



# Statistik II: Grundlagen und Definitionen der Statistik

Dr. Andreas Vlašić

Medien Institut

☎ (0621) 52 67 44

💻 [vlasic@medien-institut.de](mailto:vlasic@medien-institut.de)



## Gliederung

- ▶ **1. Hintergrund: Entstehung der Statistik**
- 2. Grundlagen der deskriptiven Statistik
  - 2.1 Skalenniveaus
  - 2.2 Die Beschreibung von Verteilungen: Lage- und Streumaße
- 3. Grundlagen der induktiven Statistik
- 4. Key Facts

## Eine kurze Geschichte der Statistik

- Einführung des Begriffs durch Gottfried Achenwall (1719-1772) als Beschreibung der „Staatszustände“ → amtliche Statistik
- Definition „Statistik“: Lehre von der Erhebung, Messung und Auswertung (quantitativer) Daten
- Weiterentwicklung der Statistik als Folge des Aufkommens von Glücksspielen und des Versicherungswesens
  - Übergang von der deskriptiven zur induktiven Statistik (Wahrscheinlichkeitstheorie)
  - Praktische Probleme (aus Agrarwissenschaft, Biologie etc.) als Motor theoretischer Weiterentwicklung
- Heute: Einsatz der Statistik in fast allen Bereichen der Wissenschaft und Praxis

## Gliederung

1. Hintergrund: Entstehung der Statistik

▶ 2. **Grundlagen der deskriptiven Statistik**

2.1 **Skalenniveaus**

2.2 Die Beschreibung von Verteilungen: Lage- und Streumaße

3. Grundlagen der induktiven Statistik

4. Key Facts

## Variablen und Skalenniveaus

Skalenart	empirische Relevanz	mögliche Aussagen	Beispiele
1. Nominalskala	keine	Gleichheit, Verschiedenheit	Geschlecht, Familienstand, Telefonnummern
2. Ordinalskala	Ordnung der Zahlen	größer-kleiner Relationen	Schulbildung, militärische Ränge, Windstärken
3. Intervallskala	Differenzen der Zahlen	Gleichheit von Differenzen	Intelligenzquotient, Temperatur, Kalenderzeit
4. Verhältnisskala	Verhältnisse der Zahlen	Gleichheit von Verhältnissen	Alter, Längenmessung, Gewichtsmessung

- In der Praxis oftmals keine Unterscheidung von Intervall- und Verhältnisskala  
→ *Metrische Skala*
- In vielen Fällen keine Unterscheidung von Nominal- und Ordinalskala  
→ *Diskrete Skala*
- Nominalskala mit zwei Ausprägungen: *Dichotome Variable*

## Gliederung

1. Hintergrund: Entstehung der Statistik

▶ 2. **Grundlagen der deskriptiven Statistik**

2.1 Skalenniveaus

2.2 **Die Beschreibung von Verteilungen: Lage- und Streumaße**

3. Grundlagen der induktiven Statistik

4. Key Facts

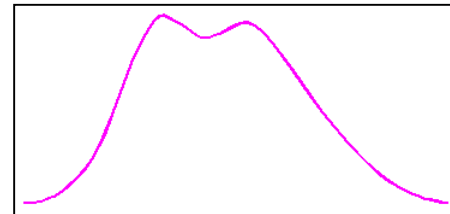
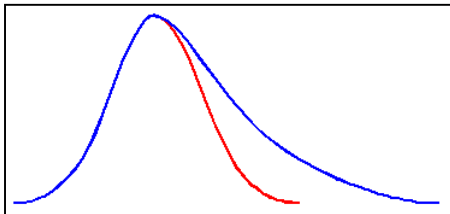
## Lagemaße

- **Modus:** häufigster Wert
  - ab Nominalskala
- **Median:** mittlerer Wert
  - ab Ordinalskala
  - robust gegenüber extremen Werten
- **Quantil/Perzentil**
  - Quartile: 25%- und 75%-Perzentil
- **Arithmetisches Mittel:** „Durchschnitt“
  - ab Intervallskala
  - anfällig für extreme Werte
- **Achtung:** Maße der zentralen Tendenz bei bimodalen Verteilungen problematisch

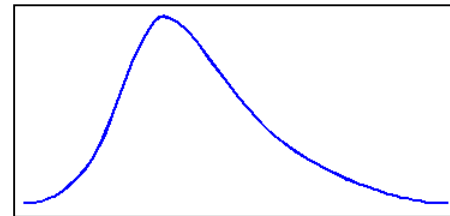
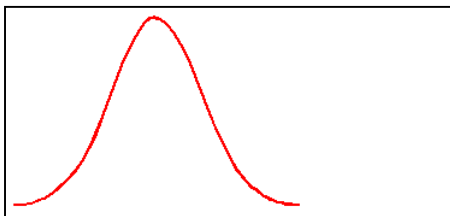
1	2	2	4	6	9	11
---	---	---	---	---	---	----

## Verteilungstypen

### ■ Eingipflig vs. mehrgipflig

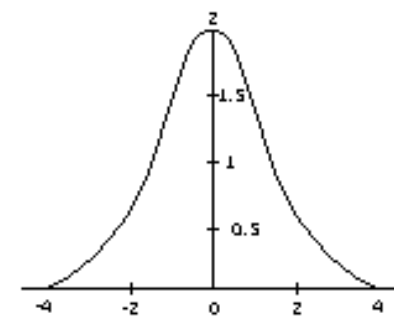


### ■ Symmetrisch vs. schief



### ■ Normalverteilung:

- Werte gruppieren sich um einen Mittelwert und fallen nach beiden Seiten hin gleichmäßig ab
- Gauß'sches Fehlergesetz





## Streuungsmaße

- Verteilungen mit gleichem Mittelwert können unterschiedlich streuen
  - Beschreibung durch Angabe von Streuungsmaßen
- *Wertebereich*
  - Differenz zwischen größtem und kleinstem Wert
- *(Inter)Quartilsabstand*
  - Abstand zwischen 1. und 3. Quartil
- *Varianz*
  - Prinzip: Streuung ist mittlere Abweichung vom Zentrum
  - mittlere quadrierte Abweichung

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

## Streuungsmaße /2

### ■ *Standardabweichung*

- Quadratwurzel der Varianz
- gleiche Maßeinheit wie zugrundeliegende Variable

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

### ■ *Z-Transformation*

- Darstellung der Daten als Abweichung vom Mittel mit Standardabweichung als Maßeinheit
- Vergleich von Variablen mit unterschiedlichen Maßeinheiten
- Standard-Normalverteilung: Mittelwert = 0, Standardabweichung = 1

## Gliederung

1. Hintergrund: Entstehung der Statistik
2. Grundlagen der deskriptiven Statistik
  - 2.1 Skalenniveaus
  - 2.2 Die Beschreibung von Verteilungen: Lage- und Streumaße
- ▶ 3. **Grundlagen der induktiven Statistik**
4. Key Facts

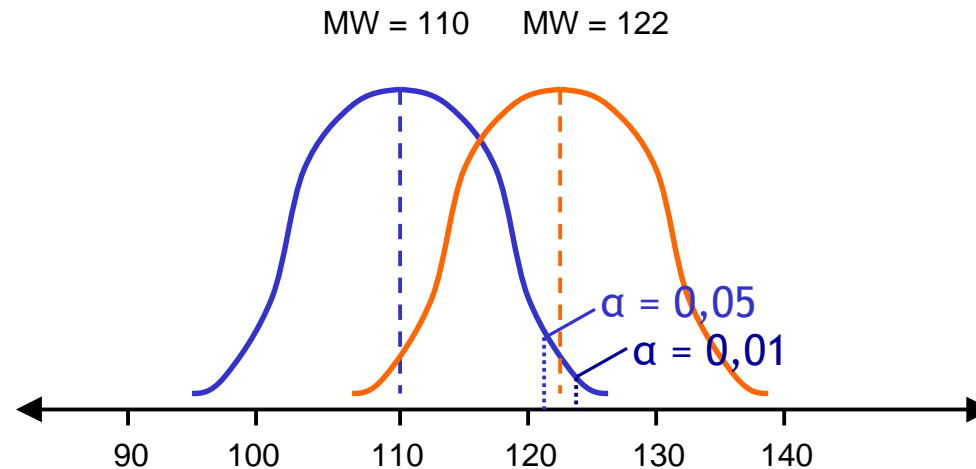
## Induktive Statistik und Signifikanz

- Problem: keine Vollerhebungen, aber Verallgemeinerung auf Grundgesamtheit → Testverfahren der *induktiven Statistik*
- Grundmodell des Signifikanztestens:
  - *Nullhypothese*  $H_0$ : Unterschiede/Zusammenhänge sind zufällig
  - *Alternativhypothese*  $H_1$ : Unterschiede/Zusammenhänge sind nicht zufällig
- *Irrtumswahrscheinlichkeit*  $p$  („Signifikanz“) wird mit Werten zwischen 0 und 1 angegeben, kann niemals 0 sein

Irrtumswahrscheinlichkeit	Bedeutung	Symbolisierung
$p > 0.05$	nicht signifikant	n.s.
$p \leq 0.05$	signifikant	*
$p \leq 0.01$	hochsignifikant	**
$p \leq 0.001$	höchst signifikant	***

## Logik des Signifikanztestens

- Grundfrage: Wie wahrscheinlich ist es, dass eine realisierte Stichprobe aus einer bestimmten Population stammt?



- Achtung: Signifikanztests sind (lediglich) ein Hilfsmittel bei der Auswertung und Interpretation von Daten
- ➔ Wichtigkeit Theorie- bzw. Hypothesen-geleiteter Datenanalyse

## Gliederung

1. Hintergrund: Entstehung der Statistik
2. Grundlagen der deskriptiven Statistik
  - 2.1 Skalenniveaus
  - 2.2 Die Beschreibung von Verteilungen: Lage- und Streumaße
3. Grundlagen der induktiven Statistik
- ▶ 4. **Key Facts**

## Key Facts

- Welche Skalenniveaus gibt es, welche Aussagen sind dabei möglich?
- Wie können statistische Verteilungen beschrieben werden?
- Welche Formen von Verteilungen gibt es?
- Welche Lage- und Streumaße gibt es, welche Aussagen ermöglichen sie?
- Welcher Logik folgen Signifikanztests?
- Was ist die Irrtumswahrscheinlichkeit?