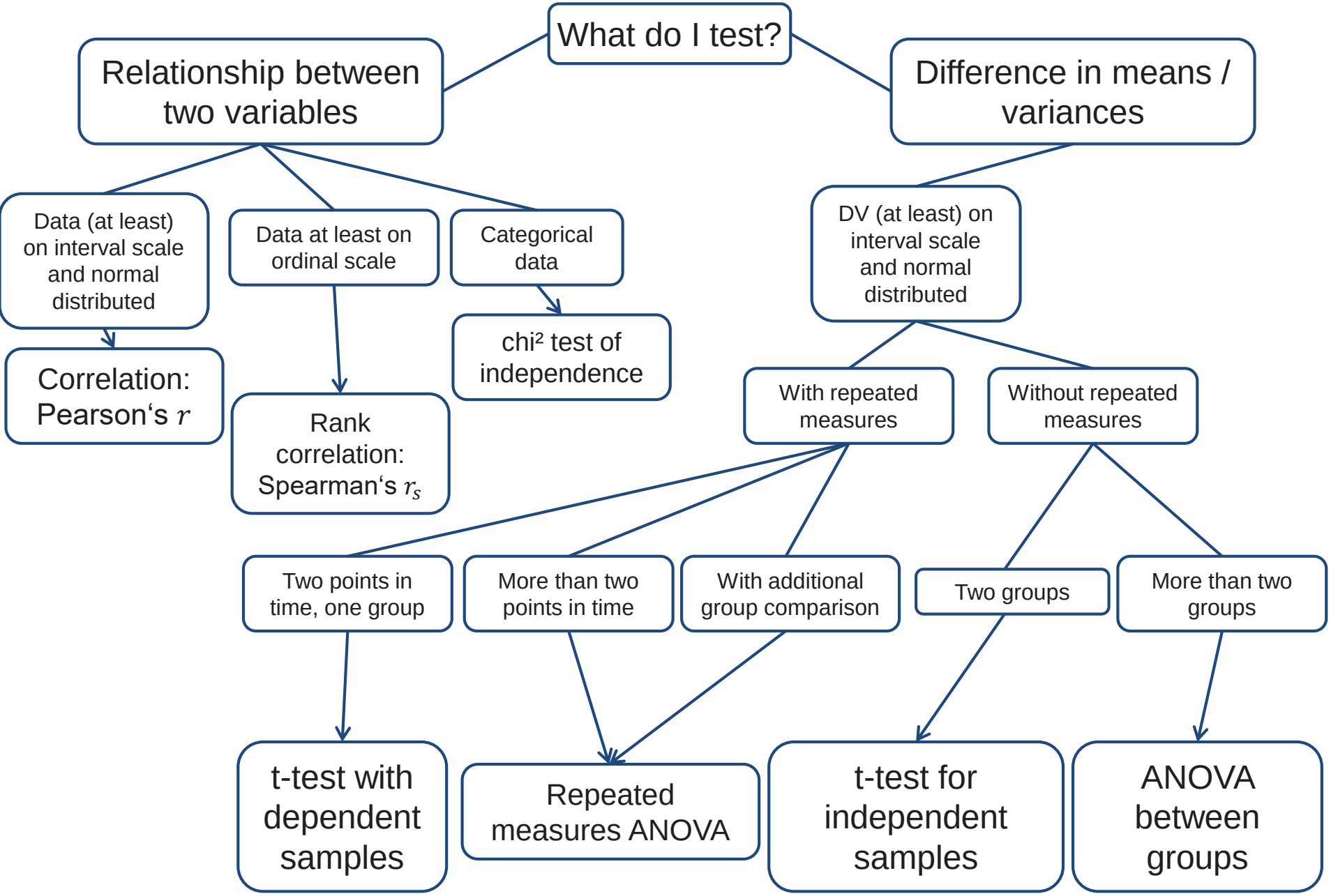
t-test with dependent samples, repeated measures ANOVA (within subjects design), effect size

Lara Kataja

Room 1.13

l.kataja@edutech.uni-saarland.de


What do I test?
Relationship between two variables
Difference in means / variances
Data (at least) on interval scale and normal distributed
Data at least on ordinal scale
Categorical data
DV (at least) on interval scale and normal distributed
Correlation: Pearson‘s r
Rank correlation: Spearman‘s r_s
chi² test of independence
With repeated measures
Without repeated measures
Two points in time, one group
More than two points in time
With additional group comparison
Two groups
More than two groups
t-test with dependent samples
Repeated measures ANOVA
t-test for independent samples
ANOVA between groups

Within or between?

Example research question: Do you learn more with text or video?

Independent samples t-test (total n = 50 subjects)

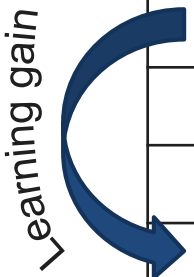
IV	Condition ‘text’	Condition ‘video’
n	25	25
DV	Knowledge test	

Dependent samples (paired samples) t-test (total n = 25 subjects)

n	25
IV	Condition ‘text’
DV1	Knowledge test
IV	Condition ‘video’
DV2	Knowledge test

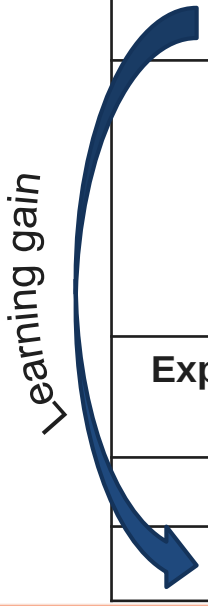
Repeated measures ANOVA: Can test for change over time AND the impact of several experimental factors simultaneously

One-way repeated measures ANOVA design (total n = 50 subjects)



DV	Knowledge test (pre)	
IV	Condition ‘text’	Condition ‘video’
n	25	25
DV	Knowledge test (post)	

Two-way repeated measures ANOVA design (total n = 100 subjects)



DV	Knowledge test (pre)			
IV	Factor 'text’			
	yes		no	
	Factor ‘video’		Factor ‘video’	
	yes	no	yes	no
Exp. groups	Text + Video	Only text	Only Video	Control group
n	25	25	25	25
DV	Knowledge test (post)			

Effect size – Why?

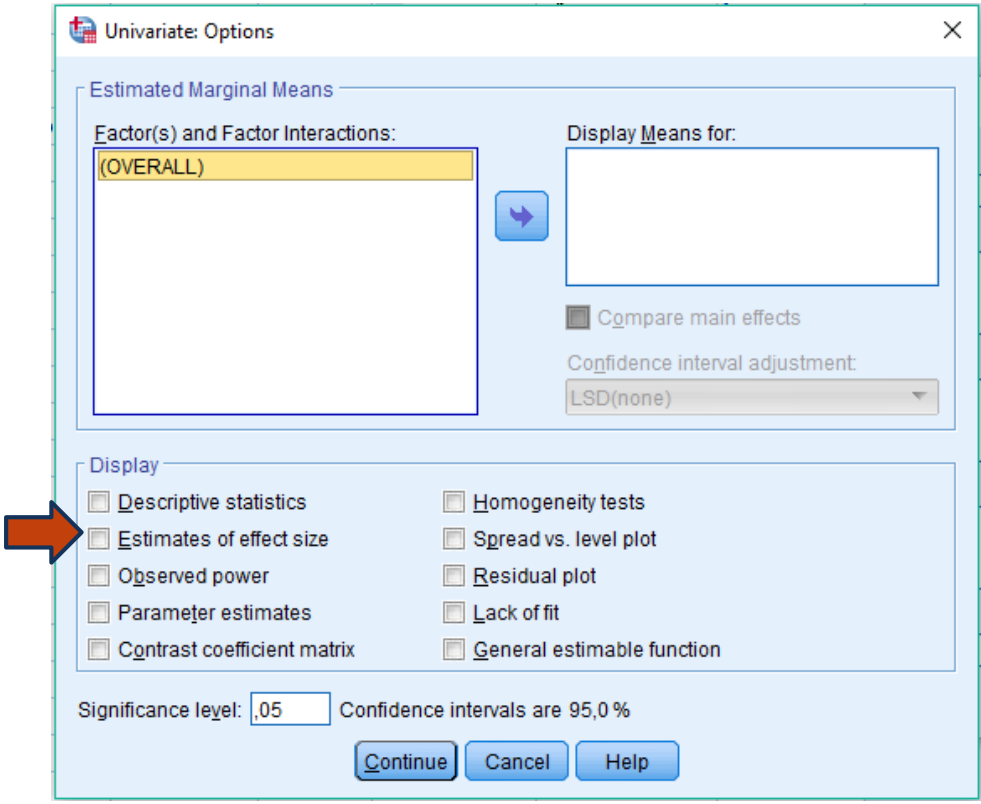
- Only telling about the ‘significance’ is not very meaningful: Significant results are also possible for practically meaningless differences

  → Additionally, report the *effect size*

  - “How large is a significant difference?”
  - “How strong is an effect?”

Effect sizes – two important measures

- η_p^2:
  - = partial eta-square
  - for ANOVAs
  - Measure of the proportion of explained variance in relationship to the overall variance (percentage)
  - You can select it in SPSS for ANOVAs



Effect sizes – two important measures

- **Cohen‘s *d*:**
  - For t-tests
  - Not selectable in SPSS
  - → Use an internet calculator, e.g.: https://www.psychometrica.de/effect_size.html

Effect sizes – Interpretation

d	r^*	η^2	Interpretation sensu Cohen (1988)	Interpretation sensu Hattie (2007)
< 0	< 0	-	Adverse Effect	
0.0	.00	.000	No Effect	Developmental effects
0.1	.05	.003		
0.2	.10	.010	Small Effect	Teacher effects
0.3	.15	.022		
0.4	.2	.039		Zone of desired effects
0.5	.24	.060	Intermediate Effect	
0.6	.29	.083		
0.7	.33	.110		
0.8	.37	.140	Large Effect	
0.9	.41	.168		
≥ 1.0	.45	.200		

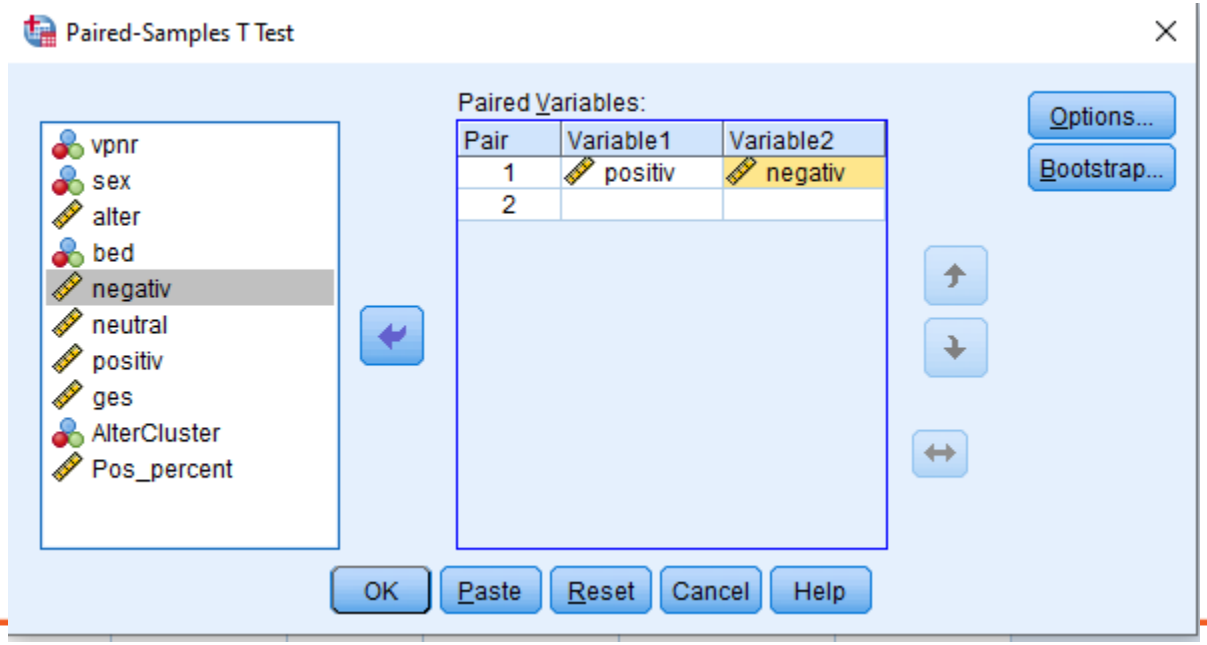
https://www.psychometrica.de/effect_size.html#cohen

Effect sizes – Reporting

- "There was a main effect for argumentation script on learning gains of knowledge about argument quality, $F(1,77) = 4.13$, $p = .046$, $\eta_p^2 = .05$, and a significant and strong main effect of subjective learning gains on argument quality, $F(1,77) = 11.99$, $p = .001$, $\eta_p^2 = .14$."
- "Helmert contrasts showed that the control was significantly better than the group awareness condition, $t(77) = 2.52$, $p = .014$, $d = .856$."

Exercise 1

- Open Beispieldatensatz
- Let’s assume that the (same) participants first studied positive adjectives, then were tested (t1), then studied negative adjectives, and were tested again (t2)
- Check with SPSS if the amount of remembered adjectives is different from t1 to t2
- Analyze > Compare Means > Paired Samples t-test



Exercise 1

Mode of testing	dependent *
Student t Value	0,462
n_1	150
n_2	
r	0,337
Effect Size *d*	0.043

https://www.psychometrica.de/effect_size.html#repeated

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	positiv	3,44	150	2,071	,169
	negativ	3,35	150	1,911	,156

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	positiv & negativ	150	,337	,000

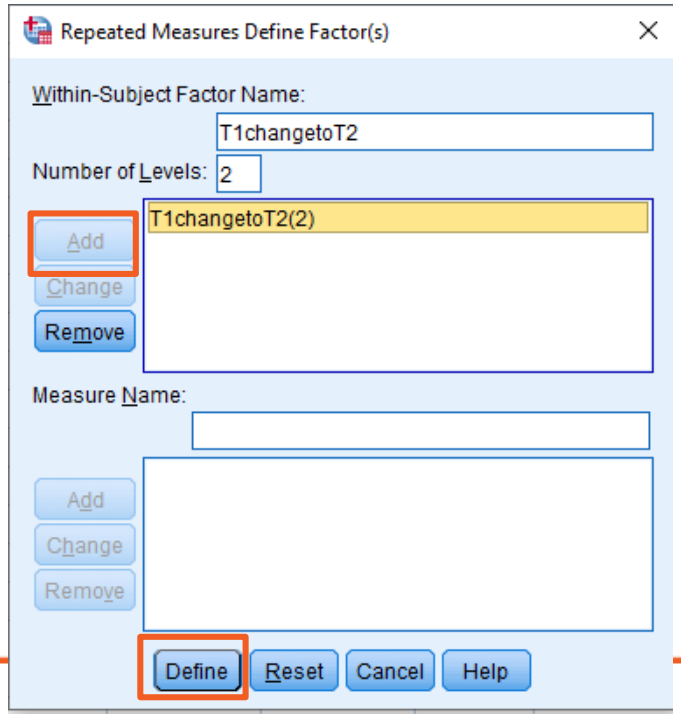
Paired Samples Test

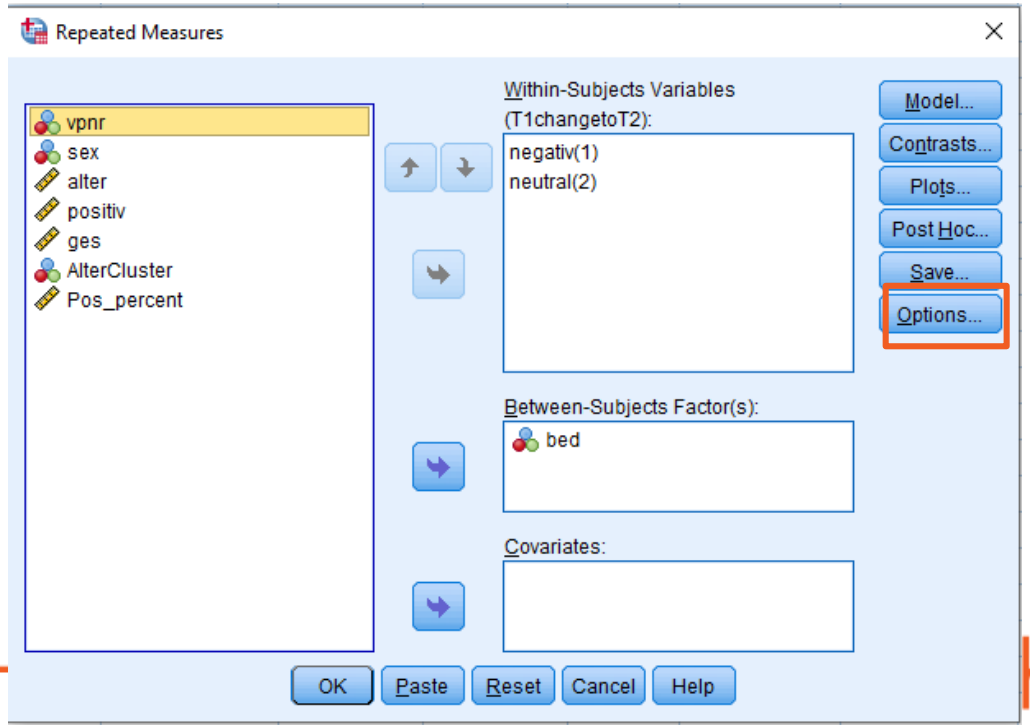
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	positiv - negativ	,087	2,297	,188	-,284	,457	,462	149	,645

There was no significant change from t1 to t2 (over time):
t (149) = .462, *p* = .645, *d* = .043

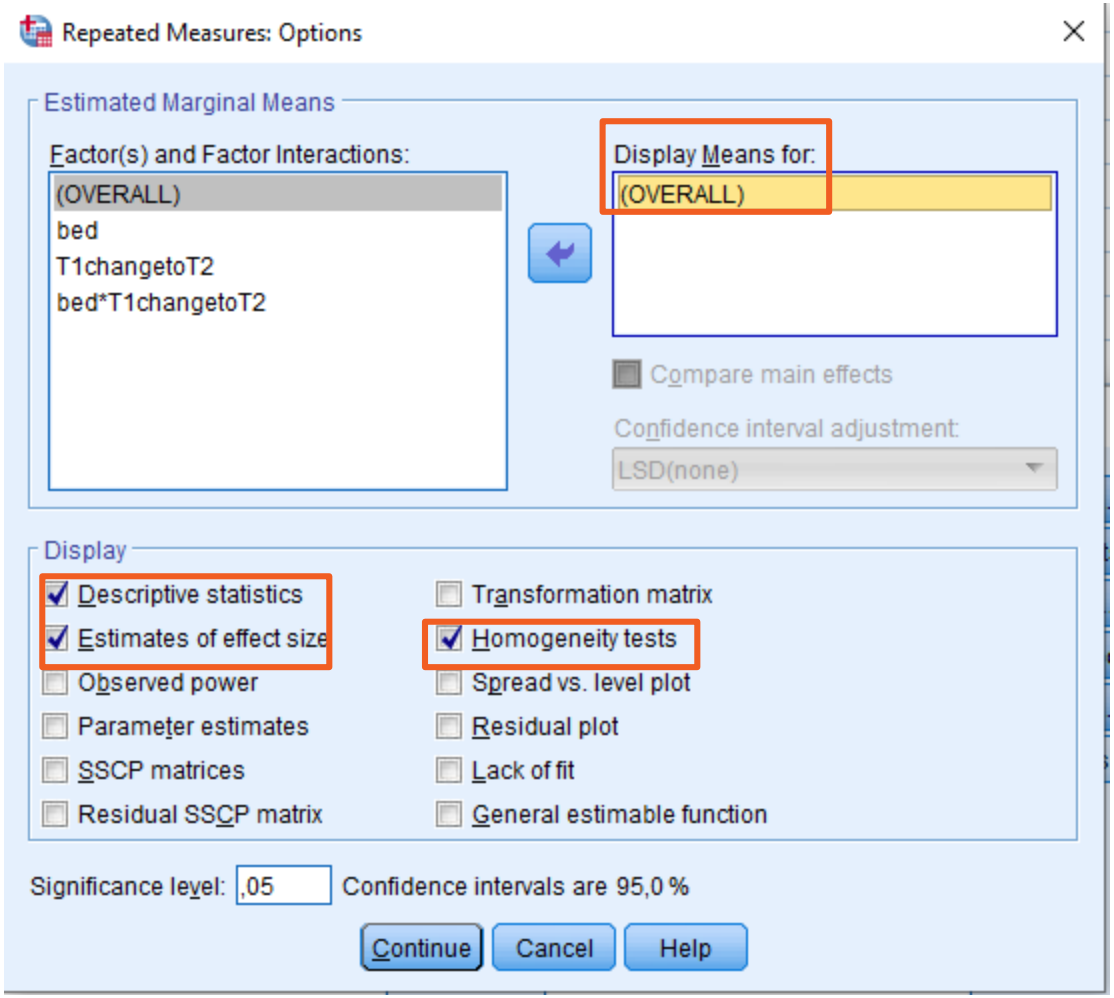
Exercise 2

- Check if a possible change from t1 (“negative”) to t2 (“neutral”) is dependent on the condition (‘bed’)
- Analyze > General Linear Model > Repeated Measures





Exercise 2


Repeated Measures: Options
Estimated Marginal Means
Factor(s) and Factor Interactions:
(OVERALL)
bed
T1changetoT2
bed*T1changetoT2
Display Means for:
(OVERALL)
Compare main effects
Confidence interval adjustment:
LSD(none)
Display
Descriptive statistics
Estimates of effect size
Observed power
Parameter estimates
SSCP matrices
Residual SSCP matrix
Transformation matrix
Homogeneity tests
Spread vs. level plot
Residual plot
Lack of fit
General estimable function
Significance level: ,05
Confidence intervals are 95,0 %
Continue
Cancel
Help

Exercise 2

Within-Subjects Factors

Measure: MEASURE_1

T1changetoT2	Dependent Variable
1	negativ
2	neutral

Between-Subjects Factors

		Value Label	N
bed	1	strukturell	50
	2	bildhaft	50
	3	emotional	50

Descriptive Statistics

	bed	Mean	Std. Deviation	N
negativ	strukturell	2,64	1,601	50
	bildhaft	3,74	1,904	50
	emotional	3,68	2,035	50
	Total	3,35	1,911	150
neutral	strukturell	1,96	1,428	50
	bildhaft	3,82	1,687	50
	emotional	4,06	2,064	50
	Total	3,28	1,973	150

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
T1changetoT2	Sphericity Assumed	,403	1	,403	,154	,696	,001
	Greenhouse-Geisser	,403	1,000	,403	,154	,696	,001
	Huynh-Feldt	,403	1,000	,403	,154	,696	,001
	Lower-bound	,403	1,000	,403	,154	,696	,001
T1changetoT2 * bed	Sphericity Assumed	14,927	2	7,463	2,841	,062	,037
	Greenhouse-Geisser	14,927	2,000	7,463	2,841	,062	,037
	Huynh-Feldt	14,927	2,000	7,463	2,841	,062	,037
	Lower-bound	14,927	2,000	7,463	2,841	,062	,037
Error(T1changetoT2)	Sphericity Assumed	386,170	147	2,627			
	Greenhouse-Geisser	386,170	147,000	2,627			
	Huynh-Feldt	386,170	147,000	2,627			
	Lower-bound	386,170	147,000	2,627			

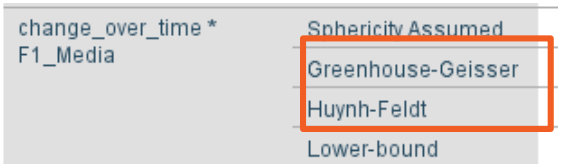
There was no significant interaction between the factor ‘bed’ and the point in time (t1 vs t2):

F (2, 147) = 2.841, p = .062, η_p^2 = .037

… repeated measures ANOVAs with multiple factors

- The SPSS output is huge, be careful to select the right tables
- The test for Sphericity* is only used when your within factor has more than two levels (rare)
- For now, you can ignore the results for multivariate tests and within subjects contrasts

* Sphericity: The variances of the differences between the within factor levels are homogeneous. If not, use a correction: Greenhousse-Geisser or Huynh-Feldt



Next week, 21.01.: SPSS workshop

Bring your laptops

In two weeks, 28.01.: Overall repetition

February 04: Final exam

Pen&paper, graded 1.0-4.0