

Study Guide

# R-Software und Tools für Financial Data Analytics

Dr. Lena Reh / Tino Werner

Wintersemester 2020/21



## INHALT

|   |    |
|---|----|
| Inhalt .....                                      | 2  |
| Kurzübersicht.....                                | 3  |
| Lehrende & Kontakt .....                          | 3  |
| Modulinhalte .....                                | 4  |
| Lernergebnisse .....                              | 4  |
| Modulkonzept .....                                | 5  |
| Modulteilnahme.....                               | 6  |
| Lernaktivitäten und Prüfungen .....               | 7  |
| Online-Aufgaben .....                             | 7  |
| Studienbegleitete benotete Prüfungsleistung ..... | 8  |
| Feedback & Noten.....                             | 8  |
| Modulabschluss.....                               | 8  |
| Wissenschaftliches Arbeiten .....                 | 9  |
| Modulablauf .....                                 | 10 |
| Fachlicher Aufbau des Moduls .....                | 13 |
| Thematische Einführung .....                      | 15 |
| Verzeichnis der weiterführenden Literatur.....    | 21 |
| Impressum.....                                    | 22 |

## MODULINHALTE

---

Der enorme Preisverfall im Bereich der Massenspeicher hat das Speichern großer Datenmengen attraktiv gemacht. Durch die Verfügbarkeit dieser großen Datenmengen und entsprechender Rechenleistung wurde es möglich, diese Datenmengen quantitativ auszuwerten, was wiederum einen Bedarf an (noch) größeren Datenmengen zur Folge hatte. Das Zusammenspiel aus der Aufbereitung und Auswertung von Daten bezeichnet man heutzutage als Data Analytics.

Dieser Kurs gibt eine Einführung in Data Analytics anhand der Statistiksoftware R. Das Modul deckt dabei die folgenden Themen ab:

- Einlesen der Daten,
- Datenaufbereitung,
- Explorative Datenanalyse,
- Auswahl und Anpassung von Modellen,
- Visualisierung der Daten und Ergebnisse sowie
- Berichten der Ergebnisse.

Im Mittelpunkt steht dabei die praktische Erfahrung mit der Umsetzung der grundlegenden Methoden der Angewandten Statistik. Die Bearbeitung der Themen erfolgt in Übungen und eigenständigen Projekten anhand von Beispielen aus dem Finanzbereich.

## LERNERGEBNISSE

---

Nach erfolgreichem Abschluss des Moduls »R-Software und Tools für Financial Data Analytics« sind die Teilnehmenden in der Lage:

- Daten aus verschiedenen Quellen zu importieren,
- das in R zur Verfügung stehende Grundinstrumentarium zu beherrschen und anzuwenden,
- durch Eigenrecherche mit Ergänzungsinfrastruktur das Grundinstrumentarium von R zu erweitern,
- Umsetzung und kritisches Hinterfragen der grundlegenden Konzepte aus dem Modul Quantitative Methoden umzusetzen und kritisch zu hinterfragen,
- Daten und Analysen in R zu visualisieren und
- Output und Ergebnisse zu präsentieren.

## FACHLICHER AUFBAU DES MODULS

---

Der fachliche Inhalt des Moduls »R-Software und Tools für Financial Data Analytics«, den Sie während der Online-Präsenzen und der Web-Seminare vermittelt bekommen, gliedert sich in die folgenden Schwerpunkte/Kapitel:

### 1. Grundlagen in R

Im ersten Abschnitt behandeln wir die ersten Schritte mit R und RStudio. Wir vertiefen die Verwendung von Paketen, wie sie im letzten Abschnitt des Study Guides weiter unten bereits erläutert wird, und klären, wie man bei Fragen Hilfe erhalten kann.

### 2. Datentypen

Es folgt die Einführung in die grundlegenden Datenstrukturen in R und wie diese in R importiert werden können. Daneben wird das Aufrufen von Funktionen erläutert.

### 3. Graphik in R

Abschnitt 3 stellt vor, wie Graphiken erstellt werden können. Neben den allgemeinen Grundlagen sowie den relevanten Funktionen des Basisumfangs von R behandeln wir den Umgang mit dem Zusatzpaket ggplot2.

### 4. Explorative Datenanalyse und Simulation in R

In Abschnitt 4 behandeln wir die Verwendung von grundlegenden empirischen Kenngrößen, theoretischen Verteilungen sowie die Simulation von Zufallsvariablen.

### 5. Testen und Schätzen in R

Wir vertiefen verschiedene Schätzverfahren (beispielsweise ML-Schätzer oder Momenten-Schätzer), Hypothesentests (beispielsweise Gauß-, t- oder  $\chi^2$ -Tests) und Konfidenzintervalle. Als Anwendungsbeispiel behandeln wir die Illustration des Gesetzes der großen Zahlen und des zentralen Grenzwertsatzes.

### 6. Regression in R

Abschnitt 6 stellt die Syntax zur Formulierung (verallgemeinerter) linearer Regressionsmodelle, die Rückgabestructur in R und sowie deren Interpretation vor.

### 7. Etwas Programmierung

In Abschnitt 7 behandeln wir die Grundlagen der Programmierung in R: Kontrollstrukturen (wie etwa For-Schleifen aber auch deren Vermeidung), das Schreiben von eigenen Funktionen in R und eigene Datenstrukturen.

### 8. Nachhaltige Software

Wir schließen das Modul ab mit einem Ausblick in die Erstellung von Dokumentation von Analysen und eigenen R-Paketen, Einsatz von Versionierungstools und der Anbindung an externe Schnittstellen.

Im Rahmen der Online-Präsenzen und der Web-Seminare werden wir regelmäßig auf grundlegende Konzepte aus dem Modul Quantitative Methoden verweisen. Die zu berechnenden Größen beziehen sich in der Regel auf grundlegende Konzepte des Moduls Quantitative Methoden. Beide Module können aber unabhängig voneinander belegt werden. Sie lernen je nach Reihenfolge der Belegung entweder erst die Anwendung oder die mathematischen Hintergründe der Konzepte.

## THEMATISCHE EINFÜHRUNG

### 1 Data Analytics

Praktisch betrachtet, kann man Data Analytics oder Datenanalyse in fünf Schritte einteilen:

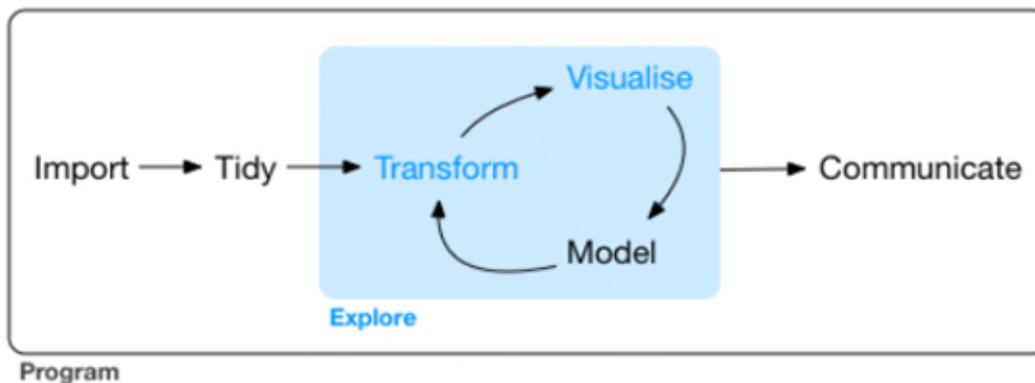


Abbildung 1: Datenanalyse in fünf Schritten, Golemund & Wickham Seite IX

- (1) Zuerst werden die Daten eingelesen (*import*), also (in unserem Fall in R) verfügbar gemacht bzw. geladen. In der Regel heißt das, dass Sie gespeicherte Daten aus Dateien, Datenbanken oder Web-APIs übernehmen und in einen Dataframe in R laden.
- (2) Nachdem Sie die Daten importiert haben, folgt das Bereinigen oder Aufbereiten (*tidy*) der Daten. Dabei speichert man die Daten in einer konsistenten Form, so dass jede Spalte einer Variablen und jede Zeile einer Beobachtung entspricht.
- (3) Im nächsten Schritt werden die Daten oftmals ein erstes Mal transformiert (*transform*). Dies beinhaltet das Einengen auf interessierende Beobachtungen (beispielsweise alle Daten aus einem bestimmten Zeitraum), Erstellen neuer Variablen, die Funktionen von vorhandenen Variablen sind (beispielsweise Berechnung der Jahresgesamtschäden) sowie das Berechnen von zusammenfassenden Statistiken (wie beispielsweise Anzahlen und Mittelwerte).
- (4) Ein nächster wesentlicher Schritt ist das Visualisieren der Daten (*visualise*). Hierbei können sich neue Fragen auftun oder Sie erkennen, dass Sie andersartige Daten erfassen müssen. Schließlich folgt das Modellieren (*model*) oder das Prüfen von Hypothesen. Die drei Schritte – Aufbereiten, Visualisieren und Modellieren – stehen dabei in keiner starren Abfolge, sondern können sich aufeinander aufbauend immer wieder abwechseln.
- (5) Der letzte und absolut entscheidende Teil der Datenanalyse ist das Kommunizieren (*communicate*) der Ergebnisse der Analysen. Den Rahmen für alle diese Tools bildet die Programmierung (*program*), mit der Sie in jedem Schritt der Datenanalyse zu tun haben werden.