Eberswalde University
for Sustainable
Development

Informationstechnologien im Wald I

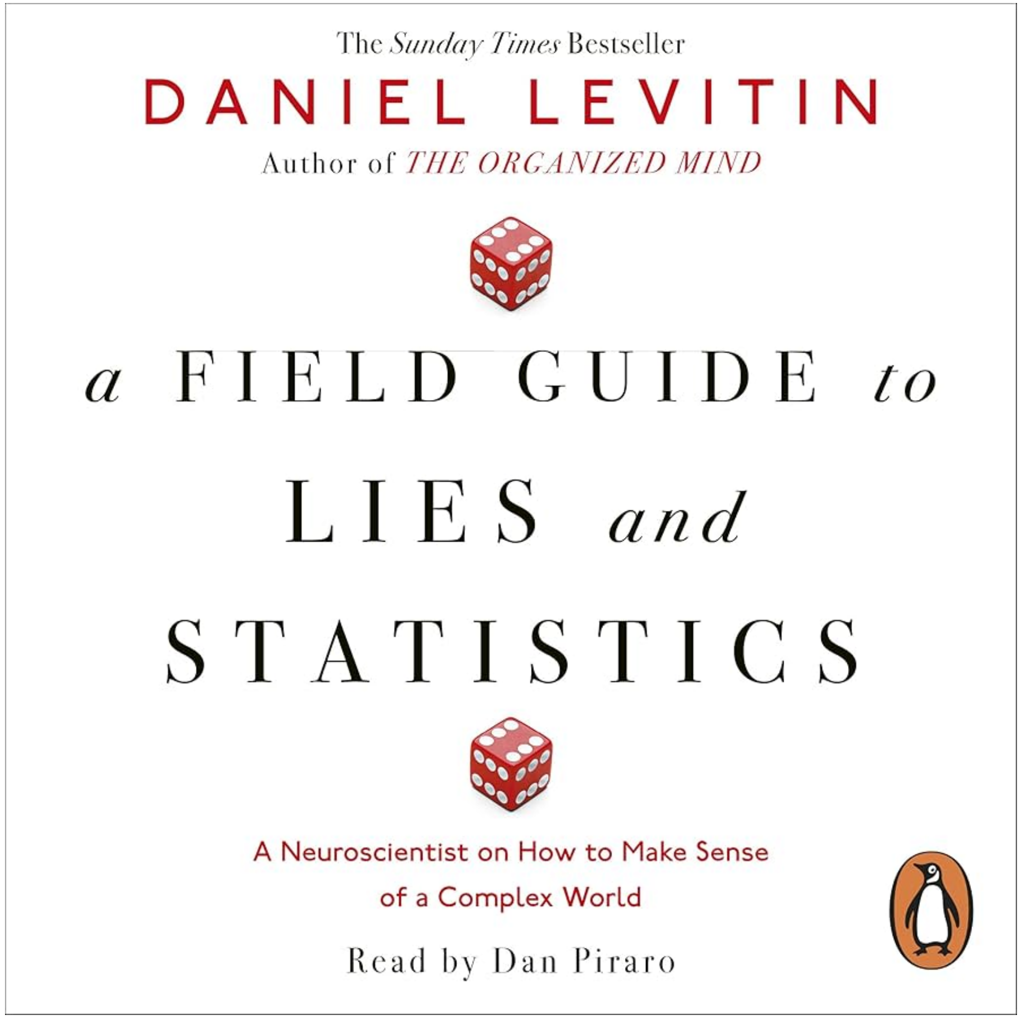
Bachelor SÖW

Sommersemester 2025

Dr. Evelyn Wallor

Inhalt

- Beispiele zum Thema *Lies & Statistics*
- Kurze Wiederholung
- Deskriptive Statistik: Lage- und Streuparameter
- Box-and-Whisker Plot



Thema *Lies & Statistics*

→ **Plausibilität**

„In the thirty-five years since marijuana laws stopped being enforced in California, the number of marijuana smokers has doubled every year.“

Beispiel aus Levitin, D. (2016)

Thema *Lies & Statistics*

→ **Plausibilität**

„In the thirty-five years since marijuana laws stopped being enforced in California, the number of marijuana smokers has doubled every year.“

$$\text{Endwert} = \text{Anfangswert} * 2^{\text{Anzahl der Jahre}}$$

Beispiel aus Levitin, D. (2016)

Thema *Lies & Statistics*

→ Behauptungen

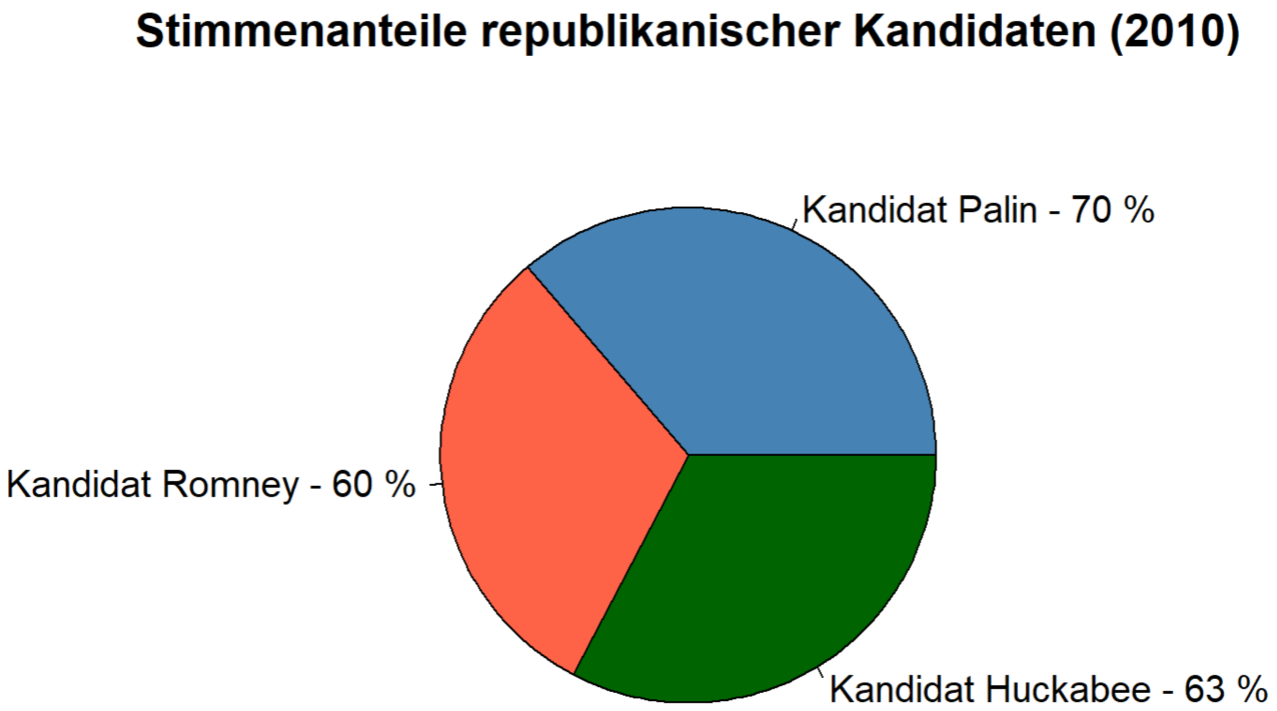
Titel: Junge Frauen sind oft gebildeter als junge Männer

- *Im Abgangsjahr 2023 waren von den 259.200 Personen mit Abitur 55% Frauen.*
- *53% der insgesamt 501.900 Hochschulabschlüsse im Prüfungsjahr 2023 entfielen auf Frauen.*
- *66% der endgültig nicht bestandenen Prüfungen wurden 2023 von männlichen Studenten abgelegt.*
- *Männer sind durchschnittlich etwas älter beim Abschluss ihres Erststudiums, nämlich 23,9 Jahre (Frauen: 23,4 Jahre).*

tagesschau.de (03.04.2025)

Thema *Lies & Statistics*

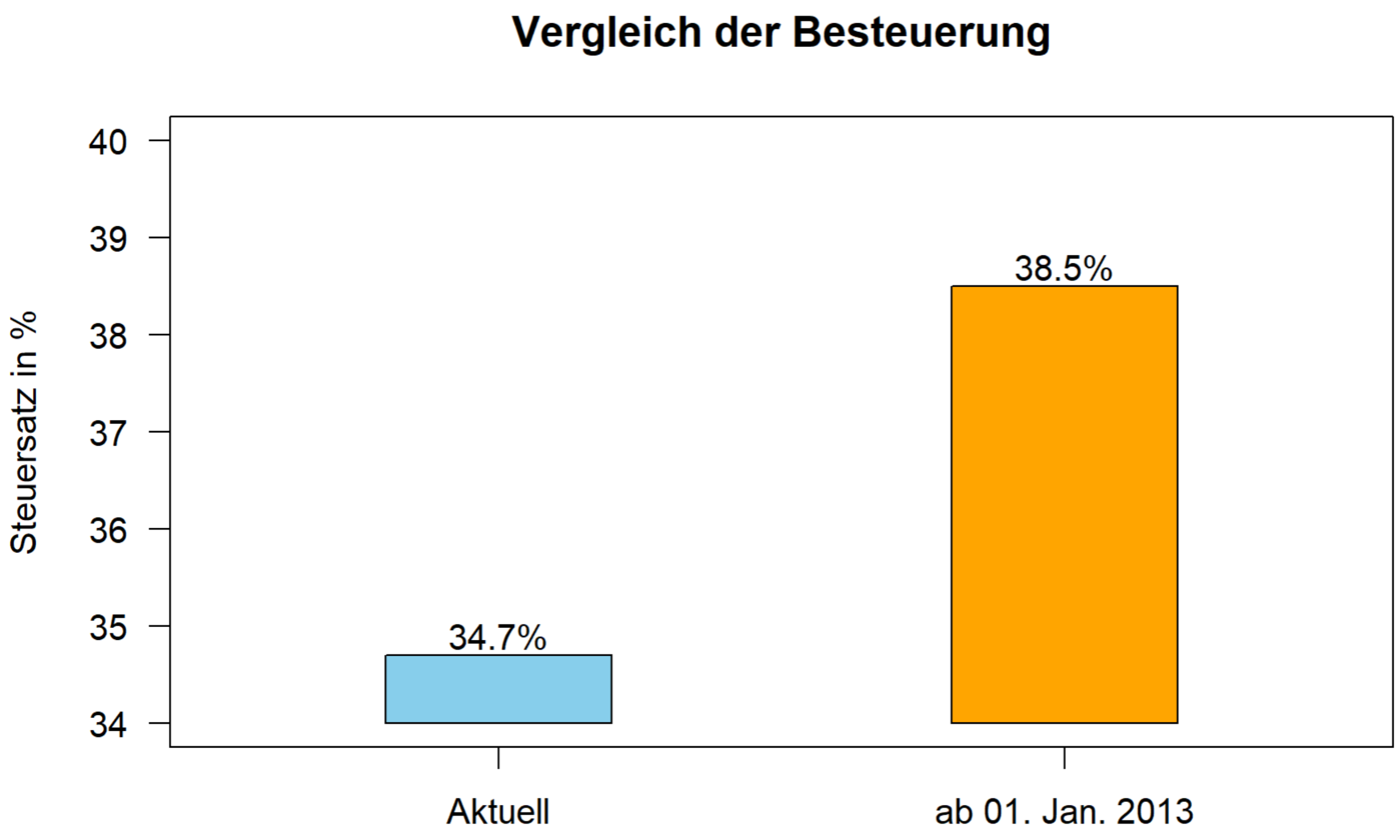
→ Visualisierung



Beispiel aus Levitin, D. (2016): Fox News (2010), eigene Darstellung

Thema *Lies & Statistics*

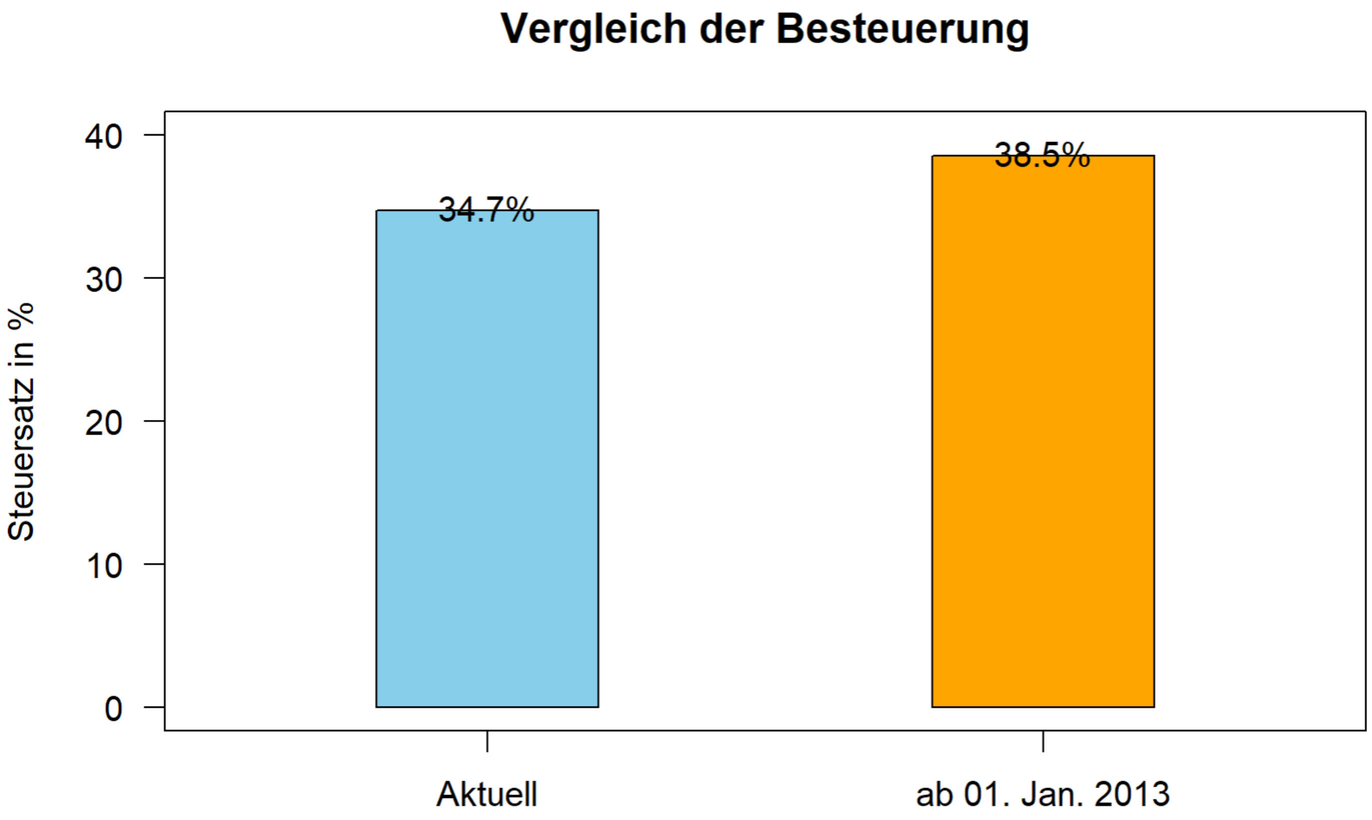
→ **Visualisierung**



Beispiel aus Levitin, D. (2016): Fox News (2012), eigene Darstellung

Thema *Lies & Statistics*

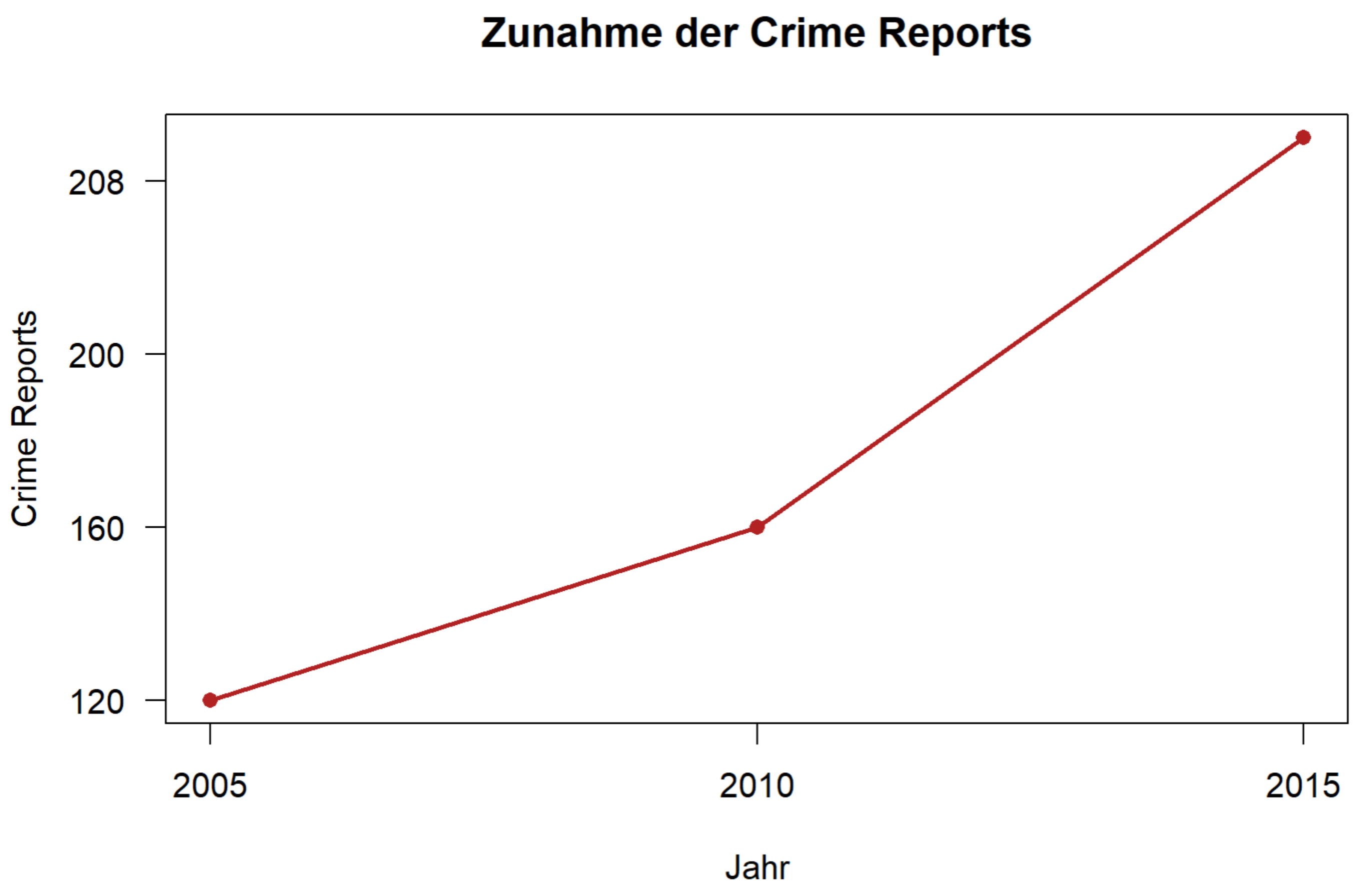
→ Visualisierung



Beispiel aus Levitin, D. (2016): Fox News (2012), berichtigt, eigene Darstellung

Thema *Lies & Statistics*

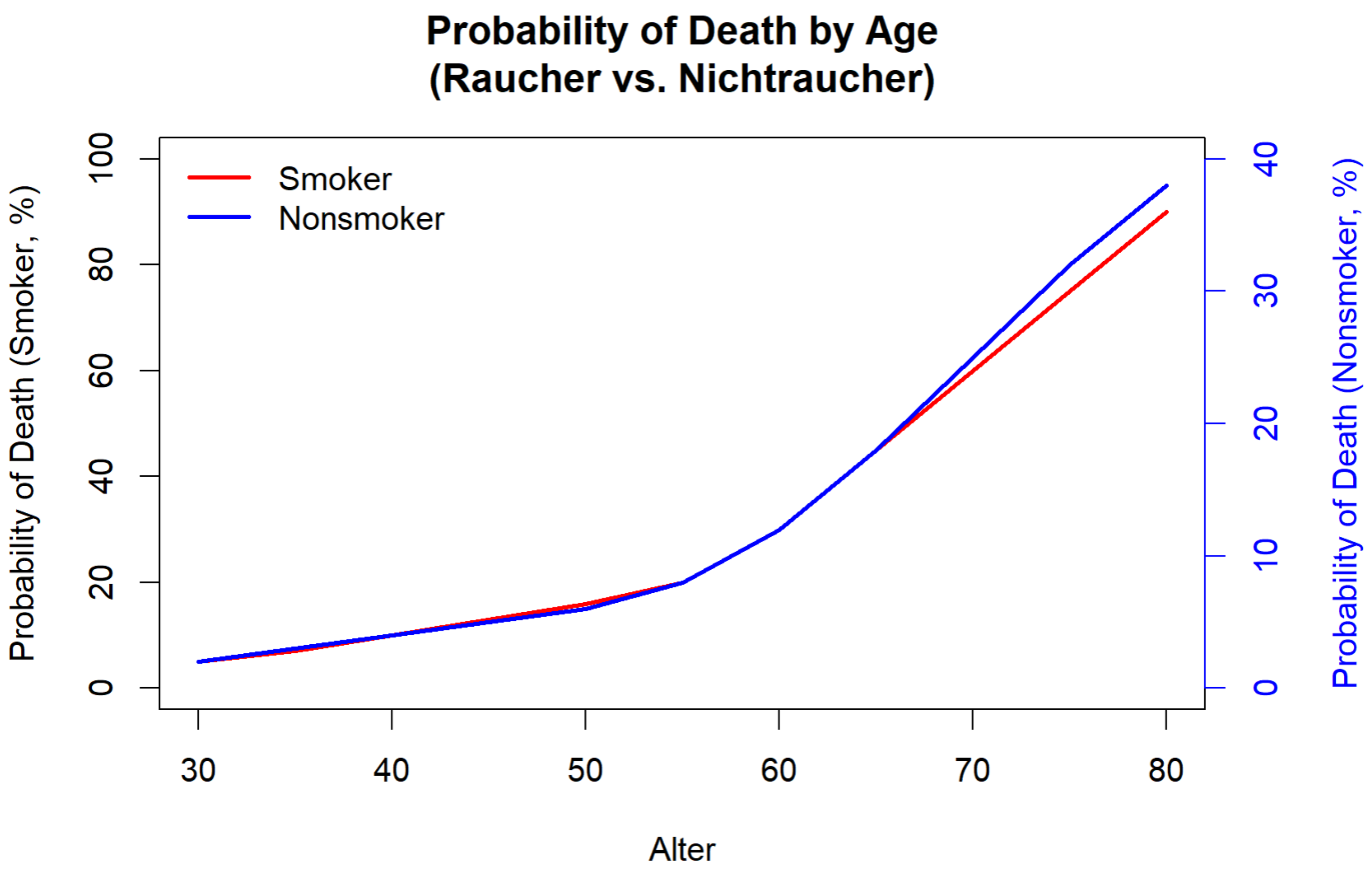
→ **Visualisierung**



Beispiel aus Levitin, D. (2016), eigene Darstellung

Thema *Lies & Statistics*

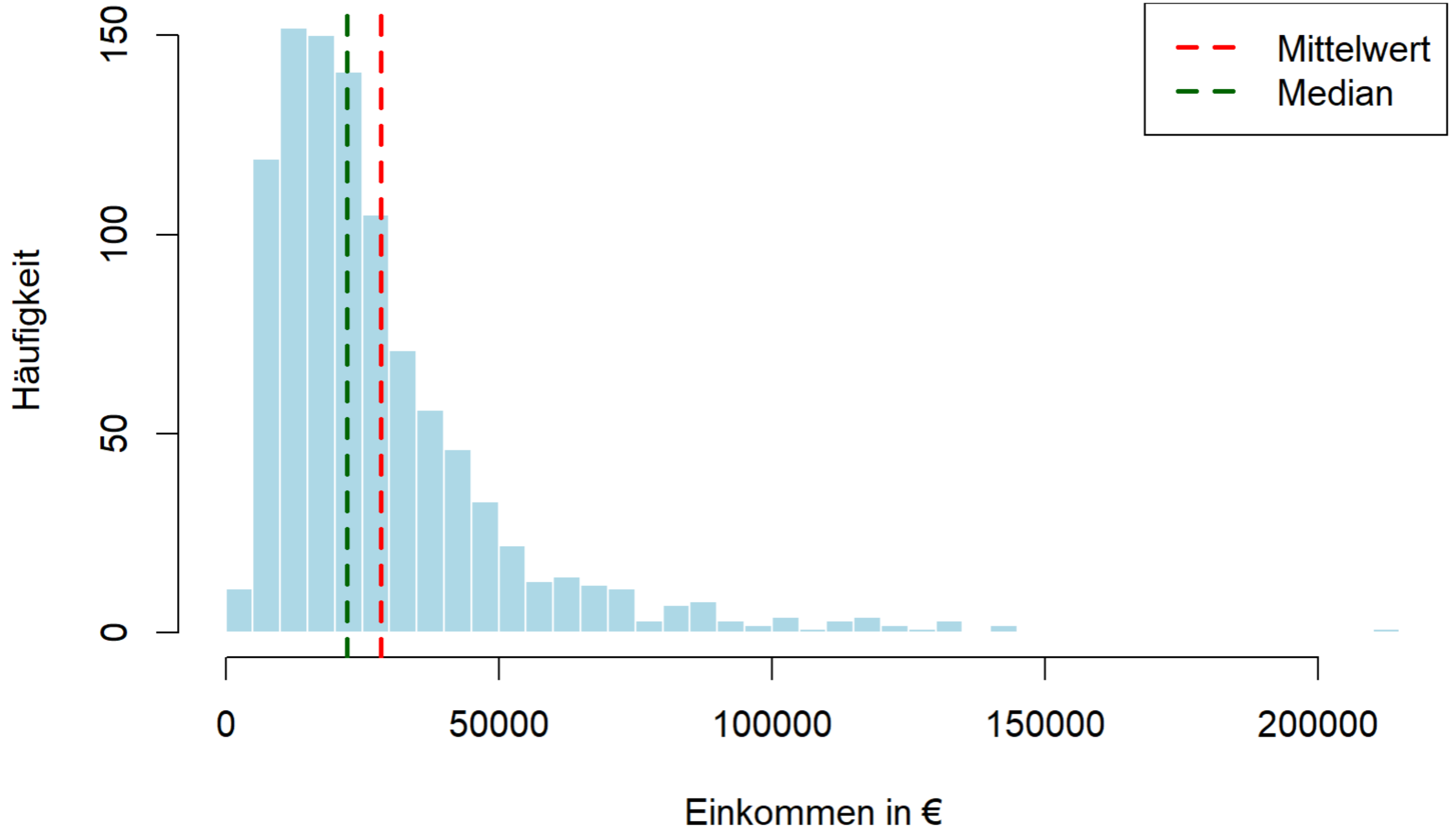
→ **Visualisierung**



Beispiel aus Levitin, D. (2016), eigene Darstellung

Thema *Lies & Statistics*

→ **Mittelwerte**



eigene Darstellung nach Levitin (2016)

Kurze Wiederholung

Deskriptive Statistik

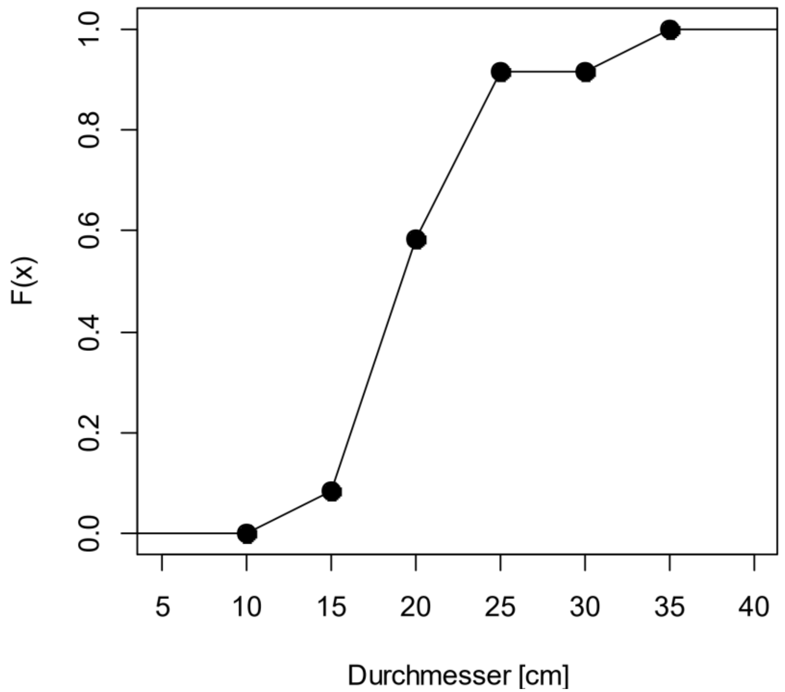
→ Beispiel Durchmesser $d_{1,3}$

Klasse	Durch-messer Intervall	Absolute Häufigkeit n_i	Relative Häufigkeit h_i	Summen-Häufigkeit N_k [absolut]	Summen-Häufigkeit H_k [relative]
1	$10 \leq x < 15$	1	0,083	1	$0{,}08\bar{3}$
2	$15 \leq x < 20$	6	0,500	7	$0{,}58\bar{3}$
3	$20 \leq x < 25$	4	0,333	11	$0{,}91\bar{6}$
4	$25 \leq x < 30$	0	0,000	11	$0{,}91\bar{6}$
5	$30 \leq x < 35$	1	0,083	12	1
	Summe	12	1.0		

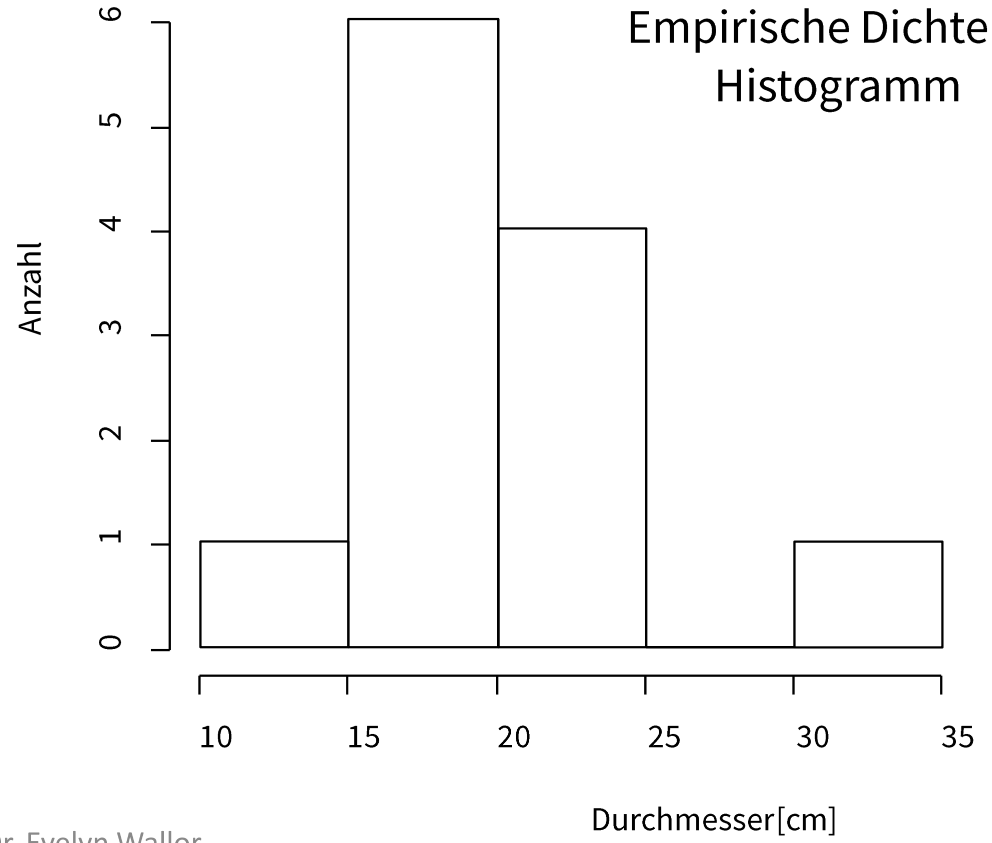
Deskriptive Statistik

→ **Häufigkeitsverteilung**

Empirische Verteilungsfunktion
„stückweise lineare Funktion“



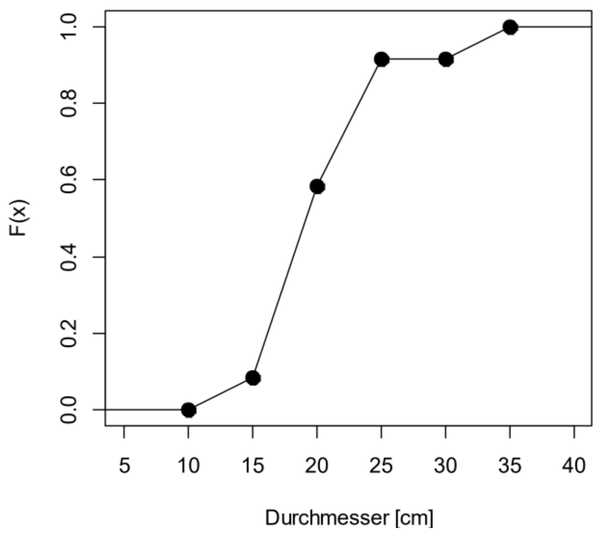
Empirische Dichte im Histogramm

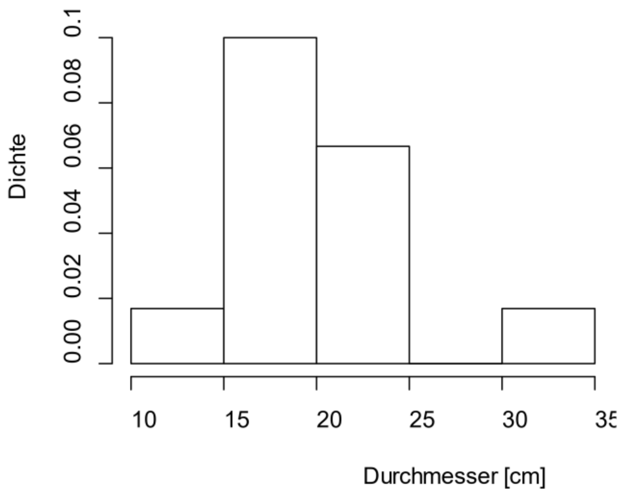


Verteilungsfunktionen

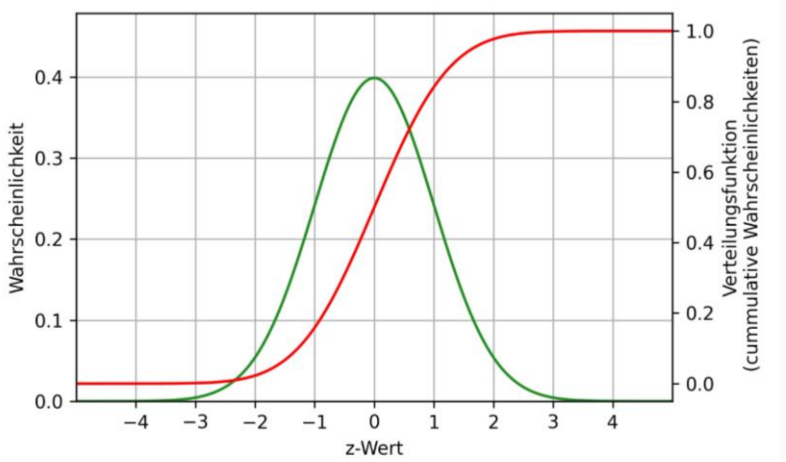
→ **Empirisch vs. theoretisch**

Empirie
(Beobachtung)





Theorie
(Wissen)
z.B. Normalverteilung $N(\mu, \sigma^2)$



Lage- und Streuparameter

Deskriptive Statistik

→ **Statistische Kennwerte (deskriptiv)**

Lageparameter

Def.

… beschreiben den Schwerpunkt oder zentrale Tendenz der Messwerte im Wertebereich der Stichprobe.

z.B.

- arithmetischer Mittelwert (metrisch)
- Modalwert oder Modus (alle Skalen)
- Median (metrisch/ordinal)

Streuparameter

Def.

… beschreiben die Variabilität der Messwerte um den jeweiligen Lageparameter.

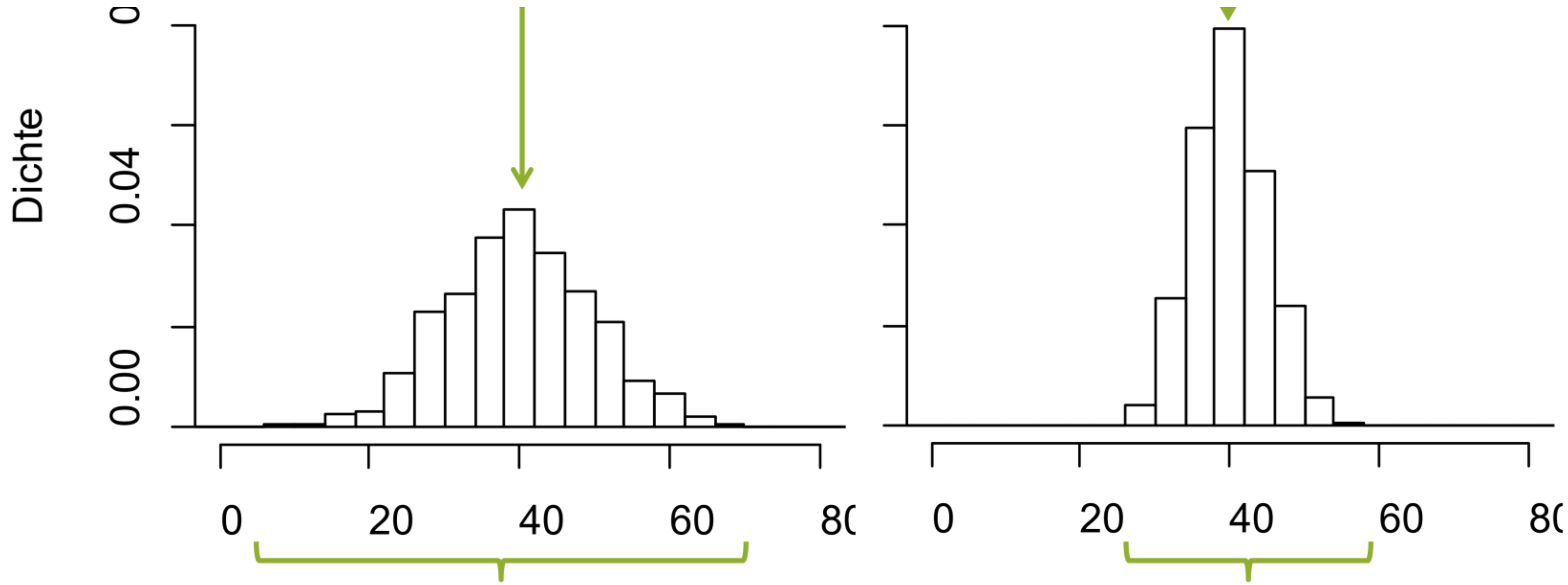
z.B.

- Varianz und Standardabweichung (metrisch)
- Interquartilsabstand (metrisch/ordinal)

Deskriptive Statistik

→ **Statistische Kennwerte (deskriptiv)**

ähnlicher Schwerpunkt bei beiden
Datensätzen



unterschiedliche Streuung

Deskriptive Statistik

→ **Lageparameter**

Den **arithmetischen Mittelwert** definieren wir mit der nachfolgenden Formel:

$$\bar{x} = \frac{1}{n} * \sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

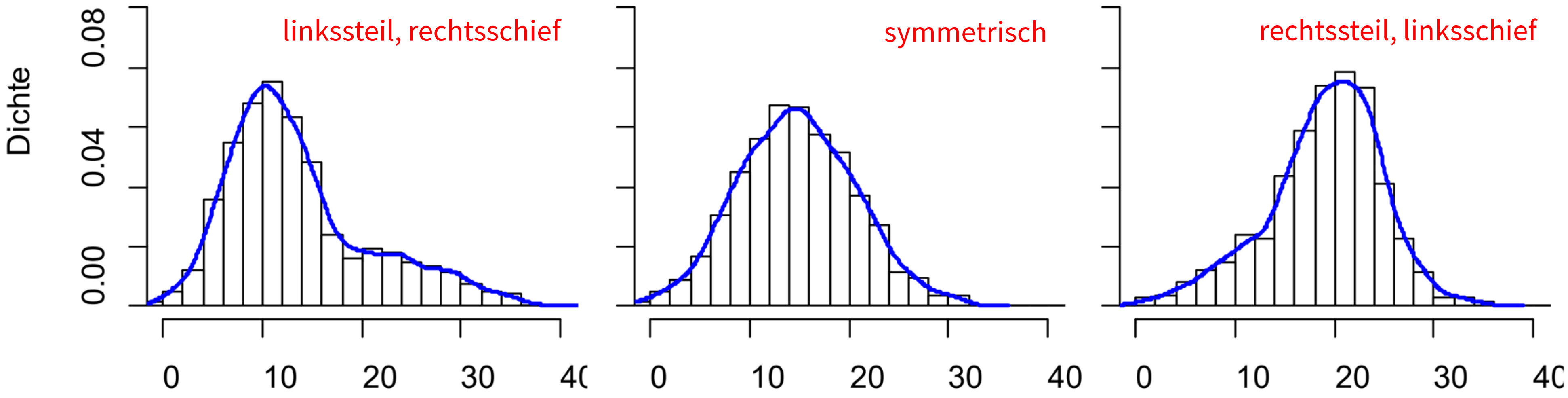
Der **Modalwert** (Modus) x_{mod} ist die Merkmalsausprägung, die in der Urliste am häufigsten vorkommt.

Der **Median** x_{med} – manchmal auch Zentralwert oder 50%-Quantil genannt – ist der mittlere Wert einer sortierten Liste. In Abhängigkeit vom Stichprobenumfang müssen folgende Fälle beachtet werden:

$$x_{med} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \textit{für ungerade Stichprobenumfänge} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & \textit{für gerade Stichprobenumfänge} \end{cases}$$

Deskriptive Statistik

→ **Lageparameter (Verteilungsmuster)**

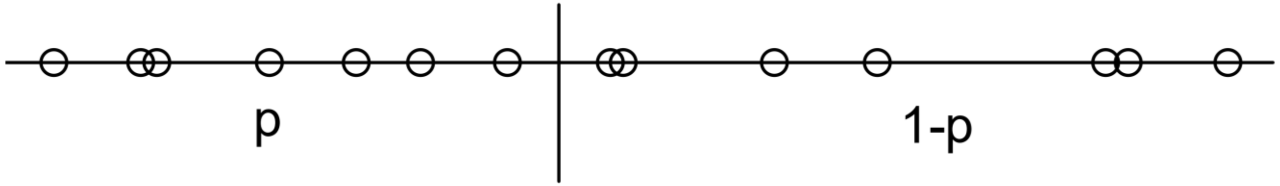


$x_{mod} < x_{med} < \bar{x}$ $\bar{x} \approx x_{med} \approx x_{mod}$ $\bar{x} < x_{med} < x_{mod}$

Deskriptive Statistik

→ **Lageparameter (Quantile)**

Als **Quantil x_p** bezeichnet man den Wert, der eine aufsteigend geordnete Liste von ***n*** Beobachtungen ungefähr im Verhältnis p zu (1-p) teilt. Der Wert für ein beliebiges Quantil liegt im Intervall [0,1].



Beispiel:

p = 0,5
(50 %)

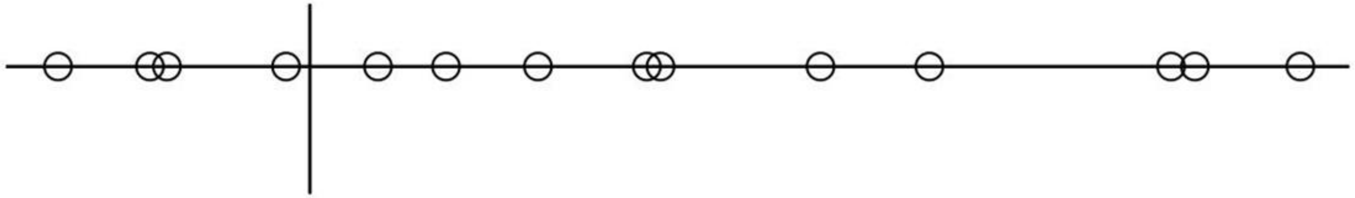
1-0,5 = 0,5
(50%)

→ In diesem konkreten Fall sprechen wir vom **Median $x_{0,5}$** oder dem 2. Quartil

Deskriptive Statistik

→ Lageparameter (Quantile)

Als **Quantil x_p** bezeichnet man den Wert, der eine aufsteigend geordnete Liste von ***n*** Beobachtungen ungefähr im Verhältnis p zu (1-p) teilt. Der Wert für ein beliebiges Quantil liegt im Intervall [0,1].



Beispiel:

p = 0,25
(25 %)

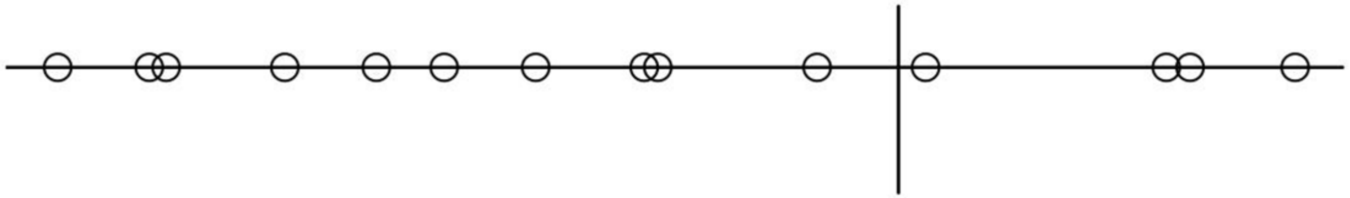
1-0,25 = 0,75
(75%)

→ In diesem konkreten Fall sprechen wir vom **1. Quartil**

Deskriptive Statistik

→ Lageparameter (Quantile)

Als **Quantil x_p** bezeichnet man den Wert, der eine aufsteigend geordnete Liste von ***n*** Beobachtungen ungefähr im Verhältnis p zu (1-p) teilt. Der Wert für ein beliebiges Quantil liegt im Intervall [0,1].



Beispiel:

p = 0,75
(75 %)

1-0,75 = 0,25
(25%)

→ In diesem konkreten Fall
sprechen wir vom **3. Quartil**

Deskriptive Statistik

→ **Lageparameter (Quantile)**

Quantile:
Hierbei handelt es sich um eine beliebige Aufteilung der Stichprobe in zwei Teile (p und 1-p)
→ z.B. $x_{0,95}$; $x_{0,99}$

Quartile:
Ein Spezialfall der Quantile, bei dem die Stichprobe in vier gleich große Bereiche geteilt wird
→ $x_{0,25}$; $x_{0.50}$; $x_{0,75}$

Quintile:
Aufteilung in 20%-Bereiche → $x_{0,20}$; $x_{0,40}$; $x_{0,60}$; $x_{0,80}$

Dezile:
Aufteilung in 10%-Bereiche →$x_{0,10}$; $x_{0,20}$; … ; $x_{0,90}$

Deskriptive Statistik

→ Lageparameter (Quantile)

Berechnung der **Quantile x_p**

$$x_p = \begin{cases} x_{(k)} & \text{für } k \text{ als nächste ganze Zahl nach } n \cdot p \\ \frac{1}{2}\left(x_{(k)} + x_{(k+1)}\right) & \text{falls } k = n \cdot p \text{ eine ganze Zahl ist} \end{cases}$$

Beispiel:
{15,8; 18,3; 19,3; 19,9; 20,9; 22,8}[1]

gesucht: $x_{0,5}$ $n \cdot p = 6 \cdot 0{,}5 = 3$ → ganze Zahl (2. Fall)

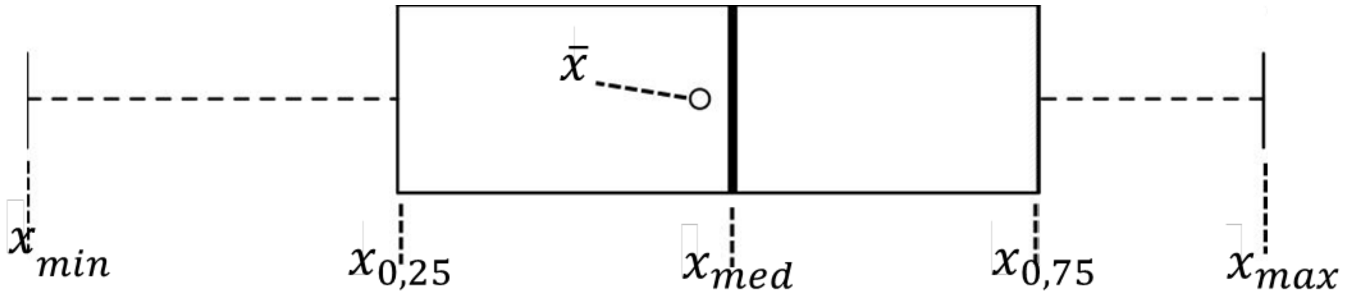
→ $x_{0,5} = 1/2(x_3 + x_{3+1}) = 1/2(19{,}3 + 19{,}9) = 19{,}6$

[1]Für derartig kleine Stichproben ist die Berechnung mancher empirischer Quantile nicht sinnvoll!

Deskriptive Statistik

→ **Lageparameter (Quantile)**

Box-Whisker-Plot (kurz: Boxplot): Hierbei handelt es sich um eine eindimensionale Darstellung der Häufigkeitsverteilung.



→ Interquartilsabstand (IQR) = $x_{0,75} - x_{0,25}$

Deskriptive Statistik

→ **Streuungsparameter (nicht im Box-and-Whisker Plot)**

Die **Varianz** (Stichproben-Varianz) ist eine Kennzahl zur Beschreibung der Streuung der Daten. Sie wird mit der nachfolgenden Formel berechnet:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Die **Standardabweichung** ist die Wurzel aus der Varianz:

$$s = \sqrt{s^2}$$

Deskriptive Statistik

→ **Streuungsparameter (im Box-and-Whisker Plot)**

Der **Interquartilsabstand** (IQR) bezeichnet Spannweite zwischen dem 1. und 3. Quartil:

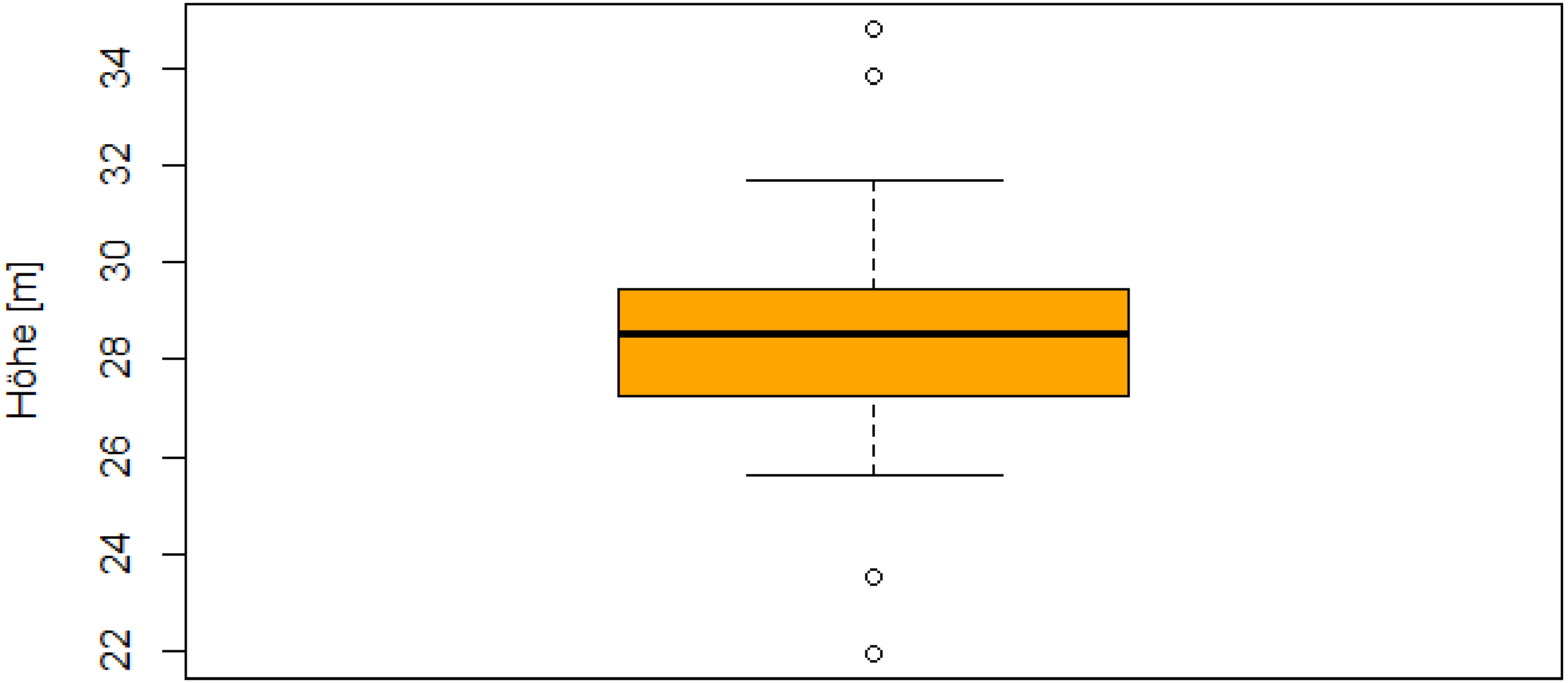
$$IQR = x_{0,75} - x_{0,25}$$

Die **Spannweite** (Range) bezeichnet den Abstand zwischen Minimum und Maximum:

$$R = x_{max} - x_{min}$$

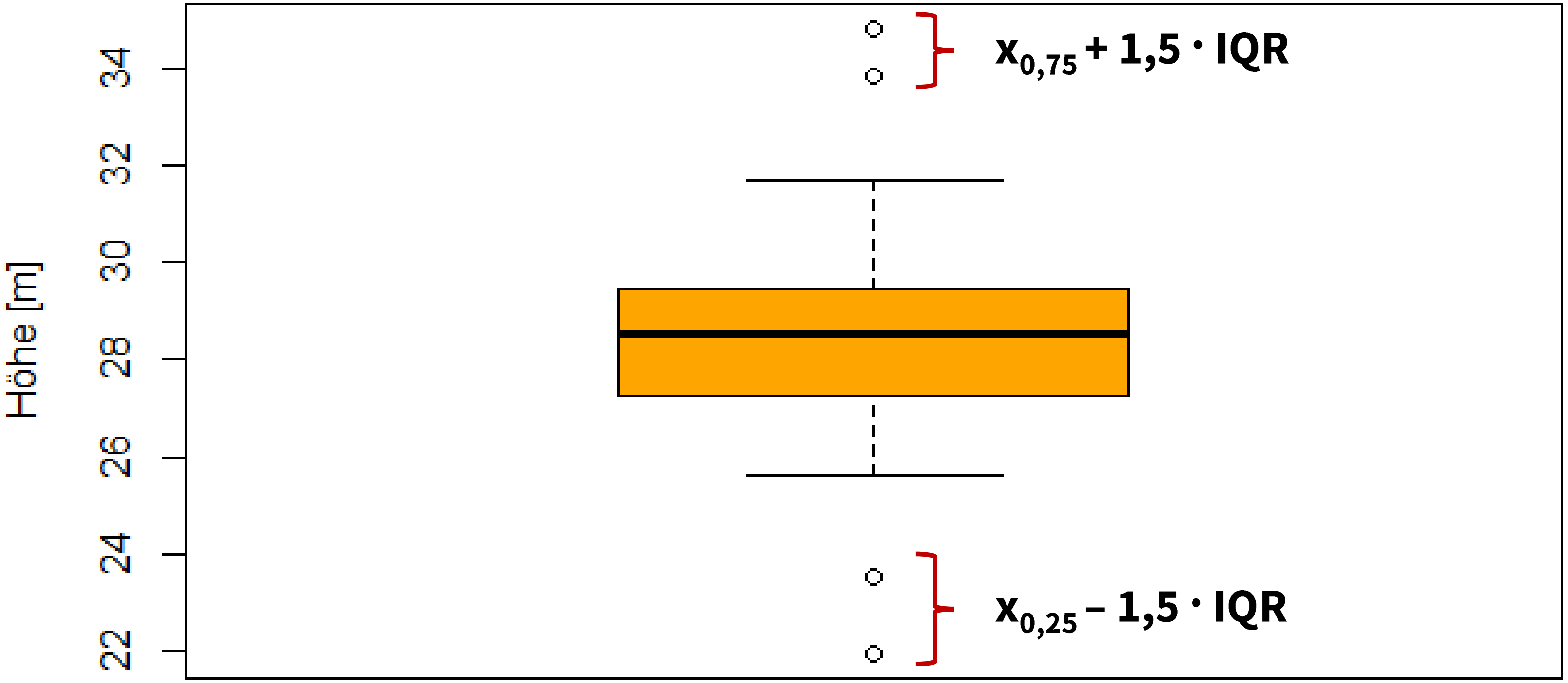
Deskriptive Statistik

→ **Deskriptive Parameter im Box-Whisker-Plot**



Deskriptive Statistik

→ Deskriptive Parameter im Box-Whisker-Plot

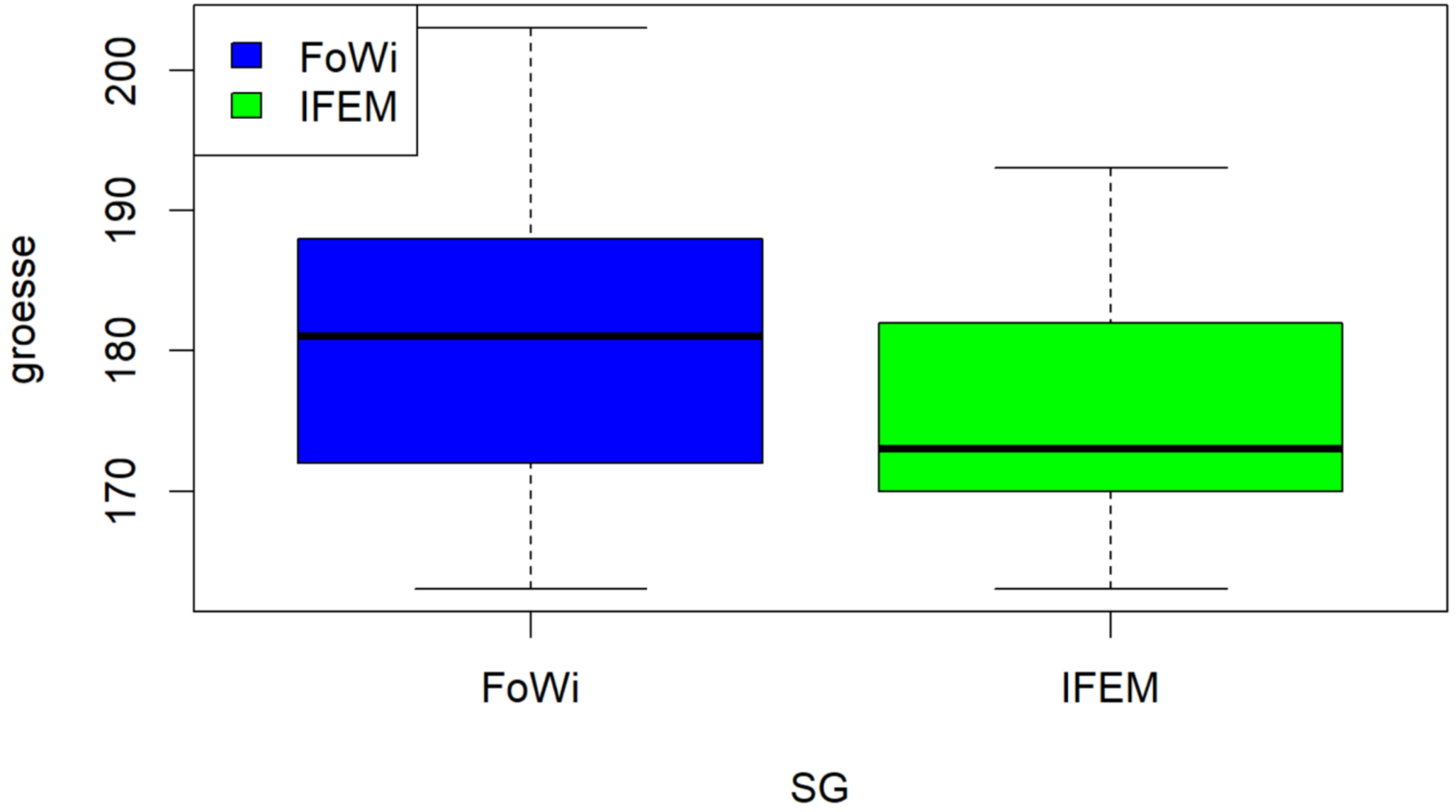


$x_{0,75} + 1{,}5 \cdot IQR$

$x_{0,25} - 1{,}5 \cdot IQR$

Deskriptive Statistik

Interpretiere!



Danke für die Aufmerksamkeit.