

# **EDUS362**

# **Quantitative Research Method**

Demo 1.

## Demo materials

- EDUS362 demo slides and tasks are based on materials drawn up by a project researcher Jukka Utriainen, an university researcher Susanna Paloniemen and a PhD student Sanna Moilanen.

## Reference

Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill Education (UK).



## Objectives of demo

- After completing the course, students are able to:
  - make a research plan in accordance with the principals of the quantitative research.
  - collect numerical research data, perform statistical analysis and report the statistical results.
  - understand the ethical principals of research work.
- In the demo, students will practice:
  - how to conduct statistical analysis using SPSS software.
  - how to report the results in words, tables and figures.
- Gain the **basic** skills necessary for writing an assignment / Master's thesis.

## Course completion and evaluation

- Doing exercises in each demo.
  - Demo materials are in Koppa demo folder.
- Instruction of assignment will be found in Koppa.
- Return the assignment no later than **4.4. 24:00** to Koppa.
- The assignment will be graded 0-5.

## Absence and replacement

- Attendance is basically mandatory, but following reasons are acceptable: 1) disease and 2) other personal reasons related to life situation (e.g., child's illness)
- Please tell your absence beforehand.
- Replacement of absence:
  - Only **one-time absence** can be replaced by an independent task.
  - Do the demo tasks independently and return by e-mail **within a week** of the demo session.

## Demo delivery

- Demo 1 Preparing the data file and descriptive statistic
- Demo 2 Crosstabs and Chi-square ( $\chi^2$ ) test
- Demo 3 Correlations
- Demo 4 Comparing two groups (t-test) and calculating effect size (Cohen *d*)
- Demo 5 Comparing three (or more) groups (one-way analysis of variance, ANOVA)

## The use of data in demo

### Research ethical commitments:

- The data set has been used ONLY in the demo.
- The data set can not be transferred to others.
- At the end of the course, please remove the data from your own directory (U-drive).
- DO NOT save the data other than your home directory.
  - USB stick, hard disk or your home computer have possibilities to be reached by unauthorized persons.



# Demo 1

## Demo 1 Contents

1. Retrieve a data file from Koppa and save it to your home directory (U-drive).
2. Open the data file in SPSS.
3. Conduct descriptive analysis using SPSS.
4. Interpret the graphic presentation of the data.

Let's go through the exercise together.

Learning by doing is the best way to learn statistics!

# Open a data file in SPSS

(in Handout p.1)

The screenshot shows the Variable View dialog box in IBM SPSS Statistics. The dialog is set to 'Variable View' and displays a list of 27 variables. The variables are organized into columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The 'Measure' column shows that 'Participation' is Nominal, 'Gender' is Nominal, 'Age' is Scale, 'Credit' is Nominal, and 'Sum\_StudyEX' is Scale. All other variables are set to Nominal. The 'Role' column shows that all variables are set to 'Input'.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Participation	String	3	0		None	None	9	Left	Nominal	Input
2	Gender	Numeric	1	0		{1, Male}...	None	8	Left	Nominal	Input
3	Age	Numeric	2	0		None	None	8	Right	Scale	Input
4	Credit	Numeric	3	0		None	None	8	Right	Nominal	Input
5	Pre_Posit	Numeric	1	0		{1, Permane...	None	8	Left	Nominal	Input
6	Specializati...	Numeric	1	0		{1, Early Chi...	None	8	Left	Nominal	Input
7	StudyEx1	Numeric	1	0		None	None	8	Right	Nominal	Input
8	StudyEx2	Numeric	1	0		None	None	8	Right	Nominal	Input
9	StudyEx3	Numeric	1	0		None	None	8	Right	Nominal	Input
10	StudyEx4	Numeric	1	0		None	None	8	Right	Nominal	Input
11	StudyEx5	Numeric	1	0		None	None	8	Right	Nominal	Input
12	StudyEx6	Numeric	1	0		None	None	8	Right	Nominal	Input
13	StudyEx7	Numeric	1	0		None	None	8	Right	Nominal	Input
14	StudyEx8	Numeric	1	0		None	None	8	Right	Nominal	Input
15	StudyEx9	Numeric	1	0		None	None	8	Right	Nominal	Input
16	FutEmp1	Numeric	1	0		None	None	8	Right	Nominal	Input
17	FutEmp2	Numeric	1	0		None	None	8	Right	Nominal	Input
18	FutEmp3	Numeric	1	0		None	None	8	Right	Nominal	Input
19	FutEmp4	Numeric	1	0		None	None	8	Right	Nominal	Input
20	FutEmp5	Numeric	1	0		None	None	8	Right	Nominal	Input
21	FutEmp6	Numeric	1	0		None	None	8	Right	Nominal	Input
22	LifeSat1	Numeric	1	0		None	None	8	Right	Nominal	Input
23	LifeSat2	Numeric	1	0		None	None	8	Right	Nominal	Input
24	LifeSat3	Numeric	1	0		None	None	8	Right	Nominal	Input
25	LifeSat4	Numeric	1	0		None	None	8	Right	Nominal	Input
26	LifeSat5	Numeric	1	0		None	None	8	Right	Nominal	Input
27	Sum_StudyEX	Numeric	8	0		None	None	13	Right	Scale	Input

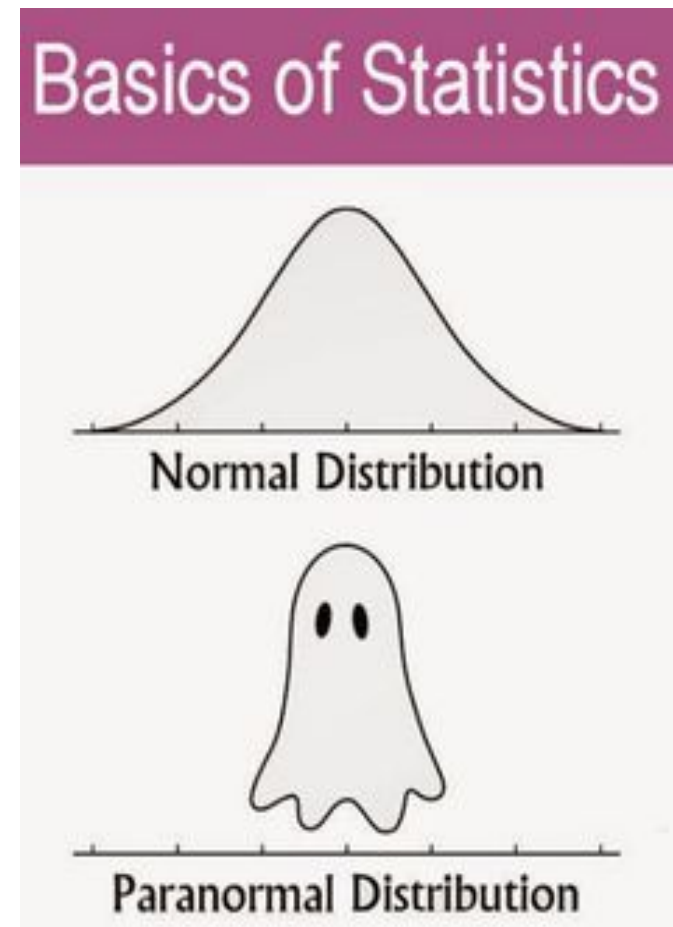
At the bottom of the dialog, there are two tabs: 'Data View' and 'Variable View', with 'Variable View' selected. The status bar at the bottom right of the window indicates 'IBM SPSS Statistics Processor is ready' and 'Unicode:OFF'.

# Categorization of variables in SPSS

- **Nominal:**
  - 2 or more categories. No order.
  - E.g. What is your gender? → Male, Female or Transgender.
- **Ordinal:**
  - 2 or more categories. It can be ordered.
  - E.g. How satisfied are you with your life?  
→ Completely satisfied, Very satisfied, Somewhat satisfied, Somewhat dissatisfied, Very dissatisfied, Completely dissatisfied
- **Scale:**
  - It can be measured along a continuum and they have numerical value.
  - E.g. Height, temperature
  - E.g. Level of life satisfaction (calculated as a mean of multiple indicators in a Life Satisfaction Questionnaire)

# Describing frequencies and percentages graphically

- **Bar charts** and **pie charts** are suitable to present nominal or ordinal variables.
  - E.g. Students' major: special needs education, adult education, early childhood education etc.
- **Histograms** are often used in analysis for continuous variables.
  - E.g. age, number of credits
  - Histogram can be used to examine whether the variable values are normally distributed data or not



## Exercise 1a: Frequency

Calculate the frequency distributions of “Gender” and create bar charts.  
(in Handout p.2)

		Gender			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	11	25.0	25.0	25.0
	Female	33	75.0	75.0	100.0
	Total	44	100.0	100.0	

Frequencies  
(Number of observations)

Percentages

Percentages,  
missing observations  
are reduced

Cumulative  
percentages

[Example of report]

”In this study, there are 33 females (75.0 %) and 11 males (25.0 %) in the sample, giving a total of 44 respondents.”

## Exercise 1b: Bar charts

- Put the numbers of observations and the percentages

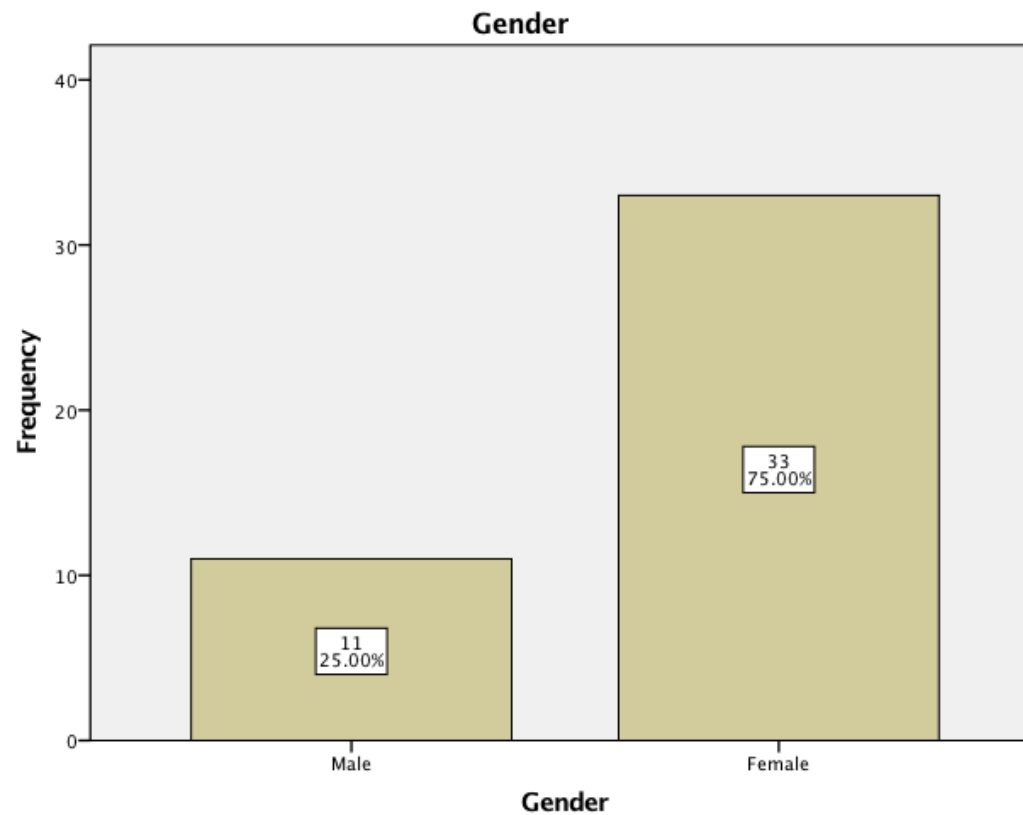
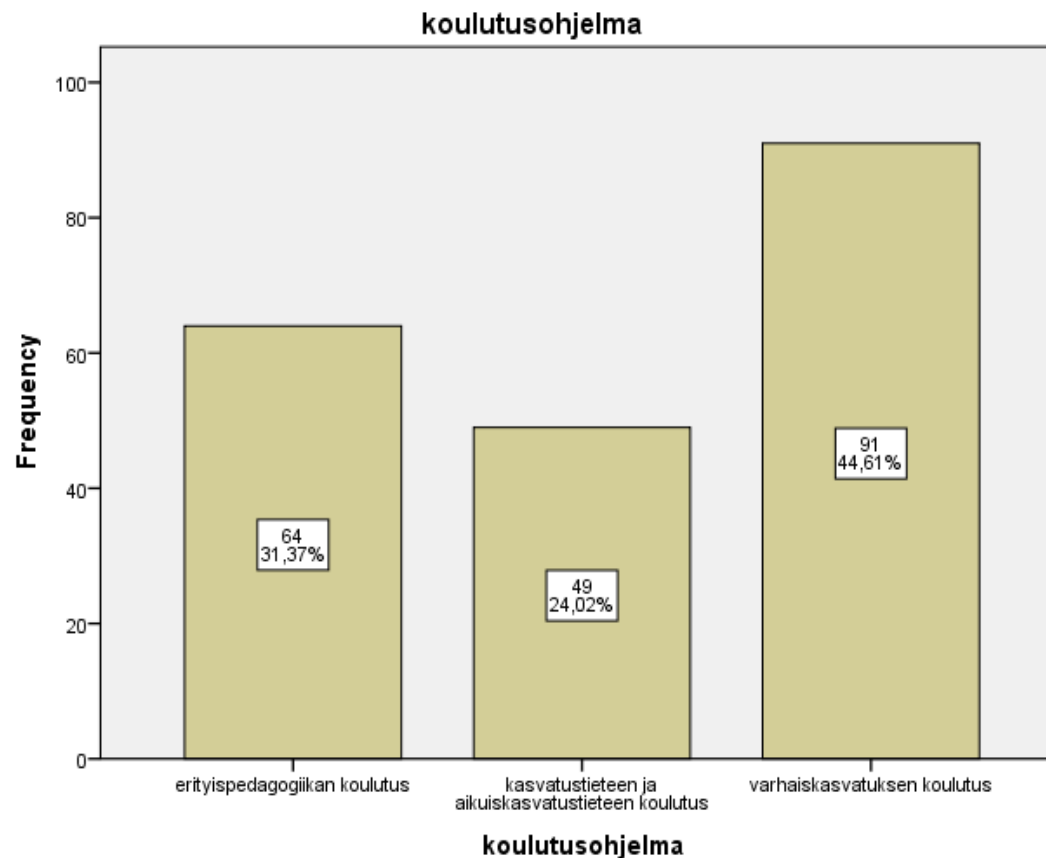


Figure 1. Distribution of respondents according to their gender (n = 44)

## Exercise 1b: Bar charts

### ■ How to report (Example)



[Example of report]  
“According to Figure 2, the most common major was early childhood education (91 participants, 44.6%). In addition, 64 participants (31.4%) were studying in special needs education and 49 participants (24.0 %) were in educational science and Adult education.”

Figure 2. Distribution of respondents according to their major (n = 204)



## SPSS output of pie chart

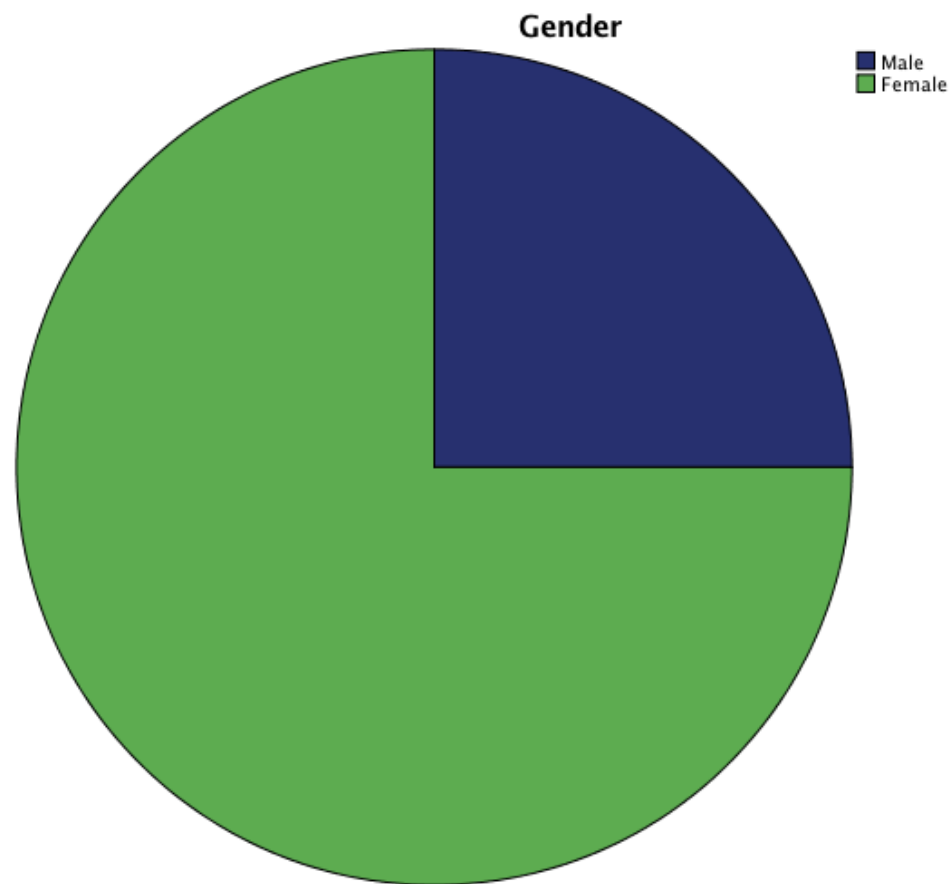
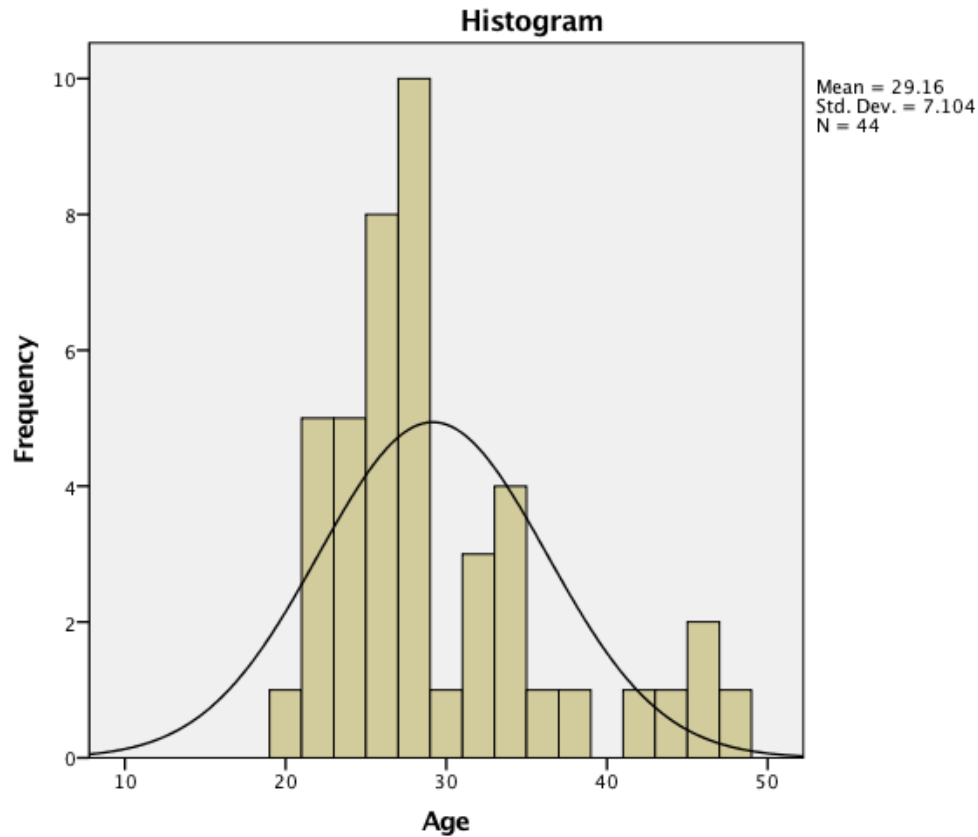


Figure 3. Distribution of respondents according to their gender (n = 44)

## Exercise 2a: Histogram

- Calculate the frequency distributions of “Age” and create a histogram.  
(in Handout p.3)



The shape of the histogram shows that the age of students is normally distributed.

Figure 4. Histogram of age variable (N = 44)

## Exercise 2b: Skewness and kurtosis

- View age-variable's skewness and kurtosis.  
(in Handout p.3)

**Descriptive Statistics**

	N Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Deviation Statistic	Skewness		Kurtosis	
						Statistic	Std. Error	Statistic	Std. Error
Age	44	20	48	29.16	7.104	1.212	.357	.823	.702
Valid N (listwise)	44								

If it is fully symmetrical, **skewness** and **kurtosis** will be zero values.  
In this data set, the former is 1.21 and the latter is 0.82.

- \*The cutoff values for skewness and kurtosis is debated among the researchers.  
E.g. 1)  $\pm 2$  for both  
2)  $\pm 0.8$  for skewness and  
 $\pm 3.0$  for kurtosis

## Exercise 2c: Measures of central tendency

- Calculate the mean, median, mode, standard deviation, and quartiles of "Age" variable. (in Handout p.3)

Statistics		
Age		
N	Valid	44
	Missing	0
Mean		29.16
Median		27.00
Mode		25 <sup>a</sup>
Std. Deviation		7.104
Percentiles	25	24.25
	50	27.00
	75	32.75

a. Multiple modes exist.  
The smallest value is shown

### Mean

- is the "average", which you add up all the numbers and then divide by the number of observations.

### Median

- is the "middle" value in the list of numbers or the average of the two middle values.

### Mode

- is the value that occurs most frequently.

### Standard deviation (SD)

- measures the clustering of numbers (the square root of the variance)
- The higher SD indicates that the data points are spread out over a wider range of values.

Which one you use will depend on a type of data you have, whether your data has outliers and/or is skewed, and what you want to show from your data.

## Exercise 2c: Measures of central tendency - How to report (Example)

Statistics		
Age		
N	Valid	44
	Missing	0
Mean		29.16
Median		27.00
Mode		25 <sup>a</sup>
Std. Deviation		7.104
Percentiles	25	24.25
	50	27.00
	75	32.75

a. Multiple modes exist.  
The smallest value is shown

Reference:  
<https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median-faqs.php>

### Mean

- "The mean age of the participants was 29.2 years old."
- The most frequently reported indicator, usually reported with standard deviation (*SD*).

### Median

- "The median age of the participants was 27 years old."

### Mode

- "The mode age of the participants was 25 years old."

### Standard deviation

- "The mean age of the participants was 29.2 years old (*SD*= 7.1)."

## Exercise 2d: Box plot

- Create box plot (in Handout p.3)

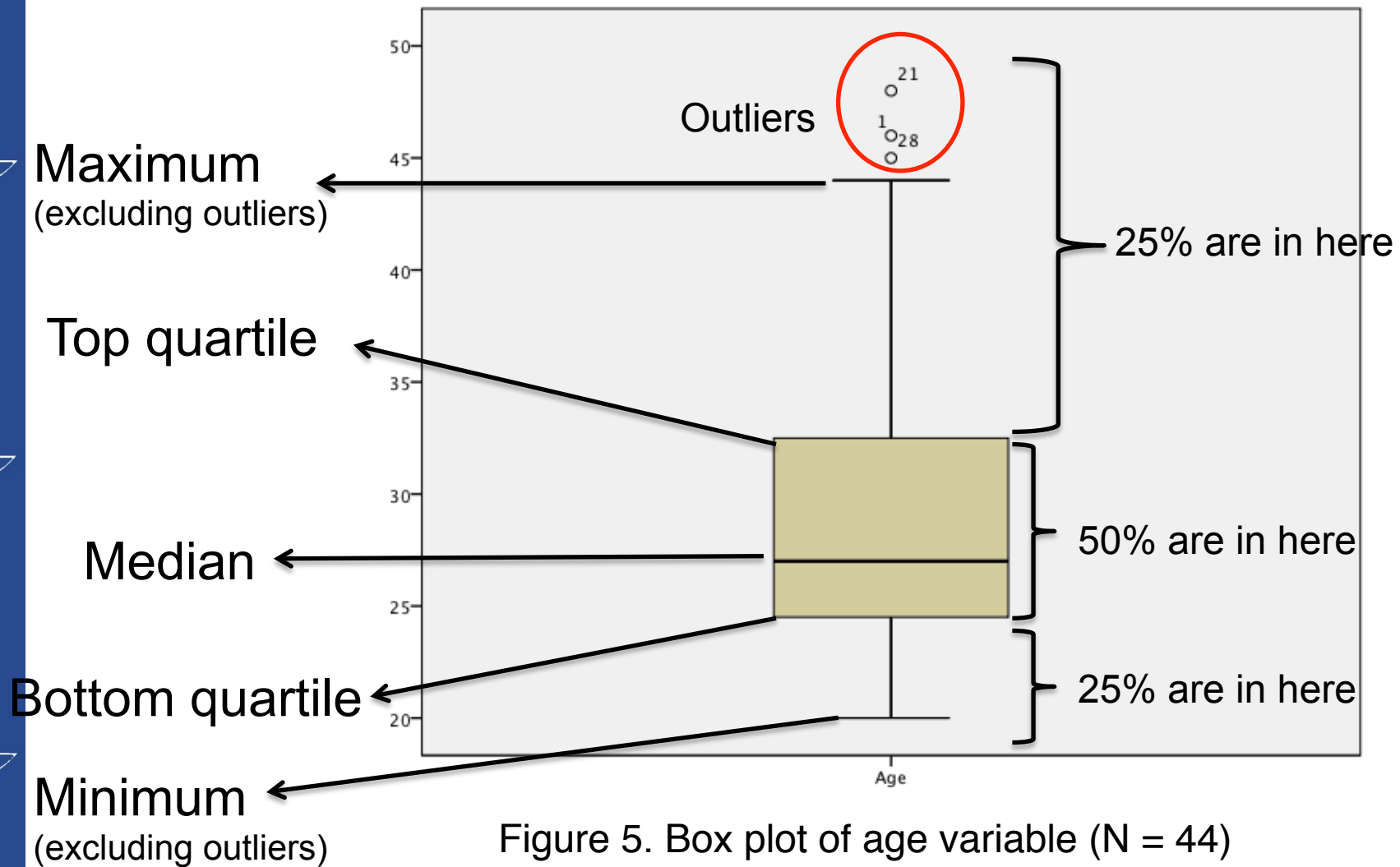


Figure 5. Box plot of age variable (N = 44)

# Exercise 2d: Quartiles

## Statistics

Age		
N	Valid	44
	Missing	0
Mean		29.16
Median		27.00
Mode		25 <sup>a</sup>
Std. Deviation		7.104
Percentiles	25	24.25
	50	27.00
	75	32.75

Bottom quartile

Median quartile

Top quartile

