Franz Kronthaler
Silke Zöllner

# Data Analysis with RStudio

## An Easygoing Introduction

Recommended in Germany

Springer Spektrum

# Data Analysis with RStudio

Franz Kronthaler · Silke Zöllner

# Data Analysis with RStudio

## An Easygoing Introduction

Springer Spektrum

Franz Kronthaler,
University of Applied Sciences Grisons
Chur, Switzerland

Silke Zöllner
Institue of Business and Regional
Economics IBR
Lucerne University of Applied
Sciences and Arts
Lucerne, Switzerland

ISBN 978-3-662-62517-0 ISBN 978-3-662-62518-7 (eBook)
https://doi.org/10.1007/978-3-662-62518-7



Planung: Iris Ruhmann

This Springer Spektrum imprint is published by the registered company Springer-Verlag GmbH, DE part of Springer Nature.
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

# Comment

R is a powerful freely available open-source tool for analyzing data and creating graphs ready for publication. In just a few years, R has become the leading statistical software in science and is now becoming even more widespread in practice and in business. R can be used to analyze data and to generate knowledge for companies and institutions that they can include in their business decisions.

The objective of the text is to introduce R—specifically RStudio—to students from different fields of study and to practitioners and enable them to use R in their everyday work. The script is not a substitute for statistical textbooks. The focus lies on the use of RStudio for data analysis, but at the same time, also some statistical knowledge is conveyed. If someone feels the need to deepen the statistical knowledge, he or she should read a textbook of statistics. At the end of the script, various textbooks are briefly described.

The main purpose however is to hand over the joy of analyzing data with RStudio!

We would like to thank Irenaeus Wolff for his critical review of the script.

# Contents


**1 R and RStudio** . . . . . 1
- 1.1 A Note on How to Use the Script . . . . . 1
- 1.2 About R and RStudio . . . . . 1
- 1.3 How to Install R and RStudio . . . . . 2
- 1.4 The Structure of RStudio . . . . . 5
- 1.5 A First Data Analysis Application with RStudio . . . . . 7
- 1.6 How to Install RStudio Packages . . . . . 10

**2 Data Analysis Basics with RStudio** . . . . . 13
- 2.1 How to Read Data with RStudio . . . . . 13
- 2.2 How to Check Data with RStudio . . . . . 15
- 2.3 Creating and Modifying Variables and Selecting Cases with RStudio . . . 17
- 2.4 Commands and Command Structure in RStudio . . . . . 21
- 2.5 Script Files and Reporting . . . . . 23
- 2.6 Time to Try . . . . . 28

**3 Data Tourism (Simulated)** . . . . . 31

**4 Describing Data with RStudio** . . . . . 35
- 4.1 Descriptive Key Figures . . . . . 35
- 4.2 Statistical Charts . . . . . 41
- 4.3 Time to Try . . . . . 57

**5 Testing Normal Distribution with RStudio** . . . . . 59
- 5.1 Graphical Ways to Check for Normal Distribution . . . . . 59
- 5.2 Numerical Ways to Check for Normal Distribution . . . . . 61
- 5.3 Time to Try . . . . . 63

**6 Testing Hypotheses with RStudio** . . . . . 65
- 6.1 One-Sample t-Test . . . . . 66
- 6.2 Two-Sample t-Test Independent Samples . . . . . 67
- 6.3 Wilcoxon Rank-Sum Test . . . . . 68
- 6.4 Two-Sample t-Test Dependent Samples . . . . . 69

6.5 Wilcoxon Signed-Rank Test . . . 71
6.6 Analysis of Variance ANOVA . . . 71
6.7 Correlation Test for Metric, Ordinal and Nominal Variables . . . 82
6.8 Time to Try . . . 85

**7 Linear Regression with RStudio** . . . 87
7.1 Simple and Multivariate Linear Regression . . . 87
7.2 Regression Diagnostic with RStudio . . . 98
7.3 Time to Try . . . 106

**8 Further Reading** . . . 107

**9 Appendix** . . . 109
9.1 Appendix 1: Questionnaire . . . 109
9.2 Appendix 2: Dataset tourism.xlsx Including Legend . . . 109
9.3 Appendix 3: How to Deal with Missing Data . . . 112
9.4 Appendix 4: Solutions for the Tasks . . . 113