

DATENANALYSE MIT R

Liebe Teilnehmerinnen und Teilnehmer,

die bereitgestellten Lehrmaterialien basieren auf Lernkarteikarten () , die sich je nach Lektion auf die dazugehörigen Textbausteine () , Programmierbeispiele () und Lehrvideos () oder auf passende Übungsaufgaben () beziehen. Dadurch ist ein orts- und zeitunabhängiges Selbststudium aller relevanten Inhalte meiner Lehrveranstaltungen und Workshops an der Universität zu Köln, der RWTH Aachen, der FOM Hochschule sowie beim Netzwerk Terrorismusforschung möglich. Einfach auf die entsprechenden Icons klicken und Sie gelangen direkt zu den Lehrmaterialien.

Ihnen viel Spaß bei der Datenanalyse mit R!

Prof. Dr. Dennis Klinkhammer

(I) GRUNDLAGEN

| Zeit | Inhalt | Lehrmaterialien |
|---------------|--|---|
| 20 Min | Einführung in die wissenschaftlichen Gütekriterien , um auf eine adäquate Datenanalyse mit R vorzubereiten. |   |
| 15 Min | Das erste Lehrvideo bietet einen Einstieg in R und RStudio sowie die Grundlagen dieser statistischen Programmiersprache. |    |
| 10 Min | Mit dem TREES Datensatz lässt sich die Datenanalyse mit R ohne statistische Vorkenntnisse demonstrieren. |    |
| 30 Min | In der 1. Übungsaufgabe werden grundlegende Funktionen der Programmiersprache R vorgestellt und zur Anwendung gebracht. |   |
| 10 Min | Datenanalysen erfordern Analysemodelle , welche in RStudio mit dem Package DIAGRAMMER angelegt werden können. |    |
| 15 Min | Der SWISS Datensatz verdeutlicht die Bedeutung theoretisch fundierter Analysemodelle für die statistische Interpretation. |    |
| 30 Min | Die 2. Übungsaufgabe stellt eine Möglichkeit der Datenanalyse mit R vor, wenn einmal kein Analysemodell vorliegen sollte. |   |
| 20 Min | Mögliche Herausforderungen bei der Datengewinnung , die bei Datenanalysen mit R zu berücksichtigen sind. |   |

(II) FORMELSAMMLUNG

| Zeit | Inhalt | Lehrmaterialien |
|---------------|--|---|
| 30 Min | Eine praktische Übersicht über die in den nachfolgenden Lehrmaterialien ausgewiesenen statistischen Formeln . |   |



(III) DATENANALYSE

| Zeit | Inhalt | Lehrmaterialien | |
|--------|---|---|---|
| 20 Min | Für ein besseres Verständnis der statistischen Formeln empfiehlt sich ein Blick auf die Skalenniveaus . |  |  |
| 15 Min | In diesem Lehrvideo werden die univariate Statistik sowie die dazugehörigen Lage- und Streuungsmaße vorgestellt. |  |  |
| 20 Min | Die Standardabweichung ermöglicht die Überführung von Rohwerten in Normwerte, die sogenannten z-Werte . |  |  |
| 20 Min | Die bivariate Statistik bezieht sich mittels Korrelation, Chi-Quadrat-Test oder t-Test auf jeweils zwei Variablen. |  |  |
| 30 Min | In der 3. Übungsaufgabe wird der Chi-Quadrat-Test über ein anschauliches Beispiel mit dem Taschenrechner ausgerechnet. |  |  |
| 30 Min | Der t-Test kann im Rahmen der 4. Übungsaufgabe direkt in RStudio zur Anwendung gebracht werden. |  |  |
| 15 Min | Die multivariate Statistik wird am Beispiel der linearen und logistischen Regression vorgestellt. |  |  |
| 20 Min | Faktorenanalyse und Clusteranalyse ermöglichen eine Komplexitätsreduktion entlang der Variablen und Fälle. |  |  |
| 30 Min | Mit der 5. Übungsaufgabe wird die Komplexitätsreduktion unter Rückgriff auf den BFI Datensatz in RStudio erprobt. |  |  |

(IV) MACHINE LEARNING

| Zeit | Inhalt | Lehrmaterialien | |
|--------|---|---|---|
| 20 Min | In der theoretischen Einführung werden das Supervised und Unsupervised Machine Learning vorgestellt. |  |  |
| 10 Min | Die praktische Einführung bereitet auf die adäquate Verwendung von Trainings- und Validierungsdaten vor. |  |  |
| 15 Min | Das CARET Package stellt Algorithmen zum Classification and Regression Training in RStudio bereit. |  |  |
| 30 Min | Verschiedene Machine Learning Algorithmen können in der 6. Übungsaufgabe miteinander verglichen werden. |  |  |
| 10 Min | Resampling präzisiert statistische Befunde und ist Grundlage eines belastbaren Machine Learnings. |  |  |
| 20 Min | Übersicht über gängige Herausforderungen in der Anwendung von Machine Learning Algorithmen. |  |  |

The End

