

**st**  
scientific tools

Reinhold Hatzinger  
Kurt Hornik  
Herbert Nagel

**R**

Einführung durch  
angewandte Statistik

**Reinhold Hatzinger  
Kurt Hornik  
Herbert Nagel**

**R**

**Einführung durch  
angewandte Statistik**

**PEARSON**  
**Studium**

---

ein Imprint von Pearson Education  
München • Boston • San Francisco • Harlow, England  
Don Mills, Ontario • Sydney • Mexico City  
Madrid • Amsterdam

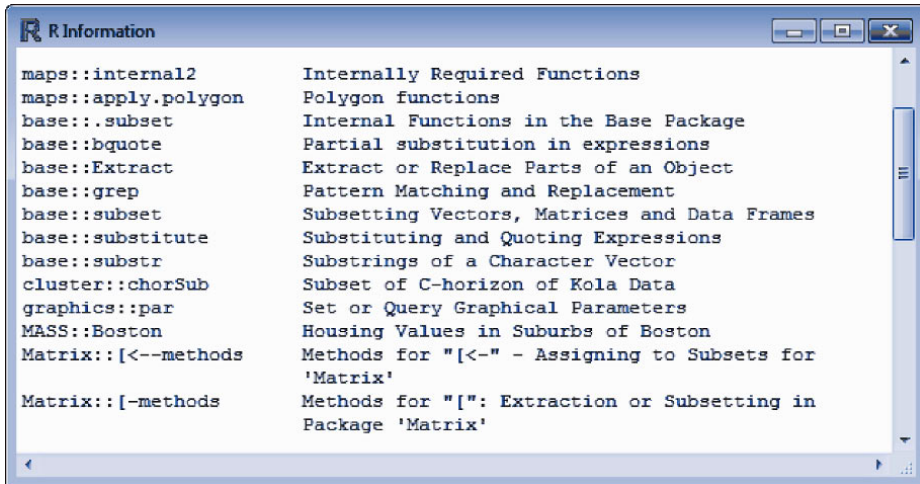


Abbildung 5.15: Ausschnitt eines Resultats von `help.search("sub")`

Durchsucht werden können dort unter anderem alle auf CRAN verfügbaren Packages und dazugehörige Dokumentationen (Functions und Vignettes), Methodenübersichten (Task Views) sowie diverse Hilfe-Foren (R-help).

Nachdem Sie die zu suchenden Begriffe eingegeben haben, öffnet sich die „R Site Search“-Webseite mit ersten Ergebnissen. Durch Anklicken diverser Auswahlpunkte bzw. Auswahlen können Sie das Suchverhalten nach Ihren Vorstellungen gestalten.

Besonders die Archive der R-help-Foren mit Hunderttausenden Beiträgen erweisen sich als Informationsquelle, wo man (nahezu) jede nur vorstellbare Information zu R findet. Oft gibt es Lösungen für verzwickte Probleme oder Hinweise, wie man solche Lösungen finden kann. Auf der Webseite <https://stat.ethz.ch/mailman/listinfo/r-help> kann man diesem Forum beitreten.

### Menüpunkt: HTML Hilfe

Das Anklicken dieses Menüpunkts öffnet eine (lokale) Informationsseite (► Abbildung 5.16).

Im Wesentlichen ist hier alles, was Sie sonst auch über das Hilfe-Menü finden, in durchblätterbarem HTML-Format verfügbar, wie z. B. die Manuals, die über das Menü Hilfe ▷ Handbücher (PDF) als PDF-Dokumente gelesen werden können, oder Search (entspricht dem Menüpunkt Durchsuche Hilfe bzw. dem Befehl `help.search()`) gibt aber dort zusätzlich eine Liste von Schlüsselwörtern. Zu den interessanten Links gehört auch Packages. Dort kann man detailliertere Information zu den lokal installierten Packages erhalten.

### Menüpunkt: Apropos ...

Hier wird eine Suche angeboten, in der man wie beim Menüpunkt Durchsuche Hilfe bzw. bei Verwendung der Funktion `help.search()` auch nach Namensteilen suchen kann. Die Suche bezieht sich dabei auf alles, was momentan im Workspace definiert ist. Das sind neben Funktionen und Daten geladener Packages (dazu gehören

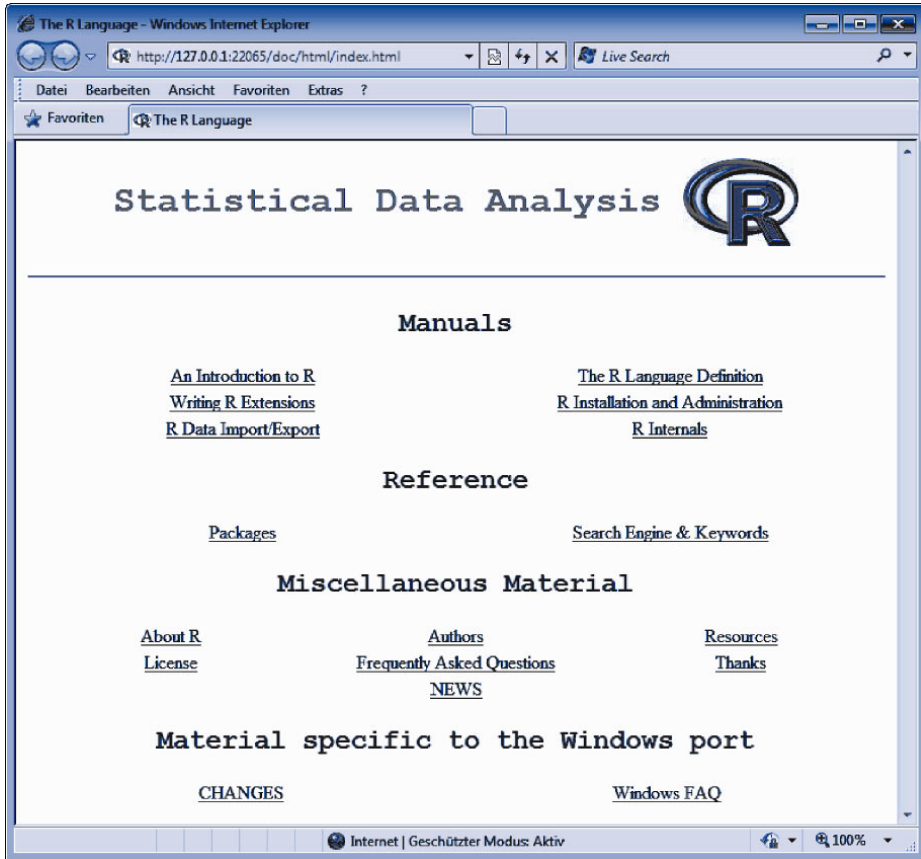


Abbildung 5.16: Die Webseite der R-HTML-Hilfe

auch die Funktionen und Datensätze aus den Standard-Packages, die automatisch geladen werden) auch definierte Variablen, Data Frames etc. Sie können eine solche Suche auch über die R-Konsole mit dem Befehl `apropos()`, also z. B. `apropos("sub")`, starten.

## 5.6.2 Dokumente, Webseiten und weiterführende Information

### Frequently Asked Questions – FAQs

Im Hilfe-Menü gibt es zwei Links zu Frequently Asked Questions-Listen. Eine ist eher allgemein gehalten und behandelt Fragen zu verschiedenen Aspekten von R. Die zweite fokussiert eher die Windowsimplementation von R. Obwohl manche Themen sehr spezifische Aspekte behandeln, kann ein Durchsehen auch für Anfänger durchaus hilfreich sein.

## Die R Manuals

Über Hilfe ▷ Handbücher (PDF) können Sie die R Manuals bekommen. Die wichtigsten sind *An Introduction to R* sowie *R Data Import ▷ Export*. Diese beschreiben die Funktionsweise von R tiefergehend als dieses Buch und eignen sich nur bedingt für Anfänger. Dennoch lohnt es sich, hin und wieder einen Blick hineinzuworfen. Die anderen Manuals behandeln eher technische Details von R, z. B. wie man Packages programmiert und dokumentiert.

## Online-Dokumente auf CRAN

Auf der CRAN-Webseite finden Sie in der Navigation links unten unter Documents den Link [Contributed](#). Auf dieser Webseite finden Sie frei verfügbare Dokumente und Bücher (meist in PDF-Form) zu den unterschiedlichsten Themen, manche auch auf Deutsch. Obwohl sich auch da die meisten Texte an eher Fortgeschrittene richten, gibt es auch einiges für Anfänger.

## Suche im WWW

Sucht man R-Material im Web über Standard-Suchmaschinen (wie Google, Yahoo oder Bing), dann bekommt man natürlich viele irrelevante Treffer, weil der Buchstabe R eine Vielzahl von Interpretationen zulässt. Besser ist es, spezialisierte Suchmaschinen zu verwenden. Gute Startpunkte sind

- <http://search.r-project.org> (auch über das R Hilfe-Menü erreichbar)
- <http://www.rseek.org>
- [http://www.dangoldstein.com/search\\_r.html](http://www.dangoldstein.com/search_r.html)

## Bücher

Wir wollen hier drei Bücher empfehlen, die als Ausgangspunkt für vertieftes Lernen bzw. als Nachschlagewerke dienen können. Eine sehr ausführliche Darstellung von R und der Verwendung auch fortgeschrittener statistischer Methoden gibt *The R Book* von Michael J. Crawley (Crawley, 2007). Schon erwähnt haben wir *das* Referenzbuch zu Grafiken von Paul Murrell, *R Graphics* (Murrell, 2005). Das (deutschsprachige) Buch *Programmieren mit R* von Uwe Ligges (Ligges, 2008) legen wir all jenen nahe, die mit R programmieren und/oder die wesentlichen Strukturen von R grundlegender verstehen möchten.

## 5.7 R-Befehle im Überblick

`?funktionsname` ruft eine Hilfeseite auf. Beispiel: `?subset`

`apropos(what)` sucht nach Objekten im Workspace. `what` kann Teil des Namens des gesuchten Objekts sein. Beispiel: `apropos("sub")`

`axis()` fügt einem Plot eine Achse hinzu.

`barplot()` erstellt ein Balkendiagramm mit horizontalen oder vertikalen Balken.

`colours()` erzeugt einen Vektor mit allen Farbnamen, die R standardmäßig kennt.

`dev.new()` erzeugt ein neues Grafikfenster.

`dir(path=".")` erzeugt einen Vektor mit allen Dateien, die im Verzeichnispfad `path` gefunden werden. `path = "."` definiert das aktuelle Working Directory.

`dotchart(x)` erzeugt einen Cleveland Dot Plot der Werte in `x`.

`example(topic)` führt den gesamten R-Code aus, der im Examples-Abschnitt auf einer Hilfeseite zu finden ist, außer der Code ist mit `dontrun` speziell „auskommentiert“.

`file.choose()` öffnet einen File-Dialog, um eine Datei auswählen zu können.

`getwd()` gibt den gesamten Pfad des aktuellen Working Directory an.

`help(topic)` öffnet eine Hilfeseite zu `topic`.

`help.search(pattern)` durchsucht die Dokumentation (Hilfeseiten) nach `pattern` (einer Zeichenkette), wobei ungefähre Einträge gesucht werden.

`hist(x)` erzeugt ein Histogramm der Werte in `x`.

`library(package)` lädt ein Package.

`lines(x, y = NULL, ...)` fügt einem Plot Linien hinzu, `x` und `y` sind Vektoren, ... sind weitere Parameter. Wird nur `x` spezifiziert, dann werden die Werte von `x` auf der `y`-Achse aufgetragen, auf der `x`-Achse die Werte 1 bis Anzahl der Elemente in `x`. Sonst werden Punkte mit den Koordinaten in `x` und `y` der Reihe nach verbunden. Unter ... kann man Grafikparameter spezifizieren.

`ls()` ergibt einen Vektor, in dem alle während einer R-Sitzung erzeugten Daten und Funktionen aufgelistet sind.

`mtext(text, side = 3)` fügt Text auf einer Seite eines Plots hinzu. Die Seiten (`side`) sind: unterhalb (1), links (2), oberhalb (3) oder rechts (4).

`objects()` wie `ls()`

`par(...)` zum Setzen und Abfragen von Grafikparametern. Will man einen Grafikparameter abfragen, dann gibt man ihn an der Stelle ... unter Hochkommas an, z. B. `par("lty")`, um den aktuellen Linientyp abzufragen. Will man einen Grafikparameter setzen, dann ist ... von der Form `name = wert`. Wenn man z. B. den Linientyp in punktiert ändern will, schreibt man `par("lty" = "dotted")`. Eine Liste aller Grafikparameter und ihrer aktuellen Werte erhält man mit `par()` (ohne Argument).

`plot(x, y, ...)` ist die Standardfunktion, um R-Objekte zu plotten. Im einfachsten Fall werden Punkte an den Koordinaten `x` und `y` gezeichnet, weitere Parameter kann man unter `...` spezifizieren. Wird nur `x` spezifiziert, dann werden die Werte von `x` auf der `y`-Achse aufgetragen, auf der `x`-Achse die Werte 1 bis Anzahl der Elemente in `x`.

`points(x, y = NULL, ...)` Fügt einem Plot Punkte hinzu. `x` und `y` sind Vektoren und spezifizieren die Koordinaten, `...` sind weitere Parameter. Wird nur `x` spezifiziert, dann werden die Werte von `x` auf der `y`-Achse aufgetragen, auf der `x`-Achse die Werte 1 bis Anzahl der Elemente in `x`. Sonst werden Punkte mit den Koordinaten in `x` und `y` der Reihe nach gezeichnet. Unter `...` kann man Grafikparameter spezifizieren. Siehe auch `lines()`.

`read.csv(file)` liest eine comma-separated Datei ein, die aus einer englischen Version von Excel oder OpenOffice.org Calc stammt, und erzeugt dabei einen Data Frame.

`read.csv2(file)` liest eine comma-separated Datei ein, die aus einer deutschen Version von Excel oder OpenOffice.org Calc stammt und erzeugt dabei einen Data Frame.

`read.spss()` liest eine SPSS-Datei ein und erzeugt dabei einen Data Frame (Package `foreign`).

`read.table(file)` liest eine Datei im Tabellenformat ein und erzeugt dabei einen Data Frame.

`rm(...)` löscht Objekte, die in einer R-Sitzung definiert wurden.

`seq(from, to, by)` erzeugt eine Sequenz der Zahlen `from` bis `to` in Schritten von `by`. Statt `by=` kann man auch `length.out=` spezifizieren. Dann hat der resultierende Vektor die unter `length.out` angegebene Länge, wobei die Schrittweite in diesem Fall automatisch bestimmt wird.

`scan()` liest Daten in einen Vektor aus der Zwischenablage oder von der Tastatur.

`setwd(dir)` setzt das Working Directory auf `dir`. Als Trennzeichen für Laufwerk-, Verzeichnis- und Dateibezeichnungen benutzt man `/` (oder in Windows alternativ einen doppelten Backslash). Beispiel: `setwd("D:/Data")`

`source(file)` liest R-Code aus einer Datei und führt ihn aus.

`text(x, y = NULL, labels, ...)` fügt Zeichenketten, die unter `labels=` spezifiziert werden, einem Plot hinzu. Siehe auch `points()`. `lines()`.

`title(main, sub, ...)` fügt einem Plot Titel und Untertitel hinzu. Unter `...` kann man Grafikparameter spezifizieren.

`write.table(x, file)` dient zur Erzeugung von Dateien im Textformat, `x` ist hierbei (vorzugsweise) eine Matrix oder ein Data Frame, kann aber auch ein Vektor sein.

`write.csv(x, file)` schreibt eine comma-separated Datei zum Einlesen in einer englischen Version von Excel.

`write.csv2(x, file)` schreibt eine comma-separated Datei zum Einlesen in einer deutschen Version von Excel.

## 5.8 Übungen

1. Besuchen Sie die Companion Webseite zu diesem Buch und laden Sie die Dateien `bspdat.csv` und `bspcode.R` herunter.
2. Setzen Sie Ihr Working Directory auf das Verzeichnis, in dem Sie die Dateien gespeichert haben.
3. Lesen Sie den Datensatz aus der Datei `bspdat.csv` ein und speichern Sie ihn in einem Objekt namens `bspdat`.
4. Untersuchen Sie die Struktur des Datensatzes mit `head()` und `str()`.
5. Der Data Frame `bspdat` enthält eine Variable `female` mit den Werten 1 für „weiblich“ und 0 für „männlich“. Fügen Sie dem Datensatz eine neue Variable namens `Geschlecht` hinzu, wozu Sie `female` verwenden. `Geschlecht` soll ein Faktor sein mit den Werten "männlich" bzw. "weiblich".
6. Stellen Sie mit `par()` ein, dass zwei Plots in einer Zeile erzeugt werden, und erstellen Sie links ein Streudiagramm mittels `plot()` für „Größe“ und „Gewicht“, wobei Sie mit `subset()` nur Frauen auswählen. Rechts machen Sie das Gleiche, wobei Sie hier nur Männer plotten.
7. Verwenden Sie `source()`, um die Datei `bspcode.R` auszuführen.
8. Woran liegt es, wenn der erzeugte Plot bei Ihnen nur die Hälfte des Fensters ausfüllt? Ist dies bei Ihnen der Fall, so schließen Sie das Grafikfenster und führen Sie `bspcode.R` nochmals aus.
9. Fügen Sie an den Koordinaten (190, 60) einen beliebigen Text ein.
10. Speichern Sie die Grafik als Metafile mit der Endung `.emf` ab.

Datenfiles sowie Lösungen finden Sie auf der Webseite des Verlags.





# TEIL II

## Kategoriale Daten

<b>6</b>	<b>Eine kategoriale Variable .....</b>	<b>147</b>
<b>7</b>	<b>Mehrere kategoriale Variablen .....</b>	<b>193</b>



# Eine kategoriale Variable

6

<b>6.1</b>	<b>Einleitung</b> .....	148
<b>6.2</b>	<b>Kommen alle Kategorien gleich häufig vor?</b> ..	152
6.2.1	Numerische Beschreibung .....	154
6.2.2	Grafische Beschreibung .....	156
6.2.3	Statistische Analyse der Problemstellung .....	159
<b>6.3</b>	<b>Entsprechen Häufigkeiten bestimmten Vorgaben?</b> .....	166
6.3.1	Numerische und grafische Beschreibung .....	167
6.3.2	Statistische Analyse der Problemstellung .....	170
<b>6.4</b>	<b>Hat ein Prozentsatz (Anteil) einen bestimmten Wert?</b> .....	172
6.4.1	Statistische Analyse der Problemstellung .....	175
<b>6.5</b>	<b>In welchem Bereich kann man einen Prozentsatz (Anteil) erwarten?</b> .....	180
<b>6.6</b>	<b>R-Befehle im Überblick</b> .....	186
<b>6.7</b>	<b>Zusammenfassung der Konzepte</b> .....	188
<b>6.8</b>	<b>Übungen</b> .....	188
<b>6.9</b>	<b>Vertiefung: Die Chi-Quadrat-Verteilung oder wie entsteht ein p-Wert?</b> .....	190

ÜBERBLICK

*Kategoriale Daten entstehen durch die Klassifikation von Beobachtungen in Kategorien. Meistens wird die interessierende Information dadurch gewonnen, dass ausgezählt wird, wie oft bestimmte Kategorien vorkommen. Dadurch können Häufigkeitsverteilungen und Prozentsätze berechnet werden. Dieses Kapitel beschäftigt sich damit, wie man prüfen kann, ob beobachtete Häufigkeiten mit bestimmten Annahmen übereinstimmen. Es werden die Grundlagen statistischen Testens besprochen. Außerdem wird behandelt, wie man für einen Prozentsatz, den man aus einer Stichprobe gewonnen hat, Bereiche ermittelt, von denen man annehmen kann, dass sie mit einer bestimmten Sicherheit den wirklichen Prozentsatz, wie er in der Population vorkommt, enthalten. Dieses Kapitel legt die Basis zum Verständnis der Inferenz- oder schließenden Statistik.*

## LERNZIELE

Nach Durcharbeiten dieses Kapitels haben Sie Folgendes erreicht:

- Sie wissen, was absolute und relative sowie beobachtete und erwartete Häufigkeiten sind, und können diese in R berechnen, tabellieren und in Form von Balken- und Kreisdiagrammen grafisch darstellen.
- Sie wissen, was eine Null- und Alternativhypothese ist und was Testen von Hypothesen bedeutet. Sie können ein- und zweiseitige Hypothesen unterscheiden und formulieren.
- Sie sind in der Lage, ein Signifikanzniveau festzulegen und einen p-Wert beim Testen einer Hypothese zu beurteilen.
- Sie können in R einen Chi-Quadrat-Test sowie einen Binomial-Test bei verschiedenen Problemstellungen berechnen. Dabei sind Sie in der Lage, das Ergebnis technisch und inhaltlich zu interpretieren.
- Sie kennen die Bedeutung der Begriffe Schwankungsbreite und Konfidenzintervall und können solche in R erstellen und grafisch darstellen.

## 6.1 Einleitung

Kategoriale Information erhält man, wenn etwas (ein Merkmal oder Charakteristikum), das man an verschiedenen Personen, Unternehmen, Pflanzen etc. (also Beobachtungseinheiten) registriert, in eine von mehreren Kategorien fällt. Dieser Prozess kann unterschiedlich verlaufen. Die Kategorien können schon von vorneherein feststehen und man braucht seine Beobachtungen nur mehr entsprechend zuzuordnen. Oder aber man sammelt die Information zunächst in freier Form (z. B. offene Fragen in einem Fragebogen), um anschließend nach einem Kategorisierungsschema (das man eventuell erst entwickeln muss) die Beobachtungen zuzuteilen. Diese beiden Vorgänge könnte man auch als **KLASSIFIKATION** bezeichnen.

Etwas anders verläuft der Prozess der **AGGREGATION**. Der Begriff kommt aus dem Lateinischen und bedeutet Anhäufung oder Vereinigung. Hier werden schon vorhandene Daten in einfachere, zusammenfassendere Strukturen transformiert. Zum Beispiel kann man die Körpergröße von Personen, die man in cm gemessen hat, zu drei Kategorien aggregieren, nämlich klein, mittel und groß. Oder, bei zugrunde liegender kategorialer Information, kann man z. B. die Berufe Tischler, Maurer etc. in



## Copyright

Daten, Texte, Design und Grafiken dieses eBooks, sowie die eventuell angebotenen eBook-Zusatzdaten sind urheberrechtlich geschützt. Dieses eBook stellen wir lediglich als persönliche Einzelplatz-Lizenz zur Verfügung!

Jede andere Verwendung dieses eBooks oder zugehöriger Materialien und Informationen, einschliesslich

- der Reproduktion,
- der Weitergabe,
- des Weitervertriebs,
- der Platzierung im Internet, in Intranets, in Extranets,
- der Veränderung,
- des Weiterverkaufs
- und der Veröffentlichung

bedarf der schriftlichen Genehmigung des Verlags.

Insbesondere ist die Entfernung oder Änderung des vom Verlag vergebenen Passwortschutzes ausdrücklich untersagt!

Bei Fragen zu diesem Thema wenden Sie sich bitte an: [info@pearson.de](mailto:info@pearson.de)

## Zusatzdaten

Möglicherweise liegt dem gedruckten Buch eine CD-ROM mit Zusatzdaten bei. Die Zurverfügungstellung dieser Daten auf unseren Websites ist eine freiwillige Leistung des Verlags. Der Rechtsweg ist ausgeschlossen.

## Hinweis

Dieses und viele weitere eBooks können Sie rund um die Uhr und legal auf unserer Website



herunterladen