

# Experimental and Statistics Software

SUMMER TERM 2021

**Prof. Dr. Thomas-Patrick Schmidt**  
Faculty of Society and Economics

This presentation is only for internal use at Rhine-Waal University of Applied Sciences in the summer term 2021.

It is not permitted to record, pass on or publish the video and audio of this lecture without the written permission of the lecturer.

# EXPERIMENTAL AND STATISTICS SOFTWARE

## Lecture 2

### Dataset | Descriptive Statistics

# COURSE OUTLINE I

- Part I: Introduction to Data Analysis with R
  - Introduction to R & Basics in R [CW 16, 17]
  - **Descriptive Statistic [CW 18, 19]**
  - Hypothesis Testing & Normal Distribution [CW 21, 22]
  - Linear Regression [CW 23, 24]
- Part II: Applied Econometrics with R (we have to pick sth.) [CW 25, 26, 27]
  - Regression with Panel Data **or**
  - Regression with Binary Dependent Variable **or**
  - Instrumental Variable Regression **or**
  - Experiments and Quasi-Experiments

# COURSE OUTLINE II

Topic	Lecture	Case Study (+1)	Submission deadline (+1)	Points
Introduction to R & Basics in R	CW 16	CW 17	CW 18 [2021 – 05 – 09 23:59 CET]	10
<b>Descriptive Statistic</b>	<b>CW 18</b>	<b>CW 19</b>	<b>CW 20</b> <b>[2021 – 05 – 23 23:59 CET]</b>	<b>20</b>
Hypothesis Testing	CW 21	CW 22	CW 23 [2021 – 06 – 13 23:59 CET]	20
Linear Regression	CW 23	CW 24	CW 25 [2021 – 06 – 27 23:59 CET]	20
Applied Econometrics	CW 25 & 26	CW 27	CW 28 [2021 – 07 – 18 23:59 CET]	30

# DATA

## PRIMARY VS. SECONDARY DATA

- data sources
  - secondary data: someone else collected the data we analyze
  - primary data: we collect our own data
- we will use (simulated) primary data
  - idea: questionnaire of a ski resort / improve the service for the guests

# DATA QUESTIONNAIRE (I): VACATION

## 1. Accomodation

In which accommodation are you staying during your current ski holiday?

1.01 ☐ less than 2 star hotel

1.02 ☐ 2 star hotel

1.03 ☐ 3 star hotel

1.04 ☐ 4 star hotel

1.05 ☐ 5 star hotel

## 2. Duration

The total duration of your current ski holiday is: <sup>2</sup>    day(s)

**3. How satisfied are you with the following offers of our ski resort?** Please mark the satisfaction on the line with a cross, 0 means very unsatisfied and 100 very satisfied.

	0 very unsatisfied	100 very satisfied
Diversity of ski slopes:	3a	
Waiting time at ski lifts:	3b	
Quality of restaurants:	3c	
Satisfaction with the resort in total:	3d	

# DATA

## QUESTIONNAIRE (II): VACATION

### 4. Price/performance ratio

How do you think about the price of the ski pass?

4.01 ☐ expensive      4.02 ☐ appropriate      4.03 ☐ cheap

---

### 5. Expenses per day

How much money do you spend per day and guest, including accommodation?:    s    \_\_\_\_\_ CHF

---

### 6. Recommendation

Would you recommend our ski area to your friends and relatives?

6.01 ☐ yes      6.02 ☐ rather yes      6.03 ☐ rather no      6.04 ☐ no

---

### 7. Future ski holiday

Is it likely that you spend your skiing holiday in our ski resort again next year?

7.01 ☐ yes      7.02 ☐ no



# DATA QUESTIONNAIRE (III): DEMOGRAPHICS

8. Sex

8.00 ☐ male

8.01 ☐ female

---

9. Country of origin

9.01 ☐ CH

9.02 ☐

GER

9.03 ☐

AUT

9.04 ☐

other country

---

10. Age

I am...

10. 

--	--

years old

---

11. Highest level of education

11.01 ☐ Secondary school

11.02 ☐ A-Level

11.03 ☐ Bachelor

11.04 ☐ Master

# KEY FIGURES

## FREQUENTLY USED MEAN VALUES

- arithmetic mean: adding up all values and then dividing them by the number of observations
- median: value that divides the values of a variable into the 50% smallest and the 50% largest values
- mode: value that occurs most frequently

# KEY FIGURES

## IMPORTANT MEASURES OF VARIATION

- range: difference between the smallest and the largest value within a variable.
- quartile deviation: difference between the 1st quartile (25% of the values are smaller than the 1st quartile) and the 3rd quartile (75% of the values are smaller than the 3rd quartile) of a variable. The 50% of the middle values are within the quartile deviation.
- standard deviation: average deviation of the individual values from the mean value
- Variance: squared standard deviation

# KEY FIGURES

## CORRELATION COEFFICIENTS OFTEN USED

- metric variables: Bravais-Pearson correlation coefficient
- ordinal variables: for the Spearman correlation coefficient
- nominal variables: contingency coefficient

# STATISTICAL CHARTS

## WHY AND WHAT

- Why
  - analyze existing patterns and relationships
  - Visualize and present facts
- What
  - Histogram
  - Pie Chart
  - Bar Chart
  - Boxplot
  - Scatterplots

# CASE-STUDY

## EXERCISE 1

- a) Find the modus for the variables accommodation, education and stay and interpret the results. Do not forget to use the legend of the data, when interpreting the results.
- b) Display the median and the mean for the variables age, expenses and stay. Interpret the results.
- c) Calculate the standard deviation, the variance and the quantiles for the variables age, expenses and stay. Interpret the results together with the results of (b).

# CASE-STUDY

## EXERCISE 1

- d) Display the median, mean, standard deviation, variance, inter quartile range, the smallest value, largest value and the first and the third quartile for the variables diversity, quality, safety, satisfaction and waitingtime with one command and compare the results.
- e) Figure out how women and men feel about the waitingtime at the ski lifts. Is there a difference between them in the sample regarding the average feeling and the variation?

# CASE-STUDY

## EXERCISE 2

- a) Calculate the correlation coefficient of Bravais-Pearson between the two variables age and expenses. Interpret the result.
- b) Calculate the correlation coefficients of Bravais-Pearson between the variables age, expenses and stay using the correlation matrix. Interpret the results.
- c) Calculate the correlation coefficient of Spearman between the two variables accommodation and education. Interpret the result.
- d) Calculate the correlation coefficients of Spearman between the variables accommodation, recommendation and education using the correlation matrix. Interpret the results.



# CASE-STUDY

## EXERCISE 3

- a) Create a histogram for the variable expenses and interpret the graph briefly. Try to label the histogram, the x-axis and the y-axis and give the histogram the color purple.
- b) Create a pie chart for the variable education and interpret the pie chart briefly. Try to add the names to the pieces.
- c) Try to create a barplot that displays the guests by country and accommodation category. Label the barplot properly and interpret the result.
- d) Create a boxplot that shows the expenses by the country of the guests. Label the boxplot and interpret the results.

# CASE-STUDY

## EXERCISE 3

- e) Draw a scatterplot for the variables age and expenses. Label it and interpret the result.
- f) Try to draw a smooth line into the graph of task (e). Is there a linear relationship (straight line) between the two variables?