

' 1. Variablen'

' 1.1 Numerische kontinuierliche Variablen'

entweder Integerwerte oder kontinuierliche Werte

```
zeit = read.table(file.path(pfadu, "zeit.txt"))  
head(zeit)  
class(zeit$Mo)
```

```
asp = read.table(file.path(pfadu, "asp.txt"))  
head(asp)  
class(asp$d)
```

'1.2 Faktoren'

Diese sind die Variablen-Namen im Data-Frame
names(asp)

```
class(asp$Vpn)
```

Ein Faktor hat eine oder mehrere Stufen: diese sind
die verschiedenen Kategorien aus denen, der Faktor besteht
levels(asp\$Kons)

Die Stufen sieht man auch wenn ein Faktor aufgerufen wird:
asp\$Kons[1:10]

'1.3 Funktionen für die Zusammenfassung numerischer Daten'

Mittelwert: mean()

```
mean(asp$d)
```

Median (oder 50% Quantil). Der mittlere Wert in der sortierten Reihenfolge
x = c(15, 19, -1, 10, 11, 18, 90000)
median(x)

das gleiche
quantile(x, .5)

Nach Sortierung sieht man, dass 15 die an der mittleren Stelle vorkommt.
sort(x)
-1 10 11 15 18 19 90000

Es gibt auch andere Bruchteile von Quantilen. Insbesondere
wird von dem interquartilen Bereich Gebrauch gemacht, um die Streuung einzuschätzen
IQR(x)

das gleich
quantile(x, .75) - quantile(x, .25)

Eine andere Messung der Streuungsgröße ist die Standardabweichung
sd(x)

```
y = c(15, 19, -1, 10, 11, 18, 20)
sd(y)
```

'1.4 Boxplot'

```
# Der Punkt ist der Median, der
# Quadrat ist der interquartile Bereich
bwplot(y)
```

```
bwplot(asp$d)
```

```
# Meistens möchte man sehen, ob eine numerische Variable
# von den Faktoren-Stufen beeinflusst wird. Hier ist die Syntax dafür:
# (entspricht: 'd gegeben Kons')
bwplot(d ~ Kons, data = asp)
```

```
# Um gleichzeitig zwei Faktoren abzubilden:
bwplot(d ~ Kons | Bet, data = asp)
```

'1.5 Barchart'

```
# Die Funktion barchart() in library(lattice) soll verwendet werden
# wenn man die Häufigkeiten von Faktoren abbilden will - also wenn
# die Ausgabe von table() abgebildet werden soll. Um zu barchart() zu gelangen,
# muss library(lattice) einmal im Session eingegeben werden - die
```

```
tab = table(asp$Kons)
tab
barchart(tab)
# oder
barchart(tab, horizontal=F)
```

```
# Gekreuzte Tabellen können auch mit barchart() abgebildet werden:
```

```
tab = table(asp$Kons, asp$Bet)
tab
```

```
#      be  un
# k  853  425
# t  450 1164
```

```
barchart(tab, horizontal=F, auto.key=T)
# meistens besser als Proportionen
p = prop.table(tab, 1)
barchart(p, horizontal=F, auto.key=T)
```

```
# oder in der anderen Reihenfolge
tab = table(asp$Bet, asp$Kons)
p = prop.table(tab, 1)
barchart(p, horizontal=F, auto.key=T)
```

```
# Für mehrere Faktoren, z.B.
# (Hier muss read.delim() statt read.table() verwendet werden,
# weil es Leerstellen in einigen Stufen vorkommen
```

```
kj = read.delim(file.path(pfadu, "kj.txt"))
names(kj)
tab = table(kj$str, kj$age, kj$gender)
# das gleiche
tab = with(kj, table(gender, age, str))
barchart(tab, horizontal=F, beside=T, auto.key=T)
p = prop.table(tab, 1:2)
barchart(p, horizontal=F, beside=T, auto.key=T)
```

```
temp = kj$age == "young"
kjyoung = kj[temp,]
tab = with(kjyoung, table(str, gender))
barchart(tab, auto.key=T, horizontal=F)
```

1.6 Fragen

```
# Für den Data-Frame dip
dip = read.table(file.path(pfadu, "dip.txt"))
```

```
# erstellen Sie einen Boxplot der Dauer
# für die Faktoren V (Vokal) und Vpn (Versuchsperson) zusammen
# Gibt es systematische Dauer-Unterschiede zwischen
# den Diphthongen und/oder zwischen den Versuchspersonen?
```

```
# Für den Data-Frame rating:
rating = read.table(file.path(pfadu, "rating.txt"))
# verwenden Sie boxplots um einzuschätzen, inwiefern Rating
# von der Grammatikalität (Gram) und Lang (E, S)
# beeinflusst werden.
```

```
# Für den Data-Frame sz
sz = read.table(file.path(pfadu, "sz.txt"))
```

```
# erstellen Sie einen Barchart, um die Häufigkeiten zu zeigen,
# mit denen /s/ oder /z/ von Sprechern aus Bayern und
# Schleswig-Holstein produziert wurden.
```

```
# Für den Data-Frame preasp
preasp = read.table(file.path(pfadu, "preasp.txt"))
```

```
# erstellen Sie einen Barchart, um einzuschätzen, inwiefern
# die Verteilung von  $\pm$ preasp (ob Pre-aspiration vorkam oder nicht)
# von dem davor kommenden Vokal (vtype) beeinflusst wurde.
```