

# Einführung in die Datenanalyse mit R - Faktoren

Jan-Philipp Kolb

4 Mai 2017

# Was sind Faktoren in R

- Faktoren können eine begrenzte Zahl von Ausprägungen annehmen
- Sie werden oft auch als kategoriale Variablen bezeichnet
- Sie sind vor allem bei der Modellierung wichtig
- Faktoren werden anders behandelt als stetige Variablen
- Wenn man diese Variablen richtig definiert werden sie auch von R richtig behandelt

<http://www.stat.berkeley.edu/~s133/factors.html>

# Beispiel Definition von Faktoren

```
data <- c(1,2,2,3,1,2,3,3,1,2,3,3,1)
fdata <- factor(data)
fdata
```

```
## [1] 1 2 2 3 1 2 3 3 1 2 3 3 1
## Levels: 1 2 3
```

labels direkt definieren

```
rdata <- factor(data, labels=c("I", "II", "III"))
rdata
```

```
## [1] I II II III I II III III I II III III I
## Levels: I II III
```

## Weitere Möglichkeit der Definition

```
levels(fdata) <- c('I', 'II', 'III')  
fdata
```

```
## [1] I  II  II  III I   II  III III I   II  III III I  
## Levels: I II III
```

# Beispiel Monate

```
mons <- c("March", "April", "January",  
          "November", "January", "September",  
          "October", "September", "November",  
          "August", "January", "November",  
          "November", "February", "May",  
          "August", "July", "December",  
          "August", "August", "September",  
          "November", "February", "April")  
  
mons <- factor(mons)  
table(mons)
```

```
## mons
```

##	April	August	December	February	January	July
##	2	4	1	2	3	1
##	May	November	October	September		
##	1	5	1	3		

## Beispiel: ordered factor

```
mons <- factor(mons, levels=c("January", "February",  
                                "March", "April", "May", "June",  
                                "July", "August", "September",  
                                "October", "November",  
                                "December"),  
              ordered=TRUE)  
  
mons[1] < mons[2]
```

```
## [1] TRUE
```

# Rücktransformation

```
fert <- c(10,20,20,50,10,20,10,50,20)
fert <- factor(fert,levels=c(10,20,50),ordered=TRUE)
fert
```

```
## [1] 10 20 20 50 10 20 10 50 20
## Levels: 10 < 20 < 50
```

```
mean(fert)
```

```
## [1] NA
```

```
mean(as.numeric(levels(fert)[fert]))
```

```
## [1] 23.33333
```

# Tabellen mit Faktoren

```
lets <- sample(letters,size=100,replace=TRUE)
lets <- factor(lets)
table(lets[1:5])
```

```
##
## a b c d e f g h i j k l m n o p q r s t u v w x y z
## 0 1 1 0 0 0 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0
```