SPSS: Up and running for academic research

Jacqueline Murphy

With thanks to Kerstin Frie and Rachel Pechey

SPSS

- SPSS is a widely used statistical software package produced by IBM
- Allows you to perform the most common statistical calculations
- Intuitive point and click interface - no coding needed (but is an option)
- SPSS is available for Windows and Mac, works in exactly the same way on both systems

Structure of this Course

- The course is structured around a series of Research Questions (RQs)
- In-person sessions:
  - Instructor-led demonstrations
  - Give-it-a-go exercises with trouble-shooting help
- Handbook:
  - Contains all computing steps for the demonstrations and exercises
  - Glossary of statistical concepts

https://skills.web.ox.ac.uk/spss-up-and-running-for-academic-research-course-pack

Research Questions	Organising Data	Descriptive Analysis	Graphing	Simple Stats	Associations	ANOVA	Output Usage
RQ1: Frequency sweet vs savoury breakfast	Compute variables Recode variables		Bar chart				
RQ2: Units breakfast	Compute variables	Descriptives		One-sample t-test			
RQ3: Effect of food types on satiety				Paired-samples t-test			
RQ4: Differences in height by gender	Compute variables			Independent-samples t-test			
RQ5: Relationship age and height	Splitting files		Scatterplot		Bivariate correlation		
RQ 6: Relationship general pain and age, controlled for health			Pie chart		Partial correlation		
RQ 7: Reaction time to emotional facial expressions	Sorting cases Selecting cases	Frequencies	Boxplot		Linear regression		
RQ 8: Effect of supermarket promotion on sales	Duplicate cases					One-way ANOVA	
RQ 9: Effect of new pain treatment		Custom Tables				Two-way ANOVA	Formatting graphs
RQ 10: Effect of weight loss programme	Merge files					Repeated /mixed ANOVA	Cleaning output Formatting tables Exporting graphs and tables
Additional Features	Find and Replace	Explore Data Crosstabs					

Before we start…

- Please feel free to ask clarifying questions during demonstrations
- Please do ask for help during the give-it-a-go exercises

- All data are adapted from SPSS example data sets or fabricated
- Please feel free to save the example data on a USB drive so you can use it at home (tip: keep a "master" copy of the original unedited data)

Statistical concepts we will be using

- Mean/Standard Deviation/Variance
- Between- vs Within-Subjects Variables
- Significance level and p-values
- Confidence Intervals
- t-tests
- Correlations
- ANOVAs
- Regressions

Getting started in SPSS (demonstration)

- Using Data in SPSS
  - Import the Import.xlsx file into SPSS
  - Open the Breakfast.sav file in SPSS
- Data vs. Output
  - Compute the frequency of a variable using Analyze -> Descriptive Statistics…
  - Take a look at the Data and Output windows
- Variable view
  - In the *Variable View* of the data window, create a new variable named age
  - Define the variable age as numeric, with 0 decimals, at scale measure

RQ1: Frequency Sweet vs Savoury Breakfast

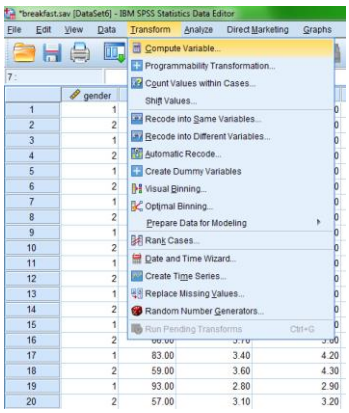
- Data set of 31 participants' breakfast consumption over 40 days
- 11 variables: gender, satiety of sweet & savoury breakfast, breakfast units split up by breakfast items: 3 x savoury, 5 x sweet

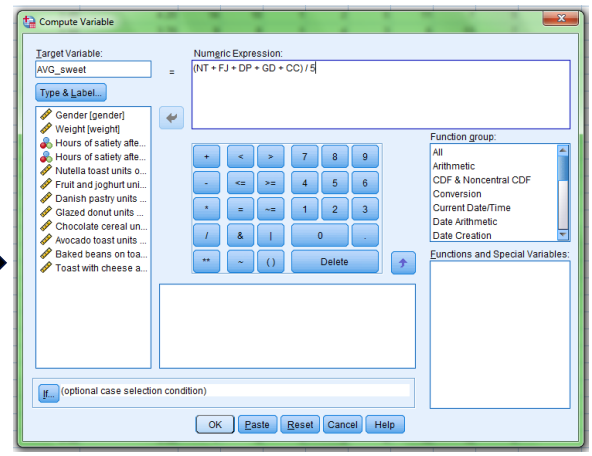
☛ Do participants consume predominantly sweet or savoury foods for breakfast?

- Compute averages sweet vs. savoury foods
- Compute difference sweet vs. savoury foods
- Recode difference score into a categorical variable
- Bar chart: Frequency predominantly sweet vs. savoury foods

RQ1: Frequency Sweet vs. Savoury Breakfast
Compute variables: average sweet items

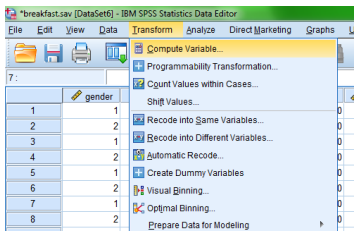
Go to *Transform* and select *Compute Variable*...

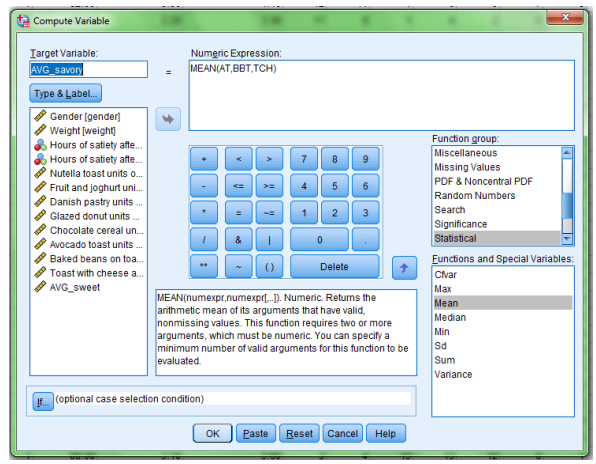




- Enter a target variable name
- Double-click on the five sweet breakfast variables to add them to the numeric expression box
- Enter "+" between the variables to add the values and put brackets around all variables, then divide by 5
- Click *OK*

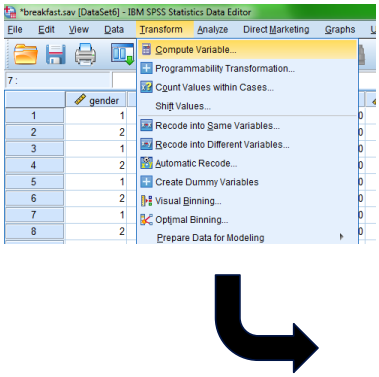
RQ1: Frequency Sweet vs. Savoury Breakfast
Compute variables: average savoury items

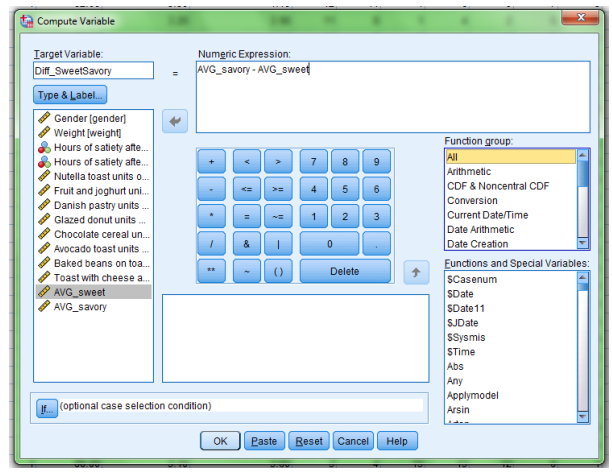




- Select *Compute Variable…* from the tab *Transform*
- Enter a target variable name
- Find the *Mean* function in the function group *Statistical*
- Double-click on all the variables you want to include in the calculation, make sure they are entered between the brackets, separated by commas
- Click *OK*

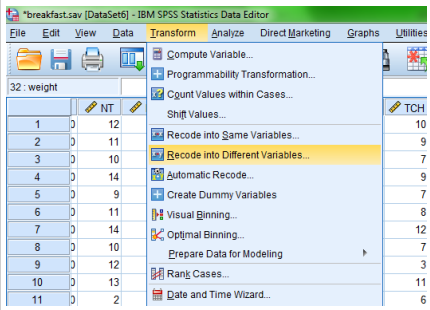
RQ1: Frequency Sweet vs. Savoury Breakfast
Compute variables: difference sweet vs. savoury



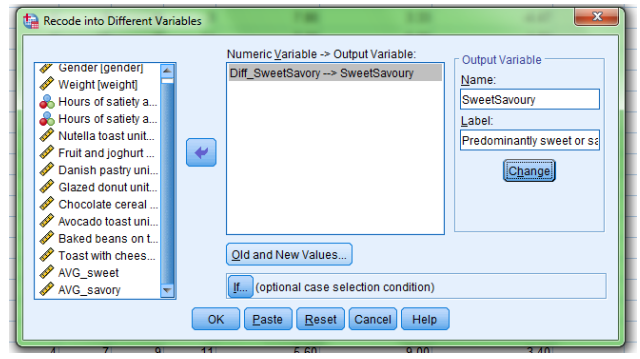


- Select *Compute Variable…* from the tab *Transform*
- Enter a target variable name
- Double-click on the variables you want to include in the calculation
- Add "-" between the variables to substract one value from another
- Click *OK*

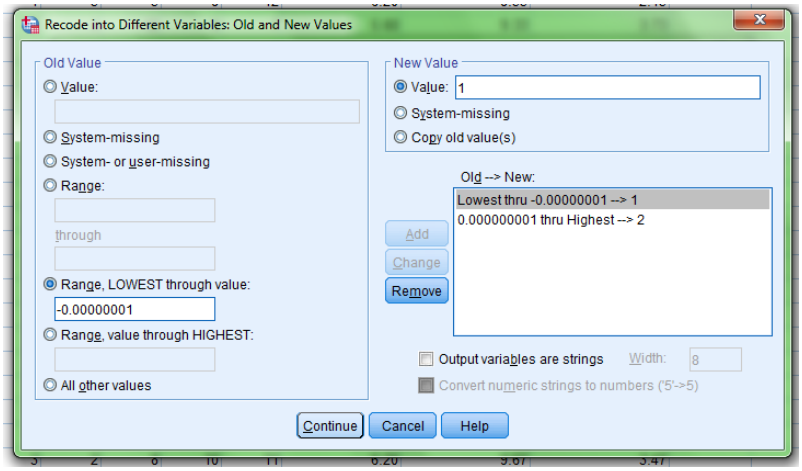
RQ1: Frequency Sweet vs. Savoury Breakfast
Recode variables: sweet vs. savoury



- Select *Recode into Different Variables…* from the tab *Transform*
- Add the variable you want to recode to the middle box
- Choose a name for the output variable (recoded values)
- Optional: add a label to remember what the variable consists of
- Click *Change*
- Click *Old and New Values…* to define the recoding

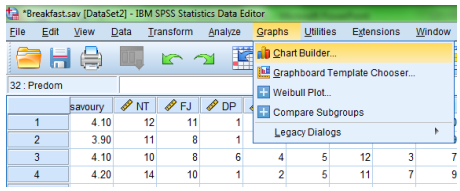


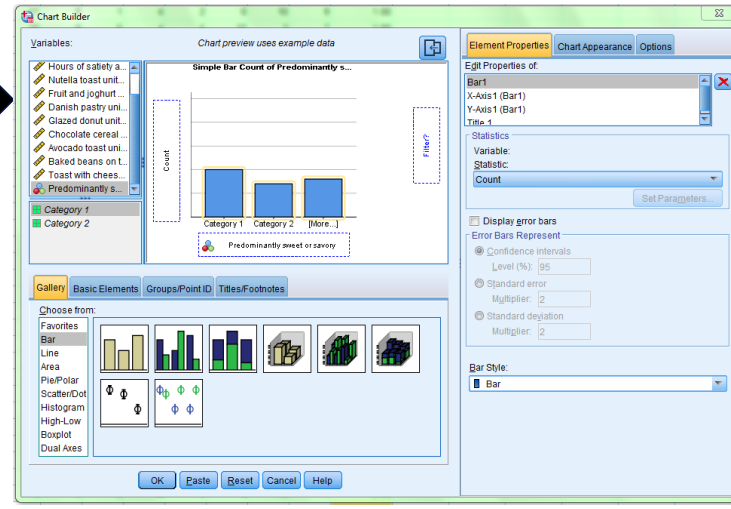
RQ1: Frequency Sweet vs. Savoury Breakfast
Recode variables: sweet vs. savoury



- A new box will have appeared
- Select *Range, LOWEST through value* and enter a negative value close to zero
- Enter "1" into the box *Value* in the section *New Value* and then click *Add* to confirm this recoding
- Select *Range, value through HIGHEST* and enter a positive value close to zero
- Enter "2"(or any other value but 1) into the box *Value* and add this recoding to the box
- Click *Continue* to get back to the previous dialogue
- Click *OK* to start the recoding

RQ1: Frequency Sweet vs. Savoury Breakfast
Bar chart: Frequency sweet vs savoury





- Select *Chart Builder* from the tab *Graphs*
- In the bottom section, choose the category *Bar* in order to see the different bar chart options
- Double-click on the simple bar chart (top left)
- Drag the recoded variable from the variable list to the x-axis
- The y-axis should automatically change to Count
- Click "OK"

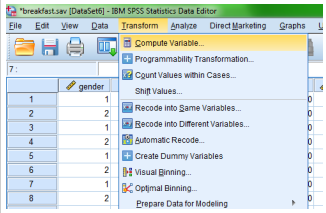
RQ2: Breakfast Units

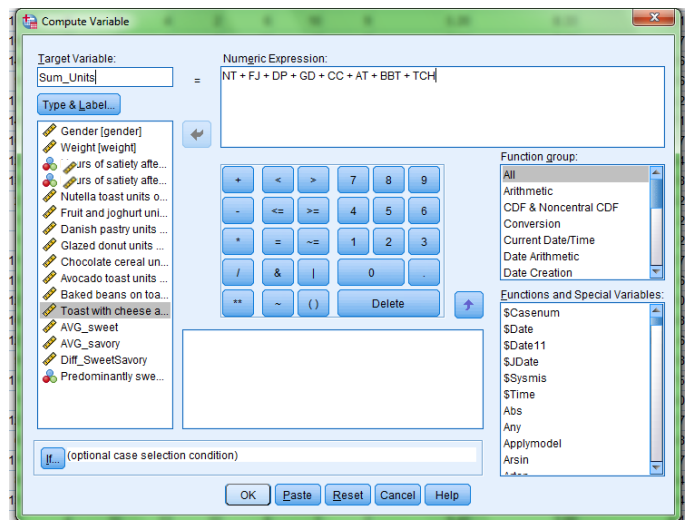
- Same data set as RQ1

☛Do participants consume more than one unit of breakfast per day?

- Compute variables: sum of breakfast items
- Get to know the data: Descriptive statistics of each variable
- One sample t-test: Did the participants consume more than one unit of breakfast a day (i.e. 40 units over 40 days)?

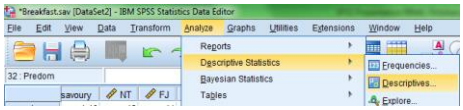
RQ2: Breakfast Units
Compute variables: sum of breakfast items

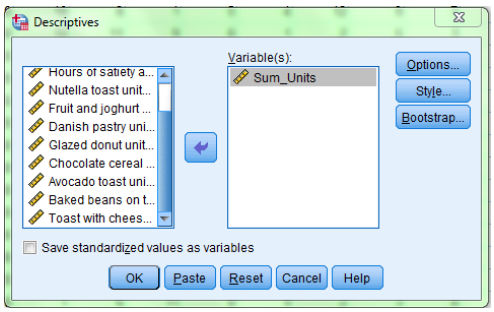


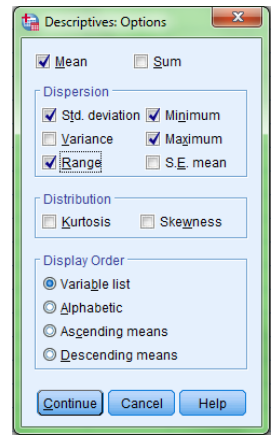


- Select *Compute Variable…* from the tab *Transform*
- Enter a target variable name *Sum-Units*
- Double-click on all breakfast variables to add them to the numeric expression box
- Enter "+" between the variables to add the values
- Click *OK*

RQ2: Breakfast Units
Descriptive Statistics

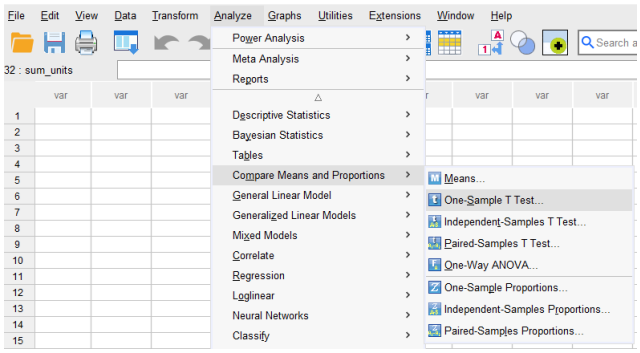




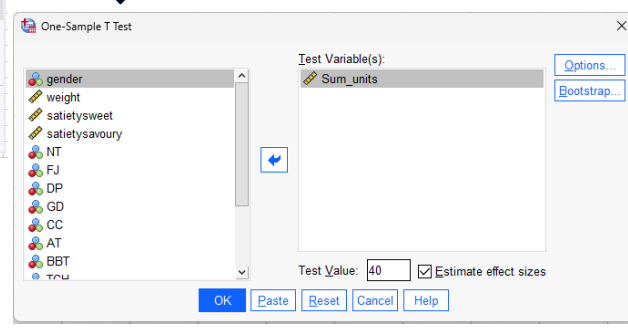


- Select *Descriptives…* under *Descriptive Statistics* in the tab *Analyze*
- Add the variable *Sum_Units* to the *Variable(s)* box
- Click on *Options…* Select the statistics of interest, then click on *Continue*
- Click *OK*

RQ2: Breakfast Units
One-sample t-test



- Select *One-Sample T Test…* under *Compare Means and Proportions* in the tab *Analyze*



- Double-click on the *Sum_Units* variable to add it to the box
- Enter the value against which you want to test the data in the *Test Value* box, in this case 40
- Click *OK*

RQ2: Breakfast Units
One-sample t-test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Sum_Units	31	56.03	6.585	1.183

One-Sample Test

Test Value = 40

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Sum_Units	13.555	30	<.001	<.001	16.032	13.62	18.45

- SPSS returns a table with statistics, a table with the test results and a table with effect sizes
- The second table gives the p-value for this two-tailed test in the column *Significance (Two-sided p)*, if this value is below 0.05, it is considered significant = the sample's mean is significantly different from the value against which it was tested
- In this case, we find that the participants ate significantly more than 40 breakfast items across the 40 days of the study

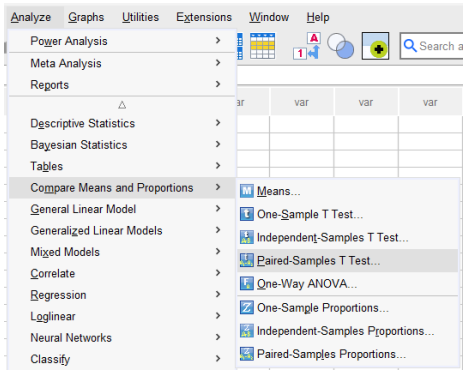
RQ3: Effect of Food Types on Satiety

- Same data set as for RQ1 and RQ2
- Data on satiety of sweet vs. savoury breakfast

☛Do sweet or savoury foods fill you up for longer?

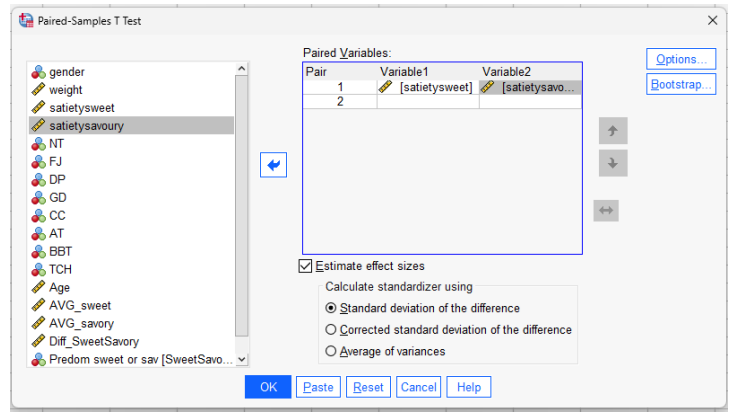
- Paired-samples t-test

RQ3: Effect of Food Types on Satiety
Paired-samples t-test



- Select *Paired-Samples T Test…* under *Compare Means and Proportions* in the tab *Analyze*

- Double-click on the two satiety variables to add them to the two columns of the box
- Click *OK*



RQ3: Effect of Food Types on Satiety
Paired-samples t-test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	satietysweet	2.981	31	.5036	.0904
	satietysavoury	3.545	31	.5662	.1017

Paired Samples Correlations

		N	Correlation	Significance One-Sided p	Significance Two-Sided p
Pair 1	satietysweet & satietysavoury	31	.838	<.001	<.001

- SPSS returns four tables, one with statistics, one with correlation results, one with the test results and one with effect sizes
- The main results are in the third table *Paired Samples Test* (see next slide)

RQ3: Effect of Food Types on Satiety
Paired-samples t-test (cont.)

Paired Samples Test

		Paired Differences					t	df	Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			One-Sided p	Two-Sided p
Pair 1	satietysweet - satietysavoury	-.5645	.3104	.0558	-.6784	-.4507	-10.125	30	<.001	<.001

- SPSS gives the p-value for this two-tailed test in the column *Two-Sided p*, if this value is below 0.05, it is considered significant, i.e. the two samples tested are different from each other
- In this case, we find a significant difference between the satiety after savoury vs sweet food, whereby the savoury food kept participants full for longer

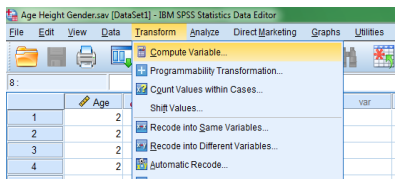
RQ4: Effect of Gender on Height

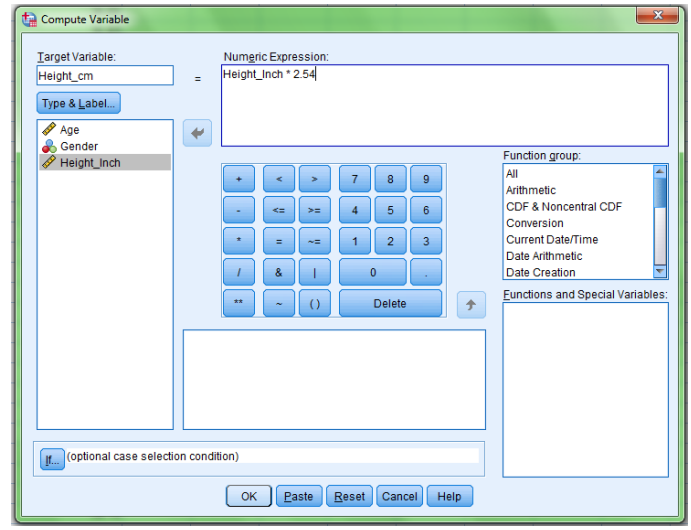
- Data set of 104 boys and girls aged 2-9
- 3 variables: Age, Height and Gender
- Height measured in inches

☛ Is there an effect of gender on height (in cm)?

- Convert height measurements from inches to centimeters
- Independent samples t-test

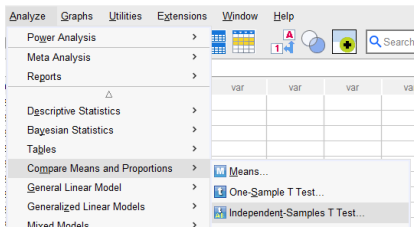
RQ4: Effect of Gender on Height
Convert Measurement

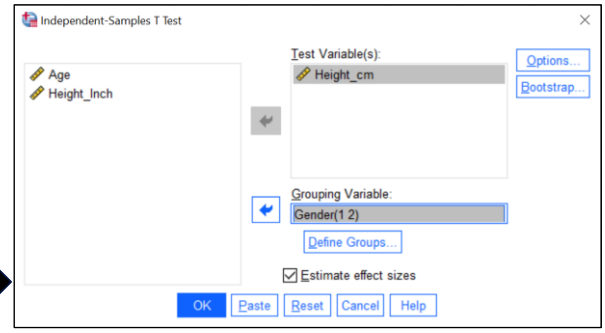


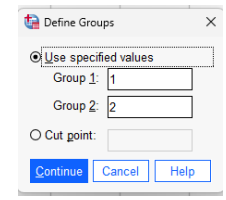


- Select *Compute Variable…* from the tab *Transform*
- Enter a target variable name
- Double-click on *Height_Inch* variable to add it to the numeric expression box
- Multiply with 2.54
- Click *OK*

RQ4: Effect of Gender on Height
Independent Samples t-test







- Select *Independent-Samples T Test…* under *Compare Means and Proportions* in the *Analyze* tab
- Select the *Height_cm* variable and add it to the *Test Variable(s)* box
- Select the *Gender* variable and add it to the *Grouping Variable* box
- Click on *Define Groups…* to open the dialogue on the right, enter the values defining the two groups, in this case 1 and 2
- Click *Continue* and then *OK*

RQ4: Effect of Gender on Height
Independent Samples t-test

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Height_cm	Male	58	108.5517	15.26281	2.00411
	Female	46	115.9565	16.26579	2.39826

- SPSS runs the Levene's Test for Equality of Variances, since this test is <u>not</u> significant (*Sig.* value above 0.05), we use the t-test outcomes of the upper row (*equal variances assumed*)

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
Height_cm	Equal variances assumed	.289	.592	-2.387	102	.009	.019	-7.40480	3.10233	-13.55826	-1.25134
	Equal variances not assumed			-2.369	93.714	.010	.020	-7.40480	3.12539	-13.61059	-1.19901

- SPSS gives the p-value for this two-tailed test in the column *Significance (Two-Sided p)*, if this value is below 0.05, it is considered significant = the two samples tested are different from each other
- In this case, we find a significant difference in height between the groups, whereby the girls are on average taller than the boys

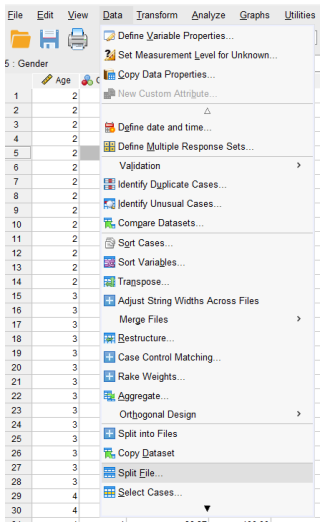
RQ5: Relationship of Age & Height, by gender

- Same data set as RQ4

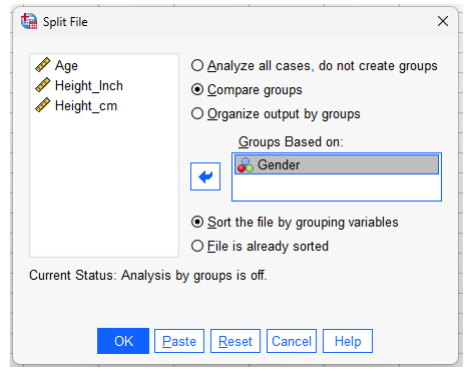
☛ Is there a correlation of age and height? Is this dependent on gender?

- Splitting file: by gender (changes how the results are displayed)
- Bivariate correlation
- Scatterplot

RQ5: Relationship of Age & Height, by gender
Splitting file by gender



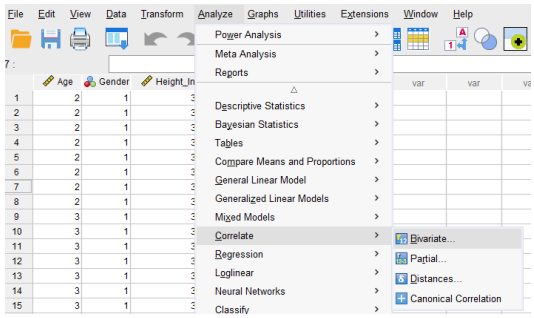
- Select *Split File…* in the tab *Data*



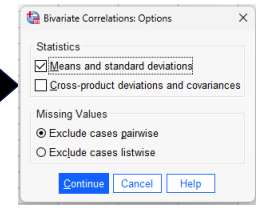
- Select the option *Compare groups*
- Double-click on the gender variable to add it to the box
- Click *OK*

- Lower right corner of the data window should now read "Split by Gender"

RQ5: Relationship of Age & Height, by gender
Bivariate correlation



- Select *Bivariate…* under *Correlate* in the tab *Analyze*
- Add the variables Height_cm and Age to the box on the right
- Optional: Click on *Options…* to add statistics to the output, then click on *Continue*
- Select the correlation coefficients you want to have calculated, in this case Pearson
- Click *OK*



RQ5: Relationship of Age & Height, by gender
Bivariate correlation

Correlations

Descriptive Statistics

Gender		Mean	Std. Deviation	N
Male	Height_cm	108.55	15.263	58
	Age	5.31	2.280	58
Female	Height_cm	115.96	16.266	46
	Age	5.43	2.344	46

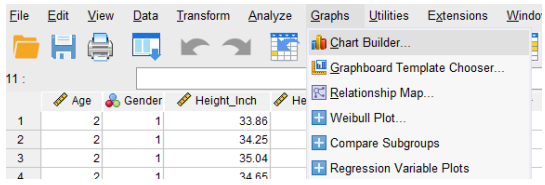
Correlations

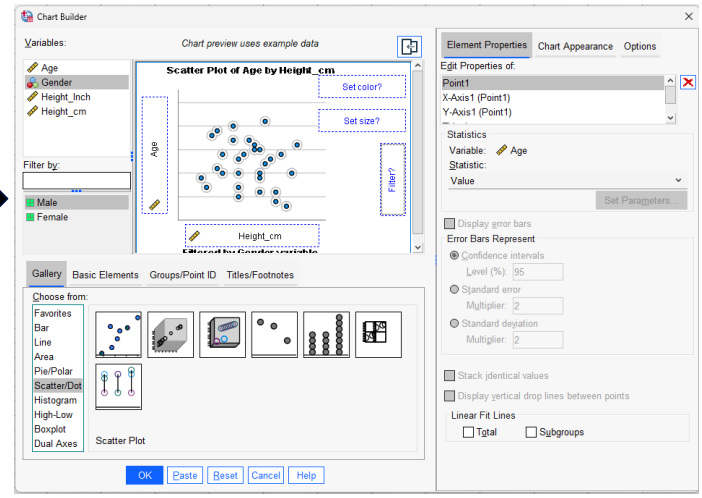
Gender			Height_cm	Age
Male	Height_cm	Pearson Correlation	1	.988**
		Sig. (2-tailed)		.000
		N	58	58
	Age	Pearson Correlation	.988**	1
		Sig. (2-tailed)	.000	
		N	58	58
Female	Height_cm	Pearson Correlation	1	.994**
		Sig. (2-tailed)		.000
		N	46	46
	Age	Pearson Correlation	.994**	1
		Sig. (2-tailed)	.000	
		N	46	46

**. Correlation is significant at the 0.01 level (2-tailed).

- SPSS returns two tables, the first shows descriptive statistics, the second presents the correlation results
- The tables are split by gender, SPSS returns correlation coefficients for males and females separately
- The first row *Pearson Correlation* is the Pearson correlation coefficient, it indicates the direction and strength of the association
- The second row *Sig. (2-tailed)* shows the p-value
- The third row *N* shows the sample size for each calculation
- In this case, we find that the significantly positive association of height and age exists for both boys and girls

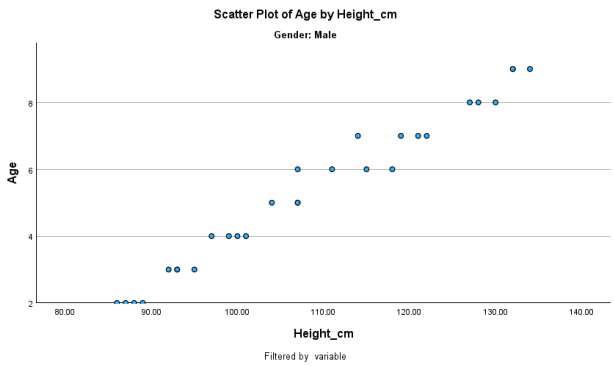
RQ5: Relationship of Age & Height, by gender
Scatterplot

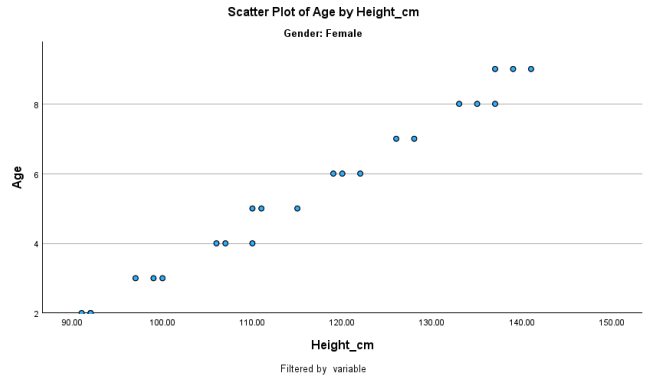




- Select *Chart Builder* from the tab *Graphs*
- In the bottom section, choose the category *Scatter/Dot* in order to see the different scatterplot options
- Double-click on the simple scatterplot (top left)
- From the list of variables, drag the variables age and Height_cm to the chart preview
- Click *OK*

RQ5: Relationship of Age & Height, by gender
Scatterplot





- Optional: double-click on the charts to look at formatting and layout options

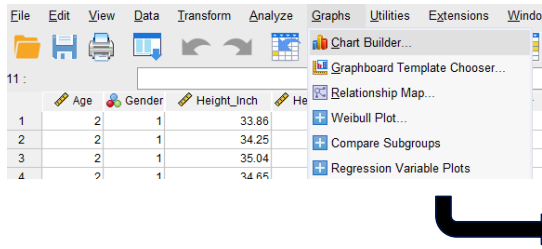
RQ6: Association of pain & age, controlling for health

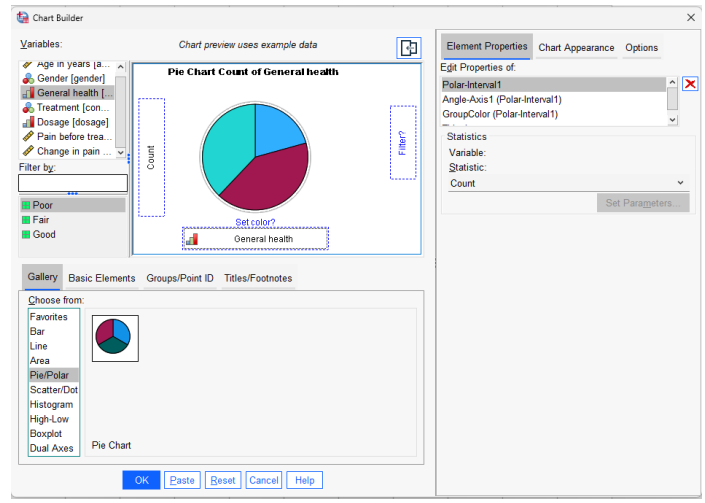
- Data set of 60 participants from a pain treatment study
- 7 variables: age, gender, general health, condition (control or treatment), dosage (low or high), pain at baseline, pain change after treatment

☛ Is there an association between pain and age, while controlling for general health?

- Pie chart: general health
- Partial correlation: age and baseline pain, controlling for general health

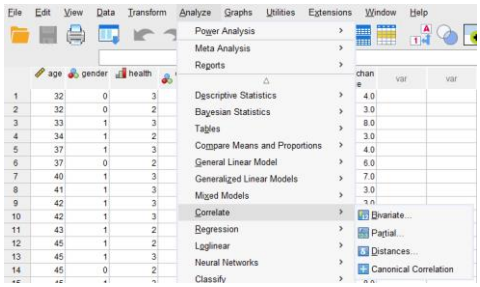
RQ6: Association of pain & age, controlling for health
Pie Chart

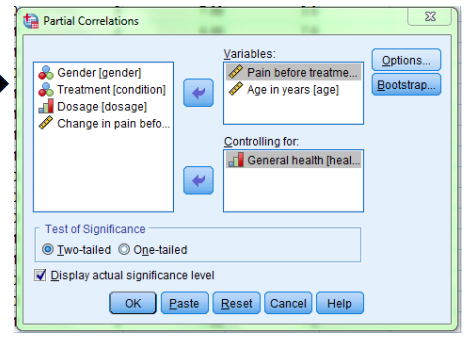


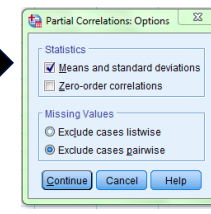


- Select *Chart Builder* from the tab *Graphs*
- In the bottom section, choose the category *Pie/Polar*
- Double-click on the depicted pie chart
- From the list of variables, drag the variable General health to the chart preview
- Click *OK*

RQ6: Association of pain & age, controlling for health
Partial Correlation







- Select *Partial…* under *Correlate* in the tab *Analyze*
- Add the variables Pain before treatment and Age in years to the *Variables* box
- Add the General health variable to the *Controlling for* box
- Optional: Click on *Options...* to add statistics to the output, then click on *Continue*
- Click *OK*

RQ6: Association of pain & age, controlling for health
Partial Correlation

Partial Corr

Descriptive Statistics

	Mean	Std. Deviation	N
Pain before treatment	7.0167	1.46706	60
Age in years	54.2333	10.76461	60
General health	2.3500	.73242	60

Correlations

Control Variables			Pain before treatment	Age in years
General health	Pain before treatment	Correlation	1.000	.511
		Significance (2-tailed)	.	<.001
		df	0	57
	Age in years	Correlation	.511	1.000
		Significance (2-tailed)	<.001	.
		df	57	0

- SPSS returns two tables, the first shows descriptive statistics, the second presents the correlation results
- The first row of the correlations table shows the correlation coefficient, it indicates the direction and strength of the association
- The second row *Significance (2-tailed)* shows the p-value
- The third row *df* shows the sample size for each calculation
- In this case, we find a significant positive relationship between age and pain, while controlling for general health

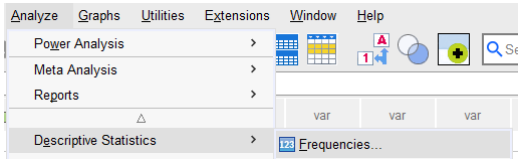
RQ 7: Reaction time to emotional facial expressions

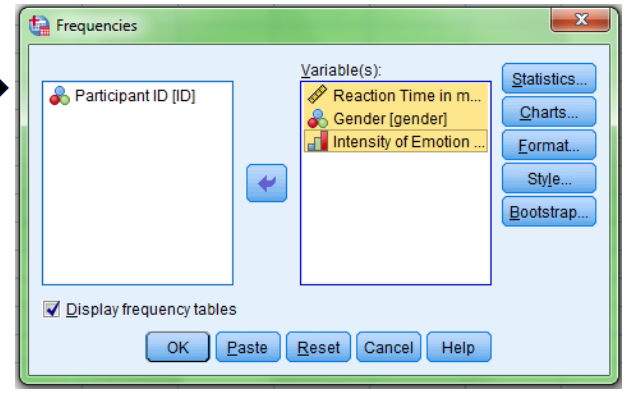
- Data set of 141 participants: reaction time study
- Participants had to identify emotional expressions of different intensities, reaction times were measured
- 3 variables: gender, reaction time and intensity of emotion

☛ Do gender and emotion intensity predict reaction time?

- Getting to know the data set: Frequencies
- Boxplot: Finding outliers
- Sorting and Selecting Cases
- Linear regression, outcome: reaction time, predictors: emotion intensity and gender

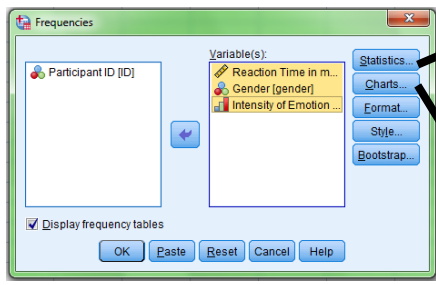
RQ 7: Reaction time to emotional facial expressions
Frequencies

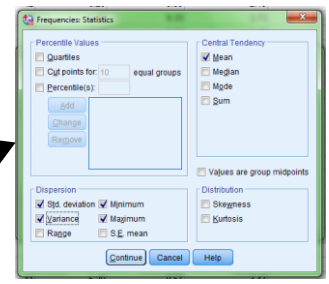




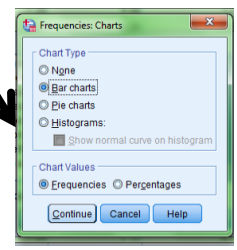
- Select *Frequencies* under *Descriptive Statistics* in the tab *Analyze*
- Double-click on the variables to add them to the box
- Open options for *Statistics* and *Charts* by clicking on the respective buttons (see next slide)

RQ 7: Reaction time to emotional facial expressions
Frequencies



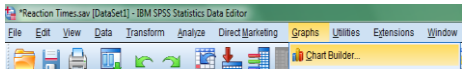


- Select *Statistics*
- Select the statistics you want to receive for the selected variables
- Click *Continue*

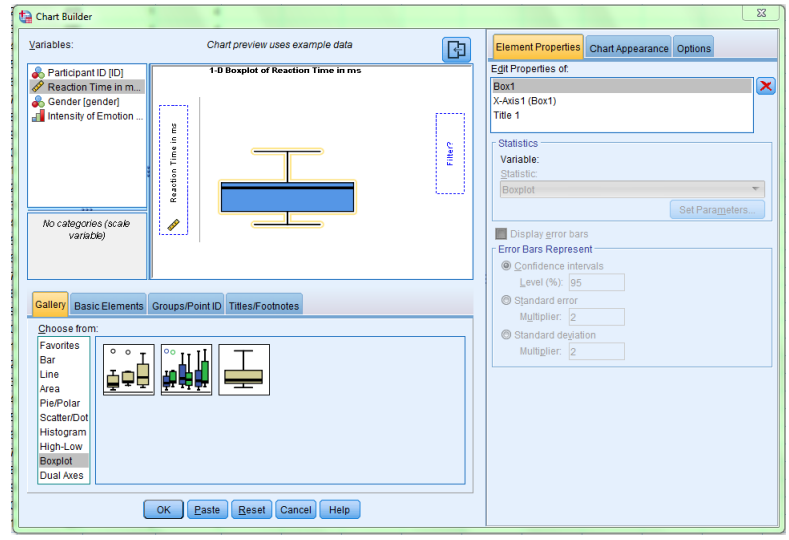


- Select *Charts*
- Select the type of chart you want to receive for the chosen variables
- Choose whether the values should be presented as frequencies or percentages
- Click *Continue*

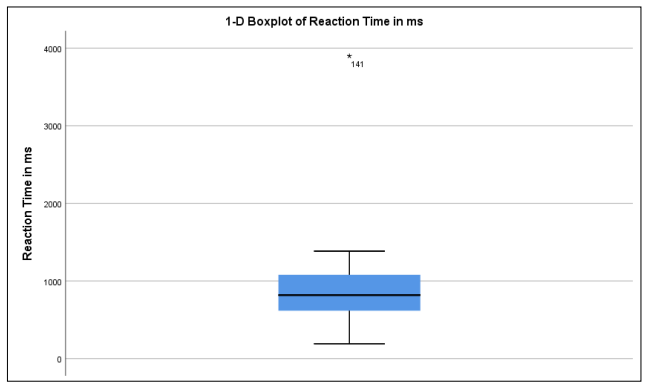
RQ 7: Reaction time to emotional facial expressions
Boxplot



- Select *Chart Builder* from the tab *Graphs*
- In the bottom section, choose the category *Boxplot*
- Double-click on the boxplot chart type on the right
- From the list of variables, drag the variable Reaction time to the main axis (labelled *X-axis* here)
- Click *OK*

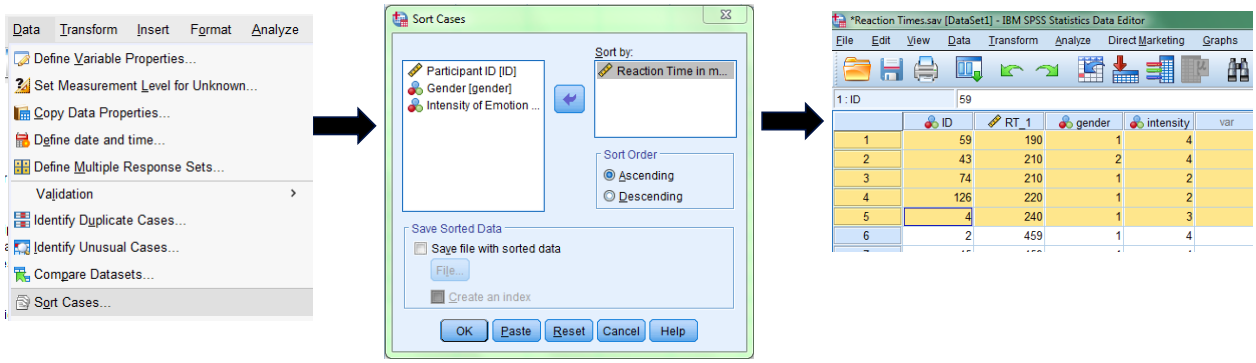


RQ 7: Reaction time to emotional facial expressions
Boxplot



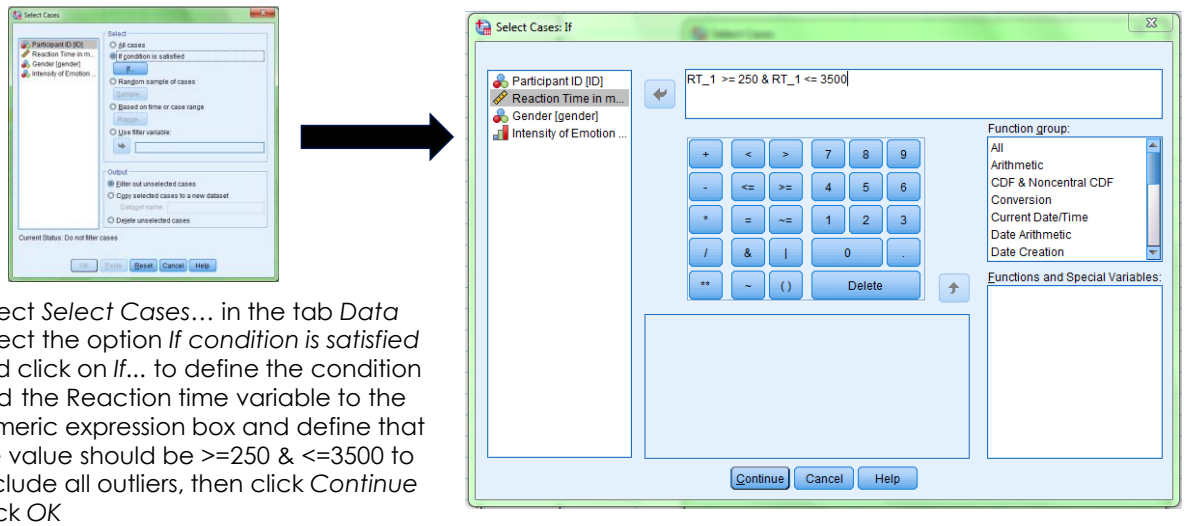
The boxplot shows there is 1 outlier in the data, around 4000ms

RQ 7: Reaction time to emotional facial expressions
Sorting Cases



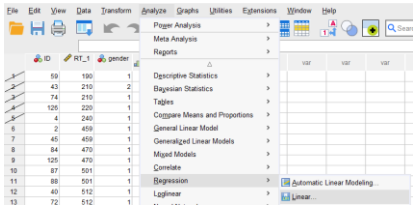
- Select *Sort Cases…* in the tab *Data*
- Add the variable Reaction Time to the *Sort by* box and select ascending sort order
- Click *OK*
- The data set should now be sorted by reaction time, we can see that 5 data points are below 250ms and 1 point is much larger than the others (3898ms) → outliers

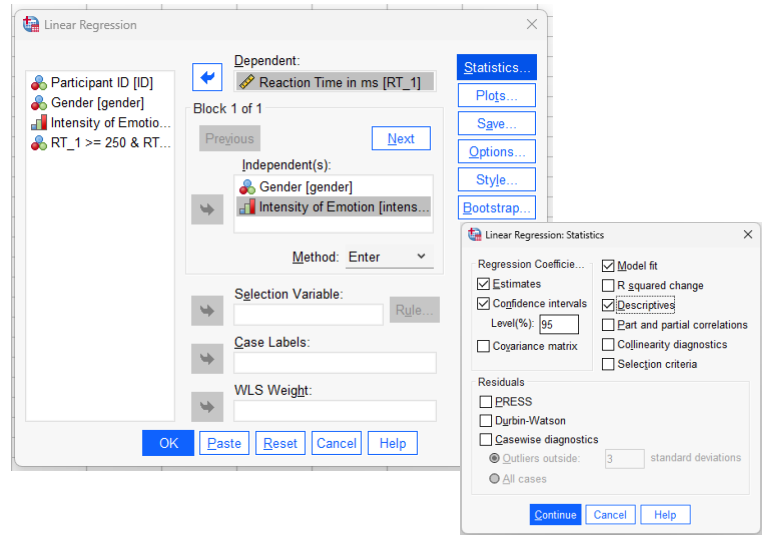
RQ 7: Reaction time to emotional facial expressions
Selecting Cases



- Select *Select Cases…* in the tab *Data*
- Select the option *If condition is satisfied* and click on *If…* to define the condition
- Add the Reaction time variable to the numeric expression box and define that the value should be >=250 & <=3500 to exclude all outliers, then click *Continue*
- Click *OK*

RQ 7: Reaction time to emotional facial expressions
Linear regression





- Select *Linear...* under *Regression* in the tab *Analyze*
- Add the Reaction Time variable to the *Dependent* box and the variables Gender and Intensity of Emotion to the Independent variables box
- Click on *Statistics...* and select *Estimates, Confidence Intervals, Model fit* and *Descriptives*, then click Continue
- Click *OK*

RQ 7: Reaction time to emotional facial expressions
Linear regression

Descriptive Statistics

	Mean	Std. Deviation	N
Reaction Time in ms	857.57	257.368	135
Gender	1.53	.501	135
Intensity of Emotion	2.61	1.099	135

Correlations

		Reaction Time in ms	Gender	Intensity of Emotion
Pearson Correlation	Reaction Time in ms	1.000	.861	-.964
	Gender	.861	1.000	-.849
	Intensity of Emotion	-.964	-.849	1.000
Sig. (1-tailed)	Reaction Time in ms	.	<.001	<.001
	Gender	.000	.	.000
	Intensity of Emotion	.000	.000	.
N	Reaction Time in ms	135	135	135
	Gender	135	135	135
	Intensity of Emotion	135	135	135

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8303655.547	2	4151827.773	957.599	<.001[b]
	Residual	572307.535	132	4335.663		
	Total	8875963.081	134			

a. Dependent Variable: Reaction Time in ms
b. Predictors: (Constant), Intensity of Emotion, Gender

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	1246.249	56.364		22.111	<.001	1134.756	1357.742
	Gender	79.316	21.448	.154	3.698	<.001	36.889	121.742
	Intensity of Emotion	-194.931	9.777	-.833	-19.937	<.001	-214.272	-175.590

a. Dependent Variable: Reaction Time in ms

- SPSS returns several tables, the first shows descriptive statistics, the second presents correlation results, after that the model summary and regression results are listed
- The first row of the *ANOVA* table shows here that the included predictors (Gender and Intensity) reliably predict the outcome (Reaction Time) ($p<0.05$)
- The second row and third row of the Coefficients table show the results for the two predictors: both gender and intensity of emotion predict reaction time ($p<0.05$)

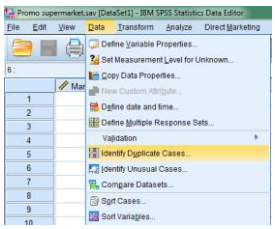
RQ 8: Effect of supermarket promotion on sales

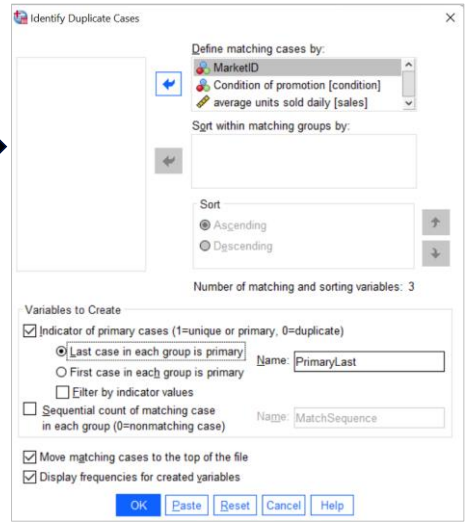
- Data from 133 supermarkets, effect of supermarket promotion on sales
- 3 promotion conditions: control, price reduction, price reduction + advertisement
- 2 variables: condition of promotion, average daily sale

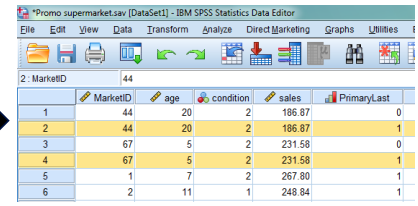
☛ Did the two types of promotion have an effect on sales?

- Identify duplicate cases
- One-way ANOVA: Did the two types of promotion have an effect on sales?

RQ8: Effect of supermarket promotion on sales
Duplicate cases





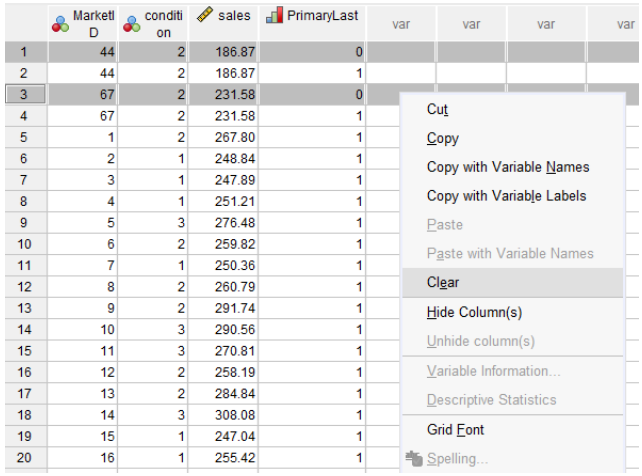


	MarketID	age	condition	sales	PrimaryLast
1	44	20	2	186.87	0
2	44	20	2	186.87	1
3	67	5	2	231.58	0
4	67	5	2	231.58	1
5	1	7	2	267.80	1
6	2	11	1	248.84	1

- Select *Identify Duplicate Cases…* in the tab *Data*
- Select all variables and add them to the box *Define matching cases by:*
- Click *OK*

- SPSS has now resorted the data set, so that duplicate cases are listed at the top
- In this case, we find that the markets 44 and 67 have duplicate entries

RQ8: Effect of supermarket promotion on sales
Duplicate cases

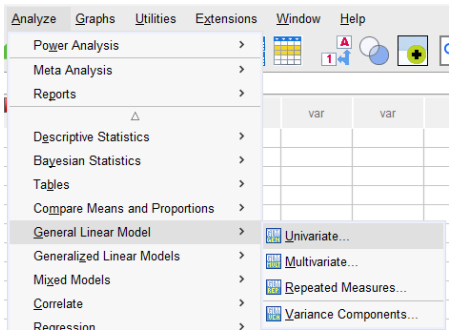


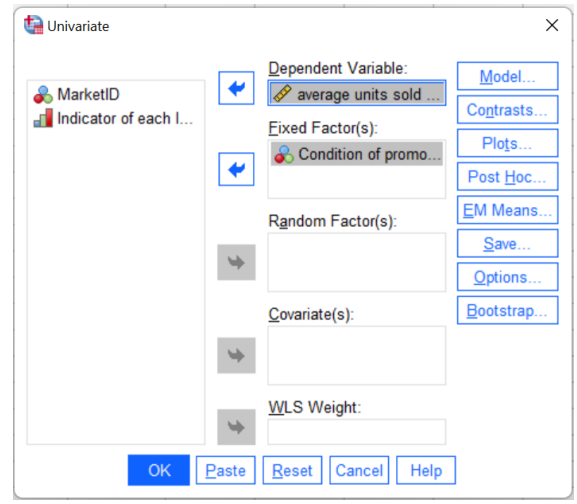
	MarketID	condition	sales	PrimaryLast
1	44	2	186.87	0
2	44	2	186.87	1
3	67	2	231.58	0
4	67	2	231.58	1
5	1	2	267.80	1
6	2	1	248.84	1
7	3	1	247.89	1
8	4	1	251.21	1
9	5	3	276.48	1
10	6	2	259.82	1
11	7	1	250.36	1
12	8	2	260.79	1
13	9	2	291.74	1
14	10	3	290.56	1
15	11	3	270.81	1
16	12	2	258.19	1
17	13	2	284.84	1
18	14	3	308.08	1
19	15	1	247.04	1
20	16	1	255.42	1

- Since markets 44 and 67 exist twice in the data set, we delete one of each
- Select the two rows you want to delete (press the Ctrl or cmd key)
- Right-click on the mouse to open a small window with several options, click on *Clear#*

Alternaitvely, you can use *Select Cases* under the *Data* tab, as in a previous question. Select *Use filter variable* and specify the new duplicate ID variable.

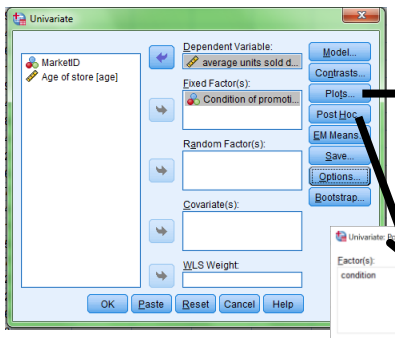
RQ8: Effect of supermarket promotion on sales
One-way ANOVA

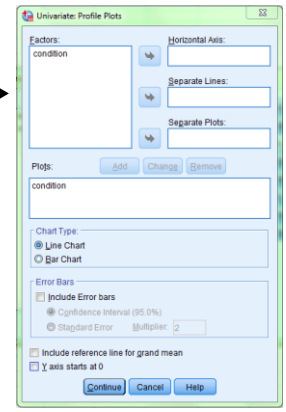


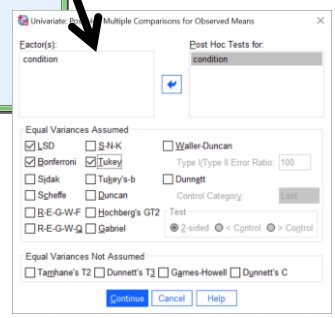


- Select *Univariate…* under *General Linear Model* in the tab *Analyze*
- Add the variable average units sold to the *Dependent Variable* box
- Add the variable Condition of promotion to the *Fixed Factor(s)* box

RQ8: Effect of supermarket promotion on sales
One-way ANOVA



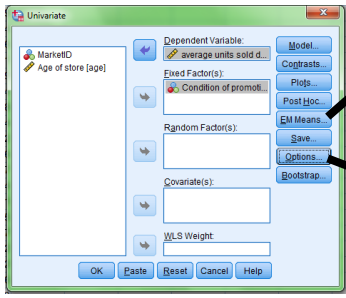


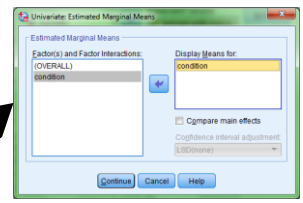


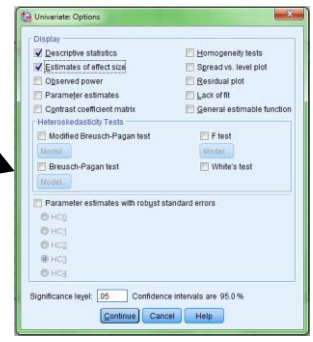
- Click on *Plots*...
- Add *condition* to horizontal axis
- Select *Add* to add the Plot

Click on *Post Hoc*... and add the variable *condition* to the box on the right, then select the post hoc tests you want to run (Tukey, Bonferroni and LSD), then click *Continue*

RQ8: Effect of supermarket promotion on sales
One-way ANOVA







- Click on *EM Means*... to add estimated marginal means to the output, then click on *Continue*

- Click on *Options*... to add statistics to the output, then click on *Continue*
- Click *OK* in the main window to start the analysis

RQ8: Effect of supermarket promotion on sales
One-way ANOVA

Between-Subjects Factors

		Value Label	N
Condition of promotion	1	control	47
	2	price reduction	45
	3	price reduction and advertisement	41

Descriptive Statistics

Dependent Variable: average units sold daily

Condition of promotion	Mean	Std. Deviation	N
control	186.7340	31.92507	47
price reduction	215.4553	35.93417	45
price reduction and advertisement	223.8083	30.45851	41
Total	207.8807	36.39105	133

- SPSS returns several tables, including frequencies in the between-subject factor, descriptive statistics and the results of the analysis
- By checking the *Sig.* value in the row *condition* of the table *Tests of Between-Subjects Effects* you can find that there is a main effect for the supermarket promotion ($p<0.05$)
- Partial Eta Squared is a measure of effect size, in this case it indicates a large effect for the promotion (>0.14)

Tests of Between-Subjects Effects

Dependent Variable: average units sold daily

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	34000.607[a]	2	17000.304	15.695	<.001	.195
Intercept	5772256.535	1	5772256.535	5329.191	<.001	.976
condition	34000.607	2	17000.304	15.695	<.001	.195
Error	140808.122	130	1083.139			
Total	5922320.703	133				
Corrected Total	174808.729	132				

a. R Squared = .195 (Adjusted R Squared = .182)

RQ8: Effect of supermarket promotion on sales
One-way ANOVA

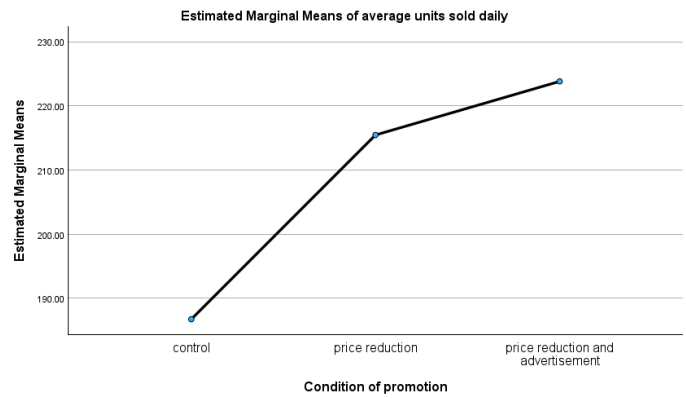
- The *Estimated Marginal Means* tables return descriptive statistics and confidence intervals for each promotion condition
- The plot makes it easy to visualise the main effect

Estimated Marginal Means

Condition of promotion

Dependent Variable: average units sold daily

Condition of promotion	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
control	186.734	4.801	177.237	196.231
price reduction	215.455	4.906	205.749	225.161
price reduction and advertisement	223.808	5.140	213.640	233.977



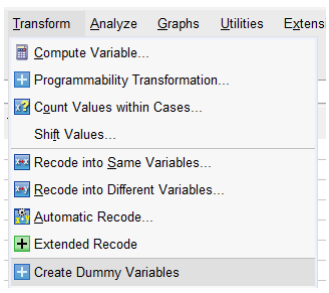
RQ8: Effect of supermarket promotion on sales
One-way ANOVA

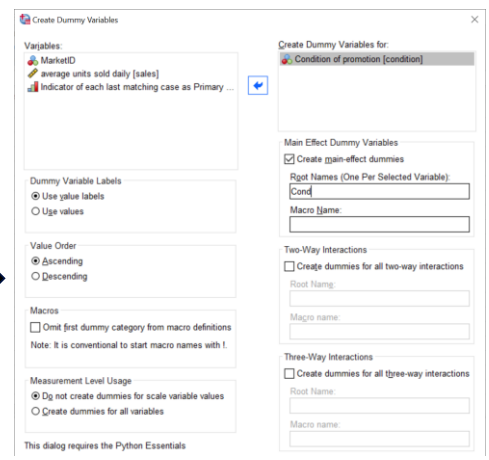
	(I) Condition of promotion	(J) Condition of promotion	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	control	price reduction	-28.7213*	6.86406	<.001	-44.9950	-12.4476
		price reduction and advertisement	-37.0743*	7.03303	<.001	-53.7486	-20.3999
	price reduction	control	28.7213*	6.86406	<.001	12.4476	44.9950
		price reduction and advertisement	-8.3530	7.10548	.470	-25.1991	8.4931
	price reduction and advertisement	control	37.0743*	7.03303	<.001	20.3999	53.7486
		price reduction	8.3530	7.10548	.470	-8.4931	25.1991
LSD	control	price reduction	-28.7213*	6.86406	<.001	-42.3010	-15.1416
		price reduction and advertisement	-37.0743*	7.03303	<.001	-50.9883	-23.1602
	price reduction	control	28.7213*	6.86406	<.001	15.1416	42.3010
		price reduction and advertisement	-8.3530	7.10548	.242	-22.4103	5.7044
	price reduction and advertisement	control	37.0743*	7.03303	<.001	23.1602	50.9883
		price reduction	8.3530	7.10548	.242	-5.7044	22.4103
Bonferroni	control	price reduction	-28.7213*	6.86406	<.001	-45.3691	-12.0735
		price reduction and advertisement	-37.0743*	7.03303	<.001	-54.1318	-20.0167
	price reduction	control	28.7213*	6.86406	<.001	12.0735	45.3691
		price reduction and advertisement	-8.3530	7.10548	.726	-25.5863	8.8803
	price reduction and advertisement	control	37.0743*	7.03303	<.001	20.0167	54.1318
		price reduction	8.3530	7.10548	.726	-8.8803	25.5863

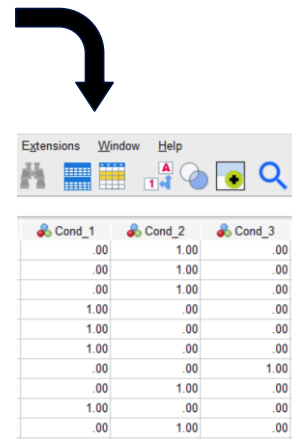
- SPSS also reports the requested *Post Hoc Tests* Tukey, LSD and Bonferroni
- All conditions are compared with each other
- SPSS reports the mean difference between the two compared groups, the standard error and significance of the comparison, as well as the confidence interval
- Here, we find significant comparisons between the control condition and the price reduction condition, as well as between the control condition and the price reduction and advertisement condition. The two experimental conditions do not differ from each other significantly, though.
- Tukey, Bonferroni and LSD come to the same conclusions. For real analyses, you would have pre-selected one of these to use

RQ8: Effect of supermarket promotion on sales
Alternative: linear regression

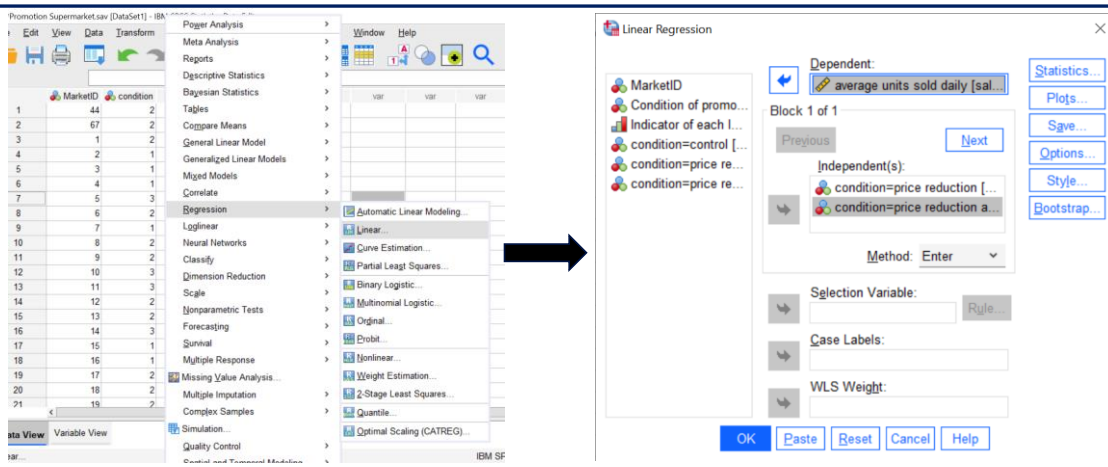
Before doing regression with a categorical predictor variable (with more than two groups) SPSS needs use to create dummy variables







RQ8: Effect of supermarket promotion on sales
Alternative: linear regression



Add two of the three dummy variables as independent variables. The dummy variable you leave out will be your reference group – the coefficients will represent differences from this reference group

RQ8: Effect of supermarket promotion on sales
Alternative: linear regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.441[a]	.195	.182	32.91108

a. Predictors: (Constant), condition=price reduction and advertisement, condition=price reduction

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34000.607	2	17000.304	15.695	<.001[b]
	Residual	140808.122	130	1083.139		
	Total	174808.729	132			

a. Dependent Variable: average units sold daily

b. Predictors: (Constant), condition=price reduction and advertisement, condition=price reduction

→ Matches the between-subjects effects table in one-way ANOVA

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	186.734	4.801		38.898	<.001
	condition=price reduction	28.721	6.864	.375	4.184	<.001
	condition=price reduction and advertisement	37.074	7.033	.472	5.271	<.001

→ Matches the mean differences in post-hoc tests table in one-way ANOVA

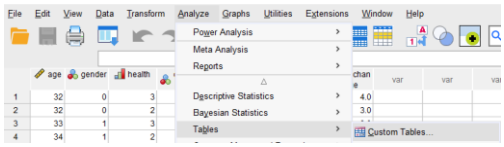
RQ 9: Effect of new pain treatment

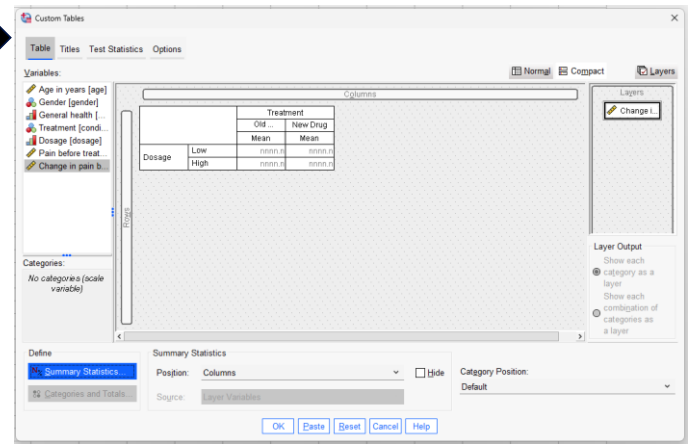
- Data set from RQ6

☛ Has the pain treatment helped the participants? Did the dosage influence the effectiveness of the treatment?

  - Custom tables: get descriptive statistics for each participant group
  - Two-way ANOVA: condition by dosage
  - Formatting graphs in output

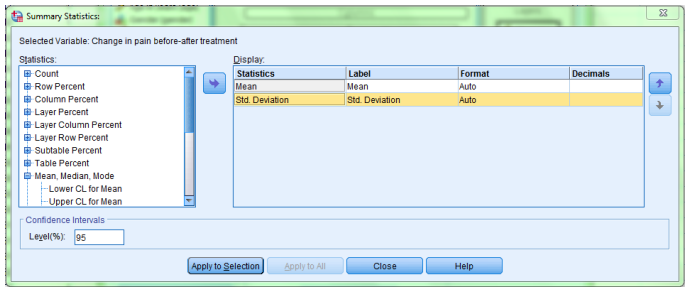
RQ 9: Effect of new pain treatment
Custom Tables

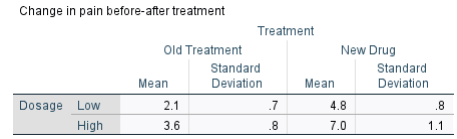




- Select *Custom Tables*... under *Tables* in the tab *Analyze*
- From the variable list on the left, pull the categorical variables by which you want to group your table to the *Columns (Treatment* variable*)* and *Rows (Dosage* variable*)* boxes in the preview
- On the top right, click on the button *Layers* to open the right hand box
- Drag the variable for which you want to receive statistics into the *Layers* box
- Select the Layers variable. On the bottom left, click on the button *Summary Statistics*...

Custom Tables



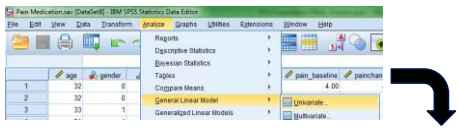


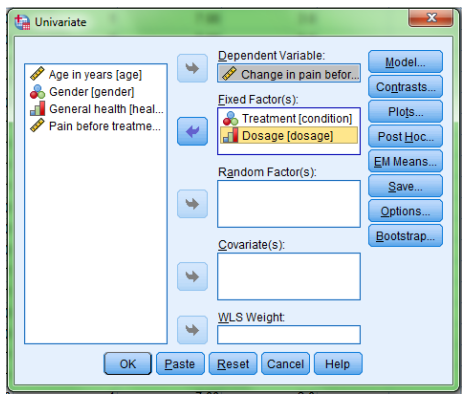
Change in pain before-after treatment

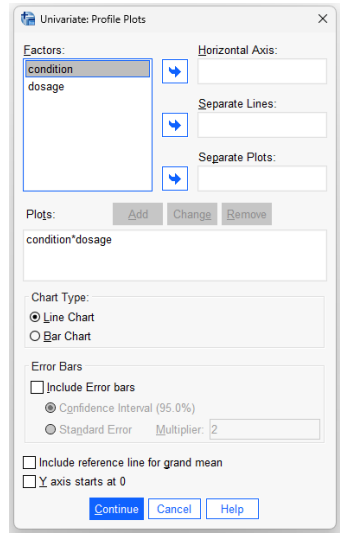
		Treatment			
		Old Treatment		New Drug	
		Mean	Standard Deviation	Mean	Standard Deviation
Dosage	Low	2.1	.7	4.8	.8
	High	3.6	.8	7.0	1.1

- From the Statistics box on the left, choose the statistics you want to receive and add them to the box on the right
- Click on *Apply to Selection* (nothing will happen), then click on *Close*
- In the main window click *OK*
- The output shows the selected statistics for each of the groups created by the categorical variables

RQ 9: Effect of new pain treatment
Two-way ANOVA

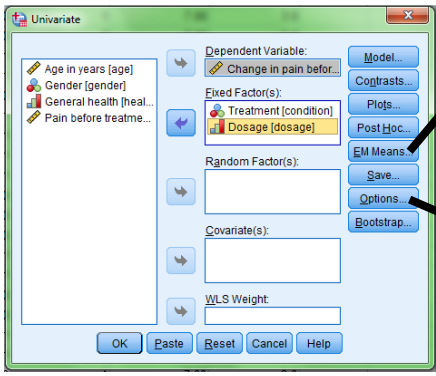




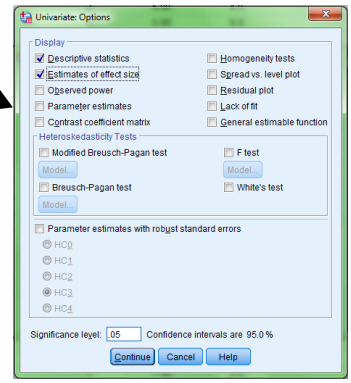


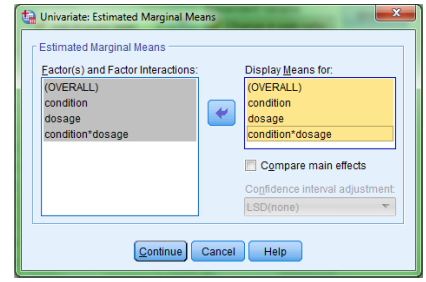
- Select *Univariate…* under *General Linear Model* in the tab *Analyze*
- Add the variable *Change in pain* to the *Dependent Variable* box
- Add the variables *Treatment* and *Dosage* to the *Fixed Factor(s)* box
- Click on *Plots…* to configure a plot for the output
- Add *dosage* to the *Horizontal Axis* and *condition* as *Separate Lines*, then *Add*
- click *Continue*
- This time we don't need post hoc tests as our two fixed factors only have two levels

RQ 9: Effect of new pain treatment
Two-way ANOVA



- Click on *EM Means...* and specify that we wantto *Display Means* for all variables





- Under *Options...* select which statistics you want to receive, in this case *Descriptive statistics* and *Estimates of effect size*

RQ 9: Effect of new pain treatment
Two-way ANOVA

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
Treatment	0	Old Treatment	30
	1	New Drug	30
Dosage	1	Low	30
	2	High	30

Descriptive Statistics

Dependent Variable: Change in pain before-after treatment

Treatment	Dosage	Mean	Std. Deviation	N
Old Treatment	Low	2.067	.7037	15
	High	3.600	.8281	15
	Total	2.833	1.0854	30
New Drug	Low	4.800	.7746	15
	High	7.000	1.0690	15
	Total	5.900	1.4468	30
Total	Low	3.433	1.5687	30
	High	5.300	1.9678	30
	Total	4.367	1.9997	60

- The *Univariate Analysis of Variance* tables present frequencies in the between-subject factor, descriptive statistics and the results of the analysis
- By checking the *Sig.* value in the row *condition* of the table on the right you can find that there is a main effect for the treatment condition, *Partial Eta Squared* shows a large effect size

Tests of Between-Subjects Effects

Dependent Variable: Change in pain before-after treatment

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	195.000[a]	3	65.000	88.925	<.001	.827
Intercept	1144.067	1	1144.067	1565.173	<.001	.965
condition	141.067	1	141.067	192.990	<.001	.775
dosage	52.267	1	52.267	71.505	<.001	.561
condition * dosage	1.667	1	1.667	2.280	.137	.039
Error	40.933	56	.731			
Total	1380.000	60				
Corrected Total	235.933	59				

a. R Squared = .827 (Adjusted R Squared = .817)

- There is also a large main effect for dosage
- The interaction (row *condition*dosage*) is not significant (p=0.137)

RQ 9: Effect of new pain treatment
Two-way ANOVA

Estimated Marginal Means

1. Grand Mean

Dependent Variable: Change in pain before-after treatment

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
4.367	.110	4.146	4.588

2. Treatment

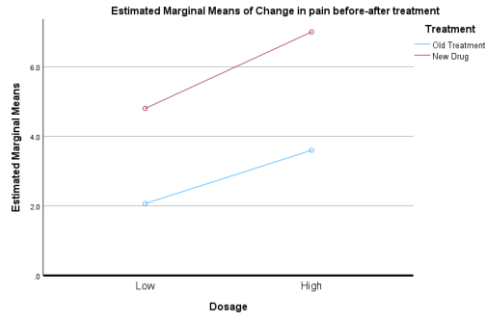
Dependent Variable: Change in pain before-after treatment

Treatment	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Old Treatment	2.833	.156	2.521	3.146
New Drug	5.900	.156	5.587	6.213

3. Dosage

Dependent Variable: Change in pain before-after treatment

Dosage	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Low	3.433	.156	3.121	3.746
High	5.300	.156	4.987	5.613



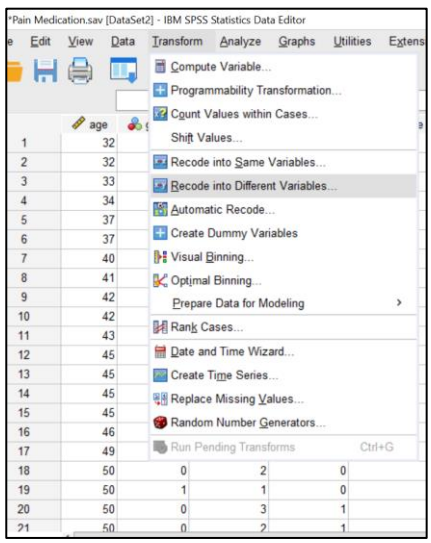
4. Treatment * Dosage

Dependent Variable: Change in pain before-after treatment

Treatment	Dosage	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Old Treatment	Low	2.067	.221	1.624	2.509
	High	3.600	.221	3.158	4.042
New Drug	Low	4.800	.221	4.358	5.242
	High	7.000	.221	6.558	7.442

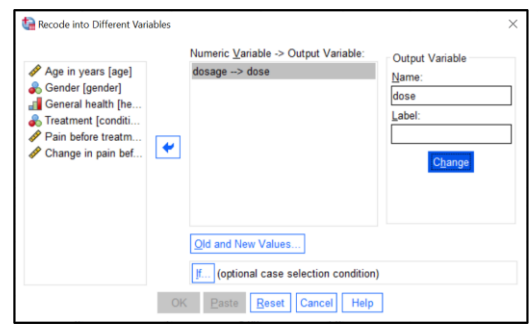
- The *Estimated Marginal Means* tables return confidence intervals for the main effect and the interaction variables
- The plot makes it easy to visualise the two main effects
- Here we have approximately parallel lines (no significant interaction)

RQ 9: Effect of new pain treatment
Alternative: linear regression



To use regression for an interaction between two binary categorical variables, we need variables coded as 0 or 1 only

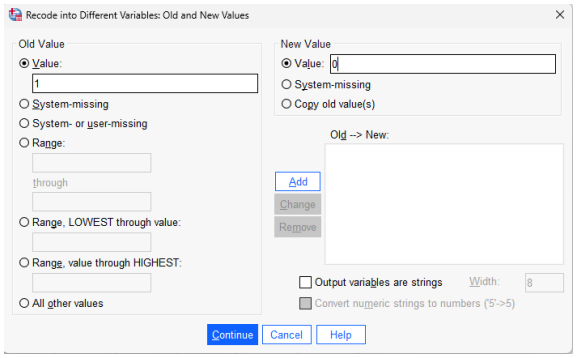
We will recode *dosage* as 0/1 instead of 1/2:
- Select *Recode into Different Variables*, under *Transform*
- Add the *dosage* variable to the *Numeric Variable* box
- Enter the new name of the recoded variable: *dose*, and select *Change*



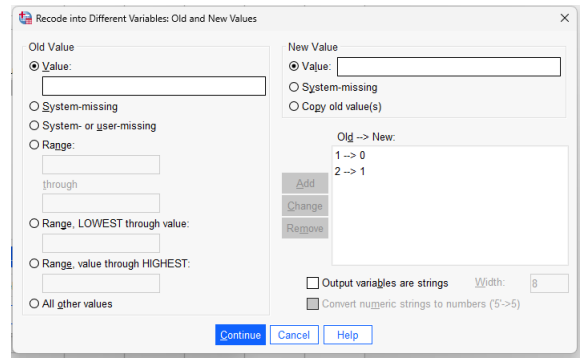
RQ 9: Effect of new pain treatment
Alternative: linear regression

- Select Old and New Values…

We will recode *dosage*=1 as *dose*=0:
- Select *Old Value*: *Value*, and enter 1
- Select *New Value*: *Value* and enter 0
- Select *Add* to add this the Old->New box

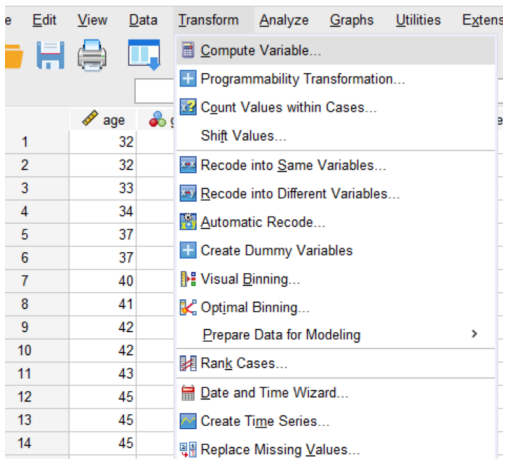


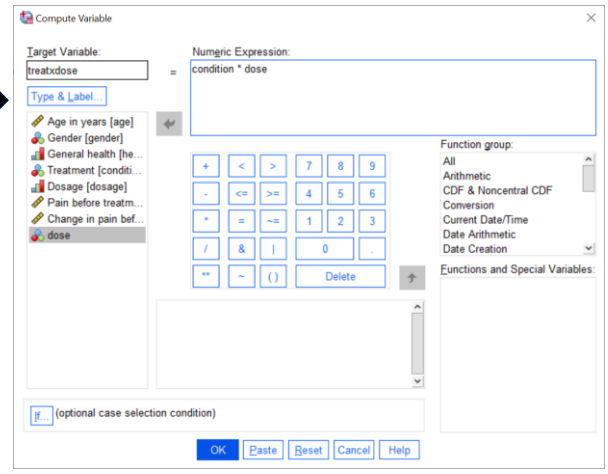
Next, we will recode *dosage*=2 as *dose*=1:
- Select *Old Value*: *Value*, and enter 2
- Select *New Value*: *Value* and enter 1
- Select *Add* to add this the Old->New box
- Select *Confirm* and then *OK*



RQ 9: Effect of new pain treatment
Alternative: linear regression

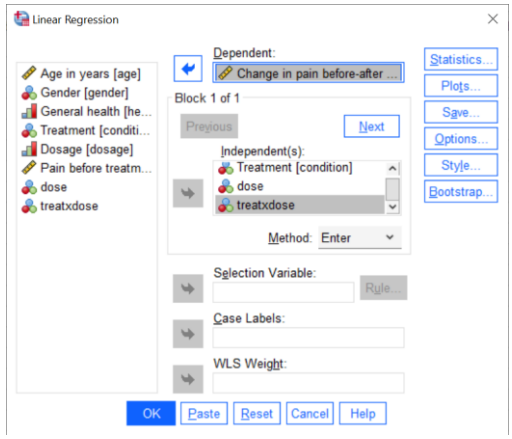
We then create an interaction term, by using *Compute Variable* to create a new variable called *treatxdose* that equals *condition* multiplied by *dose* (NOT *dosage*, since we want a binary variable)





RQ 9: Effect of new pain treatment
Alternative: linear regression

We can now put our binary categorical variables and interaction variable into the regression model.
- Run a linear regression from *Analyze* > *Regression* > *Linear* as before, using *Change in pain* as the Dependent variable, and *Treatment* and your new *dose* and *treatxdose* binary variables as Independent variables



ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	195.000	3	65.000	88.925	<.001[b]
	Residual	40.933	56	.731		
	Total	235.933	59			

a. Dependent Variable: Change in pain before-after treatment

b. Predictors: (Constant), treatxdose, dose, Treatment

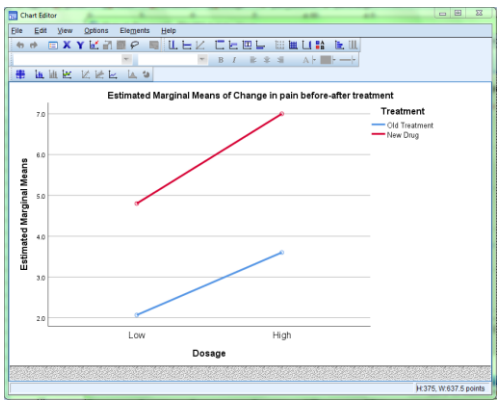
Coefficients[a]

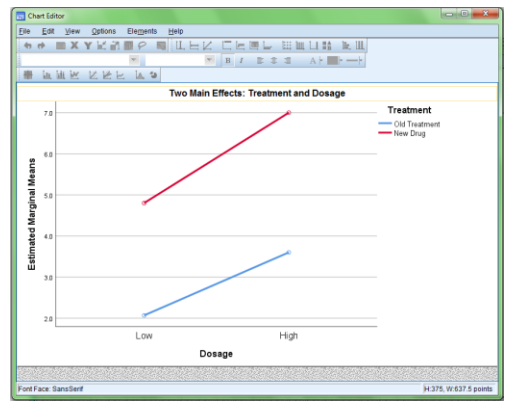
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2.067	.221		9.362	<.001
	Treatment	2.733	.312	.689	8.755	<.001
	dose	1.533	.312	.387	4.912	<.001
	treatxdose	.667	.441	.146	1.510	.137

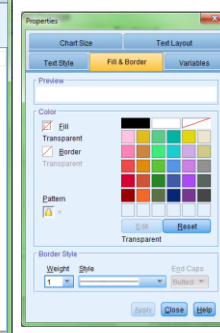
a. Dependent Variable: Change in pain before-after treatment

RQ 9: Effect of new pain treatment
Formatting graphs

- Double-click on the plot in the output to open the *Chart Editor*
- Double-click on the titles to edit, click once more to change the wording

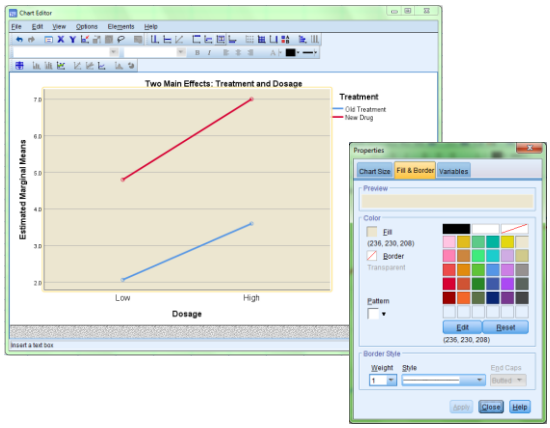




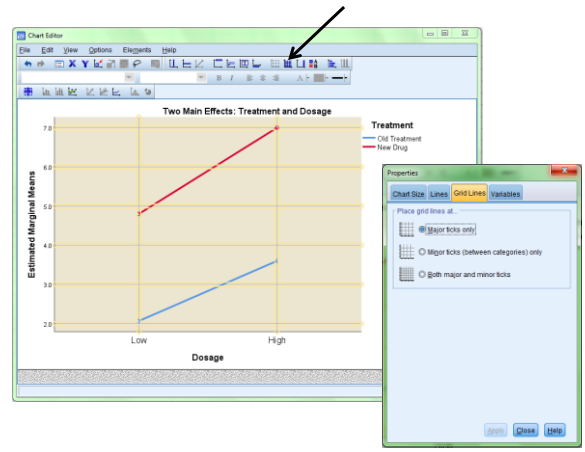


RQ 9: Effect of new pain treatment
Formatting graphs

- Double-click on the plot background to open more formatting options



- Click on the grid symbol to add a grid to the chart



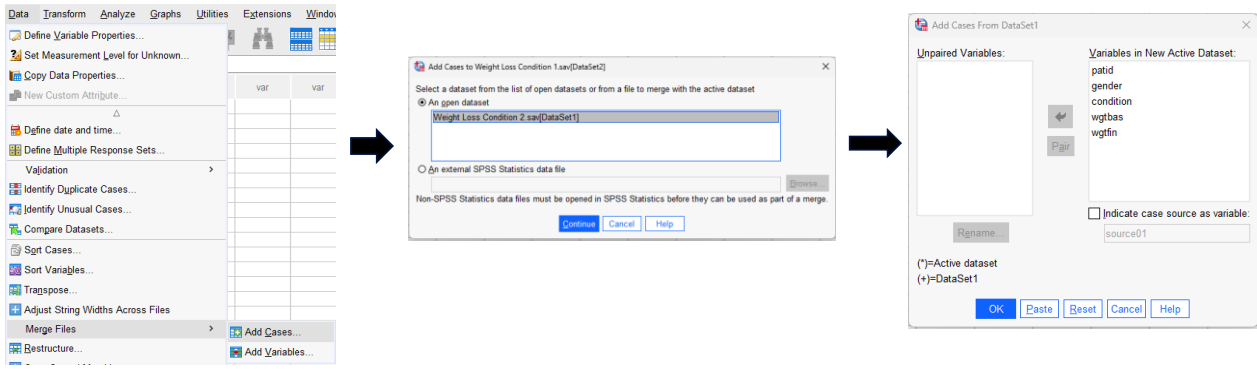
RQ 10: Effect of weight loss programme

- Data set of 64 participants who participated in a weight loss study
- 4 variables: gender, condition (control or treatment), weight at baseline, weight at finish

☛ Have participants in the weight loss treatment condition lost more weight?

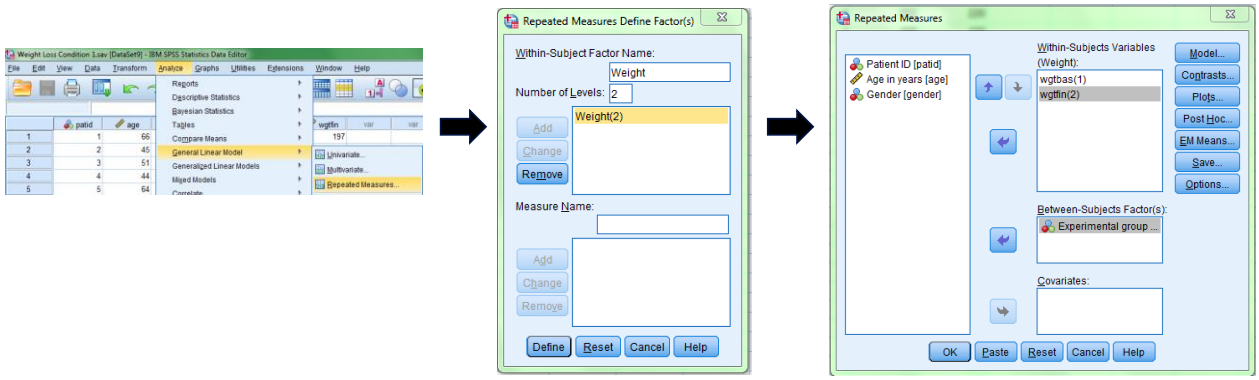
- Merge files: Weight Loss Condition 1 (control) and Condition 2 (treatment)
- Repeated/mixed ANOVA
- Cleaning output
- Formatting tables
- Exporting graphs and tables

RQ 10: Effect of weight loss programme
Merge files



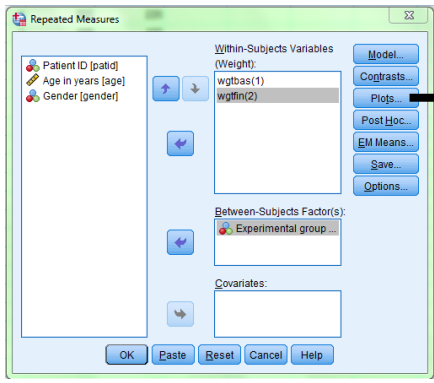
- Make sure that the two data sets that you want to merge are opened in SPSS, and that Weight Loss Condition 1.sav is open in the current data window
- Select *Add Cases…* under *Merge Files* in the tab *Data*
- Select the Weight Loss Condition 2.sav in the box and click *Continue*
- Check that there are no unpaired variables in the left box, then click *OK*

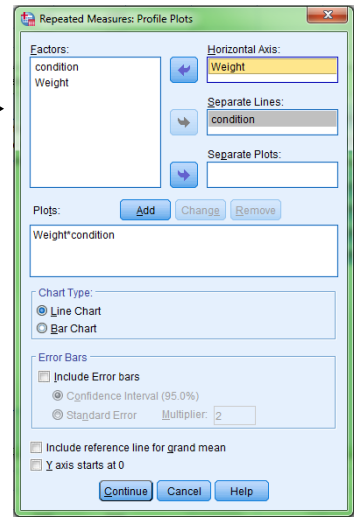
RQ 10: Effect of weight loss programme
Repeated/mixed ANOVA



- Select *Repeated Measures…* under *General Linear Model* in the tab *Analyze*
- Give a name to the Within-Subject Factor and enter "2" in the box asking for the number of levels of this variable (weight baseline and weight finish), then click *Define*
- Drag the variables wgtbas and wgtfin into the *Within-Subjects Variables* box, then add the experimental group variable to the *Between-Subjects Factor(s)*box

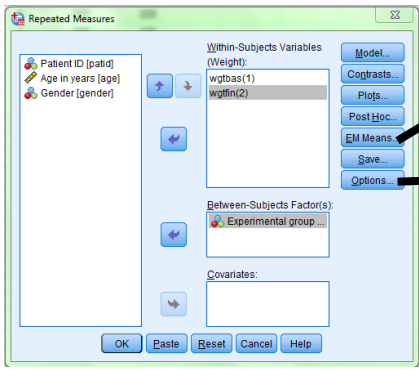
RQ 10: Effect of weight loss programme
Repeated/mixed ANOVA

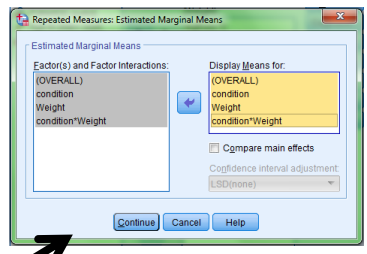


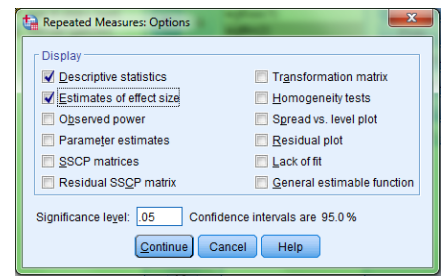


- Click on *Plots...* to configure a plot. Remember to assign the separate lines to the categorical variable. Then click *Continue*
- We don't need post hoc tests as our factors only have two levels

RQ 10: Effect of weight loss programme
Repeated/mixed ANOVA







- Go to *EM Means...* to choose the variables you would like to receive estimated marginal means for
- Under *Options...* select which statistics you want to receive, in this case *Descriptive statistics* and *Estimates of effect size*
- Click Continue and OK

RQ 10: Effect of weight loss programme
Repeated/mixed ANOVA

- The Repeated Measures ANOVA returns several tables, including frequencies in the between-subject factor, descriptive statistics and the results of the analysis
- Since we only have two levels in our within-subject factor, we can ignore the Mauchly's Test of Sphericity, as sphericity is assumed

Between-Subjects Factors

		Value Label	N
Experimental group	1	Control	32
	2	Treatment	32

Descriptive Statistics

	Experimental group	Mean	Std. Deviation	N
Weight Baseline	Control	198.72	32.908	32
	Treatment	198.38	32.928	32
	Total	198.55	32.656	64
Final weight	Control	197.63	32.750	32
	Treatment	194.13	32.953	32
	Total	195.88	32.638	64

RQ 10: Effect of weight loss programme
Repeated/mixed ANOVA

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Weight	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Weight	Linear	228.445	1	228.445	176.253	<.001	.740
Weight * condition	Linear	79.695	1	79.695	61.488	<.001	.498
Error(Weight)	Linear	80.359	62	1.296			

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	4978195.695	1	4978195.695	2303.093	<.001	.974
condition	118.195	1	118.195	.055	.816	.001
Error	134014.609	62	2161.526			

- By checking the *Sig.* value in the rows of *Weight* and *Weight*condition* of the *Tests of Within-Subjects Contrasts* table you can find that there is a main effect for weight change, as well as a significant interaction between the condition and weight change, Partial Eta Squared shows a large effect size in both cases
- The *Tests of Between-Subjects Effects* table reveals that there is no significant main effect of the condition

RQ 10: Effect of weight loss programme
Repeated/mixed ANOVA

Estimated Marginal Means

1. Grand Mean

Measure: MEASURE_1

Mean	95% Confidence Interval	
	Lower Bound	Upper Bound
197.211	188.996	205.425

Std. Error: 4.109

2. Experimental group

Measure: MEASURE_1

Experimental group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Control	198.172	5.812	186.555	209.789
Treatment	196.250	5.812	184.633	207.867

3. Weight

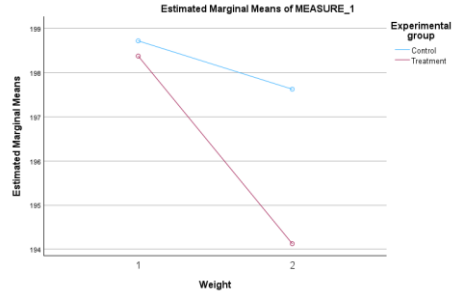
Measure: MEASURE_1

Weight	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	198.547	4.115	190.322	206.772
2	195.875	4.106	187.666	204.084

4. Experimental group * Weight

Measure: MEASURE_1

Experimental group	Weight	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Control	1	198.719	5.819	187.087	210.351
	2	197.625	5.807	186.016	209.234
Treatment	1	198.375	5.819	186.743	210.007
	2	194.125	5.807	182.516	205.734



- The *Estimated Marginal Means* tables return confidence intervals for the main effect and the interaction variables
  The plot makes it easy to visualise the interaction of the two factors

RQ 10: Effect of weight loss programme
Formatting tables

Double-click on the table you want to format to activate it

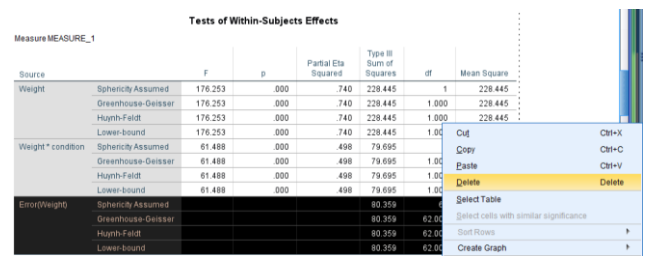
- To change the headings in the table, double-click on the heading in question and edit

- To delete unnecessary rows or columns, select the respective cells and right-click to get to the option *Delete*

Tests of Within-Subjects Effects

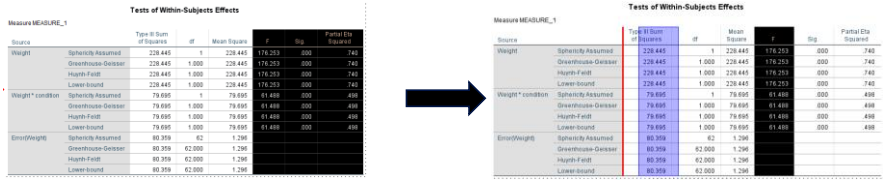
Measure MEASURE_1

Source		F	p Sig.	Partial Eta Squared	Type III Sum of Squares	df	Mean Square
Weight	Sphericity Assumed	176.253	.000	.740	228.445	1	228.445
	Greenhouse-Geisser	176.253	.000	.740	228.445	1.000	228.445
	Huynh-Feldt	176.253	.000	.740	228.445	1.000	228.445
	Lower-bound	176.253	.000	.740	228.445	1.000	228.445
Weight * condition	Sphericity Assumed	61.488	.000	.498	79.695	1	79.695
	Greenhouse-Geisser	61.488	.000	.498	79.695	1.000	79.695
	Huynh-Feldt	61.488	.000	.498	79.695	1.000	79.695
	Lower-bound	61.488	.000	.498	79.695	1.000	79.695
Error(Weight)	Sphericity Assumed				80.359	62	1.296
	Greenhouse-Geisser				80.359	62.000	1.296
	Huynh-Feldt				80.359	62.000	1.296
	Lower-bound				80.359	62.000	1.296

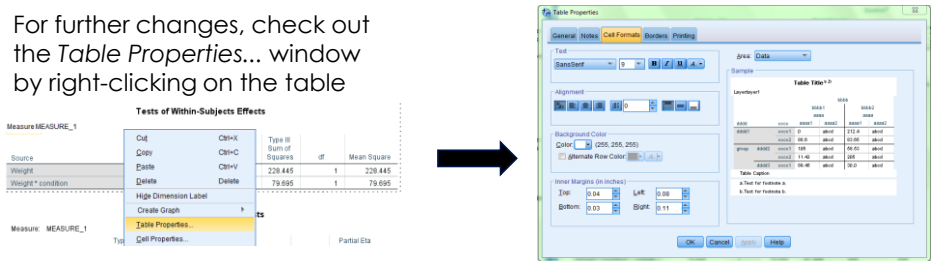


RQ 10: Effect of weight loss programme
Formatting tables

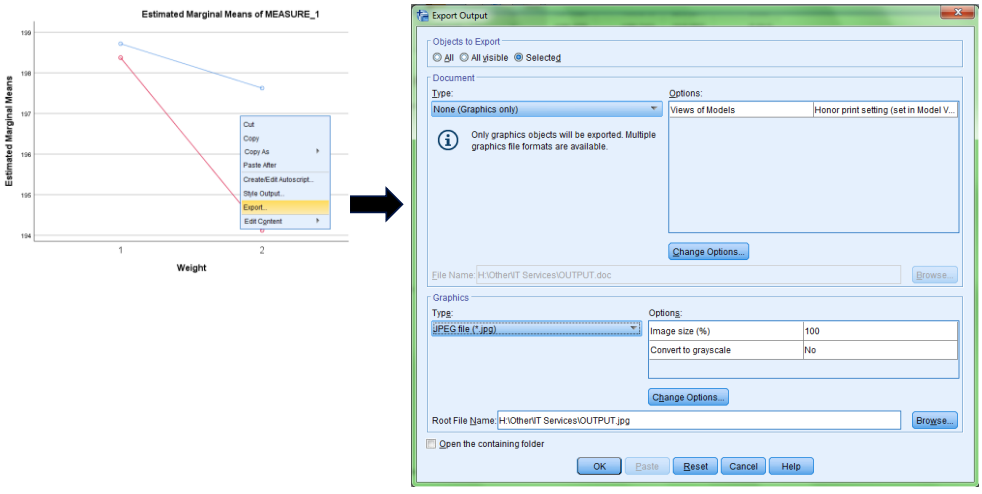
To change the order of the rows or columns, select the respective cells and drag them to the desired place



For further changes, check out the *Table Properties...* window by right-clicking on the table

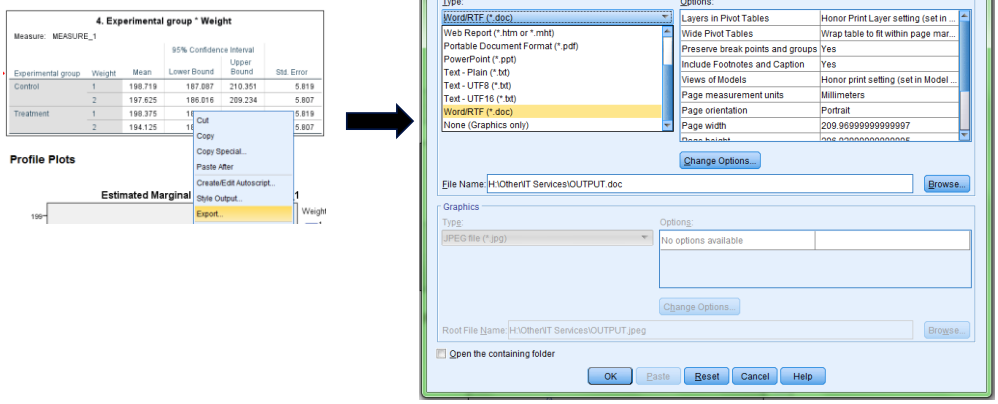


RQ 10: Effect of weight loss programme
Exporting graphs



- Right-click on the graph to show the options and click on *Export...*
- In the document section, select the type *None (Graphics only)*
- In the Graphics section, select the type of format you want to export to
- Click on *Browse...* to select the folder where you want to save the exported file to
- Click *OK*

RQ 10: Effect of weight loss programme
Exporting tables



- Right-click on the table to show the options and click on *Export...*
- In the document section, select the type of format you want to export to
- Click on *Browse...* to select the folder where you want to save the exported file to
- Click *OK*

Additional features of SPSS

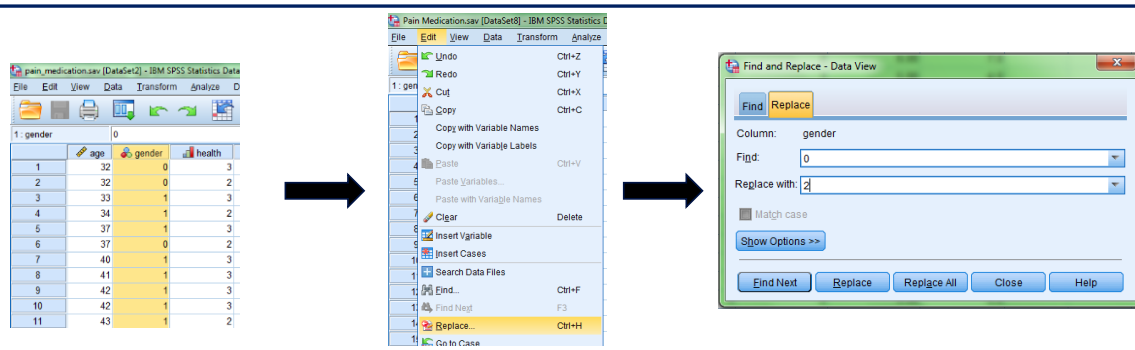
Additional Functionality

- Find and Replace
- Explore Data (including normality tests)
- Crosstabs
- Reliability Analysis
- SPSS Settings

Find and Replace

- Can exchange specific values of a column by finding and replacing them
- Similar to *recode into same variables*
- E.g. changing coding of male gender from 0 to 2 in the pain medication data

Find and Replace

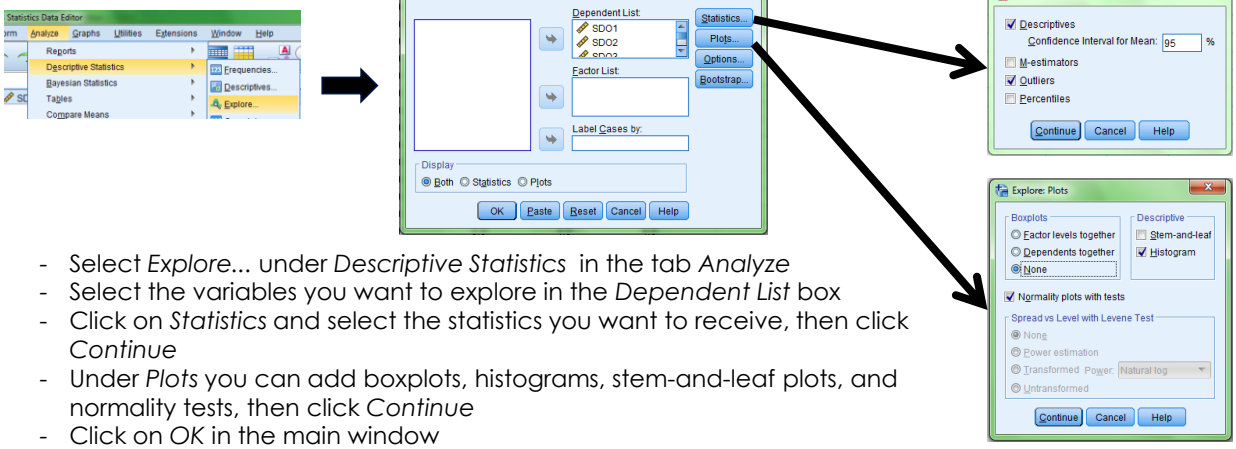


- Select the column in which you want to replace the values by clicking on the variable name
- Click on *Find...* in the tab *Edit*
- Select the tab *Replace*, then define which value you want to find and which value you want to replace it with
- Click on *Replace All* to replace the values in the selected column

Explore data

- Presents descriptive statistics
- Can return extreme values
- Can create histograms, boxplots, Q-Q plots and stem-and-leaf plots
- Can test for normal distribution
- Use Reliability data

Explore data



- Select *Explore...* under *Descriptive Statistics* in the tab *Analyze*
- Select the variables you want to explore in the *Dependent List* box
- Click on *Statistics* and select the statistics you want to receive, then click *Continue*
- Under *Plots* you can add boxplots, histograms, stem-and-leaf plots, and normality tests, then click *Continue*
- Click on *OK* in the main window

Explore data

Descriptives

			Statistic	Std. Error
SDO1	Mean		3.966	.0789
	95% Confidence Interval for Mean	Lower Bound	3.810	
		Upper Bound	4.122	
	5% Trimmed Mean		4.018	
	Median		4.000	
	Variance		.734	
	Std. Deviation		.8567	
	Minimum		2.0	
	Maximum		5.0	
	Range		3.0	
	Interquartile Range		1.0	
	Skewness		-.681	.223
	Kurtosis		.041	.442
SDO2	Mean		4.559	.0596

Extreme Values

			Case Number	Value
SDO1	Highest	1	22	5.0
		2	27	5.0
		3	28	5.0
		4	29	5.0
		5	32	5.0[a]
	Lowest	1	113	2.0
		2	102	2.0
		3	98	2.0
		4	73	2.0
		5	72	2.0[b]
SDO2	Highest	1	4	5.0
		2	6	5.0
		3	7	5.0
		4	9	5.0

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SDO1	.287	118	.000	.831	118	.000
SDO2	.362	118	.000	.624	118	.000
SDO3	.327	118	.000	.738	118	.000
SDO4	.381	118	.000	.623	118	.000
SDO5	.269	118	.000	.858	118	.000
SDO6	.326	118	.000	.685	118	.000

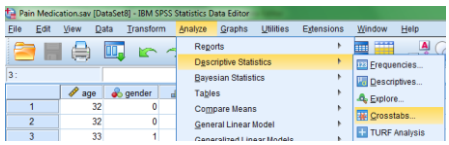
a. Lilliefors Significance Correction

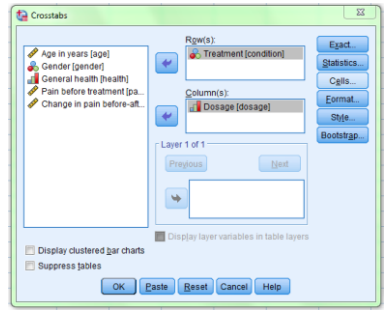
- SPSS returns a table with the standard set of descriptive statistics
- If the *Outliers* option was selected under *Statistics*, SPSS displays the most extreme values for each variable
- If the *Normality plots with tests* option was selected under *Plots,* SPSS returns Kolmogorov-Smirnov and Shapiro-Wilk test results. In both cases, significant results indicate that the distribution is significantly non-normal
- SPSS returns the requested plots

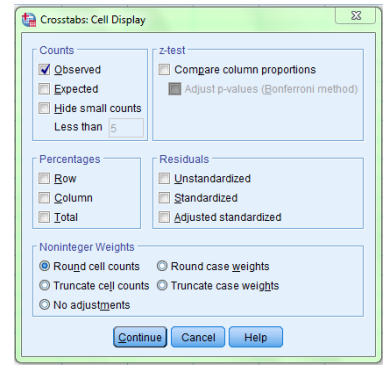
Crosstabs

- Allows you to create tables which group the sample according to categorical variables
- Simple version of custom tables
- Use pain medication data

Crosstabs





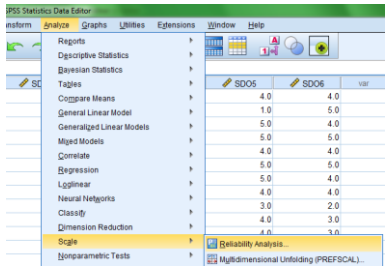


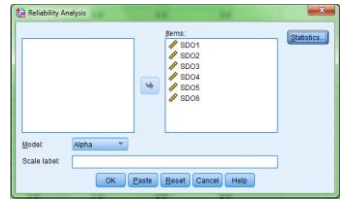
- Select *Crosstabs…* under *Descriptive Statistics* in the tab *Analyze*
- Add the variable Treatment and Dosage into the boxes for *Row(s)* and *Coloumn(s)*
- Click on *Cells...* to specify the content of the cells of the crosstabs, in this case we want the observed count in order to see how many people are in each condition, then click *Continue*
- Click *OK*

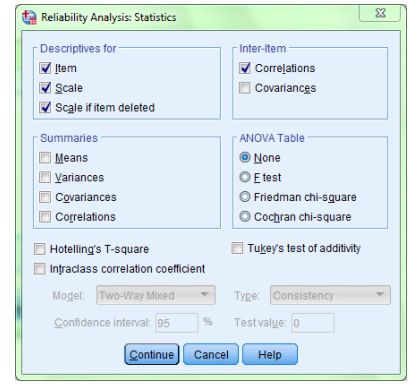
Reliability Analysis

- Allows you to assess internal reliability of a questionnaire with several items
- Returns Cronbach's Alpha
- Can request Cronbach's Alpha if items were deleted

Reliability Analysis







- Select *Reliability Analysis...* under *Scale* in the tab *Analyze*
- Select the item variables you want to test and add them to the box on the right
- Open the *Statistics...* window and select the statistics of interest, then click *Continue*
- Click *OK* in the main window

Reliability Analysis

1

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.701	.727	6

2

Item Statistics

	Mean	Std. Deviation	N
SDO1	3.966	.8567	118
SDO2	4.559	.6477	118
SDO3	4.441	.6608	118
SDO4	4.525	.8135	118
SDO5	3.788	1.0115	118
SDO6	4.449	.7578	118

3

Inter-Item Correlation Matrix

	SDO1	SDO2	SDO3	SDO4	SDO5	SDO6
SDO1	1.000	.450	.449	.320	.297	.274
SDO2	.450	1.000	.438	.459	.117	.476
SDO3	.449	.438	1.000	.265	.294	.301
SDO4	.320	.459	.265	1.000	.022	.335
SDO5	.297	.117	.294	.022	1.000	.114
SDO6	.274	.476	.301	.335	.114	1.000

4

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
SDO1	21.763	6.200	.545	.326	.622
SDO2	21.169	6.860	.583	.427	.625
SDO3	21.288	6.959	.535	.309	.637
SDO4	21.203	6.933	.389	.249	.675
SDO5	21.941	6.962	.239	.135	.743
SDO6	21.280	6.972	.428	.254	.663

5

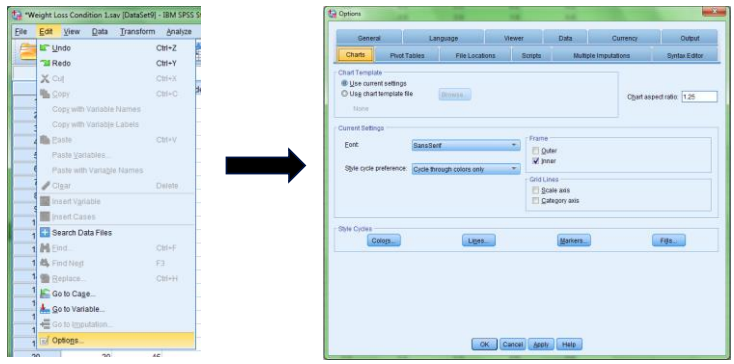
Scale Statistics

Mean	Variance	Std. Deviation	N of Items
25.729	9.259	3.0429	6

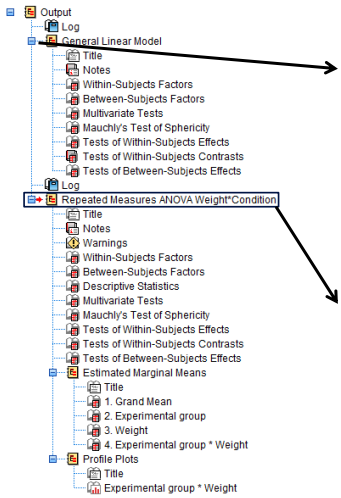
- The first table shows the overall Cronbach's Alpha, when all 6 items are included
- The second table depicts item statistics, the third table shows the requested inter-item correlations
- In the last column of table 4 you can find the potential Cronbach's Alpha when deleting the respective item
- Table 5 shows the statistics of the scale when including all 6 items

SPSS Settings

- Can set defaults for the formatting of charts, tables, outputs generally, etc.



Cleaning Output



- It is possible to delete results from the output
- Or if you just want to hide it to gain more overview, you can click on the minus symbols next to the mother tabs

- You might want to give your analyses different names so they are not all called the same
- Double-click on the rows you want to rename and start editing

Syntax window

- File which gathers all the code produced during analysis
- Can be used to keep track of actions
- Can be used to reproduce analysis easily
- Can be used to modify output (especially graphs) without having to re-do lots of steps

Further Exercises

1. For the condition variable in the pain_medication file, find and replace the control group's value from 0 to 1 and the treatment group's value from 1 to 2.
2. In the pain_medication file, create crosstabs with the categorical variables gender and general health to find out how many participants fall into each group.
3. In the breakfast file, explore the satiety variables and check whether they are normally distributed. Use a syntax file to perform this task.