Technische Universität München
TUM School of Management
Marketing and Consumer Research

SPSS exercise

As part of Quantitative Research Methods

Wintersemester 2013/2014

Outline:

1 Everything you ever wanted to know about statistics

2 Parameter testing

3 Distribution (non-parametric) testing, t-Test

4 4 Correlation analysis

5.2 Regression analysis

6 Theoretical basis of nonparametric tests

1 Everything you ever wanted to know about statistics

1.1 Preliminary notes about SPSS

- SPSS is a statistical programme for the analysis and visualisation of data
- It is possible to import data with another format into SPSS, e.g. excel data can be imported
- SPSS is frequently used in companies (e.g. market research)
- SPSS enables the user to handle a great number of data (more than with Excel) and to combine different data sets.

1.2 Change language:

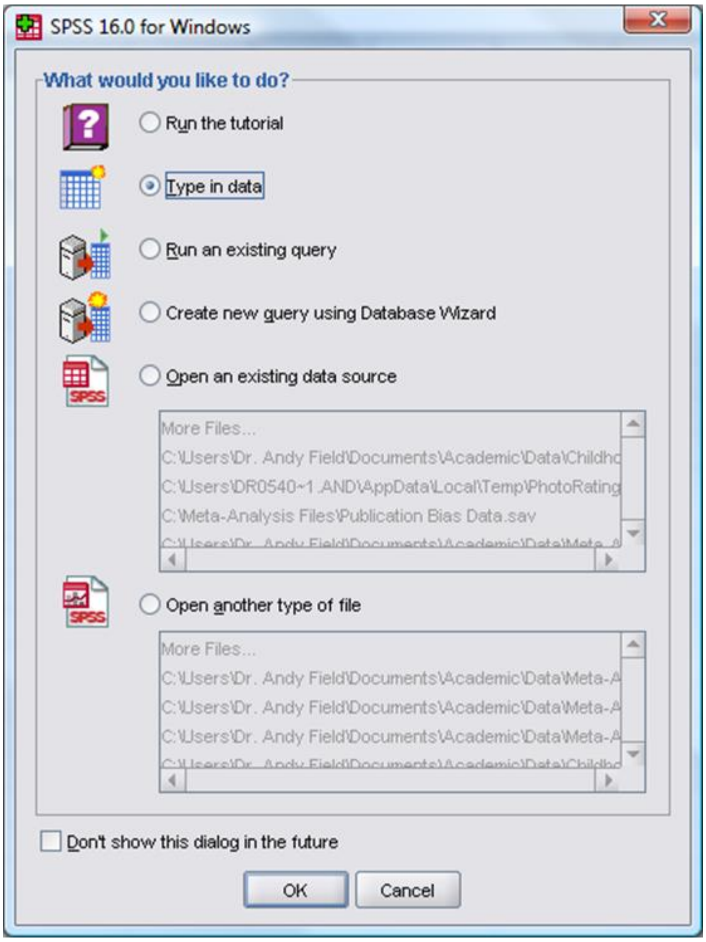
Edit>Options>General (Bearbeiten → Optionen → Allgemein)

Select Language „English" under Output as well as User Interface (Sprache "English" unter Ausbabe und Benutzeroberfläche)

1.3 Information about the data set

- Given in data descriptions (own description or external description)
    - Sometimes, the data description is already incorporated within SPSS (via labels and value labels etc.). This information can be printed
    - File→ Display Data File Information → Working File OR External File
- You should always be very accurate with your data and never work without a data description
    - When you enter your own data, always enter your information in variable view (e.g. question wordings, answer categories)

1.3 Starting SPSS



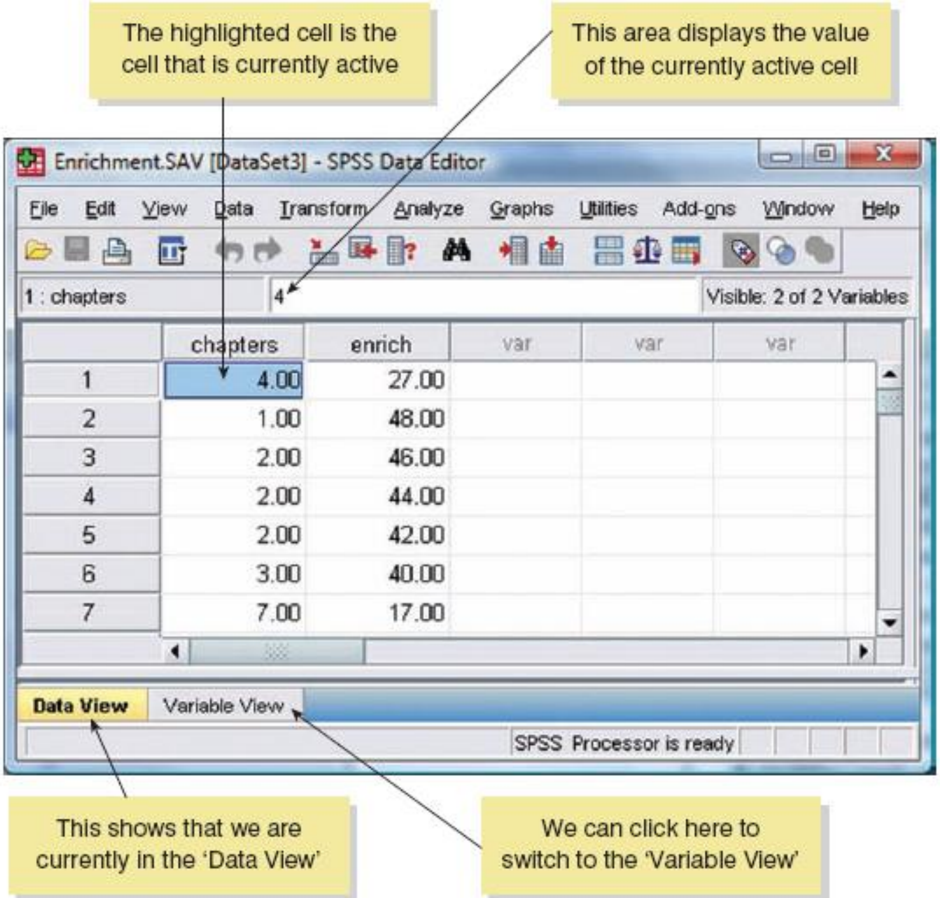
When opening SPSS you can select if you want to type in new data or use an existing data set. If you select "cancel" you will get to the main window that does not contain any data.

SPSS uses 3 windows: **data-, syntax and SPSS-viewer:**

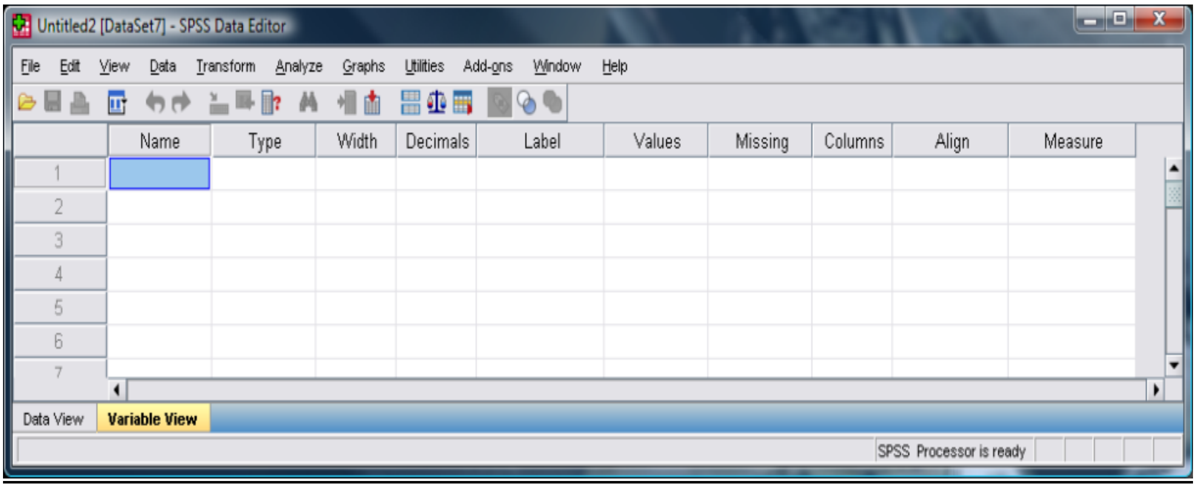
- **SPSS data-editor** (appears when the programme is opened): shows the data in the **data view** and the characteristics of the variables under **variable view (file ending .sav)**
- **SPSS Viewer** (opens automatically after an analysis): shows the results. The SPSS viewer shows the results as a hierarchy. One analysis can result in many tables. Clicking on the + or – symbol aggregates or disaggregates the results. Clicking on one object shows the respective table. **(file ending: .spv)**
- ***SPSS Syntax-Editor*** (can be opened with Aufruf über Menü FILE ⇨ NEW⇨ SYNTAX): instructions are entered in this window. The instructions are executed by clicking the ▶ **(file ending: .sps)**

1.3.1 Data Editor

FIGURE 3.3
The SPSS Data Editor



1.3.2 Variable View



- In the variable view you can enter important information for every variable
- Every variable should have a unique **name**
    - Each variable has to be named in column "**Name**".
      The default is VAR00001
- The **type** of data is important as it determines what kind of analyses you can do with the variable
- Different variable types exist
    - Numeric
        - Numbers (e.g. 7, 0, 120)
    - String
        - Letters (e.g. 'Andy', 'Idiot')
    - Currency
        - Currency (e.g. £20, $34, €56)
    - Date
        - Dates (e.g. 21-06-1973, 06-21-73, 21-Jun-1973)

If you want to create a string variable:

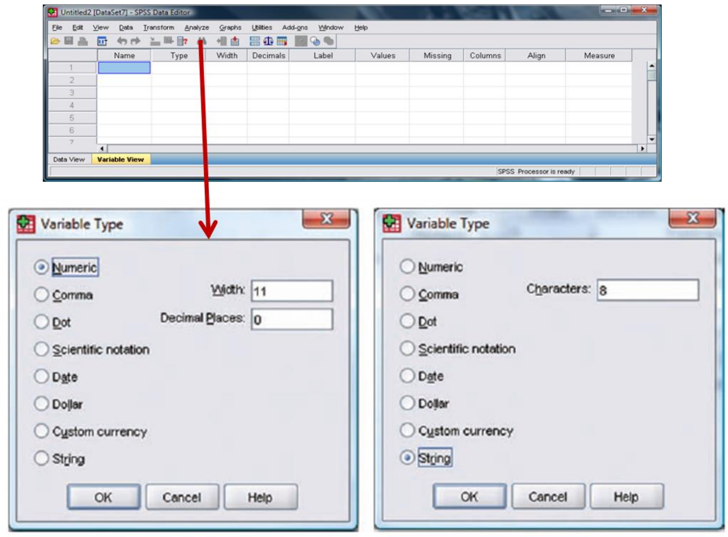


FIGURE 3.5
Defining a string variable in SPSS

Alternatively, you can define the variable as date:

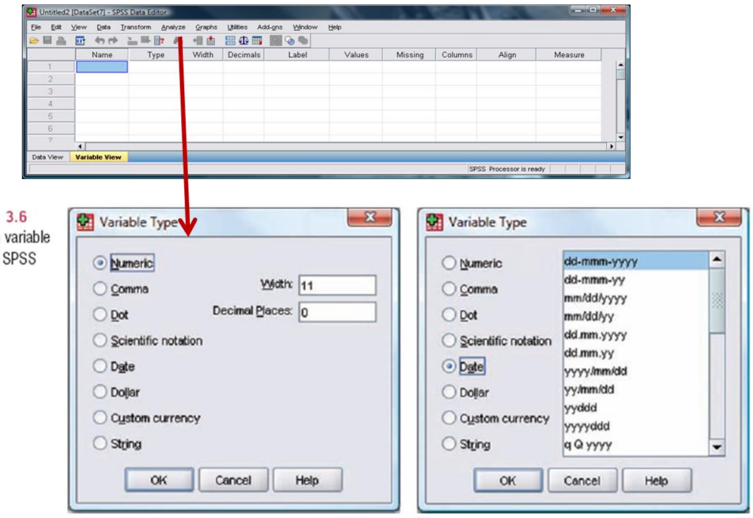


FIGURE 3.6
Defining variable types in SPSS

For coding/ categorical variables you need to define categories:

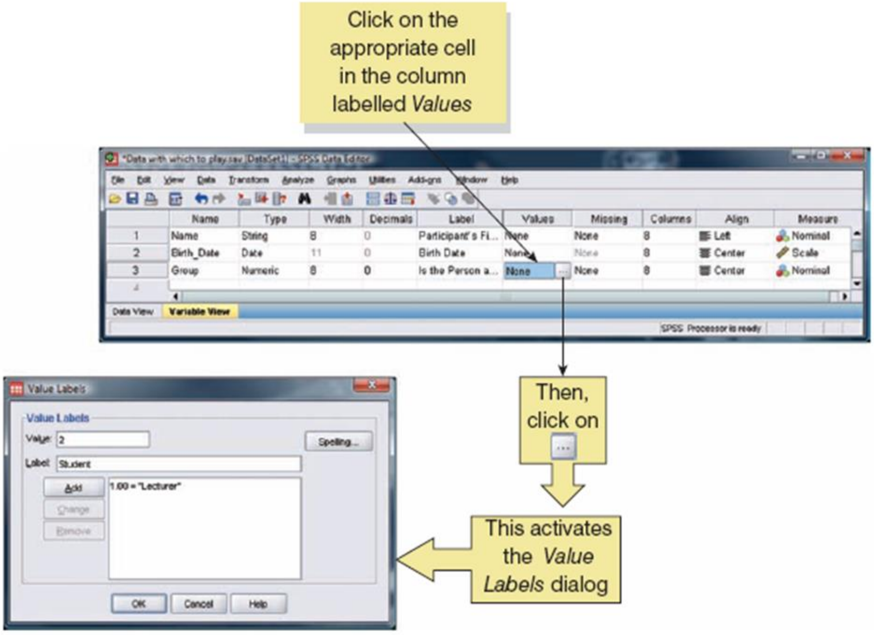


FIGURE 3.7 Defining coding variables and their values in SPSS

- Column **"Type":** SPSS can differentiate between different measurement scales (click on … to show selection)
  - Most often: numeric = numerical values or string = letters, words
- Column "labels": allows giving more information about the variable, e.g. the survey question
- Column "values": click on … to enter more data: value labels:
  - Value: 1
  - Label = Male → Add
  - Value : 2
  - Label = female → Add

1.3.3 The Viewer Window

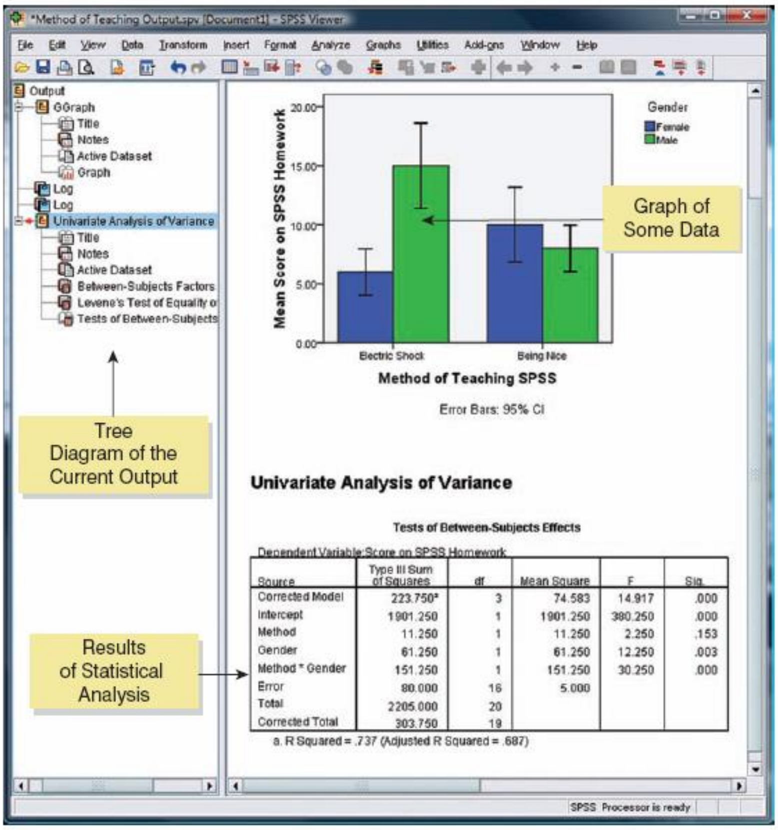


FIGURE 3.10 The SPSS Viewer

- When you work with several data sets or you work on several days on your data set but you want to save your output within one output file only, saved it the first time it opens. When you open SPSS again on another occasion, you have to open this old output manually and close the new output viewer that automatically opens. SPSS will then produce the output in the current SPSS output viewer.

1.3.4 The Syntax Window

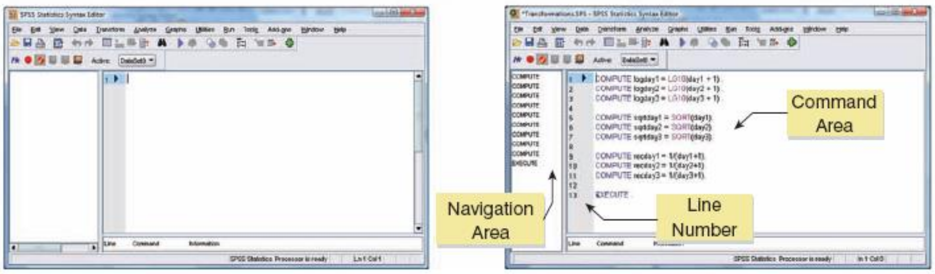


FIGURE 3.11 A new syntax window (top) and a syntax window with some syntax in it (bottom)

1.4 Opening SPSS data

- FILE/ OPEN: data/syntax/ output/ script
- Important: save all three viewers separately and in between different analyses and data adjustments

1.5 Missing values

Missing values can be problematic in empirical analyses. SPSS differentiates two kinds of missing values: **system-missing values and user-missing values**.

Recoding of missing values (BE CAREFUL)

- There are (rare) cases if it is advisable to recode missing values in the data set
- It is possible to recode missing values, e.g. to the value 0 or -99999.

Defining missing values under variable view:

- Column "Missing values": different selection options

Transform/Recode into same variables/ Select variable/ old and new values.../old value: system-missing/ new value: value: type 0 / Add/ continue/ ok

1.6 Data handling

- Data editing only under data view!!
- EDIT/ GO TO CASE…: easy navigation
- DATA/ SORT CASES: allows to order the data according to certain variable of intrerest in ascending or descending order
- **Alternative to menu navigation: using the syntax**:
    - **SORT CASES BY XY.**
    - Important: every command ends with a point!!!
    - Instead of entering the syntax by hand we can select the "PASTE" button instead of clicking ok → paste will enter the correct syntax for each command we can select in the SPSS menu in the syntax editor: useful if we need to repeat a certain command several times or we need to change the variables
- **MERGING FILES**
    - DATA / MERGE FILES/ ADD CASES: selection of cases out of an open or ext
    - Also possible to ADD VARIABLES, important to consider key variables by which the programme should order the variables (make sure that the cases are ordered the same in both data sets and both data sets contain the same number of cases)

- **SELECT CASES**
  - It can be necessary to restrict the analysis to certain subgroups in the data set, e.g. only females
  - DATA / SELECT CASES: if condition is satisfied: select if gender = 2 → continue; output: filter out unselected cases, copy selected cases to a new data set, delete unselcted cases (**ATTENTION**, make sure to have an original hard copy of the data set just in case…) → ok will select cases. DATA / SELECT CASES/ RESET to reset the selection
  - Note the new "filter" variable
  - Alternative syntax:

1.7 Transformation of variables

- In most cases it will be necessary to transform the variables of interest further to be able to use them in an analysis

EXERCISE: Transformation of variables:

DataSet: ChickFlick.sav

- Divide the variable "arousal" into 3 groups:
  - low: physical arousal up to 13
  - medium: physical arousal between 14 and 27
  - high: physical arousal higher than 27

```
compute arou=0.
if (arousal le 13) arou=1.
if (arousal ge 14 and arousal lt 28) arou=2.
if (arousal gt 27) arou=3.
execute.
```

Alternative: Transform → Recode into different variables → old and new values

DATASET ACTIVATE DataSet1.

RECODE arousal (Lowest thru 13=1) (14 thru 27=2) (28 thru Highest=3) INTO arou3.

EXECUTE.

- **Compute and execute belong together**

TEST: Transformation of variables

Data Set: Record2.sav

- Divide the variable "airplay" into the following categories: hardly ever, medium, excessive.
  - hardly ever: up to 10
  - medium: 11-30
  - excessive: more than 30 times per week
- Assign value labels to the new variable

```
compute air=0.
if (airplay le 10) air=1.
if (airplay ge 11 and airplay le 30) air=2.
if (airplay gt 30) air=3.
execute.
```

Import of an excel file into SPSS

- Copy data Excel Sheet.xlsx from the student server to your personal server
- Remember the range of data in Excel (A1:E31)
- Check “Read variable names from the first row of data”
- Enter Range
- Ok
- Save file as .sav file
- Before conducting complex analyses, it is reasonable to get a general idea about the structure and distribution of the data
- Good introduction into the data with frequency tables.
- Analyze→ Descriptive Statistics → Frequencies → Select variable(s); select “display frequency tables”
- Frequency tables, however, make only sense for variables with a manageable amount of scores

Meaninglessness

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	8.00	2	1.7	1.7	1.7
	10.00	6	5.0	5.0	6.6
	11.00	7	5.8	5.8	12.4
	12.00	5	4.1	4.1	16.5
	13.00	12	9.9	9.9	26.4
	13.63	1	.8	.8	27.3
	14.00	9	7.4	7.4	34.7
	14.43	1	.8	.8	35.5
	15.00	7	5.8	5.8	41.3
	16.00	6	5.0	5.0	46.3
	17.00	6	5.0	5.0	51.2
	18.00	8	6.6	6.6	57.9
	19.00	7	5.8	5.8	63.6
	20.00	5	4.1	4.1	67.8
	21.00	8	6.6	6.6	74.4
	22.00	6	5.0	5.0	79.3
	23.00	4	3.3	3.3	82.6
	23.91	1	.8	.8	83.5
	24.00	2	1.7	1.7	85.1
	25.00	6	5.0	5.0	90.1
	26.00	4	3.3	3.3	93.4
	27.00	2	1.7	1.7	95.0
	29.00	2	1.7	1.7	96.7
	30.00	4	3.3	3.3	100.0
	Total	121	100.0	100.0	

- In addition, there are descriptive measures which help getting an idea about the data
- **Slides PART II**

2 Parameter testing

2.1 Tables & Graphs, descriptive statistics

In order to get a first quick overview of your data and the distribution of a single variable, it is advisable to produce a tables of descriptive statistics

2.2 Descriptive measures

2.2.1 Measures of location and percentiles

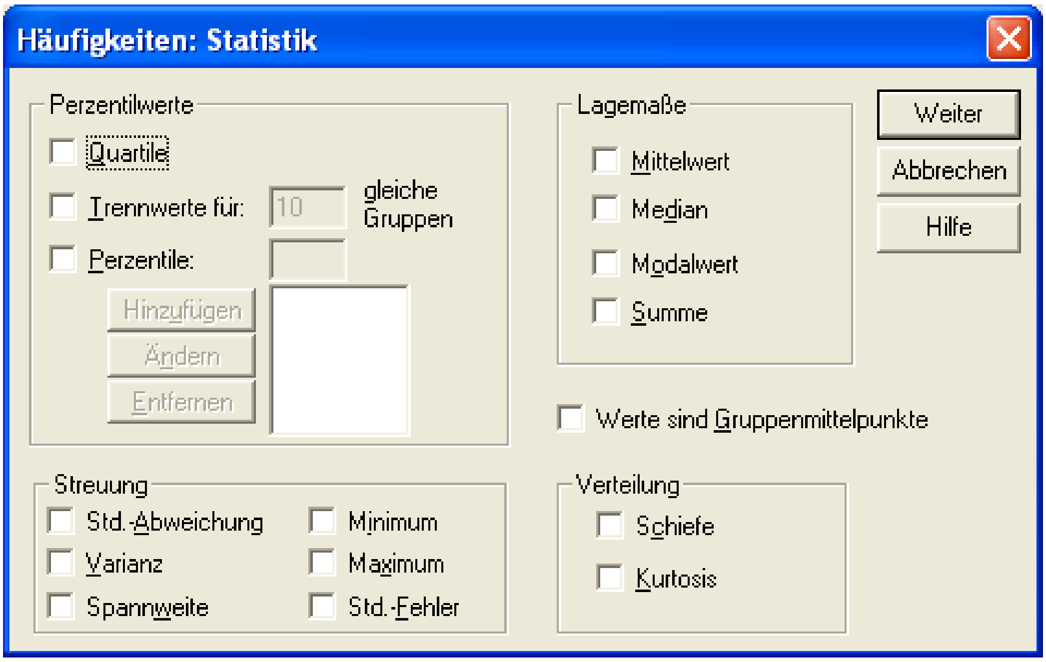
- **Arithmetic mean** (average, average mean): Sum of observations divided by total number of observations (most used)
- **Median:** Value that divides sample into two halves; Separates the higher half of a sample from the lower half, lies in the middle of a series of ordered observations = 50% quantile
- **Mode:** Value that occurs the most frequently, value with the biggest (relative) frequency; There can be more than one or no mode (if every number in the set has the same frequency)
- **Quantiles**:
    - Value which cuts x% of the distributional area. Below the quantile is a certain share of all values in a set of ordered observations
    - If the distribution is divided in *n* equal shares, there are *n* -1 quantiles
    - Quartile = sample is divided into 4 equal shares (= 3 quartiles: 25%-, 50%- 75% quartile)
    - Quintiles = sample is divided into 5 equal shares (= 4 quintiles)
    - Deciles = sample is divided into 10 equal shares (=9 deciles)
    - Percentiles = sample is divided into 100 equal shares

2.2.2 Measures of distribution

- **Variance:** main distribution measures; mean square deviation of single values from the mean
- **Standard deviation (σ, S):** positive square root of its variance
- **Variation coefficient:** relative standard deviation; ratio of the standard deviation to the mean,
- **Range:** most simple distribution measure, difference between the highest and the lowest values in the set
- **Interquartile range:** difference between the 75%-and 25% quartile

2.2.3 Measures of shape

- Deviation from normal distribution
- Skewness: Asymmetric = skewed
  - Right-skewed = positive skewed ◊ main part of distribution on the left side
  - Left-skewed= negative skewed◊ main part of distribution on the right side
- Kurtosis
  - Steepness of a distribution
  - Positive excess (sharp): maximum higher compared to normal distribution (with the same variance)
  - Negative excess (flat): maximum lower compared to normal distribution (with the same variance)
- SPSS positive value: left-skewed
- SPSS negative value: right-skewed
- SPSS = 0 : symmetric distribution
- EXERCISE: Generate descriptive statistics (quartiles, mean, standard error of the mean, median, mode, standard deviation, variance, skewness, kurtosis, minimum, maximum, sum) for meaninglessness.
- Analyze→ Descriptive Statistics → Frequencies → Select variable(s); under statistics select the statistics you want to produce: central tendency, distribution, dispersion and the 25%, 50%, 75% quartile.
- The list of variables shows all numerical and string varaibles
- The select the variable of interest and click on the arrow () to remove variable to Variable(s)
- To select certain measures, click on Statistic… which opens the following new window: frequencies: statistic



- Select the statistics of interest
- Consider whether you want to selcet: display frequency tables (!) → is not reasonable in this categroy, only reasonable for variables with manageable amount of scores (think of income)
- The default is: no statistics are calculated

Statistics

Meaninglessness

N	Valid	121
	Missing	0
Mean		17.7932
Std. Error of Mean		.49811
Median		17.0000
Mode		13.00
Std. Deviation		5.47919
Variance		30.021
Skewness		.420
Std. Error of Skewness		.220
Kurtosis		-.652
Std. Error of Kurtosis		.437
Range		22.00
Minimum		8.00
Maximum		30.00
Sum		2152.97
Percentiles	25	13.0000
	50	17.0000
	75	22.0000

!!Every output table needs an interpretation, also in the exams, although it is not explicitly requested!!

Table interpretation:

- **Mean:** Average meaninglessness in the sample
- **Median:** Value that lies in the middle of an ordered range of variables; Half of the variable is greater or smaller
- **Mode:** Meaninglessness value that occurs most often
- **Percentiles:**
  - 25%: value for meaninglessness below which there are exactly 25% of all values of the distribution, 75% of all scores lie above this value

        - 50%: meaninglessness score below which there are exactly 50% of all meaninglessness scores in the sample, 50% of all scores lie above this value = median
        - 75%: meaninglessness score below which there are 75% of meaninglessness scores in the sample, 25% lie above this score
    - **Variance and standard deviation:** Variance is not given in the same unit than the variable, therefore SD is more suitable
    - **Skewness:**
        - Pile-up left
    - **Kurtosis**
        - Rather flat (negative value)
        -

TEST: Descriptive Statistics

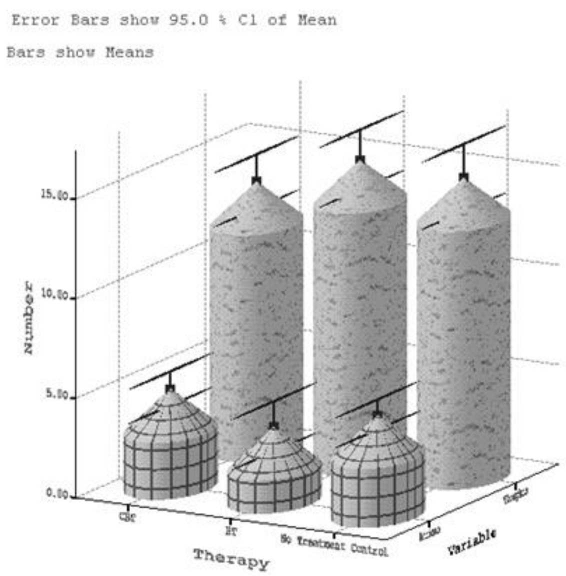
- **Data Set: Cosmetic Surgery.sav**
- **Generate quartiles for relatively happy and unhappy people (quality of life) after surgery**

Statistics

Quality of Life After Cosmetic Surgery

N	Valid	276
	Missing	0
Percentiles	25	52.10
	50	58.00
	75	67.00

- 25% Quartil: quality of life after surgery score, below which there are 25% of all scores, 75% lie above this value (relatively unhappy people after surgery)
- 50% Quartil: Median; quality of life after surgery score, below which there are 50% of all scores, 50% lie above it
- 75% Quartil: quality of life score, under which there are 75% of all scpres, 25% lie above (relatively happy people after surgery)
- Relatively unhappy people: score below 52.1; relatively happy after surgery: score above 67



2.3 Graphs

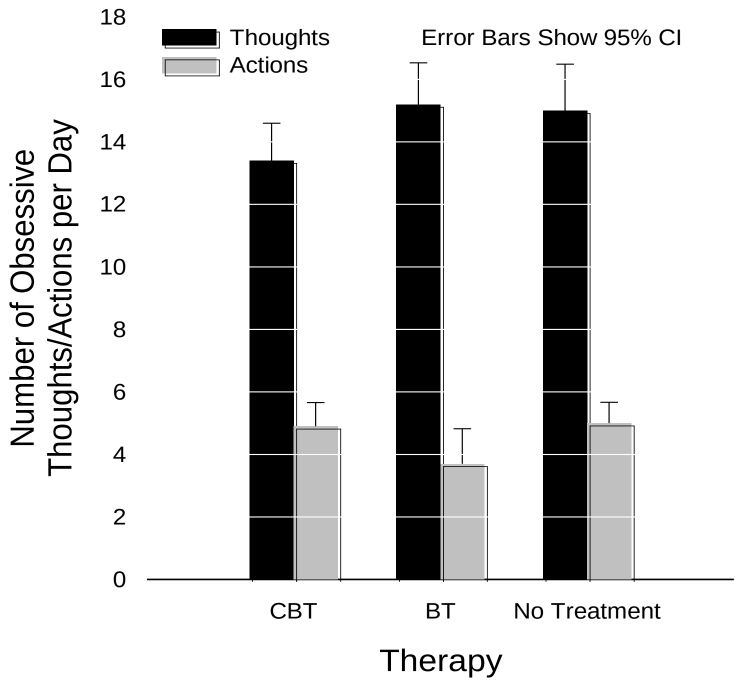
2.3.1 The art of presenting data

- Graphs should (Tufte, 2001):
  show the data.
- Induce the reader to think about the data being presented (rather than some other aspect of the graph).
- Avoid distorting the data.
- Present many numbers with minimum ink.
- Make large data sets (assuming you have one) coherent.
- Encourage the reader to compare different pieces of data.
- Reveal data.

Why is this graph bad?

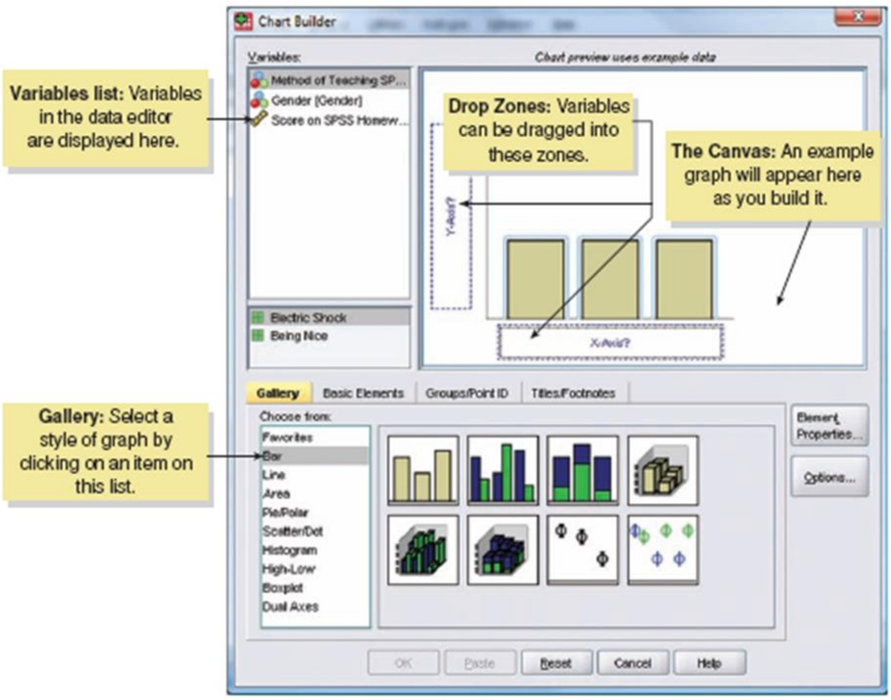
- The bars have a 3D effect, almost impossible to read
- Bars have patterns: distract the eye from the data
- Cylindrical bars: difficult to understand
- Badly labelled y-axis: number of what

Why is this graph better?



- 2-d okit: easier to compare values across therapies
- Y-axis has a more informative label
- Fewer distractions
- Minimum ink: no axis lines, no grid

2.3.2 The Chart Builder



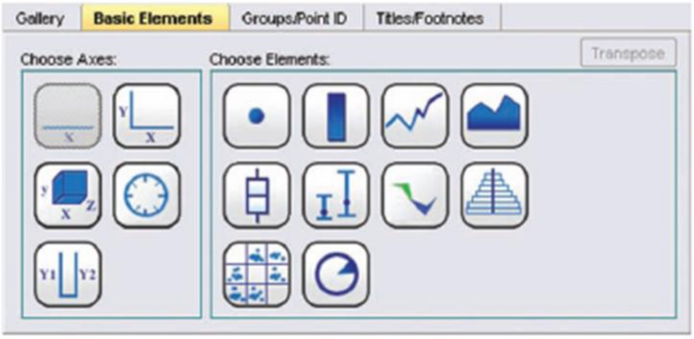
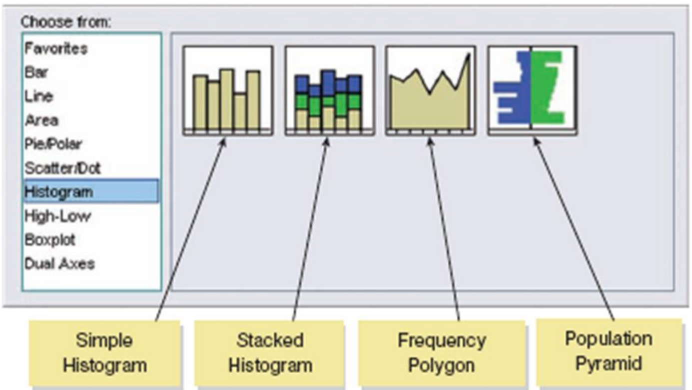


FIGURE 4.6
Building a graph from basic elements

2.3.3 Histograms: spotting obvious mistakes

FIGURE 4.7
The histogram gallery



- A histogram is a graph in which values of observations are plotted on the horizontal axis, and the frequency with which each value occurs in the data set is plotted on the vertical axis.
- Histograms plot:
    - The score (x-axis)
    - The frequency (y-axis)
- Histograms help us to identify:
    - The shape of the distribution
        - Skew
        - Kurtosis
        - Spread or variation in scores
    - Unusual scores

Histogram example

- A biologist was worried about the potential health effects of music festivals.
- Measured the hygiene of 810 concert-goers over the three days of the festival.
- Hygiene was measured using a standardised technique:
    - Score ranged from 0 to 4
        - 0 = you smell like a corpse rotting up a skunk’s arse

- 4 = you smell of sweet roses on a fresh spring day

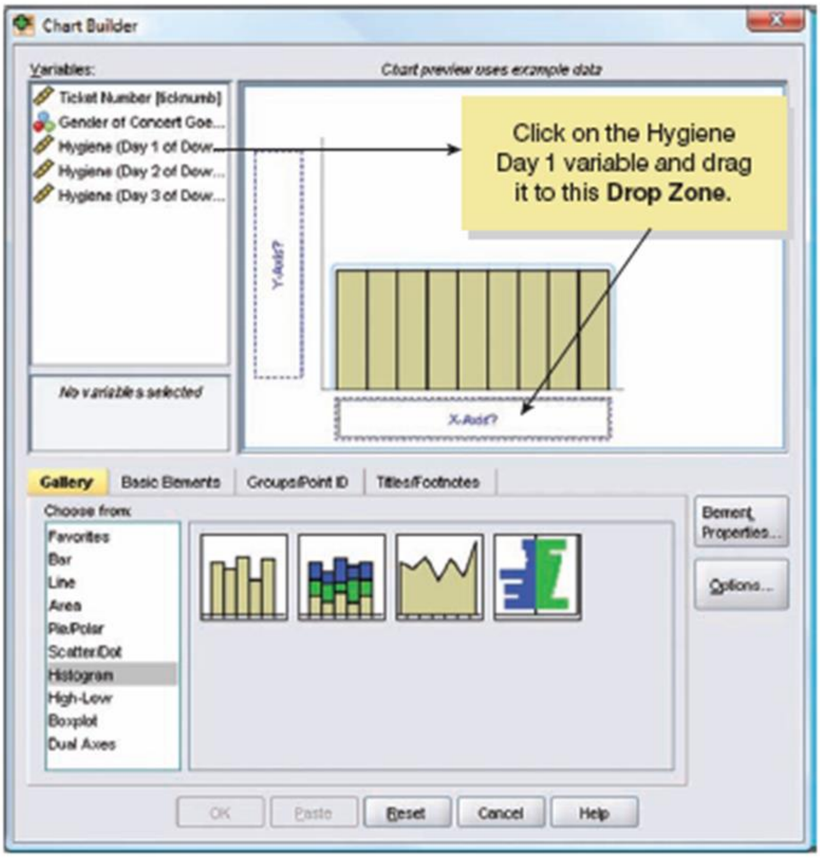


FIGURE 4.8
Defining a histogram in the Chart Builder

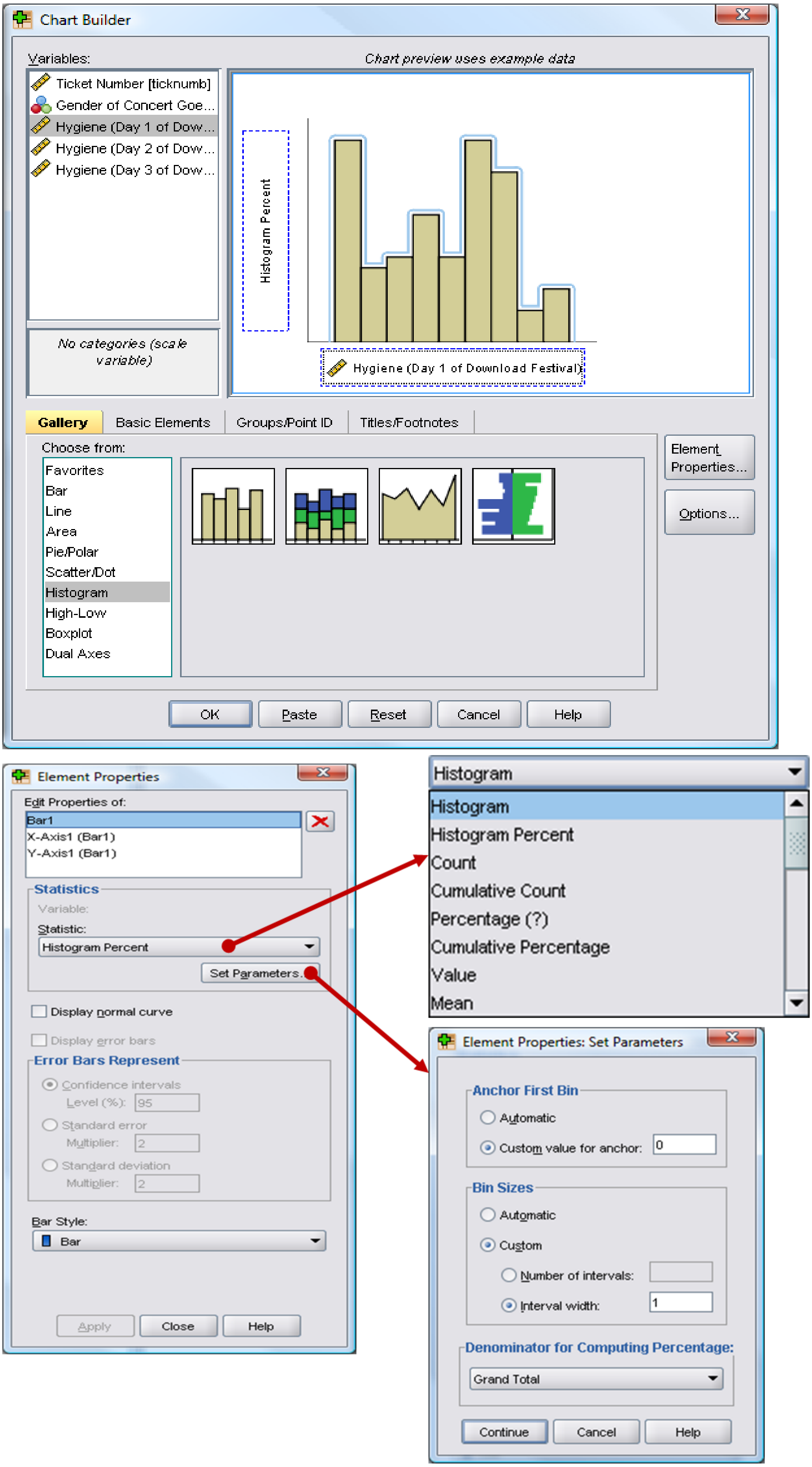
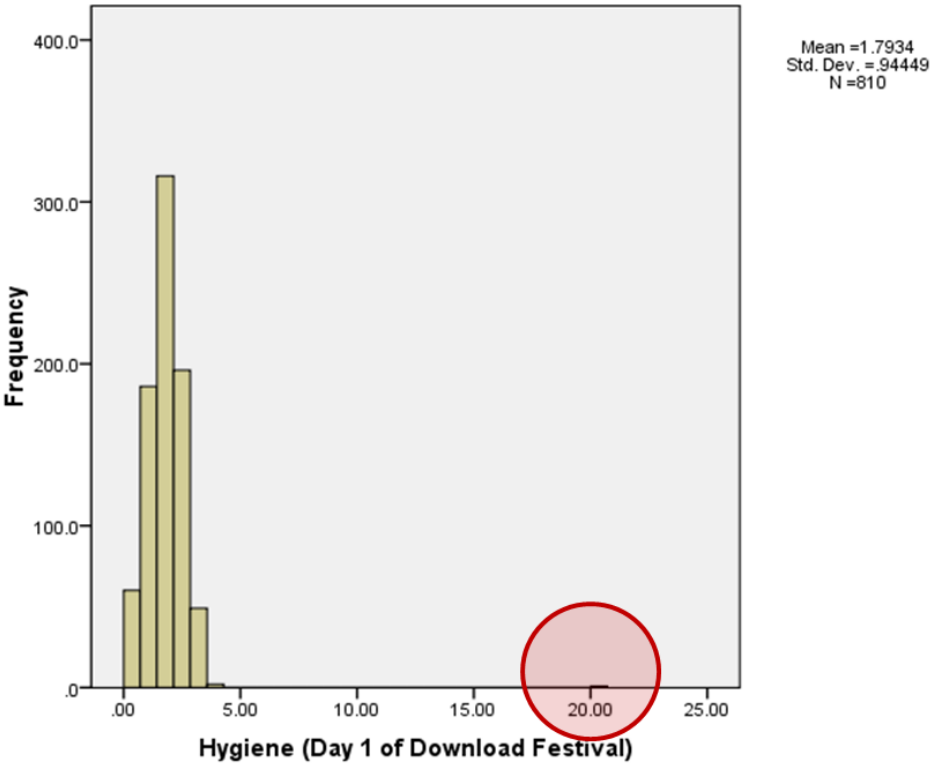


Chart Builder
Variables:
Chart preview uses example data
Ticket Number [ticknumb]
Gender of Concert Goe...
Hygiene (Day 1 of Dow...
Hygiene (Day 2 of Dow...
Hygiene (Day 3 of Dow...
No categories (scale variable)
Histogram Percent
Hygiene (Day 1 of Download Festival)
Gallery
Basic Elements
Groups/Point ID
Titles/Footnotes
Choose from:
Favorites
Bar
Line
Area
Pie/Polar
Scatter/Dot
Histogram
High-Low
Boxplot
Dual Axes
Element Properties...
Options...
OK
Paste
Reset
Cancel
Help
Element Properties
Edit Properties of:
Bar1
X-Axis1 (Bar1)
Y-Axis1 (Bar1)
Statistics
Variable:
Statistic:
Histogram Percent
Set Parameters...
Display normal curve
Display error bars
Error Bars Represent
Confidence intervals
Level (%): 95
Standard error
Multiplier: 2
Standard deviation
Multiplier: 2
Bar Style:
Bar
Apply
Close
Help
Histogram
Histogram
Histogram Percent
Count
Cumulative Count
Percentage (?)
Cumulative Percentage
Value
Mean
Element Properties: Set Parameters
Anchor First Bin
Automatic
Custom value for anchor: 0
Bin Sizes
Automatic
Custom
Number of intervals:
Interval width: 1
Denominator for Computing Percentage:
Grand Total
Continue
Cancel
Help

2.3.4 The resulting histogram



- The first thing that should leap out at you is that there appears to one case that is very different to the others
- All of the scores appear to be squashed up to one end of the distribution because they are all less than 5, except for one which has a value of 20
- This is an outlier: a score different to the rest
    - Outliers bias the mean and inflate the standard deviation
- The outlier in this diagram is particularly odd because it has a score of 20, which is above the top of our scale (hygiene scale ranges from 0 to 4), and so it must be a mistake
- How can we detect the outlier? We can go through the 851 concert goers, or do a boxplot

2.3.5 Boxplots

Boxplots are made up of a box and two whiskers.

- The box shows:
    - The median
    - The upper and lower quartile
    - The limits within which the middle 50% of scores lie.
- The whiskers show
    - The range of scores
    - The limits within which the top and bottom 25% of scores lie

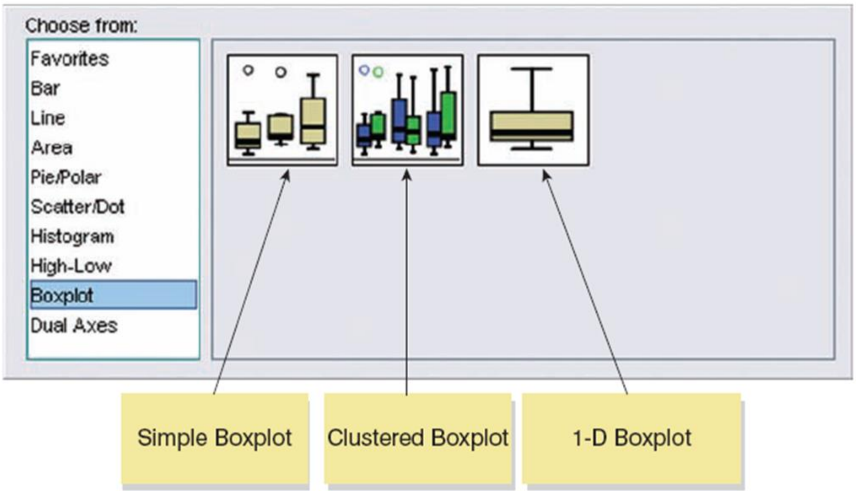
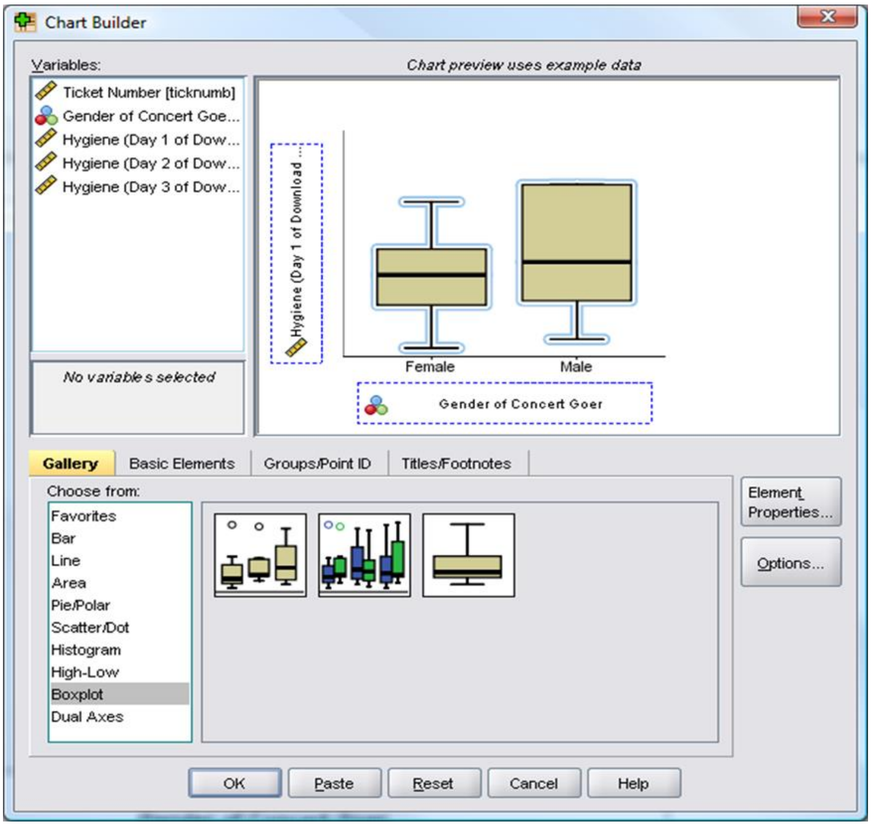
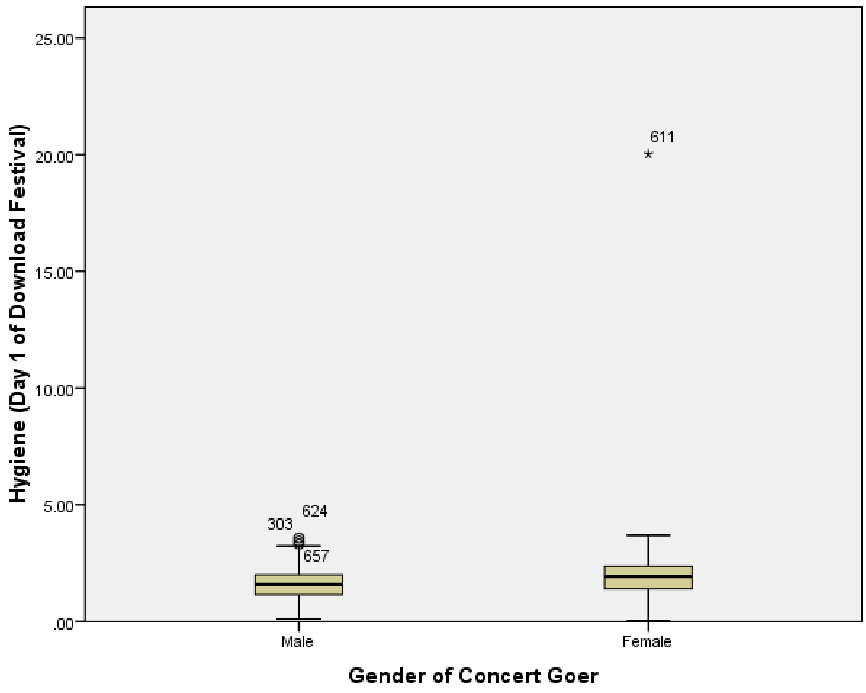


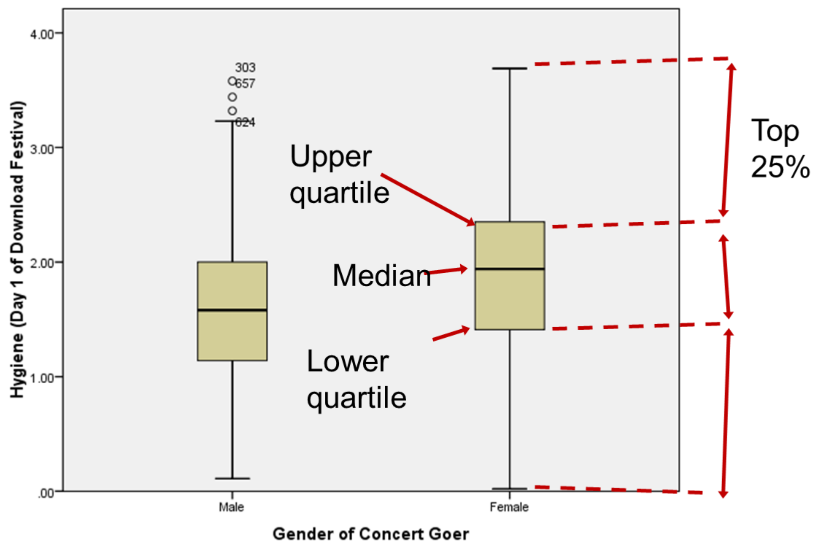
FIGURE 4.11
The boxplot gallery





The outlier that we detected in the histogram is shown up as an asterisk (*) on the boxplot and next to it is the number of the case (611) that causes this outlier

What does the boxplot show?



This is the boxplots for the hygiene scores on day 1 after the outlier has been corrected

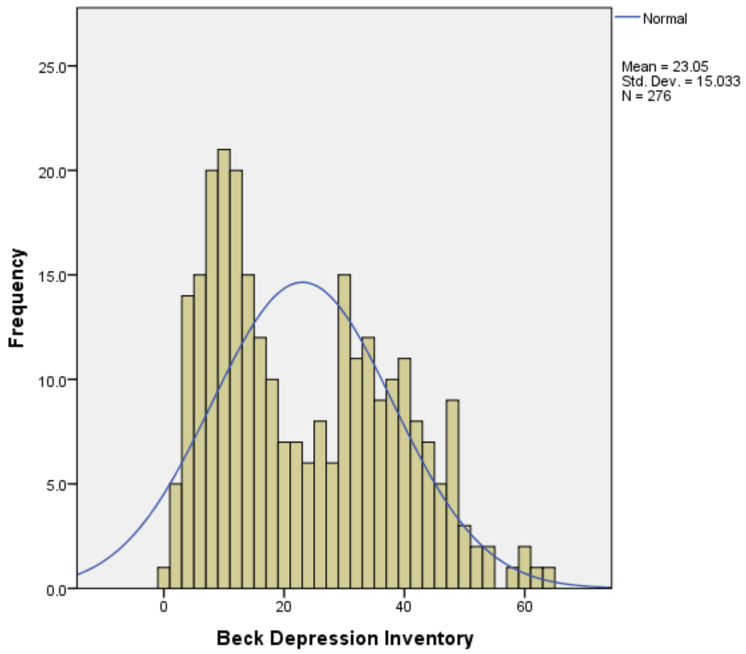
What does the boxplot represent?

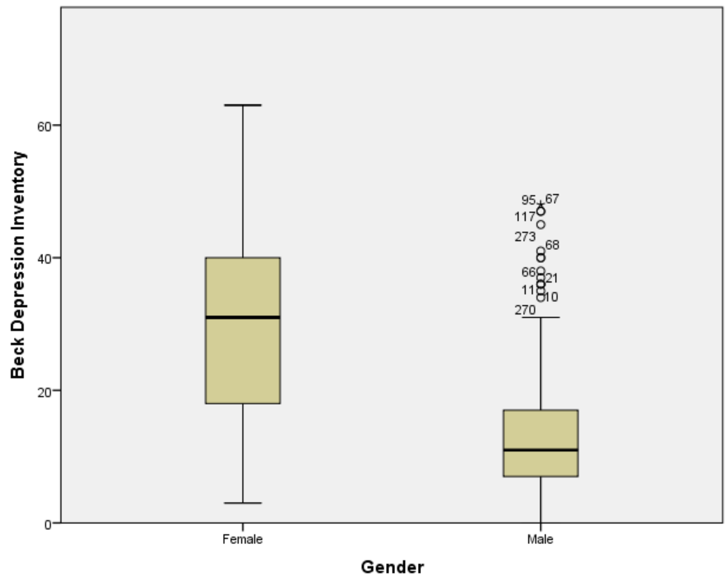
- It shows us the lowest score (the bottom horizontal line on each plot) and the highest (the top horizontal line of each plot)
- Comparing the males and the females we can see they both had similar low scores (0, or very smelly) but the women had a slightly higher top score (i.e. the most fragant female was more hygienic than the cleanest male)
- The lowest edge of the tinted box is the lower quartile
- Therefore, the distance between the lowest horizontal line and the lowest edge of the tinted box is the range between which the lowest 25% of scores fall
- This box is slightly larger for women than for men, i.e. there is more variability in the hygienic score of the bottom 25%
- The box shows the interquartile range, i.e. 50% of the scores are bigger than the lowest part of the tinted area but smaller than the top part of the tinted area
- The top edge of the tinted box shows the upper quartile
- In the middle of the box is a slightly thicker horizontal line
- It represents the median
- The median for females is higher than for males, which tells us that the middle female scored higher than the middle male

TEST: Graphs

Data Set: Cosmetic Surgery.sav

- Produce a histogram for the Beck Depression Inventory (BDI). The histogram should also display a normal distribution. Does the BDI variable follow a normal distribution?
- Produce a simple boxplot for the BDI separated by gender. Can you detect any differences?



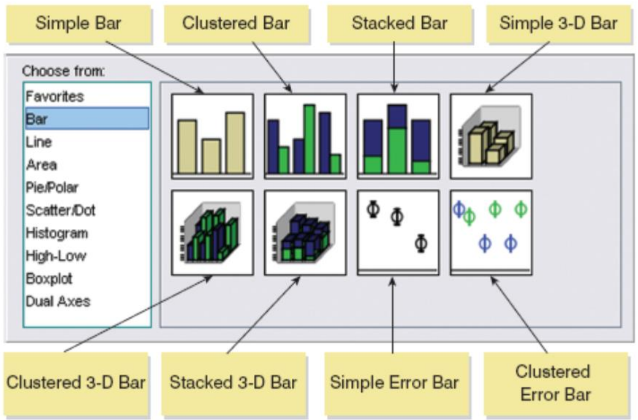


Select BDI and drag it into y-axis?, drag gender on x-axis?

2.3.6 Error bar charts

- Bar charts are the typical way for people to display means
- Chart Builder is the starting point
- The bar (usually) shows the mean score
- The error bar sticks out from the bar like a whisker.
- The error bar displays the precision of the mean in one of three ways:
    - The confidence interval (usually 95%)
    - The standard deviation
    - The standard error of the mean
- (you can select under element properties), error bars represent…

FIGURE 4.15
The bar chart gallery



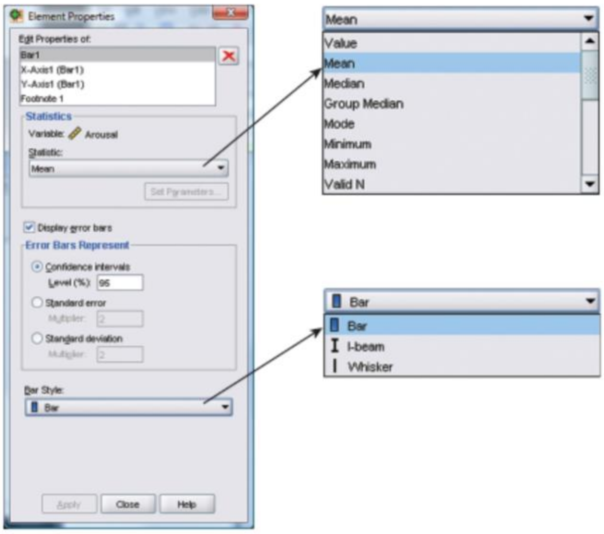
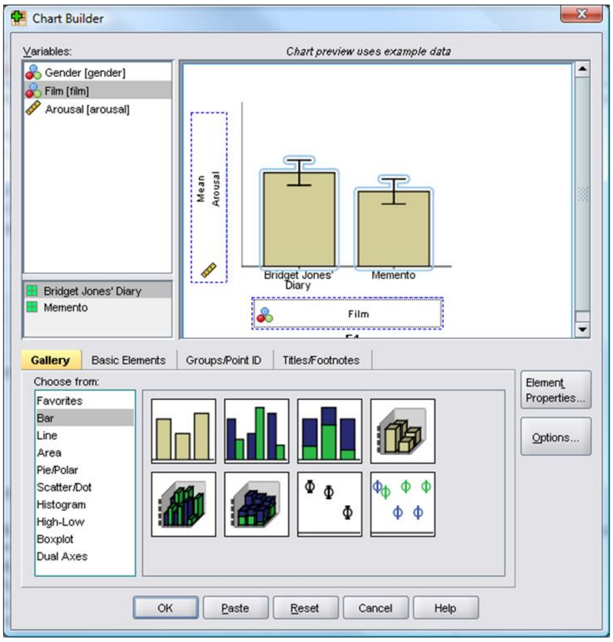
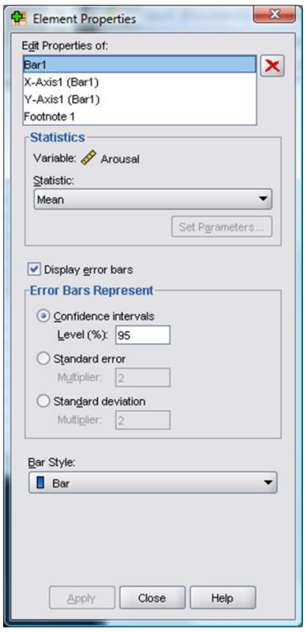


FIGURE 4.16
Element Properties of a bar chart

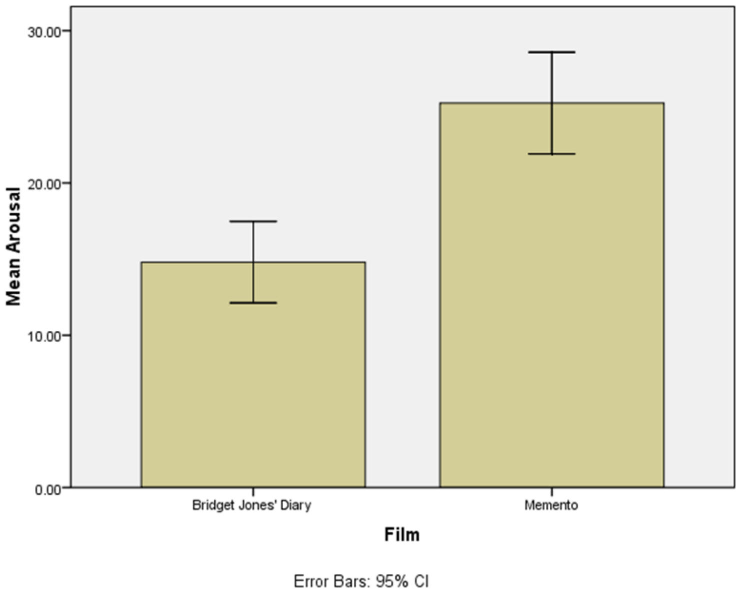
Bar Chart: One Independent Variable

- Bar charts are the usual way for people to display means
- How you create these graphs in SPSS depends on how you collected your data, whether the means come from independent cases and are, therefore, independent, or came from the same cases and so are related
- Other possibilities to create bar charts
  - Clustered bar charts for independent means
  - Simple bar chart for related means
  - Clustered bar chart for related means
  - Clustered bar charts for mixed designs



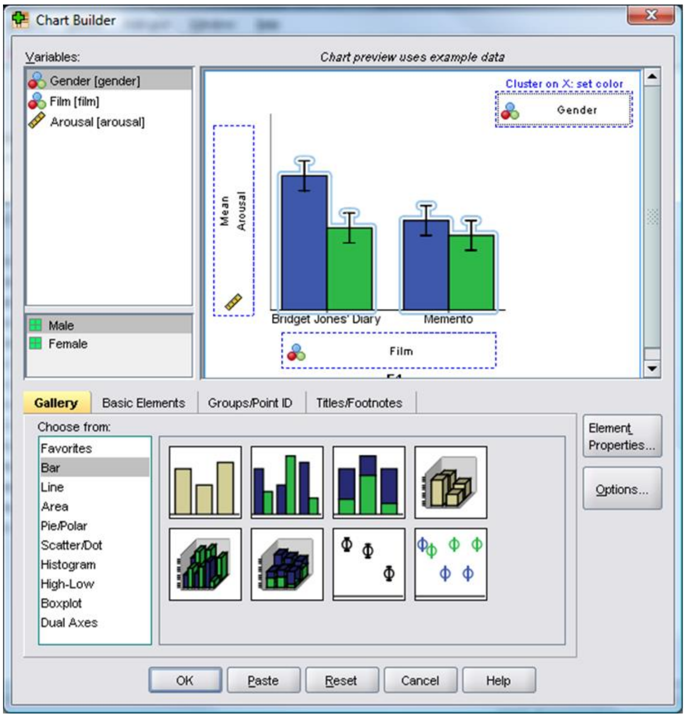


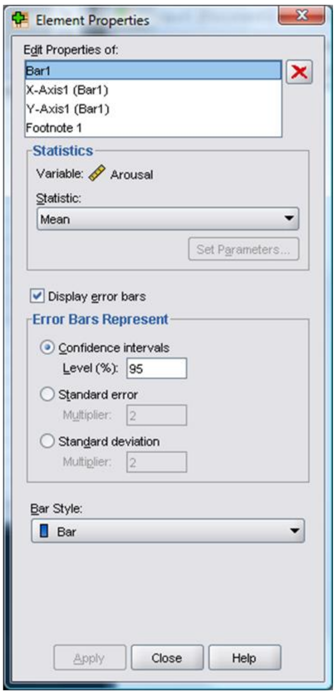
- One independent variable
  - The y-axis needs to be dependent variable, or the thing you've measured, or more simply the thing for which you want to display the mean
  - The x-axis should be the variable by which we want to split the arousal data
- Finally you can ask SPSS to add error bars to your bar chart to create an error bar chart by selecting display error bars (under element properties)
- Error bars from SPSS are suitable only for normally distributed data



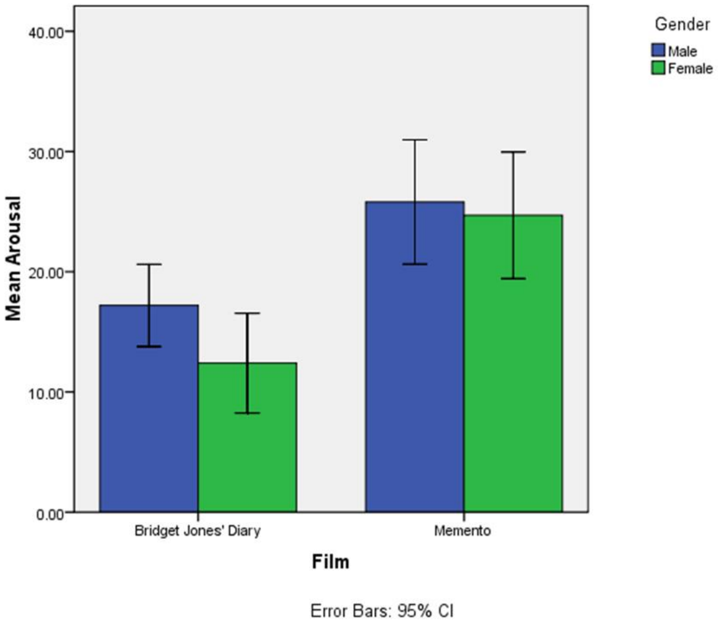
This graph displays the mean (and the confidence interval of those means) and shows us that on average, people were more aroused by Memento than they were by Bridget Jones Dairy

If we are interested in gender effects: clustered bar charts for independent means





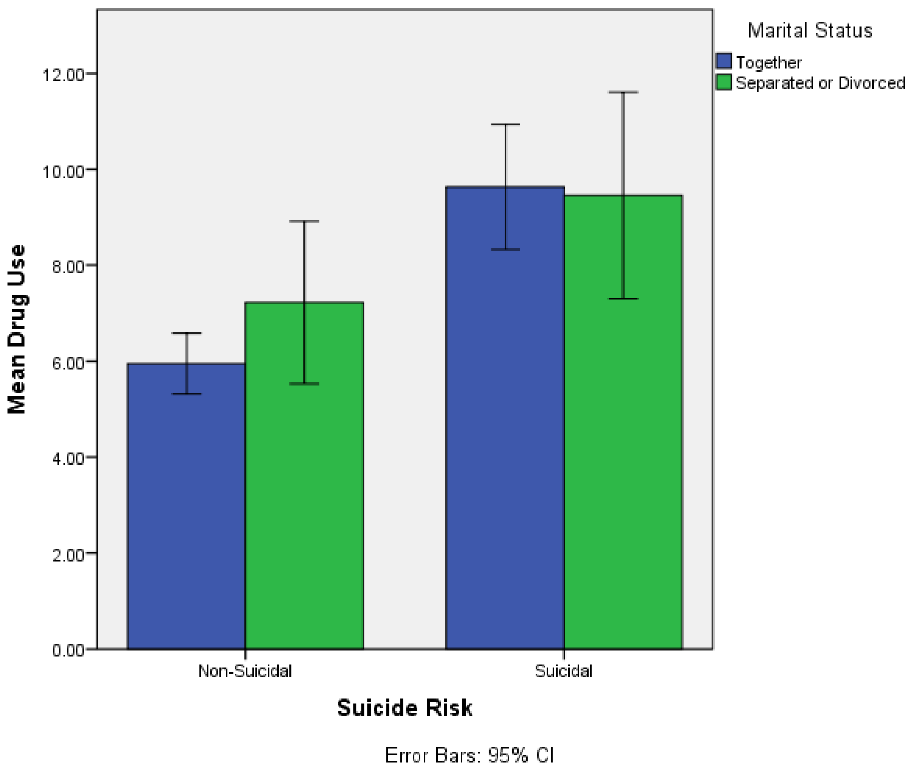
- There is now an extra drop zone: cluster on X:
- Drag our second grouping variable
    - Y-axis: arousal
    - X-axis: film
    - Cluster on: gender



- Like the simple bar chart this graph tells us that arousal was overall higher for Memento than Bridget Jones, but it also splits this information by gender
- The mean arousal for Bridget Jones shows that males were actually more aroused during this film than females
- This indicates they enjoyed the film more than the women did
- For memento, values are more comparable
- But this contradicts the idea of a chick flick
    - It actually seems that men enjoy chick flicks more than the chicks

TEST: Graph

- Data Set: Heavy Metal.sav
- Produce and error bar char for mean drug use for each of the different suicide risks splitted by martial status.



- Mean drug use (amounts of joints smoked per week) were overall higher for suicidal individuals compared to non-suicidal individuals
- For non-suicidal individuals, those living together had a lower drug use than those living separated or divorced
- For suicidal people however, mean drug use is slightly higher for those living together than for those individuals that are separated or divorced

3 Distribution (non-parametric) testing, t-Test

SPSS has different procedures for conducting a t-Test. It depends on the question which option should be used

Dependent *t*-test

Compares two means based on related data.

E.g., Data from the same people measured at different times.

Data from ‘matched’ samples.

Independent *t*-test

Compares two means based on independent data

E.g., data from different groups of people

Significance testing

Testing the significance of Pearson's correlation coefficient

Testing the significance of *b* in regression.

Dependent t-test specifics

Data arrangement

Same individuals, reactions to different conditions (in different columns)

We will not practice the dependent t-test in class, have a look at Andy Field for an example

Assumptions of the t-test

Both the independent *t*-test and the dependent *t*-test are *parametric tests* based on the normal distribution.

Can e.g. be tested with a histogram displaying a nornmal curve

Data are measured at least at the interval level.

The independent *t*-test, because it is used to test different groups of people, also assumes:

Variances in these populations are roughly equal (*homogeneity of variance*).

Will be checked with Levene's test within the SPSS procedure

Scores in different treatment conditions are independent (because they come from different people).

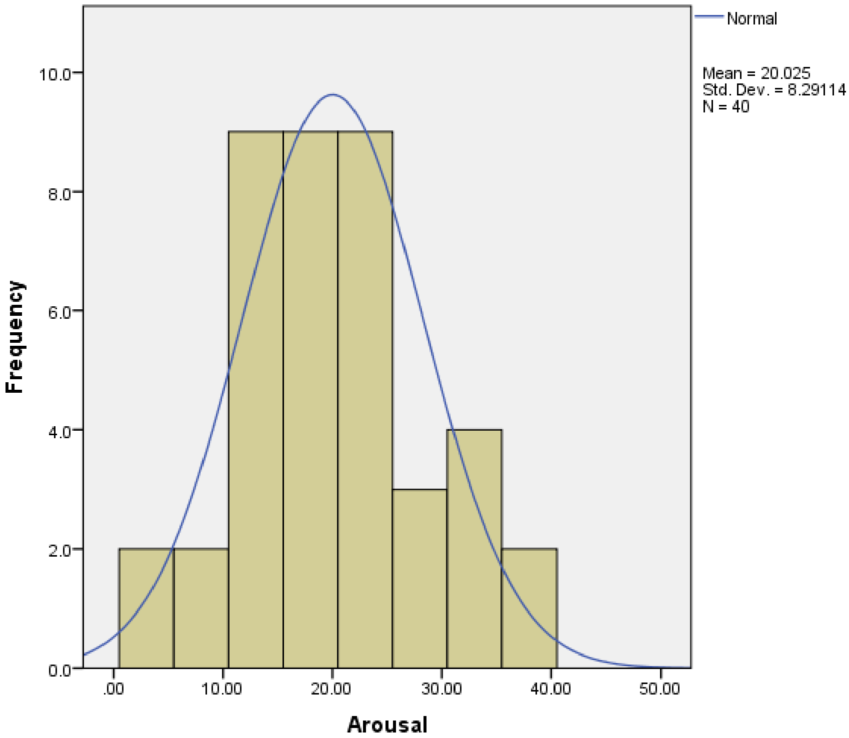
EXERCISE: Independent *t*-test

Data Set: 01_ChickFlick.sav

Does the average physical arousal during the movie differ between men and women? How big is the effect size? Also check the *t*-test assumptions.

Test for normal distribution

Histogram for arousal, arousal on x-axis, select display normal curve, apply, ok



Arousal is approximately normally distributed

Analyze → Compare means → Independent samples t-test → select test variable (arousal) → select grouping variable (gender) → define groups (1:1 (male); 2:2 (female) → ok

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Arousal	Male	20	21.5000	7.42329	1.65990
	Female	20	18.5500	9.02322	2.01765

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Arousal	Equal variances assumed	.868	.357	1.129	38	.266	2.95000	2.61270	-2.33913	8.23913
	Equal variances not assumed			1.129	36.639	.266	2.95000	2.61270	-2.34559	8.24559

3.1 Testing homogeneity of variance:

Hypotheses Levene test:

- Tests whether the variance are equal in the population
- **H0 Levene Test: variances are equal**
- **H1 Levene Test: variance are not equal**
- The maximal acceptable level of probability of error (level of significance) is 5% (alpha=0,05).

- p > 0,05: The H0 ‚Variances are equal' cannot be rejected, we can assume that variance are sufficiently equal to assume variance homogeneity. We interprete the first line in the output table (line: ‚variances are equal'
- p < 0,05: The H0 ‚variances are equal" has to be rejected. In the ouput table we interprete the line showing ‚variances are not equal"

REPORTING Levene's test*: For these data, the Levene's test is nonsignificant (because p=.357) which is greater than .05 and so we say that the assumption of homogeneity of variances is met and we work with the upper line of the t-test.*

here: p>.05, i.e. the H0 is accepted and we look at the upper line

REPORTING*: The descriptive table shows that means are different in our sample. Mean arousal for males is 21.5 and for females it is 18.55. The mean difference is 2.95. In order to test if the difference is significant, the t-test is conducted*

Definition of hypotheses

- **H0:** Physical arousal during the movie does not differ between men and women
- **H1:** Physical arousal during the movie differs between men and women
- If both means have the same size in the population, the t-value is 0
- We get a value of 1.13
- If there is no difference between the means in the population, the observed t-value can result with a probability of X
- The significance level is 26.6%, this means that when we reject the null hypothesis we make a mistake of 26.6%
- We cannot reject the null hypothesis, there is not significant difference in the mean arousal

REPORTING: On average, men experienced greater physical arousal during the film (M = 21.50, SE = 7.42), than do females (M = 18.55, SE = 9.02). This difference was not significant t(38) = 1.13, p > .05; however, it did represent a small effect r = .18.

Calculation of r: square root of (t^2/(t^2+df))

$$r = \sqrt{\frac{t^2}{t^2 + df}} = 0.18$$

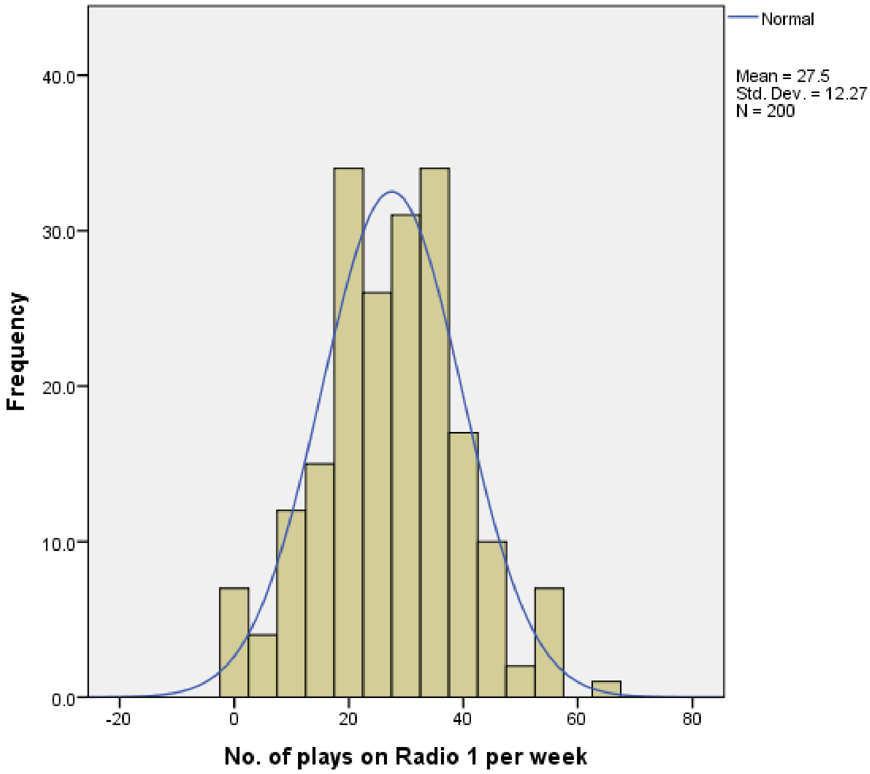
TEST: t-test

Data Set: Record.sav

Does the number of times a record is played on the radio differ dependeing on whether the band is attractive or unattractive? Also check the *t*-test assumptions.

Generate a binary variable separating attractive from unattractive bands (attractive band: attractiveness le 5; unattractive band: attractiveness gt 5)

```
compute attr=0.
if (attract le 5) attr=1.
if (attract gt 5) attr=2.
execute.
```



The variable is approximately normally distributed (the t-test is robust enough)

H0: The number of times a record is played on the radio does not differ between bands that are attractive and those that are not

H1: The number of times a record is played on the radio differs between bands that are attractive and those that are not

Group Statistics

	attr	N	Mean	Std. Deviation	Std. Error Mean
No. of plays on Radio 1 per week	1.00	26	23.50	10.782	2.115
	2.00	174	28.10	12.394	.940

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
No. of plays on Radio 1 per week	Equal variances assumed	.593	.442	-1.792	198	.075	-4.598	2.566	-9.657	.462
	Equal variances not assumed			-1.987	35.645	.055	-4.598	2.314	-9.292	.097

Testeing homogeneity of variance:

Hypotheses Levene test:

- Tests whether the variance are equal in the population
- H0 Levene Test: variances are equal
- H1 Levene Test: variance are not equal
- p > 0,05: The H0 ‚Variances are equal' cannot be rejected, we can assume that variance are sufficiently equal to assume variance homogeneity. We interprete the first line in the output table (line: ‚variances are equal'
- p < 0,05: The H0 ‚variances are equal" has to be rejected. In the ouput table we interprete the line showing ‚variances are not equal"
- here: p>.05, i.e. the H0 is accepted and we look at the upper line

REPORTING Levene's test: For these data, the Levene's test is nonsignificant (because p=.442) which is greater than .05 and so we say that the assumption of homogeneity of variances is met and we work with the upper line of the t-test.

REPORTING: The descriptive table shows that means are different in our sample. Records of attractive bands are played more often (28.10 times per week) compared to those of unattractive bands with 23.50 times per week. The mean difference is -4.6. In order to test if the difference is significant, the t-test is conducted

- If this means that there is a significant difference in the population, the t-test is conducted
- If both means have the same size in the population, the t-value is 0
- We get a value of -1.79
- If there is no difference between the means in the population, the observed t-value can result with a probability of X
- The significance level is 7.5%, this means that when we reject the null hypothesis we make a mistake of 7.5%
- We therefore are not safe to reject the null hypothesis, there is no significant difference in the number of times a record is played on the radio depending on the attractiveness of a band

REPORTING: On average, unattractive bands have a lower airplay (M = 23.50, SE = 10.78), than attractive bands (M = 28.10, SE = 12.39). This difference was not significant t(198) = -1.79, p > .05. It did represent a small effect r = .13.

4 Correlation analysis

Theses: Relationship between the value of two variables

- **H0**= there is **no linear relationship** between the variables
- **H1**= there is **a linear relationship** between the variables

Measure of the strength and direction of this relationship is the **correlation coefficient**

- Identifies linear relationships
- Expresses the strength of the relationship with a value between -1 and +1
- Positive value: positive linear relationship
- Negative value: negative linear relationship
- Correlation coefficient becomes higher, the stronger the linear relationship between the variables is
- Correlation= 0 : no linear relationship

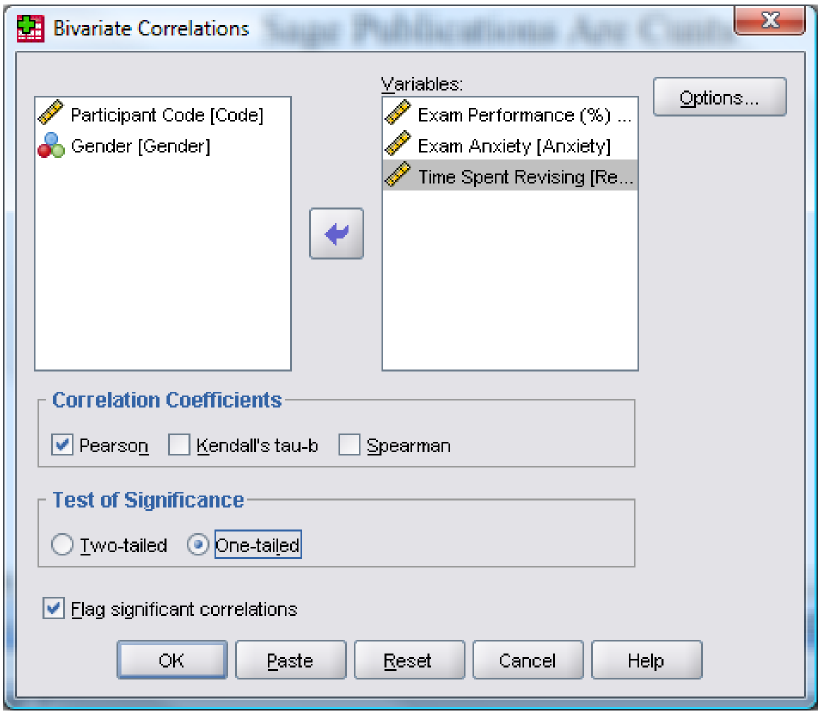
Orientation guidelines (absolute values)

Absolute value of the correlation coefficient	**Possible interpretation**
0	No correlation
> 0 bis 0,2	Very weak correlation
0,2 bis 0,4	Weak correlation
0,4 bis 0,6	Medium correlation
0,6 bis 0,8	Strong correlation
0,8 bis < 1	Very strong correlation
1	Perfect correlation

As the correlation coefficient is restricted to detect linear relationships, also with a low correlation coefficient we could have a perfect non-linear correlation

- To detect these types of effects, it is a good idea to plot the variables against each other to detect other relationship forms: use a scatter plot: this gives a first visual impression about a possible relationship
- Correlation analysis is suitable for intervall scaled variables

4.1 Doing a correlation analysis in SPSS: Analyze → Correlate

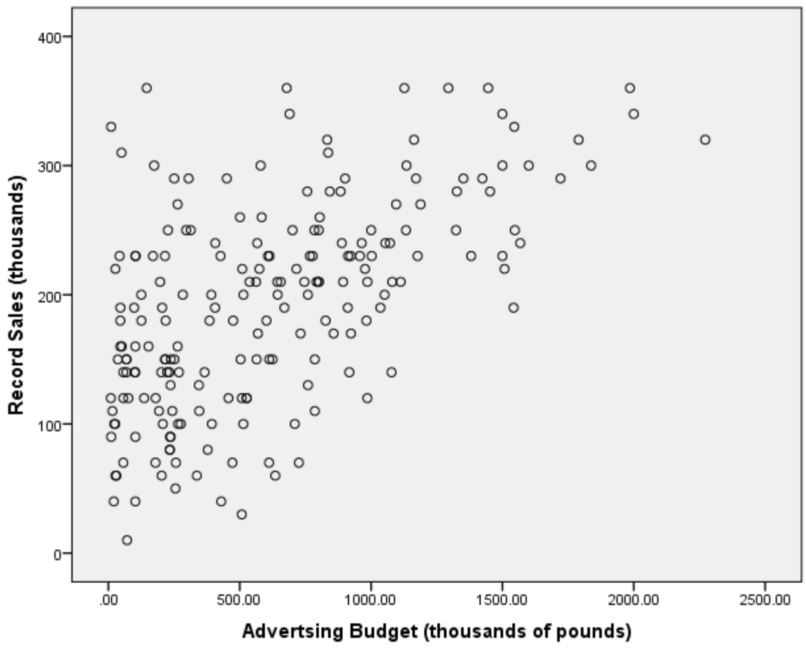


EXERCISE:

Data Set: 02_Record.sav

Is there a linear relationship between amount spent on advertising and record sales (number of copies sold). Produce a scatterplot for the two variables.

Graphs → Legacy dialog → scatter → simple scatter → select variables for X- and Y- axis, ok



We can see that there is a tendency towards a positive relationship, but we cannot see how strong this relationship is

Analyze → Correlate → bivariate, select advertising budget and record sales; keep default Pearson's correlation coefficient

- As a default/ or standard setting, the Pearson's correlation coefficient is calculated (variables need to have a ratio scale level)
- Spearman and Kendall-Tau-B for ordinally scaled variables (variable values are translated into ranks)→ **rank correlation for ordinally scaled data**
- Select variables→ ok

Correlations

		Advertsing Budget (thousands of pounds)	Record Sales (thousands)
Advertsing Budget (thousands of pounds)	Pearson Correlation	1	.578**
	Sig. (2-tailed)		.000
	N	200	200
Record Sales (thousands)	Pearson Correlation	.578**	1
	Sig. (2-tailed)	.000	
	N	200	200

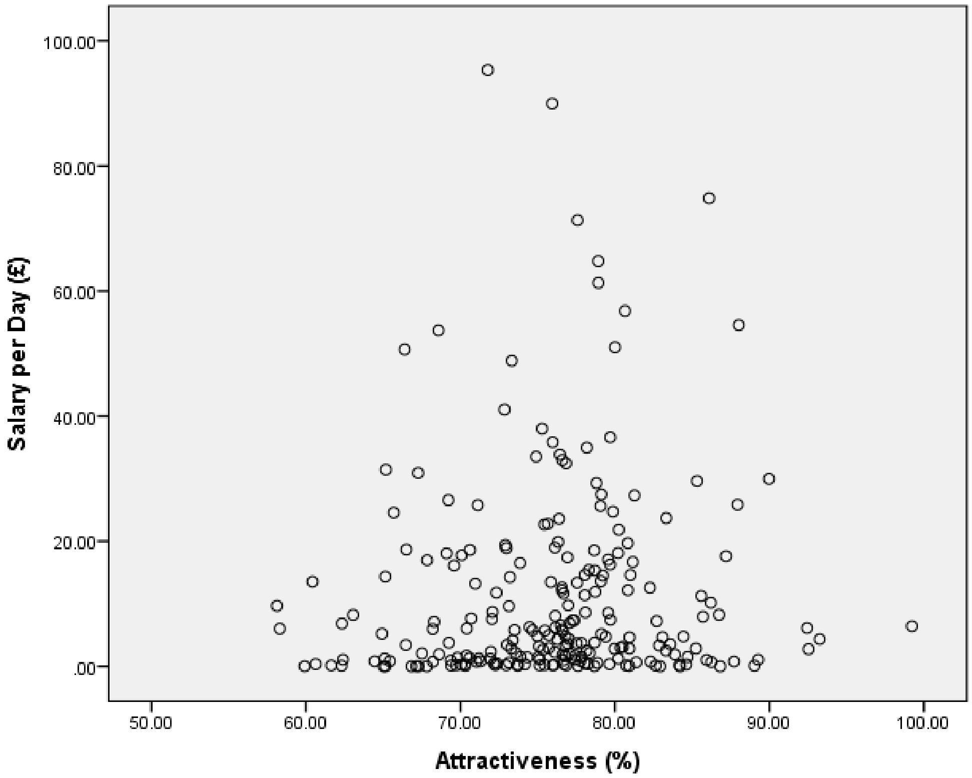
**. Correlation is significant at the 0.01 level (2-tailed).

- The null hypothesis states: There is no linear relationship between the variables.
- The alternative hypothesis is: There is a linear relationship between the variables
- The maximal acceptable level of significance is 5% (alpha=0,05).

- ***The advertising budget was significanlty correlated with record sales, r=.578, p<.001. The value for r represents a positive and large effect.***
- Null hypothesis is rejected
- The significance test examines whether there is a relationship between the variables, but does not say something about the strength of the relationship!! The strength is given with the correlation coefficient
- ***The significance test is the probability at which the correlation coefficient showing in the sample with the given sample size also shows with the same size in the population if there is in reality no relationship in the population***
- As it was unkown at the beginning of the analysis whether there is a positive or negative relationship between the variables, a 2-tailed test has been conducted (independent if there is a negative or positive relationship)

Data Set: 07_Supermodel.sav

- Is there a linear relationship between the beauty of a supermodel (attractiveness in %) and the salary per day? Produce a scatterplot for the two variables.



- Scatter plot is not very useful, seems to be a positive relation, but not very clear
- Based on the scatter plot we cannot determine whether the is a linear relationship
- Procedure: Analyze→ correlate→ bivariate

Correlations

		Salary per Day (£)	Attractiveness (%)
Salary per Day (£)	Pearson Correlation	1	.068
	Sig. (2-tailed)		.304
	N	231	231
Attractiveness (%)	Pearson Correlation	.068	1
	Sig. (2-tailed)	.304	
	N	231	231

- The null hypothesis states: There is no linear relationship between the variables.
- The alternative hypothesis is: There is a linear relationship between the variables
- The maximal acceptable level of significance is 5% (alpha=0,05).
- ***The salary per day was not significanlty correlated with the beauty of a supermodel (attractiveness in %), r=.068, p>.05. The value for r represents a positive and small effect.***
- Null hypothesis cannot be rejected
- The significance test examines whether there is a relationship between the variables, but does not say something about the strength of the relationship!! The strength is given with the correlation coefficient
- ***The significance test is the probability at which the correlation coefficient showing in the sample with the given sample size also shows with the same size in the population if there is in reality no relationship in the population***
- As it was unkown at the beginning of the analysis whether there is a positive or negative relationship between the variables, a 2-tailed test has been conducted (independent if there is a negative or positive relationship)

4.2 Cross tables

- Are used to show the joint frequency distribution of two variables
- Case groups of variables are examined, that are defined by a combination of characteristics of both variables
- Moreover, there is a statistical test which tests whether or not there is relationship between the variables
- Examines, at which probability the objective relationship of the sample also exists in the population
- **Cross tables**

- *Show the joint frequency of two variables*
- *Not only absolute frequencies (as with the single frequency table) but also the relative frequencies = expected frequencies are shown*
- *Differences in the values of the cross table and the frequency table can exist, as persons with missing values are not included in the analysis*
- *The significance test to analyse possible relationships is based on the expected frequencies of two variables*

- **Chi-Quadrat-test:**
    - Examines whether we can infer from the sample observations that there is relationship between these 2 variables also in the population or not
- **Measures of association:**
    - Aim at expressing the strength of the relationship between 2 variables with a measure
- Categorical variables are very suitable
- Analyze → Descriptive Statistics → Crosstabs

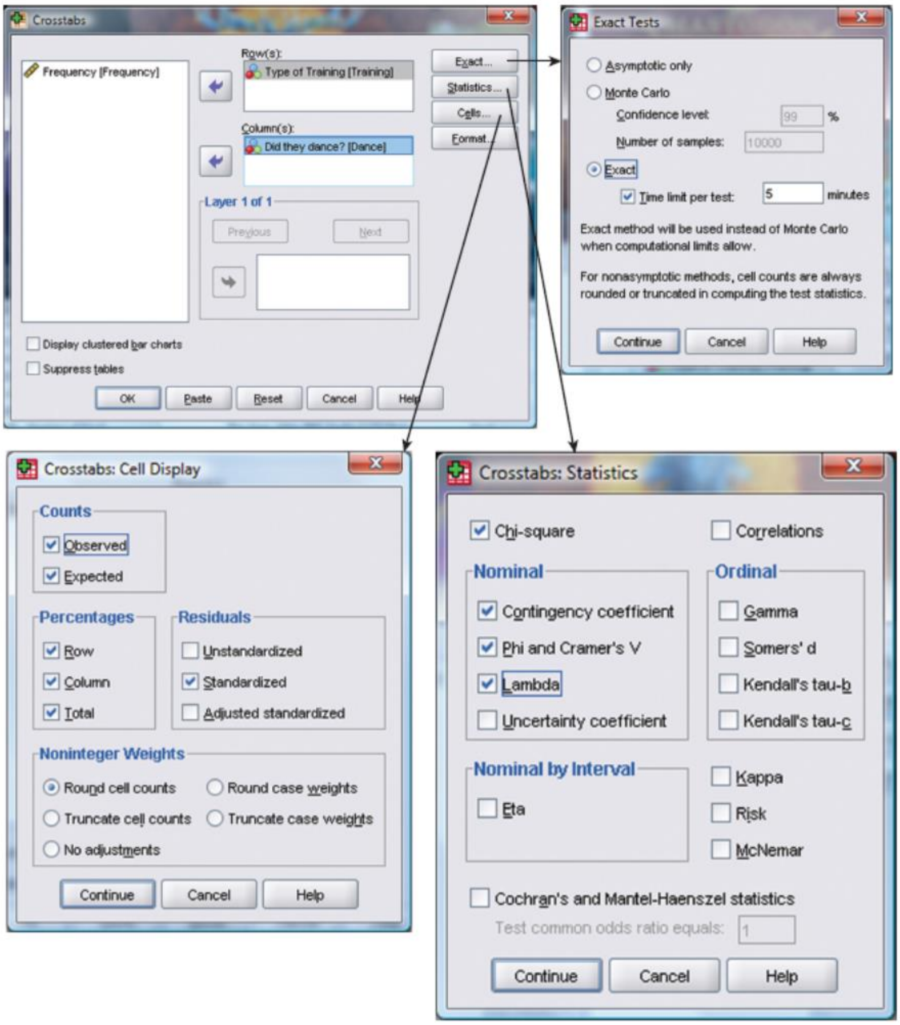


FIGURE 18.4 Dialog boxes for the *crosstabs* command

EXERCISE:

Data set: 04_CosmeticSurgery.sav

- Is there an association between the reason to have cosmetic surgery and gender?

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Reason for Surgery	276	100.0%	0	0.0%	276	100.0%

Gender * Reason for Surgery Crosstabulation

			Reason for Surgery		
			Physical reason	Change Appearance	Total
Gender	Female	Count	82	75	157
		Expected Count	55.7	101.3	157.0
		% within Gender	52.2%	47.8%	100.0%
		% within Reason for Surgery	83.7%	42.1%	56.9%
		% of Total	29.7%	27.2%	56.9%
		Std. Residual	3.5	-2.6	
	Male	Count	16	103	119
		Expected Count	42.3	76.7	119.0
		% within Gender	13.4%	86.6%	100.0%
		% within Reason for Surgery	16.3%	57.9%	43.1%
		% of Total	5.8%	37.3%	43.1%
		Std. Residual	-4.0	3.0	
Total		Count	98	178	276
		Expected Count	98.0	178.0	276.0
		% within Gender	35.5%	64.5%	100.0%
		% within Reason for Surgery	100.0%	100.0%	100.0%
		% of Total	35.5%	64.5%	100.0%

- Table contains the number of cases that fall into each combination of categories
- ***We can see that 98 of people (35.5% of total) had a cosmetic surgery due to a physical reason and of these 82 were female (83.7% of the people having a surgery for a physical reason) and 16 were male (16.3% of all people with surgery due to a physical reason).***
- ***Further, 178 people (64.5% of total) had a cosmetic surgery in order to change their appearance. Of those having a surgery to change***

appearance, 75 were female (42.1% of those having a cosmetic surgery to change appearance) and 103 were male (57.9% of those having a cosmetic surgery to change appearance).

- ***Of the females, 52.2% had a cosmetic surgery for a physical reason and 47.8% to change appearance. Similarly, of the males, 13.4% had a cosmetic surgery for a physical reason and 86.6% to change appearance***
- ***In summary, for females, the reason to have cosmetic surgery is in equal parts physical but also to change appearance. For males however, the vast majority has cosmetic surgery to change appearance***
- ***H0: the variables are independent (there is no association between gender and the reason for cosmetic surgery)***
- ***H1: the variables are dependent (there is an association between gender and the reason for cosmetic surgery)***

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	44.464[a]	1	.000		
Continuity Correction[b]	42.787	1	.000		
Likelihood Ratio	47.802	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	44.303	1	.000		
N of Valid Cases	276				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 42.25.

b. Computed only for a 2x2 table

- ***Checking the assumptions of the Chi-square test: all expected frequencies should be greater than 5. Smallest count according to footnote a is 42.25, so the assumption has been met***
- Pearson's Chi-square test examines whether there is an association between two categorical variables, test statistic tests whether the two variables are independent. If the significance value is small enough, we reject the null hypothesis that the variables are independent
- The value of the Chi-squre statistic is 44.464 and the value is highly significant ($p<.001$), indicating that gender had a significant effect on the reason to have cosmetic surgery
- Therefore we can conclude, gender is significantly associated to the reason to have cosmetic surgery, cosmetic surgery to change appearance is much more likely for males than for females

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.401	.000
	Cramer's V	.401	.000
	Contingency Coefficient	.372	.000
N of Valid Cases		276	

- The table with the symmetric measures shows the strength of the association. For Cramer's Phi has a value of 0.401 indicating a medium effect

REPORTING**: *There was a significant association between gender and the reason to have cosmetic surgery χ2 (1) = 44.464, p< .001. The size of the effect was medium*.**

This finding seems to represent the fact that based on the odds ratio, the odds of having cosmetic surgery for a physical reason were 6.81 times higher if the people were females comcpared to males

- Standardized residuals show that for females the standardized residuals were significant for both having cosmetic surgery for a physical reason or to change appearance (both values are bigger than 1.96; z=3.5 and -2.6). The plus or minus sign tells us that for females significantly more had cosmetic surgery for a physical reason and significantly less had surgery to change appearance. For males, significantly less had cosmetic surgery for physical reason but more to change appearance (z=-4.0 and 3.0, resp.). All cells contribute significantly to the overall Chi-square statistic

TEST CROSSTABS

Data Set: 08_Facebook.sav

- Is there an association between RMiP Exam Result and whether or not students looked at Facebook during lab classes?

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Looked At facebook During lab Classes * RMiP Exam Result	260	100.0%	0	0.0%	260	100.0%

Looked At facebook During lab Classes * RMiP Exam Result Crosstabulation

			RMiP Exam Result		Total
			Pass	Fail	
Looked At facebook During lab Classes	Looked at Facebook	Count	44	60	104
		Expected Count	67.2	36.8	104.0
		% within Looked At facebook During lab Classes	42.3%	57.7%	100.0%
		% within RMiP Exam Result	26.2%	65.2%	40.0%
		% of Total	16.9%	23.1%	40.0%
		Std. Residual	-2.8	3.8	
	Did Not Look at Facebook	Count	124	32	156
		Expected Count	100.8	55.2	156.0
		% within Looked At facebook During lab Classes	79.5%	20.5%	100.0%
		% within RMiP Exam Result	73.8%	34.8%	60.0%
		% of Total	47.7%	12.3%	60.0%
		Std. Residual	2.3	-3.1	
Total		Count	168	92	260
		Expected Count	168.0	92.0	260.0
		% within Looked At facebook During lab Classes	64.6%	35.4%	100.0%
		% within RMiP Exam Result	100.0%	100.0%	100.0%
		% of Total	64.6%	35.4%	100.0%

- ***Table contains the number of cases that fall into each combination of categories***
- ***We can see that 168 of people (64,6% of total) passed the exam and of these 44 looked at FB (26.2% of the students passing) and 124 did not look at FB (73.8% of students passing).***
- ***Further, 92 people (35,4% of the total) did fail. Of those failing, 60 did look at FB (65,2% of those failing) and 32 did not look at FB (34,8% of those failing).***
- ***Of those looking at FB, 42,3% did pass and 57,7% did fail. Similarly, of those not looking at FB, 79,5% passed and 20,5% did not***
- ***In summary, most students passing the exam did not look at FB during class. Of those students failing the exam, the share of students looking at FB is higher.***
- ***H0: the variables are independent (there is no association between exam result and looking at FB during class)***
- ***H1: the variables are dependent (there is an association between exam result and looking at FB during the exam)***

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	37.726[a]	1	.000		
Continuity Correction[b]	36.117	1	.000		
Likelihood Ratio	37.872	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	37.581	1	.000		
N of Valid Cases	260				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 36.80.

b. Computed only for a 2x2 table

- ***Checking the assumptions of the Chi-square test: all expected frequencies should be greater than 5. Smallest count according to footnote a is 36.80, so the assumption has been met***
- Pearson's Chi-square test examines whether there is an association between two categorical variables, test statistic tests whether the two variables are independent. If the significance value is small enough, we reject the null hypothesis that the variables are independent
- The value of the Chi-squre statistic is 37.726 and the value is highly significant ($p<.001$), indicating that there is an association between the exam result and looking at FB during class.
- Therefore we can conclude, looking at FB or not has a significant effect on the result of the exam

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-.381	.000
	Cramer's V	.381	.000
	Contingency Coefficient	.356	.000
N of Valid Cases		260	

- The table with the symmetric measures shows the strength of the association. For Cramer's Phi has a value of -0.381 indicating a medium effect

REPORTING*: *There was a significant association between looking at FB during class and the exam result $\chi2$ (1) = 37.726, p< .001. The effect size is medium.

This finding seems to represent the fact that based on the odds ratio, if the studetns looked at FB during the class the odds of their failing the exam was 5.23 times higher than if they had not looked

- Standardized residuals show that for those looking at FB the standardized residuals were significant for both failing and passing the exam (both values are bigger than 1.96; z=-2,8 and 3,8). The plus or minus sign tells us that for those looking significantly more failed and significantly less passed the exam after looking at FB. For those not looking at FB, significantly more passed but less failed (z=2,3 and -3.1, resp.). All cells contribute significantly to the overall Chi-square statistic

5 Regression analysis

5.1 ANOVA

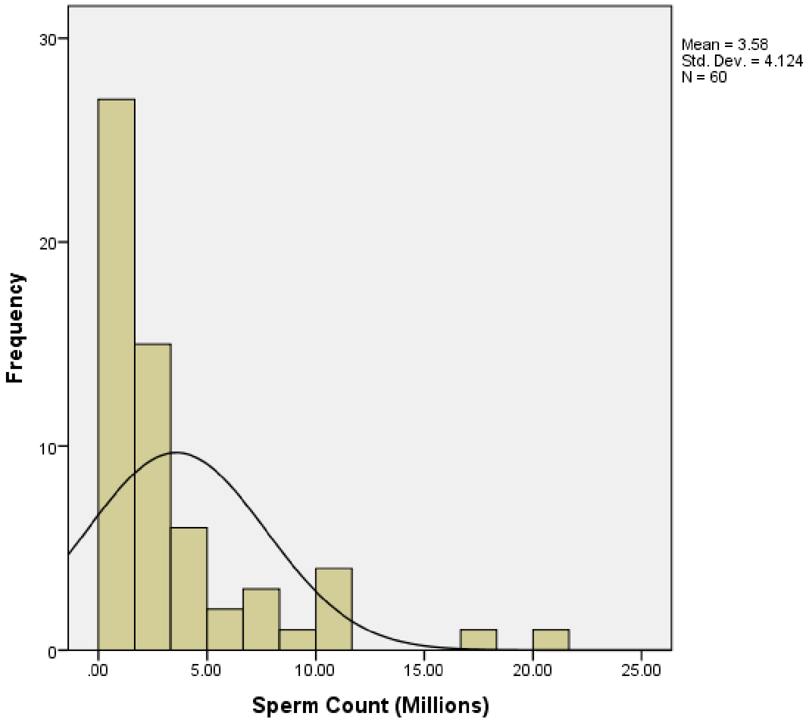
- Similar hypotheses to the t-Test:
  Are the means of different subgroups of the same variables equal?
- Different to the t-Test:
  T-test compares 2 means, ANOVA compares more than 2 means
- **Differentiation between:**
- **One-way ANOVA (1 independent variable) = ANOVA**
- **Multivariate ANOVA (> 1 independent variable)**
- **One-dimensional ANOVA (1 dependent variable)**
- **Multidimensional ANOVA (> 1 dependent variable)**
- Data requirements::
- **Variable of interest needs to have interval scale**

- **Data needs to be normally distributed**
- **Variance homogeneity: the variance of the variable of interestw needs to be equal among the different categories**

EXERCISE:

Data Set: 06_Soya.sav

- Run the following analysis excluding those males that ate only one soya meal per week (filter cases).
- Is there a difference in the mean number of sperms depending on the frequency of soya consumption (no soya meals; four soya meals per week; seven soya meals per week)?
- Select cases
    - Data → select cases → if condition is satisfied: Soya=1 | Soya=3 | Soya=4 → ok
- Check if number of soya meals is normally distributed
- Graph → Legacy Dialogs → Histogram, select display normal curve



- Variable is not normally distributed
- One option is to transform the dependent variable

```
compute sperm2=sperm*sperm.
execute.
compute lnsperm=ln(sperm).
execute.
```

- Check if transformed variables are normally distributed
- Analyze→ Nonparametric tests → Legacy dialogs → 1 sample KS → test variables: sperm count, sperm2, lnsperm

One-Sample Kolmogorov-Smirnov Test

		Sperm Count (Millions)	sperm2	lnsperm
N		60	60	60
Normal Parameters[a,b]	Mean	3.5833	29.5646	.7723
	Std. Deviation	4.12409	74.17445	1.01305
Most Extreme Differences	Absolute	.251	.346	.074
	Positive	.251	.339	.074
	Negative	-.214	-.346	-.053
Kolmogorov-Smirnov Z		1.944	2.677	.570
Asymp. Sig. (2-tailed)		.001	.000	.902

a. Test distribution is Normal.

b. Calculated from data.

- Lnsperm is normally distributed, use this transformed variable to continue with analysis
- Analyze → compare means → one-way ANOVA → select lnsperm as dependent variable → under posthoc select LSD, Tukey, Games-Howell → under options select descriptive, homogeneity of variance → ok

Descriptives

lnsperm

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
No Soya Meals	20	1.0918	1.11023	.24826	.5722	1.6114	-1.05	3.05
4 Soyal Meals Per Week	20	.9689	.97157	.21725	.5142	1.4236	-.91	2.90
7 Soya Meals Per Week	20	.2563	.76277	.17056	-.1007	.6133	-1.17	1.41
Total	60	.7723	1.01305	.13078	.5106	1.0340	-1.17	3.05

- Table of descriptives from the one-way procedure
- We see that means are different depending on the frequency of soya meals per week, highest for no soya meals, lowest for 7 times soya per week

REPORTING: The table shows the descriptive statistics of the dependent variable, the logarithm of sperm count. It shows that means are different, with the highest means for those eating no soya per week and the lowest for those eating soya 7 times per week

Test of Homogeneity of Variances

lnsperm

Levene Statistic	df1	df2	Sig.
1.876	2	57	.163

- The next part is a summary of Levene's test to test the assumption of homogeneity of variances

IMPORTANT: If assumption is broken, select the Welch and Brown-Forsythe option (robust tests of the equality of means when assumption of homogeneity of variances is broken) and report those F-ratios and Sig values as results!

REPORTING: The null hypothesis for Levene's test is: variances are equal. The alternative hyhpothesis is: variances are not equal. Levene's test for equality of variances was found to be met for the present analysis,F(2,57) = 1.876, p = .163

ANOVA

lnsperm

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.140	2	4.070	4.427	.016
Within Groups	52.409	57	.919		
Total	60.550	59			

- Main ANOVA summary table

REPORTING: The null hypothesis is: The group means are equal in the population. Mean sperm count (the ln of it) does not differ depending on the frequency of soya consumption. The alternative hypothesis is: The group means are different in the population. Mean sperm count (the ln of it) does differ depending on the frequency of soya consumption.

- Table is divided into between-group effects (effects due to the model – the experimental effect) and within-group effects (this is the unsystematic variation in the data), sum of squres, df and mean squares are reported
- The test of whether group means are the same is represented by the F-ratio
- It is greater than 1 indicating a significant result
- Finally, SPSS tells us whether this value is likely to have happened by chance
- The final column labelled Sig. indicates the likelihood of an F-ratio of this size would occur if in reality there was no effect (that is only a 1.6% chance)
- Hence, because the observed significance value is less than .05 we can say the there was a significant effect of soya consumption frequency

Multiple Comparisons

Dependent Variable: lnsperm

	(I) Number of Soya Meals Per Week	(J) Number of Soya Meals Per Week	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	No Soya Meals	4 Soyal Meals Per Week	.12289	.30323	.914	-.6068	.8526
		7 Soya Meals Per Week	.83551*	.30323	.021	.1058	1.5652
	4 Soyal Meals Per Week	No Soya Meals	-.12289	.30323	.914	-.8526	.6068
		7 Soya Meals Per Week	.71263	.30323	.057	-.0171	1.4423
	7 Soya Meals Per Week	No Soya Meals	-.83551*	.30323	.021	-1.5652	-.1058
		4 Soyal Meals Per Week	-.71263	.30323	.057	-1.4423	.0171
LSD	No Soya Meals	4 Soyal Meals Per Week	.12289	.30323	.687	-.4843	.7301
		7 Soya Meals Per Week	.83551*	.30323	.008	.2283	1.4427
	4 Soyal Meals Per Week	No Soya Meals	-.12289	.30323	.687	-.7301	.4843
		7 Soya Meals Per Week	.71263*	.30323	.022	.1054	1.3198
	7 Soya Meals Per Week	No Soya Meals	-.83551*	.30323	.008	-1.4427	-.2283
		4 Soyal Meals Per Week	-.71263*	.30323	.022	-1.3198	-.1054
Games-Howell	No Soya Meals	4 Soyal Meals Per Week	.12289	.32989	.927	-.6822	.9280
		7 Soya Meals Per Week	.83551*	.30120	.024	.0971	1.5739
	4 Soyal Meals Per Week	No Soya Meals	-.12289	.32989	.927	-.9280	.6822
		7 Soya Meals Per Week	.71263*	.27620	.037	.0375	1.3878
	7 Soya Meals Per Week	No Soya Meals	-.83551*	.30120	.024	-1.5739	-.0971
		4 Soyal Meals Per Week	-.71263*	.27620	.037	-1.3878	-.0375

*. The mean difference is significant at the 0.05 level.

- We asked SPSS to carry out post hoc tests
- This Table shows the results of the Tukey test, LSD and Games-Howell test
- If we look at the Tukey test first, it is clear from the table that each group of participants is compared to all of the remaining groups
- For each pair of groups the difference between group means is displayed, the standard error of t hat difference, the significance level of that difference and a 95% confidence interval
- First of all, the no soya meal group is compared to the 4 soya meals per week and reveals a nonsignificant difference (Sig. is greater than .05), but when compared to the 7 soya meals per week group, there is a significant difference (Sig is less than .05)
- The 4 soya meals per week is then compared to both the no soya meals and the 7 soya meals per week groups. The group means differ only between the 4 and 7 group
- The rest of the table reports the LSD and Games-Howell test. A quick inspection reveals the same patterns of results: sperm counts differs between no and 7 and 4 and 7 but not between no and 4

REPORTING*: *A one-way ANOVA was used to test the sperm count differences between three different consumption frequencies of soya meals. There was a significant effect of soya meals consumption frequency on sperm count, $F(2, 57)=4.43$, $p<.05$.

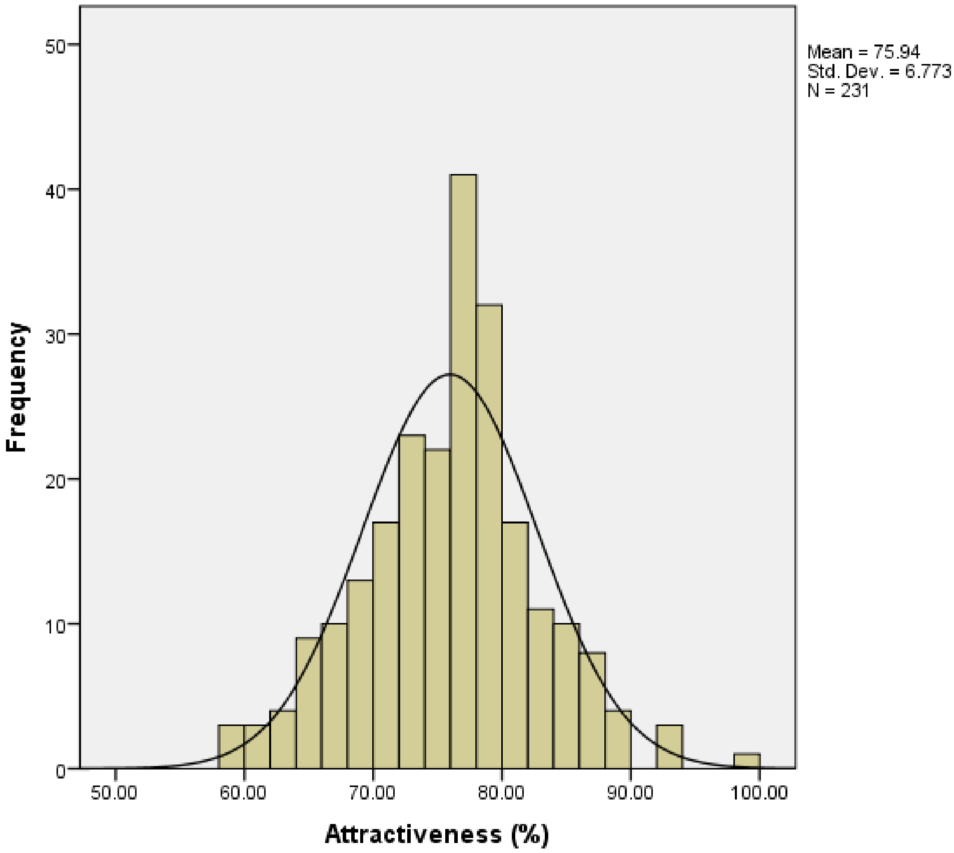
Post hoc analyses using the Tukey HSD and Games-Howell test indicated that the average sperm count was lower in the 7 soya meal per week group (M=.26, SD=1.01) than in the other two conditions (no soya meals: M=1.1, SD=1.11; 4 soya meals: M=.97, SD=.97), $p<.05$. The difference between the no soya and 4 times soya group was not significant.

TEST: Data Set: 07_Supermodel.sav

- Generate a new variable by dividing the age of the supermodel into the following categories:
    - youngster: age lt 17 years old
    - best age: between 17 and 19 years old
    - senior: older than 19 years
- Does the average beauty (attractiveness in % judged by a jury) differ by age category?

```
compute agecat=0.
if (age lt 17) agecat=1.
if (age ge 17 and age le 19) agecat=2.
if (age gt 19) agecat=3.
execute.
```

- Check for normality of dependent variable
- Histogram
- Variable is normally distributed



- Run one-way ANOVA, select posthoc and Welch and Brown-Forsythe (robust tests of the equality of means when assumption of homogeneity of variances is broken)

Descriptives

Attractiveness (%)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
1.00	72	72.8753	7.40145	.87227	71.1361	74.6146	58.13	93.25
2.00	97	77.4441	5.98557	.60774	76.2378	78.6505	59.91	99.22
3.00	62	77.1633	6.05125	.76851	75.6265	78.7000	65.17	92.53
Total	231	75.9447	6.77303	.44563	75.0667	76.8227	58.13	99.22

- Table shows descriptive statistics, it shows that the mean beauty differs by age category, especially the beauty of the youngsters is rated lower than that of the two other categories.

REPORTING*: *The table shows the descriptive statistics of the dependent variable, the attractiveness in %. It shows that means are different, with the highest means for middle age category and the lowest for the youngest

Test of Homogeneity of Variances

Attractiveness (%)

Levene Statistic	df1	df2	Sig.
3.304	2	228	.039

- The next part is a summary of Levene's test to test the assumption of homogeneity of variances

REPORTING*: *The null hypothesis for Levene's test is: variances are equal. The alternative hyhpothesis is: variances are not equal. Levene's test for equality of variances was found to be violated for the present analysis,F(2,228) = 3.304, p < .05.

- For this reason, the Robust Tests for Equality of Means are used to interprete the ANOVA results

ANOVA

Attractiveness (%)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	988.453	2	494.227	11.784	.000
Within Groups	9562.551	228	41.941		
Total	10551.004	230			

IMPORTANT: Because assumption was broken, we selected the Welch and Brown-Forsythe option (robust tests of the equality of means when assumption of homogeneity of variances is broken) and report those F-ratios and Sig values as results and not the results in this ANOVA output!

Robust Tests of Equality of Means

Attractiveness (%)

	Statistic[a]	df1	df2	Sig.
Welch	10.061	2	137.780	.000
Brown-Forsythe	11.591	2	200.401	.000

a. Asymptotically F distributed.

- Main robust tests of equality of means summary table

REPORTING: The null hypothesis is: The group means are equal in the population. Mean attractiveness (in %) does not differ depending on the age category. The alternative hypothesis is: The group means are different in the population. Attractiveness (in %) does differ depending on the age category.

The output from the above table is only valid if the equal variance assumption has been violated. From this example, using the Welch

statistic, we find that F(2, 137.78) = 10.061, p <.001. This indicates that there is a significant difference in the mean attractiveness in % and the age category.

Multiple Comparisons

Dependent Variable: Attractiveness (%)

	(I) agecat	(J) agecat	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	1.00	2.00	-4.56878*	1.00742	.000	-6.9454	-2.1922
		3.00	-4.28791*	1.12204	.001	-6.9349	-1.6409
	2.00	1.00	4.56878*	1.00742	.000	2.1922	6.9454
		3.00	.28088	1.05302	.962	-2.2033	2.7651
	3.00	1.00	4.28791*	1.12204	.001	1.6409	6.9349
		2.00	-.28088	1.05302	.962	-2.7651	2.2033
Games-Howell	1.00	2.00	-4.56878*	1.06311	.000	-7.0885	-2.0490
		3.00	-4.28791*	1.16252	.001	-7.0437	-1.5321
	2.00	1.00	4.56878*	1.06311	.000	2.0490	7.0885
		3.00	.28088	.97977	.956	-2.0422	2.6040
	3.00	1.00	4.28791*	1.16252	.001	1.5321	7.0437
		2.00	-.28088	.97977	.956	-2.6040	2.0422

*. The mean difference is significant at the 0.05 level.

We asked SPSS to carry out post hoc tests

- This Table shows the results of the Tukey test and the Games-Howell test
- If we look at the Tukey test first, it is clear from the table that each group of participants is compared to all of the remaining groups
- For each pair of groups the difference between group means is displayed, the standard error of t hat difference, the significance level of that difference and a 95% confidence interval

- First of all, the group of the youngest is compared to the group of the middle aged supermodels and reveals a significant difference (Sig. is smaller than .01), also when compared to the group of the oldest supermodels, there is a significant difference (Sig is less than .05)
- The group of the middle-aged supermodels is then compared to both the youngest and the oldest group. The group means do not differ between the middle-aged and the oldest group
- The rest of the table reports the Games-Howell test. A quick inspection reveals the same patterns of results: mean attractiveness in % differs between the youngest and the middle-aged and the youngest and the oldest, but not between the middle-aged and the oldest

REPORTING*: *A one-way ANOVA was used to test the attractivenenss differences between three different age categories of supermodels. There was a significant effect of the age category on attractiveness, F(2, 137.78) = 10.061, p <.001.

Post hoc analyses using the Tukey HSD and Games-Howell test indicated that the mean attractiveness was lower in the group of the youngest

supermodels (M=72.87, SD=7.40) than in the other two conditions (middle-aged: M=77.44, SD=5.99; oldest: M=77.16, SD=6.05), p<.05. The difference between the middle-aged and the group of the oldest supermodels was not significant.

5.2 Regression analysis

Analyses the relationship between a dependent and one or more independent variables

Goal: To describe relationships quantiatively, to explain the values of the dependent variable

Application:

- Cause analysis
- Prediction of the dependent variable
- Time series analysis

Principle: A line is searched that minimises the derivations of single values from the line = Minimisation of squared residuals (least squares)

Difference to correlation analysis: the direction of the relationship needs be clearly given (e.g. income has an impact on the expenditures for food)

5.2.1 Assumptions:

There must be causality

Variables need to have a metric scale

Introduction of independent nominally-scales variables as dummy variable

Linear relationship between dependent and independent variables

Model assumption (will not be verified in the course)

Normally distributed residuals

Linear relationship between the dependent variable and the independent variables

No multicollinearity

No heteroscedasticity

No autocorrelation

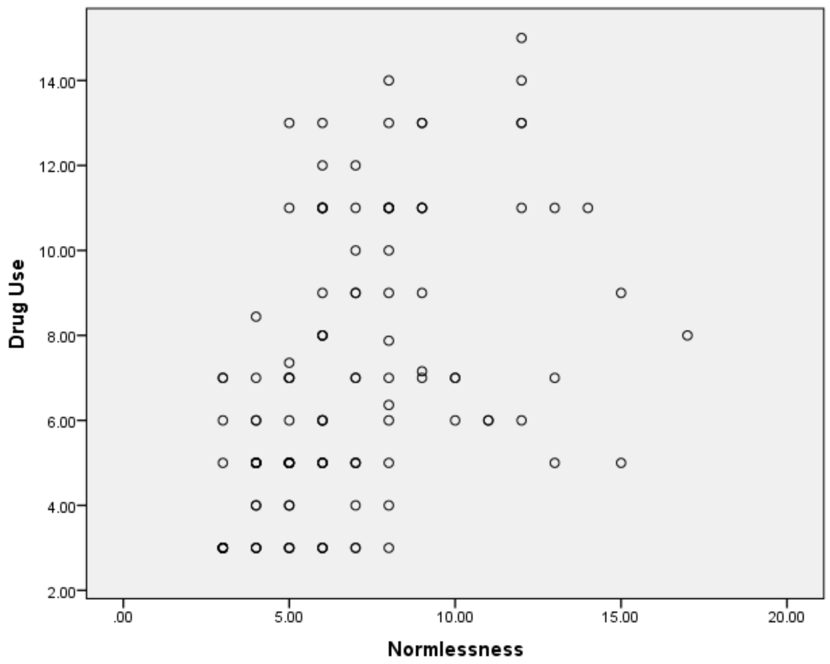
Data set: 03_HeavyMetal.sav

Does drug use (amount of joints smoked in one week) depend on age?

First of all, a scatter plot is useful

- Graphs—Y Legacy dialogs→ Scatter/Dott→ Simple scatter → →define→Y-Achse drug use→x-Achse normelssness

The selection of the axes is important! The y-axis shows the dependent, the x-axis show the independent variable



5.2.2 Model

- Y = f(X)
- Y = b0+b1*x
- Drug use=f(normlessness)
- Drug use=3.64+0.50*normlessness+e

B_0: Constant, which is the value if X=0.

B_1: Regression coefficient determines the slope of the line, thus it determines the effect of a change in X $B_1 = \Delta Y / \Delta X$

Conduct a regression analysis

- Analyze→Regression→linear

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Normlessness[b]	.	Enter

a. Dependent Variable: Drug Use

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.462[a]	.213	.207	2.86710	1.952

a. Predictors: (Constant), Normlessness

b. Dependent Variable: Drug Use

5.3 Verification of the regression function:

Quality criteria to verify the regression function

- Coefficient of determination R^2
- F-Statistic
- Standard error

Measures to verify the regression coefficients

- T-value
- Beta-value

5.3.1 Coefficient of determination: R^2

Analyses the quality of the regression function to empirical data

Measures the share of explained variation on the total variation

R^2 = explained variation / total variation

Measures how cloesly the regression line lies at the points of the scatter

Values between 0 and 1; 1= regression line is consistant with the scatter, the higher, the higher the share of explained variation on the total variation

The addition of another explaining variable can never worsen the R2, it can only improve the R2; this can tempt to include as many explaining variables in the estimation as possible

R^2 depends on the number of regressors. At a given sample size each additional regressor adds a bigger or smaller share of explanation, which might only be a coincidence. The value of R^2 can never decline with additional regressors. This is why the adjusted R2 is used.

5.3.2 Adjusted R^2

An increase in the number of explaining variables can result in a decline of the adjusted R^2

5.3.3 Standard error of the estimation

Measure about how close the predicted values come to the real values. In other words: What is the mean error that is made from estimating the regression function to estimate the dependent variable Y.

The R2 of the regression analysis shows that the normlessness explains drug use with 20.7%.

The explanatory variable available income has an explanatory power of 20.7% for the dependent variable drug use

The regression model explains the variation of the dependent variable to 20.7%

> **REPORTING: The quality of the regression is determined by the R2. Drug use can be explained with 20.7% with the variable normlessness. This means that the independent variable in the regression explains the distribution of the dependent variable to 20.7%.**

Durbin-Watson statistic (not relevant for exam)

DW tests the assumption of independent errors

For any two observations the residual terms should be uncorrelated (or independent)

This eventuality is sometimes described as lack of autocorrelation

This assumption can be tested with the Durbin-Watson test, which tests for serial correlations between errors

Specifically, it tests whether the adjacent residuals are correlated

The test statistic can vary between 0 and 4 with a value of 2 meaning that the residuals are uncorrelated

A value greater than 2 indicates a negative correlation between adjacent residuals, whereas a value below 2 indicates a positive correlation

The size of the Durbin-Watson statistic depends upon the number of predictors in the model and the number of observations

As a rule of thumb, values less than 1 or greater than 3 are definitely cause for concern, values around 2 are ok

Here, the DW value is close to 2, ok

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	265.036	1	265.036	32.242	.000[b]
	Residual	978.214	119	8.220		
	Total	1243.251	120			

a. Dependent Variable: Drug Use

b. Predictors: (Constant), Normlessness

5.4 F-Test to test the significance of the regression function

Afterwards, an F-Test is conducted. The F-test checks whether the regression function is significant, that is if at least 1 coefficient of the model is significant

Can the result from the sample be transfered to the population?

Next to the distribution, also the sample size is considered

H0: There is no linear relationship between the dependent and the independet variables, all betas =0

H1: There is a linear relationship between the dependent and the independent variables, betas are ≠ 0.

To reject H0 means that there is a linear relationship between the dependent and independent variable also in the population.

> ***REPORTING: To observe whether the regression model is significantly better in predicting the model than using the mean, an F-test is conducted. H0: There is no linear relationship between the dependent and the independet variable, all betas in the population equal 0; H1: There is a linear relationship between the dependent and the independent variable,***

betas are different from 0. Here, the F test statistic value is 32.24 and it is highly significant (p<.001). This means that at least one coefficient in the regression function is different from 0.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.636	.653		5.568	.000
	Normlessness	.500	.088	.462	5.678	.000

a. Dependent Variable: Drug Use

Coefficients table reports the test results for the single predictors in the model

5.5 Test of single regression coefficients

If the global test of the regression function was significant, that is not all coefficients are 0, and thus a sigificance between the variables has been shown, the single regression coefficients needs to be tested next. H0 is: bj = 0. If the empircal value is different from 0, it is very unlikely that the null hypothesis is correct.

If there is only one independent variable in the model, the F-test of the model is also a test of the variable whose impact is examined. It would be enough to conduct only one of the tests.

Using the t-Test, single regression coefficients are tested whether or not they are different from 0

Contrary to the F-value, a t-value can become negative that is why one has to consider the absolute value.

T-value higher than 2 for at least 30 observations (circa)

T-value shows the ratio between the coefficient and the standard deviation

Significance level: >0,05 Beta not significant

The impact of the normlessness is significant, as p=0 and therefore lower 0.005

REPORTING: Regression function: Drug Use=3.64+0.50*normlessness+e

The null hypothesis of the constant is: the constant is 0. The alternative hypothesis is: the constant is not 0. The constant shows the height of drug is if normless is 0. The t-test statistic has a value of 5.59 and is highly significant (p<.001). It shows that baseline drug use is 3.64 joints smoked per week (given that normlessness is 0).

The null hypothesis for normlessness is: beta=0. The alternative hypothesis is: beta is not 0. There is a linear relationship between drug use and normlessness. The t-test statistic has a value of 5.68 and the value is highly significant (p<.001) meaning that if normlessness increases about 1 unit, drug use increases about 0.5 units.

Single regression analysis was used to test if normlessness significantly predicted drug use. The results of the regression indicated the predictor explained 20.7% of the variance (F(1,119)=32.24, p<.001). It was found that normlessness significantly predicted drug use (beta= 0.50, p<.001).

EXERCISE II:

Data set: 03_HeavyMetal.sav

- Test the hypothesis that drug use depends on normlessness and being young. Discuss the quality of your analysis and report the regression function.
- Generate a dummy variable for young using the variable age_group=1.

    compute young=0.

    if (age_group=1) young=1.

    execute.

➔ Why not simply use binary age_group variable? Because it is not a dummy variable, needs to be 0/1 coded, data set description shows it is 1/2 coded

Dummy Variables:

By using dummies it is possible to also use qualitative normal scale variables. These normal scaled variables are transferred into binary variables which can be handled as metric variables. As however, the number of independent variables increases considerably, this technique is only possible for the dependent variable (but logit regression).

One of the dummies needs to be excluded and becomes the reference category!

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	young, Normlessness[b]	.	Enter

a. Dependent Variable: Drug Use

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.467[a]	.218	.205	2.87054	1.959

a. Predictors: (Constant), young, Normlessness

b. Dependent Variable: Drug Use

***REPORTING**: The quality of the regression is determined by the R2. Drug use can be explained with 21.8% with the variables normlessness and being young. This means that the independent variables in the regression explain the distribution of the dependent variable to 21.8%.*

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	270.934	2	135.467	16.440	.000[b]
	Residual	972.317	118	8.240		
	Total	1243.251	120			

a. Dependent Variable: Drug Use

b. Predictors: (Constant), young, Normlessness

***REPORTING*: To observe whether the regression model is significantly better in predicting the model than using the mean, an F-test is conducted. H0: There is no linear relationship between the dependent and the independet variables, all betas in the population equal 0; H1: There is a linear relationship between the dependent and the independent variables, betas are different from 0. Here, the F test statistic has a value of 16.44 and it is highly significant (p<.001). This means that at least one coefficient in the regression function is different from 0.**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.515	.669		5.255	.000
	Normlessness	.484	.090	.447	5.373	.000
	young	.451	.533	.070	.846	.399

a. Dependent Variable: Drug Use

***REPORTING*: Regression function: Drug Use=3.52+0.48*normlessness+0.45*young+e**

The null hypothesis of the constant is: the constant is 0. The alternative hypothesis is: the constant is not 0. The constant shows the height of drug if normless and young are 0. The t-test statistic has a value of 5.26 and it is highly significant (p<.001). It shows that baseline drug use is 3.52 joints smoked per week (given that the model predictors are 0).

The null hypothesis for normlessness is: beta=0. The alternative hypothesis is: beta is not 0. There is a linear relationship between drug use and normlessness. The t-test statistic has a value of 5.37 and the value is highly significant (p<.001) meaning that if normlessness increases about 1 unit, drug use increases about 0.48 units.

The null hypothesis for young is: Drug use does not differ between young and old. The alternative hypothesis is: drug use does differ between young and old. The t-test statistic has a value of 0.846 and the value is not significant (p>.05) indicating that there is no difference in drug use between young and old. (If it would be significant: Young heavy metal listeners have a drug use which is about 0.45 units higher than the drug use of the old heavy metal listeners.)

Multiple regression analysis was used to test if normlessness and being young significantly predicted drug use. The results of the multiple regression indicated the predictors explained 21.8% of the variance (F(2,118)=16.44, p<.001). It was found that normlessness significantly predicted drug use (beta= .48, p<.001) and that being young not significantly predicted drug use (beta=0.45; p>.05).

5.6 Comparison of the standardised and not-standardised beta-coefficient:

It would be wrong to infer from the fact that the coefficient of one variable is ten times higher than the coefficient of another variable that the first variable also has a ten times higher impact

It needs to be considered that the variables have a different unit, and that is why the variables are different

To compare the magnitude of the impact of single variables, the standardised coefficients are to be used (beta coefficients)

The beta coefficients are the coefficients that would result if the dependent and all of the independent variables would have been standardised (z- transformation)

The interpretation of the beta coefficients is only acceteptable if the independent variables are not correlated with each other, thus there is no multcollinearity

Regression test:

Data Set: 08_Soya.sav

Does the number of sperms depend on the number of times soya eaten per week?

Generate dummy variables for each of the 4 categories of soya

Attention: at best, a dummy variable includes not only dummies

Attention II: if the four dummy variables, only three enter the regression (the one excluded is the reference)

Attention III: DW statistic not optimal, we continue to learn about how to report regression

Attention IV: F-statistic not significant, the model is not good! Just to see how to use dummies when we have more than 2 categories!!

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	seven, four, once[b]	.	Enter

a. Dependent Variable: Sperm Count (Millions)

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.307[a]	.094	.058	4.13403

a. Predictors: (Constant), seven, four, once

REPORTING: The quality of the regression is determined by the R2. Spper count can be explained with 9.4% with the variables once, seven and four.

This means that the independent variables in the regression explain the distribution of the dependent variable to 9.4%.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	135.130	3	45.043	2.636	.056[b]
	Residual	1298.853	76	17.090		
	Total	1433.983	79			

a. Dependent Variable: Sperm Count (Millions)

b. Predictors: (Constant), seven, four, once

REPORTING: To observe whether the regression model is significantly better in predicting the model than using the mean, an F-test is conducted. H0: There is no linear relationship between the dependent and the independet variable, all betas =0; H1: There is a linear relationship between the dependent and the independent variable, betas are ≠ 0. Here, the F test statistic has a value of 2.64 and it is not significant (p>.05). This means that the regression model is not a good one (and one should not continue reporting the results)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.987	.924		5.395	.000
	once	-.382	1.307	-.039	-.292	.771
	four	-.877	1.307	-.090	-.671	.504
	seven	-3.334	1.307	-.341	-2.550	.013

a. Dependent Variable: Sperm Count (Millions)

REPORTING: Regression function: Sperm count =4.99-0.38*once-0.88*four-3.33*seven+e

The null hypothesis of the constant is: the constant is 0. The alternative hypothesis is: the constant is not 0. The constant shows the height sperm count if the predictors are all 0. The t-test statistic has a value of 5.40 and is highly significant (p<.001). It shows that baseline sperm count is 4.99 (millions) (given that the model predictors are 0).

The null hypothesis for once is: Sperm count does not differ between those eating no soya meal per week and those eating it once. The alternative hypothesis is: sperm count does differ between no and once. The t-test statistic has a value of -0.292 and the value is not significant (p>.05) indicating that there is no difference in sperm ocunt between no and once. (If it would be significant: Those eating soya meals once a week have a

sperm count that is -.382 lower than that of those eating no soya meals per week)

The null hypothesis for four is: Sperm count does not differ between those eating no soya meal per week and those eating it four times a week. The alternative hypothesis is: sperm count does differ between no and four. The t-test statistic has a value of -0.671 and the value is not significant (p>.05) indicating that there is no difference in sperm ocunt between no and once. (If it would be significant: Those eating soya meals four times a week have a sperm count that is -.671 lower than that of those eating no soya meals per week)

The null hypothesis for seven is: Sperm count does not differ between those eating no soya meal per week and those eating it seven times a week. The alternative hypothesis is: sperm count does differ between no and seven. The t-test statistic has a value of -2.55 and is significant (p<.05) indicating that there is a difference in sperm count between no and seven. Those eating soya meals seven times a week have a sperm count that is -2.55 lower than that of those eating no soya meals per week.

6 Theoretical basis of nonparametric tests

6.1 When to use non-parametric tests:

Non-parametric tests are used when assumptions of parametric tests are not met.

It is not always possible to correct for problems with the distribution of a data set

In these cases we have to use non-parametric tests.

They make fewer assumptions about the type of data on which they can be used.

Non-parmetric tests

Testing of hypotheses about unknown distributions in a population

Application:

If the variables are nominally or ordinally scaled and/ or

If the assumption about the distribution in the population are not sufficient

3 types:

Goodness of fit: Is the hypothetical distribution compatible to an observed distribution?

Test of independence: Are two characteristics independent of each other? (Chi-square-test with cross tables)

Homogeneity test: Are two or multiple samples from the same population? or Is there a significant difference between 2 distributions?

Assessing normality: Kolmogorov-Smirnov test

We don’t have access to the sampling distribution so we usually test the observed data

Central Limit Theorem

If $N > 30$, the sampling distribution is normal anyway

Graphical displays

P-P Plot (or Q-Q plot)
Histogram
Values of Skew/Kurtosis
0 in a normal distribution
Convert to *z* (by dividing value by *SE*)
Kolmogorov-Smirnov Test
Tests if data differ from a normal distribution
Significant = non-Normal data
Non-Significant = Normal data

Data Set: 06_Soya.sav

Run the following analysis including only those individuals that eat none or one soya meal per week (filter cases)

Is the sperm count normally distributed?

Filter: Data → Select cases → If condition is satisfied→ If Soya <=2 → Continue → ok

K-S test: Analyze → Descriptive Statistics → Explore → drag and drop Sperm count (Millions) to Dependent list → Statistics keep default → Plots: keep defaults and select normality plots with tests → continue → options: default → ok

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sperm Count (Millions)	40	100.0%	0	0.0%	40	100.0%

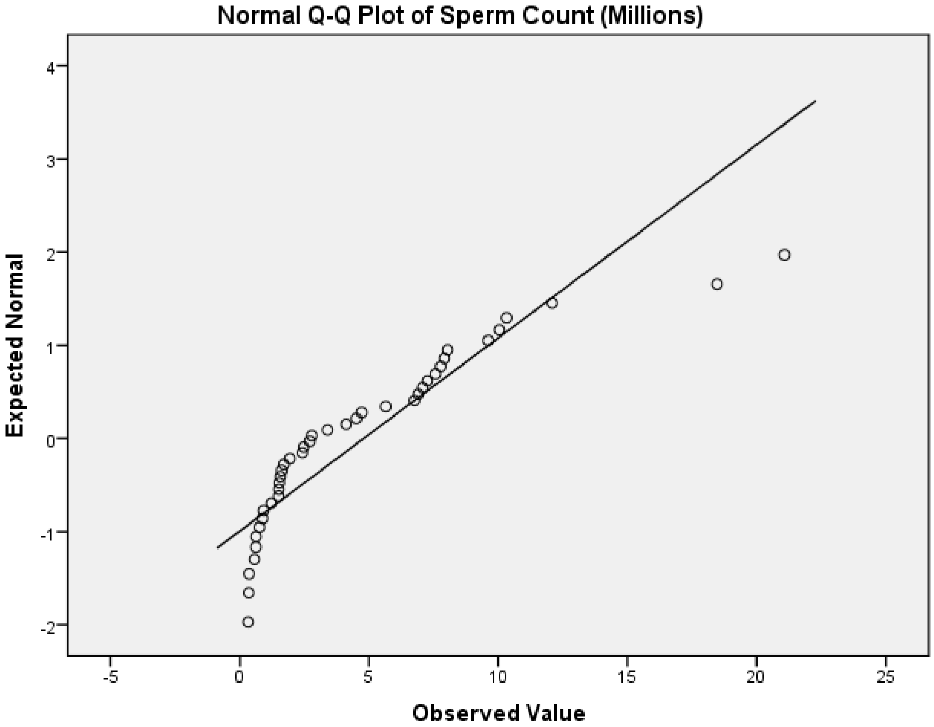
Descriptives

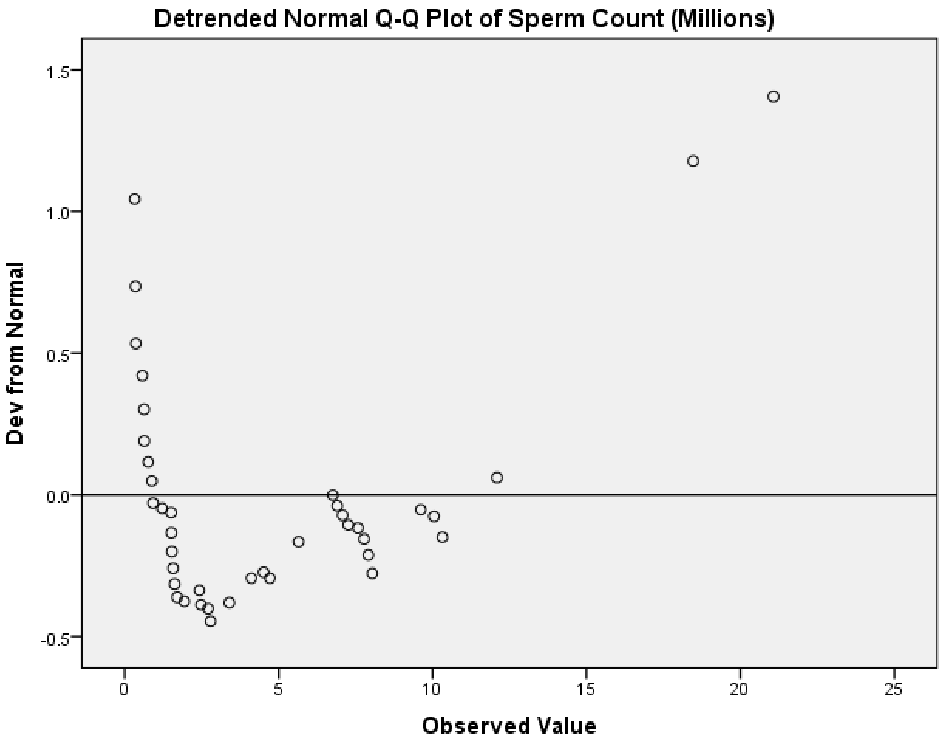
			Statistic	Std. Error
Sperm Count (Millions)	Mean		4.7960	.76269
	95% Confidence Interval for Mean	Lower Bound	3.2533	
		Upper Bound	6.3387	
	5% Trimmed Mean		4.2115	
	Median		2.7502	
	Variance		23.268	
	Std. Deviation		4.82370	
	Minimum		.33	
	Maximum		21.08	
	Range		20.75	
	Interquartile Range		6.21	
	Skewness		1.651	.374
	Kurtosis		3.085	.733

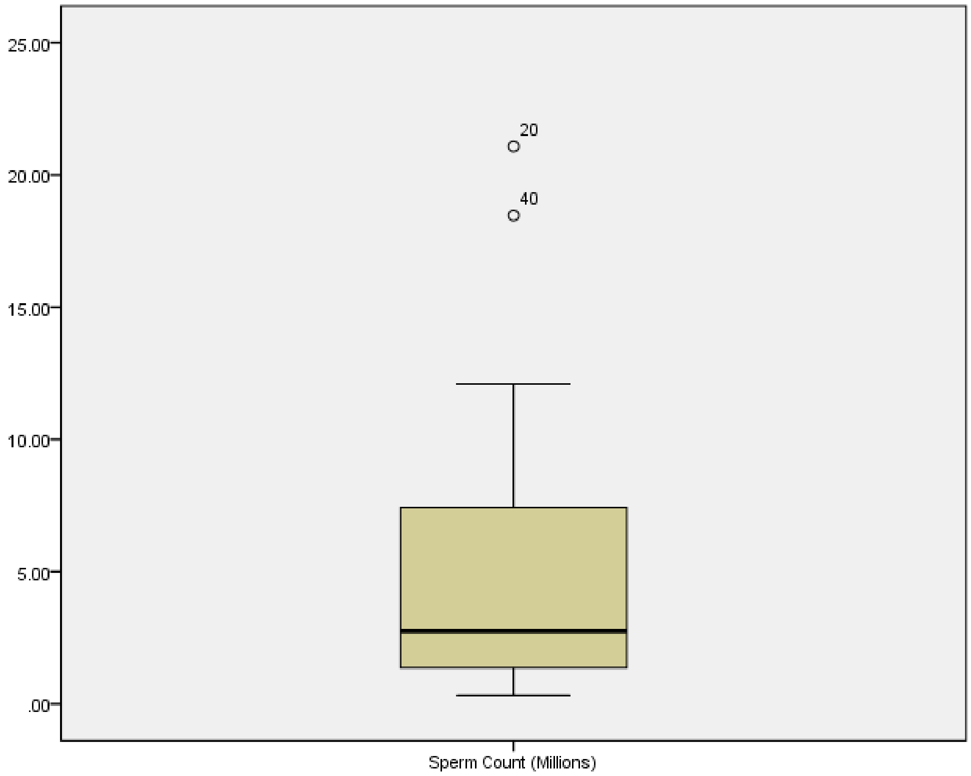
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Sperm Count (Millions)	.186	40	.001	.818	40	.000

a. Lilliefors Significance Correction







The first table produced by SPSS contains descriptive statistics (mean etc.)

The important table is that of the K-S test

This table includes the test statistic itself (D), the degrees of freedom (which should equal the sample size) and the significance value of this test

A significant value (i.e. less than $p<.05$) indicates a deviation from normality

The K-S test for the sperm count is highly significant, indicating that the distribution is not normal

SPSS also produces a normal Q-Q plot for any variables specified, it plots the value you would expect to get if the distribution were normal (expected values) against the values actually seen in the data set (observed values)

If the data are normally distributed, then the observed values (the dots) should fall exactly along the straight line

> ***REPORTING*: The sperm count in million, D (40)=0.19, p<.05 is significantly not normal.**

6.2 Mann-Whitney test:

These tests are the non-parametric equivalent of the independent *t-test.*

Use either to test differences between two conditions in which different participants have been used

***EXERCISE*:** Mann-Whitney test

Data Set: 06_Soya.sav

Keep the filter from the K-S test before

Analyze whether the sperm count in millions is different for people eating no soya meals per week compared to those eating one soya meal per week.

Data set description: p.559

Analyze → Nonpar → Legacy dialogs → 2 independent samples → sperm count as test variable → soya as grouping variable (define groups, group 1: 1 for no soya meals; group 2:2 for one soya meal per week) → test type Mann Whitney U → ok

Ranks

	Number of Soya Meals Per Week	N	Mean Rank	Sum of Ranks
Sperm Count (Millions)	No Soya Meals	20	20.95	419.00
	1 Soya Meal Per Week	20	20.05	401.00
	Total	40		

Test Statistics[a]

			Sperm Count (Millions)
Mann-Whitney U			191.000
Wilcoxon W			401.000
Z			-.243
Asymp. Sig. (2-tailed)			.808
Exact Sig. [2*(1-tailed Sig.)]			.820[b]
Monte Carlo Sig. (2-tailed)	Sig.		.819[c]
	99% Confidence Interval	Lower Bound	.809
		Upper Bound	.829
Monte Carlo Sig. (1-tailed)	Sig.		.409[c]
	99% Confidence Interval	Lower Bound	.397
		Upper Bound	.422

a. Grouping Variable: Number of Soya Meals Per Week

b. Not corrected for ties.

c. Based on 10000 sampled tables with starting seed 2000000.

***REPORTING*: Sperm count per million of individuals eating no soya meals per week (Mn rank 20.95) does not differ significantly from individuals eating one soya meal per week (Mn rank 20.05), U=191, z=-0.243, ns, r=-0.04. The relationship is very small.**

Calculating the effect size

$$r = \frac{Z}{\sqrt{N}}$$

R=-0.243/Wurzel aus 40=-0.04