

# Análise Quantitativa de Dados em Marketing

Prof. Dr João Oliveira

[joao.oliveira@iseg.ulisboa.pt](mailto:joao.oliveira@iseg.ulisboa.pt)

# Trabalho de Grupo

- Disponíveis no Fenix:
  - Instruções para o Trabalho Prático;
  - Base de dados;
  - Questionário correspondente à base de dados

# Trabalho de Grupo

- Composição dos Grupos:
  - Necessário enviar por e-mail ([joao.oliveira@iseg.ulisboa.pt](mailto:joao.oliveira@iseg.ulisboa.pt)), a composição do vosso grupo (nomes tal como aparecem no MS Teams); os grupos devem ter entre 3 a 5 alunos, sem exceção, têm todos de pertencer à mesma turma (sem exceção).
  - Data limite para constituição dos grupos: 10 de Outubro

# Today' Summary (Week 4)



- Measures of centrality (continued)
  - Mean, median, mode
- Measures of dispersion
  - Minimum, maximum, range, variance, standard deviation
- Frequencies
- Tables
- Graphs

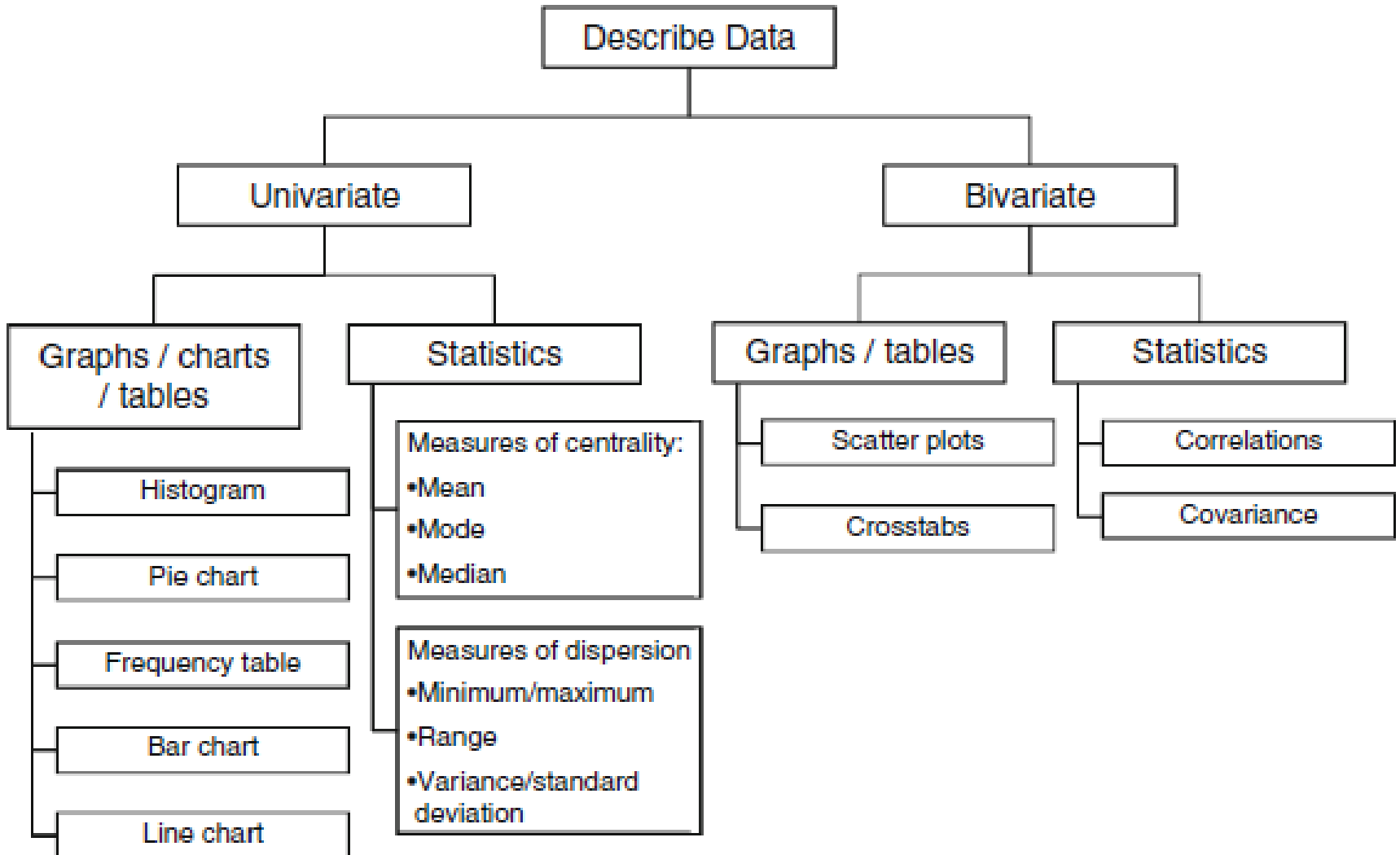


Lisbon School  
of Economics  
& Management  
Universidade de Lisboa



# Descriptives

# Descriptive Data



# Univariate measures

- Centrality-based measures
  - Mean (or average)
    - The sum of each individual observation of a variable divided by the number of observations
    - Useful when describing
      - Interval data
      - Ratio-scaled data
    - Most frequently used but is sensitive to unexpected large or small values (outliers)

# Univariate measures

- Centrality-based measures
  - Mean (or average)
    - What is the mean age in this class?
    - Is the mean a “real value”?



# Univariate measures

- Centrality-based measures
  - Median
    - The value that separates the lowest 50% of cases from the highest 50% of cases
    - Useful when describing
      - Ordinal data
    - Not sensitive to outliers

# Univariate measures

- Centrality-based measures
  - Median
    - What is the median age in this class?
    - Is the median a “real value”?

# Univariate measures

- Centrality-based measures
  - **Mode**
    - The most frequently occurring value in the dataset
    - Useful when describing
      - Ordinal data
      - Nominal data
    - Not sensitive to outliers

# Univariate measures

- Centrality-based measures
  - Mode
    - What is the mode in this class in terms of age?
    - Is the mode a “real value”?

# Univariate measures

- Centrality-based measures
  - The relationship between mean, median and mode
    - For example, if the mean is much higher than the median and mode, this suggests that the dataset contains outliers which shift the mean upwards
    - If the mean, median, and mode are more or less the same, the variable is likely to be symmetrically distributed

# Univariate measures

- Measures of dispersion
  - Minimum and maximum
    - Indicate a particular variable's highest and lowest value
  - Range
    - Difference between the highest value and the lowest value

# Univariate measures

- Measures of dispersion
  - Variance ( $s^2$ )

- Measures the sum of the squared differences between all of a variable's values and its mean divided by the number of observations minus 1

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

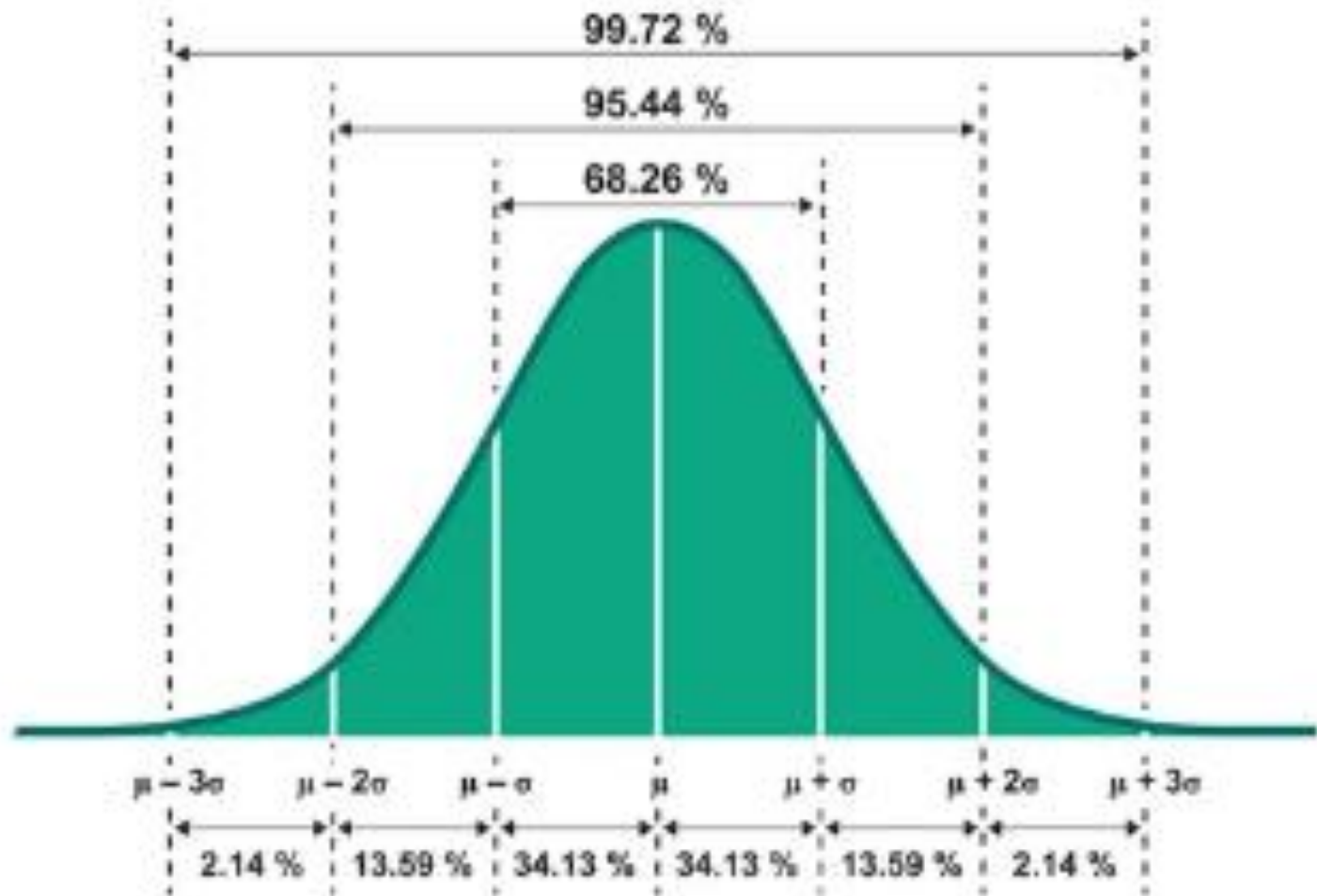
- Affected by outliers
- It tells us how strongly observations vary around the mean

# Univariate measures

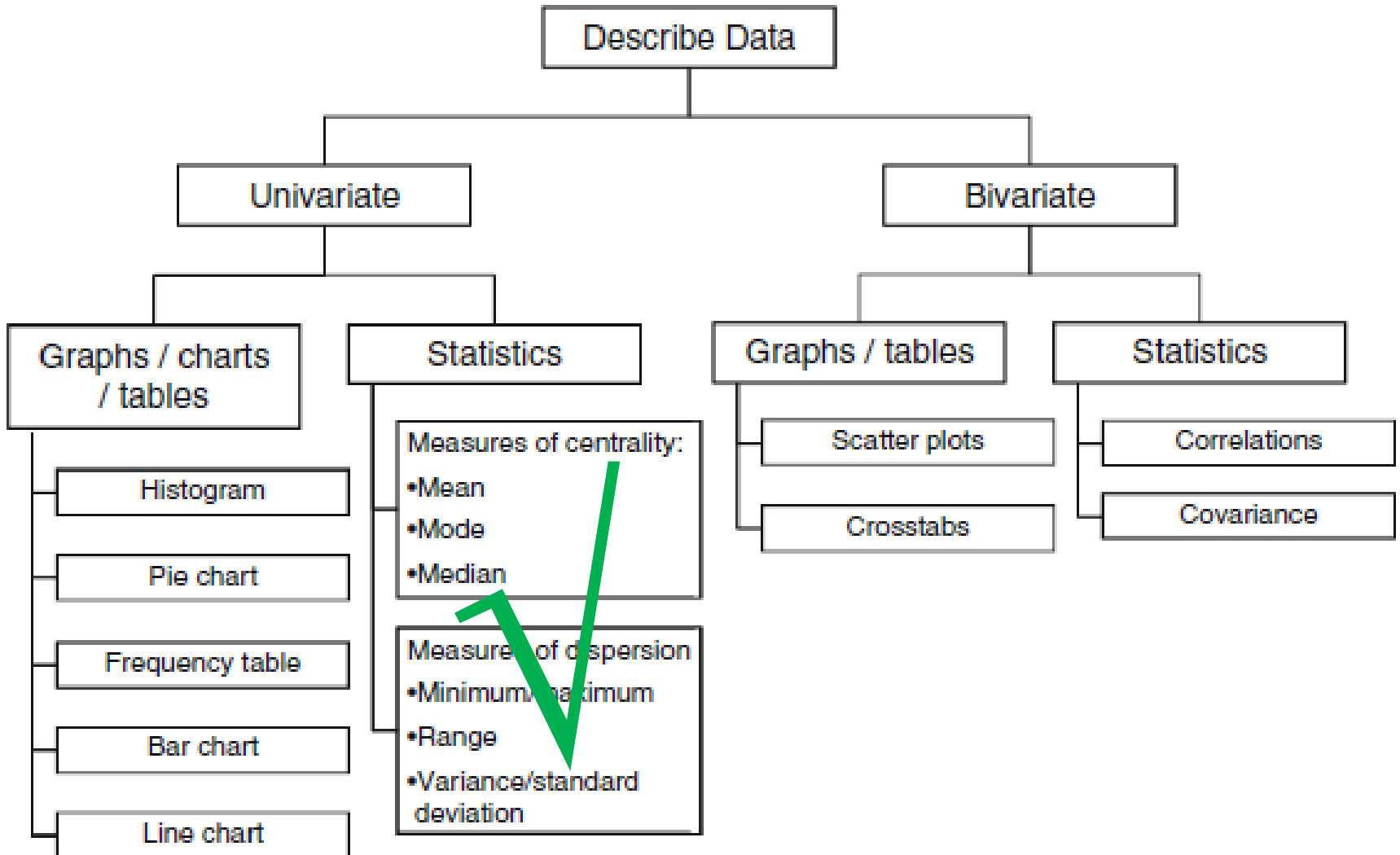
- Measures of dispersion
  - Standard deviation ( $s$ )
    - The square root of the variance
    - Measures how concentrated the data are around the mean
    - Has the same units as the original data
    - A rule of thumb for large (normally distributed) datasets
      - Approximately 68% of the data lies within  $1 \sigma$  from the mean
      - Approximately 95% of the data lies within  $2 \sigma$  from the mean
      - Approximately 99.7% of the data lies within  $3 \sigma$  from the mean



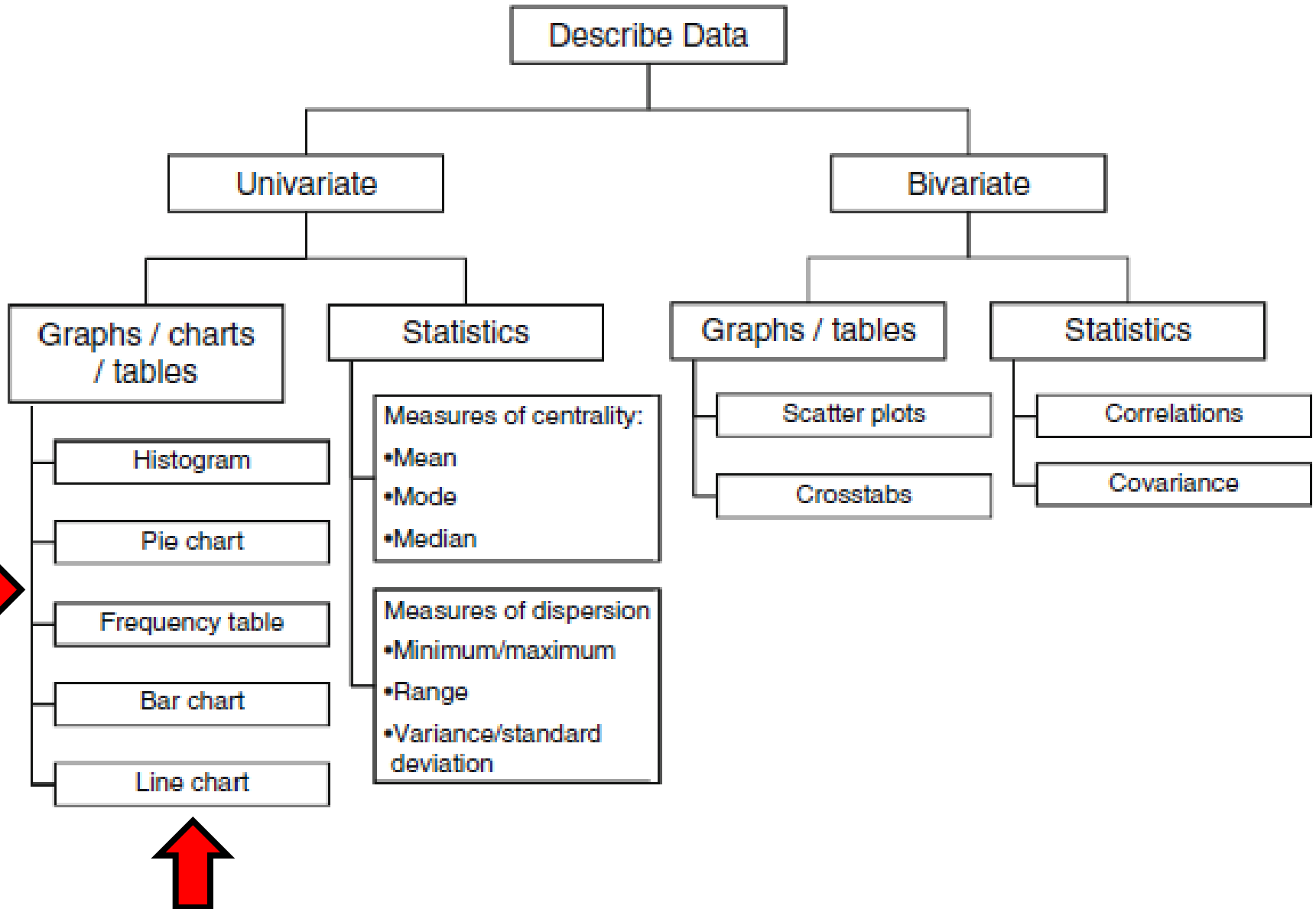
# Normal Distribution



# Descriptive Data



# Descriptive Data



# Presenting data

	Categorical		Numerical	
	Descriptive	Ranked	Continuous	Discrete
To show one variable so that any specific value can be read easily	Table/frequency distribution (data often grouped)			
To show the frequency of occurrences of categories or values for one variable so that highest and lowest are clear	Bar chart or pictogram (data may need grouping)		Histogram or frequency polygon (data must be grouped)	Bar chart or pictogram (data may need grouping)
To show the trend for a variable		Line graph or bar chart	Line graph or histogram	Line graph or bar chart
To show the proportion of occurrences of categories or values for one variable	Pie chart or bar chart (data may need grouping)		Histogram or pie chart (data must be grouped)	Pie chart or bar chart (data may need grouping)
...	...		...	...

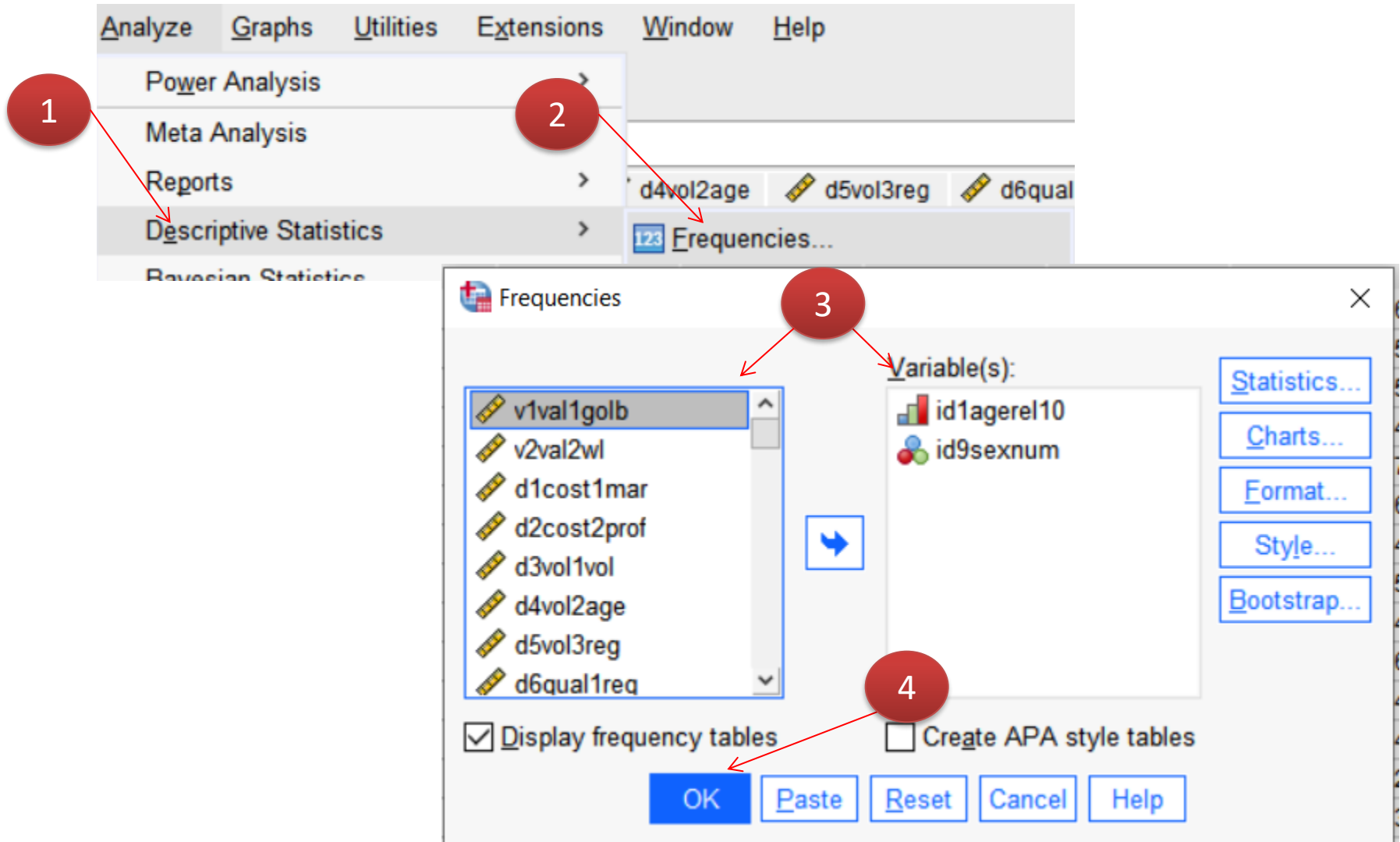
Adapted from Table 12.2

Saunders, M. L., Lewis P. & Thornhill, A.(2009). *Research methods for business students*, London: Pearson

# SPSS

- Open file “OH5 – survey analysis v3-1”

# Frequencies



The image shows a screenshot of the SPSS software interface. The 'Analyze' menu is open, and the 'Frequencies...' option is selected. The 'Frequencies' dialog box is open, showing a list of variables on the left and a 'Variable(s):' list on the right. The 'Display frequency tables' checkbox is checked. The 'OK' button is highlighted.

Numbered annotations (1-4) indicate the following steps:

1. Click on the 'Analyze' menu.
2. Click on 'Descriptive Statistics' > 'Frequencies...'
3. Click on the 'Variable(s):' list to add variables.
4. Click on the 'OK' button.

## Frequency Table

### relationship age 10 to 10 years

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 through 9 years	123	62.1	62.1	62.1
	10 through 19 years	65	32.8	32.8	94.9
	20 through 29 years	2	1.0	1.0	96.0
	30 through 39 years	8	4.0	4.0	100.0
	Total	198	100.0	100.0	

### sex-numeric

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	132	66.7	66.7	66.7
	Female	66	33.3	33.3	100.0
	Total	198	100.0	100.0	

# Descriptives

- Descriptives

1. Analyze menu

2. Descriptive Statistics

3. Variable list

4. Options... button

8. Save standardized values as variables checkbox

Variable(s): id1agerel, id6jobage, id7indage, id8age

Buttons: OK, Paste, Reset, Cancel, Help

5. Mean, Std. deviation, Variance, Range, Minimum, Maximum, S.E. mean

6. Distribution: Kurtosis, Skewness

7. Display Order: Descending means

Buttons: Continue, Cancel, Help



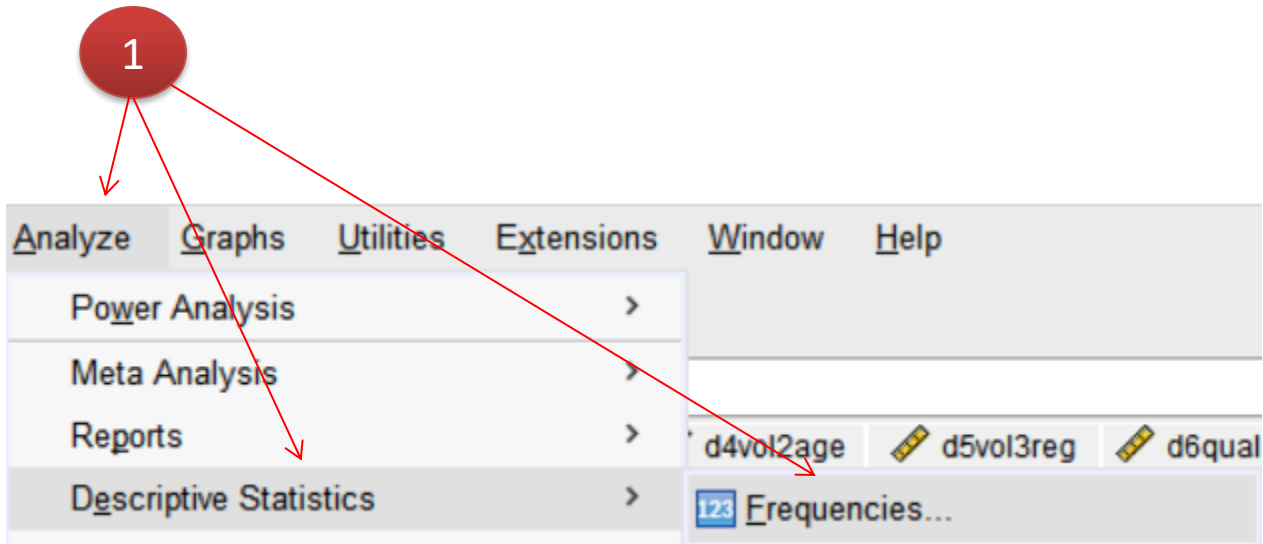
# Descriptives

- Descriptives output

## Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
id8age:A minha idade é:	198	38	61	47,22	7,201
id7indage:Trabalho nesta indústria há (anos):	198	2	44	25,42	10,918
id6jobage:Trabalho nesta empresa há (anos):	198	1	40	18,42	10,246
id1agerel:O relacionamento com este cliente tem aproximadamente (anos):	198	1	32	8,58	6,871
Valid N (listwise)	198				

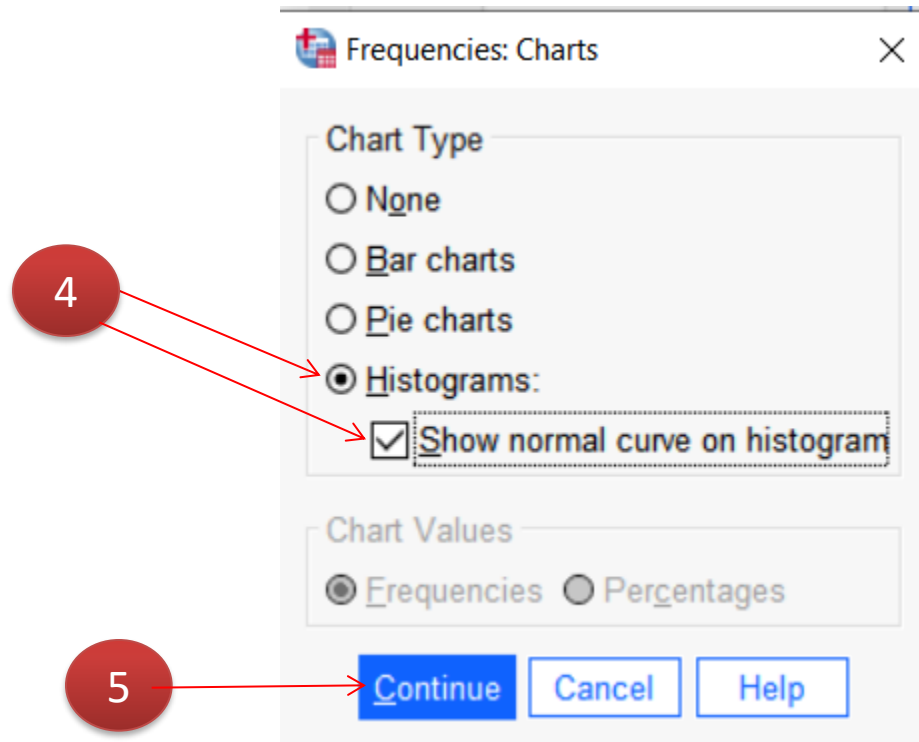
# Histograms

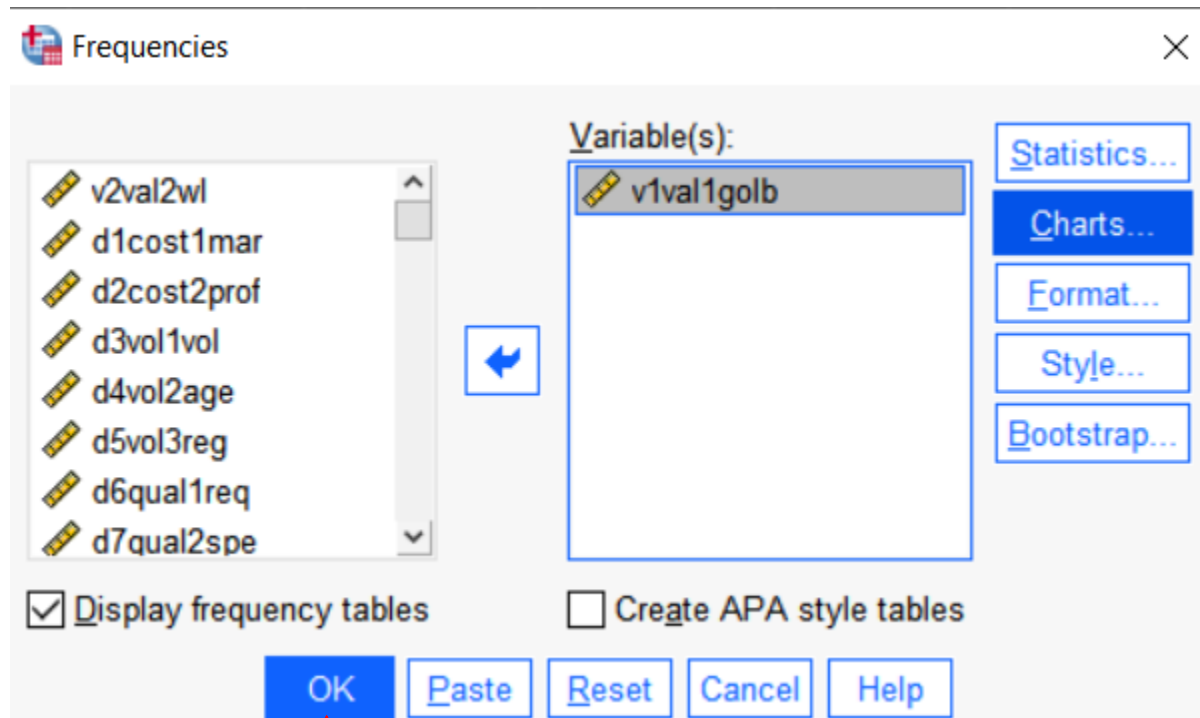


2

The image shows the 'Frequencies' dialog box in SPSS. On the left is a list of variables: v2val2wl, d1cost1mar, d2cost2prof, d3vol1vol, d4vol2age, d5vol3reg, d6qual1req, and d7qual2spe. A blue arrow points from a red circle '2' to this list. In the center, the 'Variable(s):' box contains 'v1val1golb'. Another blue arrow points from the same red circle '2' to this box. On the right, there are buttons for 'Statistics...', 'Charts...', 'Format...', 'Style...', and 'Bootstrap...'. A red circle '3' has an arrow pointing to the 'Charts...' button. At the bottom, there are checkboxes for 'Display frequency tables' (checked) and 'Create APA style tables' (unchecked), and buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.

3





6

## → Frequencies

### Statistics

v1val1golb:O relacionamento com este cliente é globalmente:

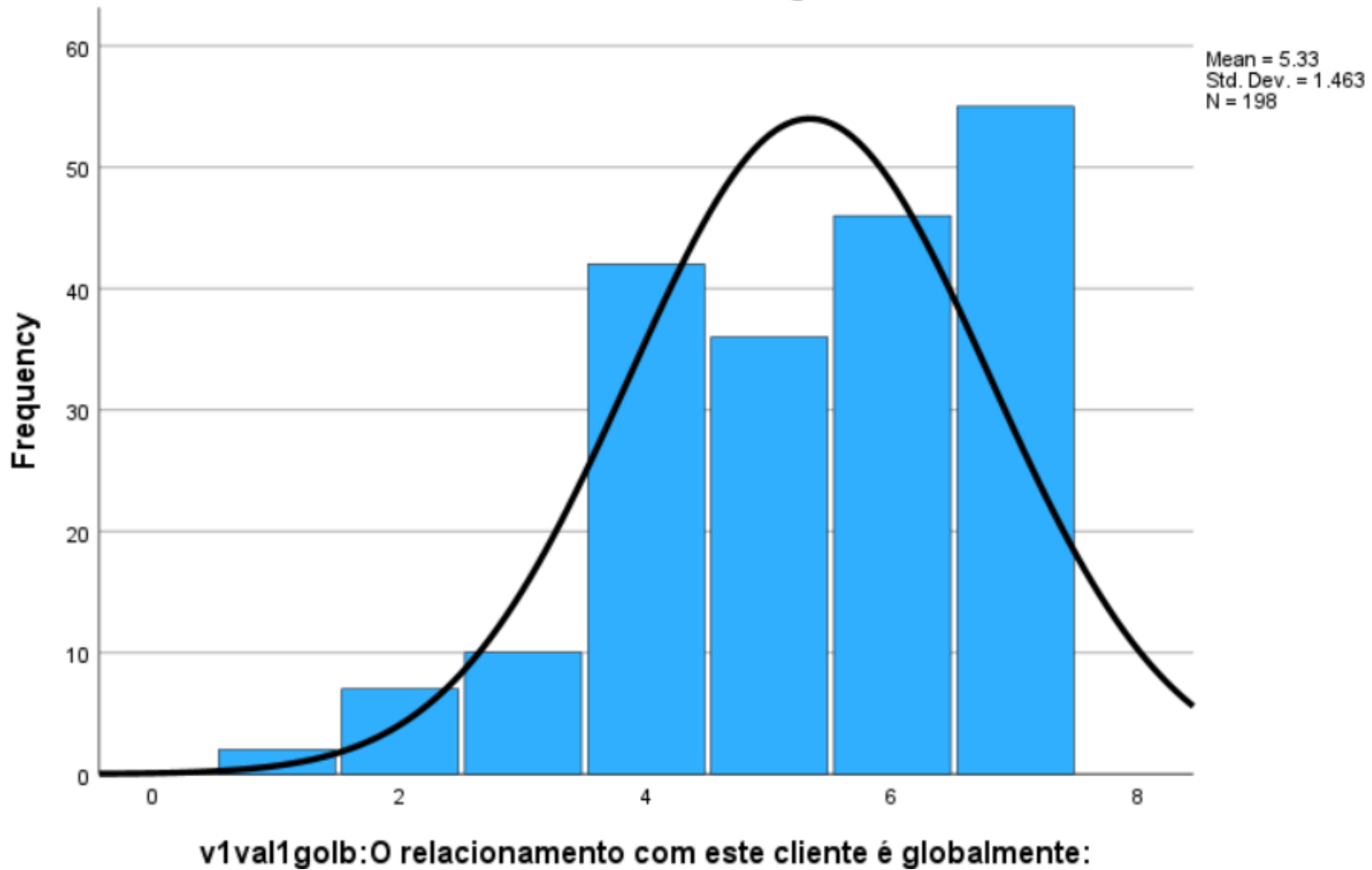
N	Valid	198
	Missing	0

**v1val1golb:O relacionamento com este cliente é globalmente:**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	2	1.0	1.0	1.0
	Disagree	7	3.5	3.5	4.5
	Slightly disagree	10	5.1	5.1	9.6
	Neutral/undecided	42	21.2	21.2	30.8
	Slightly agree	36	18.2	18.2	49.0
	Agree	46	23.2	23.2	72.2
	Strongly agree	55	27.8	27.8	100.0
	Total	198	100.0	100.0	

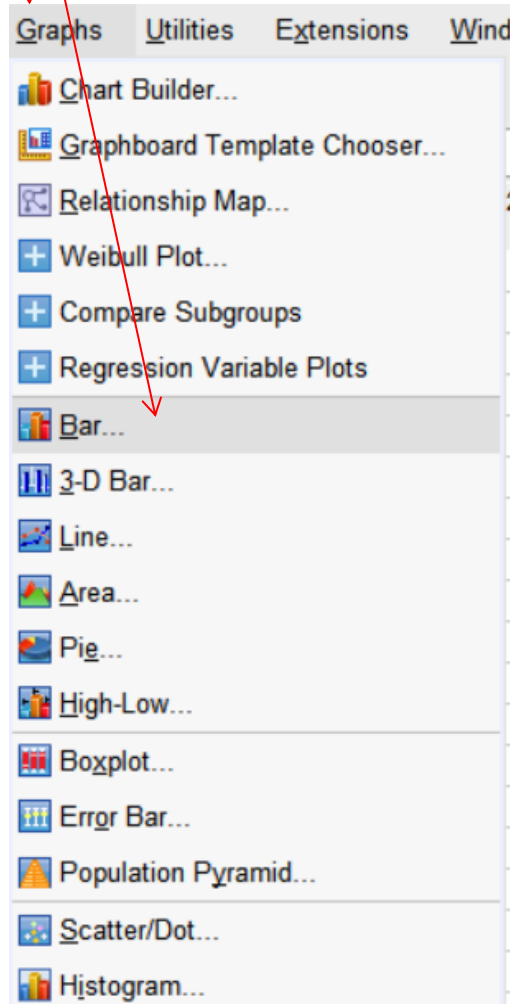
The output can be edited with a double click

Histogram

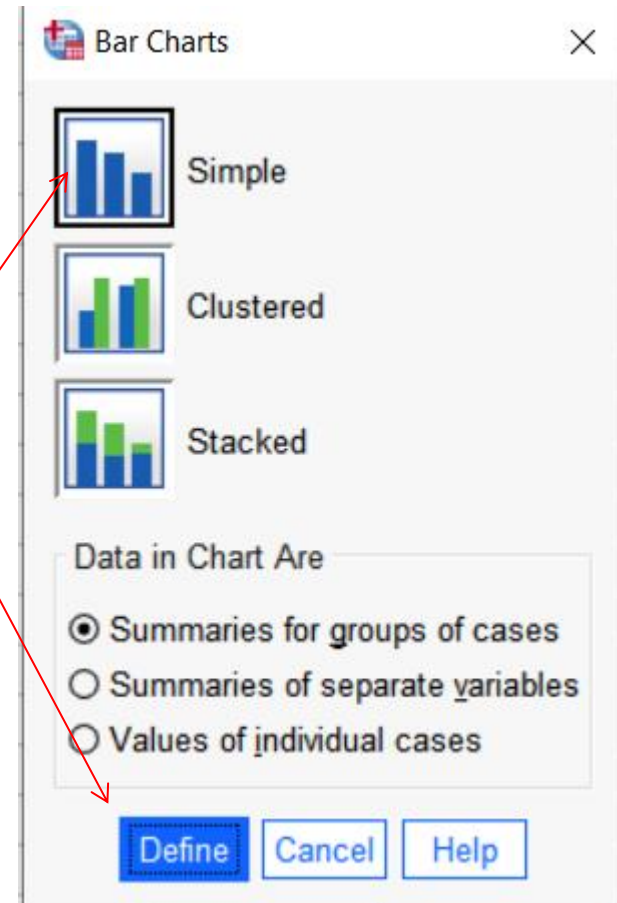


# Bar Charts

1



2





**Bars Represent**

N of cases       % of cases  
 Cum. N       Cum. %  
 Other statistic (e.g., mean)

Variable:

**Category Axis:**

**Panel by**

**Rows:**

Nest variables (no empty rows)

**Columns:**

Nest variables (no empty columns)

**Filter by:**

**Template**

Use chart specifications from:

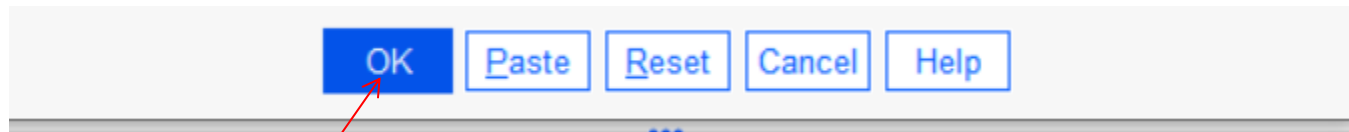
**Titles...**

**Options...**

**Variable List:**

- id1agerel
- id1agerel10
- id2fat
- id3export
- id4emp
- id4emp50
- weight1emp50
- id5job
- id5jobnum
- id5jobdummy\_1
- id5jobdummy\_2
- id5jobdummy\_3
- id5jobdummy\_4
- id6jobage
- id7indage
- id8age
- id9sexnum
- id9sexnew
- id10preexp
- id11size
- id12expeur
- id12expname

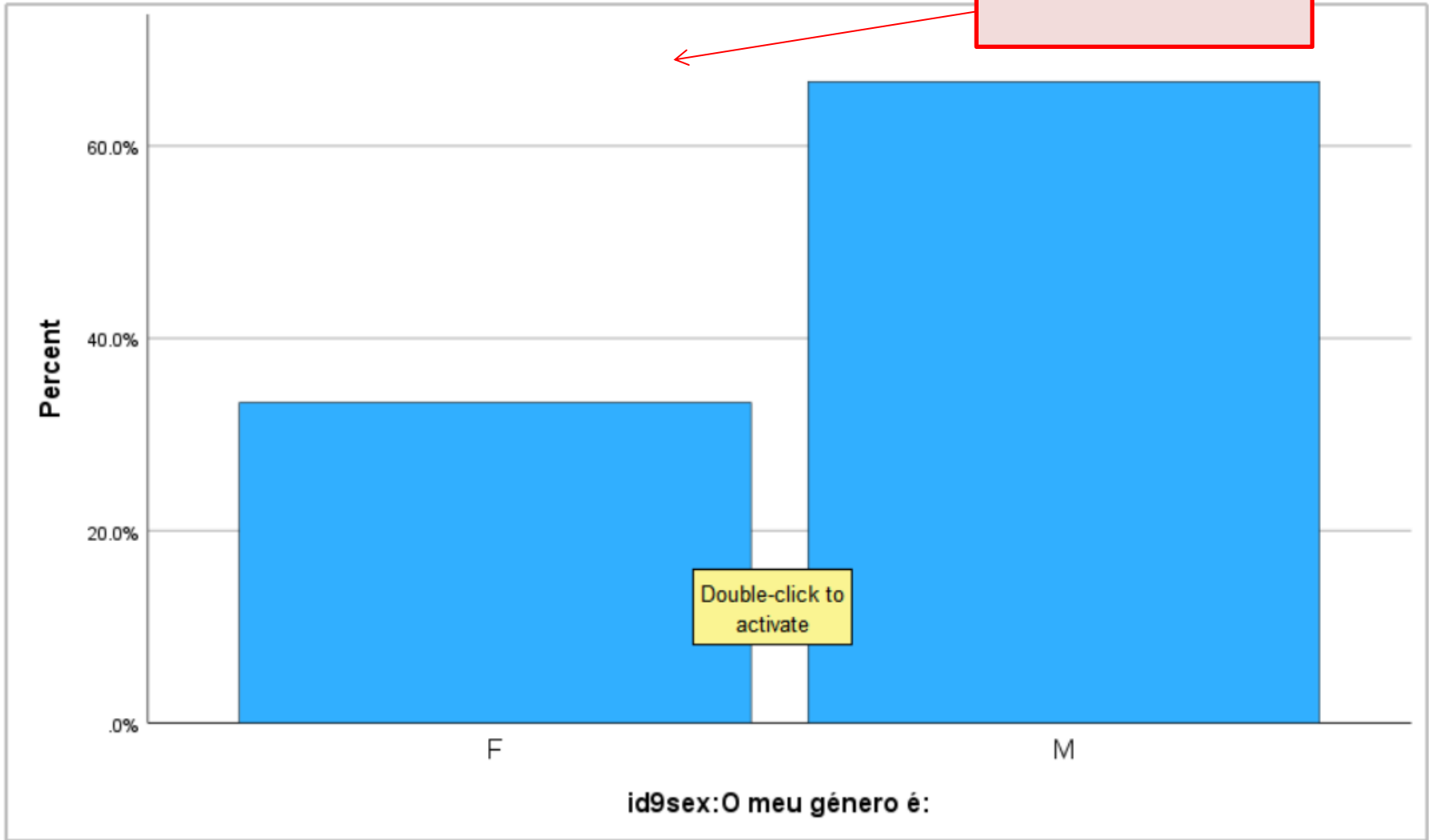
**Annotation:** A red circle with the number '3' is positioned over the 'id9sex' variable in the Category Axis field. Two red arrows point from this circle to the left and right arrow buttons of the Category Axis field.



4

→ Graph

The output can be edited with a double click



# Histograms vs Bar Charts?

## 1. Purpose:

- Histogram: Utilized to represent the distribution of continuous data
- Bar Chart: Used to represent categorical data

# Histograms vs Bar Charts?

## 2. Bars:

- Histogram: They touch each other (as data are continuous)
- Bar Chart: Separated (as categories in the data are different and not within a continuous range)

# Histograms vs Bar Charts?

## 3. X-Axis:

- Histogram: represents continuous intervals of the data
- Bar Chart: represents categories (not necessarily related or ordered)

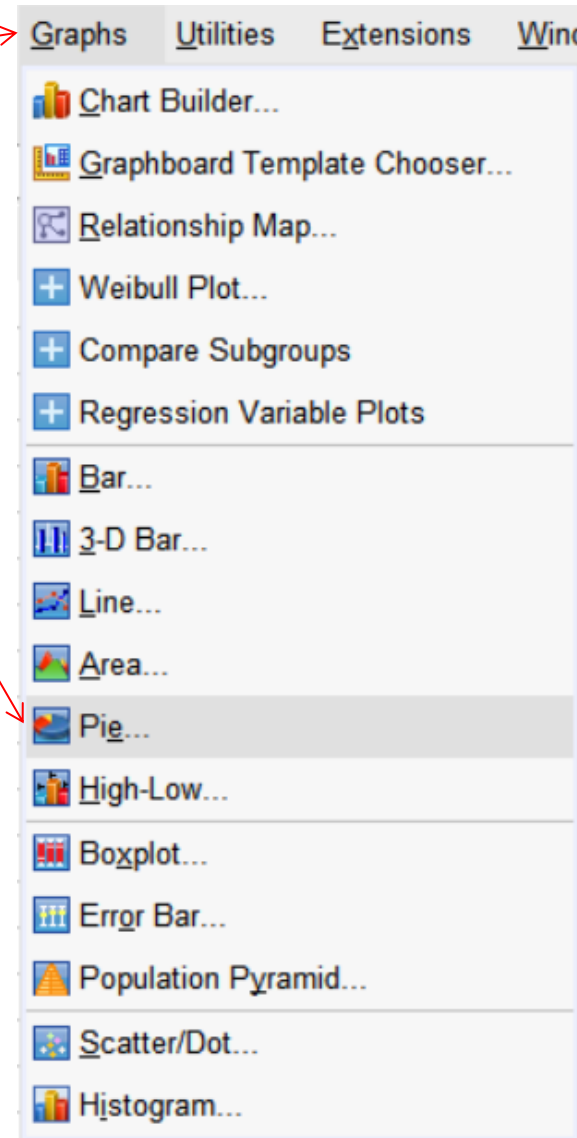
# Histograms vs Bar Charts?

## 4. Representation of Data:

- Histogram: places emphasis on the *distribution* of data and is useful for visualizing the shape of the distribution
- Bar Chart: places emphasis on *comparison* between different categories

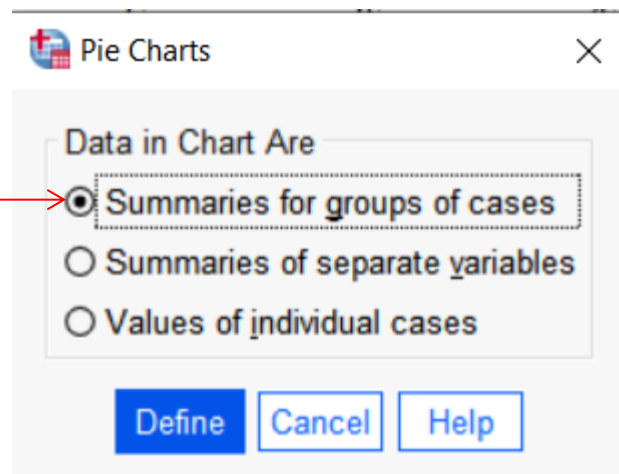
# Pie Charts

1





2



- v1val1golb
- v2val2wl
- d1cost1mar
- d2cost2prof
- d3vol1vol
- d4vol2age
- d5vol3reg
- d6qual1req
- d7qual2spe
- d8sec1dep
- d9sec2scop
- d10sec3flx
- d11sec4del
- i1mar1new
- i2mar2ref
- i3acc1info
- i4acc2good
- i5acc3bad
- i5acc3bad

Filter by:

Slices Represent

N of cases     % of cases

Sum of variable

Variable:

Titles...

Options...

Define Slices by:

3

Panel by

Rows:

Nest variables (no empty rows)

Columns:

Nest variables (no empty columns)

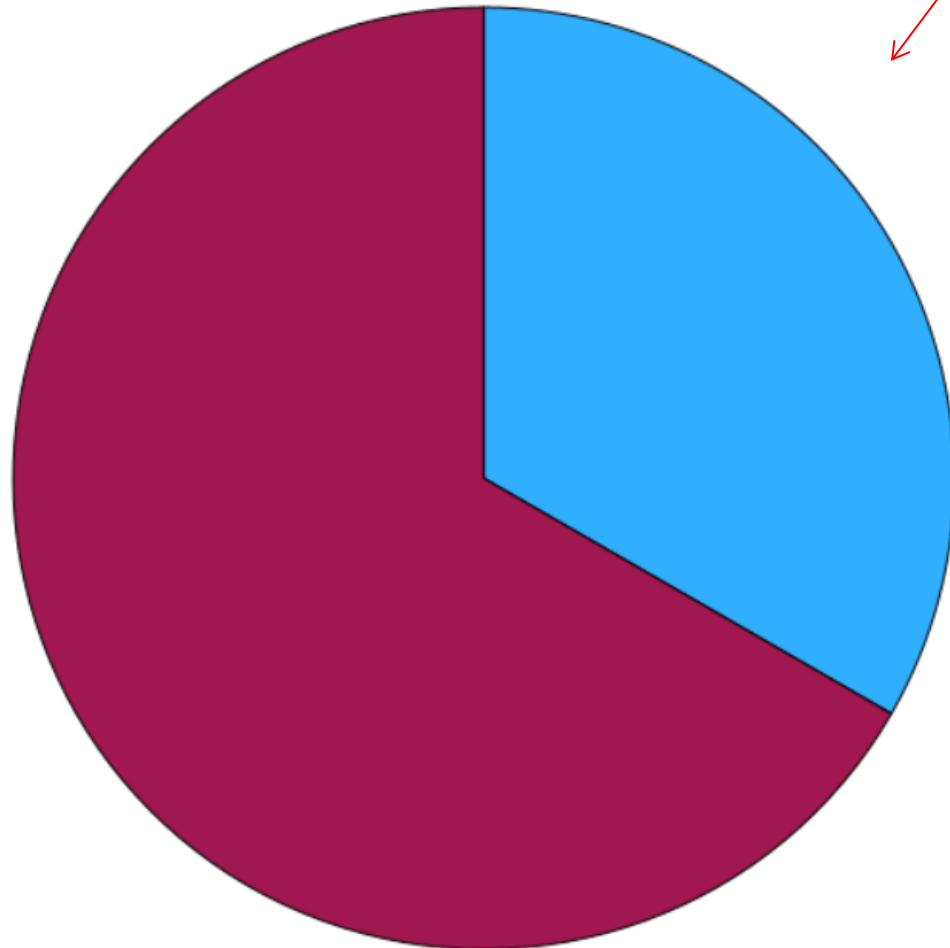
Template

Use chart specifications from:

4

OK    Paste    Reset    Cancel    Help

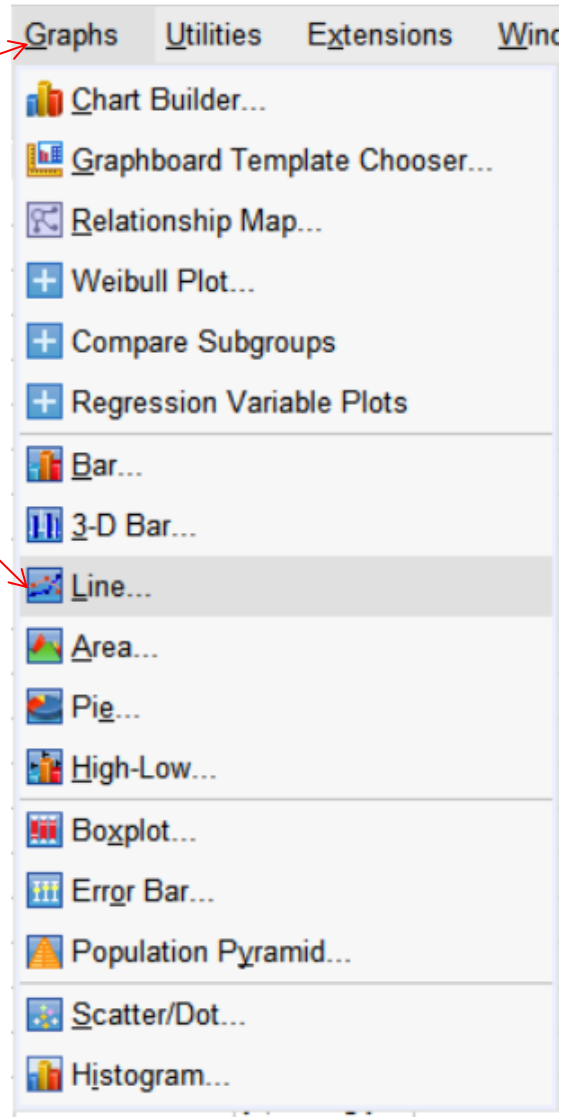
The output can be edited with a double click

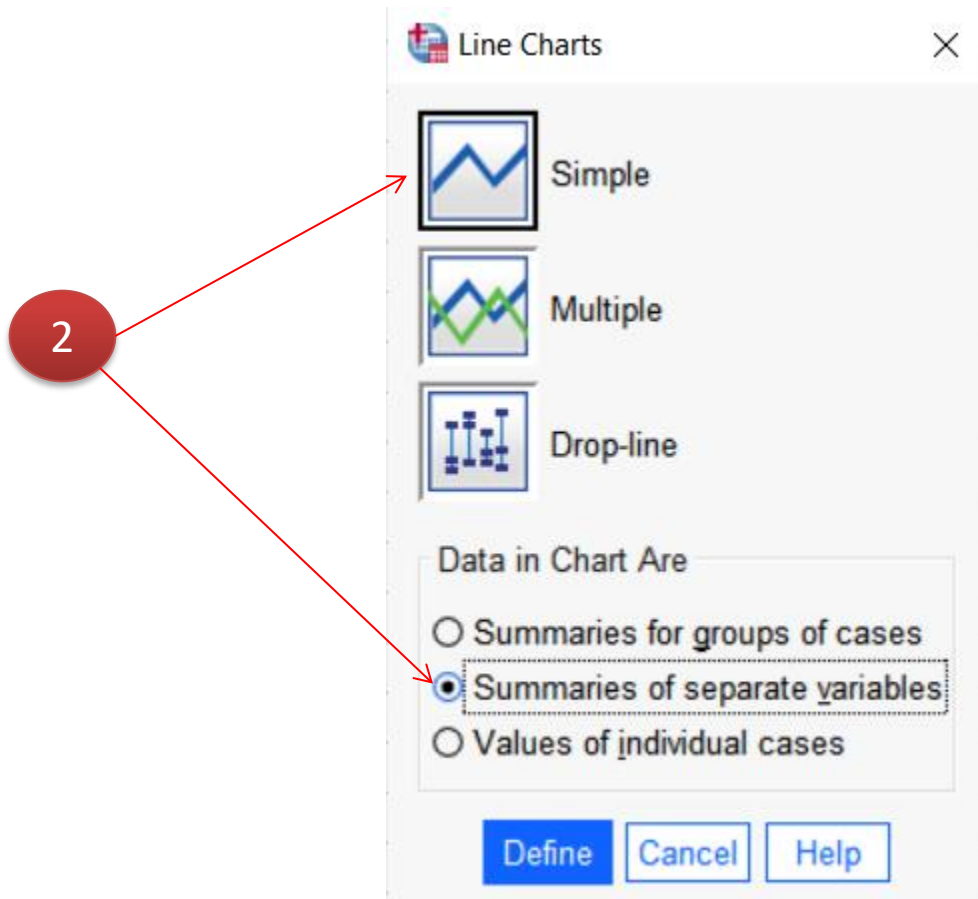


id9sex:O  
meu  
gênerô é:  
■ F  
■ M

# Line Charts

1





3

- v1val1golb
- v2val2wl
- d1cost1mar
- d2cost2prof
- d3vol1vol
- d4vol2age
- d5vol3reg
- d6qual1req
- d7qual2spe
- d8sec1dep
- d9sec2scop
- d10sec3fix
- d11sec4del
- i1mar1new
- i2mar2ref
- i3acc1info
- i4acc2good
- i5acc3bad
- i5acc3badinv

Filter by:

Line Represents:

[Empty box for Line Represents]

Change Statistic...

Titles...

Options...



Panel by

Rows:

[Empty box for Rows]

Nest variables (no empty rows)

Columns:

[Empty box for Columns]

Nest variables (no empty columns)

Template

Use chart specifications from:

File...

Line Represents:

MEAN(d2cost2prof)  
MEAN(d3vol1vol)  
MEAN(d4vol2age)  
MEAN(d5vol3reg)  
MEAN(d6qual1req)  
MEAN(d7qual2spe)  
MEAN(d8sec1dep)

Change Statistic...

Panel by

Rows:

Columns:

Nest variables (no empty rows)

Nest variables (no empty columns)

Filter by:

Template

Use chart specifications from:

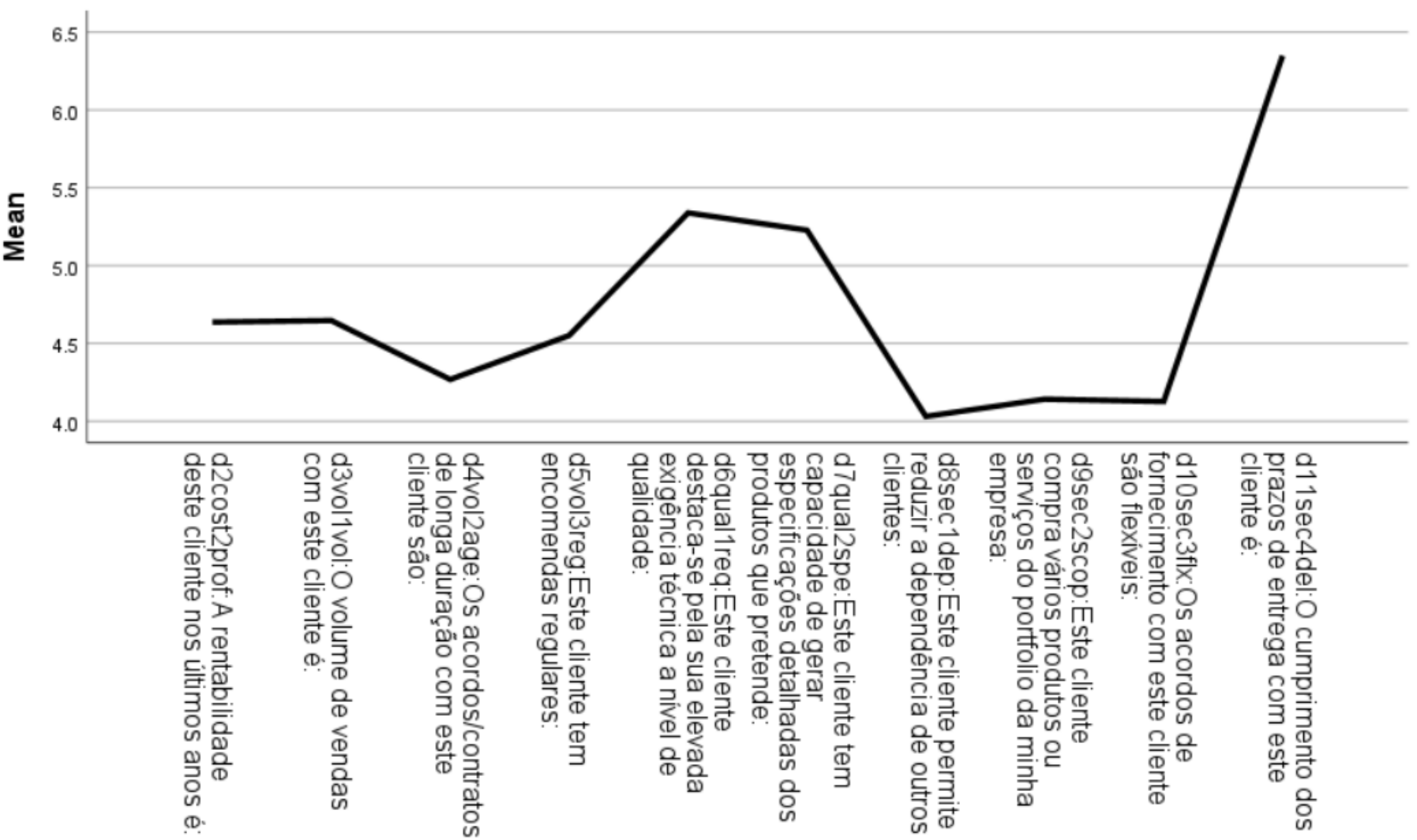
File

OK Paste Reset Cancel Help

Titles...  
Options...

v1val1golb  
v2val2wl  
d1cost1mar  
d2cost2prof  
d3vol1vol  
d4vol2age  
d5vol3reg  
d6qual1req  
d7qual2spe  
d8sec1dep  
d9sec2scop  
d10sec3fix  
d11sec4del  
i1mar1new  
i2mar2ref  
i3acc1info  
i4acc2good  
i5acc3bad  
i5acc3badinv

5







Lisbon School  
of Economics  
& Management  
Universidade de Lisboa



# So far, we've covered Univariate Data



Lisbon School  
of Economics  
& Management  
Universidade de Lisboa

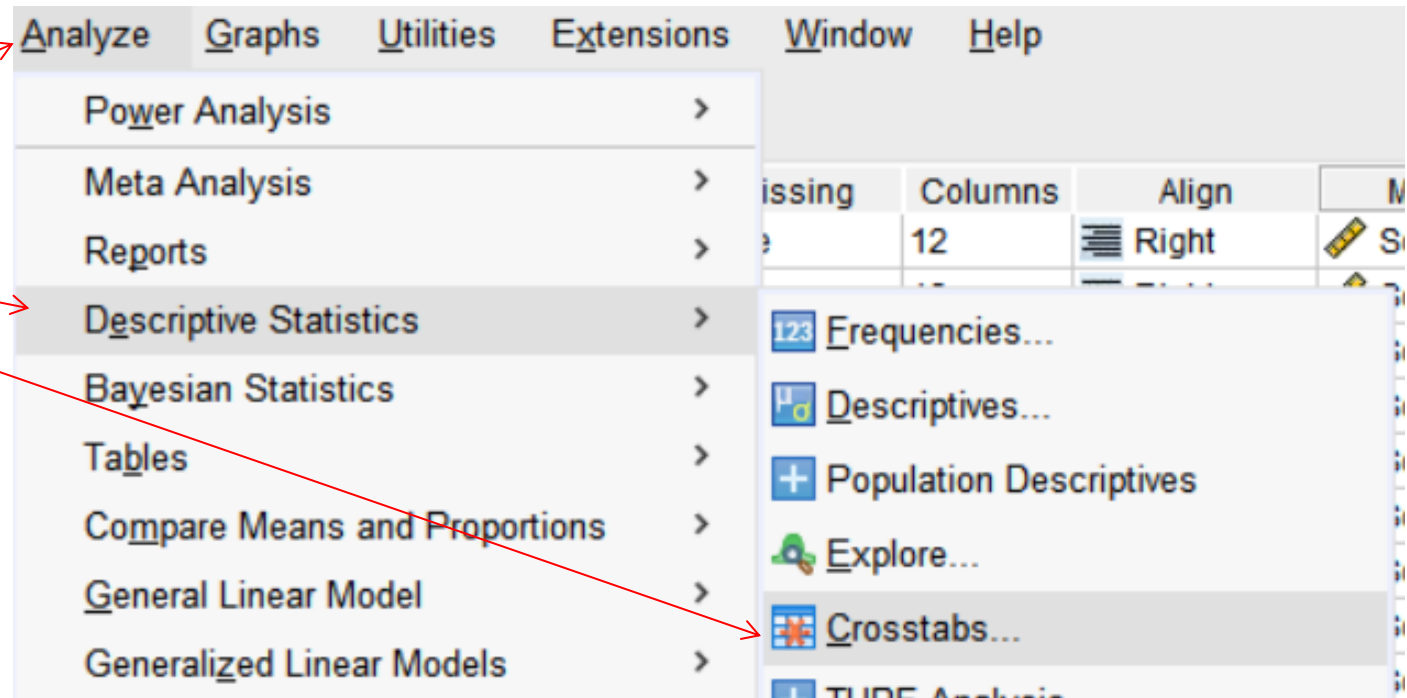


# Now let's cover Bivariate Data



# Crosstabs

(combining info of 2 variables)



Crosstabs

2

Row(s):  
id1agerel10

Column(s):  
id12exp eur

Layer 1 of 1

Previous Next

Display layer variables in table layers

Display clustered bar charts

Suppress tables

OK Paste Reset Cancel Help

3

### Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
relationship age 10 to 10 years * exports to europe	198	100.0%	0	0.0%	198	100.0%

### relationship age 10 to 10 years \* exports to europe Crosstabulation

Count

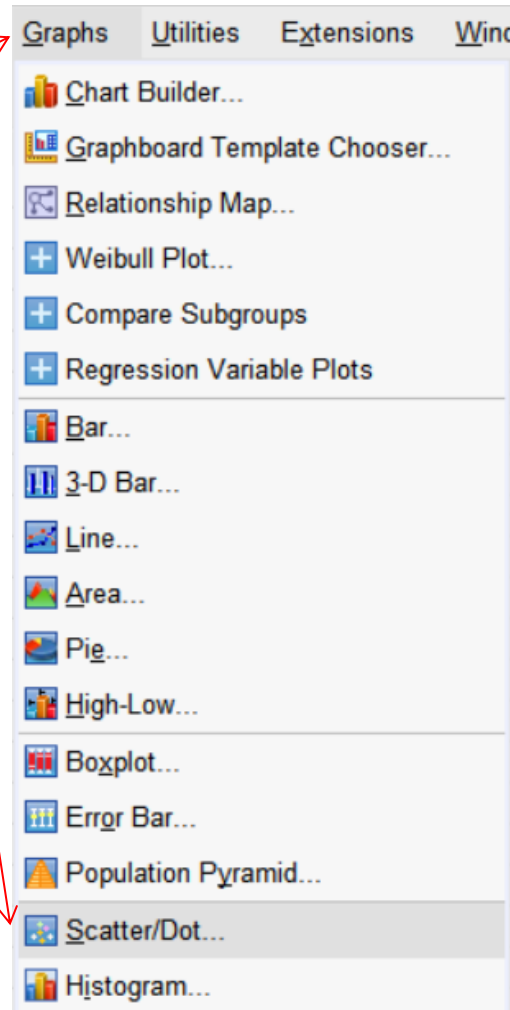
		exports to europe		Total
		no	yes	
relationship age 10 to 10 years	0 through 9 years	49	74	123
	10 through 19 years	19	46	65
	20 through 29 years	1	1	2
	30 through 39 years	2	6	8
Total		71	127	198

Helps us to see, in a table, whether for each of the 4 types of companies (according to relationship length) how many export to Europe or not.

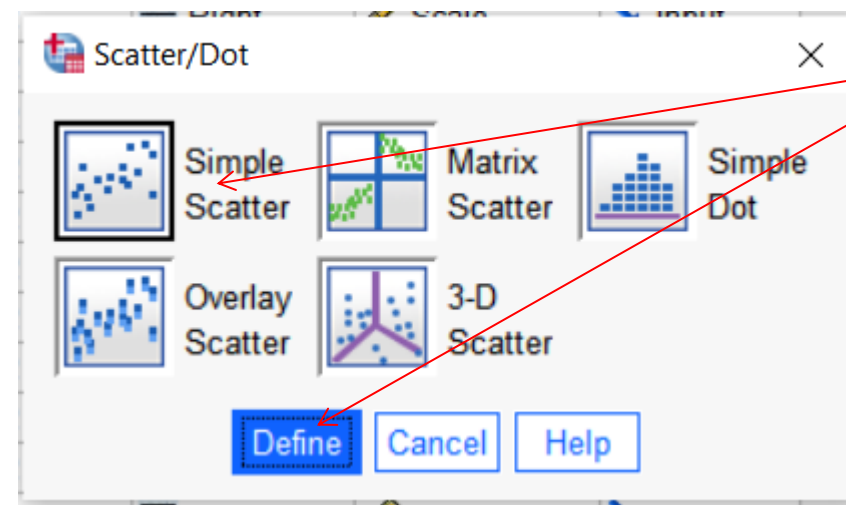
# Bivariate Graphs

(combining info of 2 variables)

# Scatter plot



We want to see in a graph the relation between the average of value and sales volume





- v1val1golb
- v2val2wl
- d1cost1mar
- d2cost2prof
- d3vol1vol
- d4vol2age
- d5vol3reg
- d6qual1req
- d7qual2spe
- d8sec1dep
- d9sec2scop
- d10sec3fix
- d11sec4del
- i1mar1new
- i2mar2ref
- i3acc1info
- i4acc2good
- i5acc3bad
- i5acc3badinv

Filter by:



Y Axis:

Titles...



X Axis:

Options...




Set Markers by:



Label Cases by:


Panel by

Rows:



Nest variables (no empty rows)

Columns:

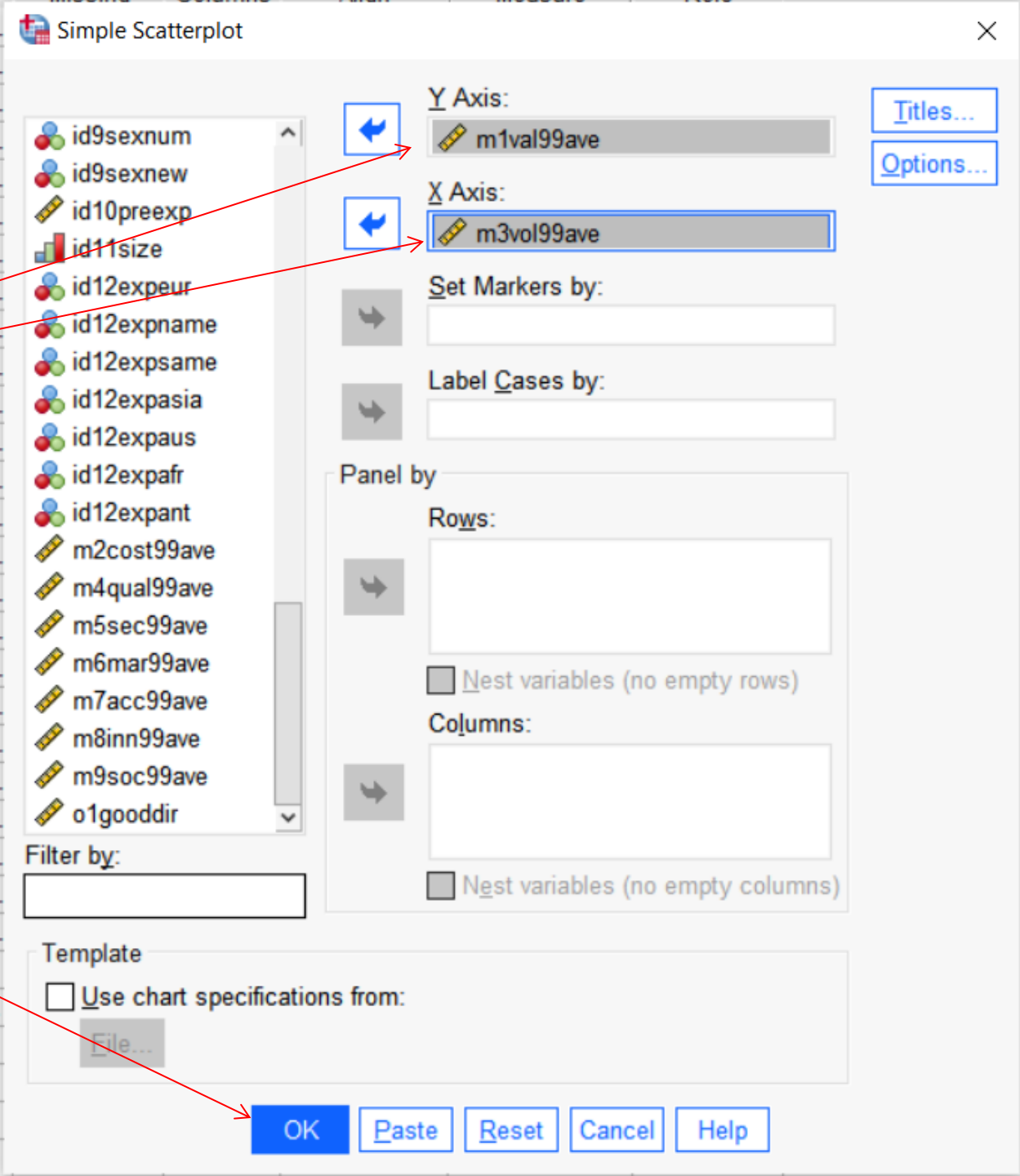


Nest variables (no empty columns)

Template

Use chart specifications from:

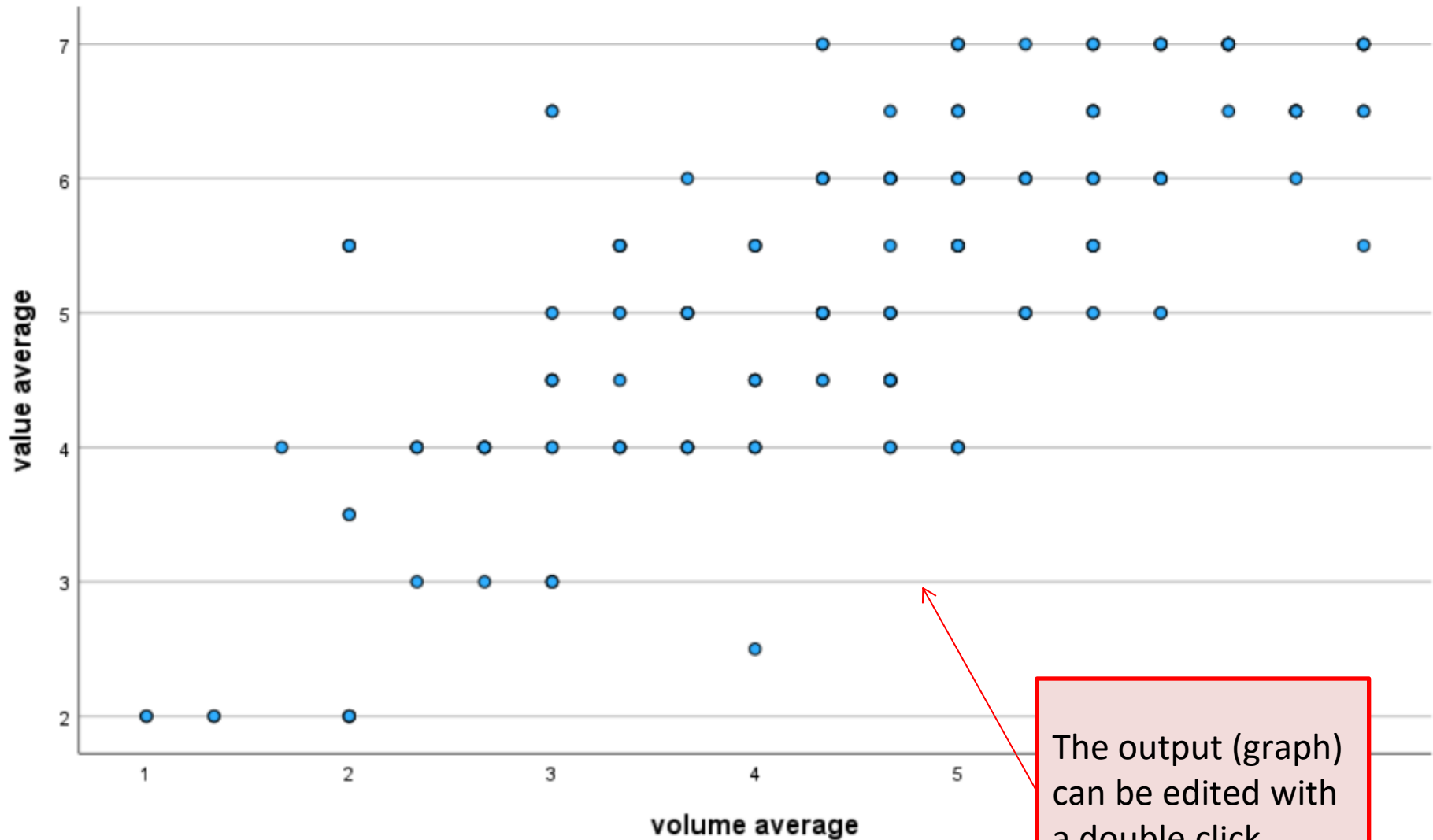
File...



4

5

→ Graph



The output (graph) can be edited with a double click

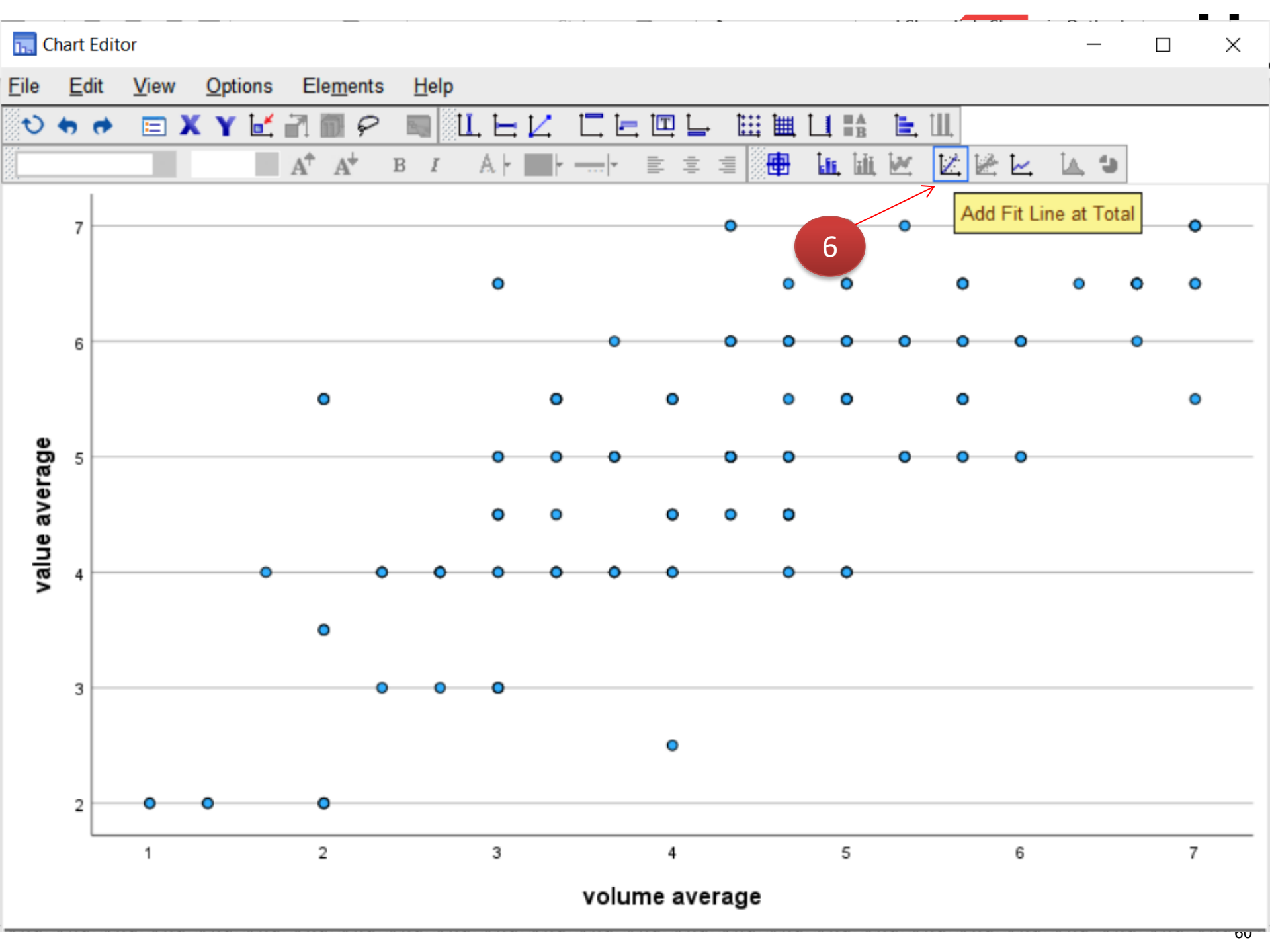
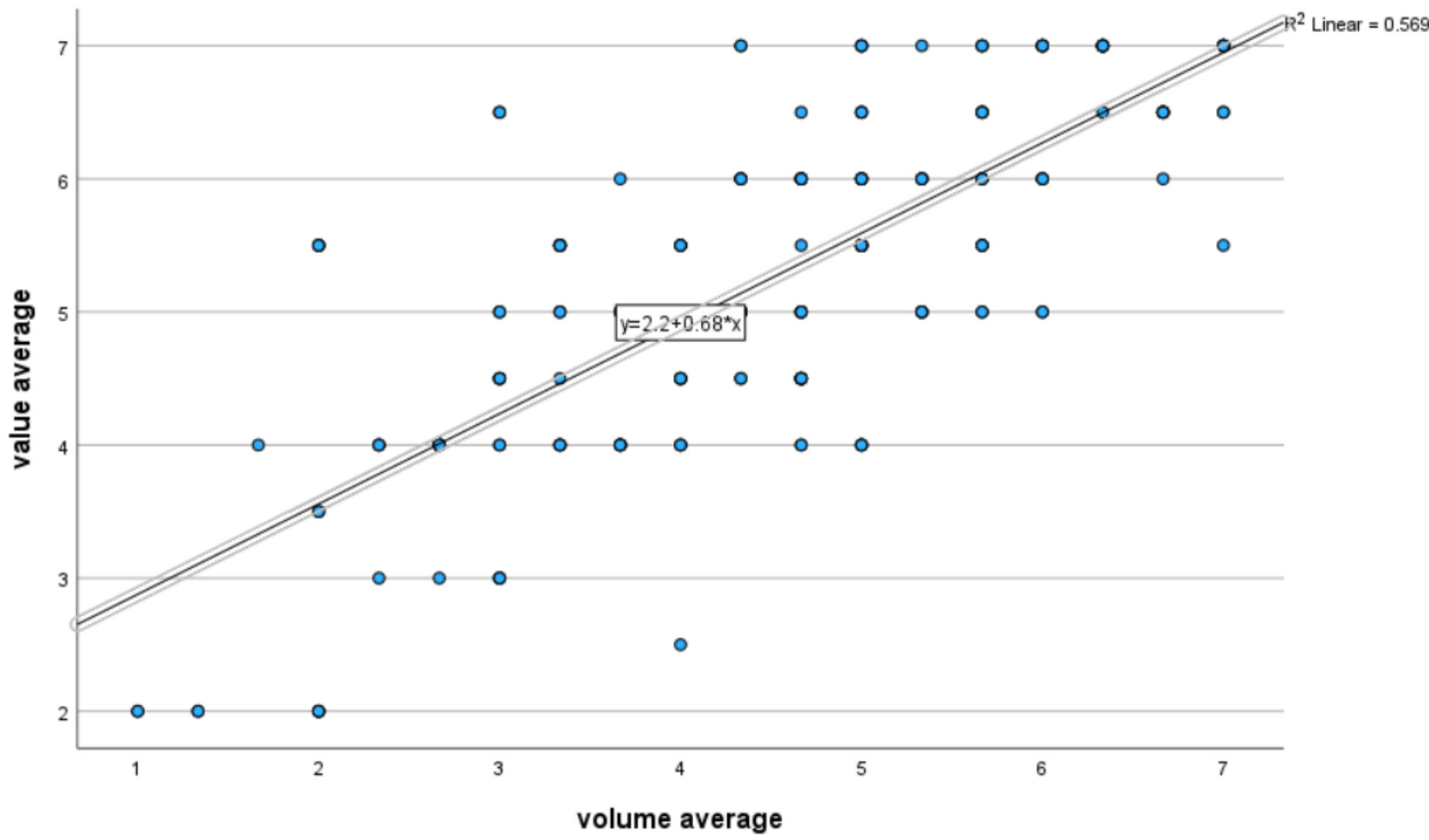


Chart Editor toolbar containing various icons for chart manipulation, including pan, zoom, and data selection tools.



# Write Down

- What topics and subtopics did you learn today?
- Which of those did you (not) understand?



Lisbon School  
of Economics  
& Management  
Universidade de Lisboa



***Thank you!***