

## Vorlesung: Statistik I

Prof. Dr. Simone Abendschön

5. Einheit

- **Abschluss univariate Datenanalyse:**
  - Boxplots
  - Standardisierung (Variationskoeffizient  $V$  und  $z$ -Transformation)

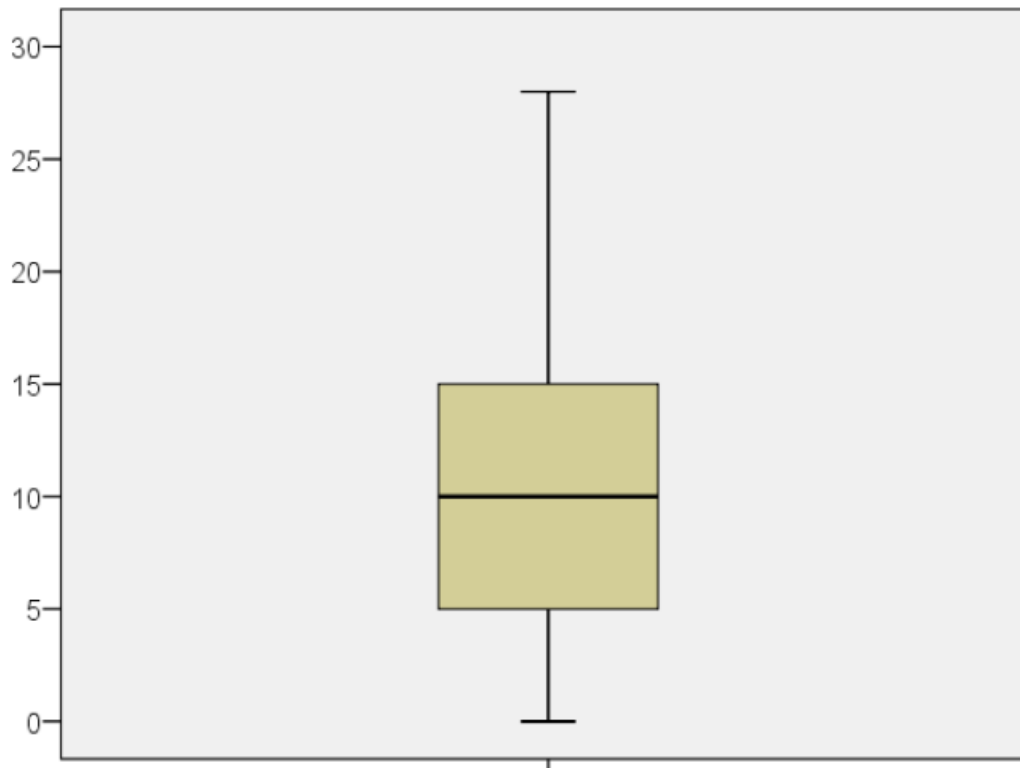
- Sie kennen die Darstellungsform des Boxplots
- Sie können ein Boxplot erstellen und interpretieren
- Sie kennen die Standardisierungsmöglichkeiten einer empirischen Verteilung mit Variationskoeffizienten und der z-Transformation

(auch Box-Whisker-Plot)

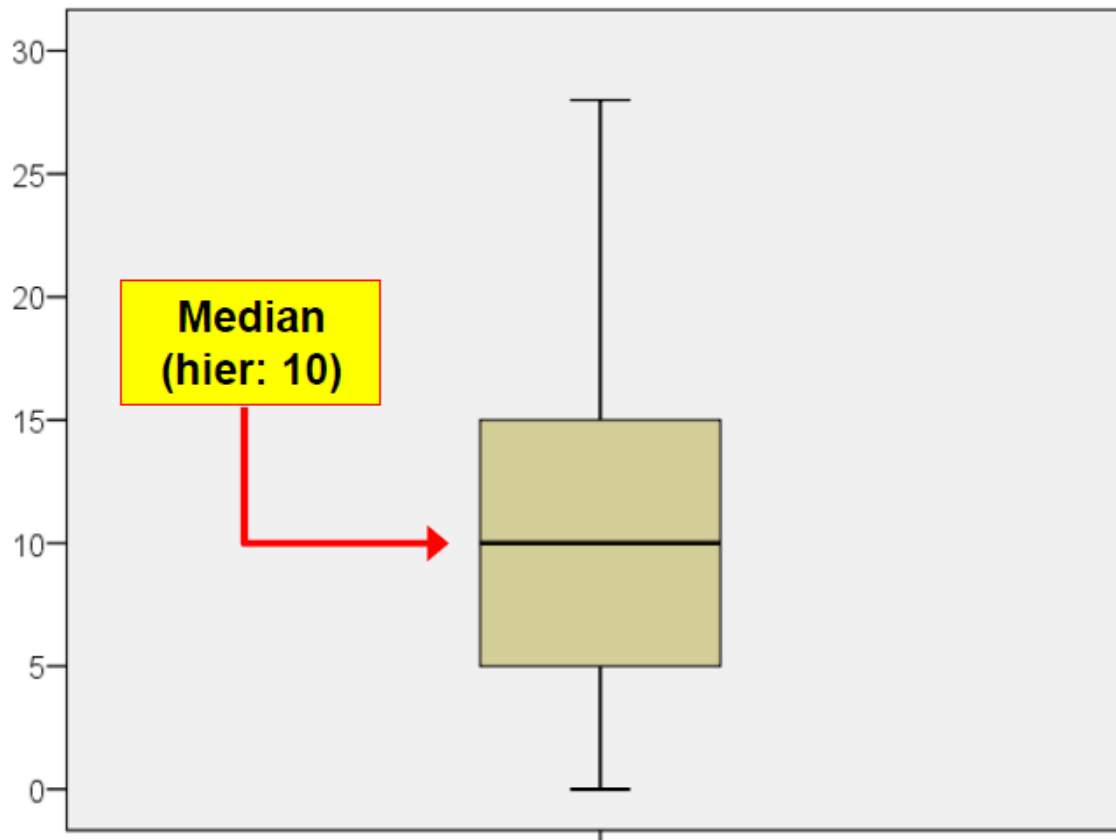
- Kompakte univariate Beschreibung einer Verteilung, Lage-/ Streumaße
- Merkmal sollte (pseudo-)metrisch skaliert sein
- Ausreißer bzw. extreme Abweichungen lassen sich gut erkennen
- Mehrere Kennwerte werden dargestellt:
  - Median
  - Unteres Quartil / Oberes Quartil → IQR
  - Variationsweite
  - Evtl. Extremwerte

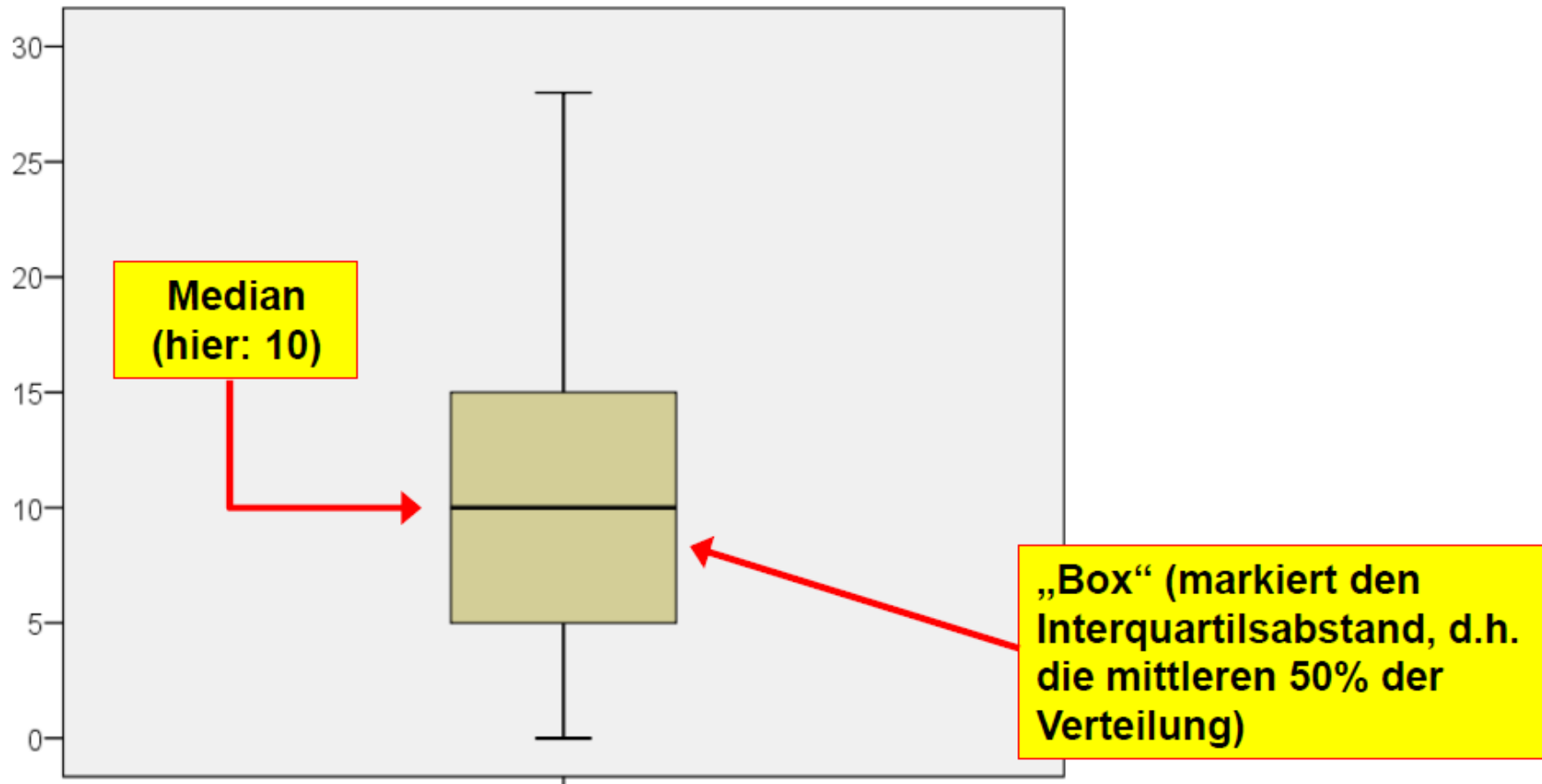
- Waagrechte und senkrechte Darstellung möglich
- Statistikprogramme i.d.R. senkrechte Darstellung

- Beispiel ohne Ausreißer, erstellt in SPSS

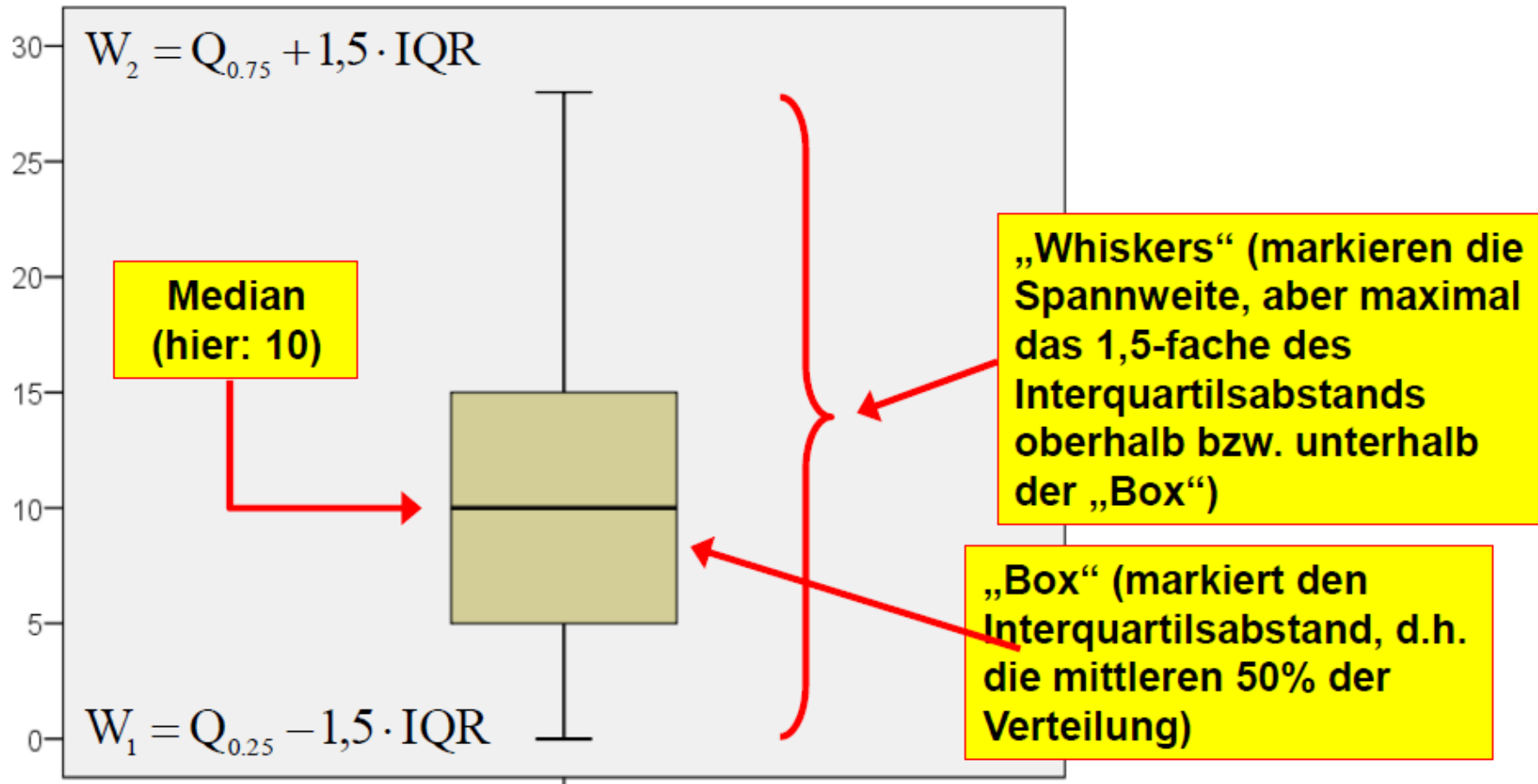


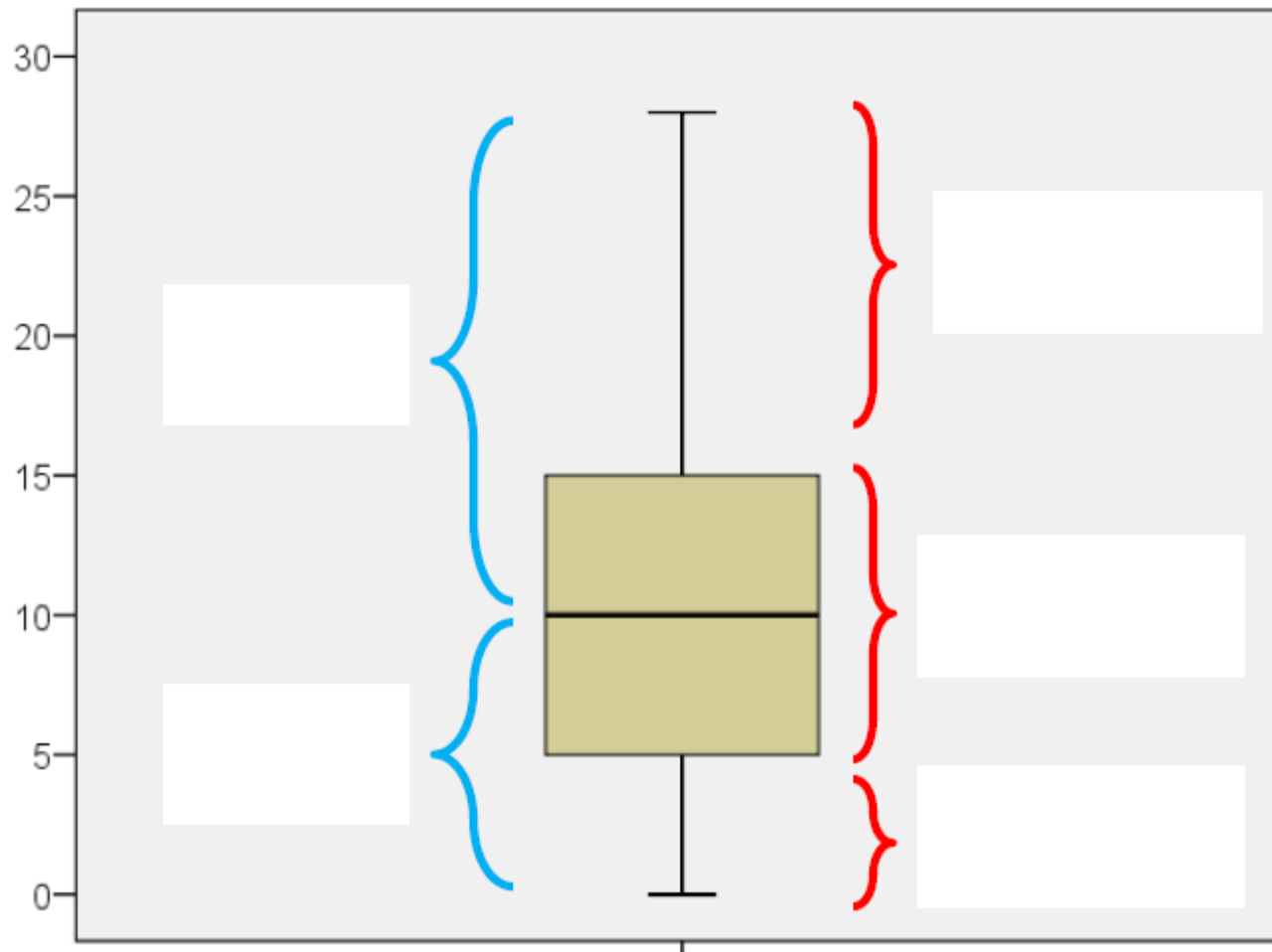
Beispiel ohne Ausreißer, erstellt in SPSS

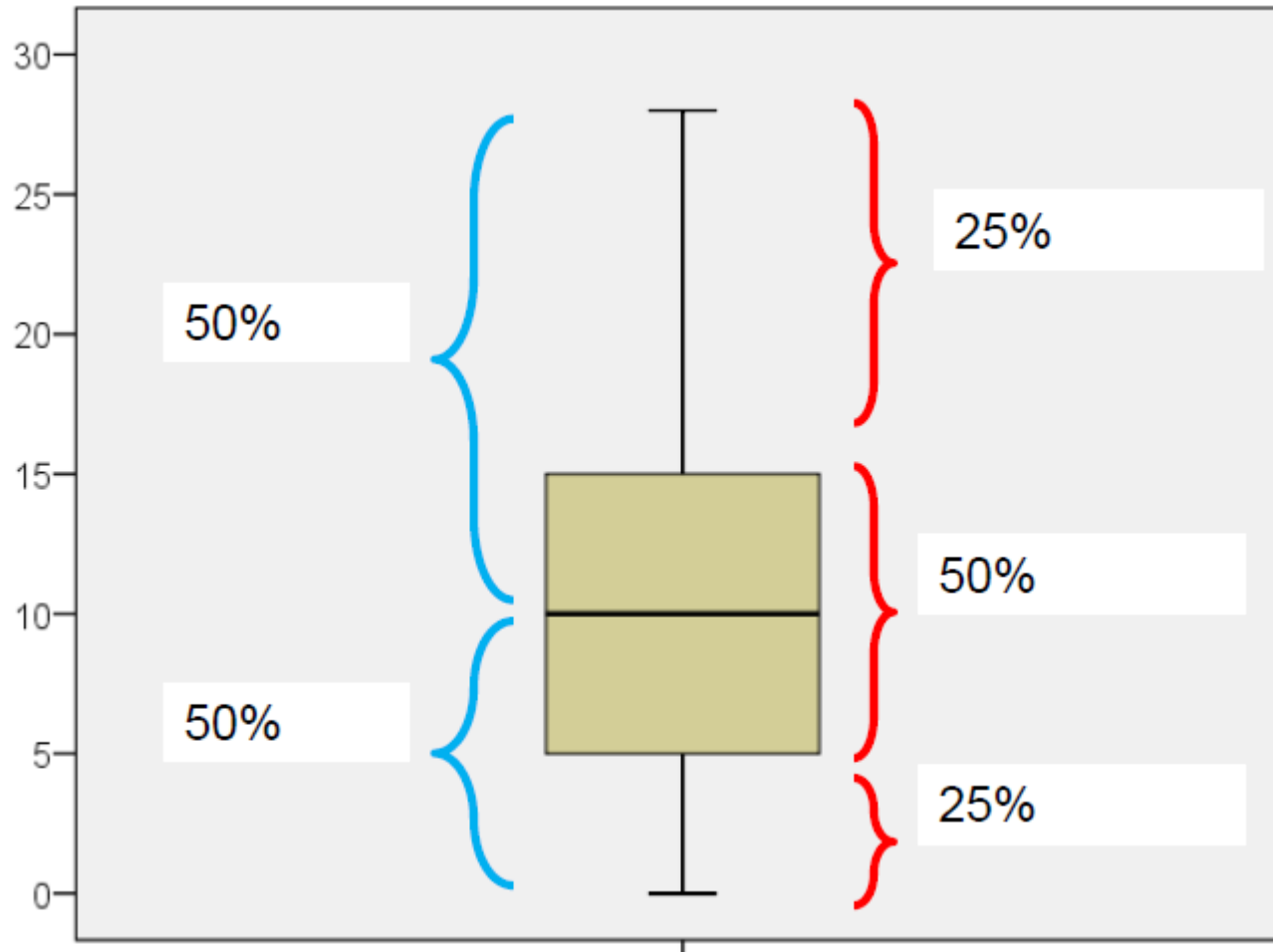












Sie haben für eine Verteilung folgende Kennwerte ermittelt:

	Wert
Minimum	=8
1. Quartil	=11
Median	=12
3. Quartil	=14
Maximum	= 16

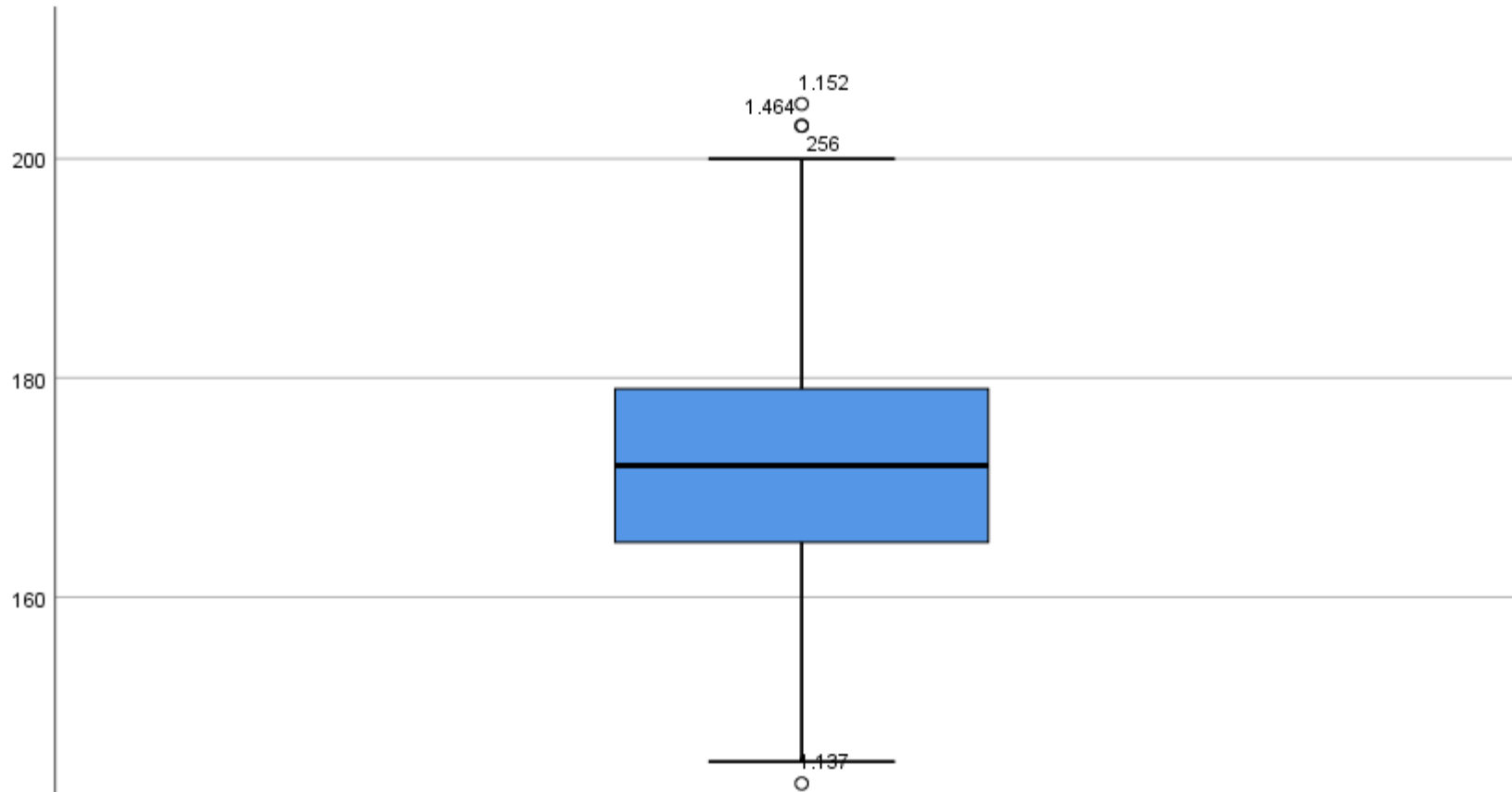
Bitte skizzieren Sie auf dieser Basis ein einfaches Boxplot (senkrecht oder waagrecht)

- Ausreißer bzw. Extremwerte einer empirischen Verteilung befinden sich außerhalb der „Whisker“ des Boxplot.
- Teilweise (SPSS) Unterscheidung zwischen Ausreißern und Extremwerten.
- Ausreißer: Werte, die um mehr als das 1.5-fache des IQR oberhalb des 3. Quartils bzw. unterhalb des 1. Quartils liegen (in SPSS durch kleine Kreise dargestellt)
- Extremwerte: Werte, die um mehr als das 3-fache des IQR oberhalb des 3. Quartils (bzw. unterhalb des 1. Quartils) liegen (in SPSS durch kleine Sternchen gekennzeichnet)

# Boxplot, Beispiel ALLBUS 2016

14

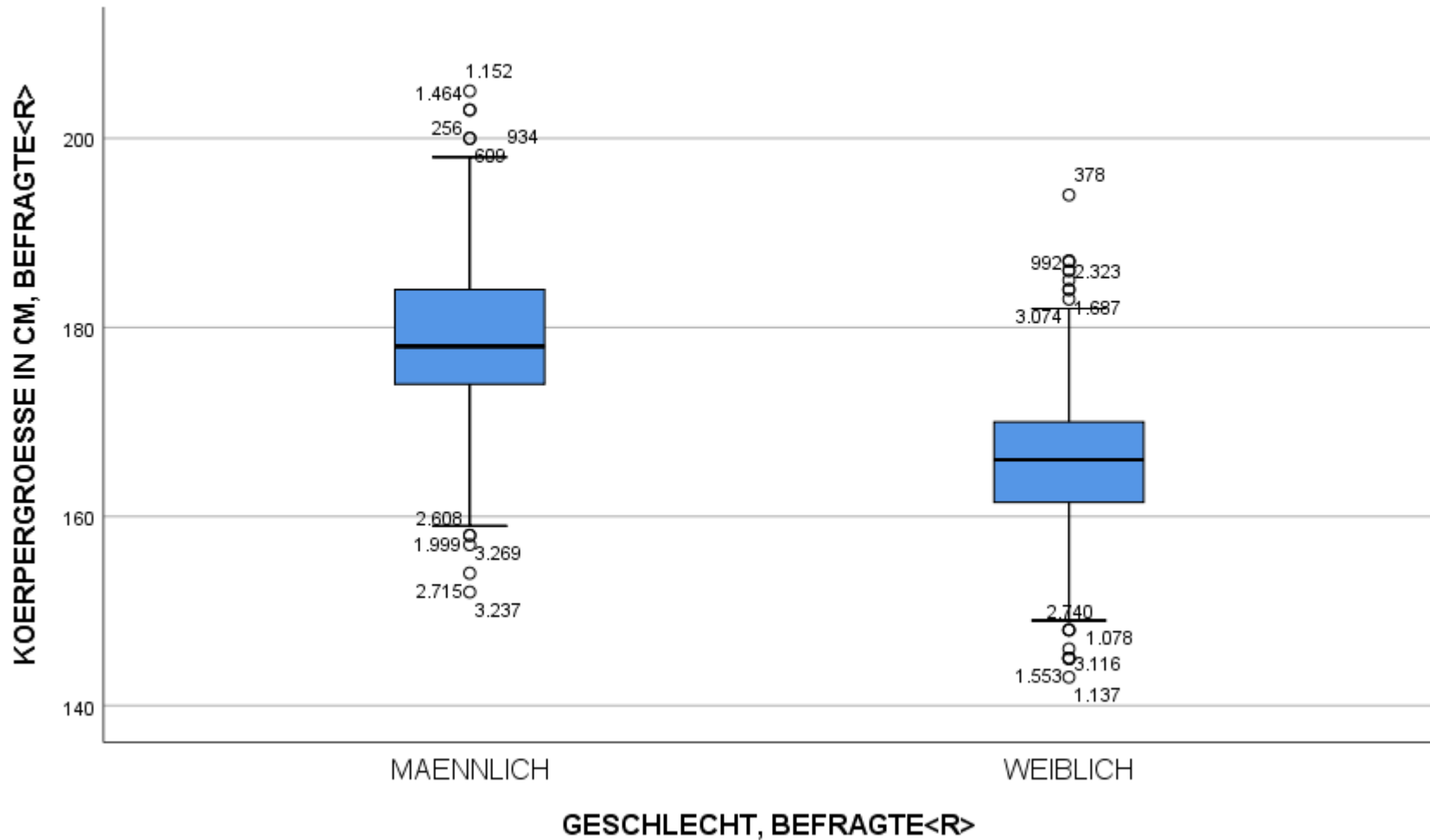
KOERPERGROESSE IN CM, BEFRAGTE<R>



# Boxplots - Vergleich zweier Gruppen

15

KOERPERGROESSE IN CM, BEFRAGTE<R>



Standardisierungsverfahren helfen, verschiedene Verteilungen miteinander zu vergleichen

- 1) **Variationskoeffizient  $V$  vergleicht Streuungen zwischen Gruppen**
- 2) Z-Transformation erlaubt Vergleich von Werten unterschiedlicher Verteilungen



Erhält man, wenn man die Standardabweichung durch den Mittelwert teilt:

$$V = \frac{s}{\bar{x}}$$

Voraussetzung:  $\bar{x}$  ungleich 0

$V > 1$ , wenn  $s > \bar{x}$

Berechnen Sie V für die folgenden beiden (Mini-)Gruppen. In welcher Gruppe streut das Einkommen stärker?

Einkommen von Studierenden:

$$x_1 = 500, x_2 = 550, x_3 = 600$$

Mittleres Einkommen von Studierenden:  
550€

Standardabweichung: 40.82

Einkommen von Millionären:

$$x_1 = 500.000, x_2 = 500050,$$

$$x_3 = 500.100$$

Mittleres Einkommen von Millionären:  
500 050 €

Standardabweichung: 40.82

Standardisierungsverfahren helfen, verschiedene Verteilungen miteinander zu vergleichen.

- 1) Variationskoeffizient  $V$  vergleicht Streuungen zwischen Gruppen
- 2) **Z-Transformation standardisiert Rohwerte, erlaubt Vergleich von Werten unterschiedlicher Verteilungen**

- Die sog. „Z-Transformation“ wandelt Rohwerte in sog. „z-Werte“ um
- Hauptanwendungen für z-Werte:
  - (1) Standardisierung von Rohwerten - Präzise Information über die *relative* Position der Rohwerte innerhalb der Verteilung
  - (2) Standardisierung von Verteilungen (ermöglicht Vergleich über verschiedene Verteilungen hinweg)

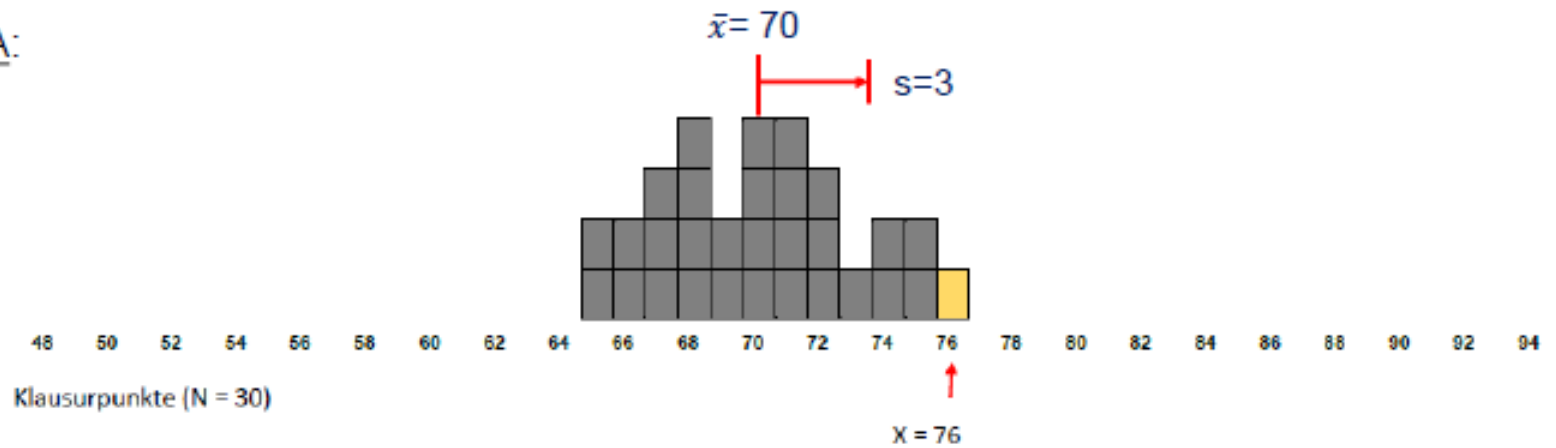
„76 Punkte in der Klausur“

Frage: Relativ gut? Relativ schlecht?

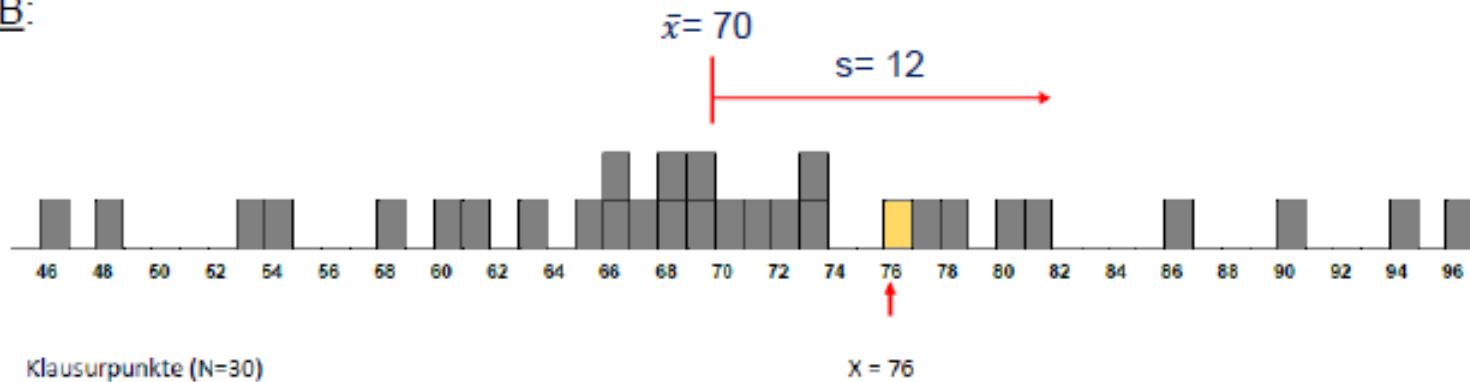
- „76 Punkte in der Klausur“ - Relativ gut? Relativ schlecht?
- Arithmetischer Mittelwert  $\bar{x} = 70$ 
  - 6 Punkte über „dem Durchschnitt“
  - vergleichsweise große oder kleine Distanz zum Mittelwert?

→ Standardabweichung  $s$ : Informationen über die durchschnittliche Abweichung vom Mittelwert

## Beispiel A:



## Beispiel B:



- Aus Rohwerten kann oft nur eingeschränkt auf ihre *relative* Position innerhalb einer Verteilung geschlossen werden
- Arithmetischer Mittelwert und Standardabweichung vermitteln zusätzliche Informationen
- Z-Transformation wandelt x-Werte in z-Werte um durch Einbezug von  $\bar{x}$  und  $s$
- ermöglicht Standardisierung der Rohwerte



# Berechnung z-Werte

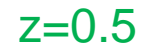
- Der z-Wert  $z_i$  eines Wertes  $x_i$  wird formal folgendermaßen berechnet:

$$z = \frac{x_i - \bar{x}}{s}$$

## Wodurch zeichnen sich z-Werte aus?

- Vorzeichen gibt an, ob sich der jeweilige Wert oberhalb (+) oder unterhalb (-) des arithmetischen Mittelwertes befindet
- Der numerische Wert informiert über die Distanz zwischen dem Rohwert und dessen Mittelwert in Standardabweichungseinheiten

”



## Beispiel:

- Für eine Verteilung einer Stichprobe von IQ-Werten mit  $\bar{x} = 100$  und  $s = 15$  würde ein Rohwert  $X = 130$  einem z-Wert von  $z = + 2.00$  entsprechen
- Positives Vorzeichen, da oberhalb des Mittelwerts
- 2 Standardabweichungen über dem Mittelwert, da 30 Punkte über dem Mittelwert

**Lisa und Bart haben jeweils an einem Leistungstest teilgenommen. Wer hat „besser“ abgeschnitten?**

Person	Wert ( $x_i$ )	Arithmetisches Mittel ( $\bar{x}$ )	Standardabweichung ( $s_x$ )
Lisa	45	25	10
Bart	60	50	25

# Übung z-Werte

1. Wie lautet der jeweilige z-Wert, der den folgenden Positionen innerhalb einer Verteilung entspricht?
  - 2 Standardabweichungen unterhalb des Mittelwertes:
  - 0.5 Standardabweichungen oberhalb des Mittelwertes:
  - 1.5 Standardabweichungen unterhalb des Mittelwertes:
2. Beschreiben Sie die Position der folgenden z-Werte innerhalb der Verteilung:
  - $z = -1.5$ :
  - $z = 0.25$ :
3. Wie lautet der jeweilige z-Wert für die nachfolgenden X-Werte ( $\bar{x} = 30$  und  $s=8$ )?
  - $X = 32$ :
  - $X = 26$ :

- **Häufigkeiten und Verteilungen eines Merkmals beschreiben, bestimmen und/oder darstellen**
- **Zentrale Kennwerte (Lage, Streumaße) können berechnet werden**
- **Boxplots und Standardisierungsverfahren sind bekannt**