

BA KW | Vorlesung

# Einführung in die Statistik

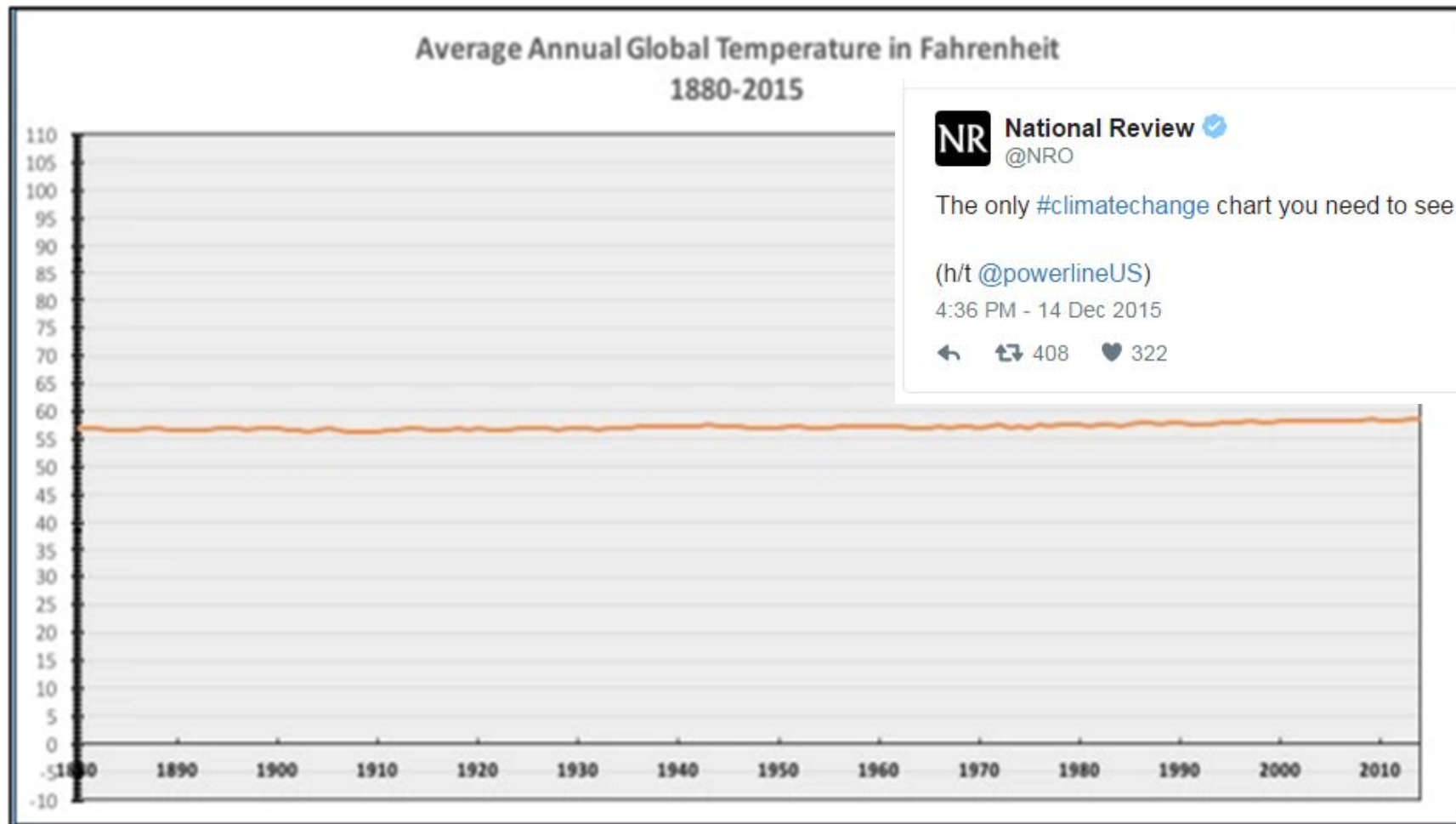
Einführung

*Prof. Thomas Hanitzsch*

# Wer braucht schon Statistik...

- Kein Voodoo mit Zahlen sondern ein Weg um Information (Daten) zu sammeln, darstellen, verarbeiten und zu analysieren
- Damit ein unerlässliches Werkzeug für die (Sozial)-Wissenschaft
- Fokus der Vorlesung:
  - Inhalt des statistischen Werkzeugkoffers
  - Umgang mit wichtigsten Werkzeugen

# Wer braucht schon Statistik...



**NR** National Review ✓  
@NRO Follow

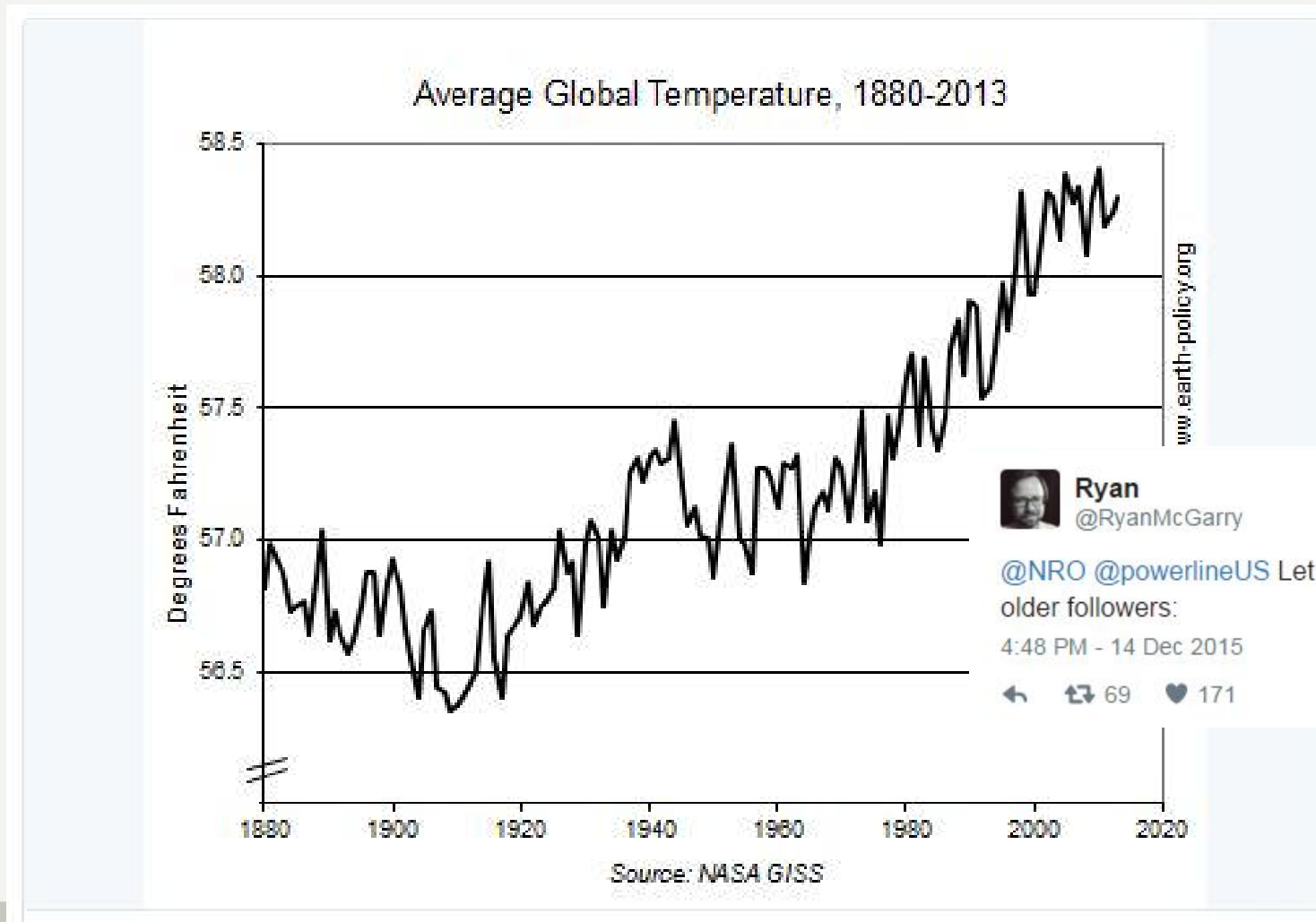
The only #climatechange chart you need to see. [natl.re/wPKpro](http://natl.re/wPKpro)

(h/t @powerlineUS)

4:36 PM - 14 Dec 2015

↩️ ↻️ 408 ❤️ 322

# Wer braucht schon Statistik...



# Programm der Vorlesung

## PROPÄDEUTIKUM

- 22.04. Einführung

## DESKRIPTIVE STATISTIK

- 29.04 Häufigkeiten
- 06.05. Streuung

## INFERENZSTATISTIK

- 13.05. Zufallsgrößen und Konfidenzintervall
- 27.05. Signifikanztest



# Programm der Vorlesung

## VERFAHREN DER ZUSAMMENHANGSMESSUNG

- 03.06.  $t$ -Test
- 10.06. Varianzanalyse
- 17.06. Korrelation
- 24.06. Regression
- 01.07. Kontingenztabellen
- 08.07. Ausblick: weitere Verfahren und Klausurvorbereitung

## PRÜFUNG

- 19.07. ***Klausur Fr. 10-12 (Anmeldung nicht vergessen!!)***



# Vorlesungsmaterial: Moodle



The screenshot shows a Moodle course page for 'Einführung in die Statistik / Datenanalyse'. The page is in German and displays course information for the semester 'SoSe 2024'. The course is organized into two main sections: '15314 Einführung in die Statistik (Vorlesung)' and '15313 Computergestützte Datenanalyse (Seminar)'. The overall responsibility is listed as 'Prof. Dr. Thomas Hanitzsch (hanitzsch@ifkw.lmu.de)'. The page includes a sidebar with navigation options like 'Teilnehmer/innen', 'Badges', and 'Bewertungen'. A search bar and a 'Foren durchsuchen' section are visible on the right. The course description is partially visible at the bottom.

<https://moodle.lmu.de/course/view.php?id=1065>

Passwort: „statistik“

# Programm Datenanalyse

DA 1	DA2	DA3	DA4	DA5	DA6	DA7	DA8	DA9	DA10	DA11
Mo. 16-18	Di. 08-10	Di. 12-14	Di. 12-14	Di. 14-16	Mi. 08-10	Mi. 10-12	Mi. 14-16	Mi. 16-18	Fr. 10-12	Fr. 12-14
Haim	Leiner	Zieringer	Schindler	Schindler	Sacher	Zieringer	Hase	Hase	Schwabl	Schwabl
22.4.	23.4.	16.4.	16.4.	16.4.	17.4.	17.4.	17.4.	17.4.	19.4.	19.4.
29.4.	30.4.	23.4.	23.4.	23.4.	24.4.	24.4.	24.4.	24.4.	26.4.	26.4.
6.5.	7.5.	30.4.	30.4.	30.4.	8.5.	8.5.	8.5.	8.5.	10.5.	10.5.
13.5.	14.5.	7.5.	7.5.	7.5.	15.5.	15.5.	15.5.	15.5.	17.5.	17.5.
27.5.	28.5.	14.5.	14.5.	14.5.	22.5.	22.5.	22.5.	22.5.	24.5.	24.5.
3.6.	4.6.	28.5.	28.5.	28.5.	29.5.	29.5.	29.5.	29.5.	31.5.	31.5.
10.6.	11.6.	4.6.	4.6.	4.6.	5.6.	5.6.	5.6.	5.6.	7.6.	7.6.
17.6.	18.6.	11.6.	11.6.	11.6.	12.6.	12.6.	12.6.	12.6.	14.6.	14.6.
24.6.	25.6.	18.6.	18.6.	18.6.	19.6.	19.6.	19.6.	19.6.	21.6.	21.6.
1.7.	2.7.	25.6.	25.6.	25.6.	26.6.	26.6.	26.6.	26.6.	28.6.	28.6.
8.7.	9.7.	2.7.	2.7.	2.7.	3.7.	3.7.	3.7.	3.7.	5.7.	5.7.
15.7.	16.7.	9.7.	9.7.	9.7.	10.7.	10.7.	10.7.	10.7.	12.7.	12.7.



# Die Klausur

- **Form:**
  - Aufgaben



- **Unterlagen und Hilfen:**
  - „Open book“ → Vorlesungsfolien, Übungsmaterialien, Lehrbücher und Notizen sind erlaubt

# Die Klausur

- **Bewertung:**
  - Bestanden ab 50% der Gesamtpunktzahl
  - Volle Punktzahl wird nur dann vergeben, wenn die einzelnen Rechenschritte dokumentiert sind
  - Dabei kann es je nach Aufgabenstellung mehrere Wege geben, um zum Ziel zu kommen
  - Punkte gibt es auch für richtige Teilergebnisse
  - Folgefehler werden als solche gewertet
  - Rundungsfehler sind einkalkuliert: Zwischenergebnisse können bis auf die 2. Nachkommastelle gerundet werden



## Ergänzende Literatur

- Küchenhoff, Helmut et al.(2006): Statistik für Kommunikationswissenschaftler (2. Aufl.). Konstanz: UVK.

Online verfügbar unter

<https://elibrary-utb-de.emedien.ub.uni-muenchen.de/doi/book/10.36198/9783838528328>



Helmut Küchenhoff  
et al.  
Statistik für  
Kommunikations-  
wissenschaftler

2. Auflage

UVK UTB



## Ergänzende Literatur

- Uhlemann, Ingrid A. (2015):  
Einführung in die Statistik für  
Kommunikationswissenschaftler.  
Springer VS.

Online verfügbar unter

<https://link-springer-com.emedien.ub.uni-muenchen.de/book/10.1007%2F978-3-658-05769-5>

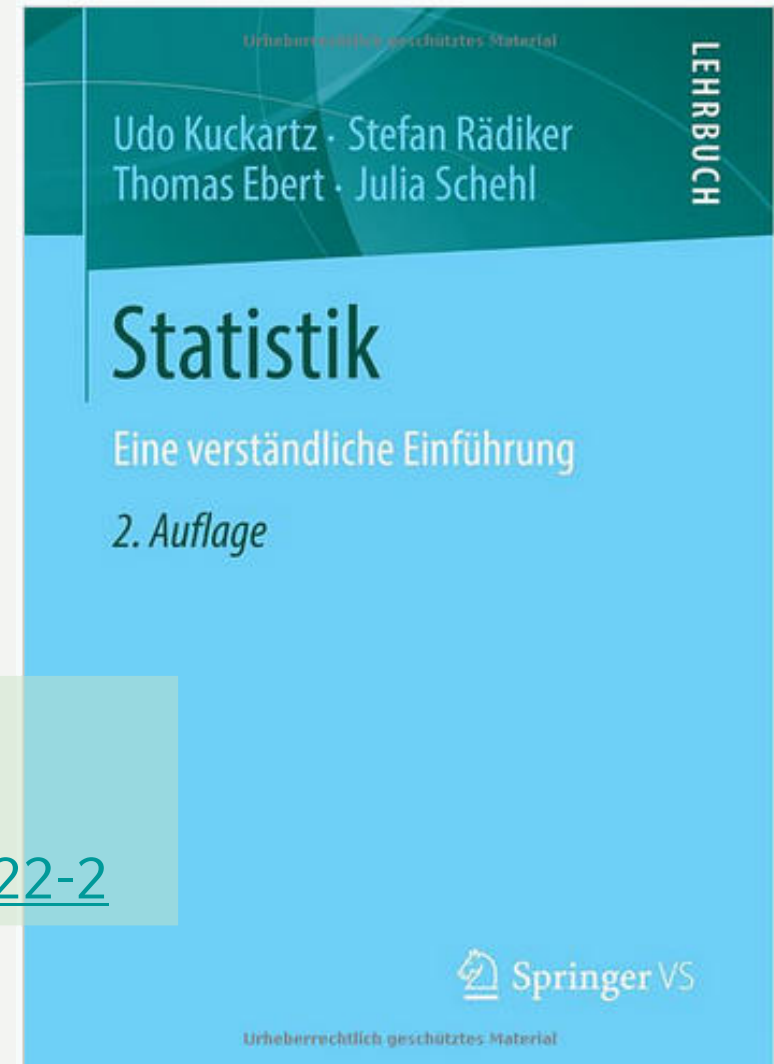


## Ergänzende Literatur

- Hartmann, F. G., Kopp, J. & Lois, D. (2022): Sozialwissenschaftliche Datenanalyse. Eine Einführung (3. Aufl.). Wiesbaden: Springer VS.

Online verfügbar unter

<https://link-springer-com.emedien.ub.uni-muenchen.de/book/10.1007/978-3-658-36322-2>



## Ergänzende Literatur

- Kuckartz, Udo et al. (2013): Statistik. Eine verständliche Einführung (2. Aufl.). Wiesbaden: VS Verlag.

Online verfügbar unter

<http://link.springer.com/emedien.ub.uni-muenchen.de/book/10.1007/978-3-531-19890-3>



## Ergänzende Literatur

- Bortz, Jürgen & Christof Schuster (2010): Statistik für Human- und Sozialwissenschaftler (7. Aufl.). Springer.

Online verfügbar unter

<http://link.springer.com.emedien.ub.uni-muenchen.de/book/10.1007/978-3-642-12770-0>



www.lehrbuch-  
psychologie.de

Bortz  
Schuster

# Statistik

für Human- und Sozialwissenschaftler

7. Auflage

 Springer

# Ergänzende Literatur

- Rasch, Björn et al. (2021): Quantitative Methoden 1/2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler (4. Aufl.). Springer.

Online verfügbar unter

<https://link-springer-com.emedien.ub.uni-muenchen.de/book/10.1007/978-3-662-63282-6>

<https://link-springer-com.emedien.ub.uni-muenchen.de/book/10.1007/978-3-662-63284-0>



B. Rasch · M. Friese · W. Hofmann · E. Naumann

## Quantitative Methoden 1

Einführung in die Statistik für Psychologen  
und Sozialwissenschaftler

4. Auflage

 Springer



B. Rasch · M. Friese · W. Hofmann · E. Naumann

## Quantitative Methoden 2

Einführung in die Statistik für Psychologen  
und Sozialwissenschaftler

4. Auflage

 Springer





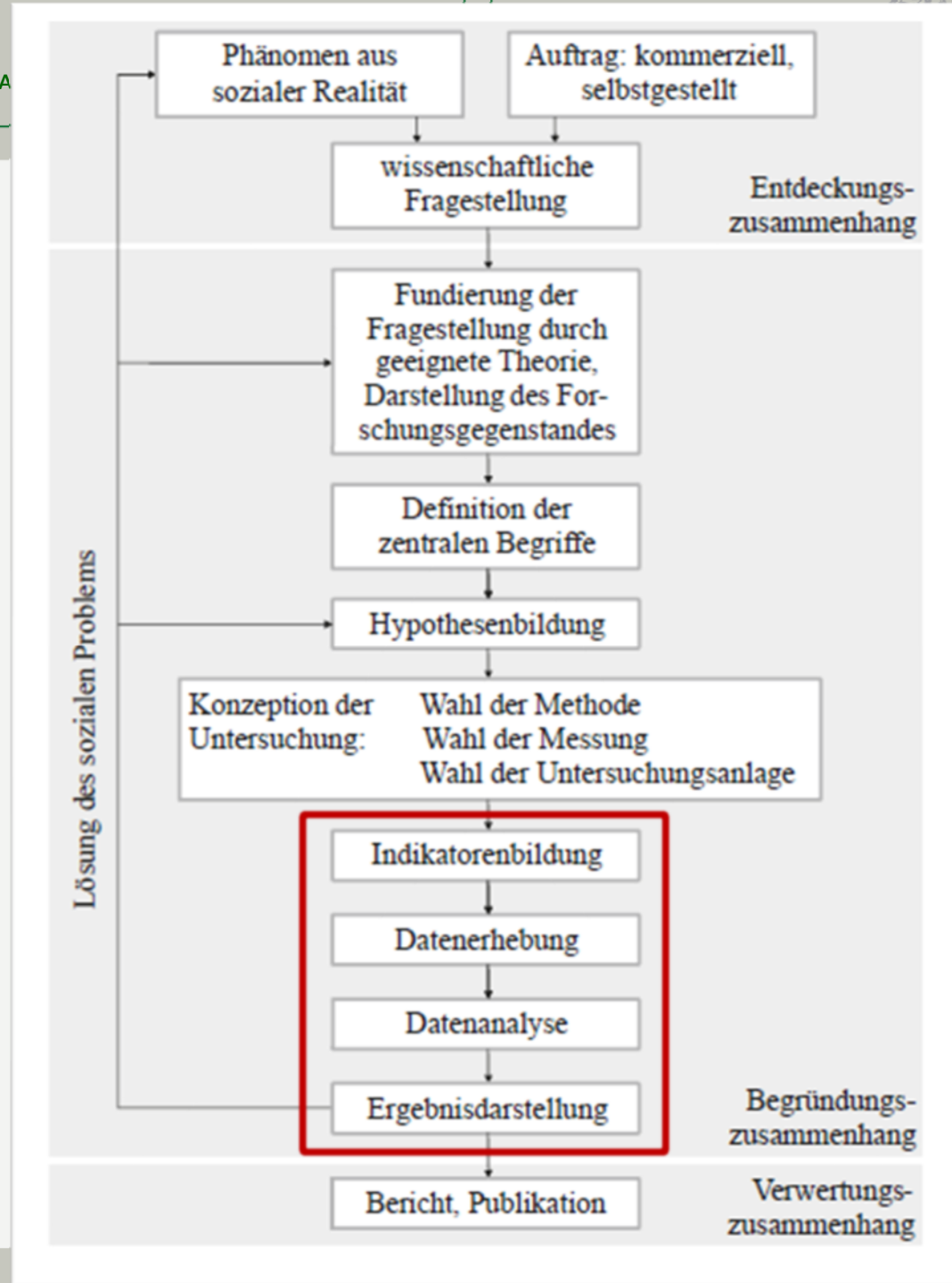
# Was ist Statistik?

„Statistik als Wissenschaft bezeichnet eine Methodenlehre, die sich mit der Erhebung, der Darstellung, der Analyse und der Bewertung von Daten auseinandersetzt.“

(Küchenhoff et al. 2006, S. 11)



# Statistik im Forschungsalltag



Quelle: Brosius et al.  
2012, S. 26

# Statistik im Forschungsalltag

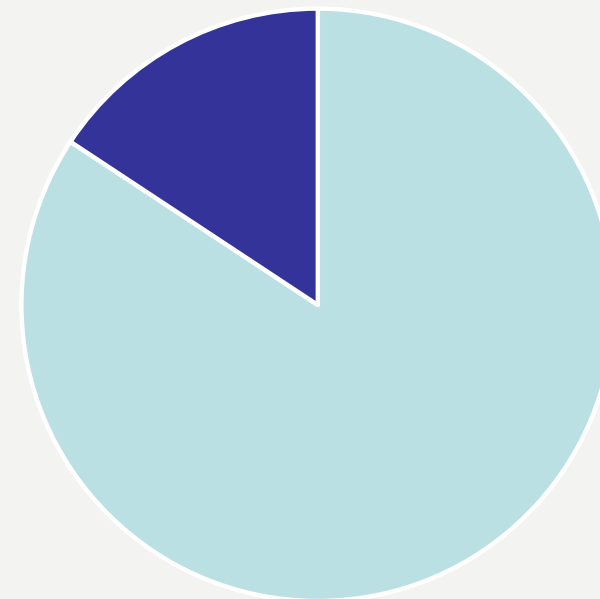
- **Indikatorenbildung**
  - Bildung neuer Skalen, Reliabilitätstests
- **Datenerhebung**
  - Stichprobenziehung
- **Datenanalyse**
  - Aufdecken und Deuten von Zusammenhängen, Schätzen von unbekanntem Parametern, Prognose
- **Ergebnisdarstellung**
  - Grafische Aufbereitung der Ergebnisse

# Was Statistik (nicht) kann

- Daten visualisieren bzw. zusammenfassen

Geschlecht	N	%
Weiblich	118	84,3
Männlich	22	15,7
<b>Gesamt</b>	<b>140</b>	<b>100</b>

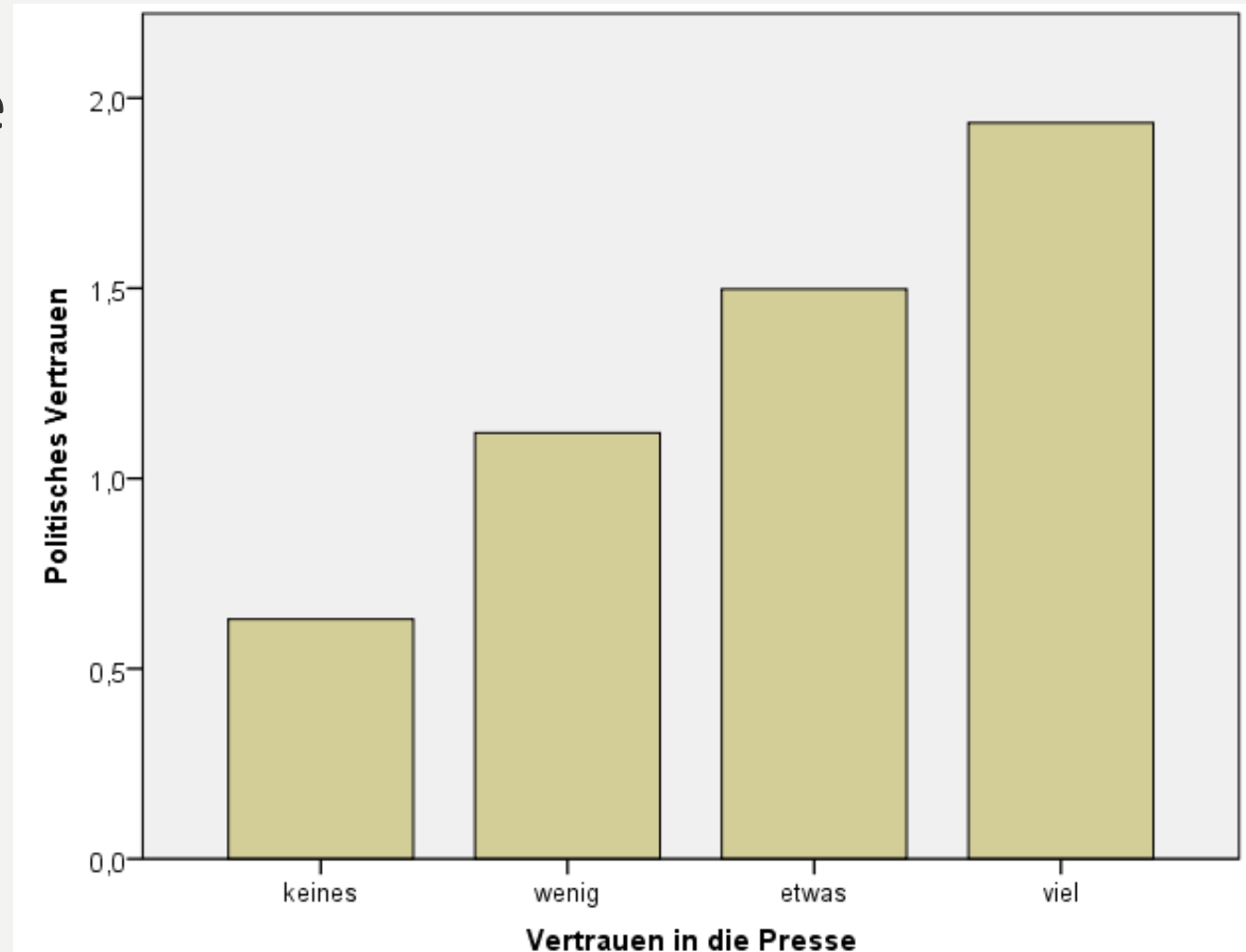
Geschlecht



■ Weiblich ■ Männlich

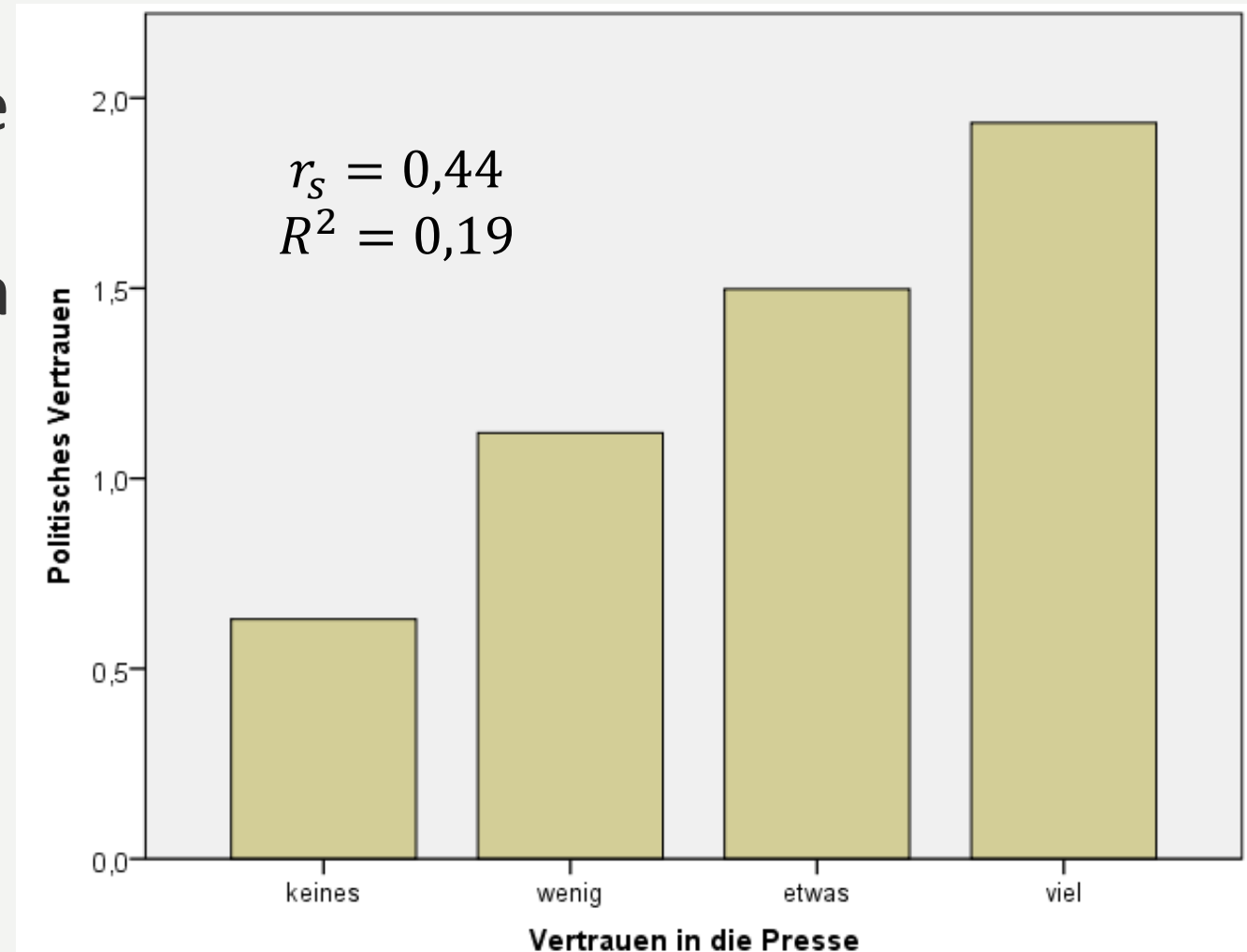
# Was Statistik (nicht) kann

- Zusammenhänge finden, **aber** Interpretation muss immer der Forscher liefern



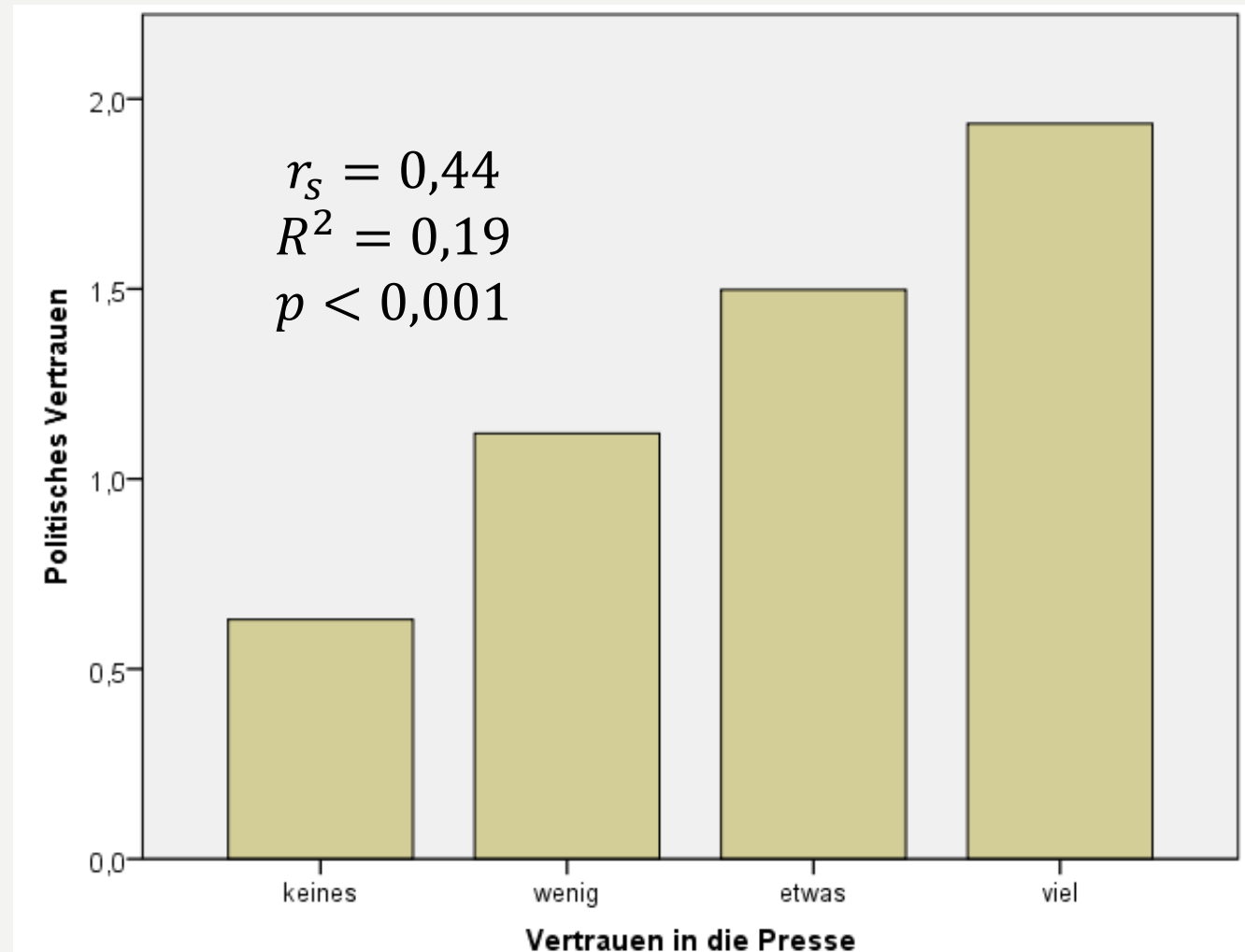
# Was Statistik (nicht) kann

- Zusammenhänge in Modellen zusammenfassen bzw. durch Modelle vereinfachen



# Was Statistik (nicht) kann

- Von einer Stichprobe auf die Population schließen



# Teilbereiche der Statistik

Bereich	Deskriptive Statistik	Inferenzstatistik
<b>Zweck</b>	<ul style="list-style-type: none"> <li>Beschreibung der empirischen Realität (anhand einer Stichprobe)</li> </ul>	<ul style="list-style-type: none"> <li>Rückschluss auf die Grundgesamtheit</li> </ul>
<b>Anwendung</b>	<ul style="list-style-type: none"> <li>Daten darstellen und visualisieren</li> <li>Zusammenhänge entdecken</li> </ul>	<ul style="list-style-type: none"> <li><b>Schätzen</b> der Verteilung und <b>Testen</b> der Zusammenhänge in der Grundgesamtheit</li> </ul>



# Die wichtigsten Begriffe

- **Grundgesamtheit:**
  - die Menge aller Objekte bzw. Subjekte, über die eine Aussage getroffen werden soll
    - z.B. „die Bild-Leser“, oder „die Wahlkampfberichterstattung“
- **Merkmalsträger:**
  - auch „Untersuchungseinheit“ bzw. „Fall“
  - jedes Element der Grundgesamtheit
    - z.B. einzelne Personen oder Zeitungsartikel

# Die wichtigsten Begriffe

- **Merkmale bzw. Variablen:**
  - Eigenschaften von Merkmalsträgern
    - z.B. Nutzungsdauer, Geschlecht
      - **Stetig:** jeder beliebige Wert innerhalb eines Intervalls möglich (z.B. Körpergröße, Mediennutzung)
      - **Diskret:** nur endlich viele Ausprägungen möglich (z.B. Geschlecht, Bildung)

# Die wichtigsten Begriffe

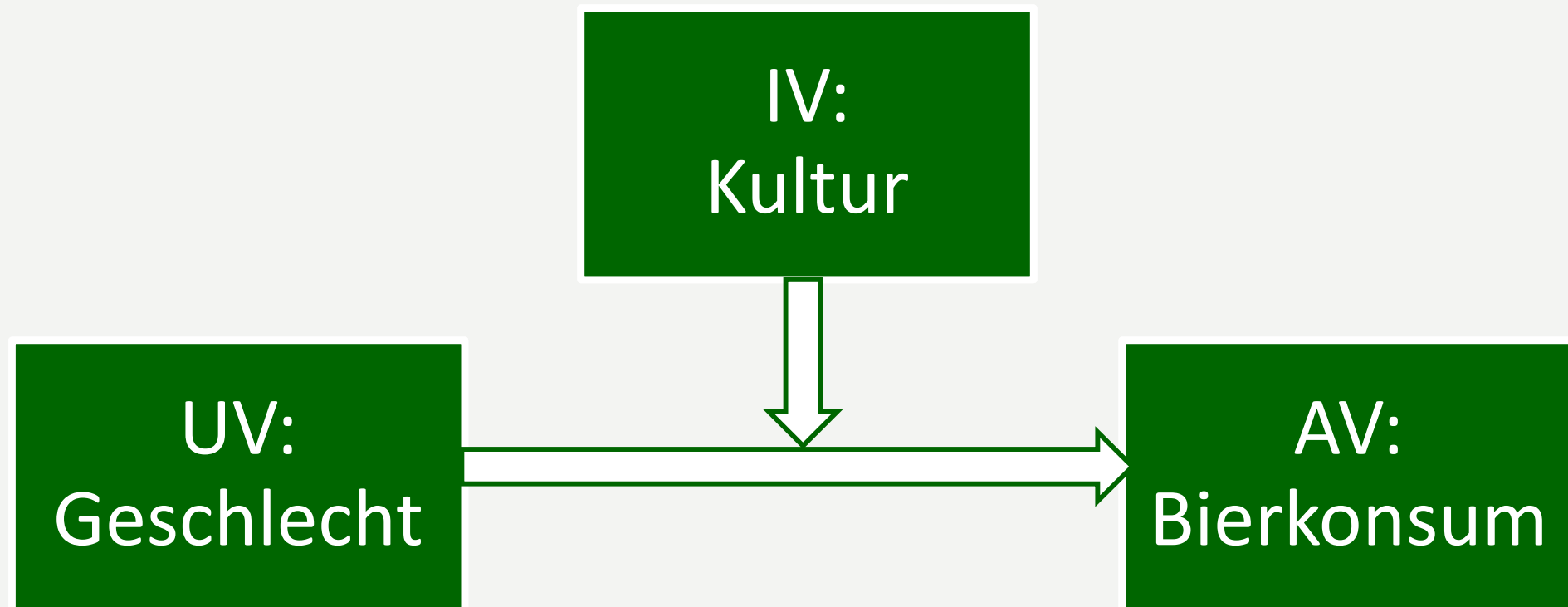
- **Ausprägungen bzw. Kategorien:**
  - Zustand des Merkmals
  - z.B. männlich/weiblich, 60 Min. TV-Nutzung
- **Wertebereich:**
  - die Menge aller möglichen Realisationen einer Variable



# Variablenlogik

- **Unabhängige Variable (UV):**
  - „erklärende“ bzw. „exogene“ Variable
  - im experimentellen Kontext wird die UV manipuliert, um ihre Wirkung auf die AV zu prüfen
- **Abhängige Variable (AV):**
  - „erklärte“ bzw. „endogene“ Variable
  - wird nur gemessen/beobachtet
- **Intervenierende Variable (IV):**
  - Zwischenvariable, „Moderator“ oder „Mediator“
  - wirkt auf die Beziehung zwischen AV und UV ein

# Variablenlogik



# Skalenniveaus

Skalierung	Funktionen	Mögliche Aussagen	Beispiele
<b>Nominal</b>	namensgebend/ benennend	Gleichheit/ Verschiedenheit	Geschlecht, Handymarke
<b>Ordinal</b>	rangordnend	Größer-Kleiner- Relationen	Schulnoten, Likert-Skalen
<b>Metrisch</b> <i>Intervall</i>	relativ (abstands- berechnend)	Gleichheit von Differenzen	Temperatur in Grad Celsius
<i>Ratio bzw. Verhältnis</i>	absolut (abstands- berechnend)	Gleichheit von Verhältnissen	Fernsehdauer in Minuten

# Skalenniveaus

- **Sonderform: „quasi-metrische“ Skala**
  - In Abweichung von der allgemeinen methodischen Konvention werden in der Praxis Bewertungsskalen (Likert-Skalen) mit mindestens fünf Abstufungen bei Auswertungen oft als „metrische“ Skalen betrachtet
  - z.B. Zustimmungsskalen:
    - 5 = stimme stark zu
    - 4 = stimme eher zu
    - 3 = teils/teils
    - 2 = lehne eher ab
    - 1 = lehne stark ab

# Verfahren der Zusammenhangsmessung und Skalenniveaus

Abhängige Variable

*Unabhängige Variable*

<b>SKALA</b>	<b>Nominal</b>	<b>Ordinal</b>	<b>Metrisch</b>
<b>Nominal</b>	<ul style="list-style-type: none"> <li>• Unabhängigkeitstest</li> <li>• Loglineare Analyse</li> </ul>	<ul style="list-style-type: none"> <li>• Unabhängigkeitstest</li> </ul>	<ul style="list-style-type: none"> <li>• Logistische Regression</li> </ul>
<b>Ordinal</b>	<ul style="list-style-type: none"> <li>• Unabhängigkeitstest</li> </ul>	<ul style="list-style-type: none"> <li>• Ordinale Korrelation</li> </ul>	<ul style="list-style-type: none"> <li>• Ordinale Korrelation</li> </ul>
<b>Metrisch</b>	<ul style="list-style-type: none"> <li>• <i>t</i>-Test (dichotom)</li> <li>• ANOVA (polytom)</li> </ul>	<ul style="list-style-type: none"> <li>• Ordinale Korrelation</li> </ul>	<ul style="list-style-type: none"> <li>• Korrelation</li> <li>• Regression</li> </ul>



# Statistische Notation: Grundlagen

$N$	Umfang der Gesamtstichprobe
$n$	Umfang einer Teilstichprobe
$X, Y$	Variable $X$ bzw. $Y$
$i$	Klasse
$x_i$	Realisation der Variable $X$ : $i$ -te Klasse
$f_i$	Absolute Häufigkeit
$p_i$ bzw. $p$	Relative Häufigkeit bzw. Anteilswert

# Statistische Notation: Schreibweisen

- Statistische Notationen können von Lehrbuch zu Lehrbuch zum Teil erheblich abweichen.

Verwechslungsgefahr besteht u.a. bei:

- Gesamt-Stichprobenumfang:  $N$  bzw.  $n$
- Absoluten Häufigkeiten:  $f_i, n_i$  bzw.  $h_i$
- Relativen Häufigkeiten:  $p_i$  bzw.  $f_i$
- Kumulierte relat. Häufigkeiten:  $F_i, F_{x_i}$  bzw.  $cf(I)$
- Empirische Varianz:  $s, \tilde{s}$  bzw.  $Var(X)$
- Fehler:  $e$  bzw.  $\epsilon$

Das griechische Alphabet ist für Parameter der Grundgesamtheit reserviert.