

Der R Commander in der Lehre

Ein Erfahrungsbericht

Marc Höglinger
WIG – Winterthurer Institut für Gesundheitsökonomie
marc.hoeglinger@zhaw.ch

27. Juni 2018

Statistik in der Lehre: FH, Konsekutiv-Master, Sozialwissenschaften/BWL

- Ziele

- Interpretieren und kritisches Nachvollziehen können von statistischen Analysen in Berichten und Artikeln
- solides Verständnis grundlegender statistischer Konzepte
- Durchführen können einfacher statistischer Analyseverfahren (t-Test und einige nicht-parametrische Tests, ANOVA, einfache Regression)

- Restriktionen

- beschränkte Zeit und Workload für Studierende
- beschränkt programmier-affine Personen
- beschränkte mathematische/theoretisch-statistische Grundlagen
- teilweise beschränkte Motivation/Einsicht in die Bedeutung von Statistik für das Studium

Studiengang: Konsekutiv-Master Public/Non-Profit Management/Management im Gesundheitswesen, Gesundheitsökonomie

Klassischer Einstiegskurs in die angewandte statistische Datenanalyse: Grundlegende Konzepte, t-Test, einige nicht-parametrische Tests, lineare Regression, kurz: logistische Regression, erweiterte Anwendungen wie RDD.

Angewandte Datenanalyse

In diesem Kurs lernen Sie, Daten für Ihre Zwecke zu nutzen: Daten korrekt zu analysieren, Zufallsbefunde von systematischen Mustern zu unterscheiden und statistische Analysen korrekt zu interpretieren.

In der Vorlesung und in den Übungen, auf Moodle und in den angegebenen Texten finden Sie einige wertvolle Hilfsmittel dazu.



Kursinhalt und -agenda

Überblick zum Kurs, den Übungen, Statistik-Software, Prüfungsinformationen, etc.



Lehrbuch Datenanalyse mit dem R Commander von Kronthaler zum Lösen der Übungen

Kronthaler, F. (2016). *Statistik angewandt. Datenanalyse ist (k)eine Kunst mit dem R Commander*.

Dieses Buch beantwortet (fast) alle Fragen zum R Commander, die beim Lösen der Übungen auftauchen. Zudem erklärt es die meisten zentralen Konzepte und statistischen Tests, die wir anschauen.

<https://moodle.zhaw.ch/course/view.php?id=19427>

Rolle eines Statistik-Programms in der Lehre

- soll helfen, Verständnis für statistische Konzepte und Analyseverfahren zu vertiefen
- möglichst steile Lernkurve für einfache Verfahren, kein grosser Aufwand per-se
- keine Ablenkung vom Inhaltlichen
- soll eigene Analysen ermöglichen (z.B. im Rahmen der Masterarbeit, aber auch später im Beruf)
- idealerweise ausbaubar für Studierende, die sich stärker mit statistischer Datenanalyse auseinandersetzen
- Im Gebrauch bis jetzt an der ZHAW (und in vielen sozialwissenschaftlichen Studiengängen, z.B. auch ETH/UZH/UniBE): SPSS, Stata, Excel

R Commander

- R Commander ist ein Menü-basiertes Interface für einen leichten Einstieg (GUI) ins Statistik Programm R. Es ist ebenfalls Open Source.
- R Commander wird als library in R installiert und aufgerufen:
`library(Rcmdr)`
- Geschrieben wurde es 2002 von John Fox, einem Soziologen, Autor eines Lehrbuchs für statistische Datenanalyse (Fox 2016), und eines Handbuchs für den R Commander (Fox 2017, nächste Seite).
- R Commander wird in einigen Lehrbüchern vorgeschlagen/berücksichtigt, u.a. in Kronthaler 2016, Field et al. 2012,
- <https://socialsciences.mcmaster.ca/jfox/Misc/Rcmdr/>

R Commander

The screenshot displays the R Commander interface, which is a graphical user interface for the R programming language. It consists of three main windows:

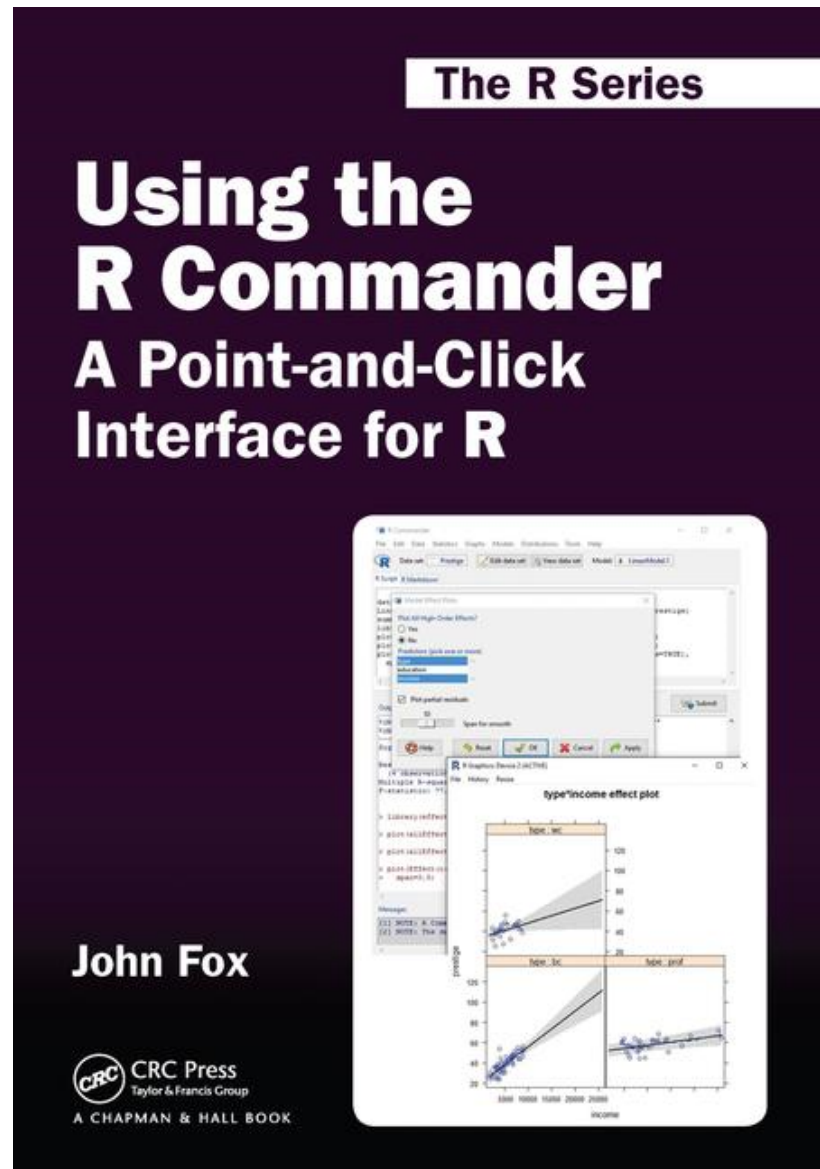
- R Console:** Shows the R command prompt with the following text:

```
Tippen Sie 'q()', um R zu verlassen.  
  
> library(Rcmdr)  
Lade nötiges Paket: splines  
Lade nötiges Paket: RcmdrMisc  
Lade nötiges Paket: car  
Lade nötiges Paket: sandwich  
Lade nötiges Paket: effects  
Lade nötiges Paket: carData  
  
Attache Paket: 'carData'  
  
The following objects are attached to the 'global' environment:  
  
  Guyer, UN, Vocab  
  
lattice theme set by effects  
See ?effectsTheme for details  
  
Rcmdr Version 2.4-2  
  
Warnung in file(.logFileName) bei:  
  kann Datei 'logfilename'  
Error in file(.logFileName, "w") :  
  cannot open file '.logfilename': No such file or directory  
>
```
- R Commander:** The main window with a menu bar (Datei, Bearbeiten, Datenmanagement, Statistik, Grafiken, Modelle, Verteilungen, Extras, Hilfe) and a toolbar. The "Datenmatrix" tab is active, showing the "Dataset" button. The "Modell" field displays "<Kein aktuelles Modell>". The "R Skript" tab is also visible, containing the following R code:

```
# Sie haben vorläufige Werte bereits den Gesundheitszustand von RaucherInnen  
# und NichtraucherInnen verglichen. Führen Sie nun einen formalen Test (t-Test)  
# durch zur Beantwortung der Frage, ob sich die beiden Gruppen tatsächlich  
# unterscheiden.  
  
# Datensatz laden  
load("C:\Users\hoeg\OneDrive - ZHAW\Lehre\2018FS-WAQM\uebungen\ISSP-health-Swiss-selecti  
# t-test (für unabhängige Stichproben) für die Variable "health" (subjektive Gesundheits  
# RaucherInnen vs. Nicht-RaucherInnen (Variable "smoking")  
  
# Grafik für Mittelwertvergleich  
with(Dataset, plotMeans(BIRTH, smoking, error.bars="conf.int", level=0.95,  
+ connect=TRUE))  
  
# Interpretation: Die Hypothese muss verworfen werden. RaucherInnen sind  
# mit einem durchschnittlichen Jahrgang von 1968 um 8 Jahre (95%-KI: 10.5;6.2)  
# jünger als NichtraucherInnen mit 1960 (P-Wert <0.01).
```
- R Graphics:** The "Device 2 (ACTIVE)" window displays a "Plot of Means". The plot shows the mean birth year for two groups: "Nicht-Raucher" (Non-smokers) and "Raucher" (Smokers). The y-axis is labeled "BIRTH" and ranges from 1960 to 1970. The x-axis is labeled "smoking". The plot shows a significant difference between the two groups, with the "Raucher" group having a higher mean birth year (approximately 1968) compared to the "Nicht-Raucher" group (approximately 1960). The plot includes error bars representing the 95% confidence interval.

Handbücher und Einstiegshilfen

Vom Autor des R Commanders



Getting Started With the R Commander*

John Fox and Milan Bouchet-Valat

Version 2.4-0 (last modified: 2017-01-24)

1 Introduction

The **R Commander** (Fox, 2005, 2017) provides a graphical user interface (“GUI”) to the open-source **R** statistical computing environment (R Core Team, 2016). This manual is a brief, basic introduction to the **R Commander**; for more extensive documentation, see Fox (2017), which has a website at socserv.mcmaster.ca/jfox/Books/RCommander/ or tinyurl.com/RcmdrBook.

R is a command-driven system, and new users often find learning **R** challenging. This is particularly true of those who are new to statistical methods, such as students in basic-statistics courses. By providing a point-and-click interface to **R**, the **R Commander** allows these users to focus on statistical methods rather than on remembering and formulating **R** commands. Moreover, by rendering the generated commands visible to users, the **R Commander** has the potential for easing the transition to writing **R** commands, at least for some users. The **R Commander**, however, accesses only a small fraction of the capabilities of **R** and the literally thousands of **R** packages contributed by users to the Comprehensive R Archive Network (CRAN). The **R Commander** is itself extensible through plug-in packages, and many such plug-ins are now available on CRAN (see Section 6.4 of this document).

This document directly describes the use of the **R Commander** under the **Windows** version of **R**. There are small differences in the appearance and use of the **R Commander** under **Mac OS X** and on **Linux** and **Unix** systems. Information about installing the **R Commander** on these platforms is available by following the link to the installation notes at the **R Commander** web page socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/index.html or at tinyurl.com/Rcmdr.

We use the following typographical conventions in this document: Names of software, such as **Windows**, **R**, the **Rcmdr** package, and the **R Commander**, are set in **boldface** type. The names of GUI elements such as menus, menu items, windows, and dialog boxes, are set in *italic* type. Variable names, names of data sets, and **R** commands are set in a **typewriter font**.

Review des R Commanders bzw. des Handbuchs im American Statistician

Using the R Commander: A Point-and-Click Interface

for R. John Fox. Boca Raton, FL: Chapman & Hall, 2016, xiv+219 pp., \$59.95(P), ISBN: 978-1-49-874190-3.

In this welcome addition to the Chapman & Hall/CRC R Series, John Fox provides a complete guide to the installation and use of his R point-and-click graphical user interface—the R Commander. The R Commander package allows students in a basic or intermediate statistics class to access the wide variety of tools in R without the increased cognitive load of learning R commands. As Fox writes in the preface of the book, “one doesn’t want a basic

statistics course to devolve into an exercise in learning how to write commands for statistical software, letting the software tail wag the statistical dog” (p. xiii). Though the software was originally intended for the introductory course, R Commander offers an array of more advanced statistical methods such as generalized linear models and nonparametric tests. Thus, the program is also practical for researchers and scientists who would like to perform statistical analyses without learning a command-based software. Since it displays the R commands in an R script file, R Commander can also be used as a starting point to learning R code.

Handbücher und Einstiegshilfen

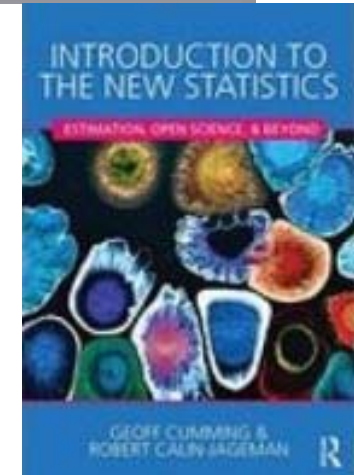
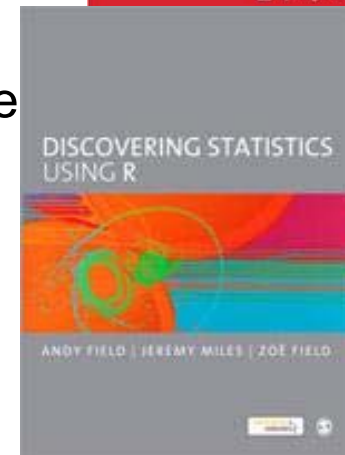
Auf Deutsch



- Kronthaler, Franz. 2016. Statistik angewandt. Datenanalyse ist (k)eine Kunst mit dem R Commander
- als e-Book frei für ZHAW verfügbar!!!

Was Sie für diese Veranstaltung brauchen

- R Commander-Anleitung
 - Kronthaler, Franz. 2016. Statistik angewandt. Datenanalyse ist (k)eine Kunst mit dem R Commander
- Statistik-Buch/Text
 - Kronthaler, Franz. 2016. Statistik angewandt. Datenanalyse ist (k)eine Kunst mit dem R Commander
 - Zusatzliteratur für einzelne Kapitel, auf Moodle abrufbar, insbesondere
 - Field, Andy, Zoë Field and Jeremy Miles. 2012. Discovering Statistics Using R.
 - Cumming, Geoff and Robert Calin-Jageman. 2017. Introduction to the New Statistics. Estimation, Open Science, and Beyond.
- Internet/Google → zum Lösen spezifischer Probleme bei der Analyse



R Commander (Rcmdr)

Methode der kleinsten Quadrate (lineare Regression)

The screenshot shows the R Commander interface with the 'Datenmanagement' menu open. The 'Importiere Daten' option is selected, leading to a submenu where 'aus Excel-Datei ...' is highlighted. A dialog box 'Excel-Datenmatrix importieren' is open, showing the name 'Daten_Alkohol' entered in the 'Name für Datenmatrix:' field. Below this, there are checkboxes for 'Variablenamen in erster Zeile des Spreadsheet' (checked), 'Zeilennamen in erster Zeile des Spreadsheet' (unchecked), and 'Konvertiere Zeichen-Variablen in Faktoren' (checked). The 'Zeichen für fehlenden Wert:' field is set to '<leere Zelle>'. At the bottom of the dialog are buttons for 'Hilfe', 'OK', and 'Abbrechen'. To the right, a file explorer window is open, showing a list of files. The file 'Daten_Alkohol.xlsx' is highlighted. Below the file explorer, the 'Dateiname:' field is empty, and the file type is set to 'Alle Dateien (*.*)'. The 'Öffnen' button is visible.

1: > Datenmanagement
> Importiere Daten
> aus Excel-Datei ...

2: Beliebigen Namen eingeben

3: Datei
Daten_Alkohol.xlsx
auswählen

11

R Commander (Rcmdr)

Methode der kleinsten Quadrate (lineare Regression)

1: > Datenmatrix betrachten

2: Alles OK?

The screenshot displays the R Commander interface. The 'Datenmatrix' tab is active, showing the loaded data 'Daten_Alkohol'. The 'Datenmatrix betrachten' button is highlighted with a red box. The 'Ausgabe' pane shows the R code used to load the data, and the 'Meldungen' pane shows a message indicating the data matrix has 50 rows and 7 columns.

```
Daten_Alkohol <-  
readXL("U:/UserHome-ZEUS/5 Lehre/MSc WAQM/2018 FS/Vorlesung und Übung/Daten ICPSR/Daten_Alk  
rownames=FALSE, header=TRUE, na="", sheet="Sheet1", stringsAsFactors=TRUE)
```

```
> Daten_Alkohol <-  
+ readXL("U:/UserHome-ZEUS/5 Lehre/MSc WAQM/2018 FS/Vorlesung und Übung/Daten ICPSR/Daten_1  
+ rownames=FALSE, header=TRUE, na="", sheet="Sheet1", stringsAsFactors=TRUE)
```

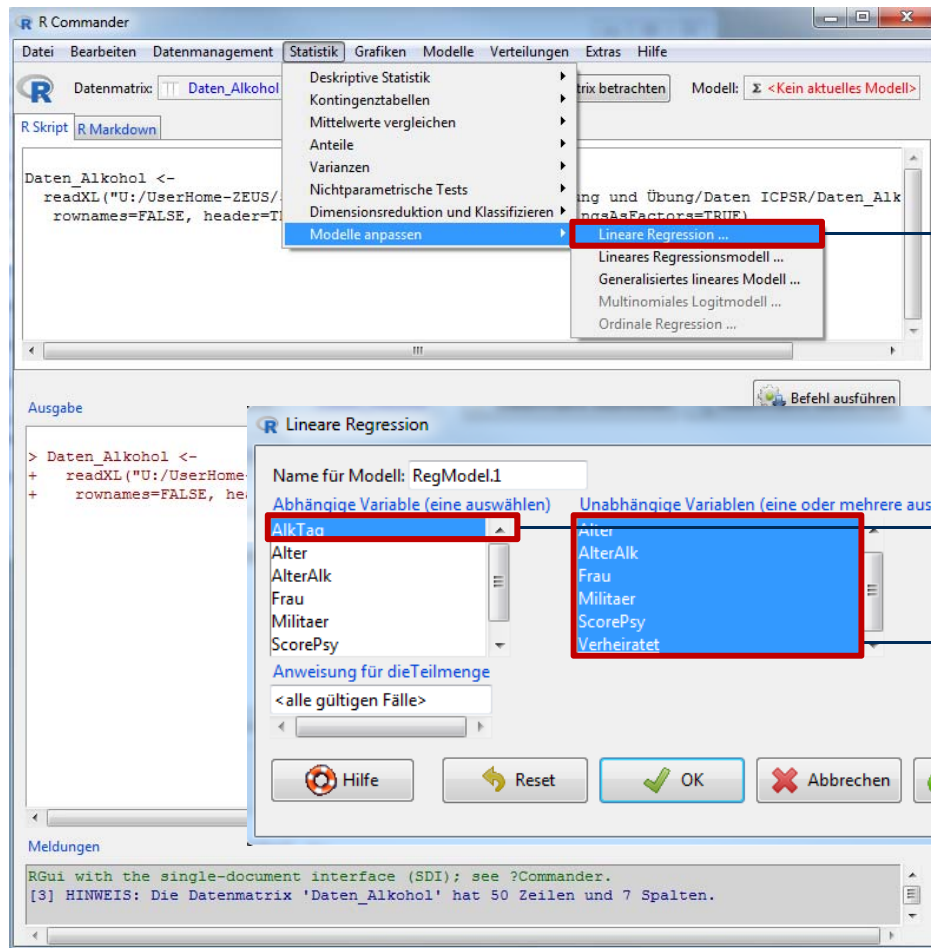
Meldungen

```
RGui with the single-document interface (SDI); see ?Commander.  
[3] HINWEIS: Die Datenmatrix 'Daten_Alkohol' hat 50 Zeilen und 7 Spalten.
```

	AlkTag	AlterAlk	Alter	ScorePsy	Militaer	Verheiratet	Frau
1	1	18	21	24	0	0	1
2	1	13	22	12	0	1	1
3	1	19	23	8	0	0	0
4	1	20	31	9	1	1	1
5	1	15	31	7	0	1	1
6	1	18	31	8	0	1	0
7	1	16	44	18	0	0	1
8	1	16	45	3	0	1	1
9	1	18	49	10	0	1	0
10	1	18	49	13	0	1	1
11	2	19	22	18	0	0	1
12	2	13	22	14	0	0	0
13	2	14	24	24	0	0	1
14	2	18	25	17	0	0	0
15	2	18	26	21	1	1	0
16	2	16	27	4	0	0	1
17	2	16	28	2	0	1	1
18	2	16	29	8	0	0	0
19	2	16	31	11	1	0	1
20	2	15	31	5	0	0	1
21	2	12	33	3	0	1	1
22	2	17	34	11	0	1	1
23	2	19	42	5	0	0	1
24	2	14	43	3	0	1	1
25	2	16	47	9	0	1	1
26	2	14	48	5	0	1	1
27	3	17	23	17	0	0	1
28	3	15	29	17	0	0	0
29	3	12	31	13	0	1	1
30	3	13	32	8	1	1	0
31	3	16	33	15	0	1	1
32	3	12	34	12	0	0	1
33	3	13	35	11	0	0	1
34	3	14	43	7	0	1	1
35	3	16	46	3	0	1	1
36	3	13	47	7	0	0	1
37	3	12	48	4	0	1	1
38	3	18	49	19	1	1	0
39	4	12	22	15	0	0	1
40	4	15	46	1	0	0	0
41	5	16	19	21	0	0	0
42	5	17	22	10	0	0	0
43	5	13	30	3	0	1	0
44	5	12	33	24	0	1	1
45	5	12	40	8	0	0	0
46	5	10	47	10	0	0	1

R Commander (Rcmdr)

Methode der kleinsten Quadrate (lineare Regression)



1: Statistik > Modelle anpassen
> Lineare Regression ...

2: Abhängige Variable auswählen

3: Unabhängige Variablen auswählen

R Commander (Rcmdr)

Schätzung von b_0 und b_j

Resultate R Commander

«Code» der linearen Regression

Resultate der linearen Regression

Regressionskoeffizienten

The screenshot shows the R Commander window with the following content:

R Skript:

```
Daten_Alkohol <-  
  readXL("U:/UserHome-ZEUS/S Lehre/MSQ WAQM/2018 FS/Vorlesung/Daten ICPSR/Daten_Alkohol.xlsx",  
    rownames=FALSE, header=TRUE, na="", sheet="Sheet1", stringsAsFactors=TRUE)  
RegModel.1 <- lm(AlkTag~Alter+AlterAlk+Frau+Militaer+ScorePsy+Verheiratet,  
  data=Daten_Alkohol)  
summary(RegModel.1)  
RegModel.2 <- lm(AlkTag~Alter+AlterAlk+Frau+Militaer+ScorePsy+Verheiratet, data=Daten_Alkohol)  
summary(RegModel.2)
```

Ausgabe:

```
> RegModel.2 <- lm(AlkTag~Alter+AlterAlk+Frau+Militaer+ScorePsy+Verheiratet, data=Daten_Alkohol)  
> summary(RegModel.2)
```

Call:
lm(formula = AlkTag ~ Alter + AlterAlk + Frau + Militaer + ScorePsy +
 Verheiratet, data = Daten_Alkohol)

Residuals:

	Min	1Q	Median	3Q	Max
	-2.6803	-0.8298	-0.1515	0.8783	2.7092

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.05733	1.44688	6.260	0.00000153 ***
Alter	0.02262	0.02181	1.037	0.3056
AlterAlk	-0.39814	0.07614	-5.229	0.00004763 ***
Frau	-1.27863	0.38861	-3.290	0.0020 **
Militaer	0.02620	0.64510	0.041	0.9678
ScorePsy	0.02152	0.03061	0.703	0.4859
Verheiratet	-0.82591	0.40098	-2.060	0.0455 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.242 on 43 degrees of freedom
Multiple R-squared: 0.526, Adjusted R-squared: 0.4599
F-statistic: 7.954 on 6 and 43 DF, p-value: 0.00008459

Meldungen

```
RGui with the single-document interface (SDI); see ?Commander.  
[3] HINWEIS: Die Datenmatrix 'Daten_Alkohol' hat 50 Zeilen und 7 Spalten.
```

Quelle: Eigene Darstellung

rigu@zhaw.ch

Angewandte quantitative Methoden

R Commander (Rcmdr)

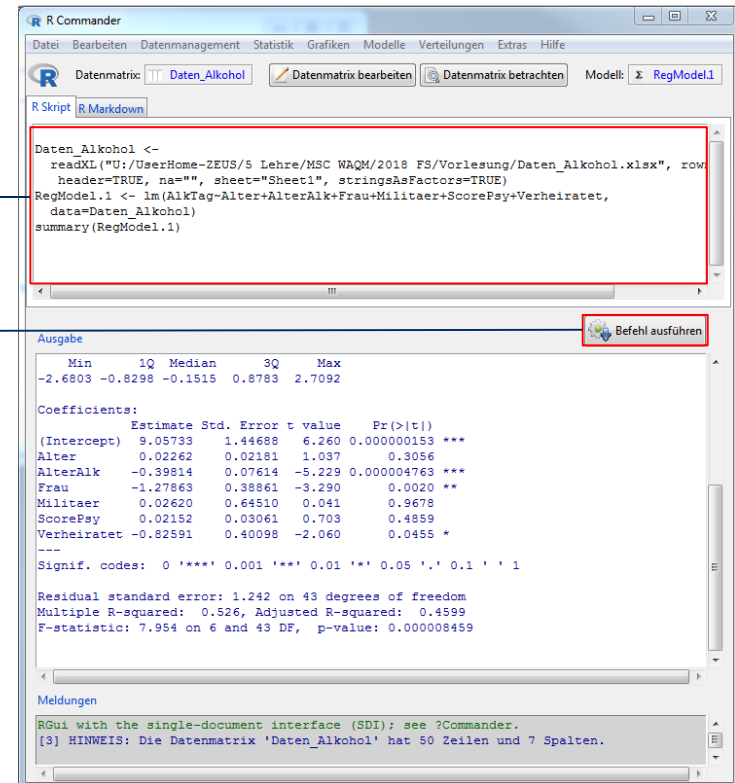
Standardisierte Regressionskoeffizienten

Vorgehen R Commander

- Leider können standardisierte Koeffizienten **nicht per Mausklick** generiert werden
- Berechnung durch **Befehlszeilen im Fenster «R Skript»**:

«R Skript» Fenster für Befehlseingaben

Befehl ausführen (oder «Ctrl+r»)



The screenshot shows the R Commander window with the 'R Skript' tab selected. The script contains the following R code:

```
Daten_Alkohol <-  
readXL("U:/UserHome-ZEUS/5 Lehre/MSQ WAQM/2018 FS/Vorlesung/Daten_Alkohol.xlsx", row  
header=TRUE, na="", sheet="Sheet1", stringsAsFactors=TRUE)  
RegModel.1 <- lm(AlkTag~Alter+AlterAlk+Frau+Militaer+ScorePsy+Verheiratet,  
data=Daten_Alkohol)  
summary(RegModel.1)
```

The output window shows the summary of the linear model:

```
Min      1Q  Median      3Q      Max  
-2.6803 -0.8298 -0.1515  0.8783  2.7092  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  9.05733    1.44688   6.260 0.00000153 ***  
Alter         0.02262    0.02181   1.037  0.3056  
AlterAlk     -0.39814    0.07614  -5.229 0.000004763 ***  
Frau        -1.27863    0.38861  -3.290  0.0020 **  
Militaer      0.02620    0.64510   0.041  0.9678  
ScorePsy      0.02152    0.03061   0.703  0.4859  
Verheiratet  -0.82591    0.40098  -2.060  0.0455 *  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 1.242 on 43 degrees of freedom  
Multiple R-squared:  0.526, Adjusted R-squared:  0.4599  
F-statistic: 7.954 on 6 and 43 DF,  p-value: 0.00008459
```

The 'Befehl ausführen' button is highlighted in the top right corner of the output window.

Vorteile R Commander

- Gratis, Open Source
- Zugänglicher Einstieg in R, insbesondere für Personen ohne Programmier-Kenntnisse
- R Code wird ausgegeben und kann bearbeitet werden – Übertritt zum Code-Schreiben ist einfach, sofern gewünscht
- Schlankes Interface mit (nur) den basalsten statistischen Analyseverfahren (im Kontrast zu SPSS)
- Dennoch eine gute Auswahl von grundlegenden Verfahren
 - deskriptive Analysen
 - diverse Grafiken
 - diagnostische Analyseverfahren zum Überprüfen von Annahmen
 - nicht-parametrische Tests
 - GLM

Nachteile R Commander

- Installation: zuerst muss R installiert werden, dann die R Commander Library → für einige Studierende schwierig
- Teilweise aufgetreten: gewisse Inkompatibilitäten zu R-Studio (?!)
- Datenmanagement ungewohnt für Excel-Sozialisierte (aber: im Prinzip kann das ganze Datenmanagement auch in Excel gemacht werden)
- Weniger bekannt und verbreitet als z.B. SPSS

Evaluation durch die Studierenden

R Commander wird uns sicherlich bei der Masterarbeit weiterhelfen.

Herr H. war sympathisch, dass er auch immer wieder am R-Commander gescheitert ist.

R-Command ist als gegeben vorausgesetzt worden so dass man anfangs vor allem an technischen und nicht inhaltlichen Dingen gescheitert ist. Später aber mit "Step by Step" Anweisungen verbessert worden.

R Commander ist ja an sich eine tolle Sache. Es wäre aber sinnvoller gewesen am Anfang eine längere "Schulung" durchzuführen und dann erst jeweils Aufgaben abzugeben.

R-Commander: Besser wäre gewesen, eine kurze Einführung (1-2h) in das Programm zu geben. Dann wären viele Fragen schneller geklärt gewesen und die Aufgaben auf nächste Woche wären ev. besser gelöst worden.

Fazit

- **BWL-Studierende an der ZHAW analysieren Daten mit R, es klappt!**
- Das Glas ist halb voll. Der Kurs lief OK, einige Anfangsschwierigkeiten, aber generell nicht schlimmer als mit einem anderen Programm wie z.B. SPSS
- Wir optimieren den Einstieg in den R Commander, leiten die Studierenden in Zukunft stärker an bei der Arbeit mit dem R Commander
- Wir verweisen intensiver auf R als nachhaltige, etablierte Statistik-Umgebung
- Zukunft: stärkerer Einsatz von R Commander/R an der ZHAW auch in anderen Studiengängen?

Literatur

- Cumming, Geoff and Robert Calin-Jageman. 2017. *Introduction to the New Statistics. Estimation, Open Science, and Beyond*. New York: Routledge.
- Field, Andy, Zoë Field and Jeremy Miles. 2012. *Discovering Statistics Using R*. Los Angeles: Sage.
- Fox, John. 2016. *Applied Regression Analysis and Generalized Linear Models*. Los Angeles: Sage Publications.
- Fox, John. 2017. *Using the R Commander. A Point-and-Click Interface for R*. Boca Raton: CRC Press, an Taylor & Francis Group.
- Kronthaler, Franz. 2016. *Statistik Angewandt. Datenanalyse Ist (K)Eine Kunst Mit Dem R Commander*. Berlin: Springer Spektrum.