Lab 1

Wiktor Soral

February 21 2017

Outline

- Course aims and structure
- How to pass the course?
- Data analysis workflow
- Data analysis tools
- Tidy data
- SPSS - recap

Course aims and structure

- The main goal of this course is to familiarize students with practical knowledge and theoretical background of more advanced data analysis methods.
- We will aim to the level where you will be able to perform by yourself analyses required for your empirical paper, master thesis, and papers in professional psychological journals.
- Each lab will begin with short introduction and discussion of the topic, next you will use your knowledge to analyse real word datasets.
- Home assignments will mostly consist of writing short empirical reports, when you will practice your knowledge even further.
- To preserve your knowledge you will write 2 exams (midterm and final)

How to pass the course?

- Students must respect the principles of academic integrity
- Attendance: 2 miss classes without excuse, 2 more classes in case of excuse (additional work), more than 4 will not pass
- Home assignments
- Midterm and final exam
- 30% * (home assignments) + 30% * (midterm score) + 40% * (final exam score) = total score
- Total score and both exam scores should be at least at the 50% level to pass the course.
- Grading scale: 95%+ = 5! 90-94% = 5 80-89% = 4.5 70-79% = 4 60-69% = 3.5 50-59% 59% = 3 below 50% = 2 (fail)

Data analysis workflow

- Collecting your data, ie. running a study
- Importing your data to stat. software and checking its integrity
- Assessing whether data are adequate to answer the research question
- Making necessary transformations, computations, filtering, etc.
- Building a statistical model that will account for the data
- Diagnosing problems with the model and the data
- Assessing whether statistical model fits the data
- If model accepted: Drawing conclusions from the model – if not repeat 3 last steps (but with different model)

- Decide whether to collect more data

Data analysis tools

- Graphical user interface, non-free: SPSS, Statistica, Stata, Excel
- Graphical user interface, free: JASP, PSPP
- Command line interface, non-free: Stata, SAS, S+, MATLAB
- Command line interface, free: R, some Python packages, Julia

Tidy data 1

- You have collected your questionnaires on gender, age, and self-esteem
- How should you put them together to work

gender	count
female	44
male	23

Tidy data 1

- You have collected your questionnaires on gender, age, and self-esteem
- How should you put them together to work

name	gender,age,SE
Dan	male,69,3.6
Brian	male,64,3.7
Sue	female,65,3.9
Hazel	female,68,3.8

Tidy data 2

- Each variable forms a column
- Each observation forms a row
- Each type of observational unit forms a table -> Each cell consists of 1 value from 1 subject

name	gender	age	self-esteem (1-4)
Dan	male	69	3.6
Brian	male	64	3.7
Sue	female	65	3.9
Hazel	female	68	3.8

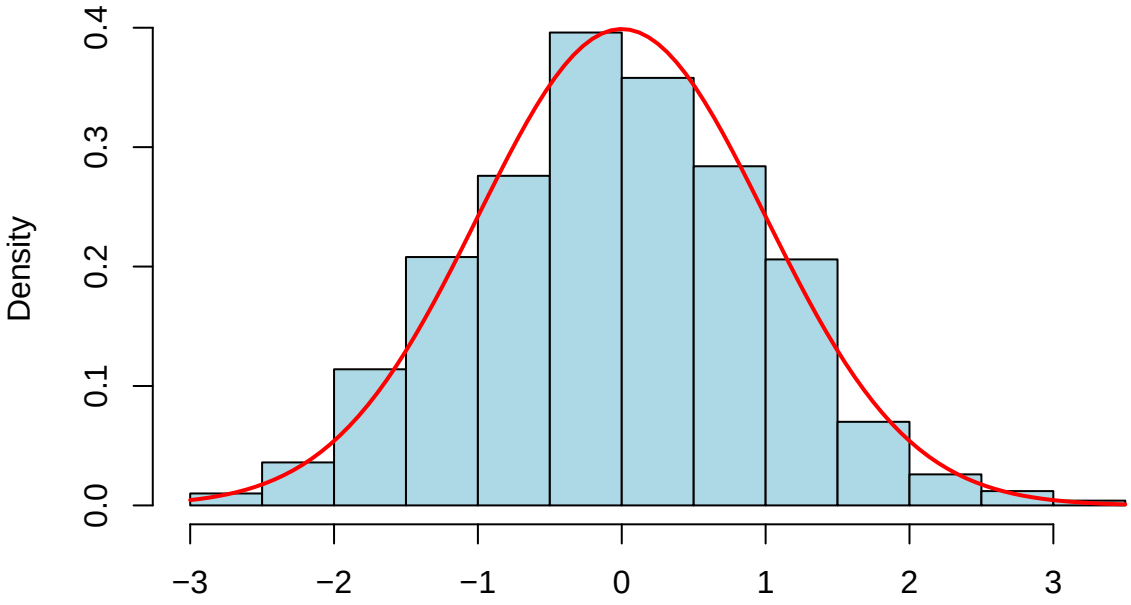
Tidy data 3

- Ordering of columns - first fixed values, then random values
- Missing values are identified

- Value meaning is indicated (where possible)
- Scale type is identified

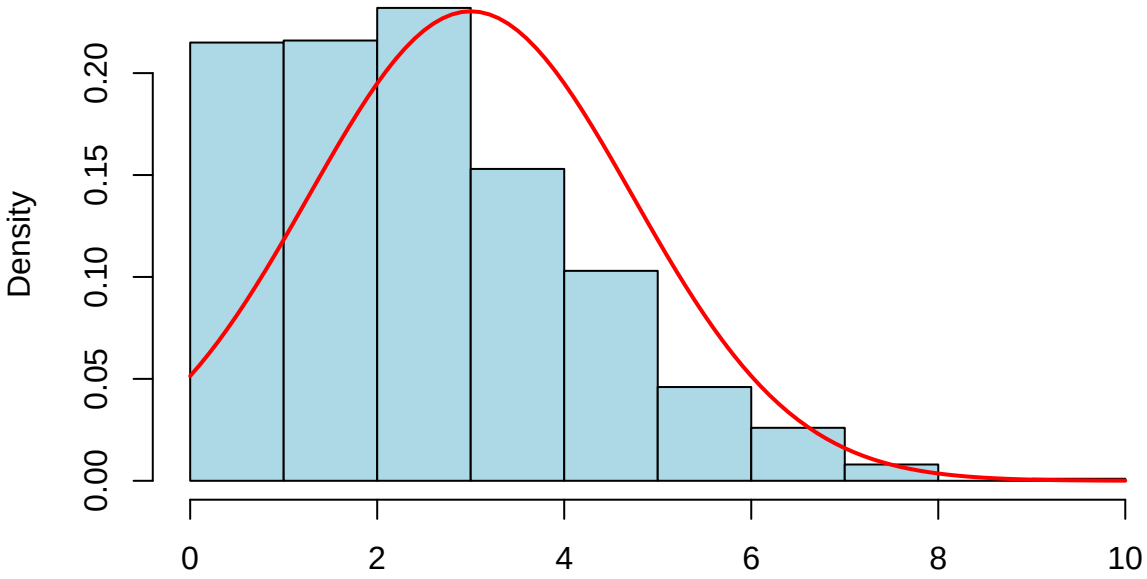
Assumptions 1

- Normally distributed data



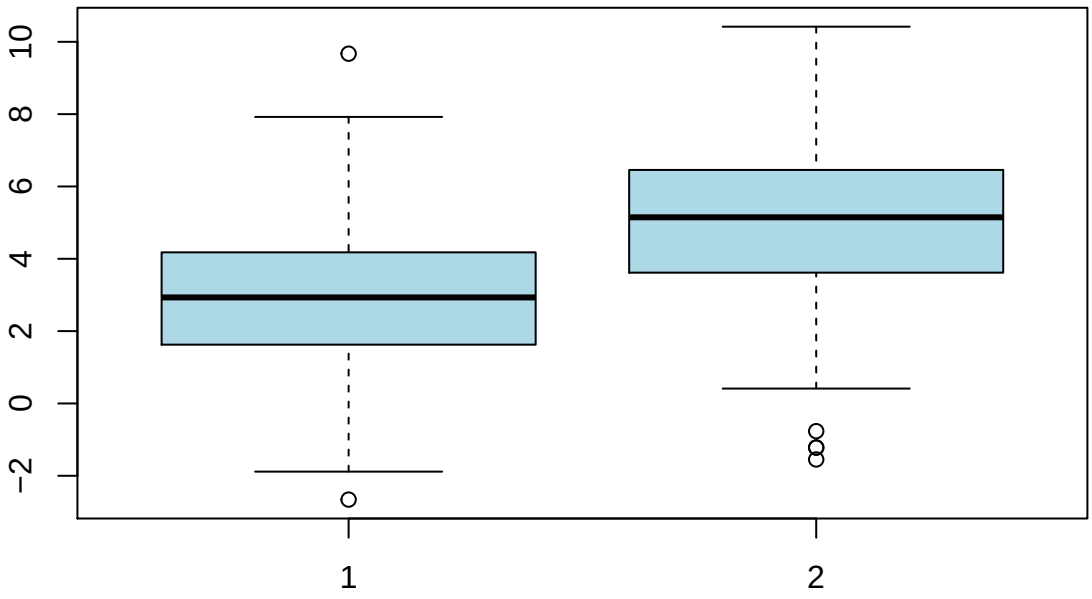
Assumptions 1

- Non-normally distributed data



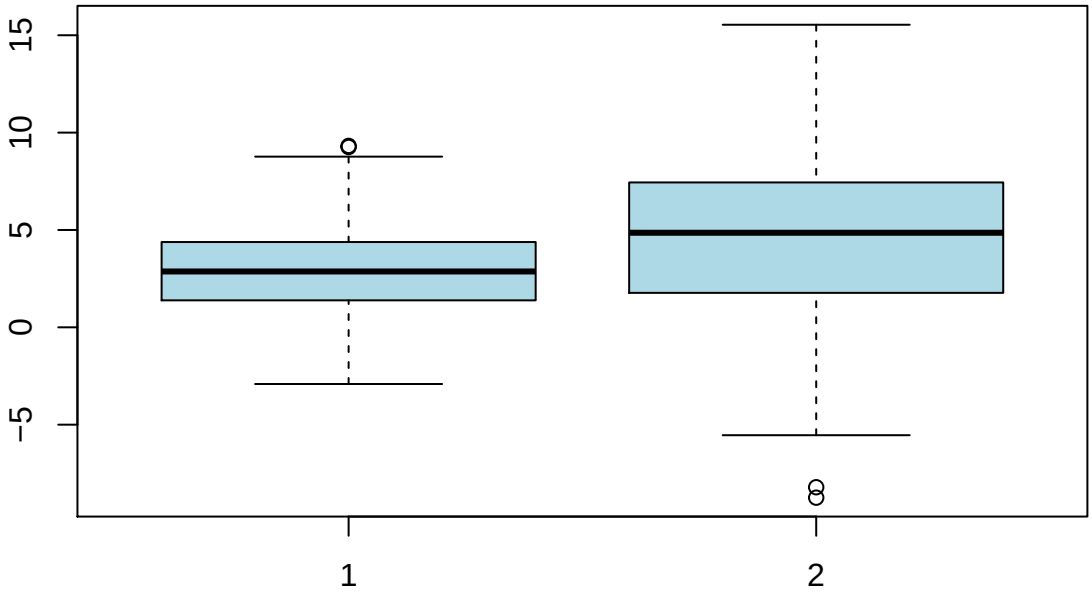
Assumptions 2

- Homogeneity of variance



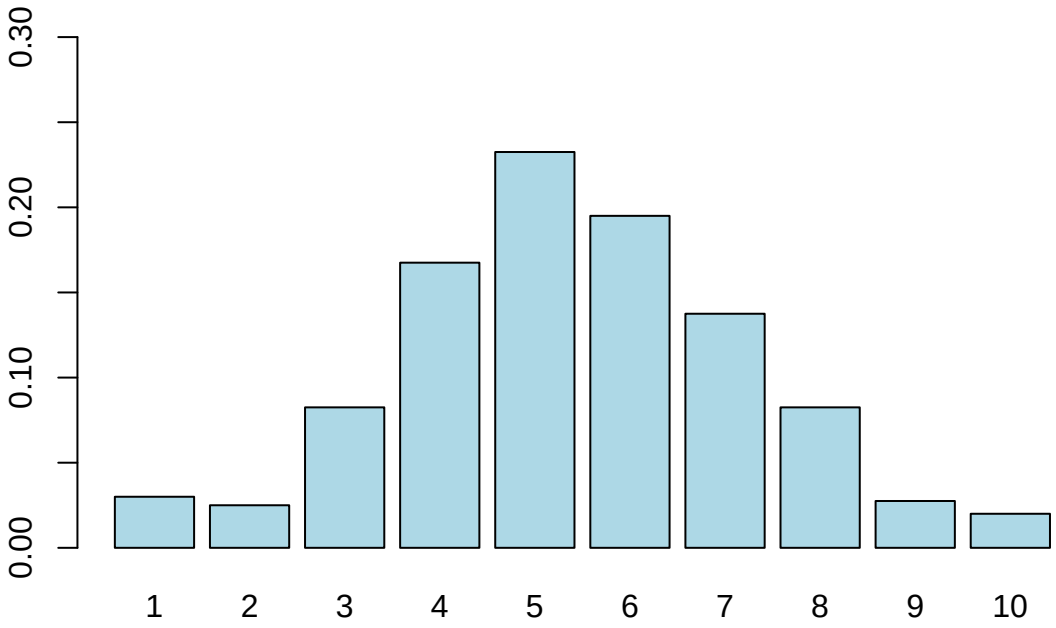
Assumptions 2

- Heterogeneity of variance



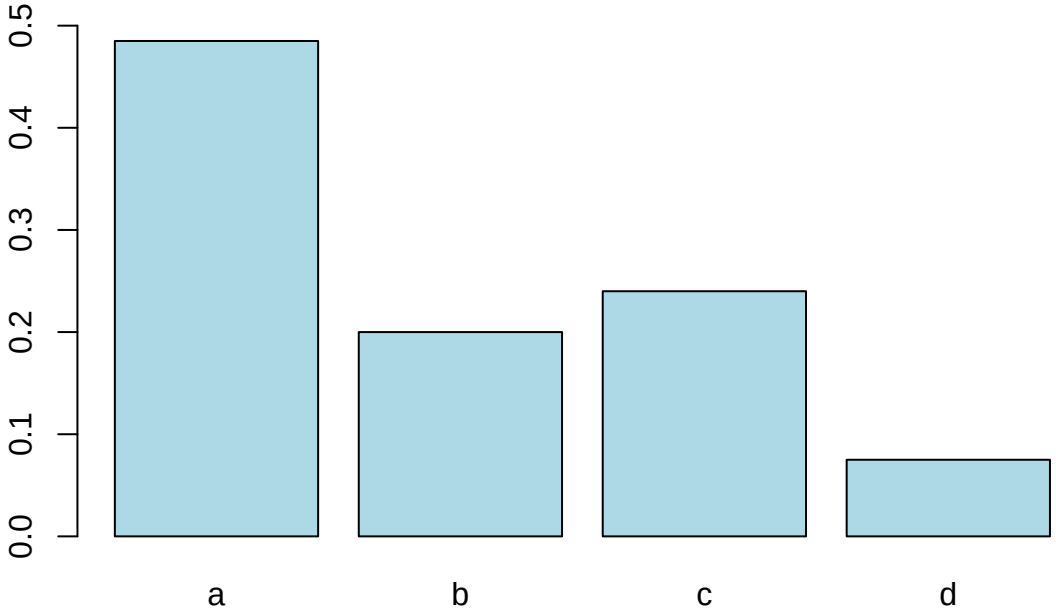
Assumptions 3

- Interval data



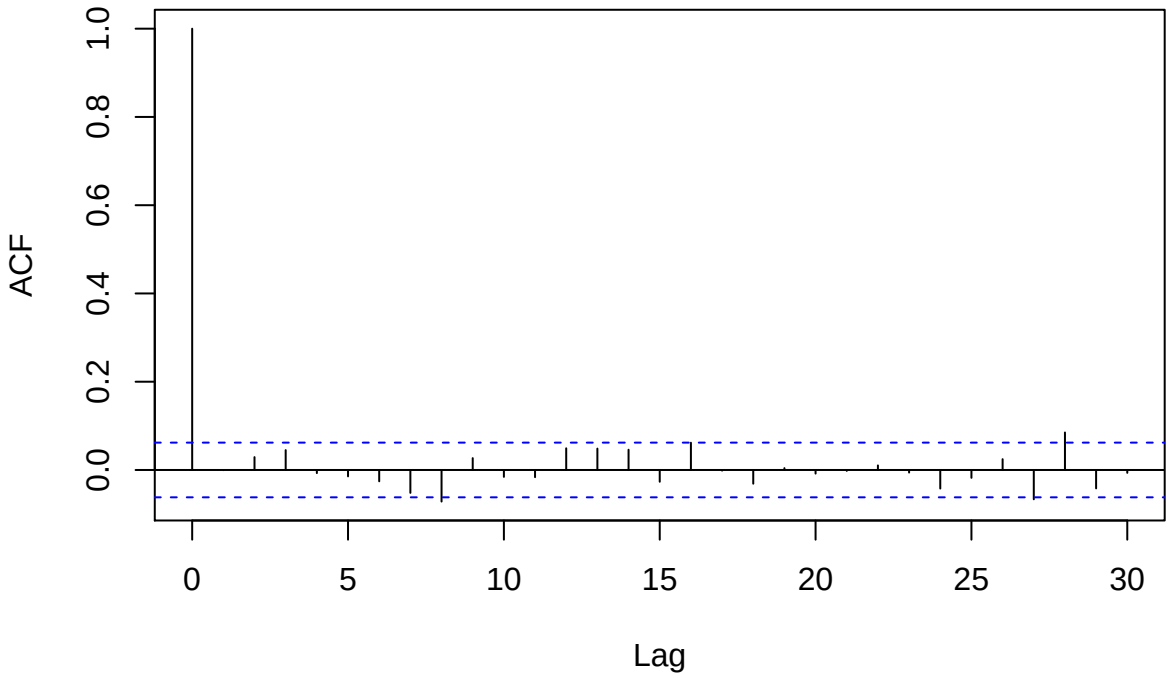
Assumptions 3

- Non-interval data



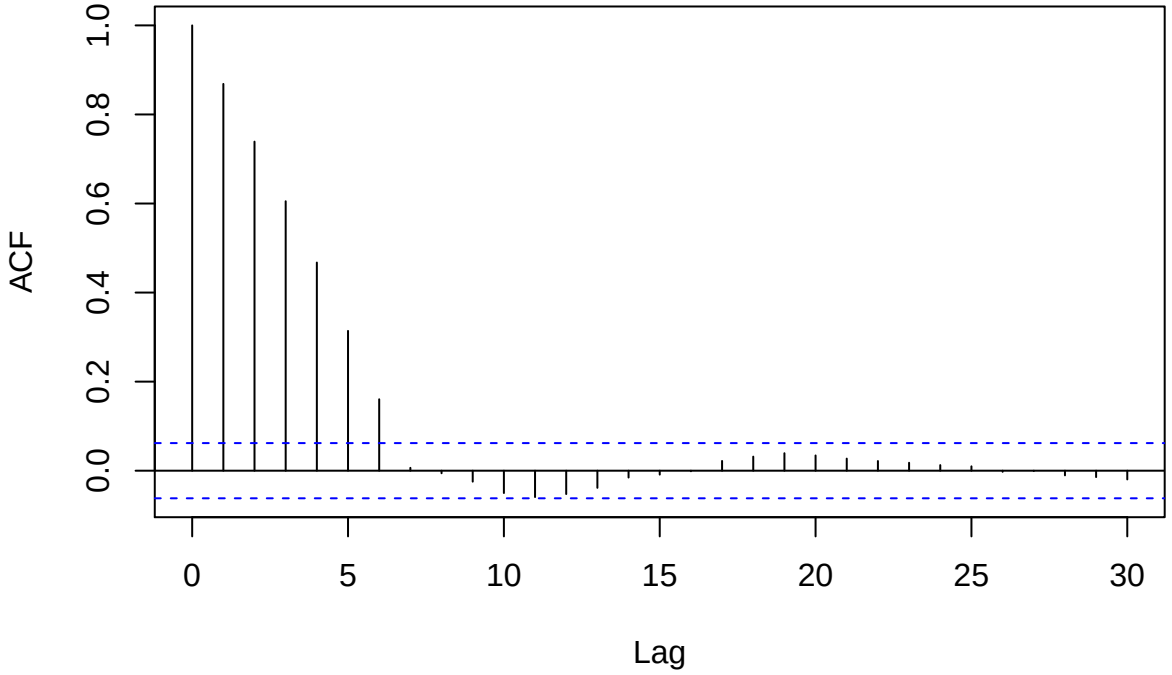
Assumptions 4

- Independence



Assumptions 4

- Non-independence



SPSS - recap

1. Selecting a subset of data, splitting data
2. Computing new values and recoding data

3. Making simple graphs
4. Checking normality with histogram and QQ-plot
5. Checking homogeneity of variance with boxplots
6. Checking independence of observations with auto-correlation plots
7. Detecting outliers