

# LV Datengestützte Analysemethoden

## 4. DATEN AUSWERTEN

Sommersemester 2018  
FH Joanneum Graz  
Studiengang Journalismus und Public Relations

Lehrender: Stefan Kasberger

**Stefan Kasberger**  
**@stefankasberger**



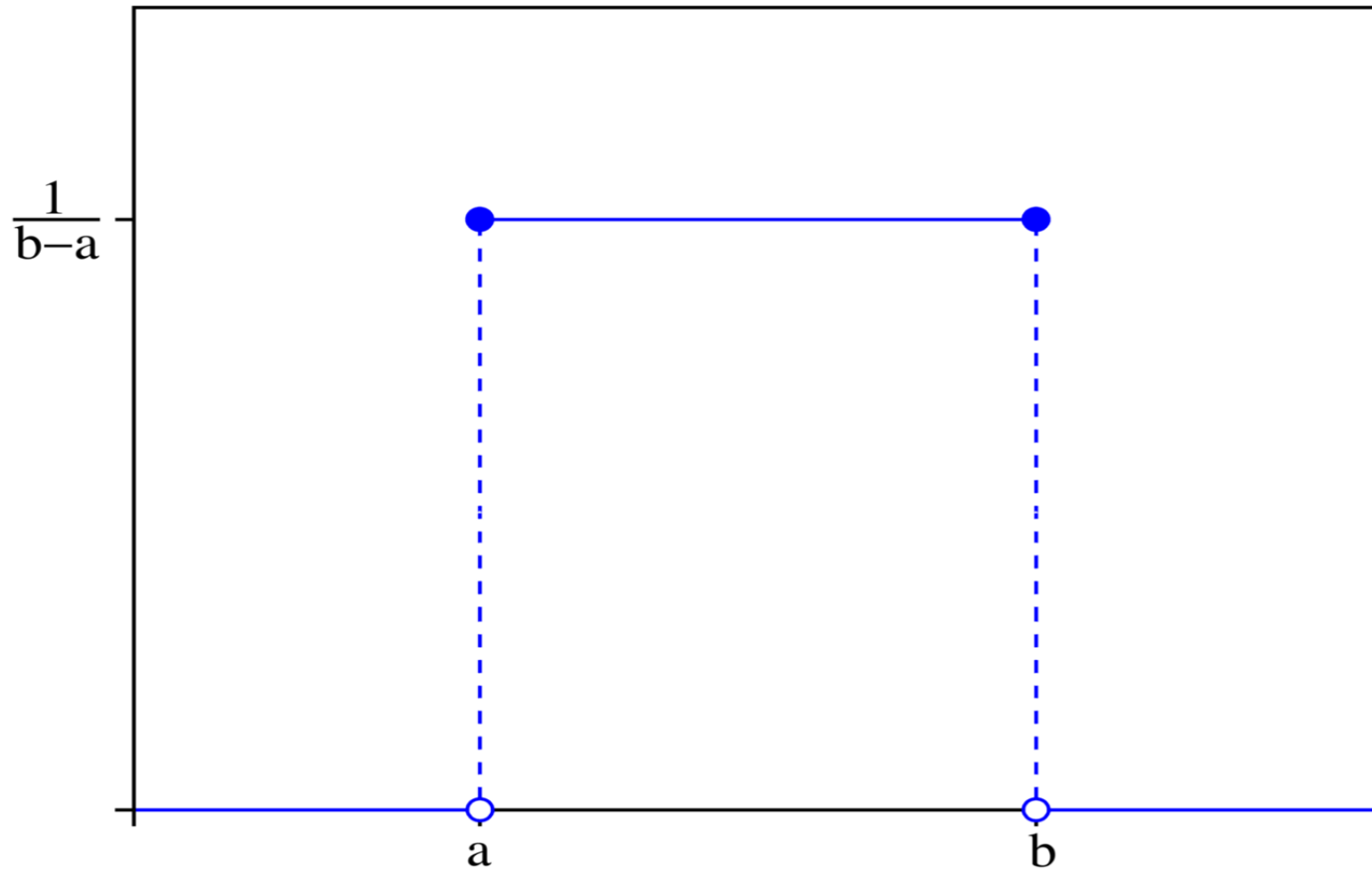
Dieses Werk ist lizenziert unter einer **Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen 4.0 International Lizenz**.

# Warum analysieren?

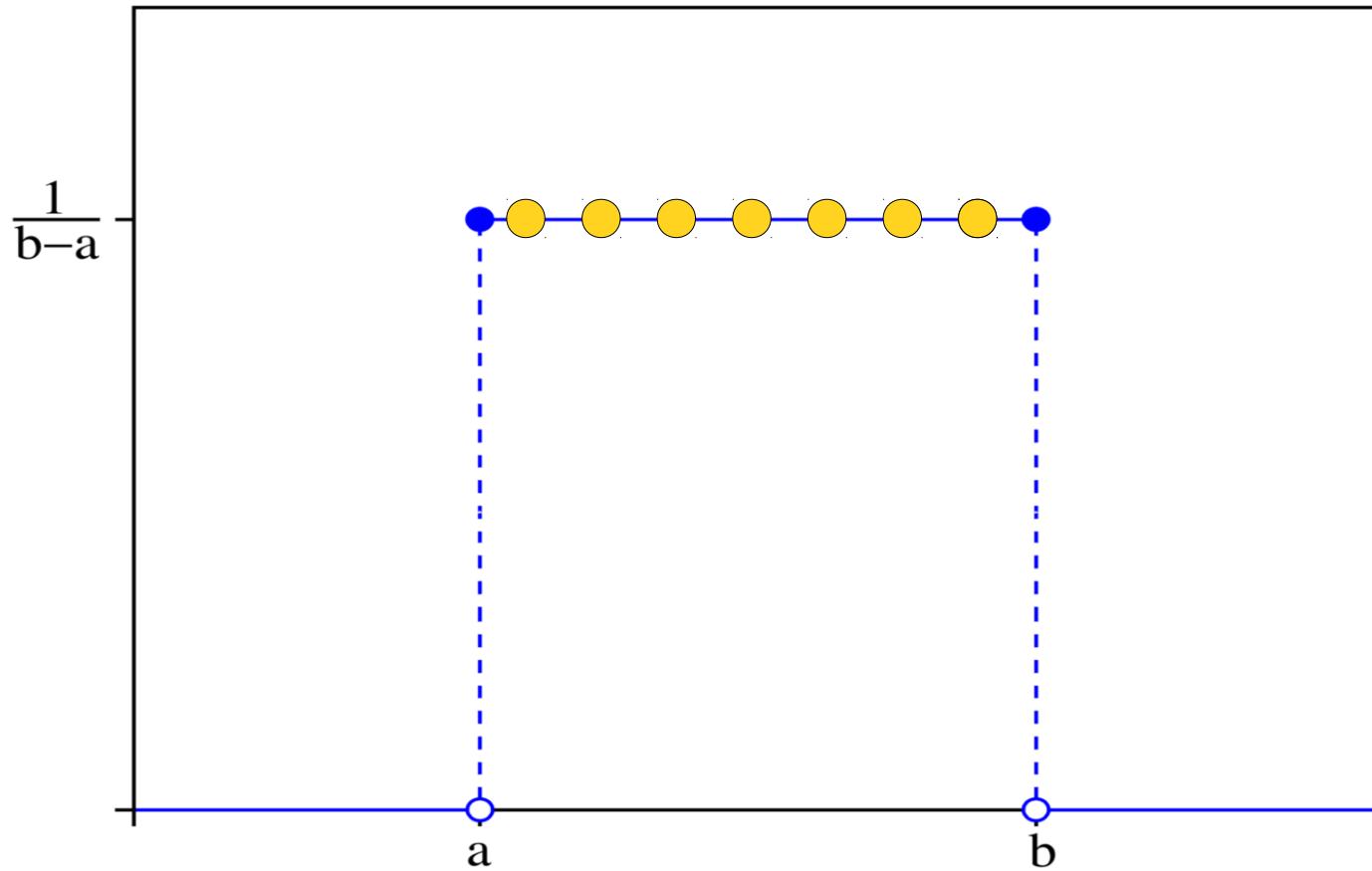
- Fragestellung beantworten
- Zusammenhänge finden
- Datensets verstehen
- Muster finden

# VERTEILUNGEN

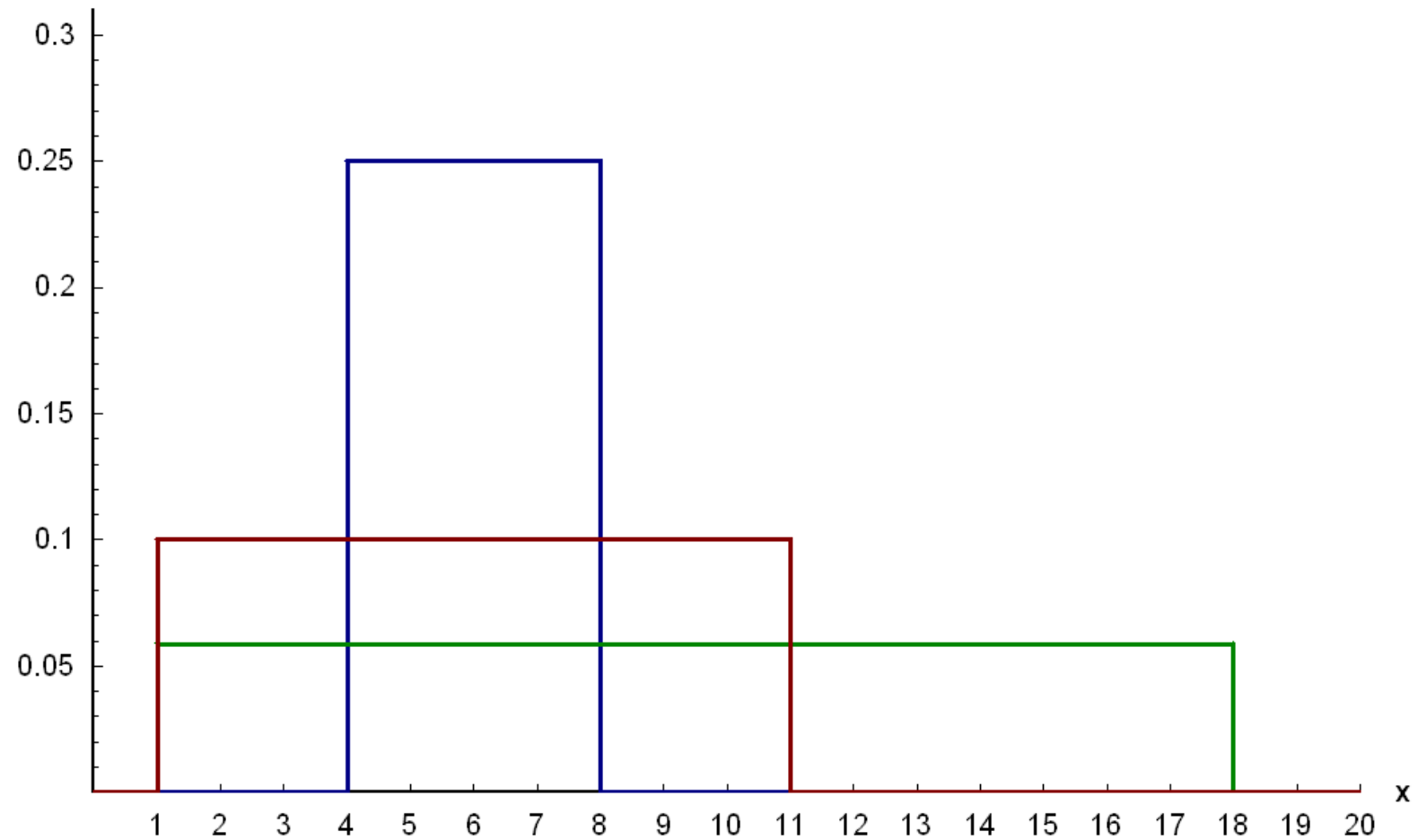
# Uniformverteilung



# Uniformverteilung



# Uniformverteilung



# Uniformverteilung

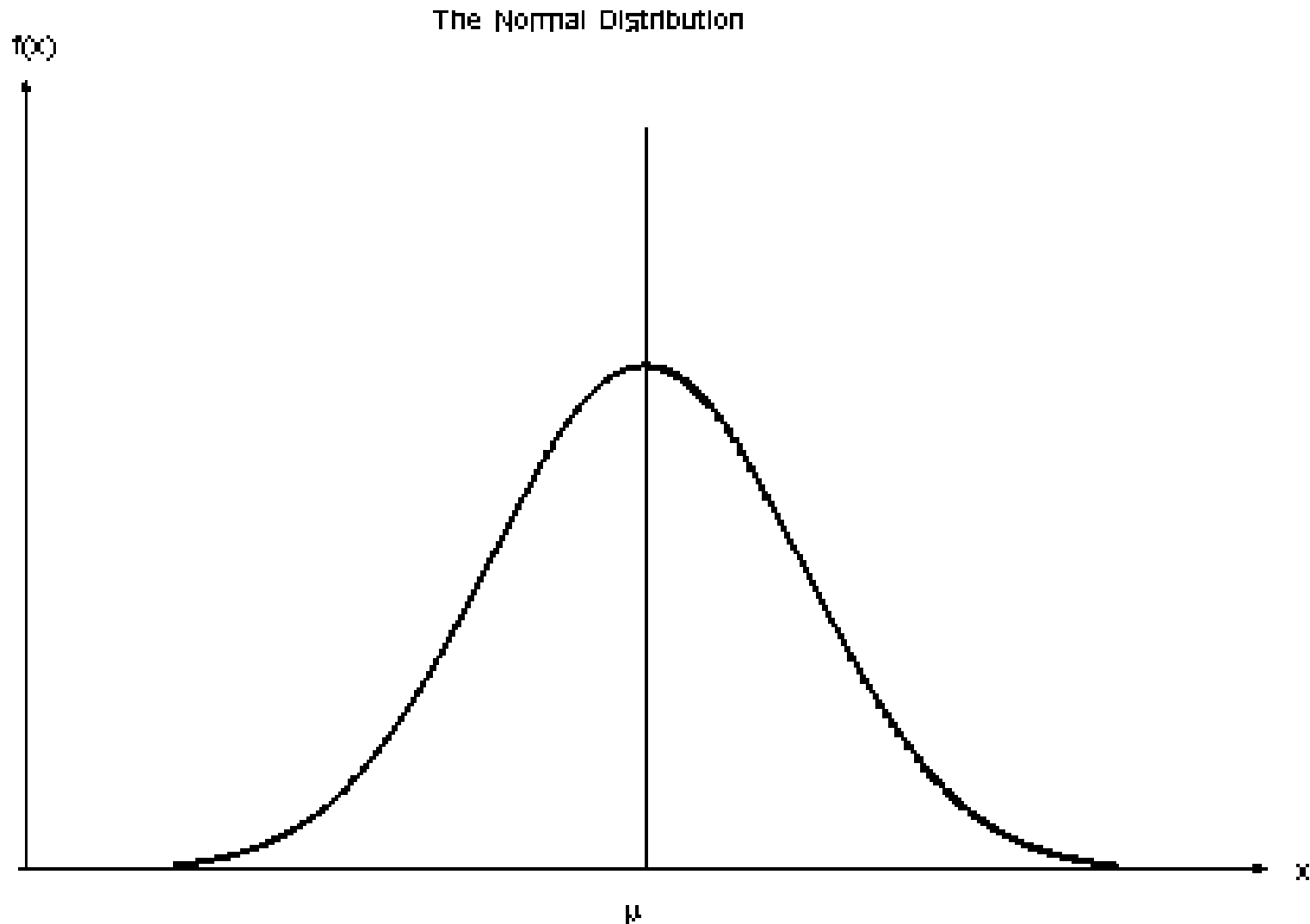


# Beispiele

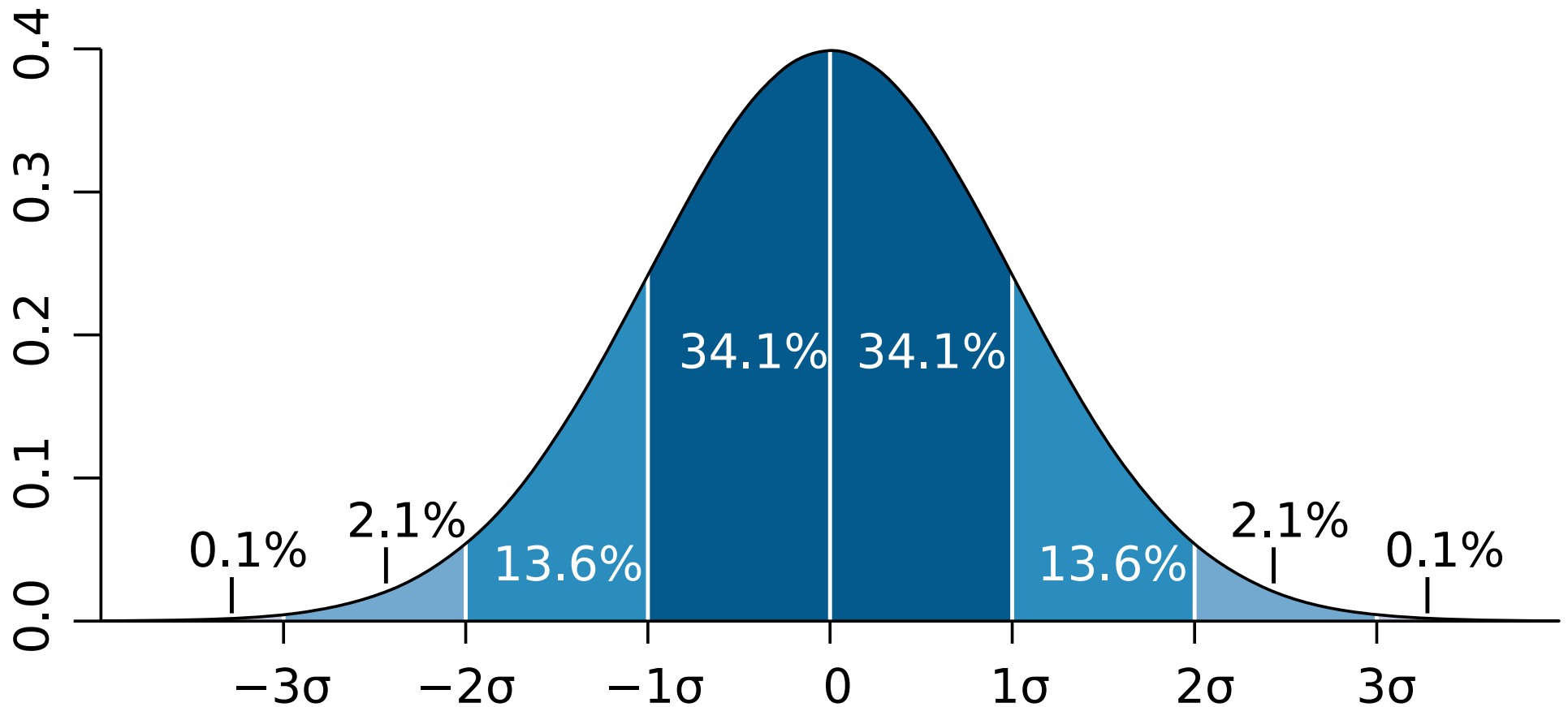
- Würfel
- Münzwurf



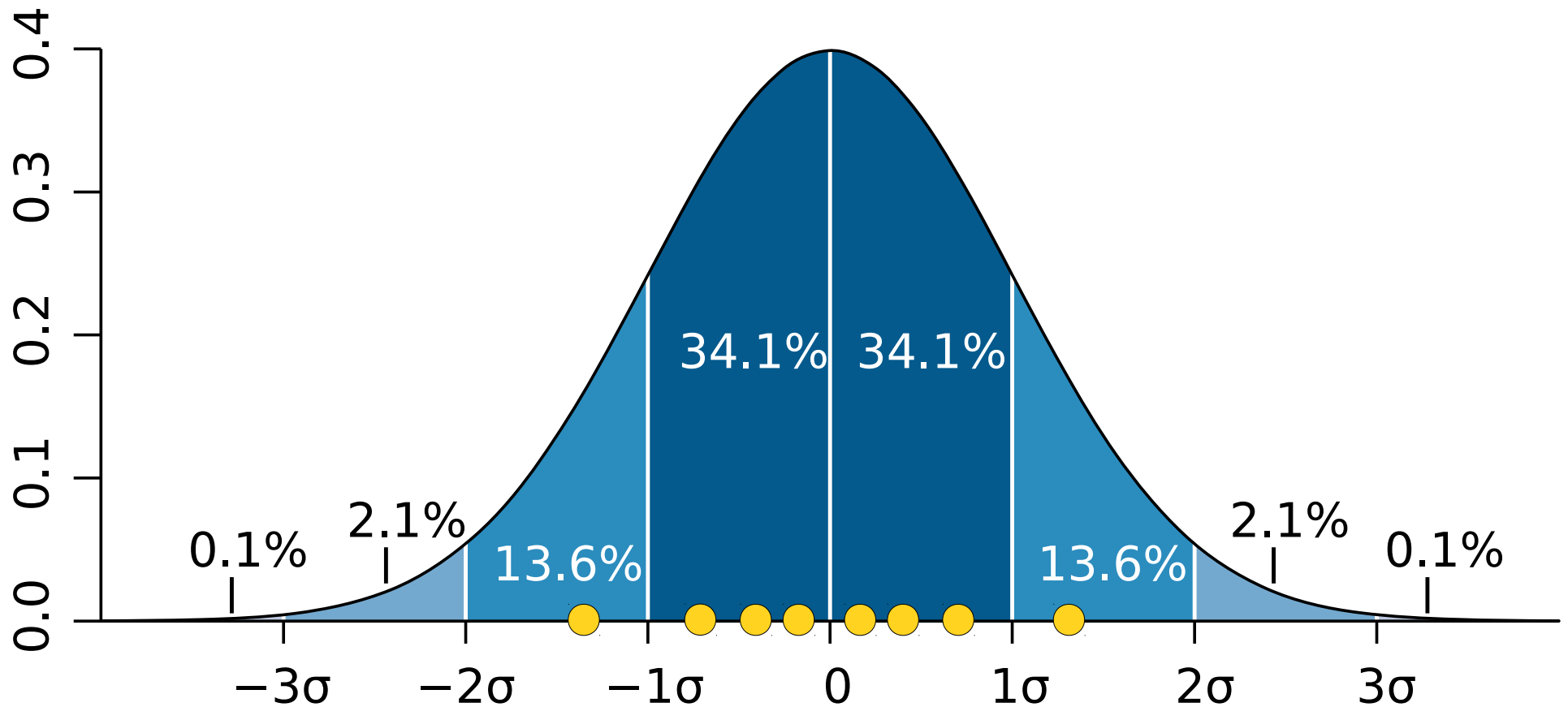
# Normalverteilung (Gauß)



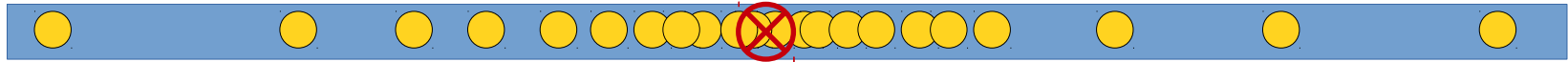
# Normalverteilung (Gauß)



# Normalverteilung (Gauß)



# Normalverteilung

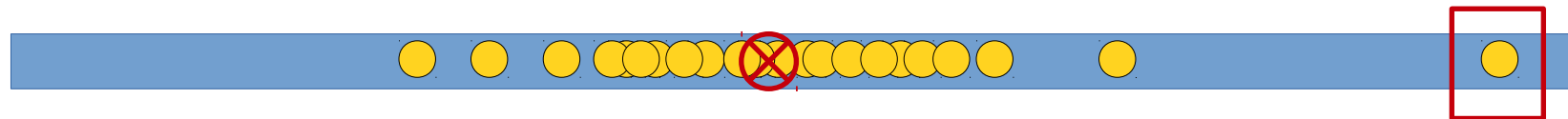


# Beispiele

- Preise eines Produktes am Markt
- Dart Wurf
- Körpergröße
- Gewicht
- IQ
- Testpunkte

# Ausreisser (Outlier)

Als Ausreißer wird ein Messwert bezeichnet, der nicht in eine erwartete Messreihe passt oder allgemein nicht den Erwartungen entspricht.



**METRIKEN**

# Min, Max und Range

- Min: kleinster Zahlenwert
- Max: größter Zahlenwert
- Range: Zahlen-Spannweite (Max - Min)

x	y
1	6
2	4
3	5
4	3
5	2
6	4



# Min, Max und Range

- Min: kleinster Zahlenwert
- Max: größter Zahlenwert
- Range: Zahlen-Spannweite (Max - Min)

Ergebnis:

Min: 2

Max: 6

Range: 4

x	y
1	6
2	4
3	5
4	3
5	2
6	4

# Mean

Summiere alle Werte und dividiere durch die Anzahl der Werte.

- Mean = Durchschnitt = Arithmetisches Mittel
- Aggregierte Größe: Ausreisser (Outlier) können verzerren.
- Bei Normalverteilung wahrscheinlich nahe des Erwartungswertes (Umso mehr Beobachtungen, umso wahrscheinlicher).

# Mean

Summiere alle Werte und dividiere durch die Anzahl der Werte.

x	y
1	6
2	4
3	5
4	3
5	2
6	4

# Mean

Summiere alle Werte und dividiere durch die Anzahl der Werte.

Mean:

$$= (6 + 4 + 5 + 3 + 2 + 4) / 6$$

$$= 24 / 6$$

$$= 4$$

x	y
1	6
2	4
3	5
4	3
5	2
6	4

# Median

1. Sortiere die Werte in Reihenfolge.
2. Wähle den Wert des Datenpunktes in der Mitte. Bei einer geraden Anzahl an Datenpunkten, nimm den Mittelwert der beiden mittleren Datenpunkte.

→ Ist robuster gegenüber Ausreißern.

# Median

Finde den Wert des in der Mitte liegenden Punktes.

x	y	y (sortiert)
1	6	2
2	4	3
3	5	4
4	3	5
5	2	6
6	4	6

# Median

Finde den Wert des in der Mitte liegenden Punktes.

Median:

$$= (4 + 5) / 2$$

$$= 4.5$$

x	y	y (sortiert)
1	6	2
2	4	3
3	5	4
4	3	5
5	2	6
6	4	6

# Modus

Modus ist der Wert, der am häufigsten im Datenset auftritt (nicht wie oft er auftritt).

- bei Gleichstand, sind alle gleich oft auftretenden Werte der Modus.
- geht auch für Nominal.
- bei Normalverteilung wahrscheinlich in der Nähe des Erwartungswertes.



# Modus

Modus ist der Wert, der am häufigsten im Datenset auftritt.

x	y
1	6
2	4
3	5
4	3
5	2
6	4

# Modus

Modus ist der Wert, der am häufigsten im Datenset auftritt.

Modus: 4

x	y
1	6
2	4
3	5
4	3
5	2
6	4

# Zusammenhang Mean & Median

Wenn Mean höher ist als Median, deutet das auf einen oder mehrere relativ große Werte im Datensatz hin (Outlier?). Die Verteilung ist somit nicht normalverteilt, sondern nach rechts verzerrt (skewed right).

# Übung

Diskutiert in den Gruppen in euren eigenen Worten was folgende Begriffe bedeuten:

- Mean
- Median
- Mode

(5min)

# Übung

- Min?
- Max?
- Range?
- Mean?
- Median?
- Modus?

(10min)

Monat	Gitarren
Januar	5
Februar	8
März	10
April	8
Mai	7
Juni	9
Juli	3
August	11
September	10
Oktober	10
November	12
Dezember	15

# In Libre Office

=MIN()

=MAX()

=AVERAGE()

=MEDIAN()

=MODE()

→ Beispiele Einkommens-Verteilung in Österreich.

# BOXPLOT

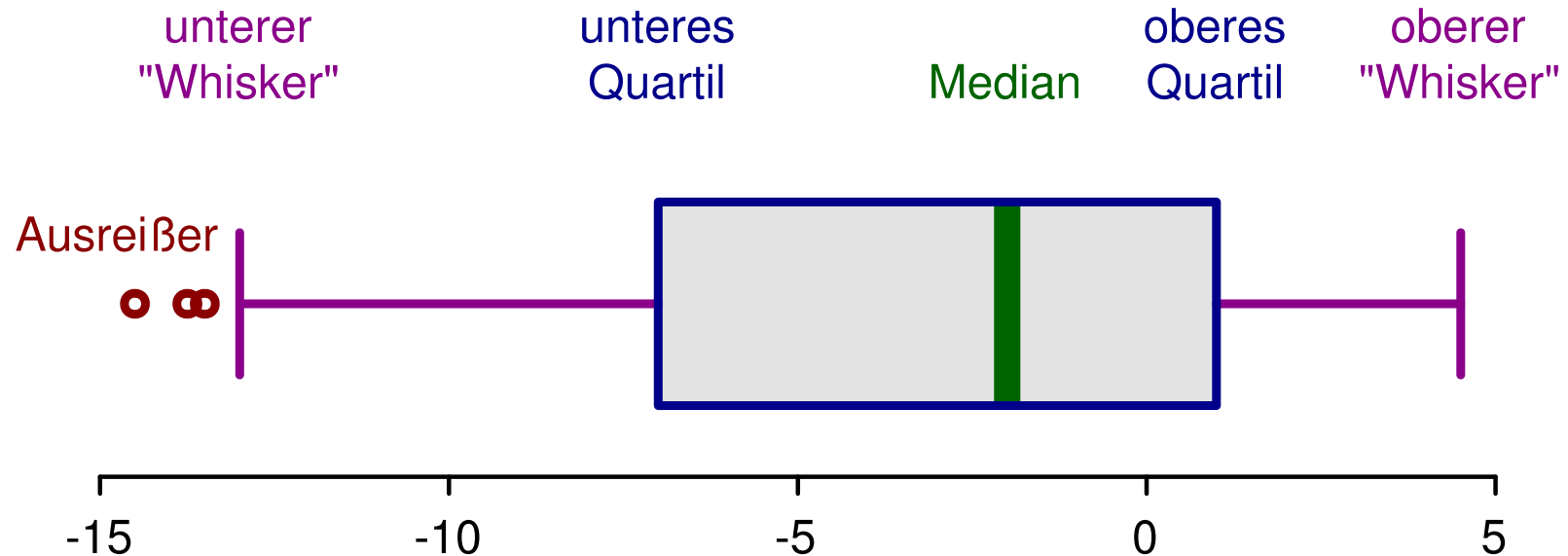
# Boxplot

Ein Box-Plot soll schnell einen Eindruck darüber vermitteln, in welchem Bereich die Daten liegen und wie sie sich über diesen Bereich verteilen.

– Wikipedia



# Boxplot



- Q1: Wert des Punktes nach 25% der Datenpunkte
- Median=Q2: Wert des Punktes nach 50% der Datenpunkte
- Q3: Wert des Punktes nach 75% der Datenpunkte
- Box: mittleren 50% der Daten (Interquartilsabstand → IQR)
- Antennen: meist letzter Datenpunkt im Abstand von 1.5x des IQR weg von Box

# Tipp: Vorgangsweise

- 1) Range ansehen
- 2) Box ansehen
- 3) Lage Median zu Box ansehen
- 4) Lage Box zu Range ansehen
- 5) Lage Antennen zu Range und Box ansehen

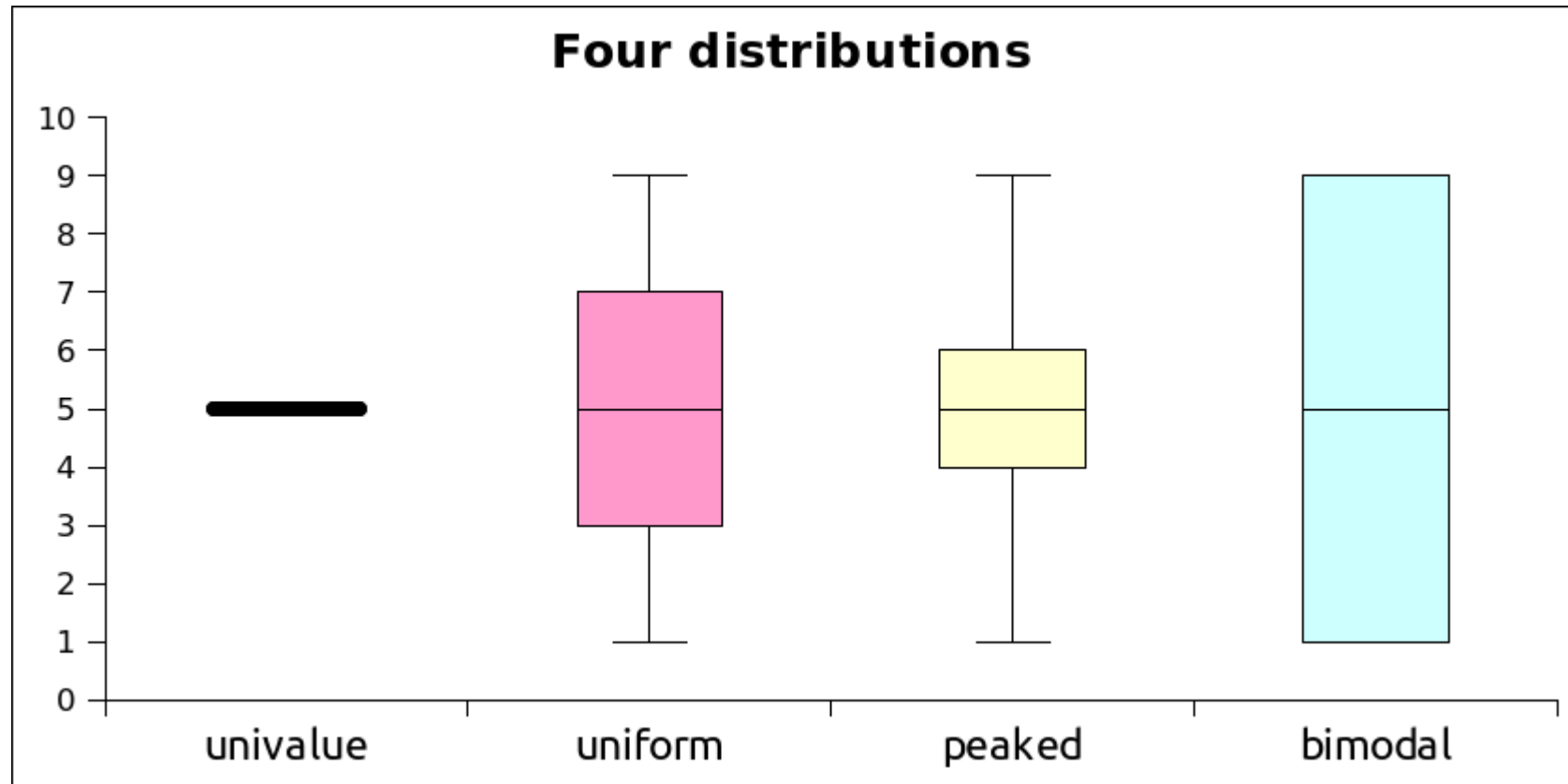
# Interpretation

## **Median ist verschoben von Box-Mittelpunkt**

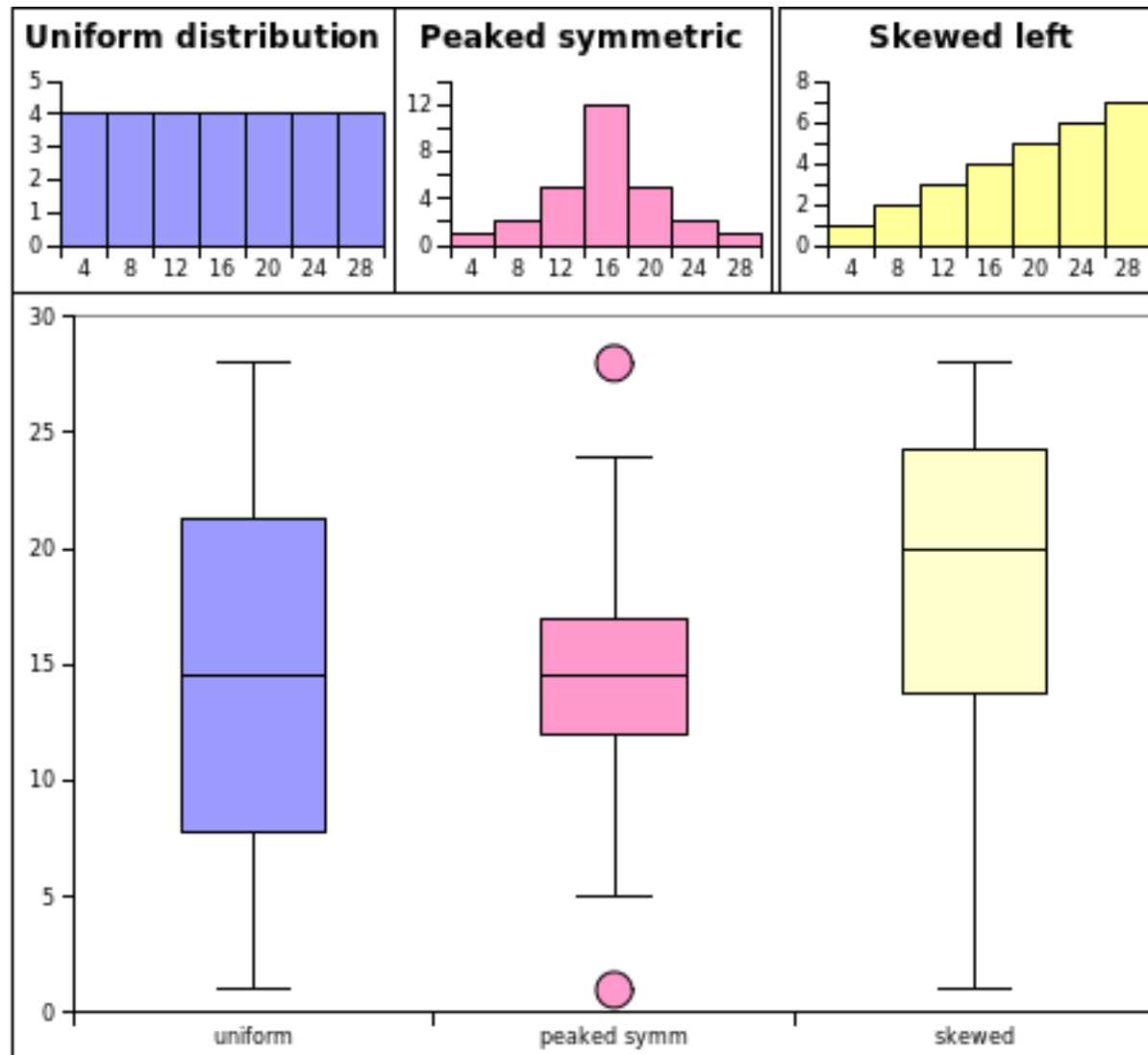
Bsp.: Wenn Median links vom Box-Mittelpunkt, sind die Daten rechts verschoben (vermutlich Ausreisser rechts). Umso weiter links, umso mehr deutet es auf starke Ausreisser hin. Gilt auch für entgegengesetzte Richtung.

**Lage Box zu Range:** Wenn Abstand gleich, sind Daten gleichverteilt. Wenn Box = schmal, sind Daten stark zentriert. Wenn Box = Range, sind Daten bimodal.

# Verteilungen

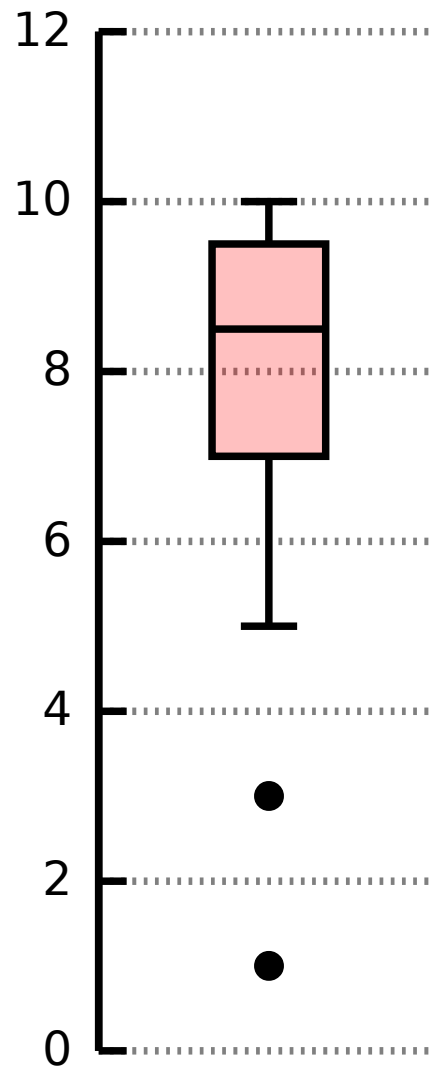


# Verteilungen



# Beispiel

Min = 1  
Max = 10  
Range = 9  
Mean = 7.75  
Median = 8.5  
Modus = 9 und 10  
Q1 = 7  
Q3 = 9.5  
IQR = 2.5

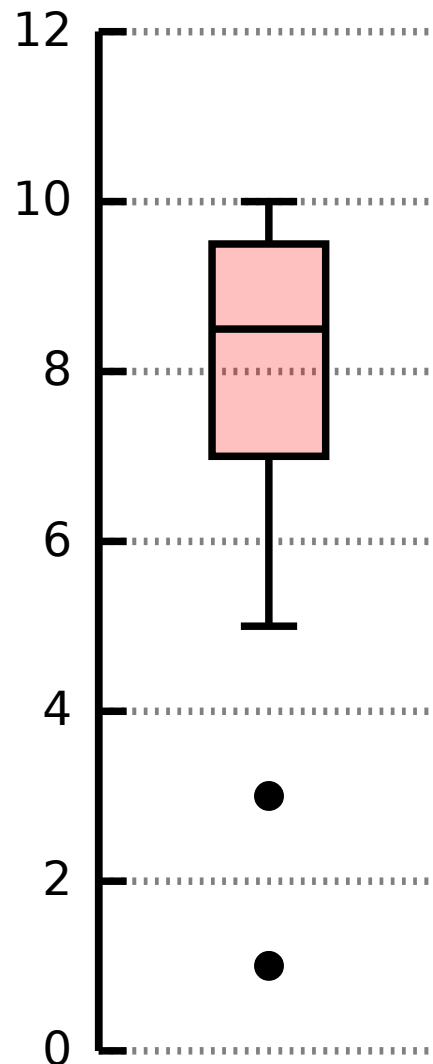


y	sort
9	1
6	3
7	5
7	6
3	7
9	7
10	7
1	8
8	8
7	8
9	9
9	9
8	9
10	9
5	9
10	10
10	10
9	10
10	10
8	10

# Beispiel

## Interpretation:

Verteilung wegen  
Ausreißern nach  
Links verschoben  
und somit nicht  
symmetrisch bzw.  
normalverteilt.



y	sort
9	1
6	3
7	5
7	6
3	7
9	7
10	7
1	8
8	8
7	8
9	9
9	9
8	9
10	9
5	9
10	10
10	10
9	10
10	10
8	10

# Beispiele

- Uniform
- Normal
- Skewed links und rechts
- Ausreisser rechts
- Bimodal



# Übung

2er Gruppen, 15min Zeit

- Sortiern
- Berechne: Min, Max, Mean, Median, Modus, Q1, Q3, IQR
- Erstelle Boxplot: [bit.ly/boxplotr](https://bit.ly/boxplotr)
- Interpretiere Daten

7

12

5

8

10

8

9

10

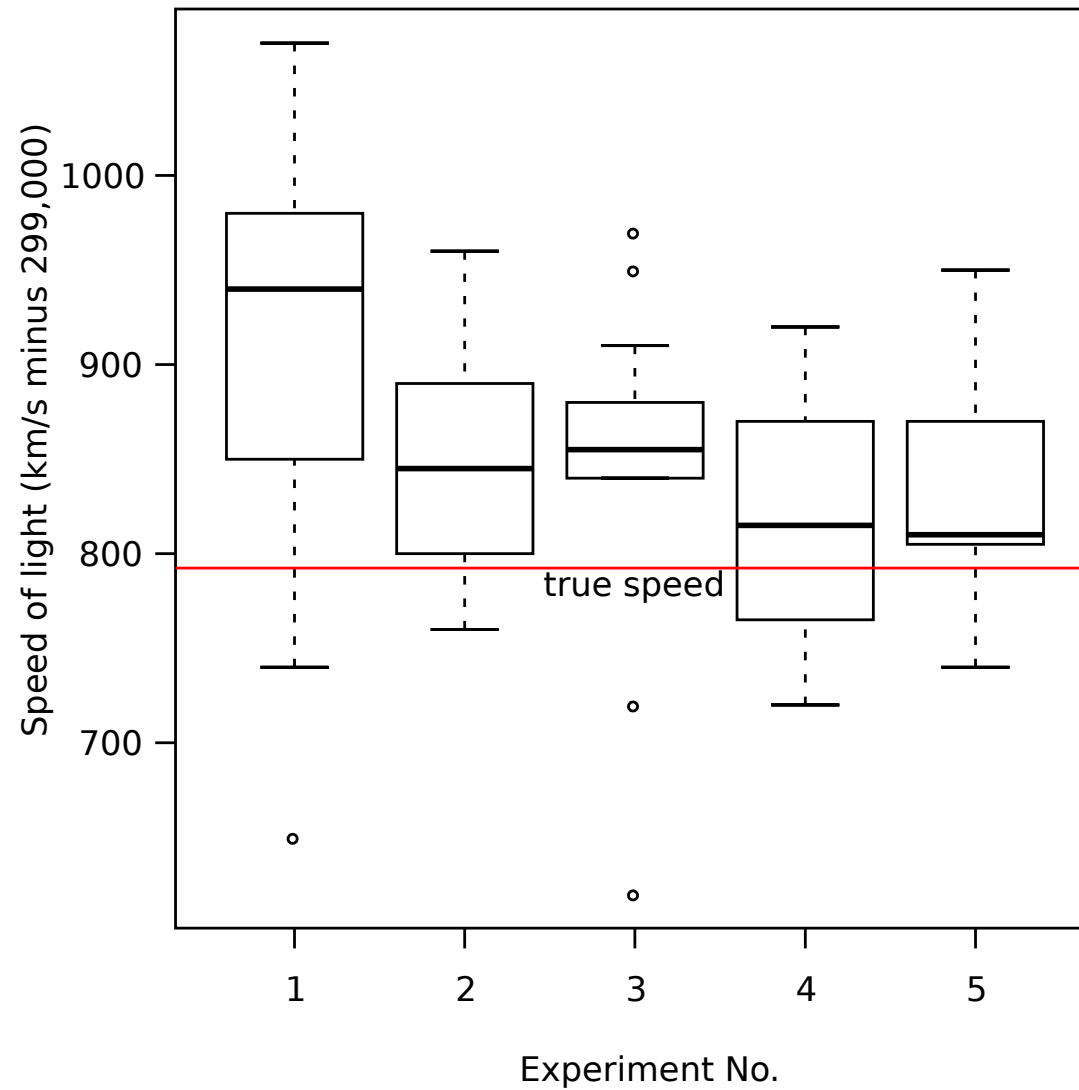
11

9

8

1

# Übung



## Kontakt

[www.offenewahlen.at](http://www.offenewahlen.at)

[@stefankasberger](https://www.instagram.com/stefankasberger)

[stefan.kasberger@okfn.at](mailto:stefan.kasberger@okfn.at)

[www.okfn.at](http://www.okfn.at)

## UrheberInnenrecht:

Dieses Werk ist, sofern nicht explizit anders angegeben, lizenziert unter einer Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen 4.0 International Lizenz.

Urheber: Stefan Kasberger (2018).

## Markenrecht:

Alle in dieser Präsentation genannten Marken und Produktnamen sind eingetragene Marken-/Warenzeichen der jeweiligen Hersteller beziehungsweise Unternehmen.