






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

## (II) FORMELSAMMLUNG













Zeit	Inhalt	Lehrmaterialien
<b>30 Min</b>	Eine praktische Übersicht über die in den nachfolgenden Lehrmaterialien ausgewiesenen statistischen <b>Formeln</b> .	 



### (III) DATENANALYSE

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

### (IV) MACHINE LEARNING

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

*The End*

