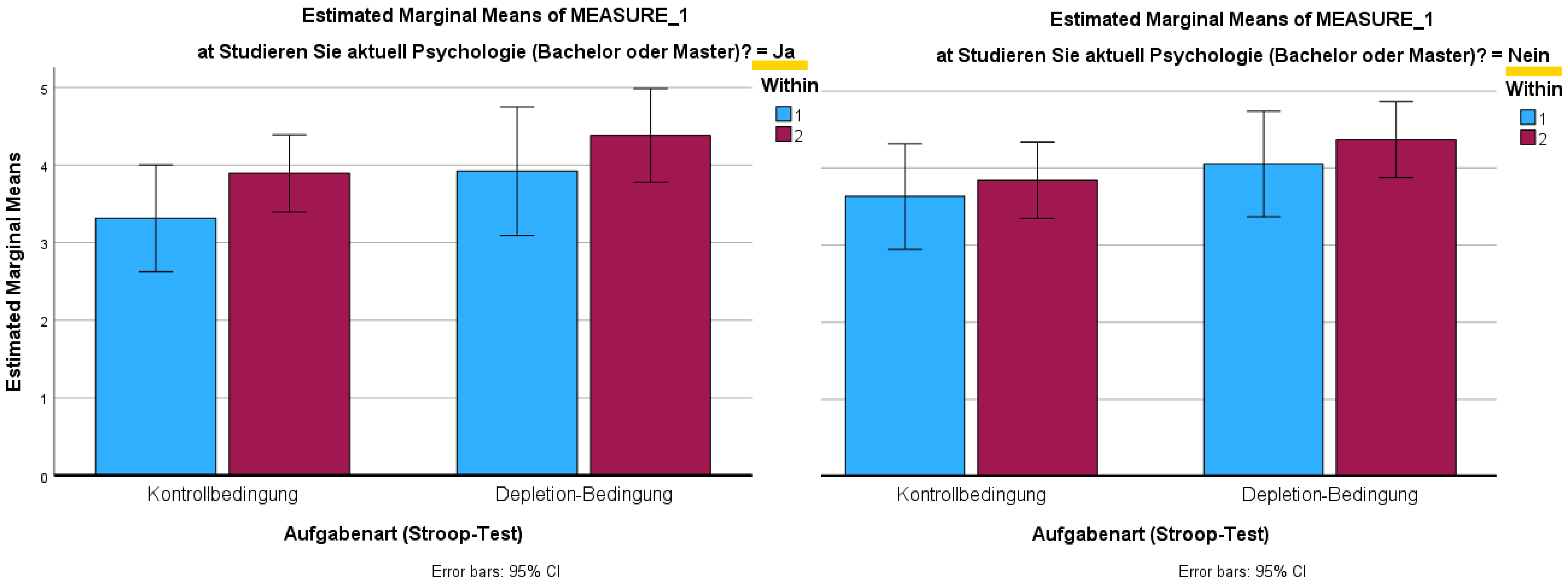
Aufgabenart (Stroop-Test)	Studieren Sie aktuell Psychologie (Bachelor oder Master)?	Within	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Kontrollbedingung	Ja	1	3,316	,344	2,629	4,003
		2	3,895	,249	3,397	4,392
	Nein	1	3,632	,344	2,945	4,318
		2	3,842	,249	3,345	4,339
Depletion-Bedingung	Ja	1	3,923	,416	3,093	4,753
		2	4,385	,301	3,783	4,986
	Nein	1	4,053	,344	3,366	4,739
		2	4,368	,249	3,871	4,866

ANOVA – 3-factors (2 betw, 1 with)

- Für die Interpretation der Ergebnisse, ist die grafische Darstellung der dreifachen Interaktion sinnvoll (falls vorhanden!)



two diagrams for the two levels of the Between-factor of interest



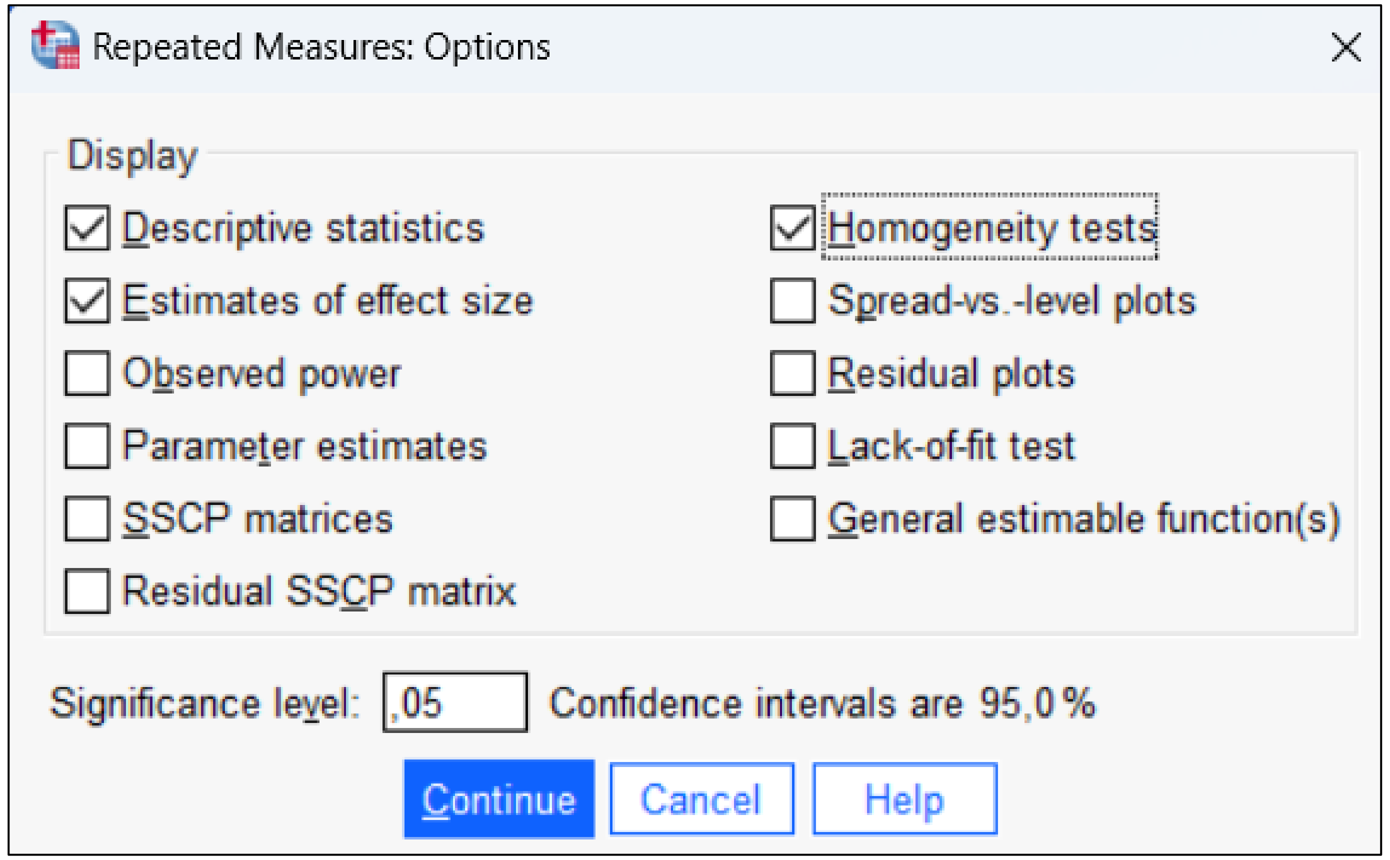
Estimates

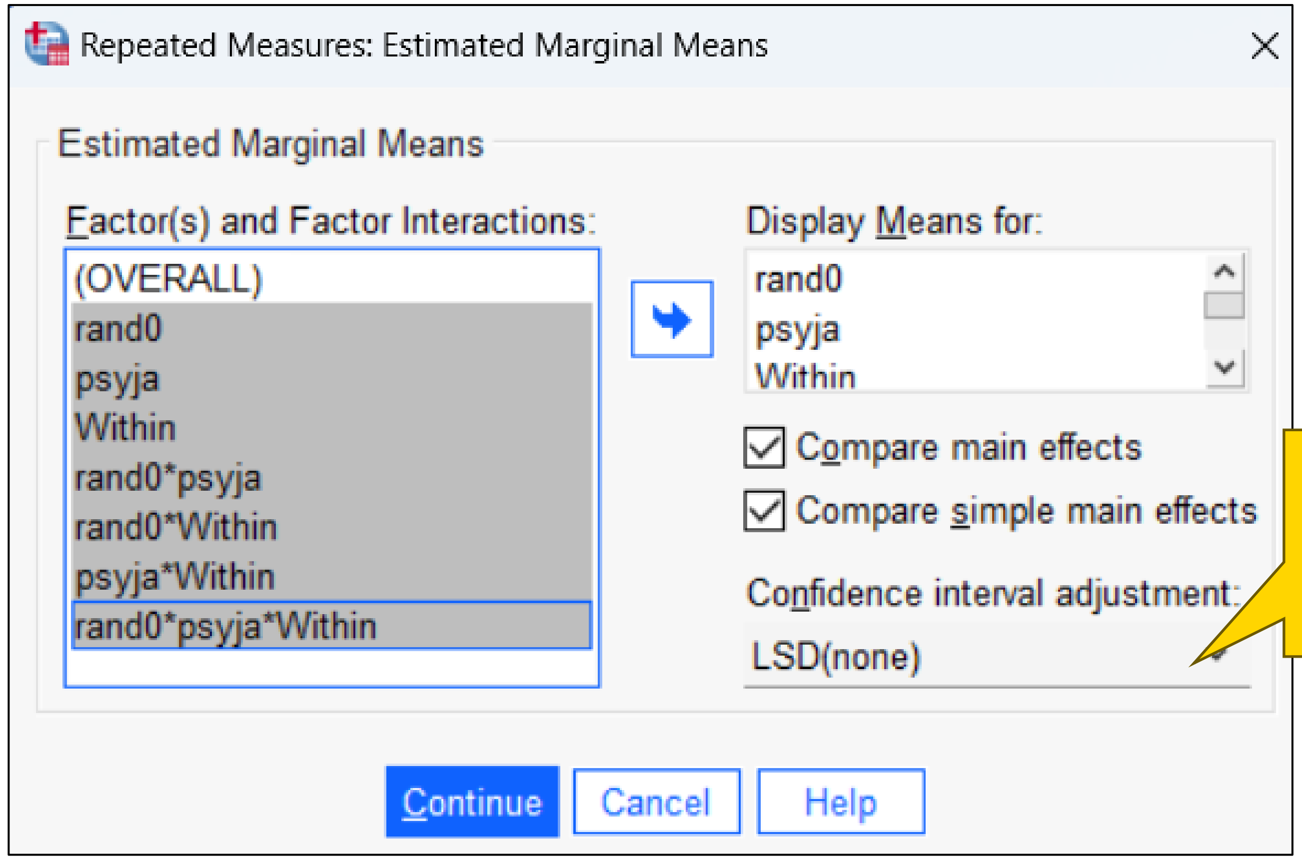
Measure: MEASURE_1

Aufgabenart (Stroop-Test)	Studieren Sie aktuell Psychologie (Bachelor oder Master)?	Within	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Kontrollbedingung	Ja	1	3,316	,344	2,629	4,003
		2	3,895	,249	3,397	4,392
	Nein	1	3,632	,344	2,945	4,318
		2	3,842	,249	3,345	4,339
Depletion-Bedingung	Ja	1	3,923	,416	3,093	4,753
		2	4,385	,301	3,783	4,986
	Nein	1	4,053	,344	3,366	4,739
		2	4,368	,249	3,871	4,866

ANOVA – 3-factors (2 betw, 1 with)

Most options are selected in the same way as for the two-factor mixed ANOVA





Better to use Bonferroni!

ANOVA – 3-factors – additional links

- https://www.researchgate.net/profile/Barbara-Tabachnick/publication/259465542_Experimental_Designs_Using_ANOVA/links/5e6bb05f92851c6ba70085db/Experimental-Designs-Using-ANOVA.pdf
  - Starting at page 358
  - Based on SPSS!

- https://labs.la.utexas.edu/gilden/files/2016/05/Statistics-Text.pdf
  - Starting at page 483
  - Shows all possible combinations of a three-way ANOVA

(Multiple) lineare Regression
- Basic idea

Basic idea - linear Regression 1/3

- Regression:
- Values of a variable Y are estimated by the values of another variable X_i
  - Linear relationship between X & Y → *„the higher X the higher Y“*
- **Unconditional Y mean**: overall mean, does not depend on X
  - e.g. ∅-*Performance on the mental rotation test for all individuals*
- **Conditional Means:** all means (of Y) with identical values in X depend on X
  - e.g. ∅-*Performance on the mental rotation test of women*
- Conditional means thus provide **additional information that reduces the prediction error**!
  - **Less variation within the** (gender) **groups** than in the overall group
  - → Sum of $\text{squares}_{\text{within the Group}}$ < Sum of $\text{squares}_{\text{overall group}}$
- The principle can be applied to metric X-variables!
  - Theoretically we have to handle ∞ mean values
  - In practice, there are as many conditional means as there are levels of X
  - **The conditional Y mean for identical X values is closer to the actual Y values than the overall mean**

Basic idea - linear Regression 2/3

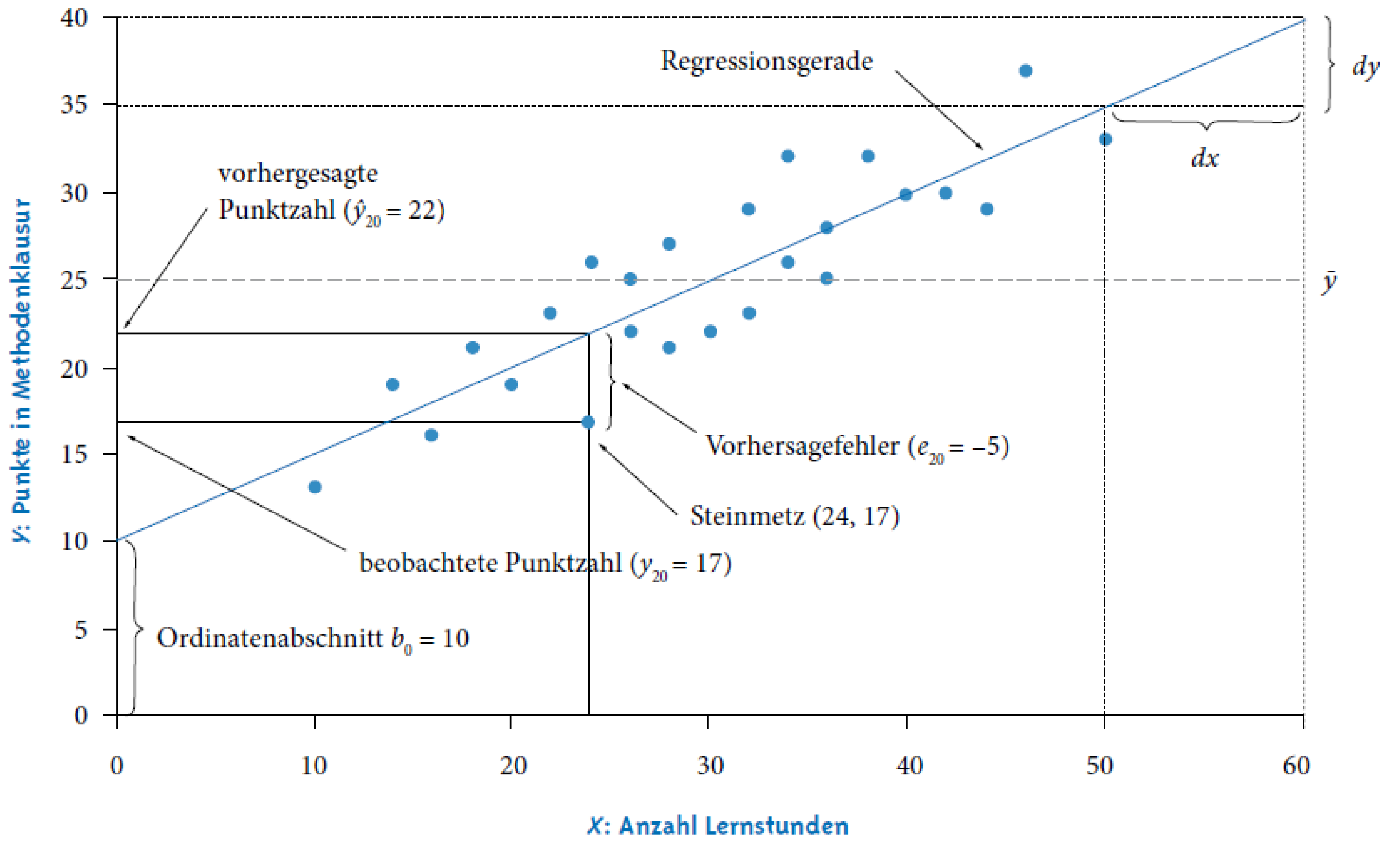
- Ideal case: all conditional means lie on a straight line = **regression line**
  - Even if the underlying linear regression model holds in a population, there are (usually) **deviations between the observed (y) values and the predicted (ŷ) values** in samples

- The sum of the squared differences between y and ŷ represents a criterion for prediction→ **ordinary least squares (OLS)**

"Less deviation = better prediction!"The goal is to minimize the residuals!
The regression line should be placed in the scatter of points in such a way that the sum of the squared deviations is minimalized

$$\sum (y - \hat{y})^2 \rightarrow min$$

Graphic from Eid et al. (201

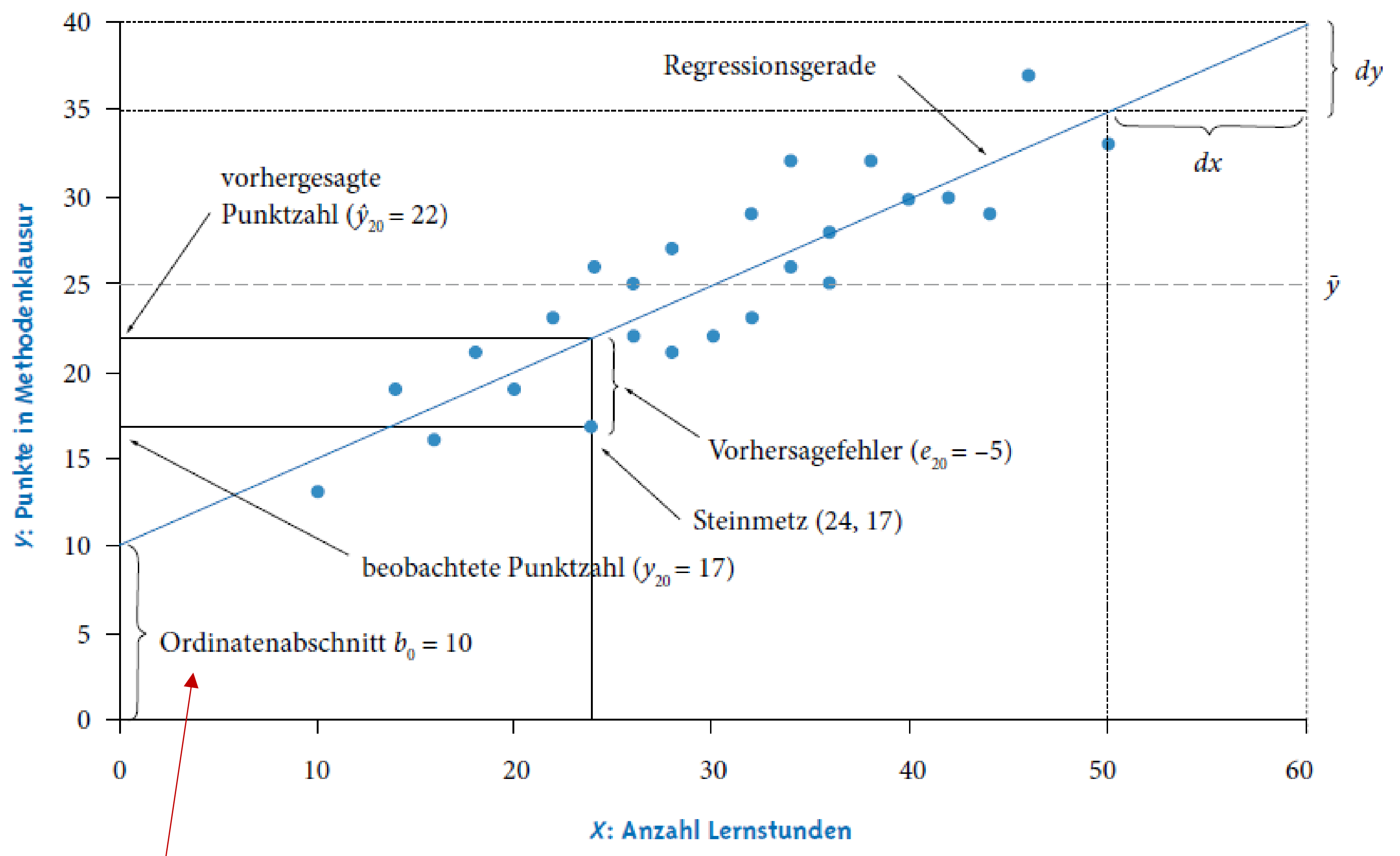


Basic idea - linear Regression 3/3

- Mathematical representation of the linear relationship between **y and ŷ**:

$$\hat{y} = b_0 + b_1 * x$$

The quotient *dy/dx* indicates the slope of the regression line= the **regression weight b_1**



To b_0 and b_1 one conceptually arrives as follows:
The least squares criterion is best fulfilled by the line whose slope (= b_1) is related to the correlation between x and y in the following way:

$$b_1 = r \cdot \frac{s_y}{s_x} = \frac{s_{xy}}{s^2 x}$$

Thus, the regression weight of two z-standardized variables is identical to the correlation coefficient!

b_0 = *the* ŷ-value, that the regression functions assign a value of 0 to an x-value

Residuals

- **Residuals/Error:** arise when the regression model does not fully or correctly represent the relationship between the independent and dependent variables
  - = Difference between the observed values of a dependent variable and the values predicted (estimated) by a statistical model
  - are a **key element in diagnosing and evaluating the fit and effectiveness of statistical models**, as they provide insight into how well the model describes the data
  - Expected value (mean) = 0, sometimes correlated with each other, sometimes with heterogeneous variance (hence often standardized/studentized)

<u>This type of residuals should not be confused with:</u>

- **Not systemetical „Residuals"/random errors/disturbances:**
  - reflect the natural variability in the data that is not explained by the model
  - Include unobserved factors and measurement errors
  - Expected value (mean) = 0, uncorrelated, homogeneous variance

regression residual

- Is the difference of y - ŷ (= *e*)
- indicate the prediction error
  - *e = 0* → measured value lays on the regression line
- **regression residual** = the remaining "residue" that could not be predicted after the best possible prediction of the y-values
- The y-values are additively composed of the ŷ-values and the error/residual

$$\hat{y} = b_0 + b_1 * x + \boldsymbol{e}$$

$$y = \hat{y} + e$$

- **For residuals, the following holds:**
  1. The sum of all regression residuals = 0
  2. The sum of all squared regression residuals = minimal
  3. The correlation between x and e = 0
  4. The correlation between ŷ and e = 0

1. "Errors average out„
2. = OLS
3. Residual values represent the part of Y that does not correlate with X
4. x and ŷ correlate perfectly with each other (because all ŷ values lie on the regression line)

Sum of Squares / Variance Decomposition

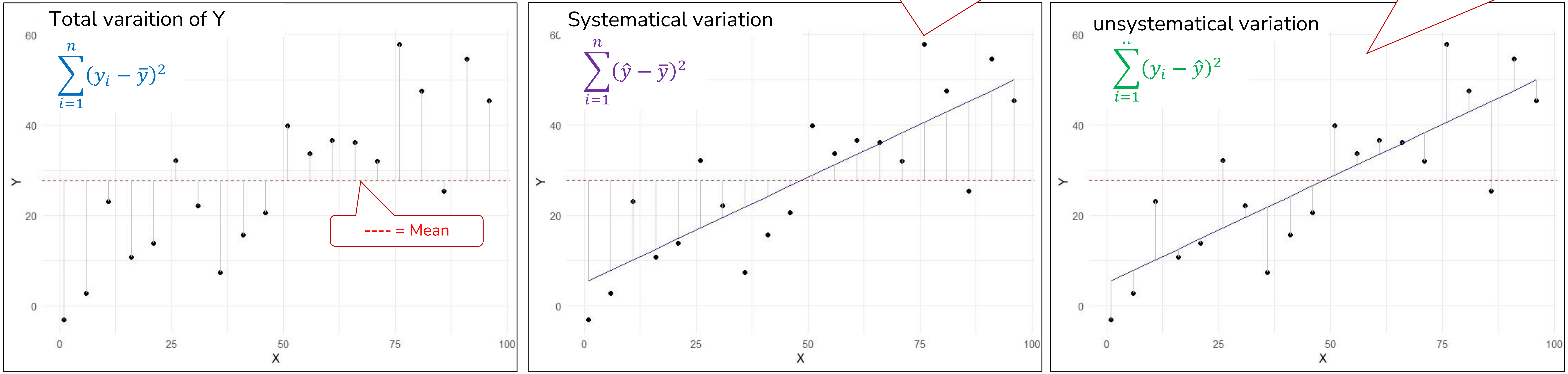
- In the context of regression analysis, the sum-of-squares decomposition is used to split the total variation (total sum of squares →TSS) of the DV into 2 parts:
  1. **systematic variation** (regression sum of squares → SSR)
  2. **unsystematic variation** (Error sum of squares → SSE)

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y} - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y})^2$$

- **TSS = SSR + SSE**
  - **To obtain the variance, we divide the sums of squares by *n* !**

deviations of predicted values to the mean

Abweichung der beobachteten Werte von den vorhergesagten Werten



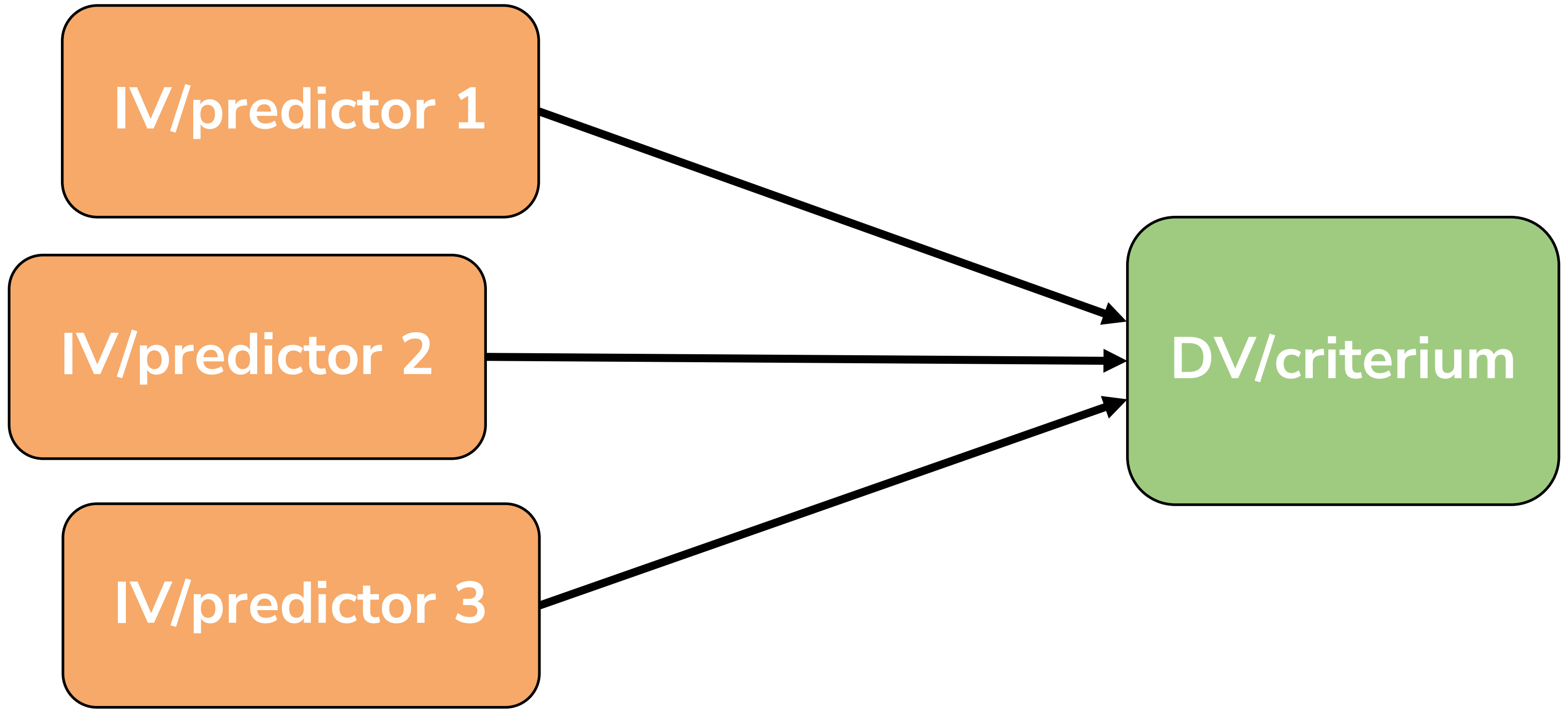
Baisc idea - MLR

- Multiple linear regression examines or models the relationship between an DV and ≥ 2 IVs

- The basic idea is to create a **linear equation that describes the DV as a function of the IVs.**
  - Behavior & experience is usually multi-determined → analysis of bivariate relationships is generally not sufficient
  - MLR extends the concept of simple linear regression and determines:

    1) The relative predictive contribution of each individual predictor (IV_1, IV_2 etc.) → **Regression weights** (*b or* β)

    2) The overall correlation of the IVs with the DV → **variance explanation**: R^2



- Based on a predefined number of predictors, we want to predict the values in the criterion.
- A linear relationship between IV and DV means that they have a common information share - one variable can partly "explain" the other→ **however, this applies in principle in both directions, which is why a causal relationship cannot automatically be assumed!**

Regression equation

Error term: represents the deviation of the actual values from the values predicted by the regression model

The **DV** that is to be predicted!

IV/predictor

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ \ldots + \beta_n X_n + \epsilon$$

- Also called „a“, „Intercept“ or „Constant“ → **y-intercept** of the regression line (the value of Y when all X's equal 0)
- In the analysis of standardized values, this has the value “0”

- Regression weights that indicate the influence of the respective UV on the AV
- **standardized β**: indicates by how many standard deviation units Y changes if X changes by 1 SD and all other predictors are held constant
  (***→ better compaibability***)
- **unstandardized *b***: indicates by how many units Y changes if X changes by 1 unit and all other predictors are held constant
  (**→ easier interpretation)**
- **Indicates the slope of the regression line**

Since we are here in the context of inferential statistics and are estimating values, the formula can also be presented in this way:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n$$

While the regression weight in the context of simple linear regression can be calculatedby $b = \frac{cov(xy)}{s_x^2}$ *he formula in the context of multiple linear regression is too complex to represent here!*

Multiple linear Regression

- Execution in SPSS

Procedure in SPSS

1. Model definition (which DV, which IVs)

2. (For the most part, the prerequisites of the MLR can only be checked AFTER implementation; corresponding options are marked during implementation)

3. Check for model quality (Globale check → R^2, *F*-Statistics)

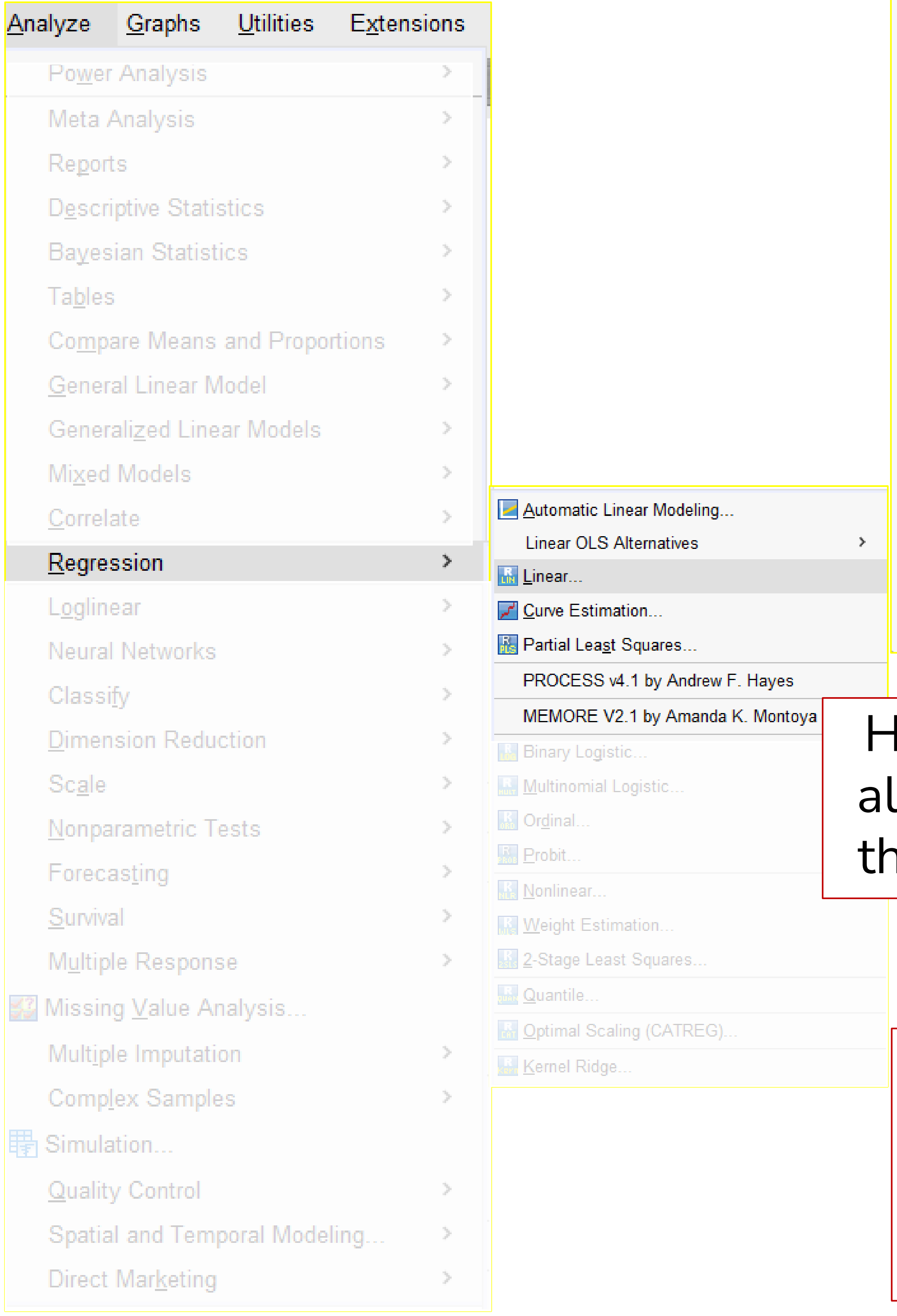
4. Check of the regression weights (Lokal check → *b*, β, *t*-values/significance)

5. Testing the model assumptions

6. Report and visualize results

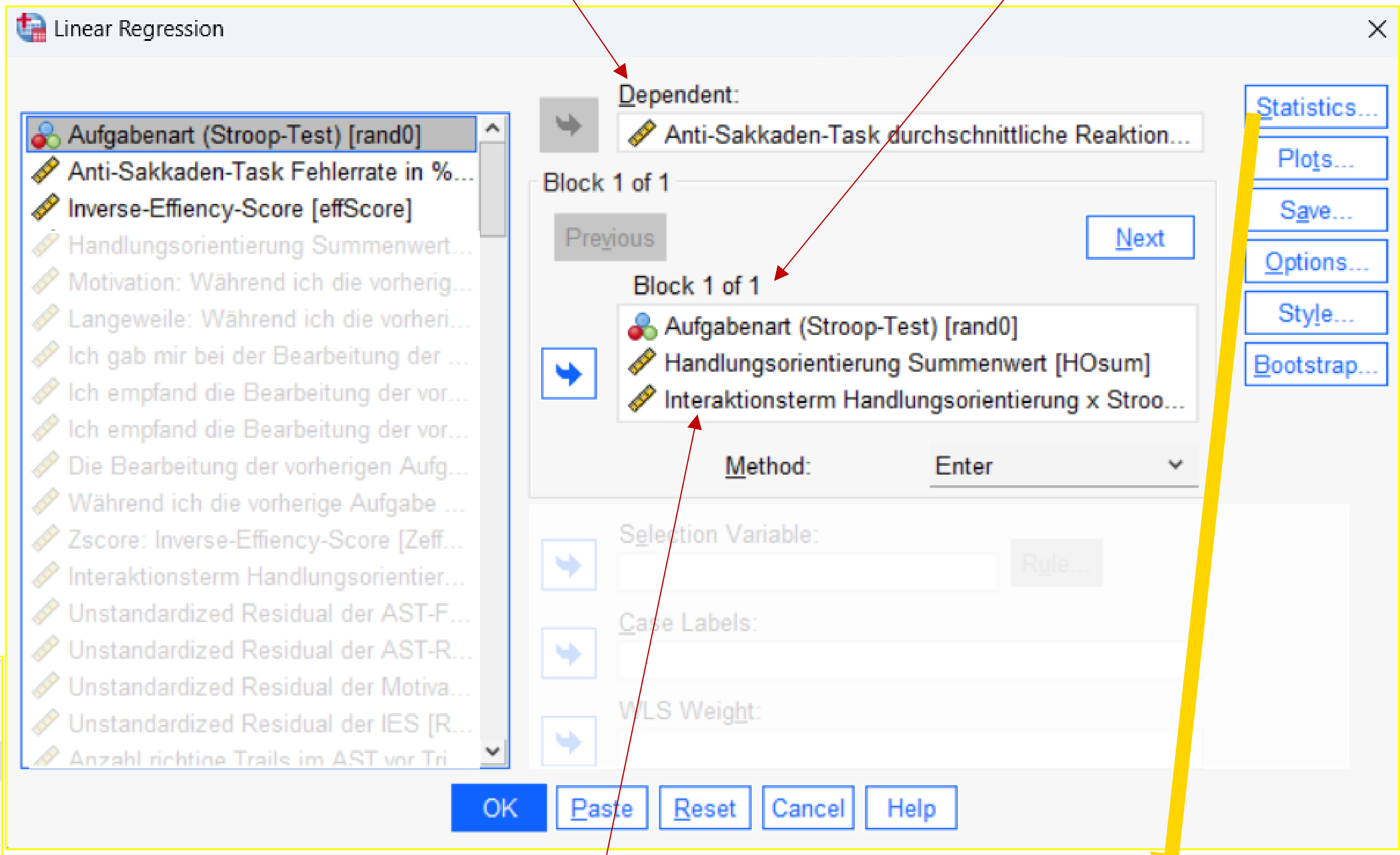
First, the implementation of a multiple linear regression with SPSS is described. For moderation analyses and mediation analyses, the additional implementation with PROCESS is described, as the PROCESS-macro provides some practical functions that significantly simplify the formulation of complex regression models. The "classic" way is necessary to check the assumptions.

MLR in SPSS



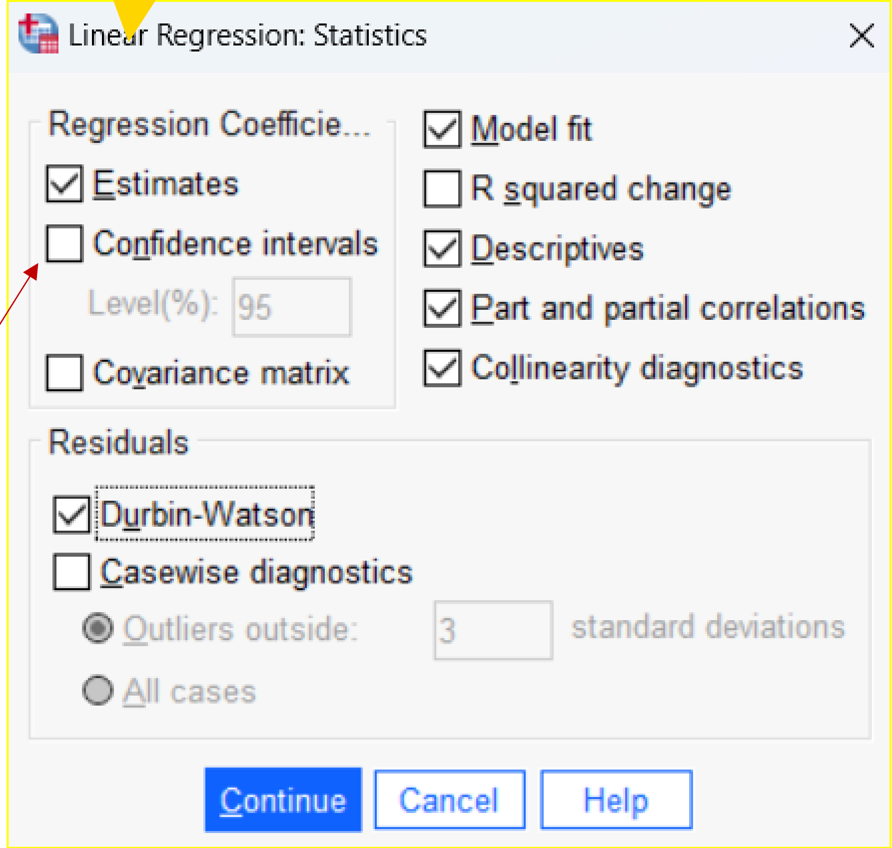
criteria/DV

predictors/IVs



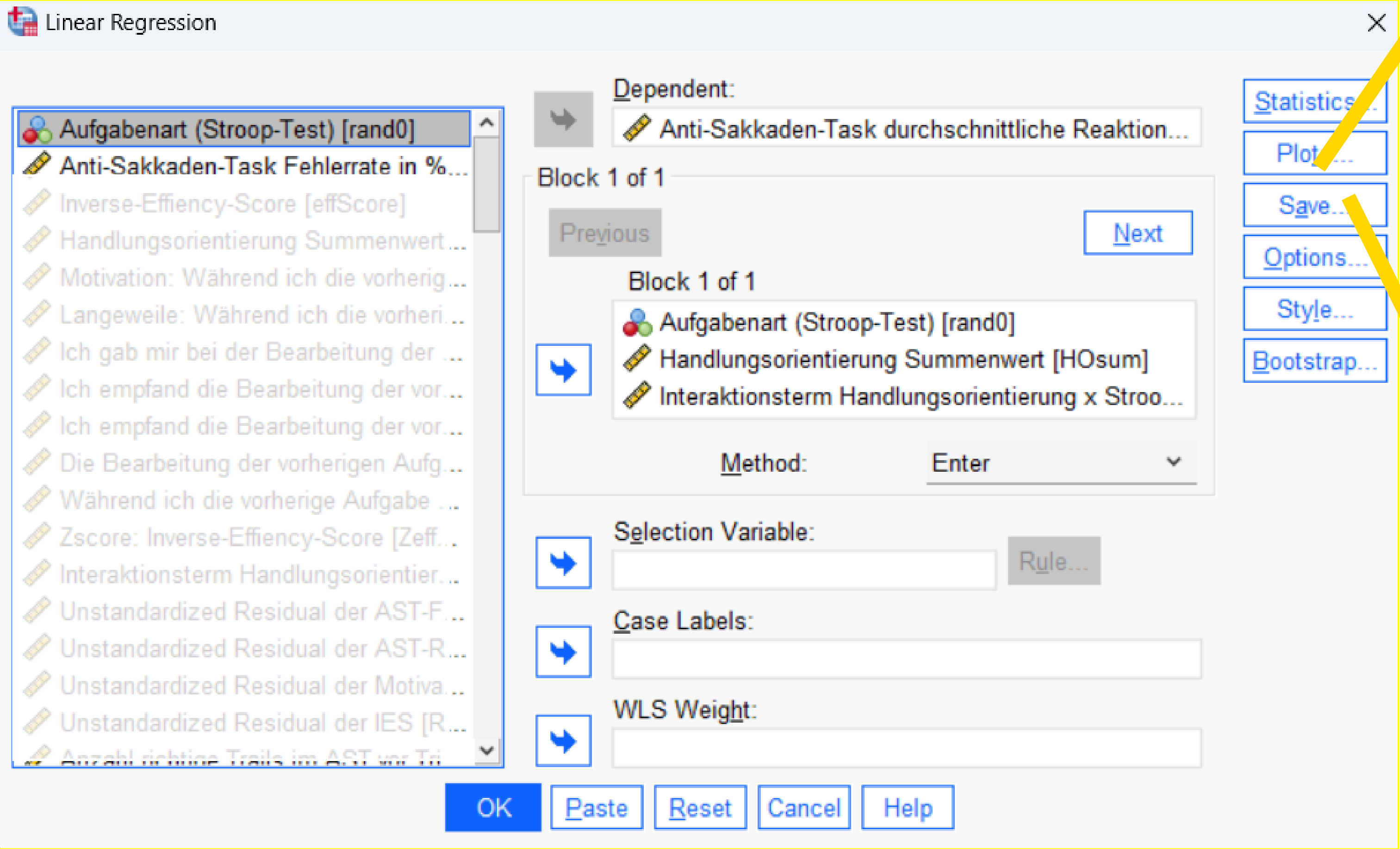
If we were to decide (based on theoretical considerations) to initially include only a certain number of predictors in the regression equation and additional predictors in a second step, the change in R^2 can provide important information!

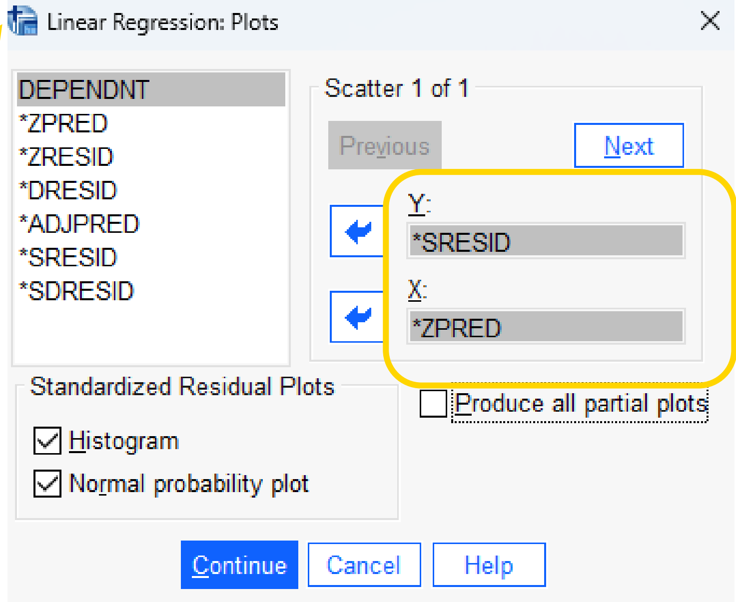
Here, an interaction/moderation is already included in the model. See the section **"Moderation analysis"**



Confidence intervals for the individual regression coefficients are frequently reported and can be requested here. Alternatively, these can be found in the PROCESS output

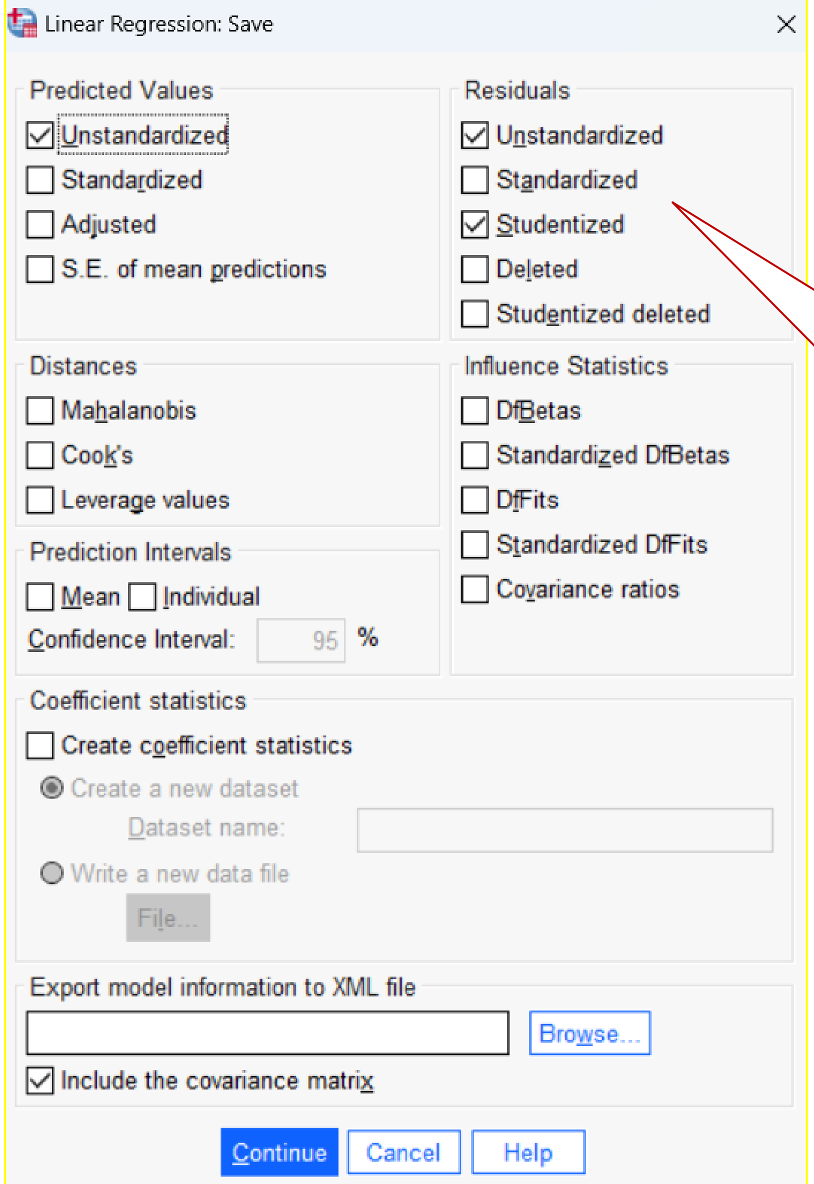
MLR in SPSS





This scatterplot is used to graphically test the assumption of **homoscedasticity**.

Alternatively, the standardized residuals (ZRESID) can be used instead of the studentized residuals (SRESID, Y-axis), whereby the latter (according to Field, 2013) are somewhat more sensitive with regard to individual deviations.



Exemplary syntax command

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT ast_rt
  /METHOD=ENTER rand0 HOsum HOxST
  /SCATTERPLOT=(*SRESID ,*ZPRED)
  /RESIDUALS DURBIN
  /SAVE PRED RESID SRESID.
```

SPSS saves the selected values as new variables. We mainly need the unstandardized residuals to check the assumption of the **normal distribution of the residuals!**

16	RES_1	Numeric	11	5	Unstandardized Residual der AST-Reaktionszeiten

model quality 1/2: R^2

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,274[a]	,075	,033	,189450	1,863

a. Predictors: (Constant), Interaktionsterm Handlungsorientierung x Stroop-Bedingung, Handlungsorientierung Summenwert, Aufgabenart (Stroop-Test)

b. Dependent Variable: Anti-Sakkaden-Task durchschnittliche Reaktionszeit

√

Explanaition on the next slide!

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	,192	3	,064	1,784	,159[b]
	Residual	2,369	66	,036		
	Total	2,561	69			

a. Dependent Variable: Anti-Sakkaden-Task durchschnittliche Reaktionszeit

b. Predictors: (Constant), Interaktionsterm Handlungsorientierung x Stroop-Bedingung, Handlungsorientierung Summenwert, Aufgabenart (Stroop-Test)

R = Correlation between the ŷ-values of the model and the actual y-values or the correlation of the criterion with the predictors contained in the model

R^2 = 0.075 → IF it is a significant overall model, 7.5% of the criterion variance is explained by all predictors together

- R^2 provides a measure of the quality of the fit of the regression function to the empirical data
- Caluculation: **dispersion explained** by the regression function (= sum of the squared deviations of the ŷ-values from ȳ) **divided** by the **total dispersion** (= sum of the squared deviations y-values from ȳ)

$R^2_{adjusted}$: provides a more conservative estimate of the model quality, as it takes into account the number of predictors in the model (as the number of predictors increases R^2 also increases without the predictors having to be significant!)

Std. error of the estimate: The square root of the residual mean squares → average deviation of the ŷ-values from the y-values (how much the data points scatter around the prediction line created by the regression model)

model quality 2/2: *F*-Statistics

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,274[a]	,075	,033	,189450	1,863

a. Predictors: (Constant), Interaktionsterm Handlungsorientierung x Stroop-Bedingung, Handlungsorientierung Summenwert, Aufgabenart (Stroop-Test)

b. Dependent Variable: Anti-Sakkaden-Task durchschnittliche Reaktionszeit

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	,192	3	,064	1,784	,159[b]
	Residual	2,369	66	,036		
	Total	2,561	69			

a. Dependent Variable: Anti-Sakkaden-T[...] durchschnittliche Reaktionszeit

b. Predictors: (Constant), Interaktionsterm [...]lungsorientierung x Stroop-Bedingung, Handlungsorientierung Summenwert, [...]rt (Stroop-Test)

- Using the ANOVA table, we check whether the predictors as a whole make a significant contribution to the prediction of the criterion

- The F-statistic provides the result from the **ratio** of the **explained dispersion** (= sum of the squared deviations of the $\hat{y}$-values from $\bar{y}$) to the **unexplained dispersion** (= sum of the squared deviations of the y-values from the $\hat{y}$-values)

We refer to this table when reporting the results, as well as to the model summary, as we can read the *F-value*, the degrees of freedom and the *p-value* here:

$R^2 = .08$, $F(3, 66) = 1.78$, $p = .159$

CAUTION! If, as in this case, the overall model is not significant, the local check of the regression coefficients is logically superfluous.For the following slides, we will focus on the local model control, assuming that the overall model is significant

local check - regression weights

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	,747	,075		9,951	<,001					
	Aufgabenart (Stroop-Test)	,259	,115	,675	2,249	,028	,117	,267	,266	,155	6,436
	Handlungsorientierung Summenwert	,009	,010	,135	,902	,370	-,064	,110	,107	,625	1,599
	Interaktionsterm Handlungsorientierung x Stroop-Bedingung	-,034	,017	-,621	-2,049	,044	-,007	-,245	-,243	,153	6,548

a. Dependent Variable: Anti-Sakkaden-Task durchschnittliche Reaktionszeit

See assumptions → multicollinearity

The local check is carried out using the:

- **unstandardized regression coefficients** → by how many units does the DV change if the respective variable changes by 1 unit and the other variables are held constant
- **corresponding *t*-tests** → significance of the regression coefficients

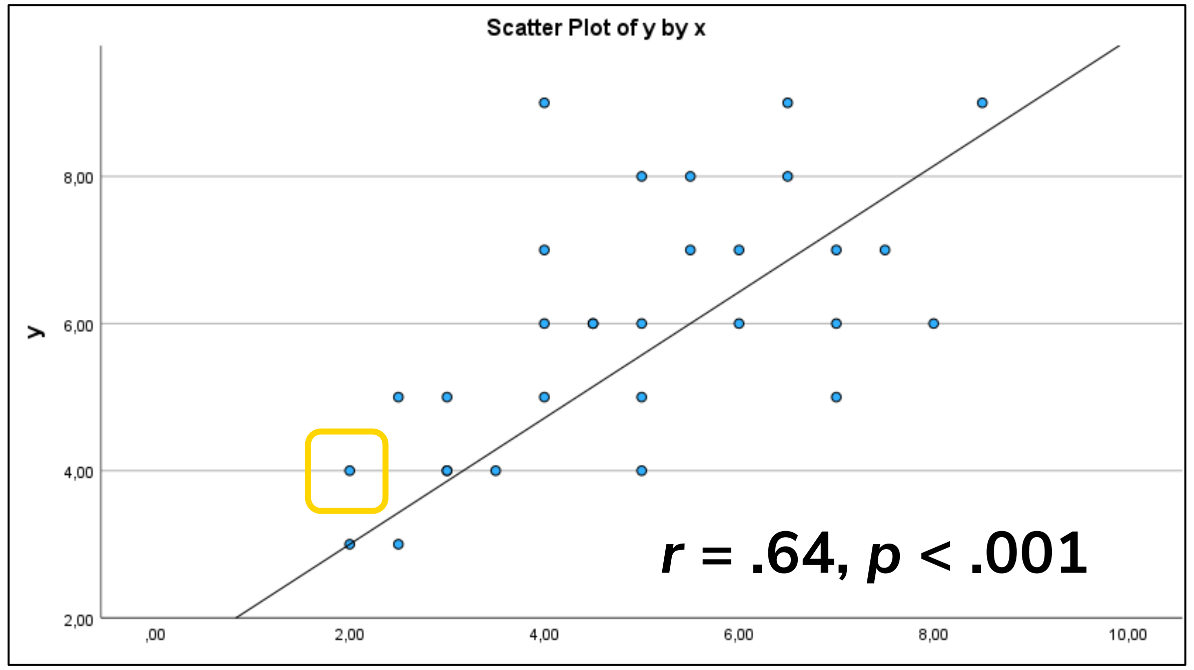
e.g.: $b = -.034$, $t(68) = -2.05$ $p = .044$

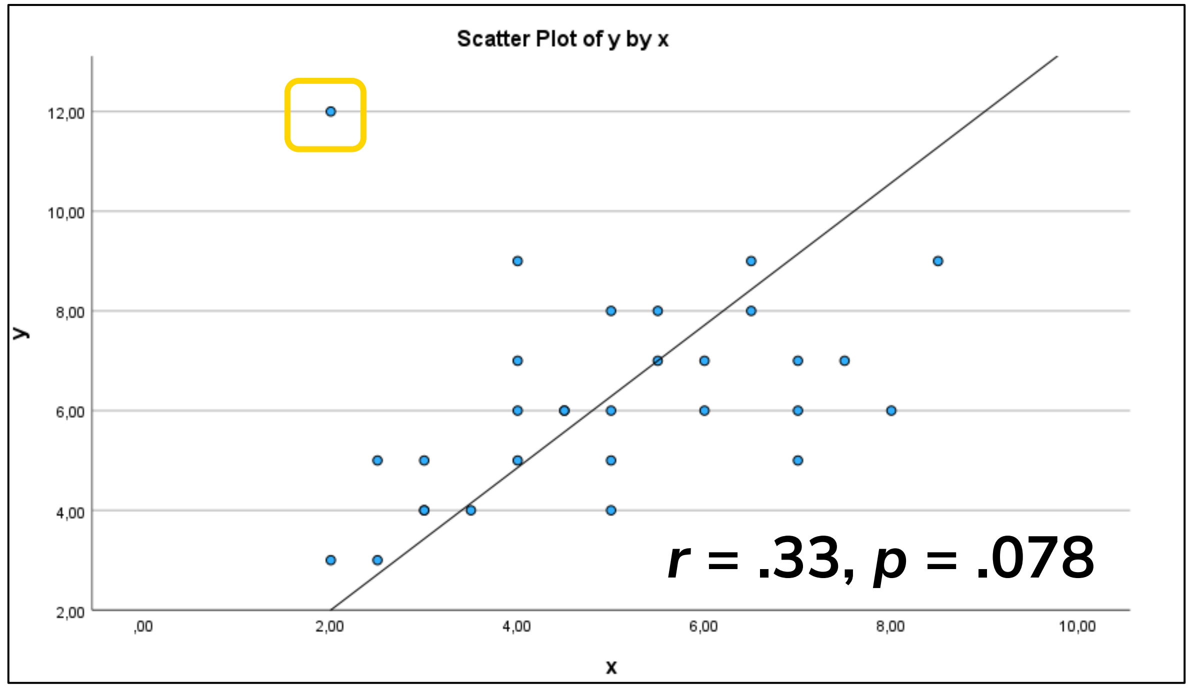
Multiple linear Regression

- Assumptions

No outliers in the data

- Like all statistical procedures based on correlative/regression analysis methods, **outliers can strongly distort the relationship between variables**
  - especially with small samples! (here: n = 30; only 1 data point was changed)



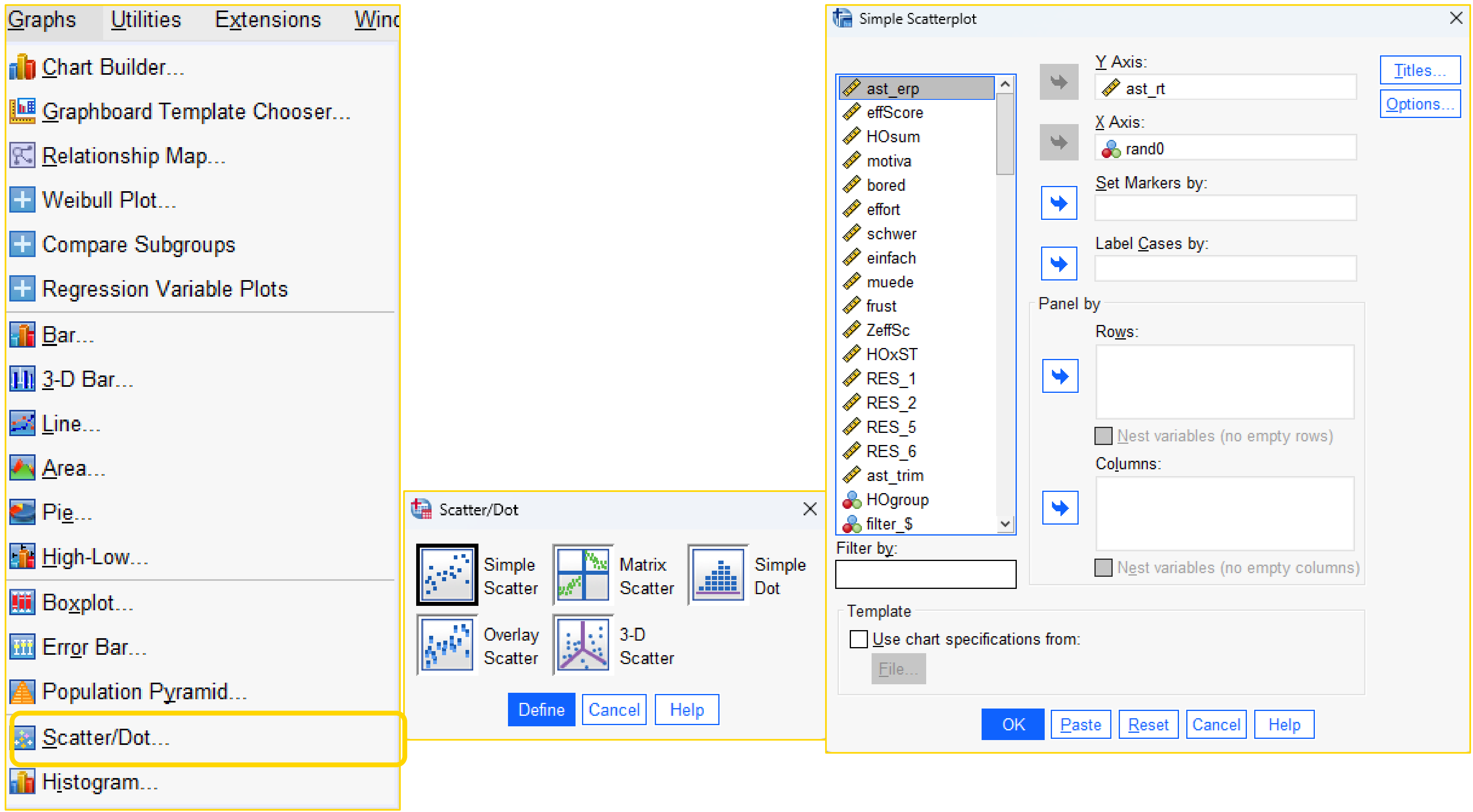


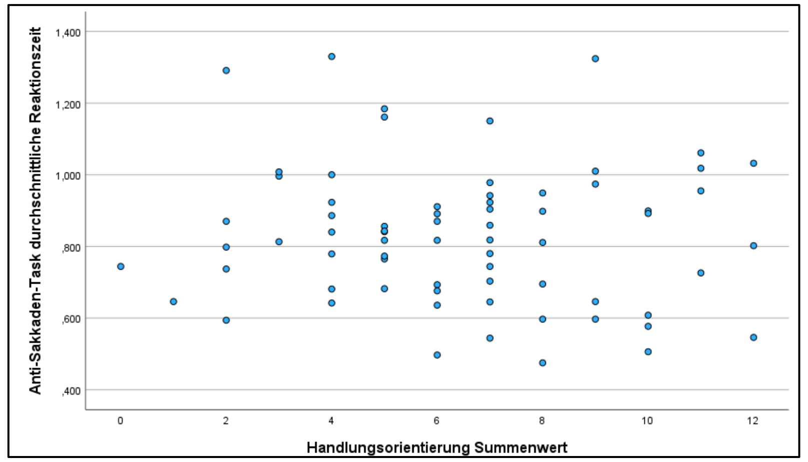
- **Verification**:
  - Scatterplot
  - Deskriptive Statistics (save and inspect z-standardized values)
  - Box-plots (*Analyze* → *Descriptive Statistics* → *Explore*)
- **Countermeasure**:
  - exclusion of values based on (several) characteristic values that can be requested as part of the regression (*case-by-case diagnosis of “Studentized deleted” residuals, Cook's or Mahalanobis distance, leverage values*)

Linearity between DV and IV(s)

- **Verification:** Scatterplots of the DV <u>against each IV</u>

Alternatively, the inspection of the scatterplot ZPRED (X) & ZRESID (Y), which is also used to analyze homoscedasticity, is recommended by authors such as A. Field (2018)



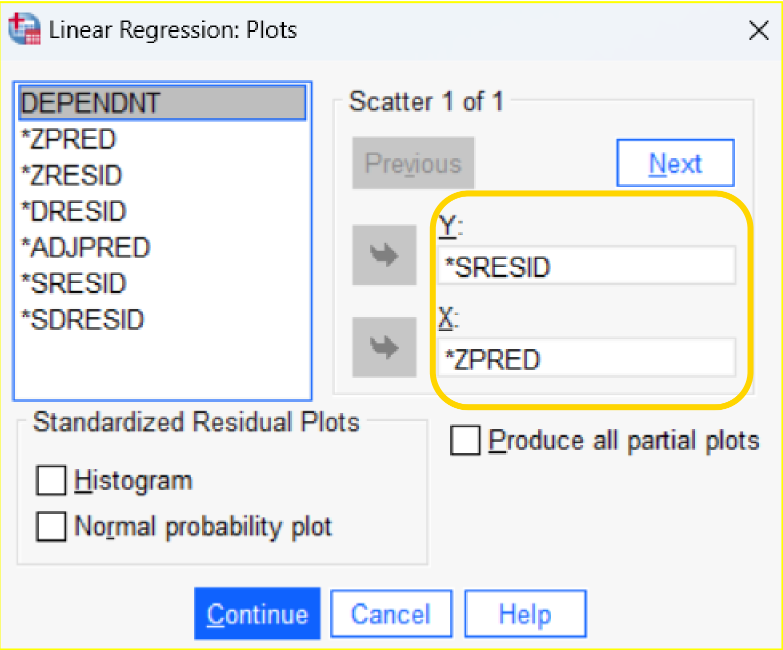


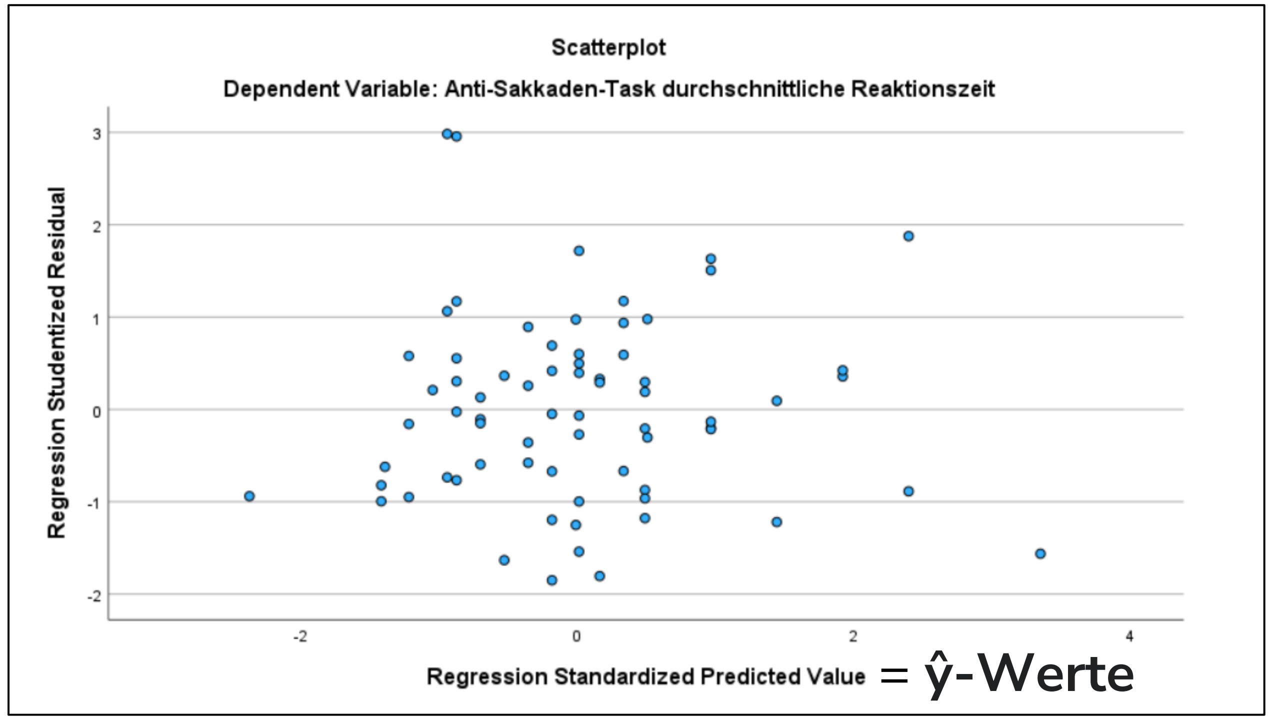
- If the violation is severe, the model parameters can no longer be interpreted meaningfully
  - Even a large sample cannot change this!

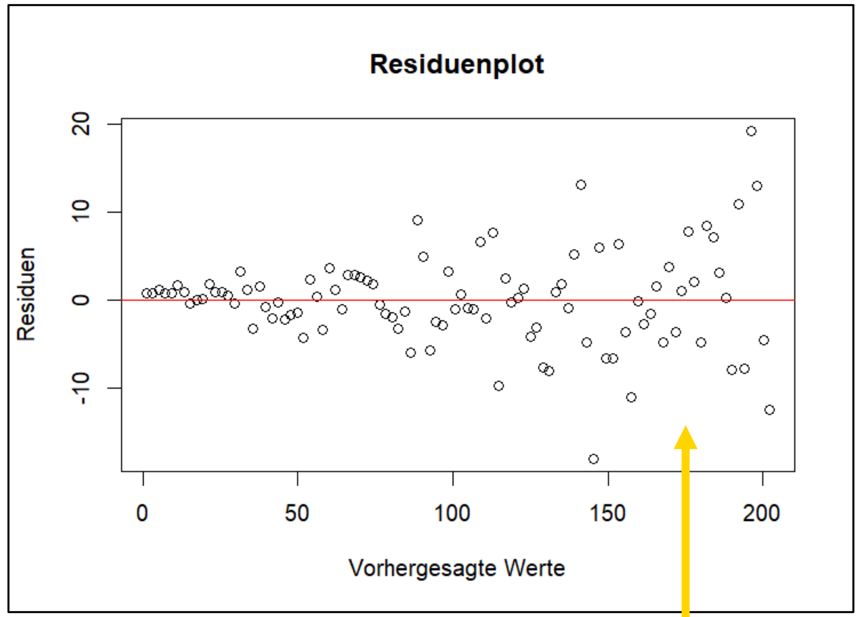
→ **Gegenmaßnahme**: use non-linear regression models

Homoscedasticity of the residuals

- **Verification**: The scatterplot is already included in the regression analysis in SPSS







- **Assumption**: The forecast by the specified model should be similarly precise along the entire DV value range
  - The variance of the residuals (errors/deviations) should be constant along the X-axis (= AV value range)

    good: uniform, "rectangular" pattern | bad: funnel-shaped, triangular pattern (= heteroscedasticity)
- Violation leads to distortion of the estimated standard errors = inaccurate *p*-values or confidence intervals
- → **Gegenmaßnahme**: transform DV/IVs and recalculate model; use PROCESS-macro with heteroscedasticity-consistent standard errors (HC3); bootstrapping

no auto-correlation

- **Verification**: is already included in the regression analysis in SPSS

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,274[a]	,075	,033	,189450	1,863

a. Predictors: (Constant), Interaktionsterm Handlungsorientierung x Stroop-Bedingung, Handlungsorientierung Summenwert, Aufgabenart (Stroop-Test)

b. Dependent Variable: Anti-Sakkaden-Task durchschnittliche Reaktionszeit

The condition is considered fulfilled if the value is between **1.5 and 2.5**

2 = no autocorrelation
0 = perfect positive autocorrelation
4 = perfect negative autocorrelation

- **Assumption**: The residuals are uncorrelated/independent in the population
  - SPSS tests whether a residual correlates with its direct neighbor
  - Particularly relevant for time series analyses or data with repeated measurements!

- Violation leads to inefficient estimation of regression coefficients and standard errors
- → **Countermeasure:** Heteroscedasticity- & autocorrelation-consistent estimators; selection of a suitable method for the data

no multicollinearity

- **Verification**: is already included in the regression analysis in SPSS

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	,747	,075		9,951	<,001					
	Aufgabenart (Stroop-Test)	,259	,115	,675	2,249	,028	,117	,267	,266	,155	6,436
	Handlungsorientierung Summenwert	,009	,010	,135	,902	,370	-,064	,110	,107	,625	1,599
	Interaktionsterm Handlungsorientierung x Stroop-Bedingung	-,034	,017	-,621	-2,049	,044	-,007	-,245	-,243	,153	6,548

a. Dependent Variable: Anti-Sakkaden-Task durchschnittliche Reaktionszeit

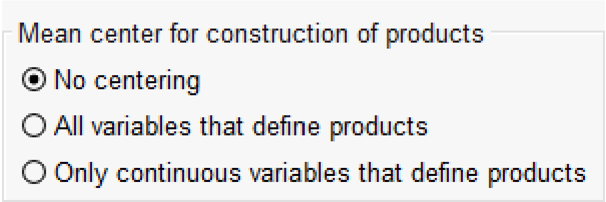
- **VIF ≤ 10** is ok (better < 5)
- Tolerance > .10 is ok

- **Variance-Inflation-Factor (VIF):** multiplicative value → variance of the estimated regression coefficient for a particular IV would be x times as large as it would be if the IV were completely uncorrelated with the other IVs in the model
- **Tolerance**: reciprocal of the VIF (VIF = $\frac{1}{Toleranz}$); value range: 0-1 → high T. = only a small part of the IV-variance can be explained by other IVs

- **Assumption**: The IVs/predictors in the model do not correlate (strongly) with each other à would mean that the value of one IV could be predicted from those of the other IVs:
  - Which IV actually contributes to the elucidation of DV? Redundant IVs?
- Violation leads to inefficient estimation of standard errors and regression coefficients
- → **Countermeasure**: exclude variable(s) from analysis; ***centering***; choice of an alternative regression method

MLR – centering of variables

- Centering is a transformation of the initial values of a variable
  - Value minus mean value→ $x - \bar{x}$
  - easy to select in the PROCESS-macro
  - in SPSS:



1. for each variable that is used for a product term, calculate a new variable that contains the mean value of this variable. e.g. “x1_mean”
2. calculate a centered variable that subtracts “x1_mean” from “x1” etc.
3. form the interaction term with the centered variables and include all variables in the analysis (instead of the original ones)

A detailed explanation can be found at: https://statistikguru.de/spss/moderation/variablen-zentrieren-oder-nicht.html

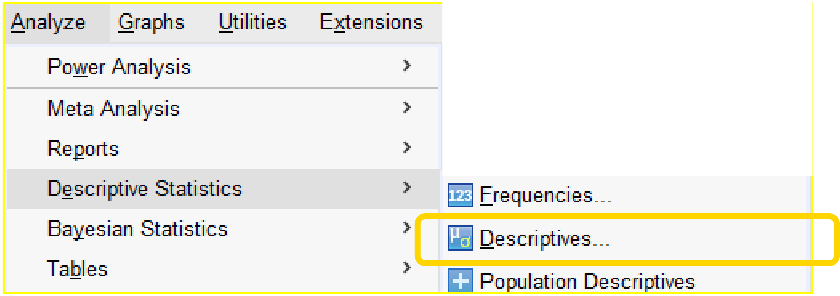
- **Conclusion of the linked website:** If the predictors do not have a natural zero point or the value zero has no meaningful significance, it is recommended to center the variable. centering facilitates the interpretation of the direct effects and does not influence the moderation effects

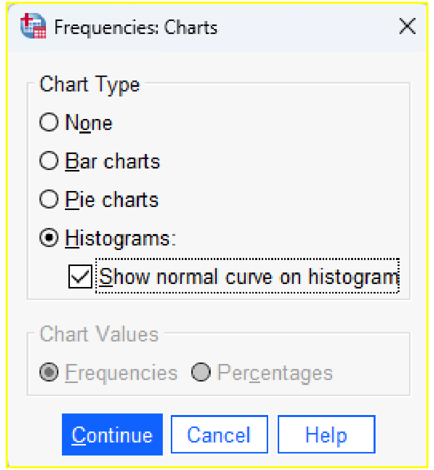
Always discuss the desired procedure with the person supervising your seminar paper/bachelor thesis!

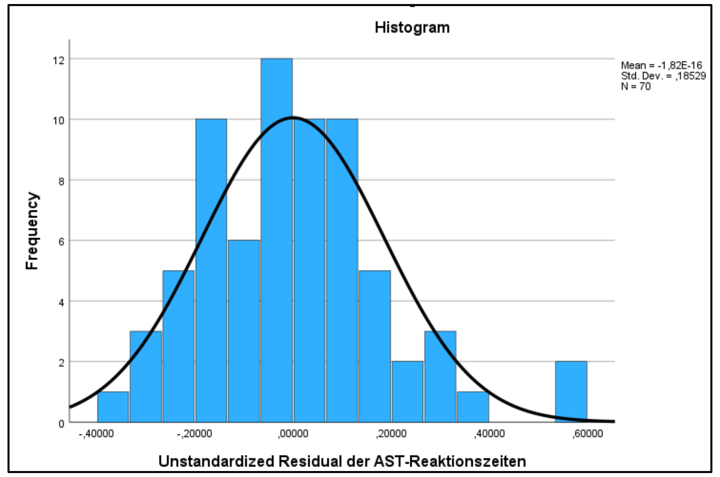
Normal distribution of the residuals 1/2

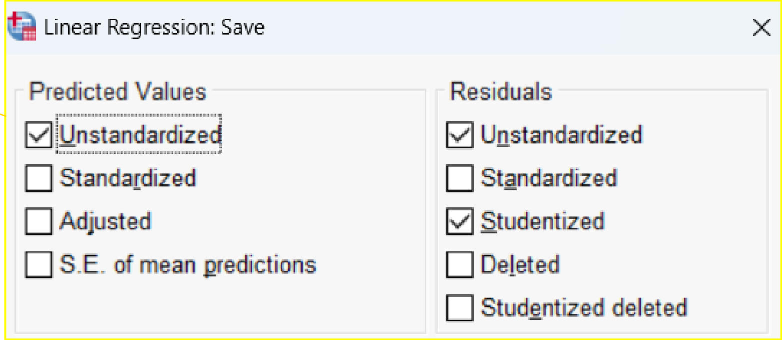
- **Verification 1:** Histogram of the standardized residuals
  - A histogram is requested under FREQUENCIES for the variable previously requested under *SAVE*

16	RES_1	Numeric	11	5	Unstandardized Residual der AST-Reaktionszeiten

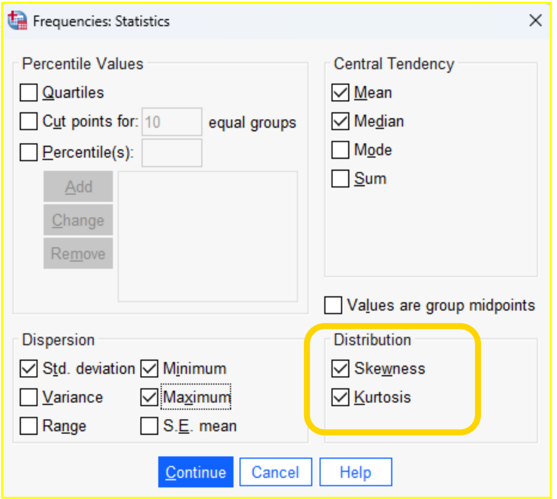








- **Verification 2:** Skewness and kurtosis of the standardized residuals
- under *FREQUENCIES* → *STATISTICS* **skewness & kurtosis** are requested. Both values are ideally **|≤ 1|,** whereby values |≤ 5 are acceptable for the kurtosis



Statistics

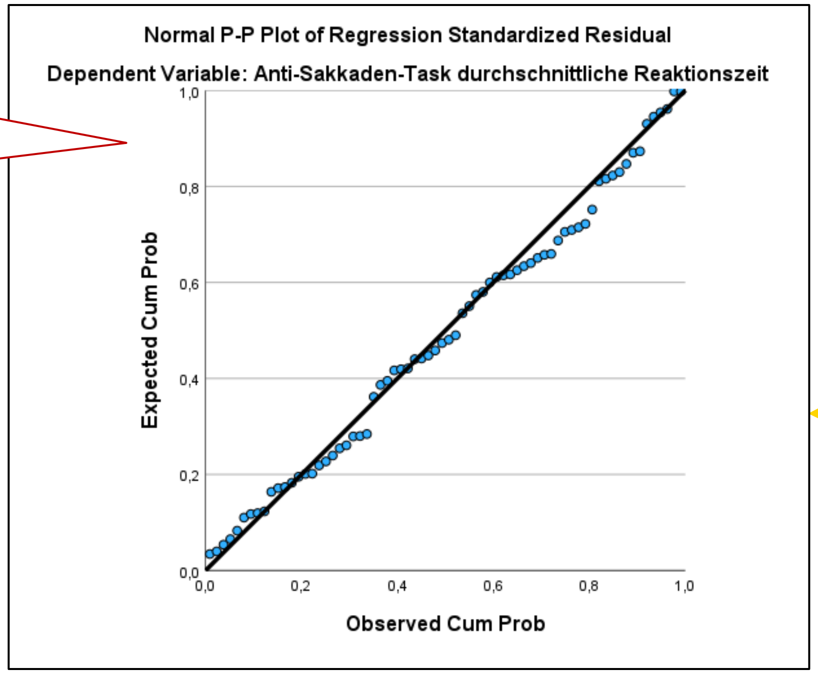
Unstandardized Residual der AST-Reaktionszeiten

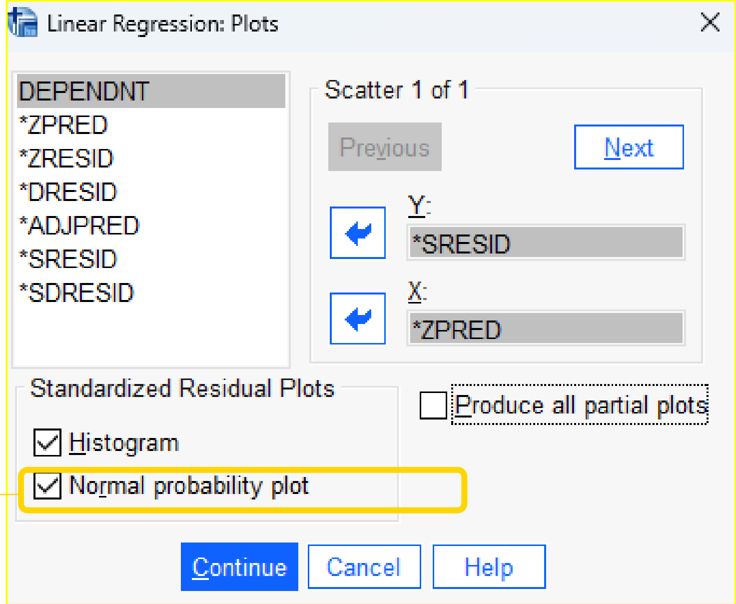
N	Valid	70
	Missing	0
Mean		,0000000
Median		-,0108437
Std. Deviation		,18528599
Skewness		,602
Std. Error of Skewness		,287
Kurtosis		,748
Std. Error of Kurtosis		,566
Minimum		-,34518
Maximum		,54624

Normal distribution of the residuals 2/2

- If the output of the residual plots has already been requested during the regression analysis, these can be used for graphical inspection

In a P-P plot, the expected probability is plotted against the observed cumulative probability. Normally distributed data lie as precisely as possible on the diagonal





- Normal distribution assumptions are the basis of many inferential statistical tests!

- **For very small samples** , the violation of the assumption can impair the significance of the significance test (less problematic for larger samples)
  - In general, the MLR is considered sufficiently robust against a violation of the normal distribution
- → **Countermeasure**: bootstrapping; transformation (*logarithmization*) of the variable

PROCESS-Macro

- Installation
- use for moderation- & (mediationsanalysis)

PROCESS: Installation 1/2

- PROCESS is a macro for SPSS developed by Andrew F. Hayes specifically designed for the analysis of mediated, moderated and conditional process models.
- It simplifies the application of regression analyses, especially for more complex models that examine interaction effects or indirect mediation effects.

Integration in SPSS

1. https://processmacro.org/download.html

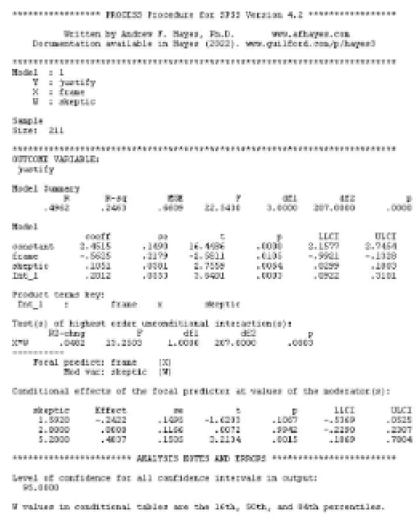
Download from the Resource Hub at CCRAM

Backup download link

"Backup download link" should automatically download the desired .zip folder.

Alternatively, you can take the detour via the website of the *Canadian Centre for Research Analysis and Methods* (step 2)

2. https://haskayne.ucalgary.ca/CCRAM/resource-hub

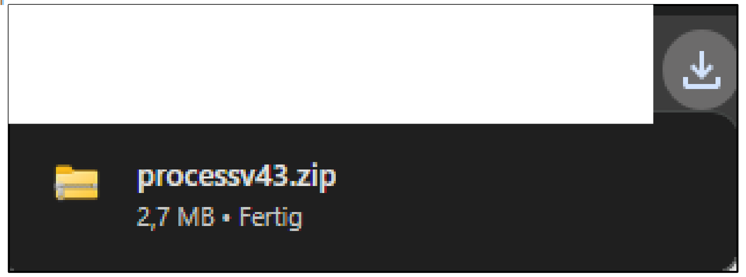


PROCESS macro for SPSS, SAS, and R

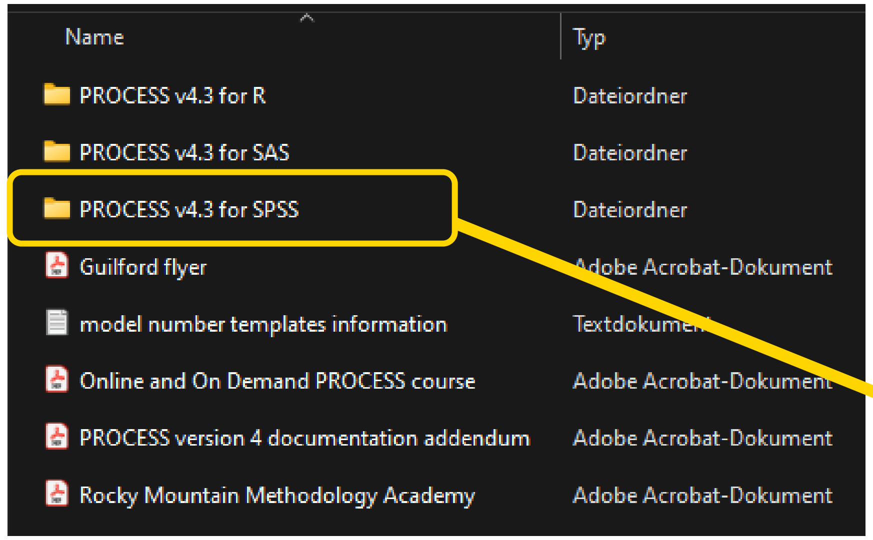
PROCESS is a computational tool invented by CCRAM expert Andrew F. Hayes. It is freely-available for SPSS, SAS, and R, and has become widely used throughout the behavioral sciences as well as in business research, medicine, and public health for easing the estimation of effects in mediation, moderation, and conditional process analysis.

Take a class on the use of PROCESS

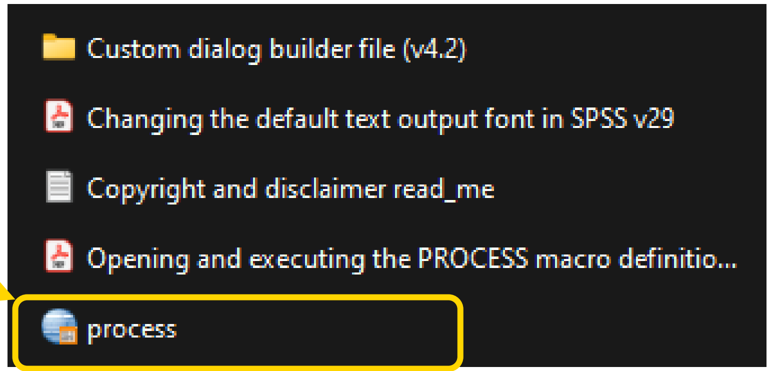
Download PROCESS v4.3



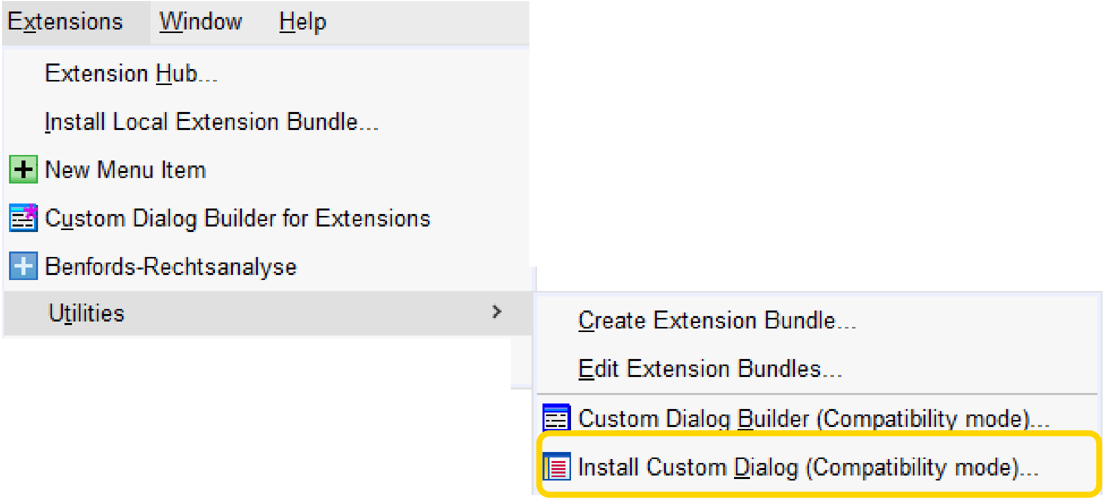
PROCESS: Installation 2/2



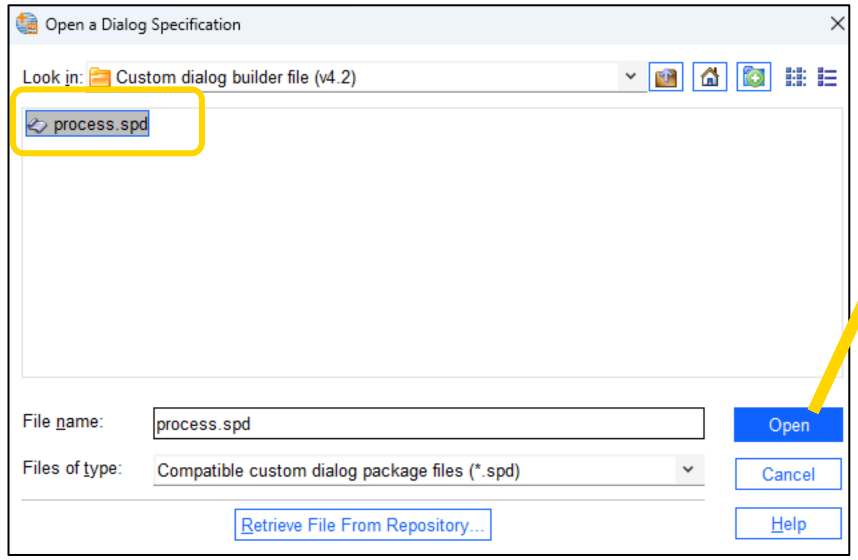
3. Unzip the .zip-folder & start „process.**sps**"-(Syntax-) file

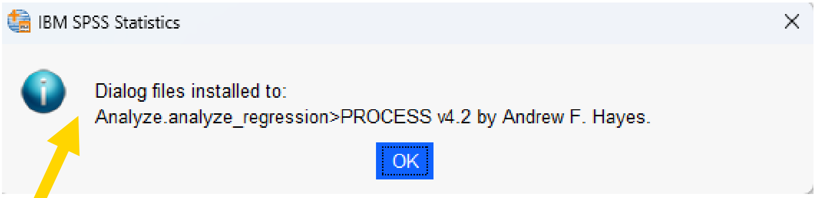


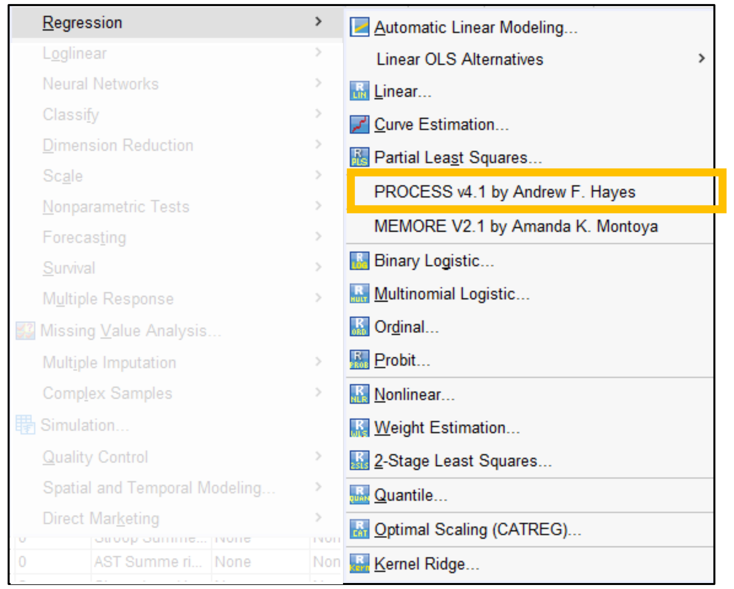
4. In the opern SPSS-Syntax: *Extensions* → *Utilities* → *Install custom Dialog* (Compatibility Mode)



5. Select the folder, in witch the file „process.**spd**" is safed + open it







6. You can select the macro under *Analyze* → *Regression*

Moderation analysis

- basic idea

- Execution in PROCESS

concept of a moderation

- The effect of a predictor on a criterion often depends on the expression of one or more other variables (= moderator variable)
  - A **moderation effect** is present when the relationship between 2 variables (IV & DV) depends on the expression of a third variable
  - Moderator variables can be quantitative and/or qualitative variables
  - Moderator variables influence the strength and/or direction of the relationship between IV & DV

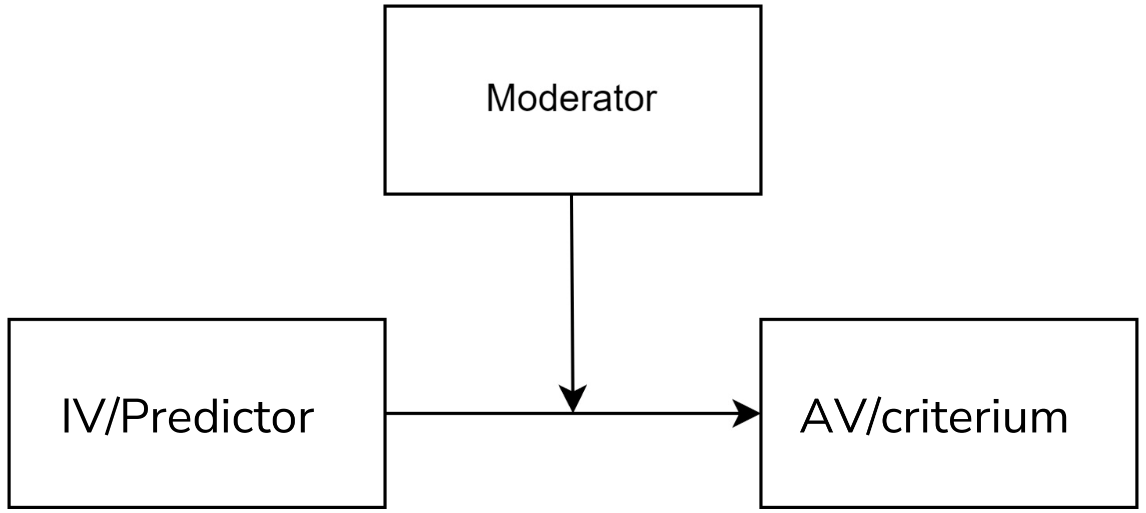
- Testing a moderation hypothesis requires the inclusion of:

  1. **the IV** (X_1) **2.** the **moderator** (X_2) **3.** the **product term** (IV * Moderator: X_1X_2)

  in the regression model

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon}$$

Mathematically, there is no difference between an independent variable and a moderator variable



concept of a moderation

- A **moderation effect** is present if the regression weight of the product term of IV and moderator becomes significant

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \boxed{\beta_3 X_1 X_2} + \epsilon$$

PROCESS-Output

```
OUTCOME VARIABLE:
 ast_rt

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      ,2739      ,0750      ,0359     1,7839     3,0000    66,0000      ,1588

Model
              coeff         se          t          p       LLCI       ULCI
constant      ,7473      ,0751     9,9507      ,0000      ,5974      ,8973
rand0         ,2593      ,1153     2,2487      ,0279      ,0291      ,4895
HOsum         ,0091      ,0101      ,9022      ,3702     -,0110      ,0293
Int_1        -,0343      ,0167    -2,0490      ,0444     -,0677     -,0009
```

SPSS-Output

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	,747	,075		9,951	<,001					
	Aufgabenart (Stroop-Test)	,259	,115	,675	2,249	,028	,117	,267	,266	,155	6,436
	Handlungsorientierung Summenwert	,009	,010	,135	,902	,370	-,064	,110	,107	,625	1,599
	Interaktionsterm Handlungsorientierung x Stroop-Bedingung	-,034	,017	-,621	-2,049	,044	-,007	-,245	-,243	,153	6,548

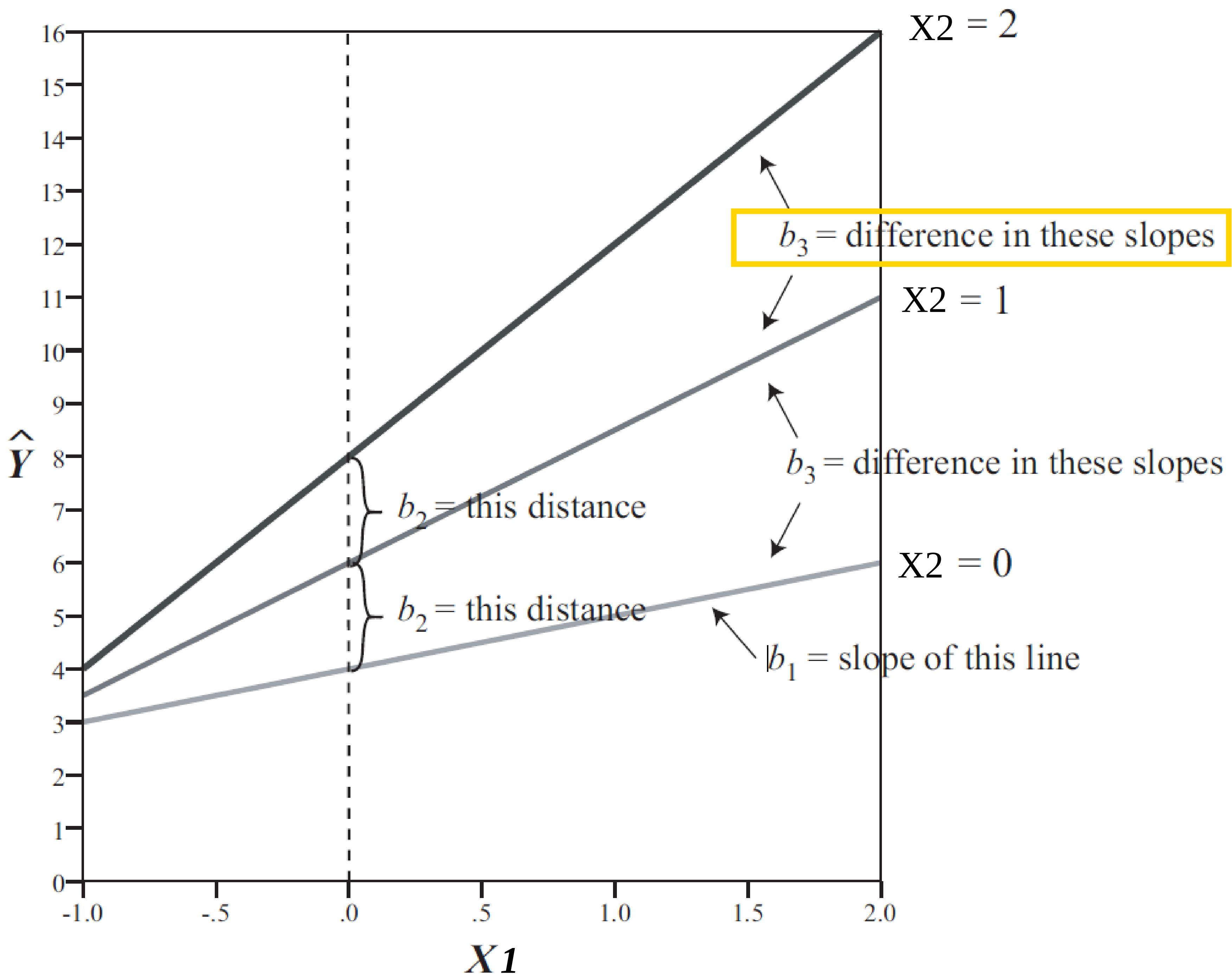
a. Dependent Variable: Anti-Sakkaden-Task durchschnittliche Reaktionszeit

- **If there is a significant interaction**, PROCESS offers corresponding follow-up analyses:
  - Conditional effects of IV on DV at certain values of the moderator
  - Graphical representation & hypothesis tests of the simple slopes (= slope of the regression line at different values of the moderator → e.g. -1 *SD*, *M*, +1 *SD*)
  - Significance region according to Johnson & Neyman

Interpretation of (unstandardized) regression-weights if a moderation effect exists

- „main effects“ → **conditional effects**
- **The interpretation of b_1 and b_2 changes** (compared to a model without interaction term!)

- b_1 is the **conditional** effect of X1 on Y when X2 = 0 (by how many units Ŷ changes when X1 increases by one unit and X2 equals 0)
- b_2 is the **conditional** effect of X2 on Y when X1 = 0 (by how many units Ŷ changes when X2 increases by one unit and X1 equals 0)
- b_3 is the number of units by which Ŷ changes based on the conditional effect of X1 when X2 increases by one unit (and vice versa)

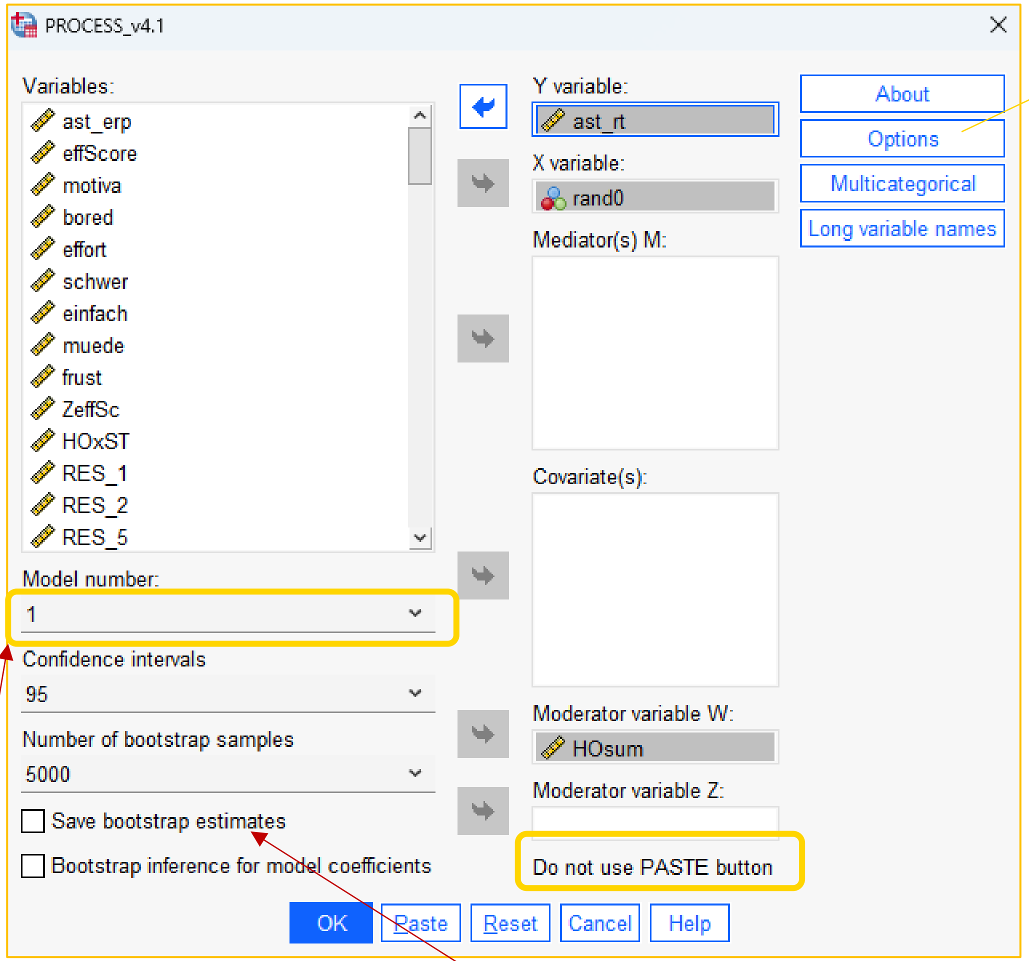


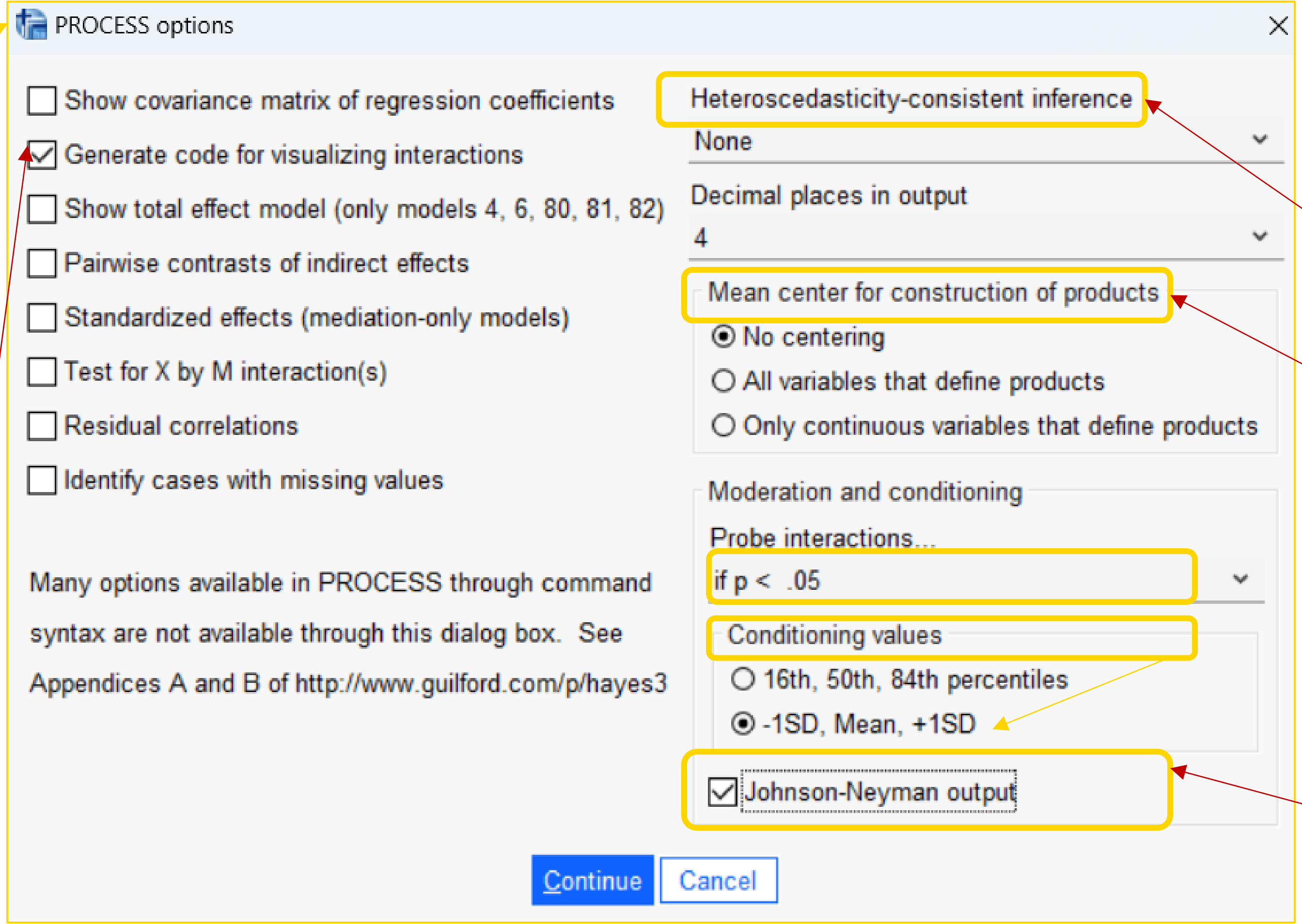
$b_1 = 1$, $b_2 = 2$, $b_3 = 1.5$

Graphic from Hayes (2018, S. 230)

Moderation analysis with PROCESS

- *Analyze → Regression → PROCESS v4.x by Andrew F. Hayes*





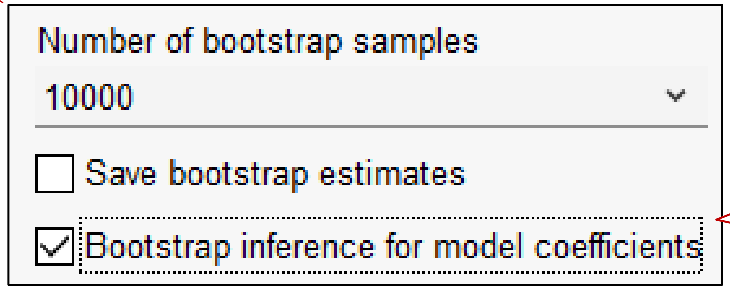
The selection of certain options depends on the respective data/analysis.

For **Heteroscedasticity**→ HC3 (Davidson-MacKinnon)

- For centering see slide "MLR - Centering of variables"
- "Conditioning values": how should the simple slopes be output?
- The „**Johnson-Neyman output**" option only makes sense for non-dichotomous moderator variables!

An overview abou all PROCESS models you can find here: https://osf.io/29c8p/download

With this option, PROCESS provides a syntax code that visualizes the interaction



You can also bootstrap your model parameters (the regression coefficients). For details, see the slides on "Bootstrapping" or (violation of) MLR assumptions

PROCESS-OUTPUT

```
Run MATRIX procedure:

***************** PROCESS Procedure for SPSS Version 4.1 *****************

          Written by Andrew F. Hayes, Ph.D.       www.afhayes.com
    Documentation available in Hayes (2022). www.guilford.com/p/hayes3

**************************************************************************
Model  : 1
    Y  : ast_rt
    X  : rand0
    W  : HOsum

Sample
Size:  70

**************************************************************************
OUTCOME VARIABLE:
 ast_rt

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      ,2739      ,0750      ,0359     1,7839     3,0000    66,0000      ,1588

Model
              coeff         se          t          p       LLCI       ULCI
constant      ,7473      ,0751     9,9507      ,0000      ,5974      ,8973
rand0         ,2593      ,1153     2,2487      ,0279      ,0291      ,4895
HOsum         ,0091      ,0101      ,9022      ,3702     -,0110      ,0293
Int_1        -,0343      ,0167    -2,0490      ,0444     -,0677     -,0009

Product terms key:
 Int_1    :        rand0    x        HOsum

Test(s) of highest order unconditional interaction(s):
       R2-chng          F        df1        df2          p
X*W      ,0588     4,1985     1,0000    66,0000      ,0444
----------
    Focal predict: rand0    (X)
          Mod var: HOsum    (W)
```

Das spezifizierte Modell

Globale Modellprüfung

Local check: main effects & interaction
"coeff" = unstand. Regression weights!

Change in R^2 by adding the interaction effect. **CAUTION – for explanatory purposes only: overall model is not significant here!**

```
*********** BOOTSTRAP RESULTS FOR REGRESSION MODEL PARAMETERS ************

OUTCOME VARIABLE:
 ast_rt

              Coeff   BootMean     BootSE   BootLLCI   BootULCI
constant      ,7473      ,7504      ,0675      ,6286      ,8930
rand0         ,2593      ,2569      ,1246      ,0115      ,5021
HOsum         ,0091      ,0085      ,0091     -,0107      ,0247
Int_1        -,0343     -,0336      ,0192     -,0701      ,0062

********************** ANALYSIS NOTES AND ERRORS ************************

Level of confidence for all confidence intervals in output:
  95,0000

Number of bootstrap samples for percentile bootstrap confidence intervals:
  10000

W values in conditional tables are the mean and +/- SD from the mean.

------ END MATRIX -----
```

Bootstrapping results

The conditional effects (simple slopes) for the selected values of the moderator (-1 SD, M, +1 SD) and their significance tests are as follows

```
Conditional effects of the focal predictor at values of the moderator(s):

      HOsum     Effect         se          t          p       LLCI       ULCI
     3,5997      ,1359      ,0647     2,1025      ,0393      ,0068      ,2650
     6,4571      ,0380      ,0459      ,8289      ,4101     -,0536      ,1296
     9,3146     -,0599      ,0678     -,8833      ,3803     -,1952      ,0755

Moderator value(s) defining Johnson-Neyman significance region(s):
      Value    % below    % above
     4,1665    25,7143    74,2857

Conditional effect of focal predictor at values of the moderator:
      HOsum     Effect         se          t          p       LLCI       ULCI
      ,0000      ,2593      ,1153     2,2487      ,0279      ,0291      ,4895
      ,6000      ,2387      ,1062     2,2484      ,0279      ,0267      ,4507
     1,2000      ,2182      ,0972     2,2441      ,0282      ,0241      ,4123
     1,8000      ,1976      ,0885     2,2329      ,0290      ,0209      ,3743
     2,4000      ,1771      ,0801     2,2108      ,0305      ,0172      ,3370
     3,0000      ,1565      ,0721     2,1709      ,0335      ,0126      ,3004
     3,6000      ,1359      ,0647     2,1024      ,0393      ,0068      ,2650
     4,1665      ,1165      ,0584     1,9966      ,0500      ,0000      ,2330
     4,2000      ,1154      ,0580     1,9887      ,0509     -,0005      ,2312
     4,8000      ,0948      ,0525     1,8072      ,0753     -,0099      ,1996
     5,4000      ,0743      ,0484     1,5348      ,1296     -,0223      ,1709
     6,0000      ,0537      ,0462     1,1634      ,2489     -,0385      ,1458
     6,6000      ,0331      ,0461      ,7195      ,4744     -,0588      ,1251
     7,2000      ,0126      ,0481      ,2615      ,7945     -,0835      ,1086
     7,8000     -,0080      ,0520     -,1534      ,8785     -,1119      ,0959
     8,4000     -,0285      ,0575     -,4967      ,6210     -,1433      ,0862
     9,0000     -,0491      ,0640     -,7670      ,4458     -,1769      ,0787
     9,6000     -,0697      ,0714     -,9759      ,3327     -,2122      ,0729
    10,2000     -,0902      ,0793    -1,1372      ,2596     -,2486      ,0682
    10,8000     -,1108      ,0877    -1,2629      ,2111     -,2859      ,0644
    11,4000     -,1313      ,0964    -1,3622      ,1778     -,3238      ,0612
    12,0000     -,1519      ,1054    -1,4418      ,1541     -,3622      ,0584

Data for visualizing the conditional effect of the focal predictor:
Paste text below into a SPSS syntax window and execute to produce plot.

DATA LIST FREE/
   rand0      HOsum      ast_rt     .
BEGIN DATA.
      ,0000     3,5997      ,7801
     1,0000     3,5997      ,9161
      ,0000     6,4571      ,8061
     1,0000     6,4571      ,8442
      ,0000     9,3146      ,8322
     1,0000     9,3146      ,7723
END DATA.
GRAPH/SCATTERPLOT=
 HOsum    WITH     ast_rt   BY       rand0    .
```

Johnson-Neyman significance region

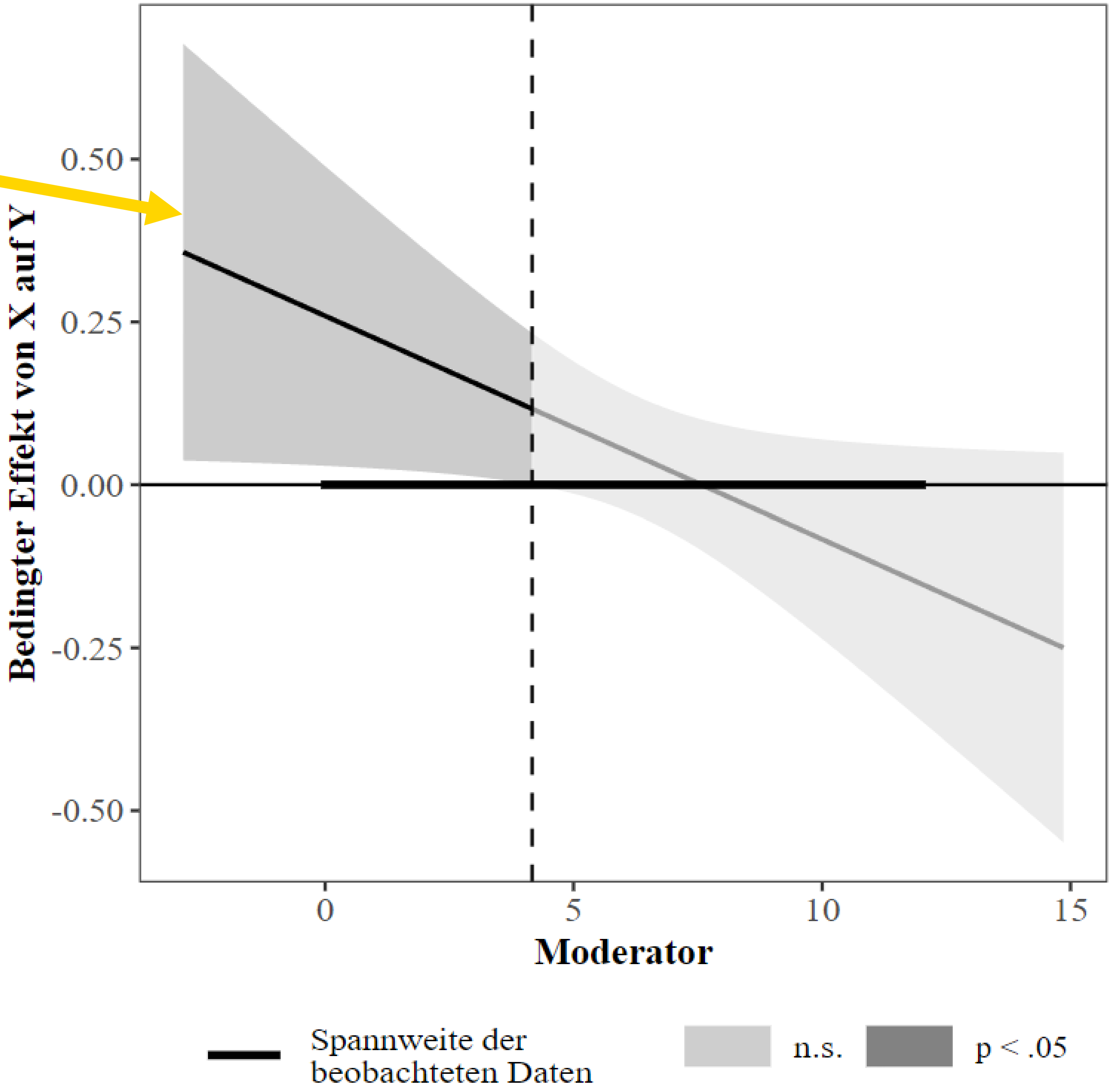
→ *A significant moderation effect can only be seen here with low values in the moderator variable*

Syntax code for creating the interaction diagram → is copied into an SPSS syntax file and executed

Johnson-Neyman-Diagramm

Conditional effect of focal predictor at values of the moderator:

HOsum	Effect	se	t	p	LLCI	ULCI
,0000	,2593	,1153	2,2487	,0279	,0291	,4895
,6000	,2387	,1062	2,2484	,0279	,0267	,4507
1,2000	,2182	,0972	2,2441	,0282	,0241	,4123
1,8000	,1976	,0885	2,2329	,0290	[illegible]	,3743
2,4000	,1771	,0801	2,2108	,0305	,0172	[illegible]
3,0000	,1565	,0721	2,1709	,0335	,0126	,3004
3,6000	,1359	,0647	2,1024	,0393	,0068	,2650
4,1665	,1165	,0584	1,9966	,0500	,0000	,2330
4,2000	,1154	,0580	1,9887	,0509	-,0005	,2312
4,8000	,0948	,0525	1,8072	,0753	-,0099	,1996
5,4000	,0743	,0484	1,5348	,1296	-,0223	,1709
6,0000	,0537	,0462	1,1634	,2489	-,0385	,1458
6,6000	,0331	,0461	,7195	,4744	-,0588	,1251
7,2000	,0126	,0481	,2615	,7945	-,0835	,1086
7,8000	-,0080	,0520	-,1534	,8785	-,1119	,0959
8,4000	-,0285	,0575	-,4967	,6210	-,1433	,0862
9,0000	-,0491	,0640	-,7670	,4458	-,1769	,0787
9,6000	-,0697	,0714	-,9759	,3327	-,2122	,0729
10,2000	-,0902	,0793	-1,1372	,2596	-,2486	,0682
10,8000	-,1108	,0877	-1,2629	,2111	-,2859	,0644
11,4000	-,1313	,0964	-1,3622	,1778	-,3238	,0612
12,0000	-,1519	,1054	-1,4418	,1541	-,3622	,0584



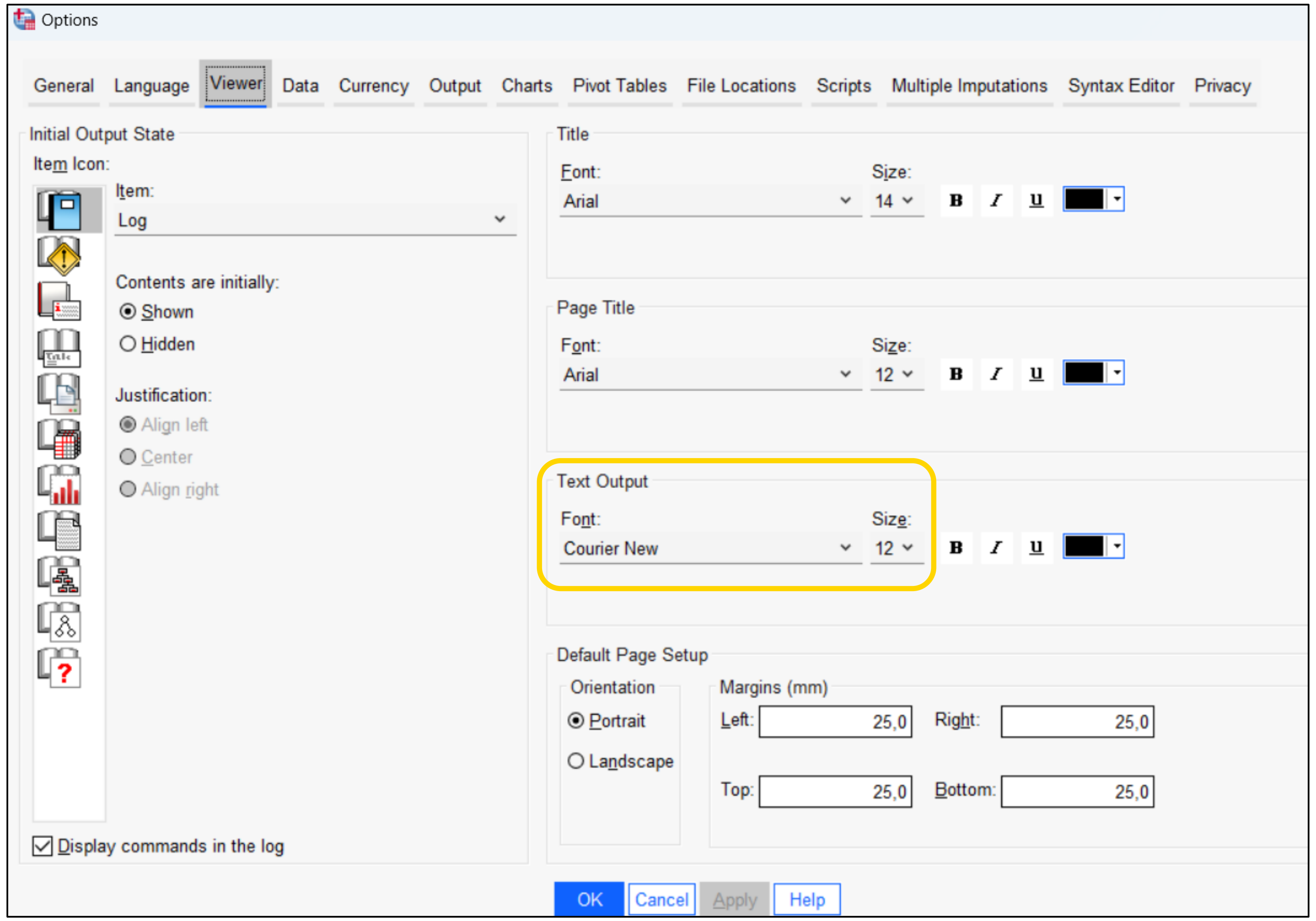
→ *A significant moderation effect can only be seen here with low values in the moderator variable!*

Created with Hemmerich, W. (2020). StatistikGuru: Johnson-Neyman-Plots creation. Retrieved from https://statistikguru.de/rechner/johnson-neyman-plots-erstellen.html

PROCESS-OUTPUT in SPSS 29

- According to the PDF file “Changing the default text output font in SPSS v29”, which was downloaded with PROCESS, the following should be noted:
  - With the introduction of SPSS 29, the default text output font has been changed and leads to a messy/shifted output of PROCESS results.
  - Return to original formating: *Edit* → *Options* → *Viewer* → *Textoutput* → *Font: „Courier New“*

PROCESS analyses already present in the SPSS output must be performed again in order to display the text output correctly!



PROCESS – interaction-diagramm

```
Data for visualizing the conditional effect of the focal predictor:
Paste text below into a SPSS syntax window and execute to produce plot.

DATA LIST FREE/
   rand0      HOsum      ast_rt     .
BEGIN DATA.
      ,0000     3,5997      ,7801
     1,0000     3,5997      ,9161
      ,0000     6,4571      ,8061
     1,0000     6,4571      ,8442
      ,0000     9,3146      ,8322
     1,0000     9,3146      ,7723
END DATA.
GRAPH/SCATTERPLOT=
 HOsum     WITH      ast_rt    BY        rand0     .
```

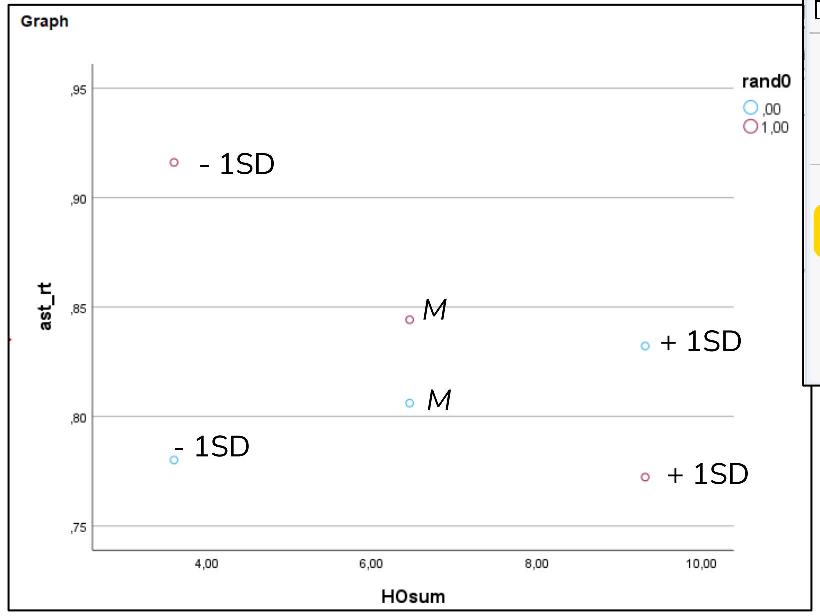

```
DATA LIST FREE/
   rand0      HOsum      ast_rt     .
BEGIN DATA.
      ,0000     3,5997      ,7801
     1,0000     3,5997      ,9161
      ,0000     6,4571      ,8061
     1,0000     6,4571      ,8442
      ,0000     9,3146      ,8322
     1,0000     9,3146      ,7723
END DATA.
GRAPH/SCATTERPLOT=
 HOsum     WITH      ast_rt    BY        rand0     .
```

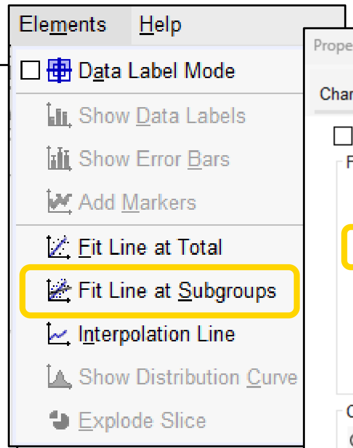

```
1  DATA LIST FREE/
2    rand0   HOsum   ast_rt  .
3  BEGIN DATA.
4     ,0000   3,5997   ,7801
5    1,0000   3,5997   ,9161
6     ,0000   6,4571   ,8061
7    1,0000   6,4571   ,8442
8     ,0000   9,3146   ,8322
9    1,0000   9,3146   ,7723
10 END DATA.
11 GRAPH/SCATTERPLOT=
12  HOsum   WITH   ast_rt  BY    rand0
```

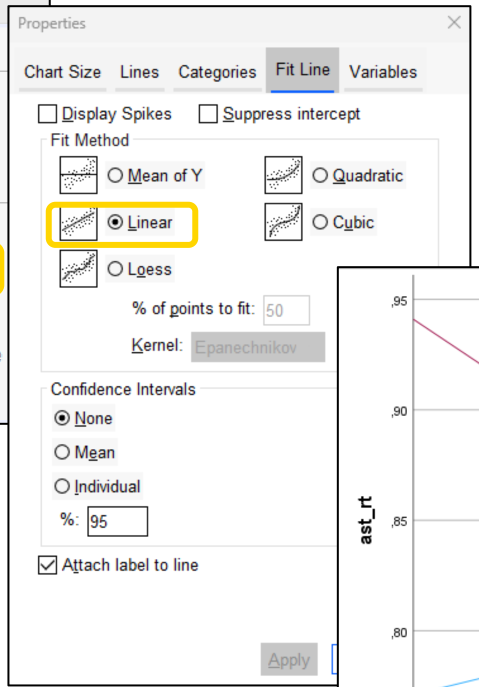
1. Double-click on output
2. Select the section
3. Copy & paste into syntax file
4. Execute syntax command

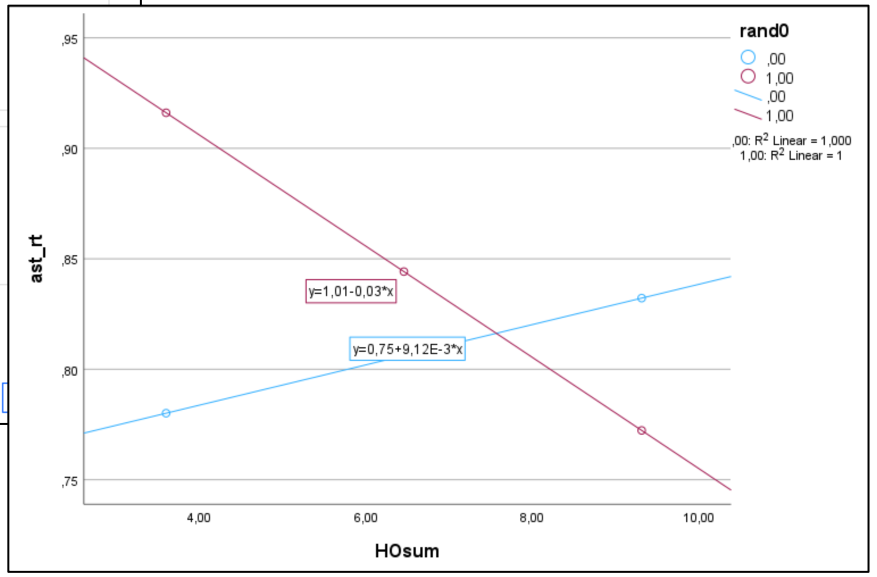
In order to be able to analyze the interaction pattern more precisely, a second version of the graphic can be requested in which simple slopes are plotted for certain characteristics of the IV. To do this, the code is copied and only the last line is adapted (the variable name of the IV and the moderator are swapped):
rand0 WITH ast_rt BY HOsum

5. double-click the graphic in the output→ Editor
6. Elements → Fit Line at Subgroups → Linear









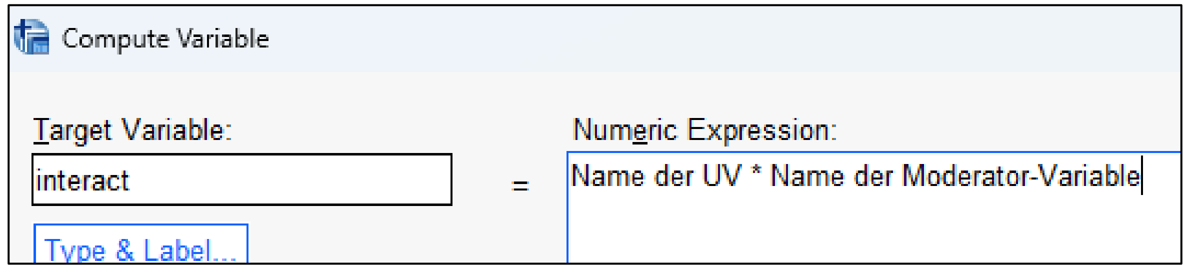
Since SPSS graphics are generally not welcome / not permitted for written reports, it makes sense to copy the values into Excel, for example, and create an APA-compliant graphic there!

Interpretation of MLR/moderation analyses and further information

Further information for moderation analyses

- In this example, a dichotomous IV and a metric moderator variable (and their interaction term) were included in the regression model.
  - If you use a dichotomous moderator, there is no need to output the Johnson & Neyman significance regions
  - The use of a categorical variable (as IV or moderator → dummy coding) is not explicitly considered here. Instructions can be found on the Internet.

- PROCESS CANNOT be used to Therefore, **the interaction term IV * moderator** must first be **created in SPSS in the form of a new variable**
  - *Transform → Compute variable*
  - This variable is then entered into the regression together with IV and moderator in the regression



Interpretation of results 1/2

- **1. global model check**
  - R^2: How much variance of the DV can be explained by the specified model (the predictors); rule of thumb according to Cohen (1988):
    - $R^2 = .02$ → low / weak variance explanation
    - $R^2 = .13$ → medium / moderate variance explanation
    - $R^2 = .26$ → high / strong variance explanation
    - *p*-value of the overall model (interprtation is certanly known)

$R^2 = .08$, $F(3, 66) = 1.78$, $p = .159$

Model Summary

R	R-sq	MSE	F	df1	df2	p
,2739	,0750	,0359	1,7839	3,0000	66,0000	,1588

CAUTION! If the overall model is not significant, the local check is not necessary! For 2. we assume that we have a significant overall model!

- **2. local model check** – main effects and interaction effects
  - „`coeff`" = unstandardized regression weights
  - Interpretation: see section on "MLR basic idea"
  - R2-chng/ΔR²: how much additional variance of the DV is caused by the additionally clarified by the addition of the interaction term
    - Identical procedure to a hierarchical regression

e.g.: $b = 0.26$, $SE = 0.12$, $t(66) = 2.25$, $p = .028$, 95%-CI[0.03, 0.49]

Model

	coeff	se	t	p	LLCI	ULCI
constant	,7473	,0751	9,9507	,0000	,5974	,8973
rand0	,2593	,1153	2,2487	,0279	,0291	,4895
HOsum	,0091	,0101	,9022	,3702	-,0110	,0293
Int_1	-,0343	,0167	-2,0490	,0444	-,0677	-,0009

Product terms key:
Int_1 : rand0 x HOsum

Test(s) of highest order unconditional interaction(s):

	R2-chng	F	df1	df2	p
X*W	,0588	4,1985	1,0000	66,0000	,0444

Interpretation of results 2/2

- Significance of the simple slopes
  - Here for values -1 *SD*, *M*, +1 *SD*
  - Graphical representation!

Conditional effects of the focal predictor at values of the moderator(s):

HOsum	Effect	se	t	p	LLCI	ULCI
3,5997	,1359	,0647	2,1025	,0393	,0068	,2650
6,4571	,0380	,0459	,8289	,4101	-,0536	,1296
9,3146	-,0599	,0678	-,8833	,3803	-,1952	,0755

- Johnson & Neyman-significanace regions
  - In principle, there can also be 2 significance regions, in which case both threshold values are reported

Make sure that the variable values also make sense in terms of content! Take this into account especially for metric variables whose values are less finely graded than is shown in the output!

Conditional effect of focal predictor at values of the moderator:

HOsum	Effect	se	t	p	LLCI	ULCI
,0000	,2593	,1153	2,2487	,0279	,0291	,4895
,6000	,2387	,1062	2,2484	,0279	,0267	,4507
1,2000	,2182	,0972	2,2441	,0282	,0241	,4123
1,8000	,1976	,0885	2,2329	,0290	,0209	,3743
2,4000	,1771	,0801	2,2108	,0305	,0172	,3370
3,0000	,1565	,0721	2,1709	,0335	,0126	,3004
3,6000	,1359	,0647	2,1024	,0393	,0068	,2650
4,1665	,1165	,0584	1,9966	,0500	,0000	,2330
4,2000	,1154	,0580	1,9887	,0509	-,0005	,2312
4,8000	,0948	,0525	1,8072	,0753	-,0099	,1996
5,4000	,0743	,0484	1,5348	,1296	-,0223	,1709
6,0000	,0537	,0462	1,1634	,2489	-,0385	,1458
6,6000	,0331	,0461	,7195	,4744	-,0588	,1251
7,2000	,0126	,0481	,2615	,7945	-,0835	,1086
7,8000	-,0080	,0520	-,1534	,8785	-,1119	,0959
8,4000	-,0285	,0575	-,4967	,6210	-,1433	,0862
9,0000	-,0491	,0640	-,7670	,4458	-,1769	,0787
9,6000	-,0697	,0714	-,9759	,3327	-,2122	,0729
10,2000	-,0902	,0793	-1,1372	,2596	-,2486	,0682
10,8000	-,1108	,0877	-1,2629	,2111	-,2859	,0644
11,4000	-,1313	,0964	-1,3622	,1778	-,3238	,0612
12,0000	-,1519	,1054	-1,4418	,1541	-,3622	,0584

Estimation and interpretation of the parameters of centered predictors

Problematic for interpretation if the variable value “0” is not meaningful in terms of content!

non-centered predictors

- b_1 is the conditional effect of X1 on Y when **X2 = 0** (by how many units $\hat{Y}$ changes when X1 increases by one unit and X2 equals 0)
- b_2 is the conditional effect of X2 on Y when **X1 = 0** (by how many units $\hat{Y}$ changes when X2 increases by one unit and X1 equals 0)

centered predictors

- b_1 is the conditional effect of X1 on Y when **X2 is equal to ist mean** (by how many units $\hat{Y}$ changes when X1 increases by one unit and X2 is equal to $\overline{X2}$)
- b_2 is the conditional effect of X2 on Y when **X1 is equal to ist mean** (by how many units $\hat{Y}$ changes when X2 X1 increases by one unit and X1 is equal to $\overline{X1}$)

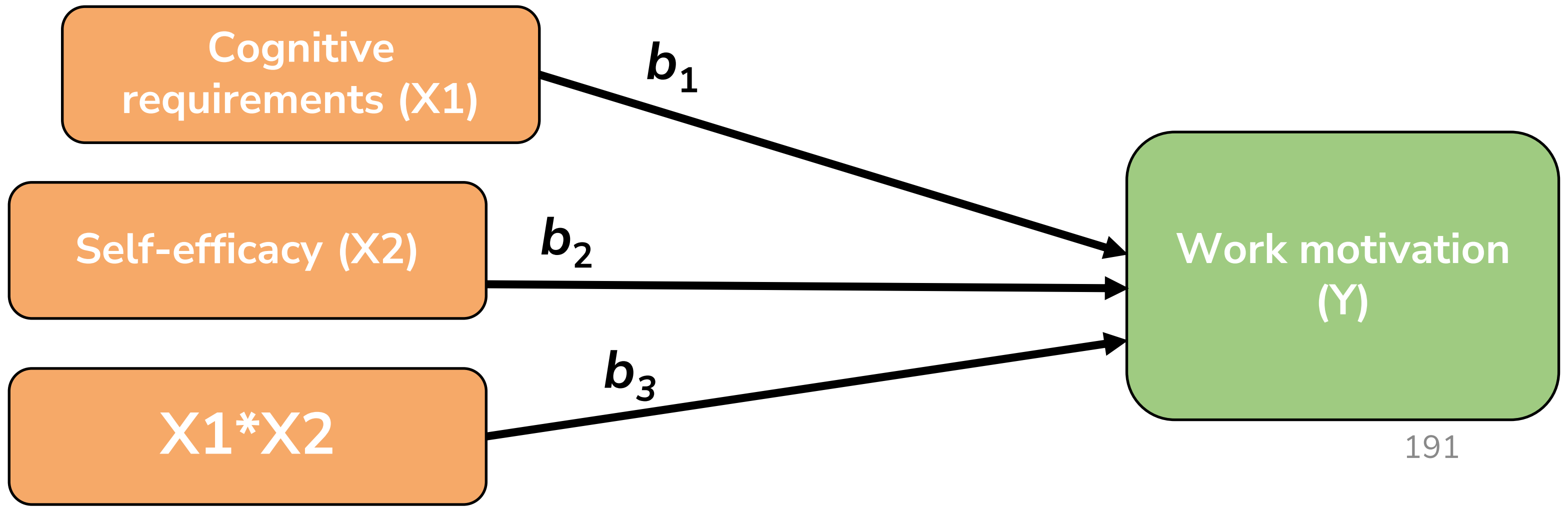
Centering only affects the regression coefficients, standard errors, t- and p-values of X1 and X2.
The values for X1*X2 remain the same.
For a detailed explanation, see Hayes (2018).

Example – simple Moderation

- <u>Research question</u>: Is the relationship between the **cognitive requirement to structure one's own work activity** (X1) and **work motivation** (Y) moderated by the **extent of self-efficacy** (X2)?
- Possible hypotheses:
  1. “Cognitive demands are positively associated with work motivation.” (b_1)
  2. “Self-efficacy experiences are positively associated with work motivation.” (b_2)
  3. “The effect of cognitive demands on work motivation is moderated by the level of self-efficacy/changes depending on self-efficacy” (b_3)

At this point, we would like to point out once again that:
a moderator only becomes a moderator through the (plausible) theoretical assumption of a moderation relationship. From a mathematical point of view, there is no difference in the way X1 and X2 are included in the statistical analysis!

- <u>Analysis</u>: Multiple linear Regression → Moderation analysis
- What possible <u>effects</u>?
  - Conditional effect „main effect“ of X1 on Y (b_1)
  - Conditional effect „main effect“ of X2 on Y(b_2)
  - Interaction effect „Moderation“ from X1*X2 to Y (b_3)



Example – simple Moderation

- "Is the relationship between the **cognitive requirement to structure one's own work activity** (X1) and **work motivation** (Y) moderated by the **level of self-efficacy** (X2)?"
  1. "Cognitive demands are positively associated with work motivation."(b_1)
  2. "Self-efficacy experiences are positively associated with work motivation." (b_2)
  3. "The effect of cognitive demands on work motivation is moderated by the level of self-efficacy/changes depending on self-efficacy"(b_3)

→ **You carry out ONE analysis to test the three hypotheses!**

→ It would not be correct to carry out separate correlations or t-tests/ANOVAs (for dichotomous predictors) for the "main effects"

→ results (fictitious and without *F*- or *t*-values):

$R^2 = .21, p < .001$	Specified total model is significant	→ Continue to test the regression coefficients
$b_1 = 0.19, p = .013$	Significant sonditional effect of X1 on Y	→ Cognitive demands are positively associated with work motivation
$b_2 = 0.29, p < .001$	Significant sonditional effect of X2 on Y	→ Self-efficacy is positively associated with work motivation
$b_3 = 0.10, p < .001$	Significant interaction effect of X1*X2 on Y	→ The relationship between cognitive demands and work motivation is moderated by the level of self-efficacy

Be careful with terms that imply causality, especially in cross-sectional study designs!
(Even though we often draw unidirectional arrows in statistical models!)

Mediation analysis

- Basic idea

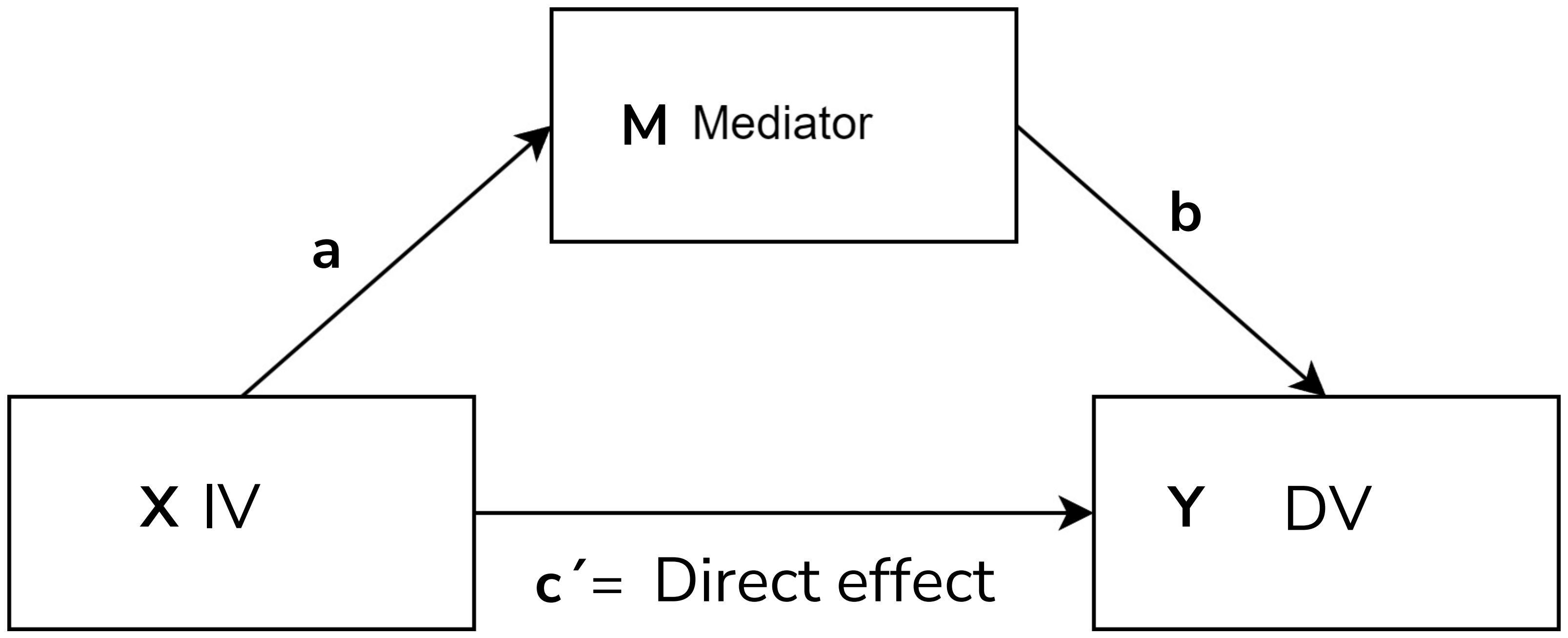
- Execution in PROCESS

Basic idea of a Mediation

- The effect that a predictor (X/IV) has on a criterium (Y/DV) is explained by one or more third variables (mediator):
  - fully explained = total mediation → c´ is not significant
  - partially explained = partial mediation → after considering M, a significant effect of X on Y remains
- A Mediator is both a IV and DV at the same time
- Caution: The implied causal relationship is not established by statistical analysis alone, but must be supported by an appropriate research design!

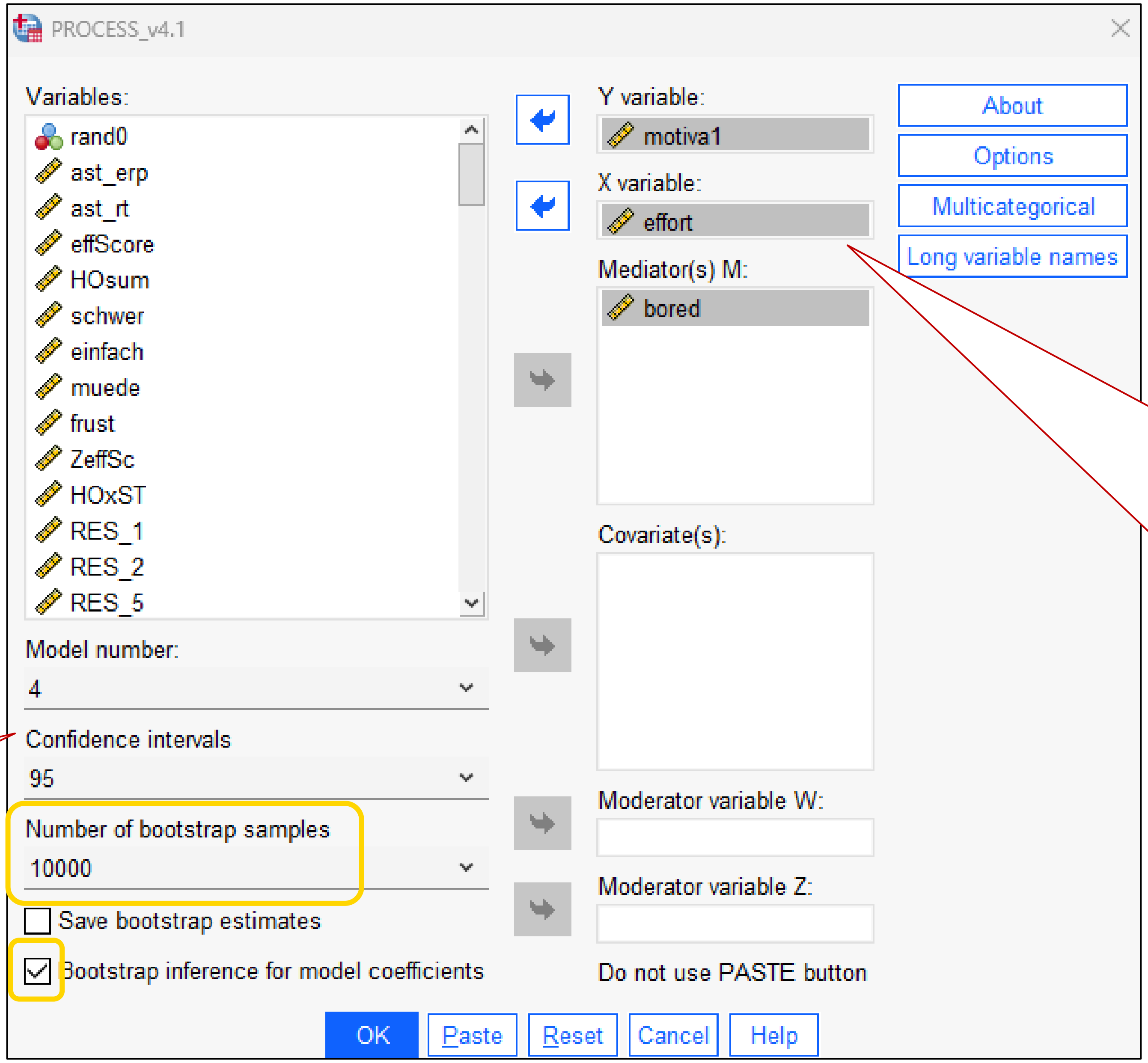
- (total effect) **c = c´+ a*b**

- **a*b:** The product of paths a and b is referred to as the **indirect** effect of X on Y through M, and it quantifies the extent of the mediation



Mediation in PROCESS 1/2

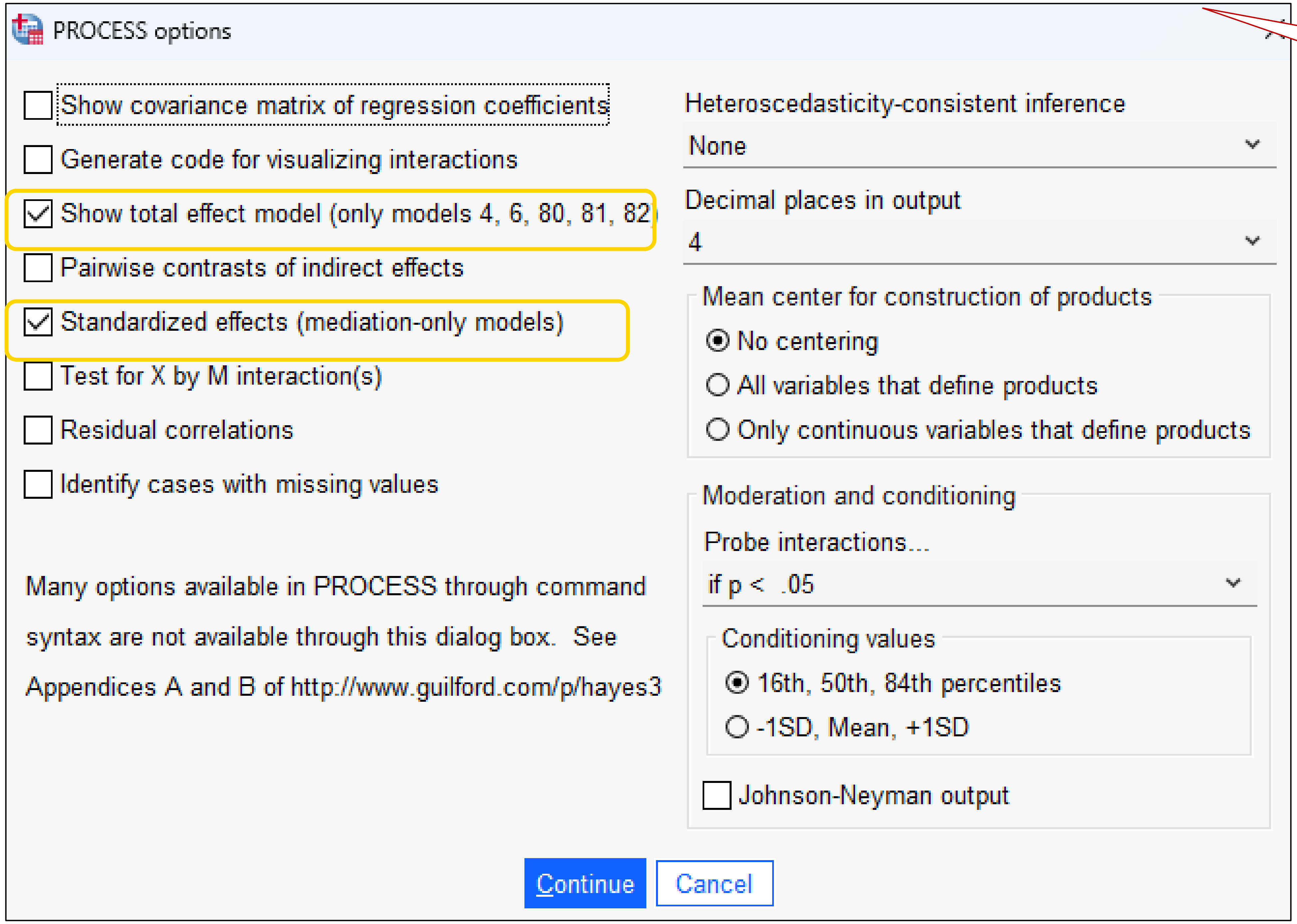
The prerequisites for mediation are similar to those of MLR. Additionally, we can generally bootstrap our coefficients!



- **Just like in a moderation analysis, specify the model in the dialog window**
- **X can be either a categorical predictor (→ dummy coding) or a metric predictor.**

model nr.4 is the „classic mediation modell“(see provious slide)

Mediation in PROCESS 2/2

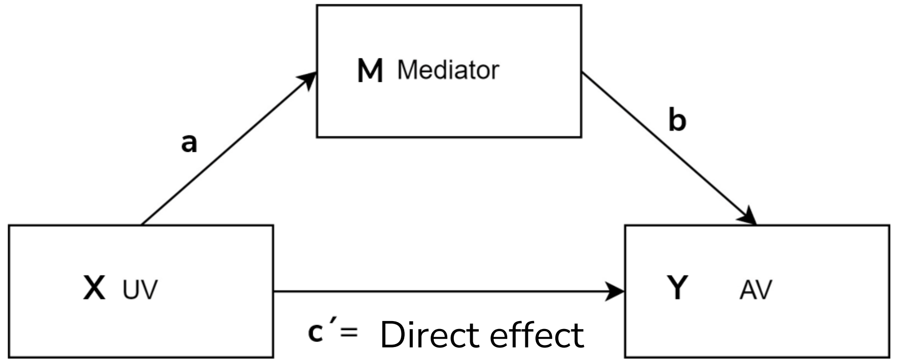


- Select accordingly under "Options.
- For heteroscedastic data, choose the appropriate correction
- Depending on the type of mediation model (e.g., multiple mediation), additional options may be required → *Pairwise contrasts*

Mediation in PROCESS

```
Run MATRIX procedure:

***************** PROCESS Procedure for SPSS Version 4.1 *****************

          Written by Andrew F. Hayes, Ph.D.       www.afhayes.com
    Documentation available in Hayes (2022). www.guilford.com/p/hayes3

**************************************************************************
Model  : 4
    Y  : motival
    X  : effort
    M  : bored

Sample
Size:  70

**************************************************************************
OUTCOME VARIABLE:
 bored

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      ,4089      ,1672     2,1843    13,6494     1,0000    68,0000      ,0004

Model
              coeff         se          t          p       LLCI       ULCI
  constant   7,0833     1,0201     6,9437      ,0000     5,0477     9,1188
1 effort     -,6947      ,1880    -3,6945      ,0004    -1,0700     -,3195

Standardized coefficients
             coeff
effort      -,4089

**************************************************************************
OUTCOME VARIABLE:
 motival

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      ,5222      ,2727     1,6749    12,5576     2,0000    67,0000      ,0000

Model
              coeff         se          t          p       LLCI       ULCI
2 constant   2,0484     1,1678     1,7541      ,0840     -,2825     4,3793
  effort      ,4936      ,1804     2,7357      ,0080      ,1335      ,8538
3 bored      -,2881      ,1062    -2,7134      ,0085     -,5001     -,0762

Standardized coefficients
             coeff
effort       ,3123
bored       -,3098
```

a = 1
c‘= 2
b = 3
c = 4
a*b = 5



Important question: Is there still a (significant) direct effect of X on Y after accounting for the mediator or indirect effect?

```
************************** TOTAL EFFECT MODEL ****************************
OUTCOME VARIABLE:
 motival

Model Summary
          R       R-sq        MSE          F        df1        df2          p
      ,4390      ,1927     1,8316    16,2338     1,0000    68,0000      ,0001

Model
              coeff         se          t          p       LLCI       ULCI
  constant    ,0074      ,9341      ,0079      ,9937    -1,8566     1,8714
4 effort      ,6938      ,1722     4,0291      ,0001      ,3502     1,0374

Standardized coefficients
             coeff
effort       ,4390

************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y **************

1 Total effect of X on Y
     Effect         se          t          p       LLCI       ULCI       c_cs
      ,6938      ,1722     4,0291      ,0001      ,3502     1,0374      ,4390

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI      c'_cs
      ,4936      ,1804     2,7357      ,0080      ,1335      ,8538      ,3123

Indirect effect(s) of X on Y:
          Effect     BootSE   BootLLCI   BootULCI
5 bored    ,2002      ,0980      ,0209      ,4039

Completely standardized indirect effect(s) of X on Y:
          Effect     BootSE   BootLLCI   BootULCI
bored      ,1267      ,0620      ,0129      ,2538

*********** BOOTSTRAP RESULTS FOR REGRESSION MODEL PARAMETERS ************

OUTCOME VARIABLE:
 bored

              Coeff   BootMean     BootSE   BootLLCI   BootULCI
constant     7,0833     7,1353     1,0283     5,1169     9,1755
effort       -,6947     -,7044      ,1872    -1,0726     -,3329

----------

OUTCOME VARIABLE:
 motival

              Coeff   BootMean     BootSE   BootLLCI   BootULCI
constant     2,0484     1,9362     1,3312     -,8985     4,2939
effort        ,4936      ,5099      ,1912      ,1719      ,9197
bored        -,2881     -,2814      ,1217     -,5075     -,0312

********************** ANALYSIS NOTES AND ERRORS ************************

Level of confidence for all confidence intervals in output:
  95,0000

Number of bootstrap samples for percentile bootstrap confidence intervals:
  10000
```

If 0 is not included in the confidence interval, the path is significant

Purely fictional data example!

Mediation – presenting results

- No general standards available
- Recomendations:
  - Graphical representation + reporting results in the text or APA-table for MLR
  - Path coefficients (a, b, a*b, c‘) → (un)standardized, SE, bootstrapping confidence interval
  - Effect size:
    - Dichotomous predictor → partially standardized effect (unstd. a*b / SD_Y)
    - Continuous predictor → fully standardized effect (unstd. a*b / SD_Y / SD_X)

- Further guidelines for result presentation can be found online
  - e.g. http://www.regorz-statistik.de/inhalte/tutorial_mediator_bootstrapping_process.html#:~:text=Beim%20partiell%20standardisierten%20Effekt%20wird,an%20der%20abh%C3%A4ngigen%20Variable%20standardisiert.

Mediation with categorical predictors

- A detailed description can be found, for example, in A. Field (2018).
- **Please feel free to reach out by email if you have any questions regarding your analysis!**

- Key points to remember:
  - A categorical variable is dummy-coded
  - One group serves as the reference/baseline for all other groups → 0-coding

- Video-tutorial: https://www.youtube.com/watch?v=r2_zw4G55X0

Bootstrapping

Bootstrapping - Grundidee

- If normal distribution of data given → inference to normally distributed sample characteristic distribution (which we use for many inferential statistical methods)
  - Particularly problematic with small samples *(does this actually reflect the population?)*
- Bootstrapping methods are used to estimate the distribution properties on the basis of the sample

1. The sample is treated as a population
2. A specific number of subsamples (bootstrap samples) is drawn from the sample

   → **With „laying back“ the cases!**
3. In each bootstrap sample, the parameter of interest is calculated (e.g., 10,000 times)
4. The calculated values are sorted in ascending order, and the range that captures 95% of these values is determined → serves as an estimate of the 95% confidence interval of the parameter (not dependent on the assumption of normality!)
5. The standard deviation of the bootstrap samples serves as the standard error

Caution: The parameter estimates deviate (slightly) from each other due to the described random process during implementation!

References and Links

Usefull Links

- https://www.esci.thenewstatistics.com/
  - Shows different visualisations
- https://www.psychometrica.de/effektstaerke.html
- https://rpsychologist.com/viz
  - Shows different visualisations

References

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* (2. Auflage). New York: Routledge.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2017). *Statistik und Forschungsmethoden: Lehrbuch. Mit Online-Material* (Originalausgabe, 5., korrigierte Aufl.). Beltz.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146
- Field, A. (2013, 2018). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll* (4th edition). *MobileStudy*. Sage.
- Moosbrugger, H., & Kelava, A. (Eds.). (2020). *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion* (3rd ed.). Springer Berlin Heidelberg.

- Slides from „VO Psychologische Statistik 1 (WS 23)“
- Slides from „VO Psychologische Statistik 2 (SS 24)“
- Slides from „Forschungsmethodik und wissenschaftliches Arbeiten (SS 21)“