



R-Kurs: Tag 2

Lukas Mödl, Matthias Becher, Erin Sprünken

Institut für Biometrie und Klinische Epidemiologie

Charité - Universitätsmedizin Berlin, Berlin

erin-dirk.spruenken@charite.de

January 17, 2022



- 1 Laden von Daten
- 2 Deskreptive Statistik
- 3 Datenaufbereitung, Kovertierung und Verwendung
- 4 Plots

Laden

- ▶ `load()`
- ▶ `read.table()`
- ▶ `read.csv()`

Optionen bei `read.csv()`

Wenn man CSVs in R lädt kann man verschiedene Parameter verwenden um der Funktion zu sagen wie die CSV formatiert ist. Die wichtigsten werden hier vorgestellt

- ▶ `header(TRUE/FALSE)` : Zeigt an ob in der CSV Spaltenname in der ersten Reihe stehen
- ▶ `sep` : Welches Zeichen wird verwendet um Spalten zu trennen. Default ist ",". Es werden aber auch häufig ";" oder "\t" verwendet
- ▶ `dec` : Welches Zeichen wird bei Dezimalzahlen verwendet "." oder ","
- ▶ Beispiel: `read.csv("data.csv", header=TRUE, sep=";", dec=",")`

Summary()

```
> summary(data)
```

Spalte1	Spalte2	Spalte3	Spalte4
Min. : 1.00	a:25	Length:100	Mode :logical
1st Qu.: 25.75	b:25	Class :character	FALSE:50
Median : 50.50	c:25	Mode :character	TRUE :50
Mean : 50.50	d:25		
3rd Qu.: 75.25			
Max. :100.00			

Funktionen für die Deskreptive Statistik

▷ Mean = `mean()`

▷ Median = `median()`

▷ Minimum = `min()`

▷ Maximum = `max()`

▷ Standard Deviation = `sd()`

▷ Variance = `var()`

▷ Quantile = `quantile()`

▷ Correlation = `cor()`

▷ Covariance = `cov()`

▷ Crosstable = `table()`

Konvertieren von Daten

- ▶ Numeric \Leftrightarrow `as.numeric()`
- ▶ Character \Leftrightarrow `as.character()`
- ▶ Factor \Leftrightarrow `as.factor()`
- ▶ Date \Leftrightarrow `as.Date()`
- ▶ Logical \Leftrightarrow `as.logical()`

Indizierung

Häufig möchte man nur bestimmte Elemente eines Vektors, einer Liste oder eines Data Frames auswählen. Um das zu tun gibt es mehrere Möglichkeiten. Die direkteste ist es, die Indizes zu verwenden. Angenommen wir haben den Vektor `x <- c(1, 2, 3, 4, 5)`

- ▶ Einen bestimmten Wert auswählen \Leftrightarrow `x[1]`
- ▶ Mehrere Werte auswählen \Leftrightarrow `x[c(1, 3, 5)]`
- ▶ Eine Reihe von Werten auswählen \Leftrightarrow `x[1:3]`
- ▶ Einen bestimmten Wert weglassen \Leftrightarrow `x[-1]`

Indizierung von Listen und Data Frames

Liste

- ▶ `x[1]`
- ▶ `x[[1]]`
- ▶ `x[[1]][1]`

Data Frame

- ▶ `x[1,]`
- ▶ `x[,1]`
- ▶ `x[, "Spalte1"]`
- ▶ `x$Spalte1`

Filtern

Häufig kommt es vor, dass wir unsere Daten filtern möchten um beispielsweise nur die Männer bzw. Frauen zu untersuchen oder nur Patienten ab einem bestimmten Alter zu betrachten. In R gibt es verschiedene Befehle mit denen man das erreichen kann.

- ▶ `which()`

- ▶ `data[which(data$Sex == "M"),]`

- ▶ `%in%`

- ▶ `data[which(data$Color %in% c("blue", "red")),]`

- ▶ `subset()`

- ▶ `subset(data, Age < 50,)`

Nach mehreren Sachen gleichzeitig Filtern

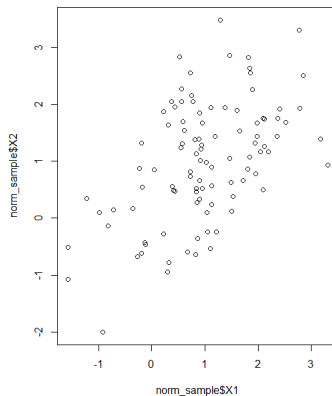
Manchmal möchte man nach mehreren Spalten gleichzeitig filtern. Anstatt das nacheinander zu tun, kann man auch mehrere Filter mit "&" verbinden.

Zum Beispiel:

```
▶ data[which(data$Sex == "W" & data$Age > 50),]
```

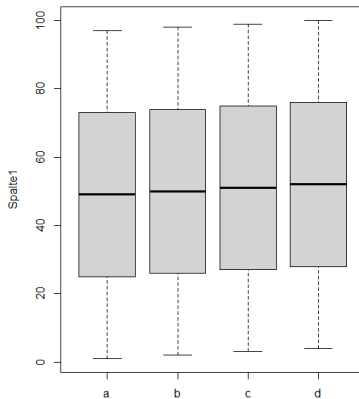
Scatterplot

▷ `plot(data$Spalte1, data$Spalte2)`



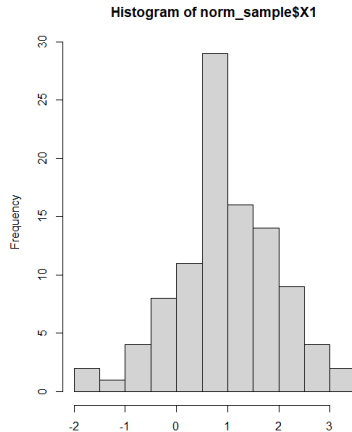
Boxplot

▷ `boxplot(Spalte1 ~ Spalte2, data)`



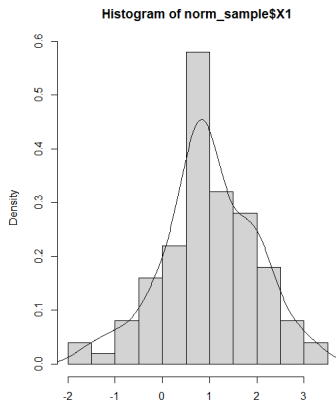
Histogram

▷ `hist(norm_sample$X1)`



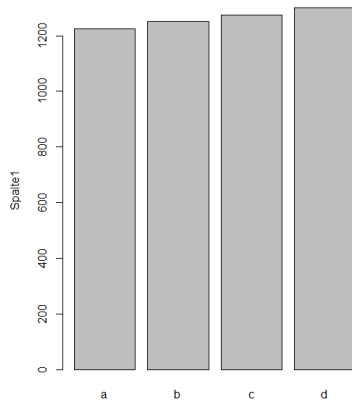
Histogram mit Density

```
▷ hist(norm_sample$X1, probability = T)  
  lines(density(norm_sample$X1))
```



Barplot

▷ `barplot(Spalte1 ~ Spalte2, data)`



Speichern von Plots

