

02 - Univariate Statistik I | Nachbereitung der Sitzung 2

Aufgabe 1 - Der Datensatz

Gehe zu folgender Seite und klicke auf den Link zu *Datensatz Tutorat* und trage deine Daten in das Spreadsheet ein.

https://github.com/ch-bu/tutorat_statistik/blob/master/links_ressourcen.md

- shoesize: Deine Schuhgrösse
- height: Deine Grösse in Zentimetern
- gender: 0 = männlich, 1 = weiblich

Aufgabe 2 - Den Datensatz laden

Lade den Datensatz in R. Achte darauf, dass du das richtige Working-Directory bestimmt hast.

```
setwd("C:/Users/Christian/Dropbox/Christian_HiWi/_Statistik I_Tutorat/02_Univariate deskriptive Statistik I")  
getwd() # Arbeitsverzeichnis ansehen
```

```
## [1] "C:/Users/Christian/Dropbox/Christian_HiWi/_Statistik I_Tutorat/02_Univariate deskriptive Statistik I"
```

```
studis <- read.table("studidaten.csv", header = T, sep = ",") # Datensatz laden  
View(studis) # Datensatz betrachten
```

Aufgabe 3 - Skalierungen

Lasse dir die Variablenamen des Datensatzes in R anzeigen.

```
names(studis)
```

```
## [1] "shoesize" "height" "gender"
```

Bestimme die Skalierung fuer jede der drei Variablen:

shoesize: Verhaeltnisskaliert (es gibt einen natuerlichen Nullpunkt)

height: Verhaeltnisskaliert (es gibt einen natuerlichen Nullpunkt)

gender: Nominalskaliert

Aufgabe 4 - Zentrale Tendenzen

Als naechstes moechten wir die Masse der zentralen Tendenz der drei Variablen finden. Bestimme zuerst, welche Masse der zentralen Tendenz fuer welche Variable angemessen sind. Kennzeichne dies durch ein Kreuz.

| | Modus | Median | Mittelwert |
|----------|-------|--------|------------|
| shoesize | X | X | X |
| height | X | X | X |
| gender | X | | |

Aufgabe 5 - Zentrale Tendenz bestimmen

Bestimme alle *moeglichen* zentralen Tendenzen der Variablen.

```
attach(studis)
mean(shoesize)
```

```
## [1] 39.46
```

```
median(shoesize)
```

```
## [1] 39
```

```
sort(table(shoesize))
```

```
## shoesize
## 36 40 41 43 46 39 38
##  1  1  1  1  1  3  5
```

```
shoesize
Mittelwert = 39.46
Median = 39
Modus = 38
```

```
mean(height, na.rm = T)
```

```
## [1] 169.8
```

```
median(height,na.rm = T)
```

```
## [1] 168.5
```

```
sort(table(height))
```

```
## height
## 160 169 177 178 183 163 170 168
##    1    1    1    1    1    2    2    3
```

```
height
Mittelwert = 169.75
Median = 168.5
Modus = 168
```

```
sort(table(gender))
```

```
## gender
##    0    1
##    2   11
```

```
gender
Mittelwert = Macht keinen Sinn, da nominalskaliert
Median = Macht keinen Sinn, da nominalskaliert
Modus = 1 (je nach Kodierung maennlich oder weiblich)
```

Aufgabe 6 - Haeufigkeiten darstellen

- a) Trage die absoluten Haeufigkeiten der Variable `height` in R ab.

```
table(height)
```

```
## height
## 160 163 168 169 170 177 178 183
##    1    2    3    1    2    1    1    1
```

- b) Trage die relativen Haeufigkeiten der Variable `height` in R ab.

```
table(height) / length(height) # Achtung falsch! Da es einen NA im Datensatz gibt, wird die Laenge des
```

```
## height
##      160      163      168      169      170      177      178      183
## 0.07692 0.15385 0.23077 0.07692 0.15385 0.07692 0.07692 0.07692
```

```
# oder
prop.table(table(height)) # Diese Funktion ist richtig
```

```
## height
##      160      163      168      169      170      177      178      183
## 0.08333 0.16667 0.25000 0.08333 0.16667 0.08333 0.08333 0.08333
```

- c) Trage die relativen Haeufigkeiten der Variable `gender` in R ab. BONUS: Speichere die Tabelle in eine Variable und benutze die Variable, um die relativen Haeufigkeiten auszurechnen.

```
relGender <- table(gender) / length(gender)
relGender
```

```
## gender
##      0      1
## 0.1538 0.8462
```

d) Trage die kumulierten Haeufigkeiten der Variable **shoesize** in R ab.

```
relShoe <- table(shoesize) / length(shoesize)
cumsum(relShoe)
```

```
##      36      38      39      40      41      43      46
## 0.07692 0.46154 0.69231 0.76923 0.84615 0.92308 1.00000
```

```
# oder
cumsum(table(shoesize) / length(shoesize))
```

```
##      36      38      39      40      41      43      46
## 0.07692 0.46154 0.69231 0.76923 0.84615 0.92308 1.00000
```

Aufgabe 7 - Grafische Darstellung -> Theorie

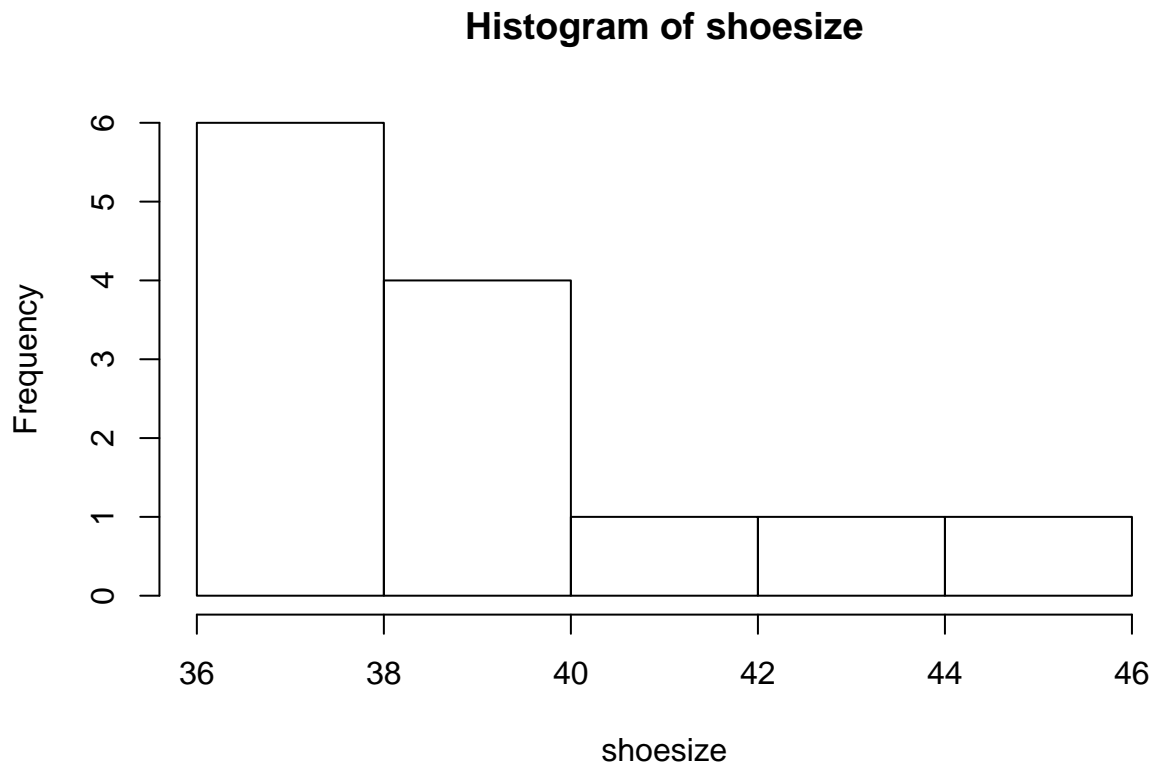
Welche grafische Darstellungen sind fuer welche Skalierungen angemessen? Kennzeichne dies durch ein Kreuz.

| | Nominal | Ordinal | Intervall | Verhaeltnis |
|--------------------|---------|---------|-----------|-------------|
| Histogramm | | | X | X |
| Balkendiagramm | X | X | | |
| Kreisdiagramm | X | X | | |
| Stammblattdiagramm | | (X) | X | X |

Aufgabe 8 - Grafische Darstellung -> Praxis

- a) Waehle eine grafische Darstellung, die fuer die Variable **shoesize** angemessen ist und erzeuge sie in R. Bestimme wenn moeglich die Schiefe und die Gipflichkeit der Verteilung.

```
hist(shoesize)
```



Grafische Darstellung: Histogramm

Schiefe: rechtsschief

Gipfligkeit: unimodal

- b) Wähle eine grafische Darstellung, die fuer die Variable **height** angemessen ist und erzeuge sie in R.
Bestimme wenn moeglich die Schiefe und die Gipfligkeit der Verteilung.

```
stem(height)
```

```
##
## The decimal point is 1 digit(s) to the right of the |
##
## 16 | 033
## 16 | 8889
## 17 | 00
## 17 | 78
## 18 | 3
```

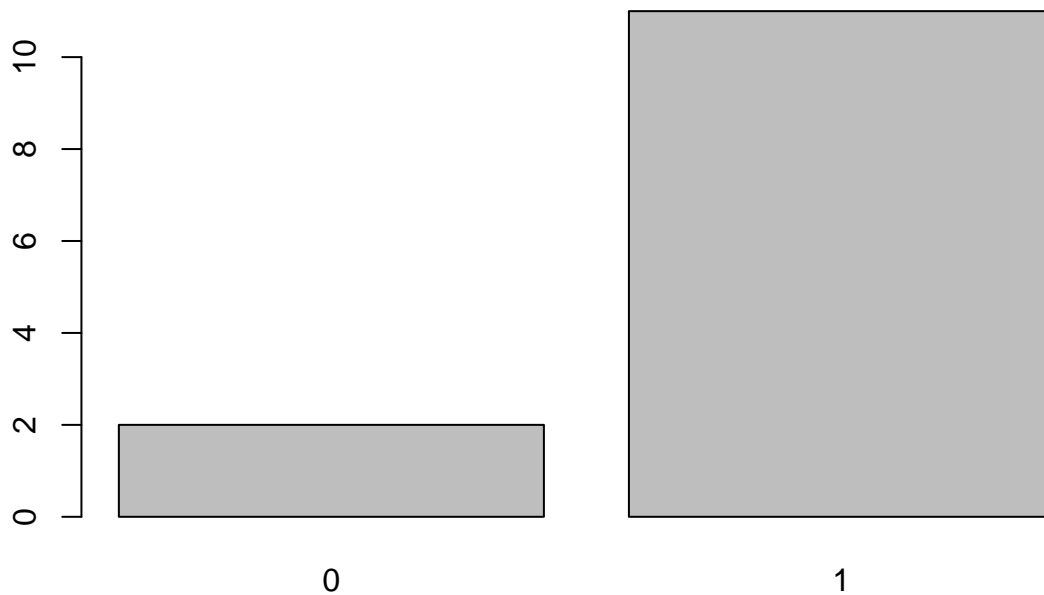
Grafische Darstellung: Stammblattdiagramm

Schiefte: Tendenz rechtsschief

Gipfligkeit: bimodal (es gibt Maenner und Frauen in der Stichprobe)

- c) Waehle eine grafische Darstellung, die fuer die Variable **gender** angemessen ist und erzeuge sie in R.
Bestimme wenn moeglich die Schiefe und die Gipfligkeit der Verteilung.

```
barplot(table(gender))
```



Grafische Darstellung: Barplot

Schiefte: keine Schiefe, da nominalskaliert

Gipfligkeit: im strengen Sinne keine Gipfligkeit. Es gibt lediglich mehr Frauen als Maenner

Aufgabe 9 - Schiefe von Verteilungen

Finde mindestens zwei Beispiele fuer eine rechtsschiefe Verteilungen.

Beispiel 1: Dauer einer Arbeitstätigkeit im Studium.

Beispiel 2: Pro-Kopf-Einkommen von arbeitenden Menschen.

Finde mindestens zwei Beispiele fuer eine linksschiefe Verteilungen,

Beispiel 1: Wahrscheinlichkeit der Sterblichkeit nach Lebensalter in Deutschland.

Beispiel 2: Abiturnoten (1 bis 15) bei denen die Mehrheit gut abgeschlossen hat.