

Session start question

<https://app.sli.do/event/kugjanfy>

Join at
slido.com
#80294





Empirical Research Methods 1

**Repeated measures tests: t-test with dependent samples, repeated measures ANOVA (within subjects design).
Effect size**

Miguel Rejón,

Room 1.14

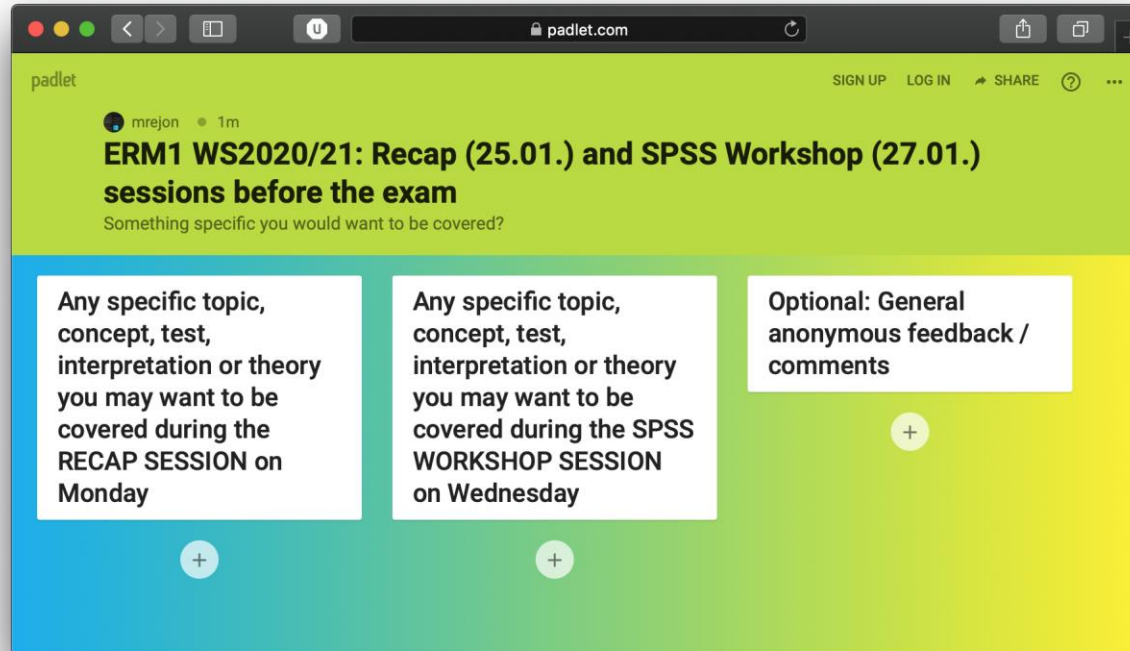
m.rejon@edutech.uni-saarland.de



Before we start 1/3: Info and tools

- ◇ Slido for session start question and anonymous Q&A
- ◇ MS Teams for comments, questions, etc
- ◇ SPSS for exercise livecast
- ◇ Last topic session and 2nd to last Monday session before the exam
 - ◇ Next week: recap week
 - ◇ Next next week: no sessions
 - ◇ Nex next next week: Exam! (08.02.)

Before we start 2/3: recap sessions next week



<https://padlet.com/mrejon/y71gbr4qt5yx720d>

Contribute until Wednesday Jan 20

Add (“+” button), comment, or like posts

Before we start 3/3: Important info about the exam!



Miguel Angel Rejon Zamudio Wednesday 5:37 PM



ERM1 exam: 08.02.2021

hey everyone, just to let you know that I got confirmation about the date for the in-person exam. Here are the details:

Day: Monday February 8, 2021

Time: from 9:50 to 11:40

Location: Raum Hörsaal I, Room number 0.14 in building B3.1

Only pen, pencil and eraser are required and allowed.

Don't forget to bring your facemask!

If you won't be able to attend, please let me know ASAP!

[See less](#)



Miguel Angel Rejon Zamudio 11:44 AM

Important additional information about the exam

I know the exam will take place until February, but here's some important info that you can already know

-Don't forget to register to the exam on the LSF. You have until 05.02. to register. If you don't know how to, Claudio did a very good screenshot tutorial of how to do it. The file is called "Apply_Exam_Lsf" and is in the Files tab of the EduTech official Team

-Given that the exam will be in-person here at the uni, we need to follow the following indications:

+Symptoms: please check the "FragencheckPräsenz" pdf in the Files tab of this Team and ensure you don't have any of those symptoms one day before the exam day

+Face mask: you need to always have your face mask covering your mouth and nose, even when you're outside of the building. Only after being seated you can take it off

+Distance: we need to keep a distance of 1.5m between each other. If you arrive early to the room, avoid close congregations. You can be there and talk but with enough distance between each other. Same case after the exam, if you want to hang out outside of the classroom, try to do it in small groups and with enough distance

If you have questions, let me know.

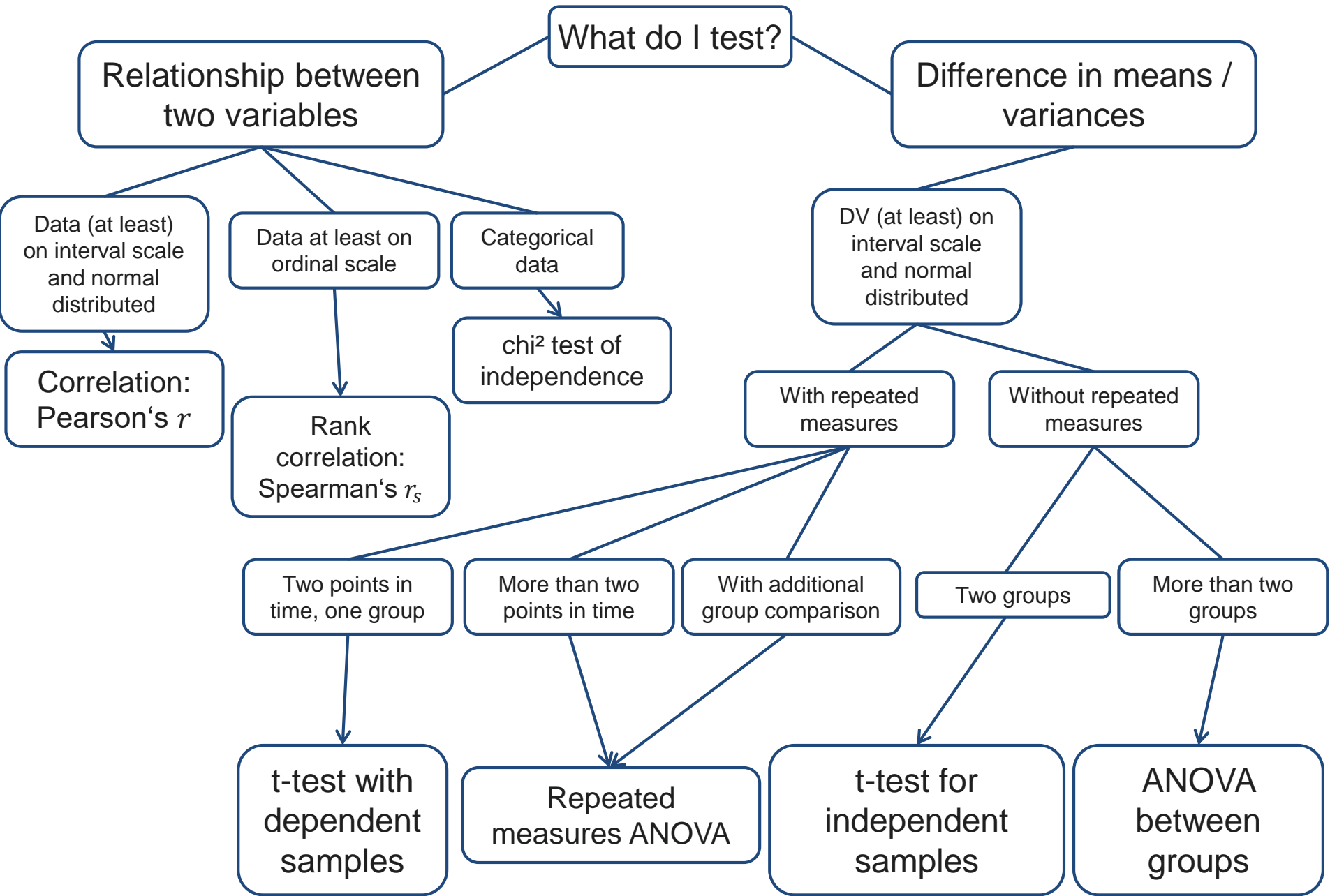
Enjoy the snowy weekend!

[See less](#)

← Reply

Agenda

- ◇ Updated “What do I test?”
- ◇ Synonyms for “Within subjects”
- ◇ Within vs between subjects designs
- ◇ Effect size: definition; for t-tests; for ANOVAS
- ◇ Today’s topics on SPSS



Within subjects, repeated measures, paired (samples), dependent samples. They all mean the same

- ◇ Both for ANOVAs and t-tests
- ◇ Same participants, individuals or objects are tested more than once. This can mean one of 2 things:
 1. Same participants are assigned to only 1 level but are measured (in terms of the DV) more than once. A.k.a. Pre- post-test study. E.g., Attitude change before and after attending a workshop.
 2. Same participants are assigned to more than 1 level and are measured (in terms of the DV) once for each level. Measurement is normally done immediately after exposure to each level. A.k.a. Within-subjects study. E.g., Task completion using 3 different apps.
- ◇ **You can them both:** You can have independent samples and dependent samples analyses on your same data; provided that the experiment is well designed and the RQQ and hypotheses are clearly defined.

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)

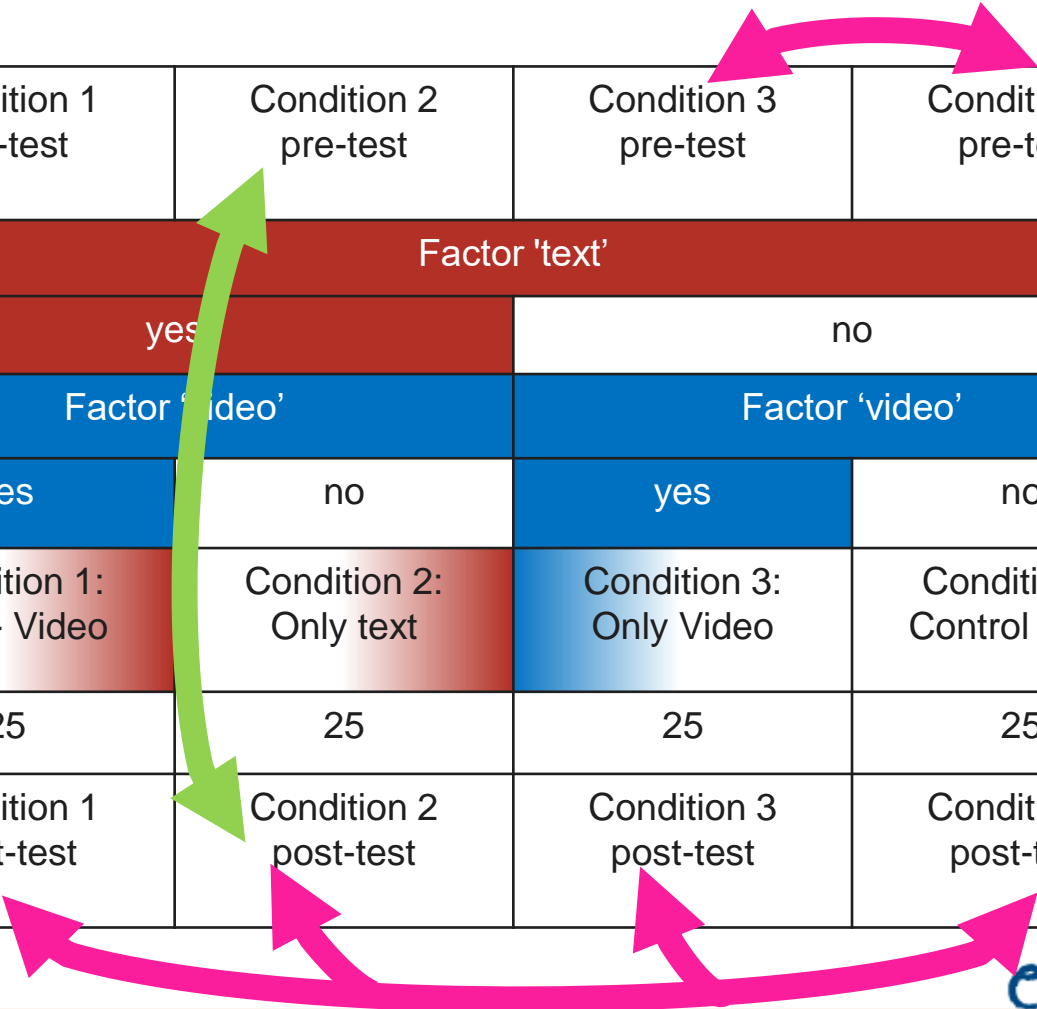
n	25
IV	Condition 'text'
DV1	Knowledge test
IV	Condition 'video'
DV2	Knowledge test

Within or between? You can have them both

Pre- post-test experimental design with a 2x2 factorial design.

Within comparisons. **Between comparisons**

DV	Condition 1 pre-test	Condition 2 pre-test	Condition 3 pre-test	Condition 4 pre-test
IV	Factor 'text'			
	yes		no	
	Factor 'video'		Factor 'video'	
	yes	no	yes	no
Exp. groups	Condition 1: Text + Video	Condition 2: Only text	Condition 3: Only Video	Condition 4: Control group
n	25	25	25	25
DV	Condition 1 post-test	Condition 2 post-test	Condition 3 post-test	Condition 4 post-test



Effect size – Why?

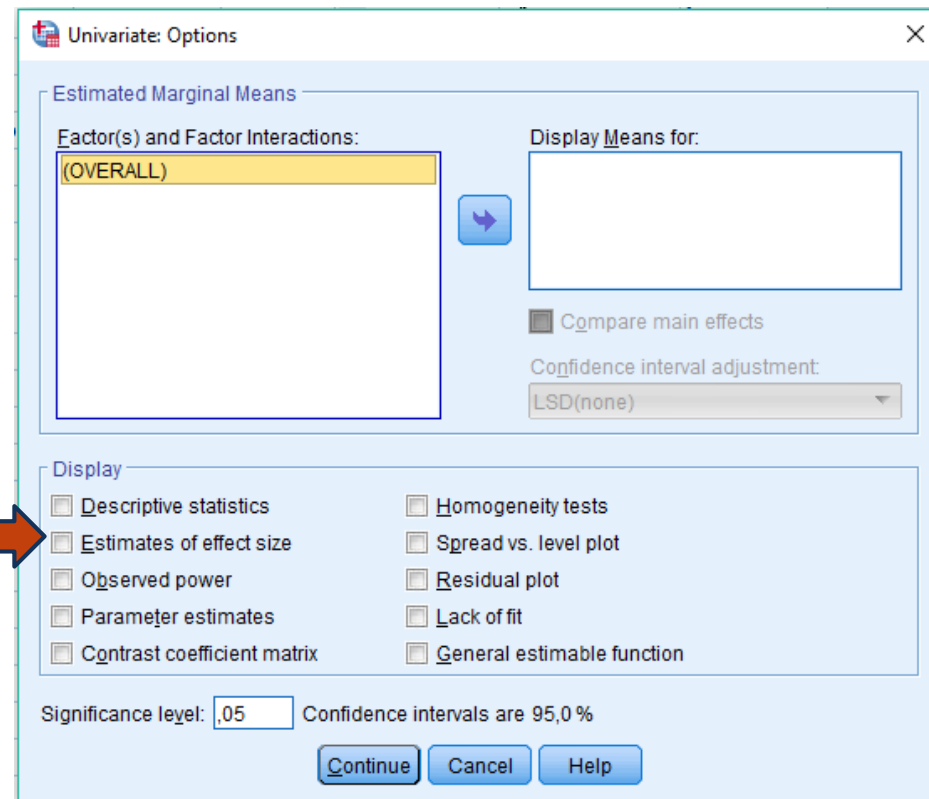
- Only telling about the ‘significance’ is not enough: Significant results are also possible for practically meaningless differences
 - Additionally, report the *effect size*
 - How large is a significant difference?
 - How strong is an effect? How to compare it with effects found in other studies/ data?

Effect sizes – two important measures



- **Cohen's d :**
 - For t-tests
 - Available in the newest versions of SPSS but not selectable in older versions → Use an internet calculator, e.g.:
https://www.psychometrica.de/effect_size.html

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 – Interpretation

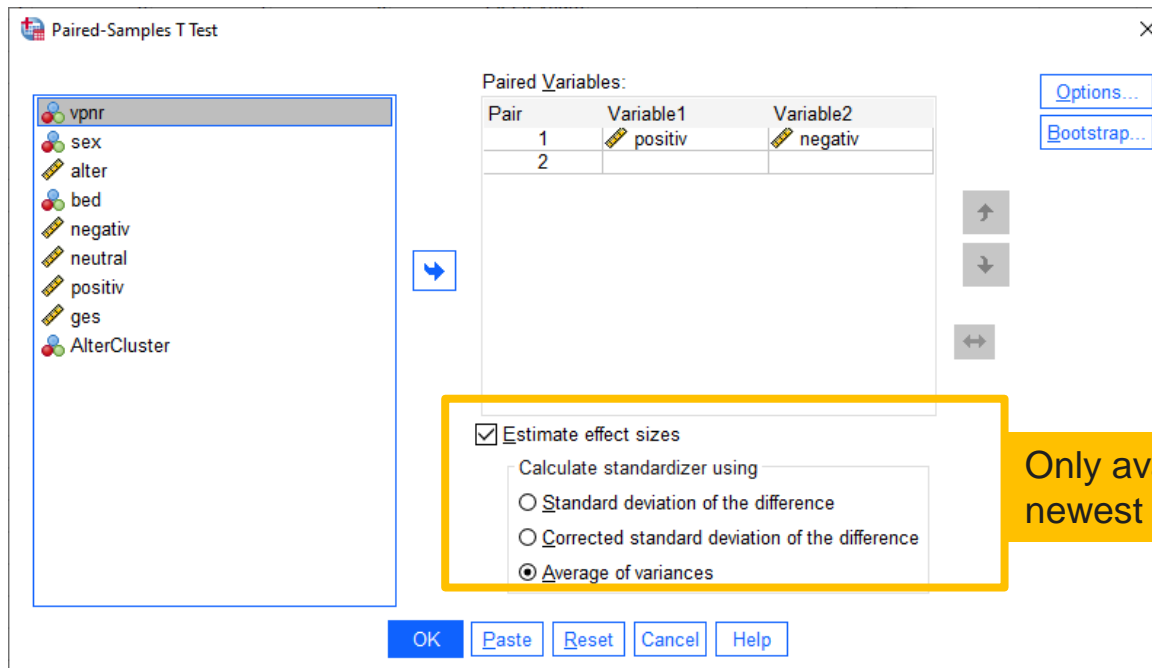
 d	r[*]	 η²	Interpretation sensu Cohen (1988)	Interpretation sensu Hattie (2007)
< 0	< 0	-	Adverse Effect	
0.0	.00	.000	No Effect	Developmental effects
0.1	.05	.003		
0.2	.10	.010	Small Effect	Teacher effects
0.3	.15	.022		
0.4	.2	.039		
0.5	.24	.060	Intermediate Effect	Zone of desired effects
0.6	.29	.083		
0.7	.33	.110		
0.8	.37	.140	Large Effect	
0.9	.41	.168		
≥ 1.0	.45	.200		

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. Paired samples t-test

- ◇ Open Beispieldatensatz_S10
- ◇ 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



Only available on the newest versions of SPSS

Exercise 1 . Paired samples t-test

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

Mode of testing	dependent * ▾
Student t Value	0,462
n ₁	150
n ₂	
r	0,337
Effect Size <i>d</i>	0.043

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

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

Exercise 1 . Paired samples t-test

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

Mode of testing	dependent * ▾
Student t Value	0,462
n ₁	150
n ₂	
r	0,337
Effect Size <i>d</i>	0.043

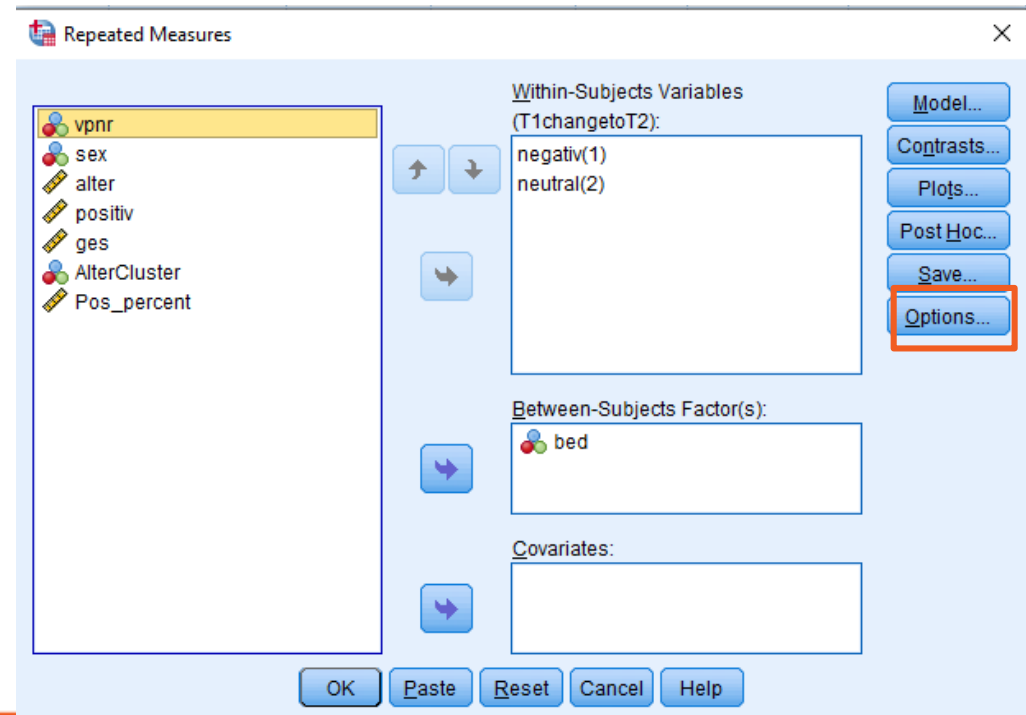
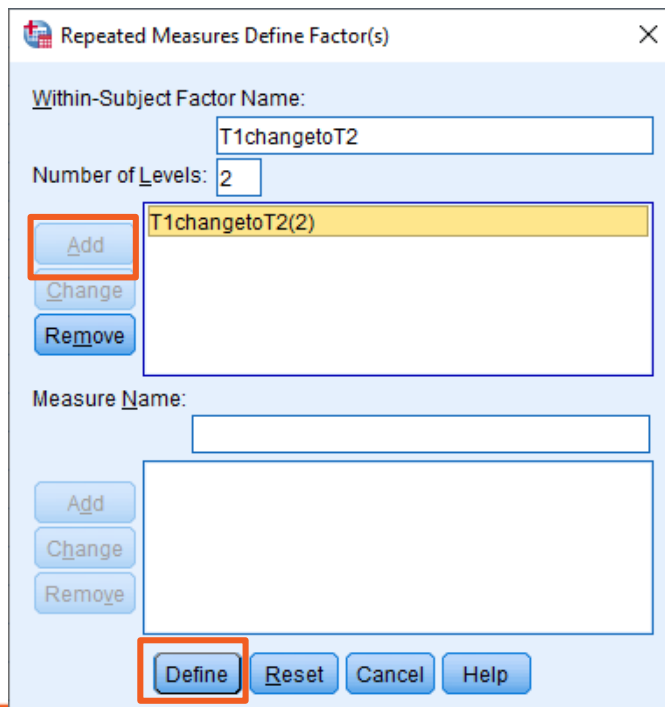
Paired Samples Effect Sizes						
			Standardizer ^a	Point Estimate	95% Confidence Interval	
					Lower	Upper
Pair 1	erinnerte positive Adjektive - erinnerte negative Adjektive	Cohen's d	1.993	.043	-.117	.204
		Hedges' correction	1.998	.043	-.116	.203

a. The denominator used in estimating the effect sizes.
 Cohen's d uses the square root of the average variance of measures.
 Hedges' correction uses the square root of the average variance of measures, plus a correction factor.

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

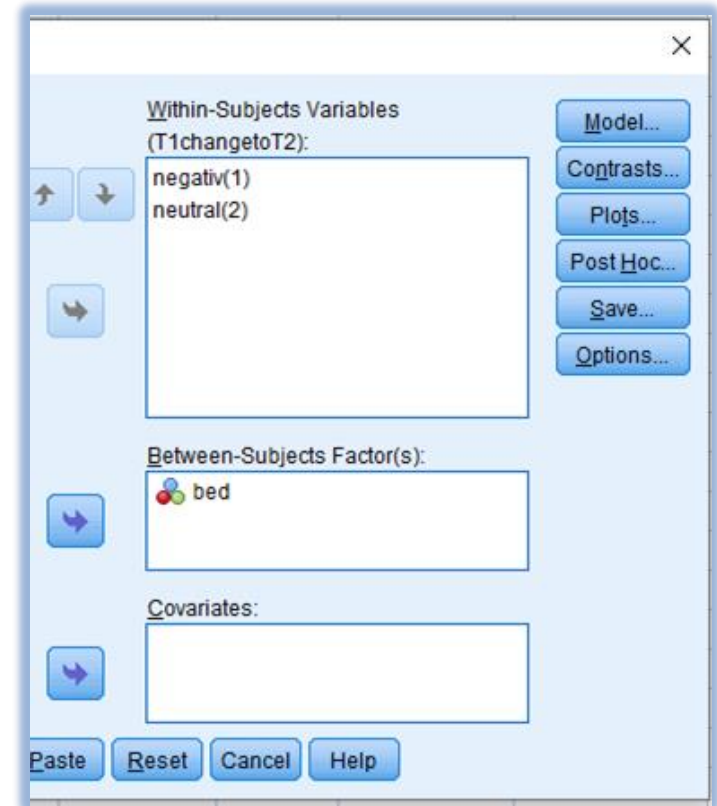
Exercise 2. Repeated measures ANOVA with independent samples analysis

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



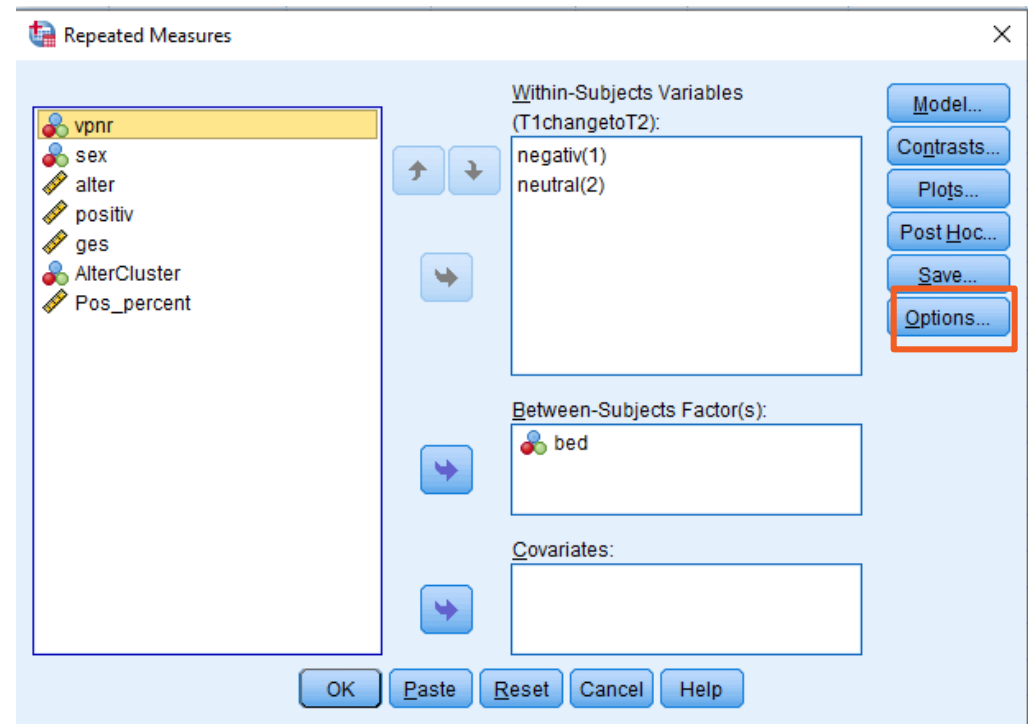
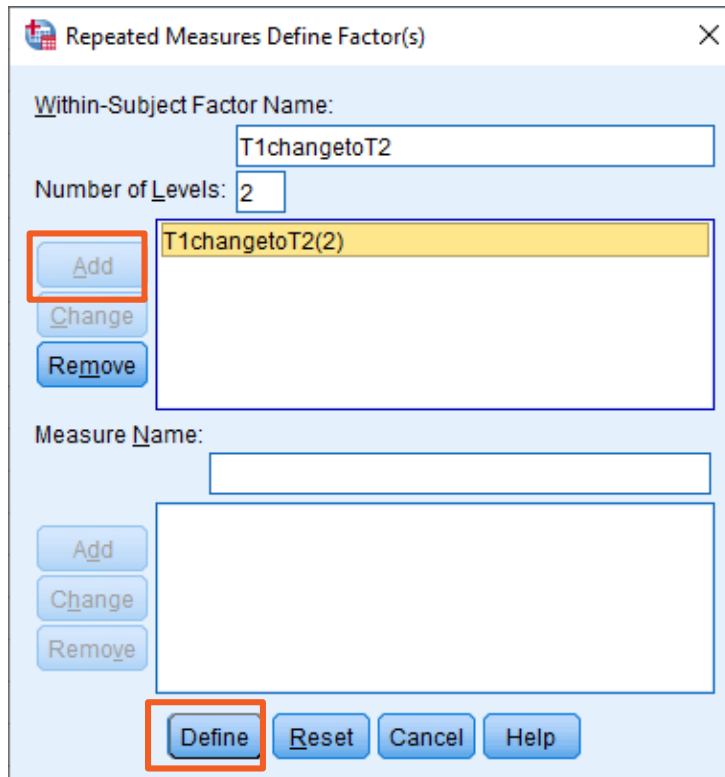
Exercise 2. Repeated measures ANOVA with independent samples analysis

- ◇ Why repeated measures? Because we will compare 2 measurements across the same data (“negative” and “neutral”)
- ◇ Why independent samples? Because we want to see the impact of the condition (“bed”) on the measured data

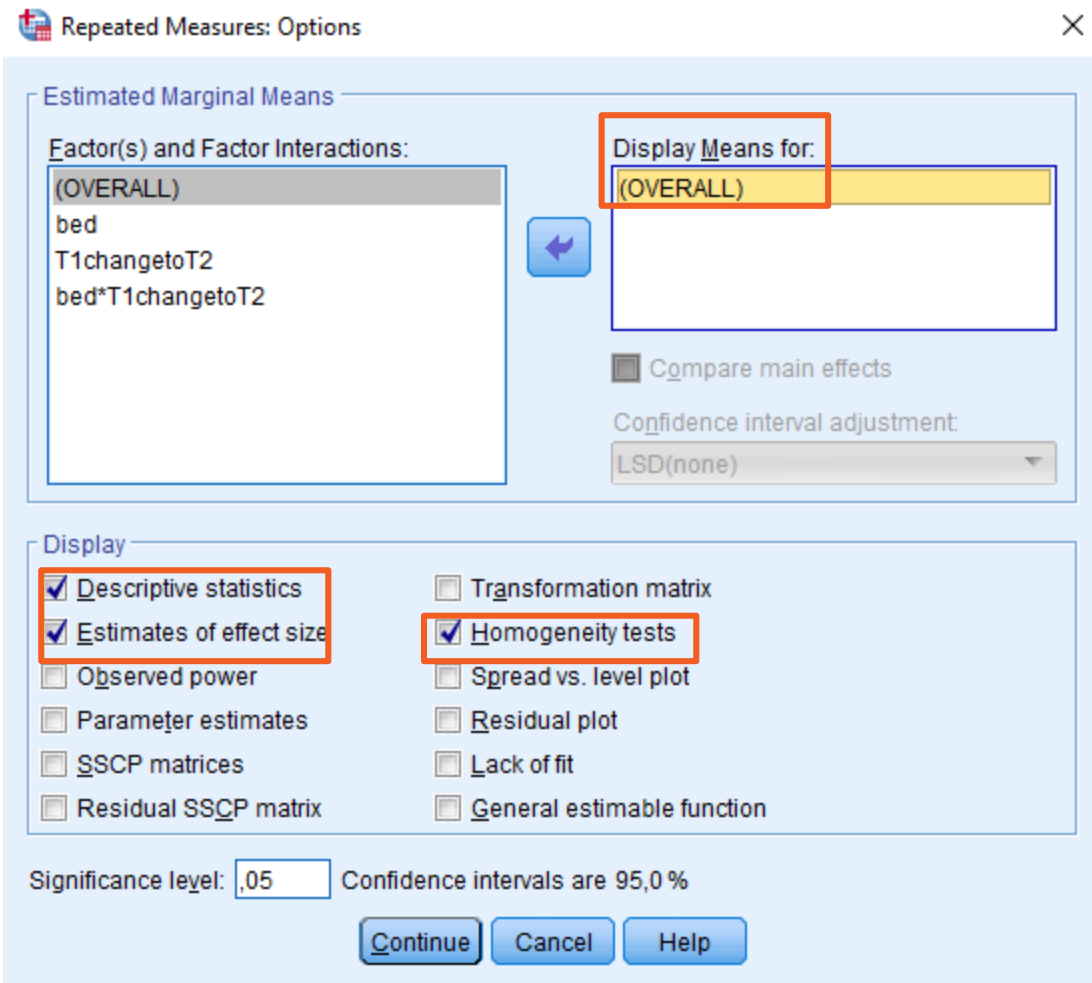


Exercise 2. Repeated measures ANOVA with independent samples analysis

- ◇ Analyze > General Linear Model > Repeated Measures



Exercise 2. Repeated measures ANOVA with independent samples analysis



The image shows the 'Repeated Measures: Options' dialog box in SPSS. The 'Estimated Marginal Means' section has a list of factors on the left: '(OVERALL)', 'bed', 'T1changetoT2', and 'bed*T1changetoT2'. A blue arrow button points from this list to a 'Display Means for:' list on the right, which contains '(OVERALL)'. Below this, there is a checkbox for 'Compare main effects' and a dropdown for 'Confidence interval adjustment' set to 'LSD(none)'. The 'Display' section contains several checkboxes: 'Descriptive statistics' (checked), 'Estimates of effect size' (checked), 'Observed power' (unchecked), 'Parameter estimates' (unchecked), 'SSCP matrices' (unchecked), 'Residual SSCP matrix' (unchecked), 'Transformation matrix' (unchecked), 'Homogeneity tests' (checked), 'Spread vs. level plot' (unchecked), 'Residual plot' (unchecked), 'Lack of fit' (unchecked), and 'General estimable function' (unchecked). At the bottom, the 'Significance level' is set to '.05' and 'Confidence intervals are 95,0 %'. There are 'Continue', 'Cancel', and 'Help' buttons at the bottom.

Note: On the latest versions of SPSS, this step is now splitted and Estimated Marginal Means has its own button

Exercise 2. Repeated measures ANOVA with independent samples analysis

Within-Subjects Factors

Measure: MEASURE_1 ← DV

T1changetoT2		Dependent Variable
1		negativ
2		neutral

← Groups of the IV

Between-Subjects Factors

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

← Levels of our factor for the indep. samples analysis part

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

... 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 factor (i.e. the “within-subjects variable”) 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: Greenhouse-Geisser or Huynh-Feldt

change_over_time * F1_Media	Sphericity Assumed
	Greenhouse-Geisser
	Huynh-Feldt
	Lower-bound

Exercise 2. Repeated measures ANOVA with independent samples analysis

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
T1 changetoT2	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
T1 changetoT2 * 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(T1 changetoT2)	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):

$$F(2, 147) = 2.841, p = .062, \eta_p^2 = .037$$