



TECHNISCHE UNIVERSITÄT  
IN DER KULTURHAUPTSTADT EUROPAS  
CHEMNITZ

Professur Psychologie digitaler Lernmedien

Institut für Medienforschung

Philosophische Fakultät



## Statistik I

# Zentrale Tendenz, Streuung und Verteilung

Matrix Reloaded (2003). Warner Bros. Pictures / Village Roadshow Pictures.

# Überblick

- Mittelwert
- Standardabweichung
- Weitere Maße der zentralen Tendenz
- Weitere Streuungsmaße
- z-Standardisierung
- Normalverteilung
- Schiefe und Kurtosis

# Einführung

- **Definition:** Statistik als Suche nach Mustern in Zahlen
- **Vorgehen:** Wie sucht man in riesigen Datensätzen aus Zahlen nach Mustern?
- **Zusammenfassung (Aggregation)** von Zahlen zu charakteristischen statistischen Kennwerten
- **Zwei Arten von statistischen Kennwerten**
  - **Maße der zentralen Tendenz:** Maße fassen die Messwerte zusammen
  - **Streuungsmaße:** Maße geben Auskunft über die Unterschiedlichkeit der Messwerte
- **Mittelwert** als Maß der zentralen Tendenz und **Standardabweichung** als Streuungsmaß am gebräuchlichsten

# Mittelwert

- **Mittelwert** (engl. mean,  $M$ ): „Durchschnitt“ aller Messwerte
- **Mathematisch**: Summe aller Werte dividiert durch deren Anzahl  $n$

- **Formel**:

$$M = \frac{\sum_{i=1}^n x_i}{n}$$

- **Eigenschaften**

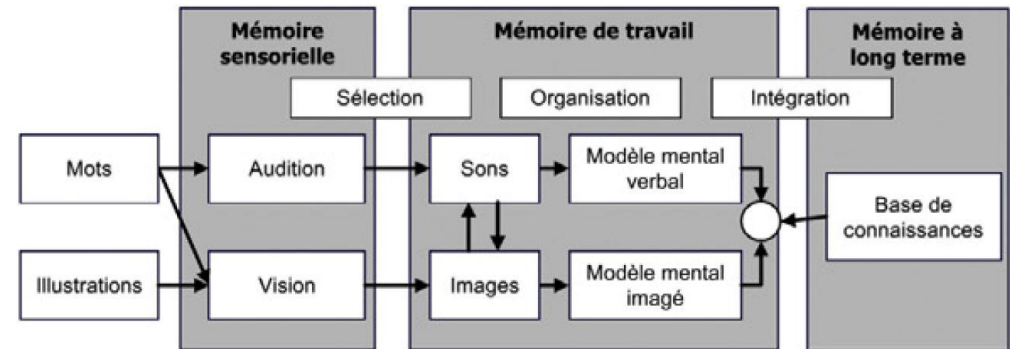
- Summe der gerichteten (mit Vorzeichen versehenen) Abweichungen vom Mittelwert ergibt Null
- Summe der quadrierten Abweichungen ergibt ein Minimum

- **Voraussetzung**: Mindestens Intervallskalenniveau der Daten

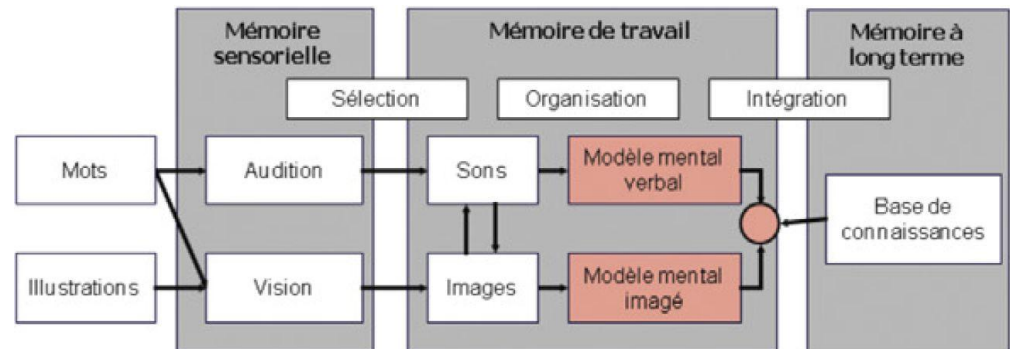
# Beispiel: Mittelwertberechnung beim Signalisierungseffekt (Jamet, 2014)

- $N = 32$ ; 69% ♀;  $\bar{X} = 22.4$  Jahre ( $SD = 2.1$ )
- Lernmaterial: CTML
- Einfaktorielles, zweifachgestuftes Design
  - Ohne Signalisierungen
  - Mit Signalisierungen
- U. a. Lernleistung als abhängige Variable

## Ohne Signalisierungen



## Mit Signalisierungen



Quelle: Jamet (2014)

# Beispiel: Mittelwerteberechnung

- Welche Versuchsbedingung hat zu höheren Lernleistungen geführt?

VPN	Bedingung	Lernleistung
1	Ohne Signalisierungen	8
2	Ohne Signalisierungen	4
3	Ohne Signalisierungen	6
4	Ohne Signalisierungen	2
5	Mit Signalisierungen	9
6	Mit Signalisierungen	7
7	Mit Signalisierungen	4
8	Mit Signalisierungen	8

Mittelwerte für die beiden Gruppen

$$M = (8 + 4 + 6 + 2) : 4 = 5$$

$$M = (9 + 7 + 4 + 8) : 4 = 7$$

# Standardabweichung

- **Streuungsmaße:** Neben dem Mittelwert (oder anderen Maßen der zentralen Tendenz) immer auch ein Streuungsmaß (z. B. Standardabweichung) angeben
- **Standardabweichung (engl. standard deviation, *SD*):** Grob gesagt die durchschnittliche Abweichung der Messwerte von ihrem Mittelwert
- **Formel:**
$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$
- **Eigenschaften**
  - Standardabweichung  $\geq 0$
  - Große Abweichungen werden stärker berücksichtigt als kleine
- **Voraussetzung:** Mindestens Intervallskalenniveau der Daten

# Beispiel: Berechnung der Standardabweichungen

- Wie stark weichen die Lernleistungen innerhalb der beiden Versuchsbedingungen voneinander ab?

VPN	Bedingung	Lernleistung	$\bar{x}$	$(x_i - \bar{x})^2$
1	Ohne Signalisierungen	8	5	9
2	Ohne Signalisierungen	4	5	1
3	Ohne Signalisierungen	6	5	1
4	Ohne Signalisierungen	2	5	9
5	Mit Signalisierungen	9	7	4
6	Mit Signalisierungen	7	7	0
7	Mit Signalisierungen	4	7	9
8	Mit Signalisierungen	8	7	1

Standardabweichungen für die beiden Gruppen

$$SD = \sqrt{\frac{9 + 1 + 1 + 9}{4 - 1}}$$

$$SD = 2.58$$

$$SD = \sqrt{\frac{4 + 0 + 9 + 1}{4 - 1}}$$

$$SD = 2.16$$

# Berechnung von Mittelwert und Standardabweichung

Wie lauten Mittelwert und Standardabweichung für den dargestellten Datensatz?

- A:  $M = 3$  und  $SD = 3$
- B:  $M = 3$  und  $SD = 4$
- C:  $M = 4$  und  $SD = 3$
- D:  $M = 4$  und  $SD = 4$
- E:  $M = 3$  und  $SD = 2$

VPN	1	2	3	4	5	6	7	8
Wert	7	-3	2	9	1	2	0	6

# Beispiele für Mittelwerte und Standardabweichungen in Fachzeitschriften

There was also a main effect of cueing condition on the retention of signalled information, with the cueing group ( $M = 2.65$ ,  $SD = 1.66$ ) performing better than the no-cueing group ( $M = 1.34$ ,  $SD = 1.23$ ),  $F(1, 30) = 6.44$ ,  $MSE = 13.78$ ,  $p = .03$ ,  $\eta_p^2 = 0.18$ .

Quelle: Jamet (2014)

Learners who received a personalized narrated animation scored significantly higher on the retention test ( $M = 8.15$ ,  $SD = 2.59$ ) than learners who received a formal version of the narrated animation ( $M = 6.93$ ,  $SD = 2.66$ ),  $F(1, 206) = 11.41$ ,  $p < .01$ , two-sided). The effect size (Cohen's  $d$ ) was 0.46 on retention.

Quelle: Rey und Steib (2013)

Eighty students of the University of Tübingen, Germany participated in the study (64 female, 16 male, average age:  $M = 24.0$  years,  $SD = 2.5$ ) for either payment or course credit. Partic-

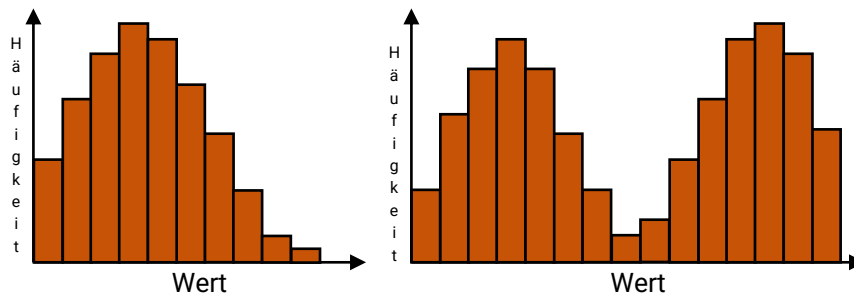
Quelle: Huff, Bauhoff und Schwan (2012)

# Weitere Maße der zentralen Tendenz

- **Modalwert (bzw. Modus):** Wert, der in einer diskreten Verteilung am häufigsten vorkommt
- **Beispiel:** Der Modalwert ist hier 3

VPN	1	2	3	4	5	6	7	8
Wert	3	7	5	4	4	3	7	3

- **Unimodale vs. bimodale Verteilung**



- **Voraussetzung:** Mindestens Nominalskalenniveau der Daten

# Weitere Maße der zentralen Tendenz

- **Medianwert (bzw. Median):** Wert, der in der Mitte eines sortierten Datensatzes liegt
- **Beispiel:** Der Medianwert ist hier 4

VPN	1	2	3	4	5	6	7	8	9
Wert	3	3	3	4	4	5	7	7	84

- **Eigenschaften**
  - Summe der Abweichungsbeträge ist minimal
  - Median halbiert die Verteilung
- **Voraussetzung:** Mindestens Ordinalskalenniveau der Daten

# Berechnung von Maßen der zentralen Tendenz

Wie lauten Modus, Median und Mittelwert für den dargestellten Datensatz?

- A: Modus = 0, Median = 2, Mittelwert = 2
- B: Modus = 2, Median = 2, Mittelwert = 3
- C: Modus = 2, Median = 3, Mittelwert = 3
- D: Modus = 3, Median = 3, Mittelwert = 4
- E: Modus = 2, Median = 2, Mittelwert = 2

VPN	1	2	3	4	5	6	7	8
Wert	-3	0	1	2	2	6	7	9

# Weitere Streuungsmaße

- **Variationsbreite (Spannweite, engl. range):** Statistischer Wert, der den Streuungsbereich der Messwerte angibt
- **Mathematisch:** Differenz aus dem größten und dem kleinsten Wert (bei nominalskalierten Daten hingegen die Anzahl der Kategorien)
- **Beispiel:** Die Variationsbreite beträgt hier 81 ( $84 - 3$ ).

VPN	1	2	3	4	5	6	7	8	9
Wert	3	3	3	4	4	5	7	7	84

- **Eigenschaften:** Hohe Ausreißerabhängigkeit
- **Voraussetzung:** Mindestens Nominalskalenniveau der Daten

# Weitere Streuungsmaße

- **Varianz:** Grob gesagt die quadrierte durchschnittliche Abweichung der Messwerte von ihrem Mittelwert
- **Mathematisch:** Die quadrierte Standardabweichung
- **Formel:**
$$SD^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$
- **Eigenschaften**
  - Varianz  $\geq 0$
  - Große Abweichungen werden stärker berücksichtigt als kleine
- **Voraussetzung:** Mindestens Intervallskalenniveau der Daten



# Berechnung von Streuungsmaßen

Wie lauten Range, Varianz und IQR für den dargestellten Datensatz?

- A: Range = 6, Varianz = 4, IQR = 6
- B: Range = 12, Varianz = 16, IQR = 0
- C: Range = 6, Varianz = 6, IQR = 6
- D: Range = 12, Varianz = 16, IQR = 6
- E: Range = 12, Varianz = 6, IQR = 6

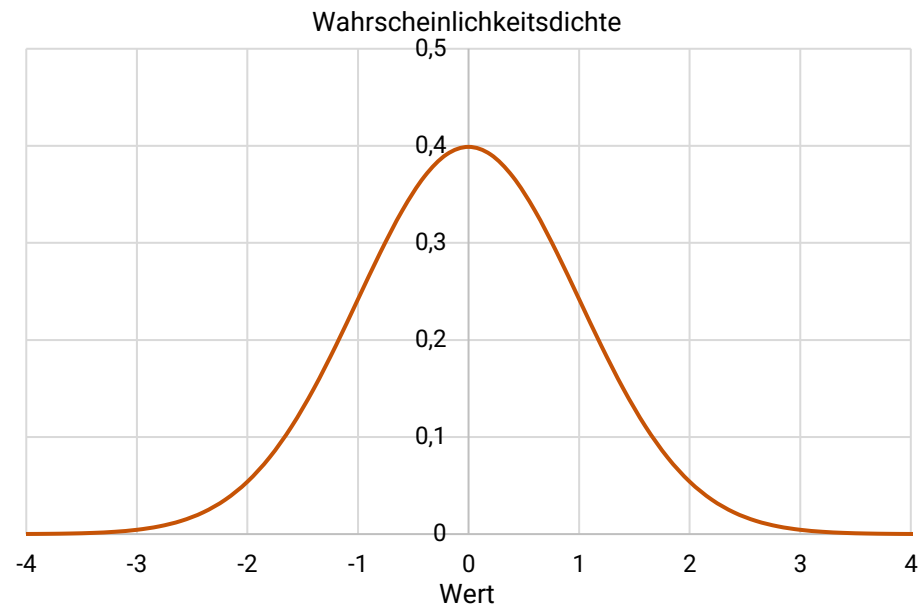
VPN	1	2	3	4	5	6	7	8
Wert	-3	0	1	2	2	6	7	9

# z-Standardisierung

- **z-Standardisierung (z-Transformation):** Abweichungen vom Mittelwert werden durch die Standardabweichung geteilt
- **Formel:** 
$$z_i = \frac{x_i - \bar{x}}{s}$$
- **Verteilung dieser z-Werte:** Mittelwert von 0 und Standardabweichung von 1
- **Voraussetzung:** Mindestens Intervallskalenniveau der Ausgangswerte
- **Verwechslungsgefahr:** z-Standardisierung nicht mit Fishers Z-Transformation verwechseln!
- **z-Standardisierung lineare Transformation:** Verändert die Form der Verteilung nicht

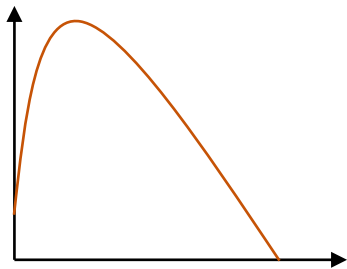
# Normalverteilung

- **Normalverteilung:** Symmetrische, glockenförmige Kurve, die sich der x-Achse asymptotisch nähert und deren Mittelwert, Median und Modus identisch sind
- **Standardnormalverteilung:** Normalverteilung mit einem Mittelwert von 0 und einer Standardabweichung von 1 (z-Werte)
- **Flächenanteile zwischen**
  - -1 und +1: Etwa 68.3%
  - -2 und +2: Etwa 95.4%



# Schiefe

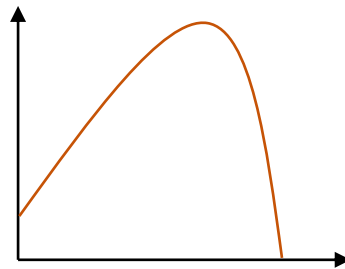
- **Schiefe (engl. skewness oder skew):** Maß für die Asymmetrie einer Verteilung
- **Verteilungen mit unterschiedlichen Schiefen**



**Rechtsschief**

Linkssteil

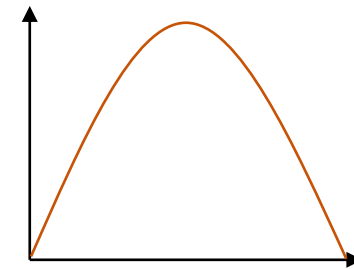
Schiefe  $> 0$



**Linksschief**

Rechtssteil

Schiefe  $< 0$



**Symmetrisch**

Keine Schiefe

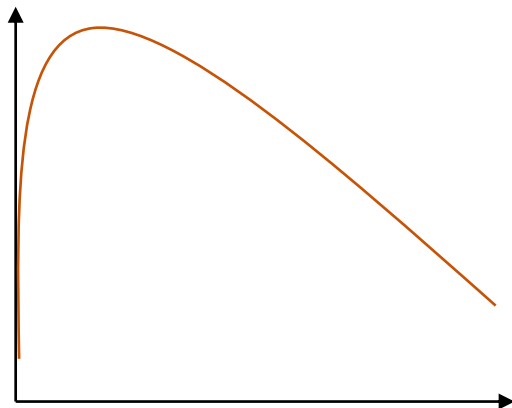
Schiefe  $= 0$

- (x-Achse: Werte; y-Achse: Wahrscheinlichkeitsdichte)

# Schiefe

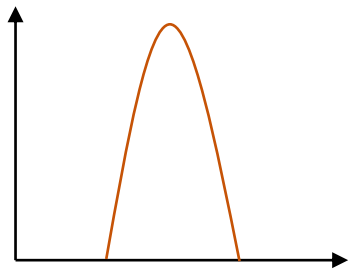
Wo befinden sich Mittelwert, Median und Modalwert in der unten dargestellten rechtsschiefen Verteilung?

- A: Mittelwert  $>$  Median  $>$  Modalwert
- B: Mittelwert  $<$  Median  $<$  Modalwert
- C: Median  $>$  Mittelwert  $>$  Modalwert
- D: Median  $<$  Mittelwert  $<$  Modalwert
- E: Median = Mittelwert = Modalwert



# Kurtosis

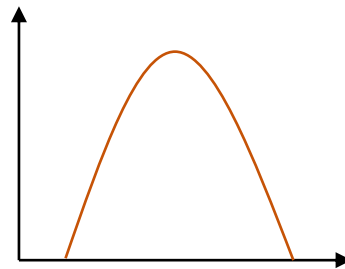
- **Kurtosis (Wölbung)**: Maß für die „Flachheit“ bzw. „Steilheit“ einer Verteilung
- **Exzess**: Teilweise synonym zur Kurtosis verwendet, aber eigentlich gilt:  $\text{Exzess} = \text{Wölbung} - 3$
- **Verteilungen mit unterschiedlicher Kurtosis**



Steilgipflig

Positiver Exzess

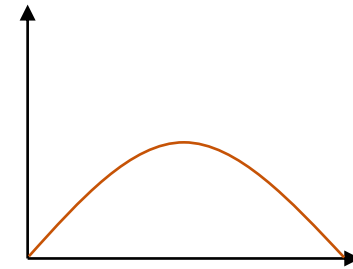
$$\text{Exzess} > 0$$



Normalgipflig

Neutraler Exzess

$$\text{Exzess} = 0$$



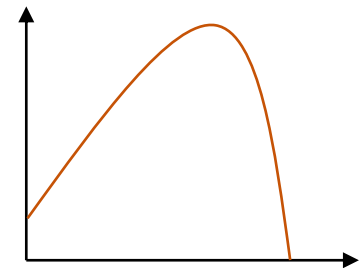
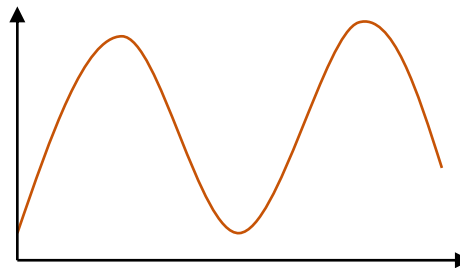
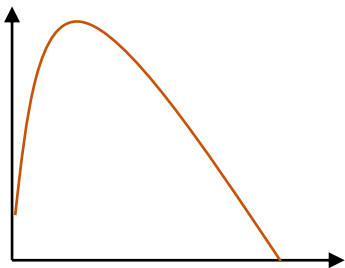
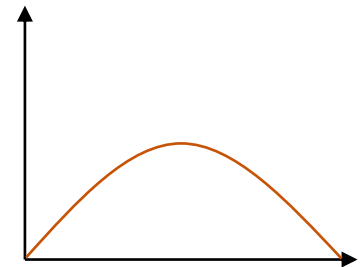
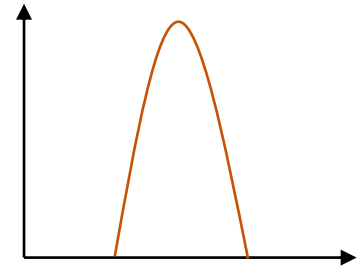
Flachgipflig

Negativer Exzess

$$\text{Exzess} < 0$$

# Zusammenfassung

- **Mittelwert:** Gebräuchlichstes Maß der zentralen Tendenz
- **Standardabweichung:** Gebräuchlichstes Streuungsmaß
- **Weitere Maße der zentralen Tendenz:** Modalwert und Medianwert
- **Weitere Streuungsmaße:** Variationsbreite, Varianz und Interquartilbereich
- **z-Standardisierung**
- **Verschiedene Verteilungen**



# Prüfungsliteratur

- Leonhart, R. (2022). *Lehrbuch Statistik. Einstieg und Vertiefung* (5. Auflage). Bern: Huber.
  - Maße der zentralen Tendenz und der Dispersion (S. 41–79)
- Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2021). *Quantitative Methoden 1: Einführung in die Statistik für Psychologie, Sozial- & Erziehungswissenschaften* (5. Aufl.). Heidelberg: Springer.
  - Deskriptive Statistik (S. 1–20)
  - Inferenzstatistik (S. 23–33)

# Weiterführende Literatur

- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.
  - Statistische Kennwerte (S. 25–36)
- Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden* (5. Aufl.). Weinheim: Beltz.
  - Univariate Deskriptivstatistik (S. 127–168)
- Sedlmeier, P., & Renkewitz, F. (2018). *Forschungsmethoden und Statistik: Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (3. Aufl.). München: Pearson.
  - Lage- und Streuungsmaße (S. 187–206)