

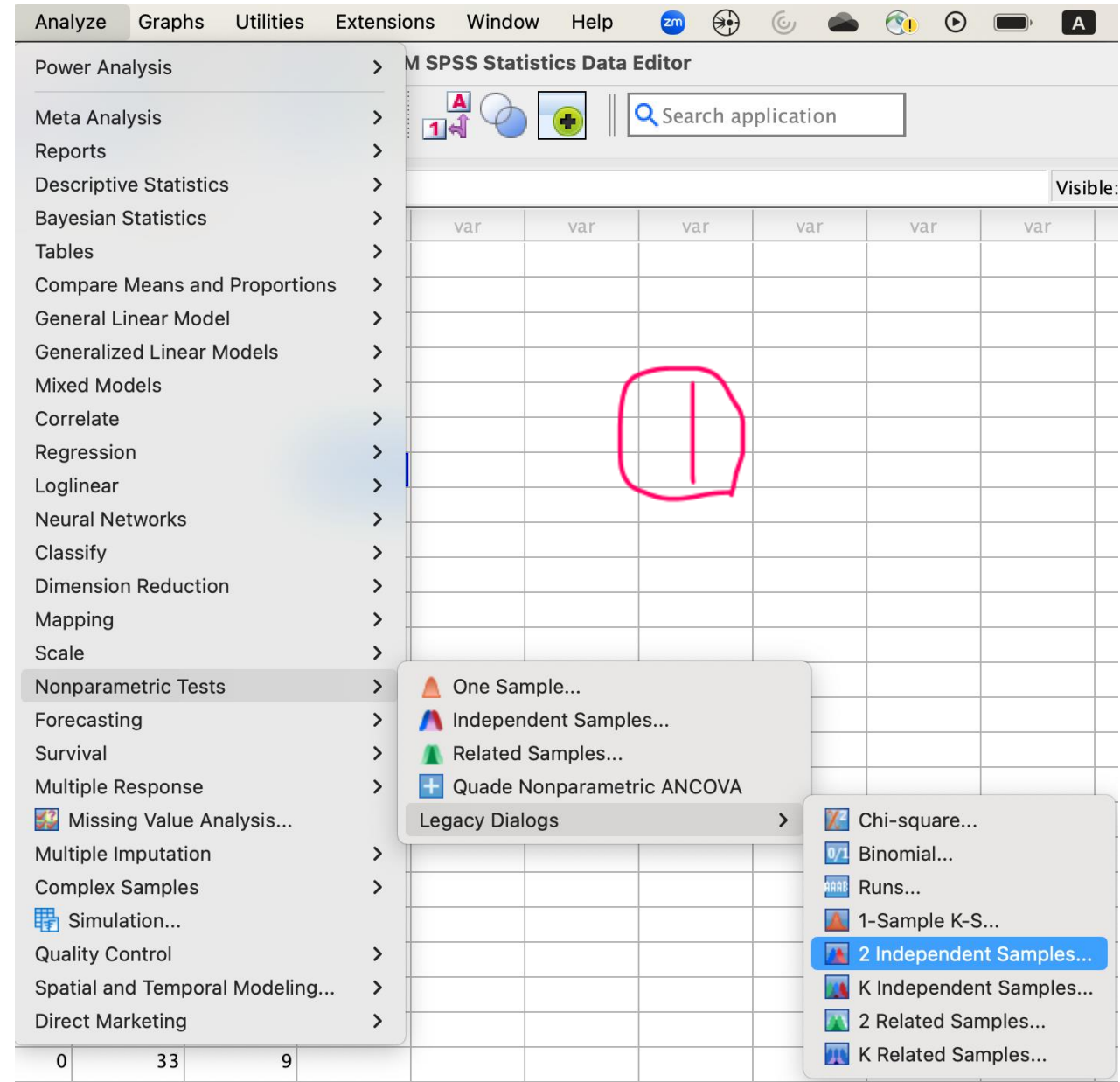
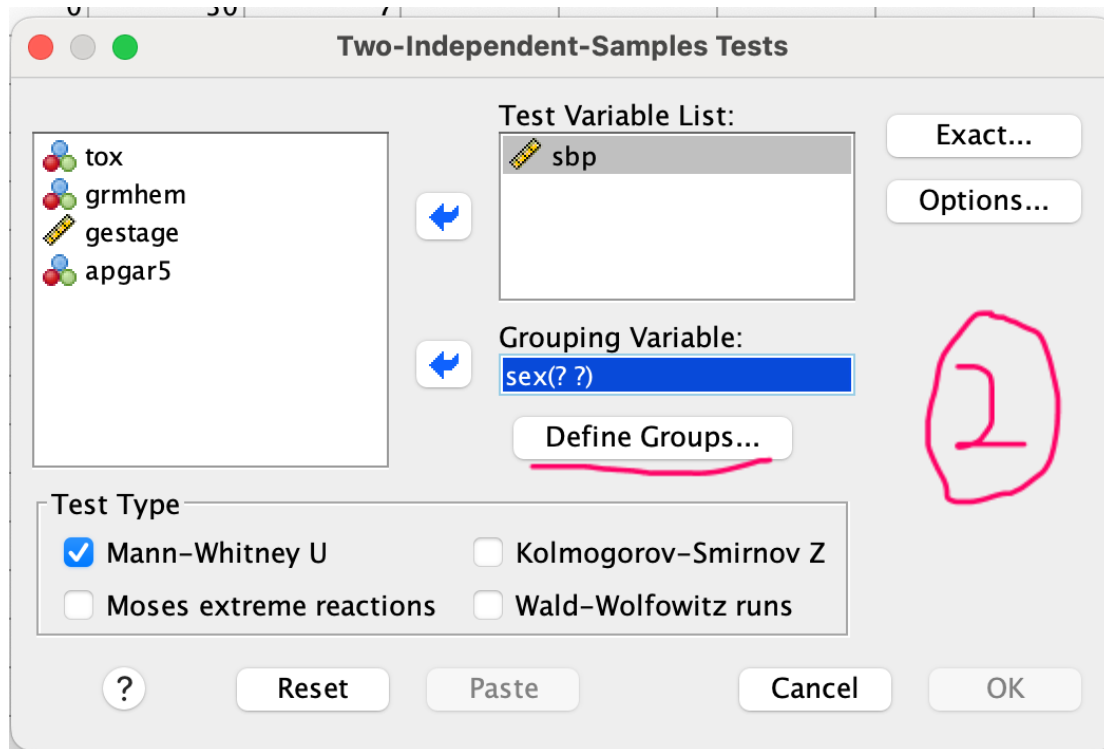
# Hypothesis testing 1-2

Mann Whitney U-test, paired t-test

# Mann-Whitney U test

- Also known as the **Wilcoxon rank-sum test**
- A non-parametric test used to compare two independent groups when the data is ordinal or continuous.
- Let's again check if systolic blood pressure (sbp) level differs between boys and girls (sex).

# Mann-Whitney U test



# Mann-Whitney U test

sex	tox	grmhem	gestage	apgar5	var	var	var
1	0	0	29	7			
1	0	0	31	8			
0	0	0	33	0			
0	0	0	31	8			
0	1	0	30	7			
1	0						
1	1						
0	0						
0	0						
0	1						
0	0						
0	0						
0	0						
0	0						
0	0						
1	0						
1	0						

Two-Independent-Samples Tests

Test Variable List:

sbp

Two Independent Samples: Define Gro...

Group 1: 0

Group 2: 1

Test Type

?

Cancel

Continue

If groups are coded differently (ex. 1 vs. 2),  
then you have to enter 1 and 2, respectively.

# Compare with independent samples t-test

➔ NPar Tests

## Mann-Whitney Test

T-Test

### Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
sbp	0	56	46.46	11.145	1.489
	1	44	47.86	11.806	1.780

### Ranks

	sex	N	Mean Rank	Sum of Ranks
sbp	0	56	50.06	2803.50
	1	44	51.06	2246.50
	Total	100		

### Test Statistics<sup>a</sup>

	sbp
Mann-Whitney U	1207.500
Wilcoxon W	2803.500
Z	-.170
Asymp. Sig. (2-tailed)	.865

a. Grouping Variable: sex

$P > 0.05$ ... what's the null hypothesis?

## Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Significance One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference
sbp	Equal variances assumed	.079	.779	-.607	98	.273	.545	-1.399	2.305
	Equal variances not assumed			-.603	89.858	.274	.548	-1.399	2.321

# Reporting results

## Example statement of results [\[ edit \]](#)

In reporting the results of a Mann–Whitney  $U$  test, it is important to state:<sup>[12]</sup>

- A measure of the central tendencies of the two groups (means or medians; since the Mann–Whitney  $U$  test is an ordinal test, medians are usually recommended)
- The value of  $U$  (perhaps with some measure of effect size, such as [common language effect size](#) or [rank-biserial correlation](#)).
- The sample sizes
- The significance level.

In practice some of this information may already have been supplied and common sense should be used in deciding whether to repeat it. A typical report might run,

"Median latencies in groups E and C were 153 and 247 ms; the distributions in the two groups differed significantly (Mann–Whitney  $U = 10.5$ ,  $n_1 = n_2 = 8$ ,  $P < 0.05$  two-tailed)."

# Reporting results




- Please write a sentence reporting the results of the Mann-Whitney U test demonstrated in previous slides.










# Paired t-test

- This is essentially the same as one-sample t-test which is testing the null hypothesis that the population mean *difference* is zero.
- We will use sleep data for this task (sleep.csv).
  - Compares reaction times before and after sleep deprivation for 10 subjects. Reaction time measured before and after sleep deprivation in milliseconds (ms).



# Load the dataset

	 extra	 group	 ID
1	.7	1	1
2	-1.6	1	2
3	-.2	1	3
4	-1.2	1	4
5	-.1	1	5
6	3.4	1	6
7	3.7	1	7
8	.8	1	8
9	.0	1	9
10	2.0	1	10
11	1.9	2	1
12	.8	2	2
13	1.1	2	3
14	.1	2	4
15	-.1	2	5
16	4.4	2	6
17	5.5	2	7
18	1.6	2	8
19	4.6	2	9
20	3.4	2	10

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	extra	Numeric	4	1		None	None	8	 Right	 Scale	 Input
2	group	Numeric	1	0		None	None	8	 Right	 Nominal	 Input
3	ID	Numeric	2	0		None	None	8	 Right	 Nominal	 Input

# Explore the data

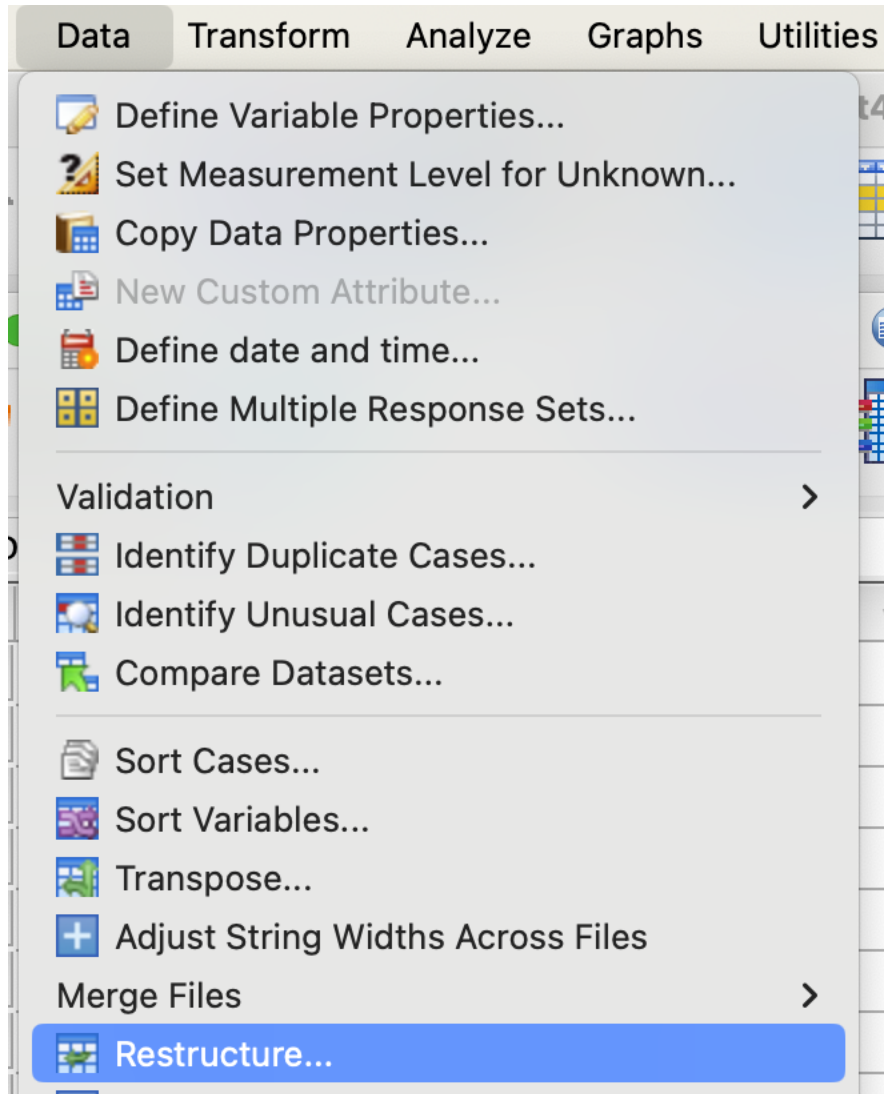
- 1. What are the sample means and standard deviations of the reaction time?

	Group 1	Group 2
Sample mean		
Sample standard deviation		

# Explore the data

- 2. Can you check if the samples can be assumed to follow normal distributions? What are your supporting materials? Use plots (ex. histogram, boxplot, Q-Q plot) or test results (ex. Shapiro-Wilk, Kolmogorov-Smirnov) to support your claim.

# Restructure your data

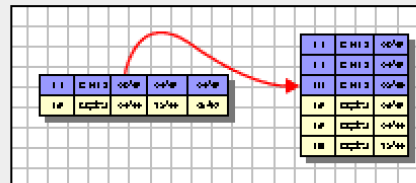


## Welcome to the Restructure Data Wizard!

This wizard helps you to restructure your data from multiple variables (columns) in a single case to groups of related cases (rows) or vice versa, or you can choose to transpose your data.



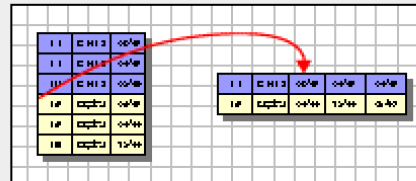
The wizard replaces the current data set with the restructured data. Note that data restructuring cannot be undone.



What do you want to do?

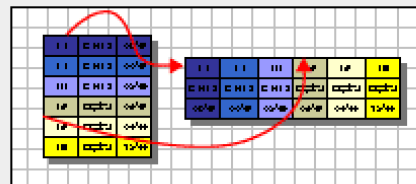
- ☐ Restructure selected variables into cases

Use this when each case in your current data has some variables that you would like to rearrange into groups of related cases in the new data set.



- ☒ Restructure selected cases into variables

Use this when you have groups of related cases that you want to rearrange so that data from each group are represented as a single case in the new data set.



- ☐ Transpose all data

All cases will become variables and selected variables will become cases in the new data set. (Choosing this option will end the wizard, and the Transpose dialog will appear.)

# Restructure your data

## Cases to Variables: Select Variables

Data from case groups in the current file will be restructured into single cases in the new file.

Choose variables that identify case groups by moving those variables into the Identifier Variable list. Optionally you can also choose Index Variables.

- i** The variables that remain in the list of Variables in the Current File either contain data that vary within a case group or data that do not vary.  
A variable with data that vary will become a group of new variables in the restructured file. A variable with data that do not vary will be copied into the new file.

Variables in the Current File:

extra



Identifier Variable(s):

ID



Index Variable(s):

group

## Cases to Variables: Sorting Data

The variables that you used to identify case groups in the current file need to be sorted before the file can be restructured. If you are not sure about your data, select "Yes".

	1	2	3	
2	1	3	.006	
3	1	1	.010	
1	1	1	.003	
2	1	1	.008	
2	1	2	.007	
1	1	2	.004	
1	1	3	.002	

	1	2	3	
1	1	1	.003	
1	1	2	.004	
1	1	3	.002	
2	1	1	.008	
2	1	2	.007	
2	1	3	.006	
3	1	1	.010	

Sort the current data?

☒ Yes – data will be sorted by the Identifier and Index variabl...

☐ No – use the data as currently sorted

# Restructure your data

## Cases to Variables: Options

In this step you can set options that will be applied to the restructured data file.

### Order of New Variable Groups

- ☐ Group by original variable (for example: w1 w2 w3, h1 h2 h3)
- ☒ Group by index (for example: w1 h1, w2 h2, w3 h3)

**Either fine**

### Case Count Variable

- ☐ Count the number of cases in the current data used to create a new case

Name:

Label:




### Indicator Variables

- ☐ Create indicator variables

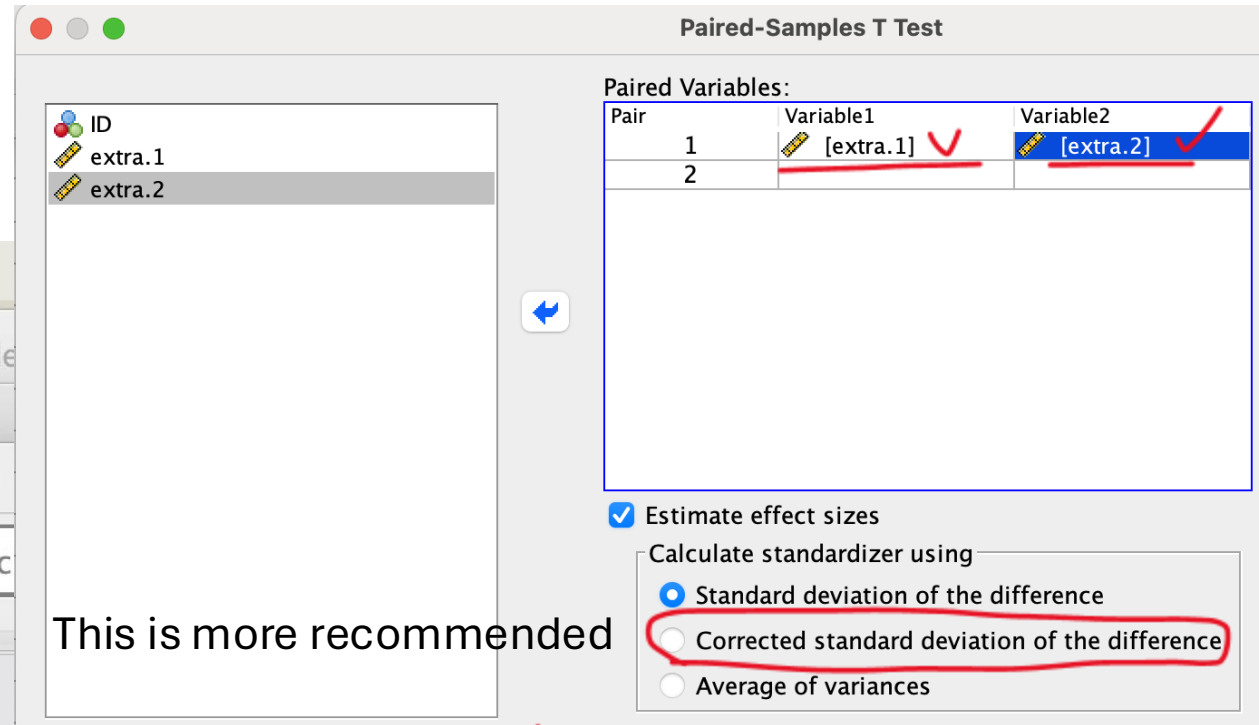
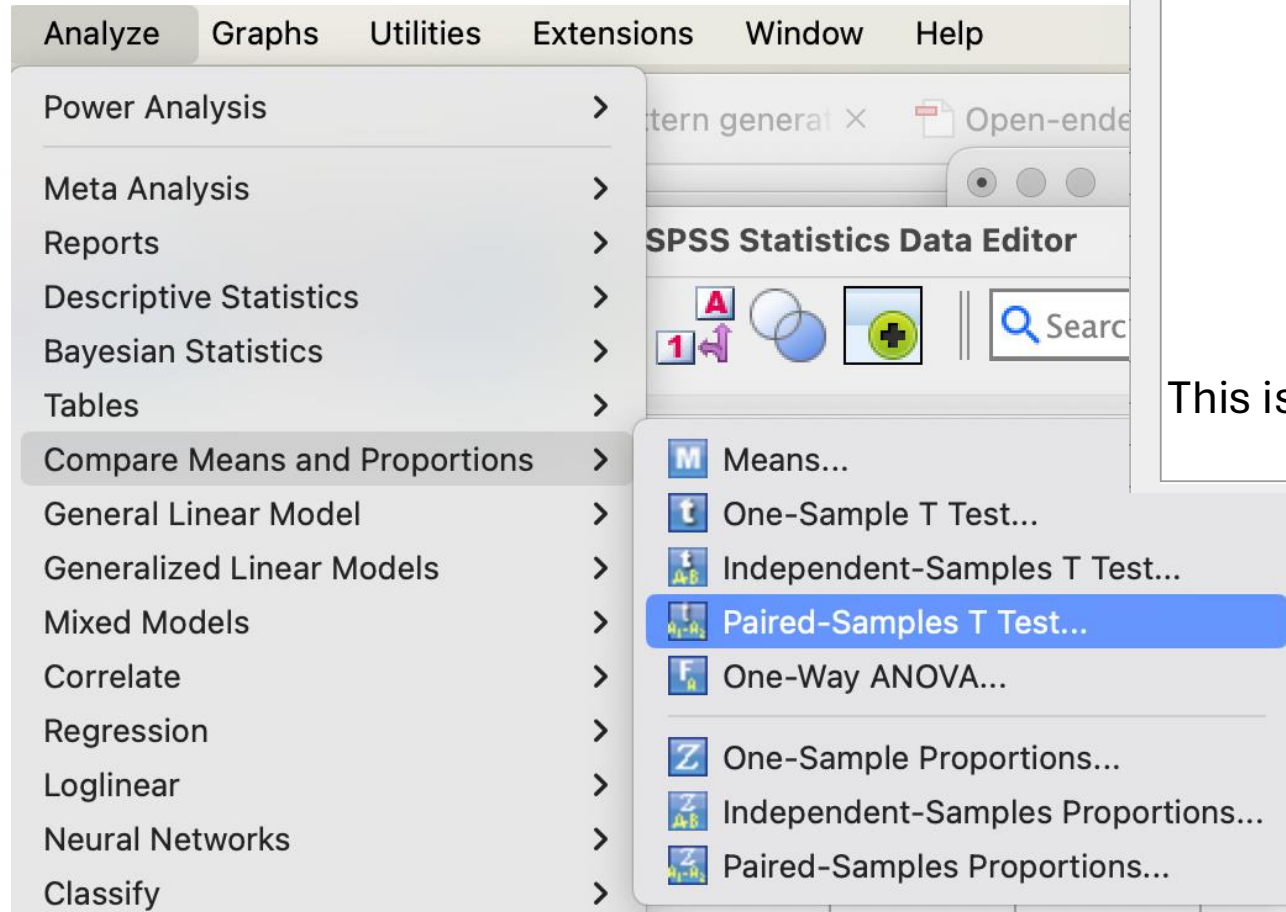
Root Name:

ind

## Final output

	 ID	 extra.1	 extra.2
2	2	-1.6	.8
3	3	-.2	1.1
4	4	-1.2	.1
5	5	-.1	-.1
6	6	3.4	4.4
7	7	3.7	5.5
8	8	.8	1.6
9	9	.0	4.6
10	10	2.0	3.4

# Carry out paired t-test



# Understand your results

## T-Test

[DataSet7] /Users/joh/Downloads/Untitled4.sav

### Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	extra.1	.750	10	1.7890	.5657
	extra.2	2.330	10	2.0022	.6332

Are they the same as what you reported in **slide 4**?

### Paired Samples Correlations

		N	Correlation	Significance One-Sided p	Two-Sided p
Pair 1	extra.1 & extra.2	10	.795	.003	.006

Pearson's r

I don't know why SPSS spits One-Sided p. Check the Two-Sided p.

### Paired Samples Test

		Paired Differences		95% Confidence Interval of the Difference		Significance	
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	
Pair 1	extra.1 - extra.2	-1.5800	1.2300	.3890	-2.4599	-.7001	
							t
							df
							One-Sided p
							Two-Sided p

This is what you report w.r.t. t-test

### Paired Samples Effect Sizes

		Standardizer <sup>a</sup>	Point Estimate	95% Confidence Interval	
				Lower	Upper
Pair 1	extra.1 - extra.2	Cohen's d	1.9217	-.822	-1.456
		Hedges' correction	2.1028	-.751	-1.172

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

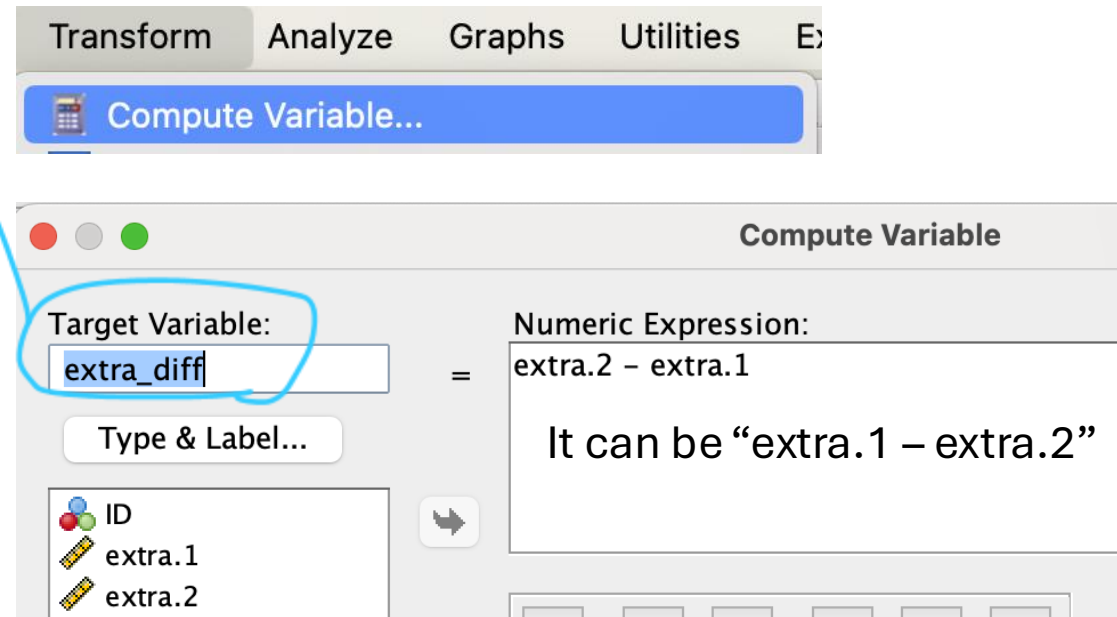
Again, better report positive values (ex. point estimate: 0.822, 95% CI: 0.188 ~ 1.456)

		Standardizer <sup>a</sup>	Point Estimate	95% Confidence Interval	
				Lower	Upper
Pair 1	extra.2 - extra.1	Cohen's d	1.9217	.822	1.456
		Hedges' correction	2.1028	.751	1.331



# Replicate results using one sample t-test?

14 : extra_diff				
	ID	extra.1	extra.2	extra_diff
1	1	.7	1.9	1.20
2	2	-1.6	.8	2.40
3	3	-.2	1.1	1.30
4	4	-1.2	.1	1.30
5	5	-.1	-.1	.00
6	6	3.4	4.4	1.00
7	7	3.7	5.5	1.80
8	8	.8	1.6	.80
9	9	.0	4.6	4.60
10	10	2.0	3.4	1.40



# Replicate results using one sample t-test?

14 : extra\_diff

	ID	extra.1	extra.2	extra_diff
1	1	.7	1.9	1.20
2	2	-1.6		
3	3	-.2		
4	4	-1.2		
5	5	-.1		
6	6	3.4		
7	7	3.7		
8	8	.8		
9	9	.0		
10	10	2.0		

One-Sample T Test

Test Variable(s):  
extra\_diff

Test Value: 0 ☒ Estimate effect sizes

Options...  
Bootstrap...

Reset Paste

If you set this value other than 0, what does that mean?

# Replicate results using one sample t-test?

## T-Test

### One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
extra_diff	10	1.5800	1.23000	.38896

### One-Sample Test

Test Value = 0

	t	df	Significance		Mean Difference	95% Confidence Interval of the Difference	
			One-Sided p	Two-Sided p		Lower	Upper
extra_diff	4.062	9	.001	.003	1.58000	.7001	2.4599

### One-Sample Effect Sizes

		Standardizer <sup>a</sup>	Point Estimate	95% Confidence Interval	
				Lower	Upper
extra_diff	Cohen's d	1.23000	1.285	.415	2.118
	Hedges' correction	1.34591	1.174	.379	1.936

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

### Paired Samples Test

		Paired Differences			95% Confidence Interval of the Difference		Significance			
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	extra.2 - extra.1	1.5800	1.2300	.3890	.7001	2.4599	4.062	9	.001	.003

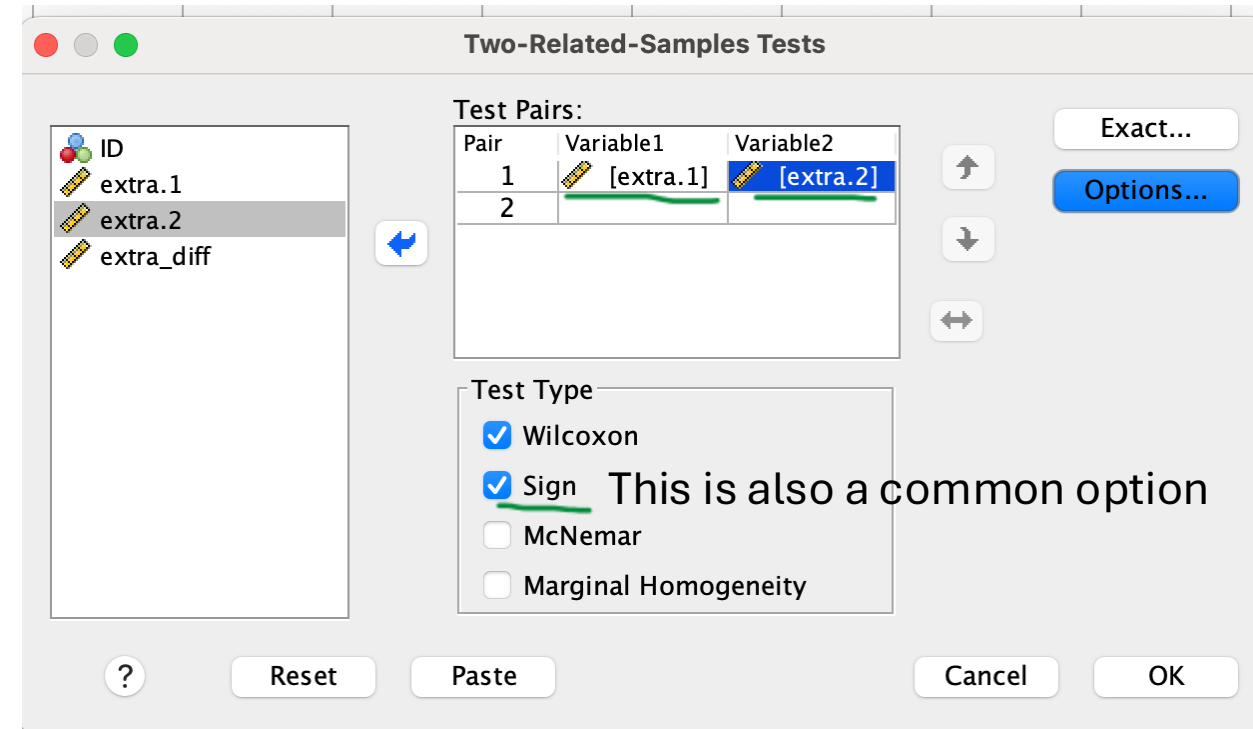
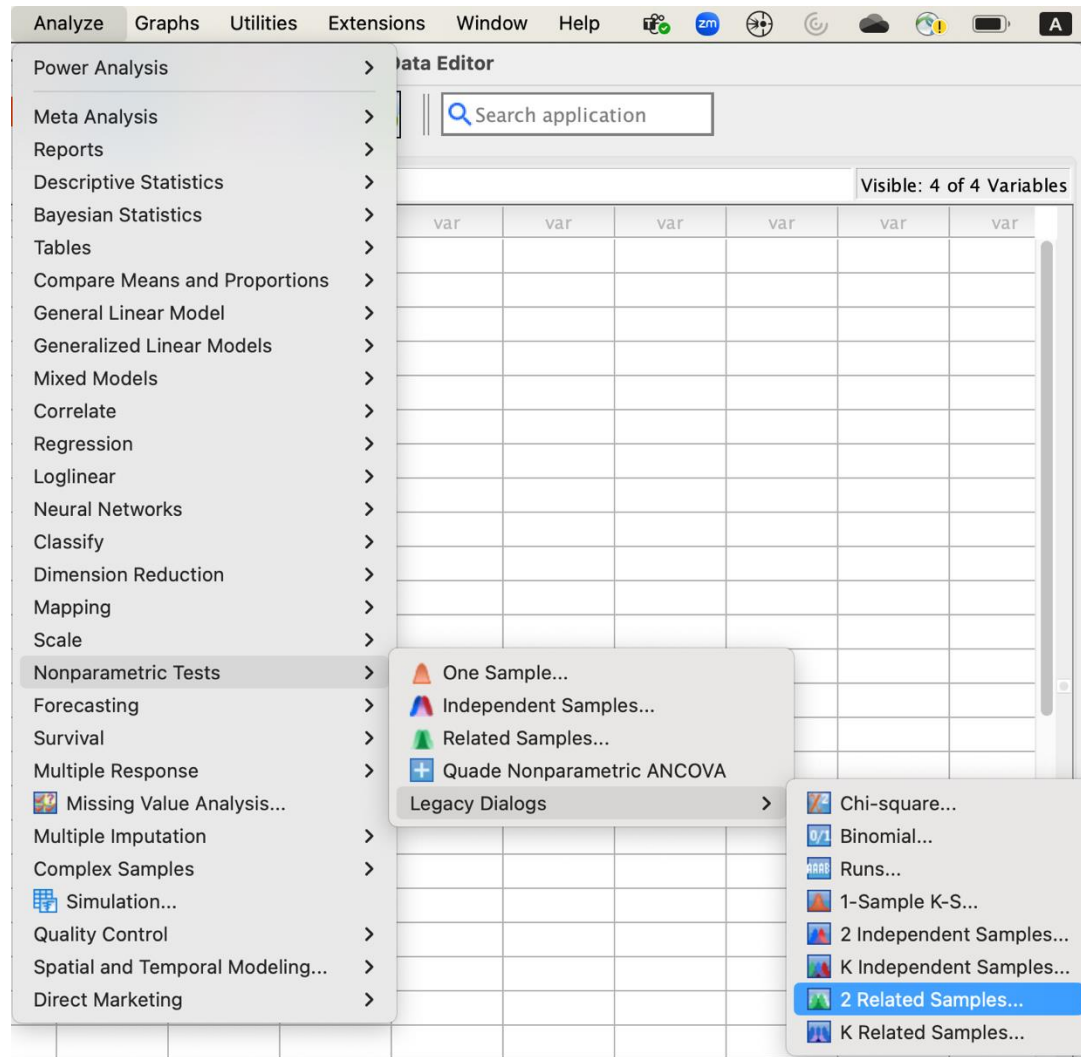
Why are the two different?  
Can you answer?

### Paired Samples Effect Sizes

		Standardizer <sup>a</sup>	Point Estimate	95% Confidence Interval	
				Lower	Upper
Pair 1	extra.2 - extra.1	Cohen's d	1.9217	.822	1.456
		Hedges' correction	2.1028	.751	1.331

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

# Non-parametric: Wilcoxon Signed-Rank Test



# Reporting results

## ➔ NPar Tests

[DataSet9]

### Wilcoxon Signed Ranks Test

		Ranks		
		N	Mean Rank	Sum of Ranks
extra.2 – extra.1	Negative Ranks	0 <sup>a</sup>	.00	.00
	Positive Ranks	9 <sup>b</sup>	5.00	45.00
	Ties	1 <sup>c</sup>		
	Total	10		

a. extra.2 < extra.1

b. extra.2 > extra.1

c. extra.2 = extra.1

### Test Statistics<sup>a</sup>

	extra.2 – extra.1
Z	-2.668 <sup>b</sup>
Asymp. Sig. (2-tailed)	.008

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

You can report results in a similar fashion to how you reported results for Mann-Whitney U test.

Compare the medians of the two samples, report Statistic (Z) and its significance, and each sample size.

# Covered all relevant 2-sample comparisons

