

Statistische Software (R)

Paul Fink, M.Sc.

Institut für Statistik
Ludwig-Maximilians-Universität München

Einführung



Formeln und Code

Formeln sind der bei weitem effizienteste und eleganteste Weg, um mathematische Zusammenhänge auszudrücken:

$$1 + e^{i\pi} = 0$$

Ähnliches gilt für Programmiersprachen und Computer:

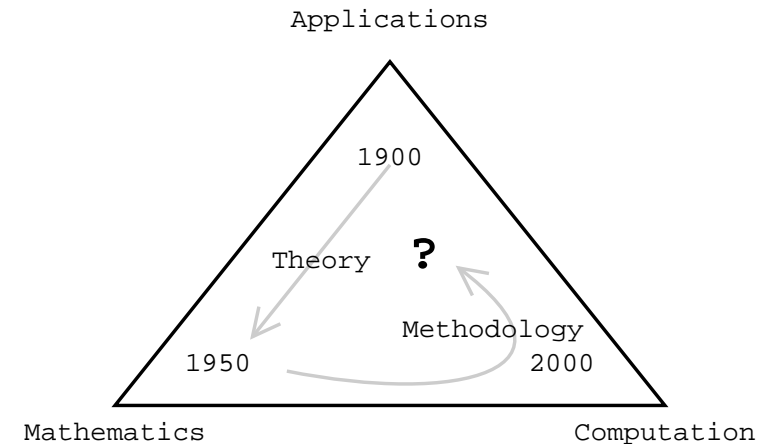
```
R> all.equal(exp(1i*pi)+1, 0+0i)
[1] TRUE
```

Das Schöne an Statistik ist,

- daß wir mit Mathematik unsere Theorien beschreiben,
- und Computer verwenden, um Daten zu analysieren.

Statistik und Computer

Efron (2001): Statistik im 20. Jahrhundert



Fink: Statistische Software (R) SoSe 2014

1

Was ist Statistiksoftware?

- Eine Programm, das Zahlen als Input nimmt, und daraus Tabellen (und Grafiken) erzeugt?
- Ein(e) (Sammlung von) Programm(en) für Exploration, Inferenz und Modellierung?
- Ein Werkzeug zur Verwaltung, Manipulation und Analyse von Daten?
- Ein Kommunikationsmedium mit CPU (Grafikkarte, Drucker, ...) des Rechners?

Kommunikationsformen

Zeichensprachen: Leicht zu lernen, oft universell verständlich, in Komplexität und Ausdrucksform limitiert. Bsp: Verkehrszeichen, GUIs, etc.

Schriftsprachen: Schwer zu lernen, Festlegung auf Sprache, in Komplexität und Ausdrucksform nahezu unlimitiert. Bsp: Deutsch, UNIX-Shell, etc.

Relevanz für Datenanalysen

- Sehen Sie Statistiksoftware (auch) als Hilfsmittel, um dem Computer mitzuteilen, wie Sie Ihre Daten analysieren wollen.
- Für einfache Analysen reichen einfache Kommunikationsformen, aber selbst im komplexesten GUI haben Sie irgendwann das letzte Untermenü erreicht.
- In programmierbaren Umgebungen hängt die Decke des Möglichen bedeutend höher.

Relevanz für Datenanalysen

- R-Befehle sind Ihre Anweisungen an den Computer.
- Kommandos, die Sie direkt am Prompt tippen, ähneln verbalen Zurufen: schnell kommuniziert und verarbeitet, schnell vergessen.
- Dinge, die man sich länger merken will, schreibt man sich auf → dasselbe gilt für Datenanalysen.
- **Tipp 1:** Schreiben Sie Ihre Analysen in einem Editor, sparen Sie nicht mit Kommentaren.

Was ist R?

R ist eine kostenlose Software-Umgebung für statistische Datenanalyse und Graphiken. Es beruht auf einer Implementation der Sprache S. Anfänglich wurde R von Ross Ihaka und Robert Gentleman (Univ. Auckland) entwickelt und wird seit Mitte der 90er Jahre von einem Entwickler-Kollektiv (R-Core) betreut.

Informationen zu R:

<http://www.R-project.org>

Vorteile von R

- Open Source
(Alle Algorithmen und Funktionen im Quelltext nachvollziehbar)
- schnelle Entwicklung
- leichte Erweiterbarkeit mit Paketen
- auf (fast) jedem Betriebssystem / jeder Plattform lauffähig
- keine Lizenzgebühren

Nachteile von R

- keine vollständige grafische Benutzeroberfläche
- Viele Pakete, einige Funktionen gibt es mehrfach

Wie installiere ich R?

Fertige R Distributionen sind für Windows, MacOS X und viele Linux-Versionen auf CRAN (**C**omprehensive **R** **A**rchive **N**etwork) erhältlich:

- Die Website <http://www.R-project.org> aufrufen und auf „download R“ klicken
- Einen nahe gelegenen Server auswählen
- Dem Link mit dem benötigten Betriebssystem folgen
- Bei Windows: „base“ auswählen und R installieren

Texteditoren

- Damit Tipp 1 nicht mühsam wird, braucht man einen Editor, der R unterstützt (Syntax Markierung, Klammerpaare zählen, Einrückung, Verbindung zum R Prozess, ...).
- R ist kein Texteditor.
- Welcher Editor für wen optimal ist, hängt stark von persönlichen Präferenzen und vom Arbeitsstil ab.
- Empfehlung: RStudio

Texteditoren

Tinn-R : Als kostenloser Editor für Windows bietet er Syntax-Highlighting und eine Anbindung an R:

<http://sourceforge.net/projects/tinn-r/>

RStudio : Der Editor funktioniert unter Windows, Mac und Linux und bietet ebenso Syntax-Highlighting und eine Anbindung an R (empfehlenswert):

<http://www.rstudio.com/>

Plugins : R-Plugins für Notepad++, Emacs, Vim, etc..

Eine Übersicht über weitere Editoren findet man unter http://www.sciviews.org/_rgui/projects/Editors.html .

RStudio als Editor

Strg + Enter

führt einen Befehl (die aktuelle Befehlszeile) aus dem R-Skript in der Konsole aus. Es können auch ganze Abschnitte markiert werden und gemeinsam abgeschickt werden.

Strg + Tab

ermöglicht die Auto-Vervollständigung der R-Befehle, mit F1 gelangt man zu den zugehörigen Hilfeseiten.

RStudio

RStudio ist ein kostenloser Editor für R und noch viel mehr als das. RStudio kombiniert

- die R-Konsole
- ein Editorfenster für R-Skripts
- Plots und Hilfeseiten
- Übersicht der getätigten Befehle (History), Workspace-Übersicht
- und vieles mehr!

in einem Fenster.

⇒ IDE: Integrated **D**evelopment **E**nvironment

RStudio kann von <http://rstudio.org/> heruntergeladen werden, dort finden sich auch viele nützliche Infos zum Umgang mit RStudio.

HILFE!!!

R kennt viele Befehle aber wie ging das noch mal?

Hilfe zur Selbsthilfe! Auf den Hilfeseiten ist alles dokumentiert!

Wie kommt man dahin?

Das magische Zeichen ist "?", stellt man es vor einen Befehl gelangt man zu dessen Hilfeseite

`R> ?median`

HILFE gefunden !!!

Häufige Anwendungsfälle:

- Ich soll den Median ausrechnen, kenne aber den Befehl nicht.
⇒ Nachschlagen in Suchmaschinen/Büchern und ausprobieren

- Ich denke, dass der Befehl "med" enthält.
⇒ `apropos()`.

```
R> apropos("med")  
[1] ".__C__namedList" "elNamed"      "elNamed<-"    "median"  
[5] "median.default"  "medpolish"    "runmed"
```

- Ich weiß nicht mehr welche Argumente der Befehl `median` annimmt oder wie sie heißen.
⇒ `?median`