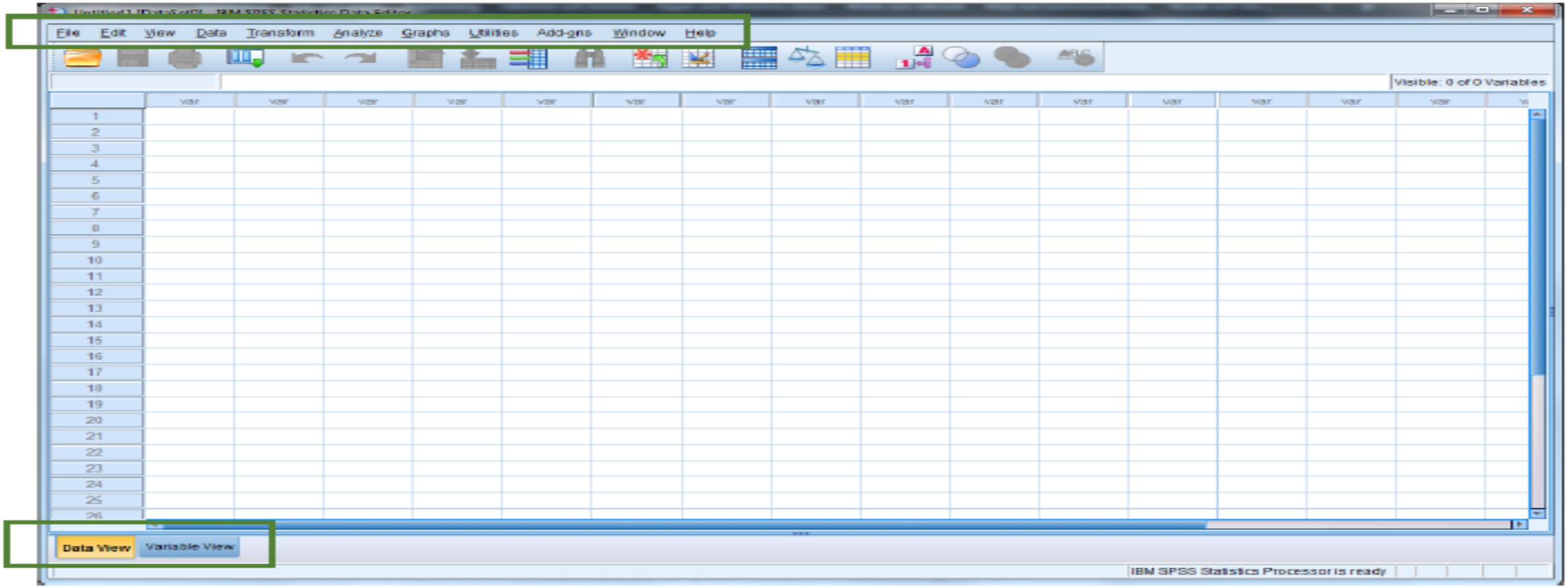
A guide to quantitative methods Almquist, Ashir & Brännström

Outline

- 1.1 General functions
- 1.2 Variable view
- 1.3 Creating a new data set
- 1.4 Data view
- 1.5 Syntax
- 1.6 Output

1.1 GENERAL FUNCTIONS

The Menu bar ("File", "Edit" and so on) is located in the upper area.



In the lower left corner, two tabs are available: Data View and Variable View. When you start SPSS, Variable View is default.

File types in SPSS

File types

SPSS uses three types of files with different functions and extensions:

Type	Extension	Content
Data set	.sav	Data and variables
Syntax	.sps	Commands and comments
Output	.spv	Results

Options

The SPSS menu works similar to the menus in many other programs, such as Word or Excel. Some useful options are listed below:

Option	Description
Open a file	Go to File\Open and choose Data, Syntax or Output. Browse your hard drive to locate the file. Then click on Open.
Save a file	Go to File\Save As. Type in a descriptive name and then click Save. You can also choose to save only some of the variables into the new data set by clicking Variables (before saving) and then ticking the boxes next to the variables you want to keep.
Overwrite a file	Go to File\Save.
Import a data set	Go to File\Open\Data and choose the desired format next to Files of type. Browse the file and click on Open.

1.2 Variable view

In Variable View, different columns are displayed. Each line corresponds to a variable. A variable is simply a quantity of something, which varies and can be measured, such as height, weight, number of children, educational level, gender and so forth.

Column	Function
Name	Name of the variable. It is your own choice, but make it understandable and do not use numbers or symbols as the first letter since SPSS will not accept it. Moreover, you cannot use spaces in the name. For example: "edu_level"
Type	Indicates the variable type. The most common is Numeric (only accepts numerical data, for example age or number of children) and String (also accepts letters, e.g. for qualitative questions). Typically, all responses in a questionnaire are transformed into numbers. For example: "Man"=0 and "Woman"=1, or "Non-smoker"=1, "Ex-smoker"=2 and "Current smoker"=3.

Width	Corresponds to the number of characters that is allowed to be typed in the data cell. Default for numerical and string variables is 8, which only needs to be altered if you want to type in long strings of numbers or whole sentences.
Decimals	Default is 2 for numerical variables and will automatically be displayed as .00 in the data view, if not otherwise specified.
Label	The description of the variable. Use the question that the variable is based upon or something else accurately describing the variable. For example: "What is your highest level of education?"
Values	Here you can add labels to each response alternative. For example: For the variable gender, "Men" are coded as 0 and "Women" are coded as 1. Through the option Values you tell SPSS to label each number according to the correct response. Next to Value (below Value Labels), type in "0" and next to Label, type in "Men". Then click Add. Next to Value (below Value Labels), type in "1" and next to Label, type in "Women". Then click Add.
Missing	By default, missing values will be coded as "." (dot) for numerical variables in the data set. For missing values in String variables, cells will be left blank.

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											

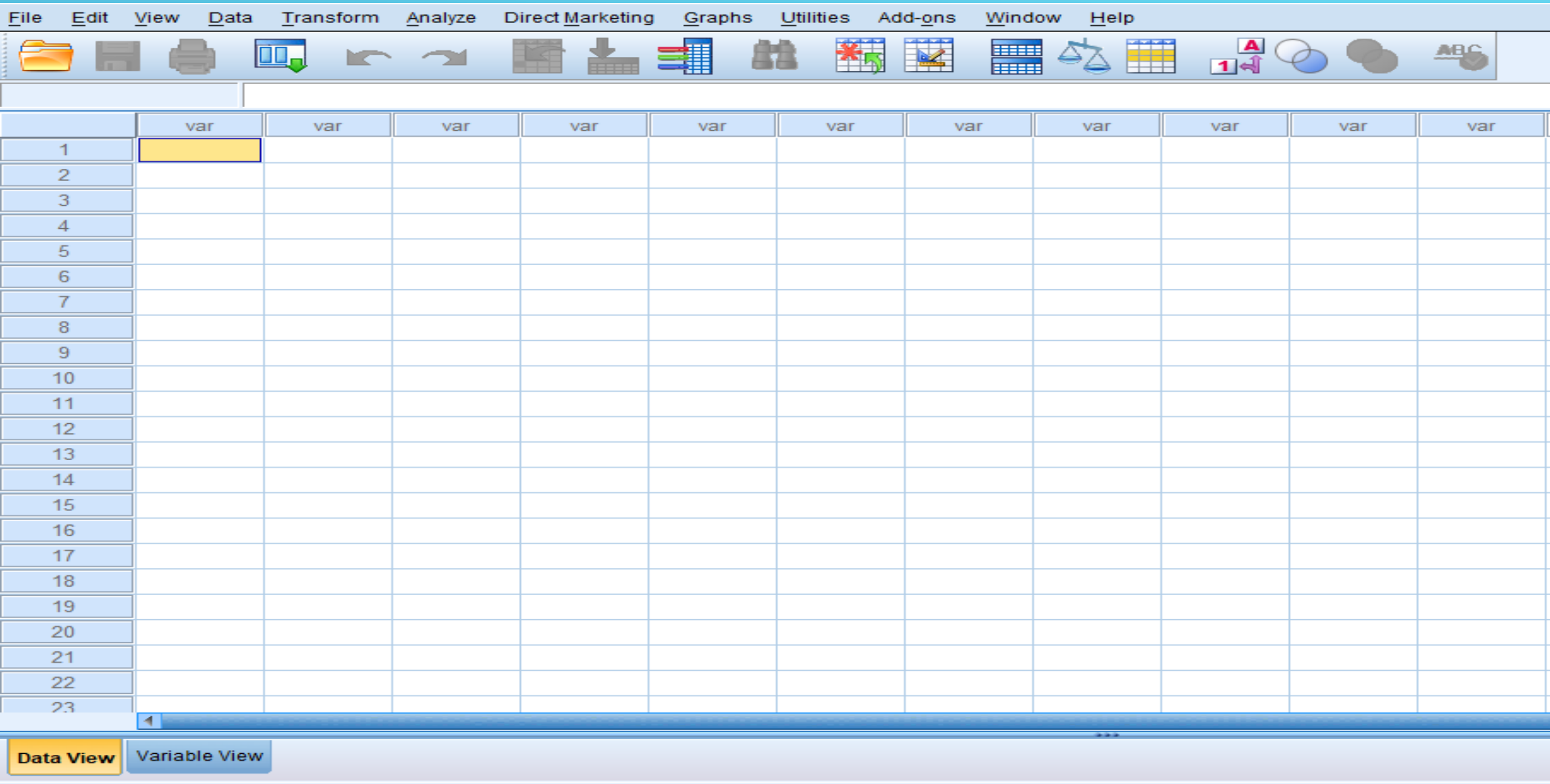
Data View | Variable View

1.3 Creating a new data set

If you have a questionnaire, you can easily create the corresponding data structure in Variable View in SPSS. For example:

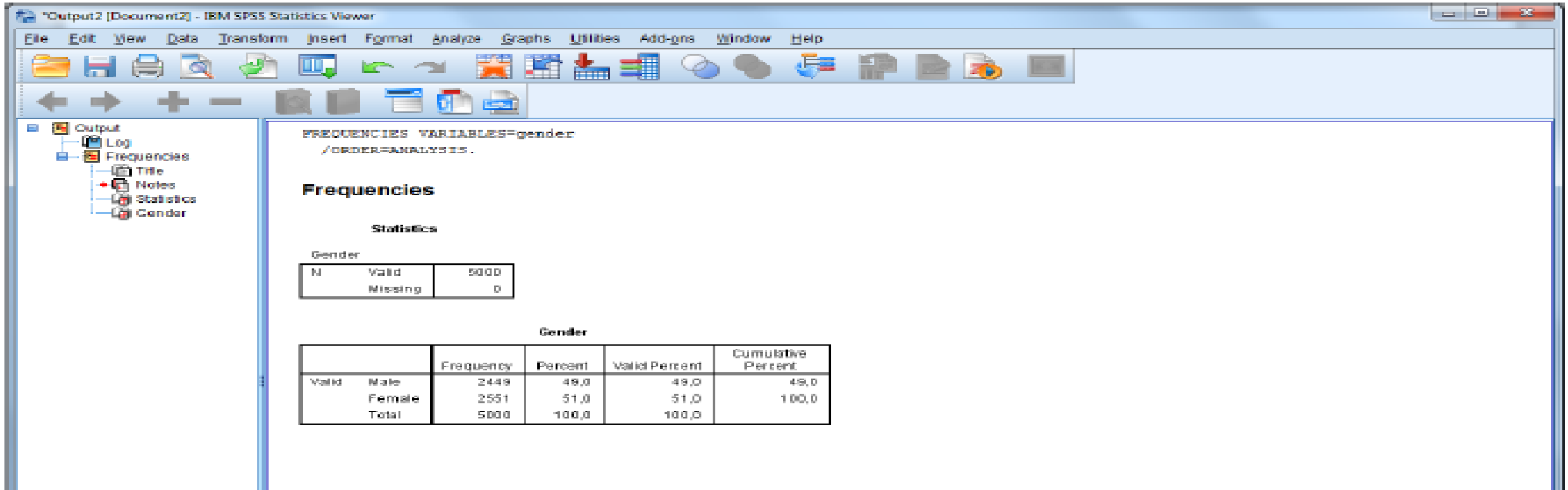
Name	Type	Width	Decimals	Label	Values	Missing
gender	Numeric	1	0	Gender	0=Man 1=Woman	None
srh	Numeric	1	0	Self-rated health	1=Poor 2=Fair 3=Good 4=Excellent	None
income	Numeric	10	0	Disposable income	None	None

Data view


File
Edit
View
Data
Transform
Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Window
Help
var
Data View
Variable View

1.6 Output

Everything you order in SPSS (e.g. graphs, tables, or analyses) ends up in a window called Output. In the area to the left, all the different steps are listed. It is possible to collapse specific steps by clicking on the box with the minus sign (and expand it again by clicking on the same box, now with a plus sign). In the area to the right, your actual output is shown. First, you see the syntax for what you have ordered, and then you get the tables or graphs related to the specific command.



3. DESCRIPTIVE STATISTICS

When we know about the measurement scale and the distribution of the variables in our data set, we can decide on how to best describe our variables. In this type of exploratory data analysis, we use a set of tables and graphs as well as measures of central tendency and variation. Here, we will address the following types of description:

Outline
3.1 Tables
3.2 Graphs
3.3 Measures of central tendency
3.4 Measures of variation

Type of variable	
Categorical (nominal/ordinal)	Frequency table
	Cross table
	Bar chart
	Pie chart
	Mode
Continuous (ratio/interval)	Histogram
	Scatterplot
	Mean (if normal distribution)
	Median (if skewed distribution)
	Min
	Max
	Range
	Standard deviation

3.1 Tables

Tables are useful if one wants to see the distribution of values for categorical (nominal or ordinal) variables. Here, we will discuss frequency tables and cross tables.

A frequency table is a simple but very useful description of one variable and gives us both the frequency and various types of percentages of individuals with the different values.

Column	Content
Frequency	The number of individuals in the different categories.
Percent	The distribution of percent also taking into consideration any missing information. This means that if some individuals would have missing information about gender, the percentages in this column would be dependent upon that.
Valid Percent	Same as Percent but does not take missing into account. This column is what we primarily focus on.
Cumulative percent	Adds the percentages from top to bottom.

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Man	7	46,7	46,7	46,7
	Woman	8	53,3	53,3	100,0
	Total	15	100,0	100,0	

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Man	7	46,7	50,0	50,0
	Woman	7	46,7	50,0	100,0
	Total	14	93,3	100,0	
Missing	System	1	6,7		
Total		15	100,0		

			Gender		Total
			Man	Woman	
Health	Good	Count	2	3	5
		% within Gender	28,6%	37,5%	33,3%
	Poor	Count	5	5	10
		% within Gender	71,4%	62,5%	66,7%
Total		Count	7	8	15
		% within Gender	100,0%	100,0%	100,0%

			Gender		Total
			Man	Woman	
Health	Good	Count	2	3	5
		% within Health	40,0%	60,0%	100,0%
	Poor	Count	5	5	10
		% within Health	50,0%	50,0%	100,0%
Total		Count	7	8	15
		% within Health	46,7%	53,3%	100,0%

3.3 Measures of central tendency

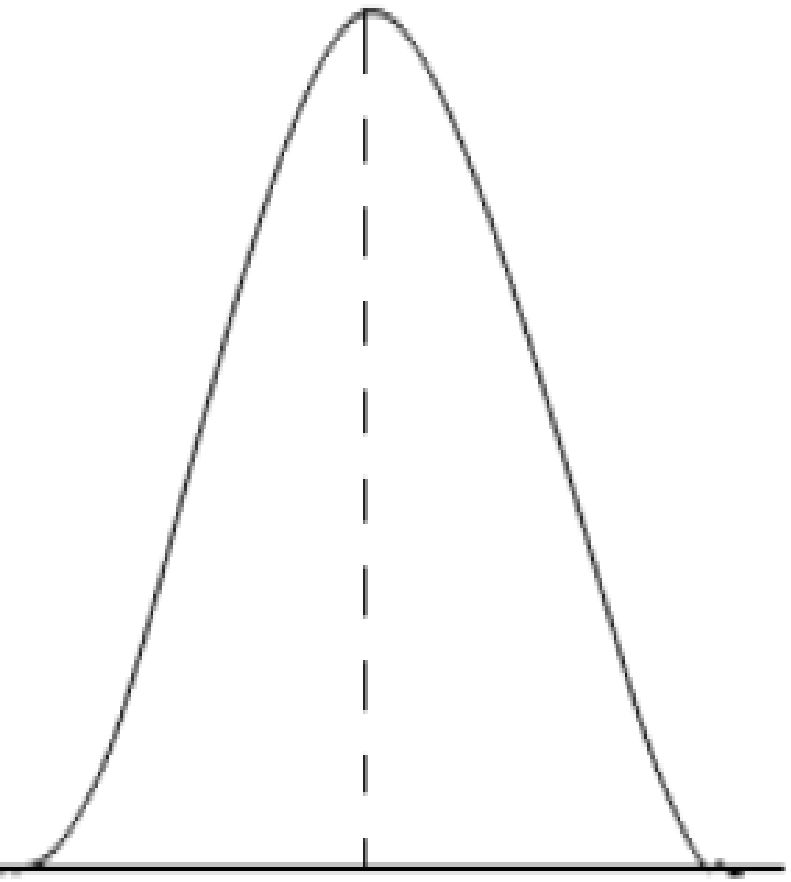
Central tendency can be defined as measures of the location of the middle in a distribution. The most common types of central tendency are:

Measure	Definition
Mean	The average value
Median	The value in the absolute middle
Mode	The most frequently occurring value

The mean is perhaps the most commonly used type of central tendency and we get it by dividing the sum of all values by the number of values.

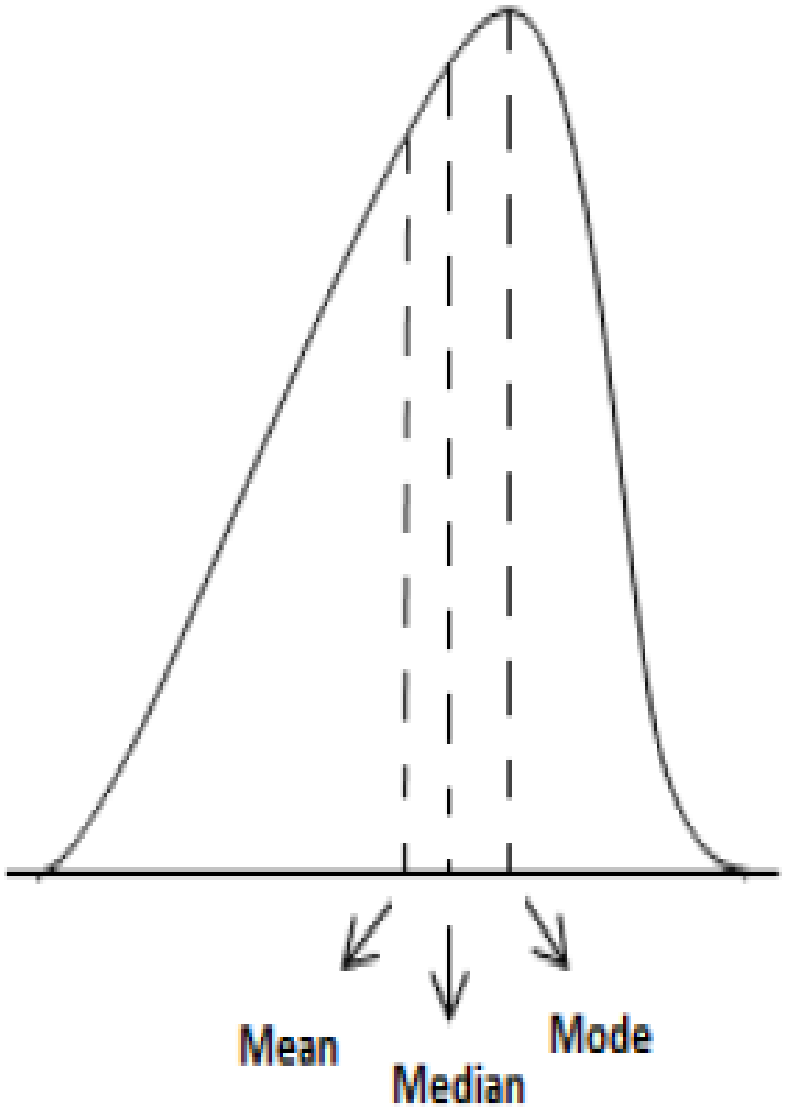
Scale	Type	Central tendency
Nominal	Categorical	Mode
Ordinal		
Ratio	Continuous	Normal distribution: Mean; Skewed distribution: Median
Interval		

Normal distribution

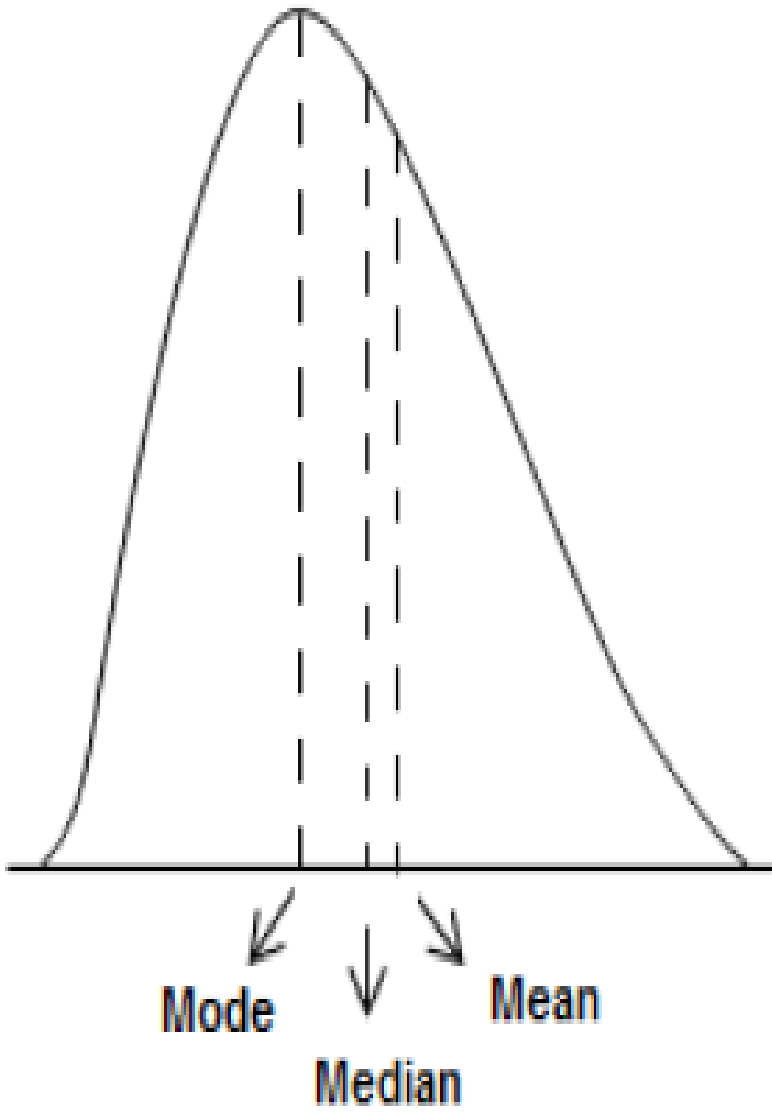


Mean
Median
Mode

Negatively skewed



Positively skewed



3.4 Measures of variation

Besides the mean, the median and the mode, we may use some measures of variation to describe our variables further. Here are some of the most common measures of variation:

Measure	Definition
Min	The lowest value
Max	The highest value
Range	The difference between the lowest and highest value
Standard deviation	The dispersion of values from the mean

4. PRODUCING DESCRIPTIVE STATISTICS IN SPSS

This part of the guide will describe how to use SPSS to produce the various tables and graphs as well as measures of central tendency and variation previously discussed.

Outline
4.1 Descriptives
4.2 Frequency table
4.3 Cross table
4.4 Bar chart
4.5 Pie chart
4.6 Histogram
4.7 Scatterplot
4.8 Edit graphs

4.1 Descriptives

Quick facts	
Number of variables	At least one
Scale of variable(s)	Continuous (ratio/interval) or ordinal

Types of statistic	
Mean	Mean value
Sum	Sum of all values
Stddev	Standard deviation
Min	Minimum (smallest) observed value
Max	Maximum (largest) observed value
Variance	Variance
Range	The difference between the minimum value and the maximum value
Semean	Standard error of the mean
Kurtosis	Kurtosis and standard error of kurtosis
Skewness	Skewness and standard error of skewness

Descriptives: Function

1. Go to the Menu bar, choose Analyze\Descriptive Statistics\Descriptives.
2. This will open up a new window called Descriptives.
3. A small window will open, where you see two boxes.
4. In the left box, all your variables are displayed. Here you choose the variable(s) you want to get the measures of central tendency and/or variation for.
   In other words, if you want to, you can choose several variables here, and SPSS will produce descriptives for all of them.
5. Click on the small arrow between the boxes to transfer the variable(s) to the box to the right.
6. Click on Options.
7. Tick the boxes for the measures you want to have.
8. Click on Continue to close the small window.
9. Click on OK to close the Descriptives window in order to get the results in your Output window.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Age in years	5000	61	18	79	46,94	17,703
Valid N (listwise)	5000					

4.2 Frequency table

Quick facts	
Number of variables	At least one (one table will be produced for each variable)
Scale of variable(s)	Categorical (nominal/ordinal)

1. Go to the Menu bar, choose Analyze\Descriptive Statistics\Frequencies.
2. This will open up a new window called Frequencies.
3. A small window will open, where you see two boxes. In the left box, all your variables are displayed.
4. In the left box, you choose the variable(s) you want to get a frequency table for.
   In other words, you may choose several variables here, and SPSS will produce frequency tables for all of them.
5. Click on the small arrow between the boxes to transfer the variable(s) to the box to the right.
6. To order statistics for your variable, click on Statistics. In the new window called Frequencies: Statistics tick the boxes for the measures you want SPSS to show. Click on Continue.
7. To order a graph, click on Charts. In the new window called Frequencies: Charts tick the box for the graph you want SPSS to show. Also make sure to tick the box Percentages. Click on Continue.
8. Click on OK.

Building type

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Single family	2246	44,9	44,9	44,9
	Multiple family	1554	31,1	31,1	76,0
	Townhouse	908	18,2	18,2	94,2
	Mobile home	292	5,8	5,8	100,0
	Total	5000	100,0	100,0	

4.4 Bar chart

Quick facts	
Number of variables	Simple bar chart: one (it is possible to split/panel the chart by a second variable) Clustered bar chart: two (it is possible to split/panel the chart by a third variable)
Scale of variable(s)	Categorical (nominal/ordinal)

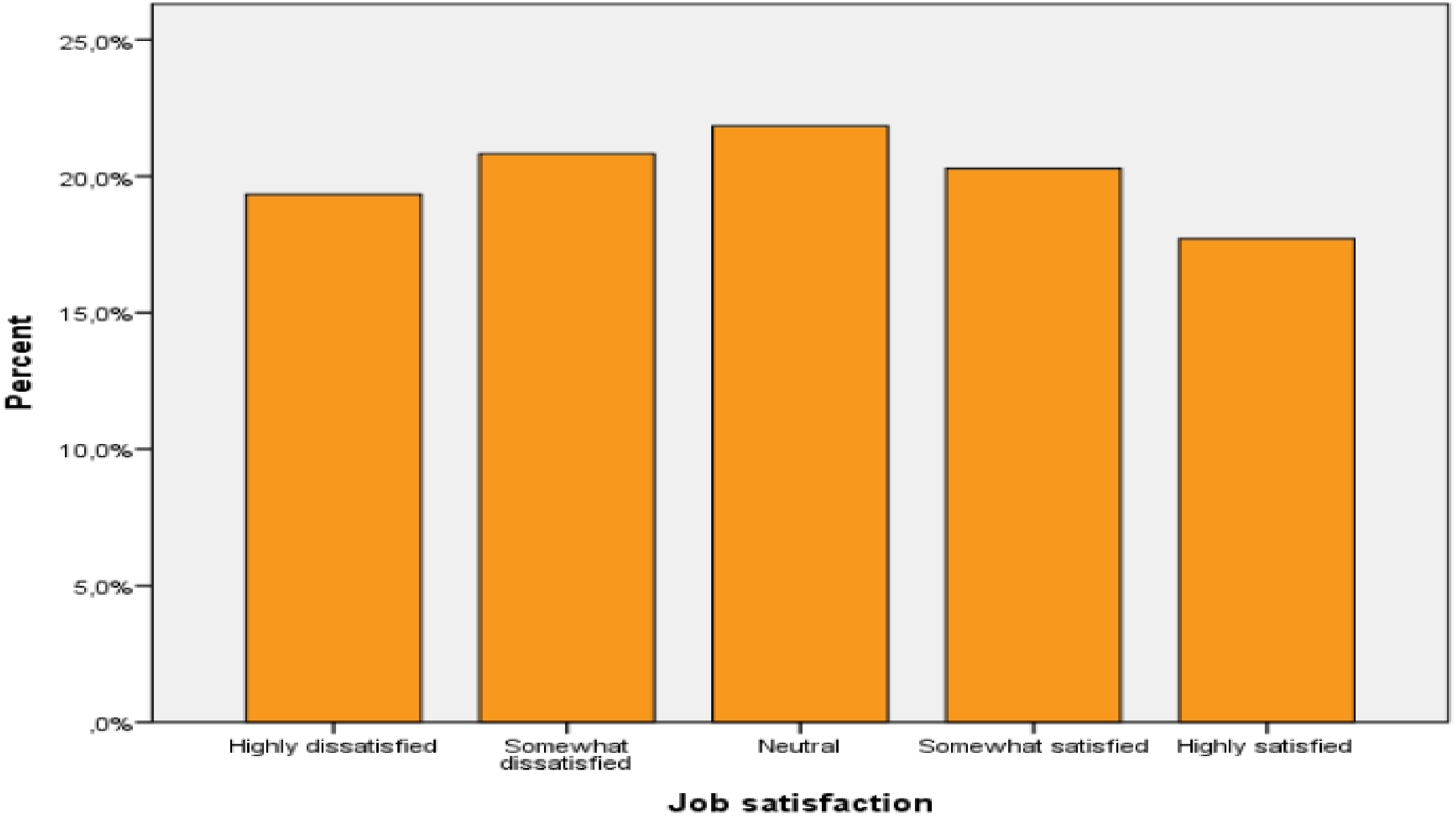
Bar chart: Function

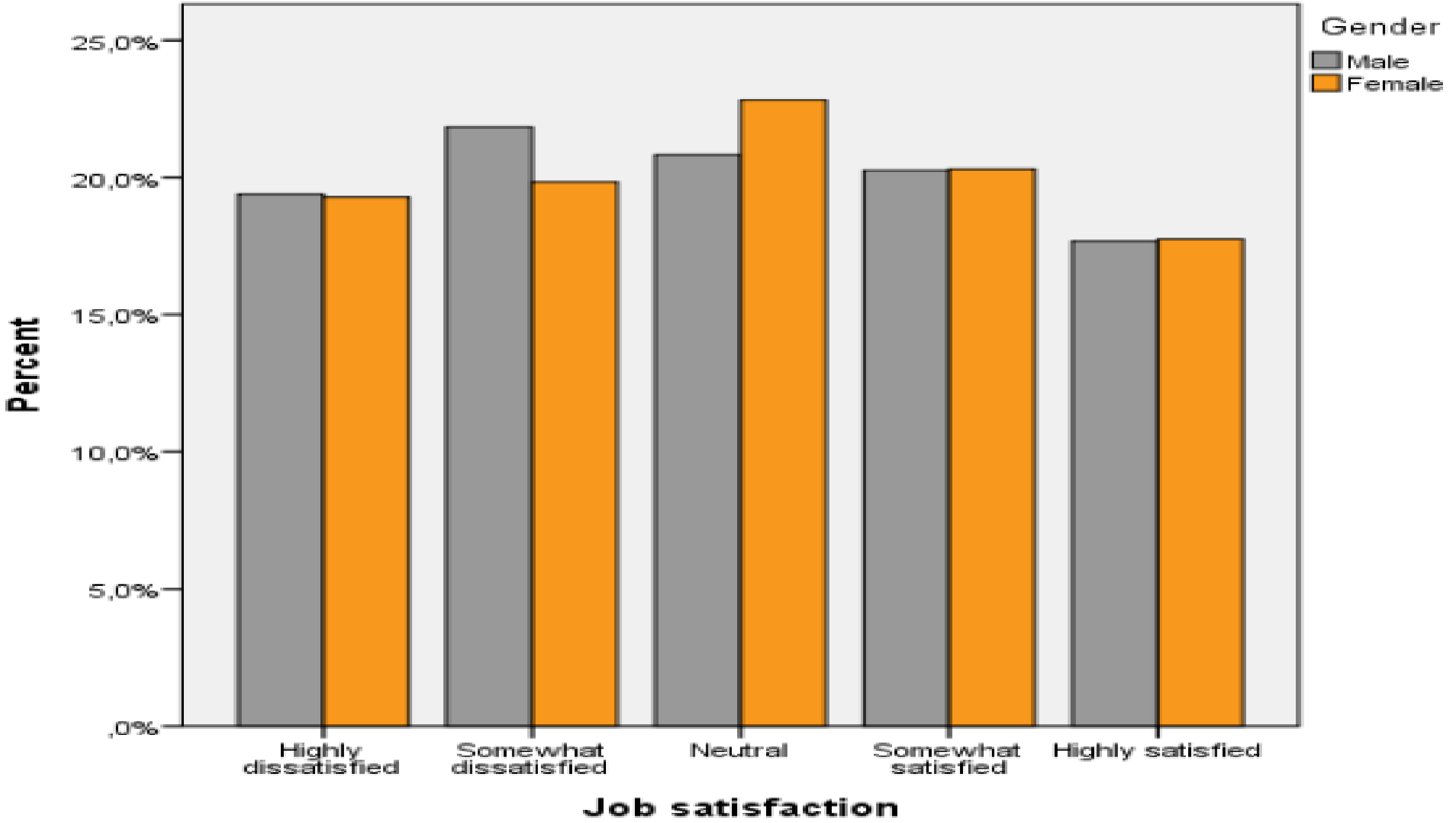
Simple bar chart

1. Go to the Menu bar, choose Graphs\Legacy Dialogs\Bar.
2. A small window will open, where you click on Define.
3. A new window called Define Simple Bar: Summaries for Groups of Cases will open.
4. Tick the option % of cases.
5. In the left box, all your variables are displayed. Here, you select the variable you want SPSS to show a bar chart for. Click on the arrow next to Category Axis.
6. Click on OK to close the Define Simple Bar: Summaries for Groups of Cases window in order to get the results in your Output window.

Clustered bar chart

1. Go to the Menu bar, choose Graphs\Legacy Dialogs\Bar.
2. A small window will open, where you choose the option Clustered.
3. Click on Define.
4. A new window called Define Clustered Bar: Summaries for Groups of Cases will open.
5. Tick the option % of cases.
6. In the left box, all your variables are displayed. Here, you select the variable you want SPSS to show a bar chart for. Click on the arrow next to Category Axis.
7. In the left box, you select the variable you want the previous variable to be clustered by. Click on the arrow next to Define Clusters by.
8. Click on OK to close the Define Simple Bar: Summaries for Groups of Cases window in order to get the results in your Output window.


25,0%
20,0%
15,0%
10,0%
5,0%
,0%
Percent
Highly dissatisfied
Somewhat dissatisfied
Neutral
Somewhat satisfied
Highly satisfied
Job satisfaction

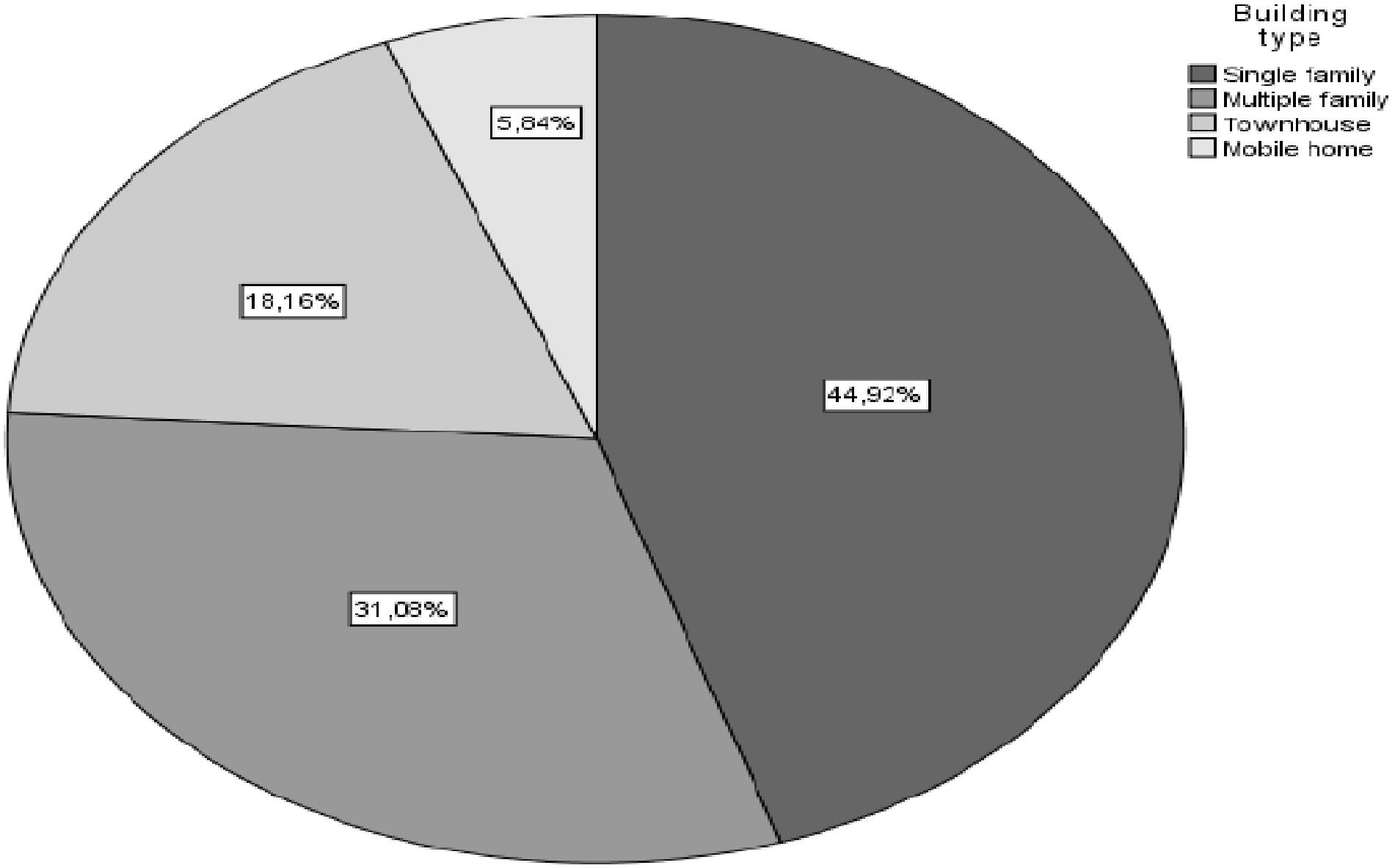

Gender
Male
Female
25,0%
20,0%
15,0%
10,0%
5,0%
,0%
Percent
Highly dissatisfied
Somewhat dissatisfied
Neutral
Somewhat satisfied
Highly satisfied
Job satisfaction

4.5 Pie chart

Quick facts	
Number of variables	One (it is possible to split/panel the chart by a second variable)
Scale of variable(s)	Categorical (nominal/ordinal)

Pie chart: Function

1. Go to the Menu bar, choose Graphs\Legacy Dialogs\Pie.
2. A small window will open, where you click on Define.
3. A new window called Define Pie: Summaries for Groups of Cases will open.
4. Tick the option % of cases.
5. In the left box, all your variables are displayed. Here, you select the variable you want SPSS to show a pie chart for. Click on the arrow next to Define slices by.
6. Click on OK to close the Define Pie: Summaries for Groups of Cases window in order to get the results in your Output window.

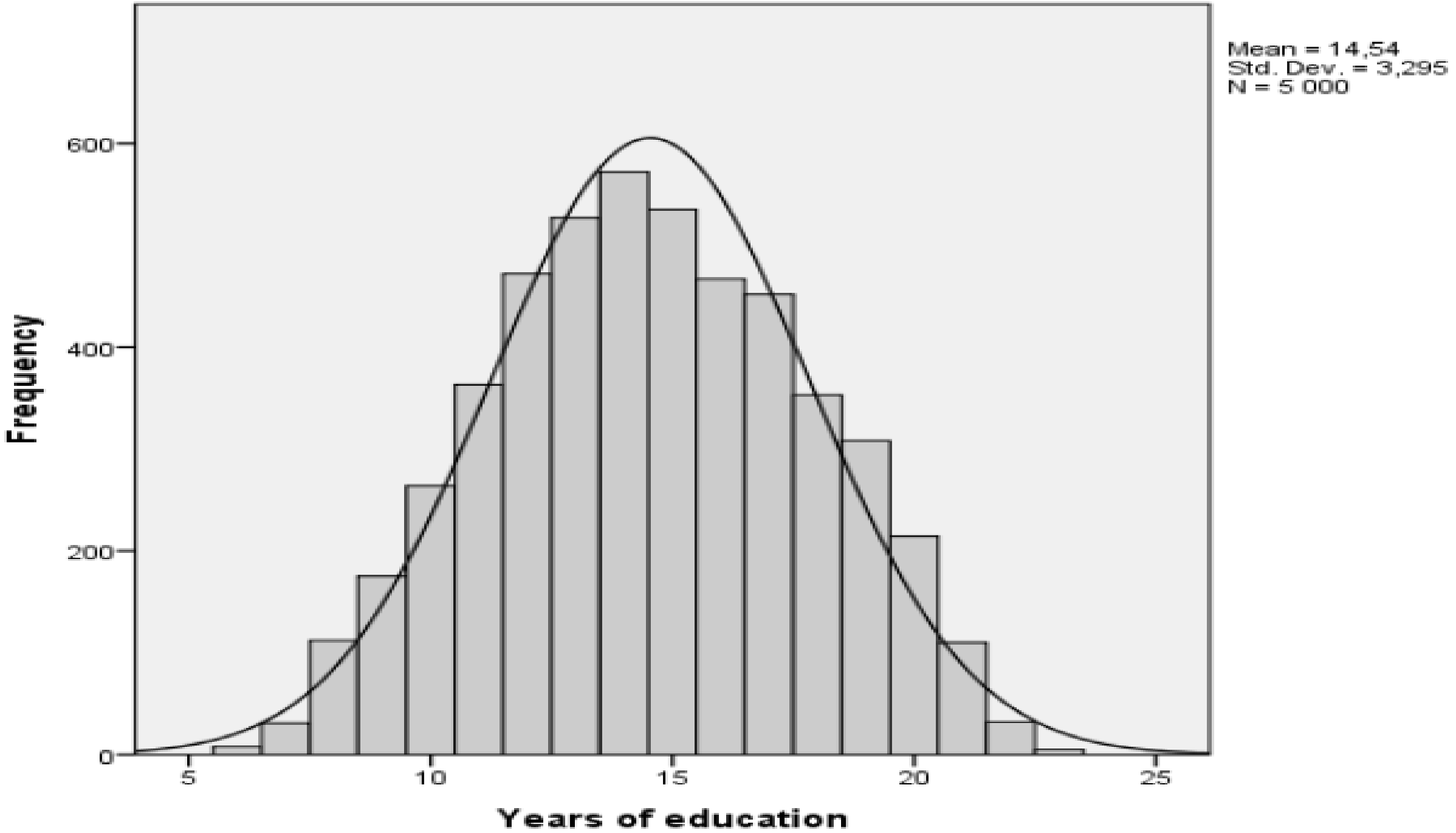

Building type
Single family
Multiple family
Townhouse
Mobile home
44,92%
31,08%
18,16%
5,84%

4.6 Histogram

Quick facts	
Number of variables	One (it is possible to split/panel the chart by a second variable)
Scale of variable(s)	Continuous (ratio/interval)

Histogram: Function

1. Go to the Menu bar, choose Graphs\Legacy Dialogs\Histogram.
2. A new window called Histogram will open.
3. In the left box, all your variables are displayed. Here, you select the variable you want SPSS to show a histogram for. Click on the arrow next to Variable.
4. Tick the option Display normal curve to include a normal curve fitted to the data.
5. Click on OK to close the Histogram window in order to get the results in your Output window.

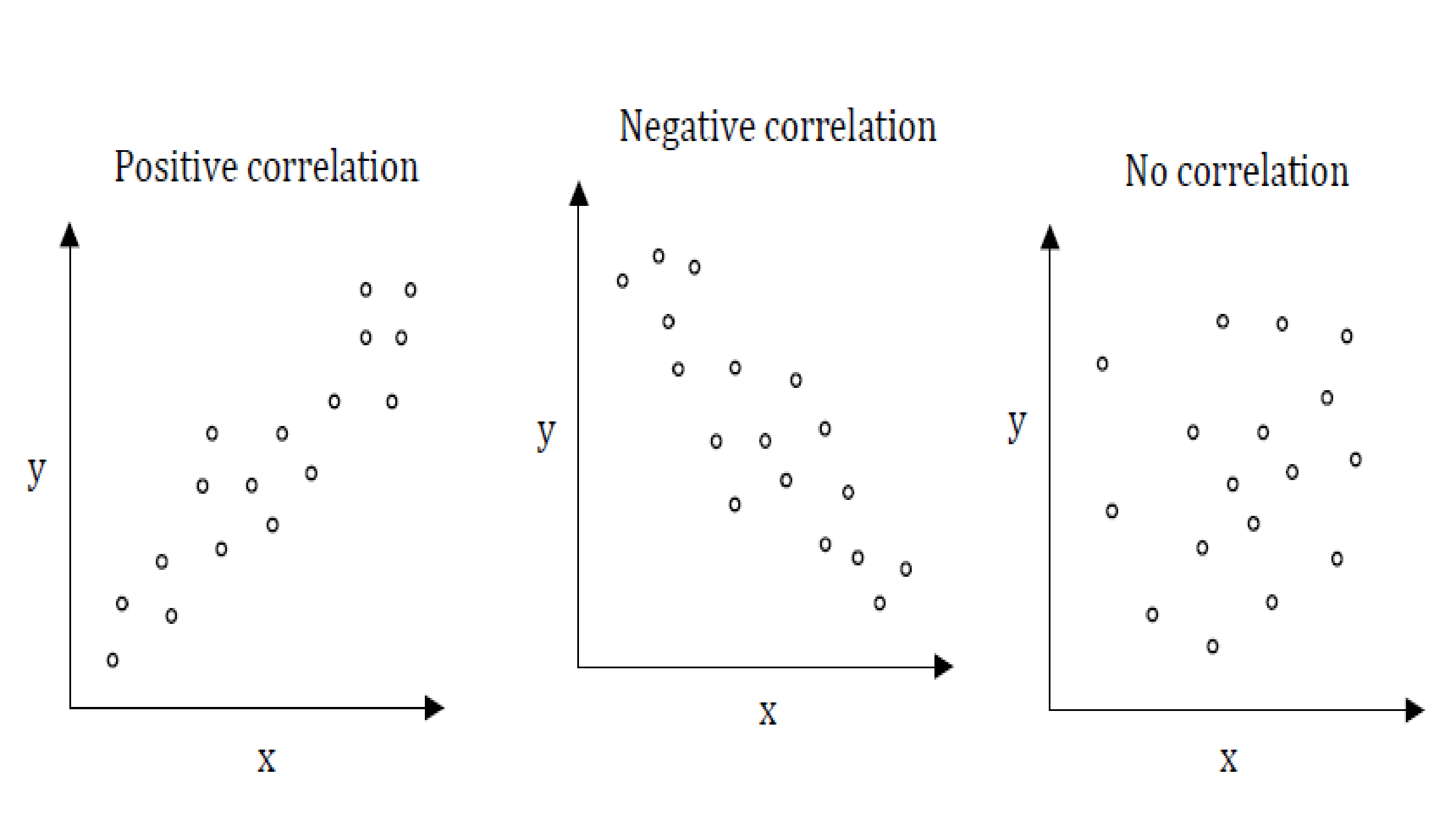

Mean = 14,54
Std. Dev. = 3,295
N = 5 000
Frequency
600
400
200
0
5
10
15
20
25
Years of education

4.7 Scatterplot

Quick facts	
Number of variables	Two (it is possible to split/panel the chart by a third variable)
Scale of variable(s)	Continuous (ratio/interval)

Scatterplot: Function

1. Go to the Menu bar, choose Graphs\Legacy Dialogs\Scatter/Dot.
2. A small window will open, where you click on Define.
3. A new window called Simple Scatterplot will open.
4. In the left box, all your variables are displayed. Move the variables you want to use to the Y Axis and X Axis by highlighting them and clicking on the arrow next to the axis you want them on.
5. Click on OK to close the Simple Scatterplot window in order to get the results in your Output window.


Positive correlation
y
x
Negative correlation
y
x
No correlation
y
x

6. STATISTICAL SIGNIFICANCE

Outline
6.1 Hypothesis testing
6.2 P-values
6.3 Confidence intervals
6.4 Discussion

Hypotheses

Let us return to the matter of statistical significance: what is it really? Well, for example, if we find that cats are smarter than dogs, we want to know whether this difference is "real". Hypothesis testing is how we may answer that question. We start by converting the question into two hypotheses:

Hypotheses		
Null hypothesis	(H_0)	There is no difference
Alternative hypothesis	(H_1)	There is a difference

Outcomes

There are two possible outcomes of hypothesis testing:

Outcomes of hypothesis testing	
Reject H_0 in favour of H_1	Suggests that the alternative hypothesis *may* be true (but it does not prove it)
Do not reject H_0	Suggests that there is not sufficient evidence against H_0 in favour of H_1 (but it does not prove that the null hypothesis is true)

Note that we are never able to decide from hypothesis testing that we should reject or accept H_1. However, rejecting H_0 may lead us to suggest that H_1 might be accepted.

Type I and type II errors

		Conclusion	
		Reject H_0 in favour of H_1	**Do not reject H_0**
"Truth"	**H_0**	*Type 1 error*	*Right decision*
	H_1	*Right decision*	*Type II error*

Type I errors are generally considered to be more serious that type II errors. Type II errors are often due to small sample sizes.

6.2 P-values

The probability value – or p-value – helps us decide whether or not the null hypothesis should be rejected. There are some common misunderstandings about p-values:

Significance levels		
$p<0.05$	Statistically significant at the 5 % level	*
$p<0.01$	Statistically significant at the 1 % level	**
$p<0.001$	Statistically significant at the 0.1 % level	***

7. COMPARE MEANS

Outline
7.1 T-test: independent samples
7.2 T-test: paired samples
7.3 One-way ANOVA

7.1 T-test: independent samples

Quick facts	
Number of variables	One independent (x) One dependent (y)
Scale of variable(s)	Independent: categorical with two values (binary) Dependent: continuous (ratio/interval)

The independent samples t-test is a method for comparing the mean of one variable between two (unrelated) groups. For example, you may want to see if the income salary of teachers differs between men and women, or if the score of a cognitive test differs between children who have parents with low versus high education.

Mean income salary among men

Mean income salary among women

Assumptions

Checklist	
Continuous dependent variable	Your dependent variable should be continuous (i.e. interval/ratio). For example: Income, height, weight, number of years of schooling, and so on. Although they are not really continuous, it is still very common to use ratings as continuous variables, such as: "How satisfied with your income are you?" (on a scale 1-10) or "To what extent do you agree with the previous statement?" (on a scale 1-5).
Two unrelated categories in the independent variable	Your independent variable should be categorical and consist of only two groups. Unrelated means that the two groups should be mutually excluded: no individual can be in both groups. For example: men vs. women, employed vs. unemployed, low-income earner vs. high-income earner, and so on.
No outliers	An outlier is an extreme (low or high) value. For example, if most individuals have a test score between 40 and 60, but one individual has a score of 96 or another individual has a score of 1, this will distort the test.

T-test: independent samples: Function

1. Go to the Menu bar, choose Analyze\Compare Means\Independent-Samples T Test.
2. In the left box, all your variables are displayed. You choose the variable you want to have as your dependent variable and transfer it to the box called Test Variable(s).
3. Then you choose the variable you want as your independent variable and transfer it to the box called Grouping Variable.
4. Click on Define Groups…
5. Specify which values the two categories in the independent variable have.
6. Click on Continue.
7. Click on OK.

Output/Step 1

The table called Group Statistics sums the statistics for the variables in the t-test. Here, it can be interesting to look at each groups' mean value. As can be seen, those who are not retired have a slightly higher mean value for hours spent watching TV: 19.89 compared to 18.21 for those who are retired.

Group Statistics

	Retired	N	Mean	Std. Deviation	Std. Error Mean
Hours spent watching TV last week	No	4268	19,89	4,745	,073
	Yes	732	18,21	7,018	,259

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
Hours spent watching TV last week	Equal variances assumed	108,569	,000	8,170	4998	,000	1,680	,206	1,277	2,083
	Equal variances not assumed			6,238	849,229	,000	1,680	,269	1,152	2,209

7.2 T-test: paired samples

Quick facts	
Number of variables	Two (reflecting repeated measurement points)
Scale of variable(s)	Continuous (ratio/interval)

Happiness score before summer vacation

Happiness score after summer vacation

Assumptions

Checklist	
Continuous variables	Your two variables should be continuous (i.e. interval/ratio). For example: Income, height, weight, number of years of schooling, and so on. Although they are not really continuous, it is still very common to use ratings as continuous variables, such as: "How satisfied with your income are you?" (on a scale 1-10) or "To what extent do you agree with the previous statement?" (on a scale 1-5).
Two measurement points	Your two variables should reflect one single phenomenon, but this phenomenon is measured at two different time points for each individual.
Normal distribution	Both variables need to be approximately normally distributed. Use a histogram to check (see Section 4.6).
No outliers in the comparison between the two measurement points	For example, if one individual has an extremely low value at the first measurement point and an extremely high value at the second measurement point (or vice versa), this will distort the test. Use a scatterplot to check (see Section 4.7).

T-test: paired samples: Function

1. Go to the Menu bar, choose Analyze\Compare Means\Paired Samples T Test.
2. In the left box, all your variables are displayed. You choose the variable you want to have as your dependent variable and transfer it to the box called Paired variables.
3. Then you choose the variable you want as your independent variable and transfer it to the box called Paired variable.
4. Click on OK.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Unemployment days in 2003	8,12	4971	23,286	,330
	Unemployment days in 2005	11,31	4971	44,103	,626

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Unemployment days in 2003 - Unemployment days in 2005	-3,190	43,017	,610	-4,386	-1,994	-5,228	4970	,000

7.3 One-way ANOVA

Quick facts	
Number of variables	One independent (x) One dependent (y)
Scale of variable(s)	Independent: categorical (nominal/ordinal) Dependent: continuous (ratio/interval)

Mean number of ice cones per week during May in Swedish children ages 5-10

Mean number of ice cones per week during June in Swedish children ages 5-10

Mean number of ice cones per week during July in Swedish children ages 5-10

Assumptions

Checklist	
Continuous dependent variable	Your dependent variable should be continuous (i.e. interval/ratio). For example: Income, height, weight, number of years of schooling, and so on. Although they are not really continuous, it is still very common to use ratings as continuous variables, such as: "How satisfied with your income are you?" (on a scale 1-10) or "To what extent do you agree with the previous statement?" (on a scale 1-5).
Two or more unrelated categories in the independent variable	Your independent variable should be categorical (i.e. nominal or ordinal) and consist of two or more groups. Unrelated means that the groups should be mutually excluded: no individual can be in more than one of the groups. For example: low vs. medium vs. high educational level; liberal vs. conservative vs. socialist political views; or poor vs. fair, vs. good vs. excellent health; and so on.
No outliers	An outlier is an extreme (low or high) value. For example, if most individuals have a test score between 40 and 60, but one individual has a score of 96 or another individual has a score of 1, this will distort the test.

One-way ANOVA: Function

1. Go to the Menu bar, choose Analyze\Compare Means\One-way ANOVA.
2. In the left box, all your variables are displayed. You choose the variable you want to have as your dependent variable and transfer it to the box called Dependent list.
3. You also choose the variable you want as your independent variable and transfer it to the box called Factor.
4. Go to the box Option. Tick the boxes called Descriptive, Homogeneity of variance test, Brown-Forsythe, Welch and Means Plot.
5. Click on Continue and then on OK.

Descriptives

Household income in thousands

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Energy bar	1554	46,7651	39,96608	1,01383	44,7765	48,7537	9,00	424,00
Oatmeal	1549	60,7366	61,95588	1,57419	57,6488	63,8244	9,00	780,00
Cereal	1897	57,1687	60,04177	1,37854	54,4651	59,8723	9,00	1073,00
Total	5000	55,0406	55,54475	,78552	53,5006	56,5806	9,00	1073,00

Test of Homogeneity of Variances

Household income in thousands

Levene Statistic	df1	df2	Sig.
46,458	2	4997	,000

ANOVA

Household income in thousands

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	165270,934	2	82635,467	27,064	,000
Within Groups	15257739,82	4997	3053,380		
Total	15423010,76	4999			

Robust Tests of Equality of Means

Household income in thousands

	Statistic[a]	df1	df2	Sig.
Welch	35,404	2	3209,671	,000

a. Asymptotically F distributed.

8.1 Chi-square

Quick facts	
Number of variables	Two
Scale of variable(s)	Categorical (nominal/ordinal)

	No smoking	*Occasional smoking*	*Frequent smoking*
Men (age 15-24)	85 %	10 %	5 %
Women (age 15-24)	70 %	20 %	10 %

Assumptions

First, you have to check your data to see that the assumptions behind the chi-square test hold. If your data "passes" these assumptions, you will have a valid result.

Checklist	
Two or more unrelated categories in both variables	Both variables should be categorical (i.e. nominal or ordinal) and consist of two or more groups. Unrelated means that the groups should be mutually excluded: no individual can be in more than one of the groups. For example: low vs. medium vs. high educational level; liberal vs. conservative vs. socialist political views; or poor vs. fair, vs. good vs. excellent health; and so on.

Chi-square: Function

1. Go to the Menu bar, choose Analyze\Descriptive Statistics\Crosstabs.
2. A small window will open, where you see one big box and three small boxes. In the left box, all your variables are displayed.
3. Here, you choose two variables: one to be the Row variable, and one to be the Column variable.
4. Move your variables to the Row and Column boxes by using the arrows.
5. Click on Statistics.
6. Tick the box for Chi-square.
7. Click on Continue.
8. Tick the box called Suppress tables located below the box containing all variables.
9. Click on OK.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	,230[a]	2	,891
Likelihood Ratio	,230	2	,891
Linear-by-Linear Association	,168	1	,682
N of Valid Cases	5000		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 758,70.

9. CORRELATION ANALYSIS

Outline

9.1 Correlation analysis

9.1 Correlation analysis

Quick facts	
Number of variables	Two or more
Scale of variable(s)	Continuous (ratio/interval)

A correlation analysis tests the relationship between two continuous variables in terms of: a) how strong the relationship is, and b) in what direction the relationship goes. The strength of the relationship is given as a coefficient (the Pearson product-moment correlation coefficient, or simply Pearson's r) which can be anything between -1 and 1. But how do we know if the relationship is strong or weak? This is not an exact science, but here is one rule of thumb:

Strength		
Negative	**Positive**	
-1	1	**Perfect**
-0.9 to -0.7	0.7 to 0.9	**Strong**
-0.6 to -0.4	0.4 to 0.6	**Moderate**
-0.3 to -0.1	0.1 to 0.3	**Weak**
0	0	**Zero**

Thus, the coefficient can be negative or positive. These terms, "negative" and "positive", are not the same as good and bad (e.g. excellent health or poor health; high income or low income). They merely reflect the direction of the relationship.

Direction	
Negative	As the values of Variable 1 increases, the values of Variable 2 *decreases*
Positive	As the values of Variable 1 increases, the values of Variable 2 *increases*

Note however that correlation analysis does not imply anything about causality: Variable 1 does not *cause* Variable 2 (or vice versa). The correlation analysis only says something about the degree to which the two variables co-vary.

Assumptions

First, you have to check your data to see that the assumptions behind the correlation analysis hold. If your data "passes" these assumptions, you will have a valid result.

Checklist	
Two continuous variables	Both variables should be continuous (i.e. interval/ratio). For example: Income, height, weight, number of years of schooling, and so on. Although they are not really continuous, it is still rather common to use ratings as continuous variables, such as: "How satisfied with your income are you?" (on a scale 1-10) or "To what extent do you agree with the previous statement?" (on a scale 1-5).
Linear relationship between the two variables	There needs to be a linear relationship between your two variables. You can check this by creating a scatterplot (described in Section 4.7).
No outliers	An outlier is an extreme (low or high) value. For example, if most individuals have a test score between 40 and 60, but one individual has a score of 96 or another individual has a score of 1, this will distort the test.

Correlation analysis: Function

1. Go to the Menu bar, choose Analyze\Correlate\Bivariate.
2. A new window called Bivariate Correlations will open.
3. In the left box, all your variables are displayed. Highlight the variables for which you want to test the correlation, and then transfer them to Variables.
4. Click on OK.

Note that it is possible to include more than two variables, and in that case you will get correlation coefficients for each pair of variables.

Correlations

		Job satisfaction	Age in years
Job satisfaction	Pearson Correlation	1	,413**
	Sig. (2-tailed)		,000
	N	5000	5000
Age in years	Pearson Correlation	,413**	1
	Sig. (2-tailed)	,000	
	N	5000	5000

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