

Statistik I

Einheit 3: Statistische Kennwerte (2)

31.10.2024 | Prof. Dr. Stephan Goerigk

Wiederholung Einheit 2 - Was können wir schon?

- Kenntnis der **Skalenniveaus**
- Kenntnis univariater **Maße der zentralen Tendenz**

Kompetenzen:

- Berechnen von **absoluten/relativen Häufigkeiten** und Darstellung in **Häufigkeitstabelle**
- Berechnen von Summen und Notation mit **Summenzeichen**.
- Berechnen von Modus, Mittelwert (mehrere Gruppen, aus Häufigkeitstabelle), Median.

Agenda für Heute:

Ziel: Effiziente Beschreibung **einer Variablen (**univariate** Statistik)**

Was wir danach wissen werden:

- Kenntnis weiterer univariater **statistischer Kennwerte**

Was wir danach können werden:

- Berechnung von **Streuungsmaßen**

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Wozu statistische Kennwerte?

- Bestimmte Eigenschaften einer Verteilung numerisch wiedergeben
- Aus vielen Einzelwerten wenige Werte bilden, die gesamte Verteilung beschreiben
- Werte idealerweise so bestimmen, dass Verteilung aus den Kennwerten rekonstruiert werden könnte

Zwei häufige Arten statistischer Kennwerte:

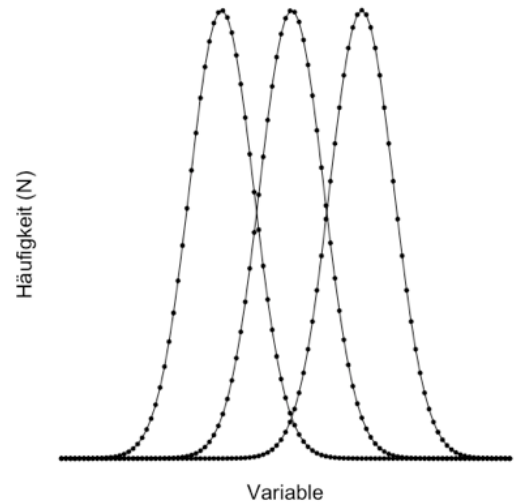
1. **Maße der zentralen Tendenz** (Lagemaße)
 - repräsentieren alle Einzelwerte der Verteilung zusammenfassend
2. **Streuungsmaße** (Dispersionsmaße)
 - geben Auskunft über Variation der Messwerte

Skalenniveaus und statistische Kennwerte

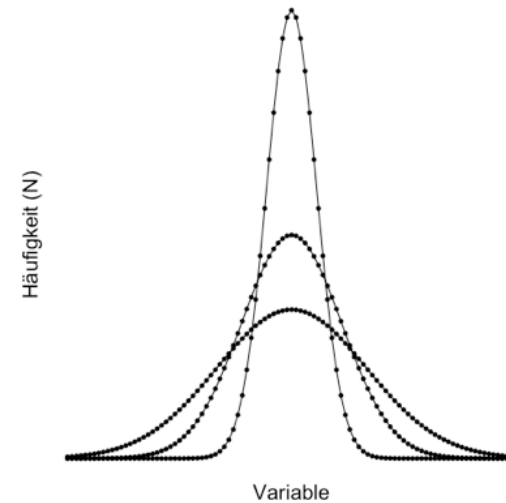
Statistische Kennwerte

Zentrale Tendenz und Streuung

Unterschiedliche Lagemaße, gleiche Streuung



Gleiche Lagemaße, unterschiedliche Streuung



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

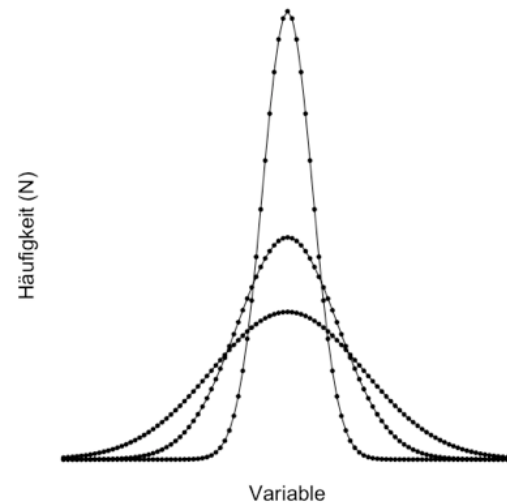
- Maße der zentralen Tendenz können Verteilung nicht vollständig beschreiben
- Sehr unterschiedliche Verteilungen können das selbe Maß der zentralen Tendenz haben
- Streuungsmaße beschreiben, wie stark die einzelnen Werte einer Verteilung von der zentralen Tendenz abweichen
- Sind unimodale Verteilungen mit gleichem Mittelwert in ihrer Form verschieden, dann muss dies an der unterschiedlich starken Streuung der Variablen liegen

Skalenniveaus und statistische Kennwerte

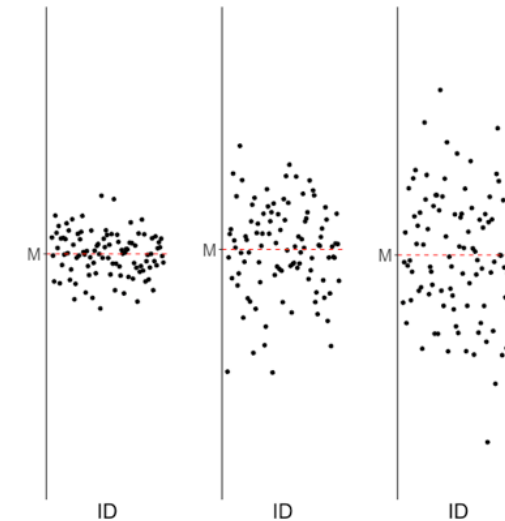
Statistische Kennwerte

Streuungsmaße:

Gleiche Lagemaße, unterschiedliche Streuung



Darstellung der Einzelpersonen



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Spannweite R (aka Variationsbreite, *engl. range*):

- Größe des Bereichs in dem die Messwerte liegen
- Differenz aus größten und kleinsten Wert
- stark von **Ausreißern** beeinflusst
- nur für **metrische** Variablen sinnvoll

Beispiel: Spannweite einer Variable X mit ($N = 10$)

ID	1	2	3	4	5	6	7	8	9	10
X	10	11	9	10	12	10	10	11	8	9

Formel:

$$R = x_n - x_1$$

$$R = 12 - 8 = 4$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Varianz s^2 bzw. σ^2 (engl. *variance*):

- **wichtigstes Streuungsmaß** in der Psychologie
- (wie Mittelwert) nur für **metrische Daten**
- Summe der Abweichungen um den Mittelwert
- Abweichungen werden vor dem summieren **quadriert** (2 Gründe)
 1. Positive und negative Werte heben sich nicht auf
 2. Größere Abweichungen werden stärker berücksichtigt (Potenzfunktion)
- Varianz nimmt größere Werte an, je stärker Messwerte vom Mittelwert abweichen
- **Problem der Varianz:** Durch das Quadrieren geht ursprüngliche Einheit der Variable verloren

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Varianz (s^2):

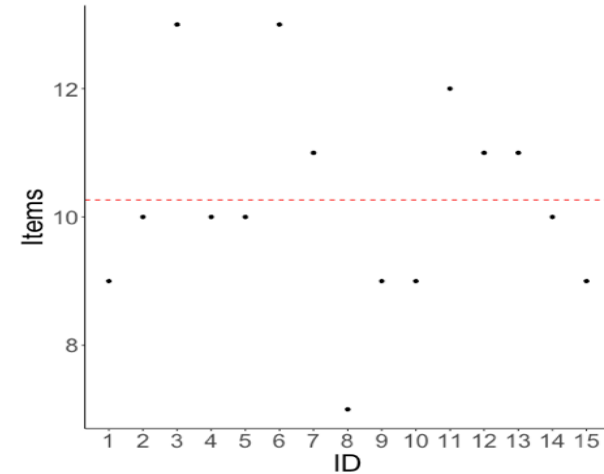
Beispiel: Varianz einer Variable **erinnerte Items** in Gedächtnisaufgabe ($N = 15$)

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Items	9	10	13	10	10	13	11	7	9	9	12	11	11	10	9

Formel:

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

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



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

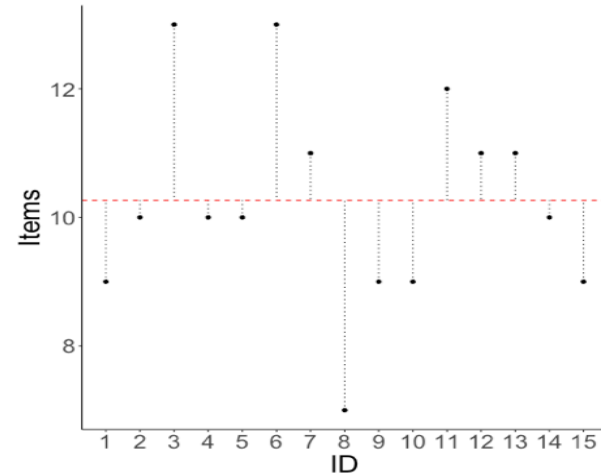
Varianz (s^2):

Beispiel: Varianz einer Variable **erinnerte Items** in Gedächtnisaufgabe ($N = 15$)

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Items	9	10	13	10	10	13	11	7	9	9	12	11	11	10	9

Formel:

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



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Varianz (s^2):

Beispiel: Varianz einer Variable **erinnerte Items** in Gedächtnisaufgabe ($N = 5$)

ID	1	2	3	4	5
Items	9	10	13	10	10

Lösungsweg:

$$\bar{x} = \frac{9 + 10 + 13 + 10 + 10}{5} = \frac{52}{5} = 10.4$$

$$\sigma^2 = \frac{(9 - 10.4)^2 + (10 - 10.4)^2 + (13 - 10.4)^2 + (10 - 10.4)^2 + (10 - 10.4)^2}{5 - 1} = \frac{9.2}{4} = 2.3$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Standardabweichung s bzw. σ (engl. *standard deviation*):

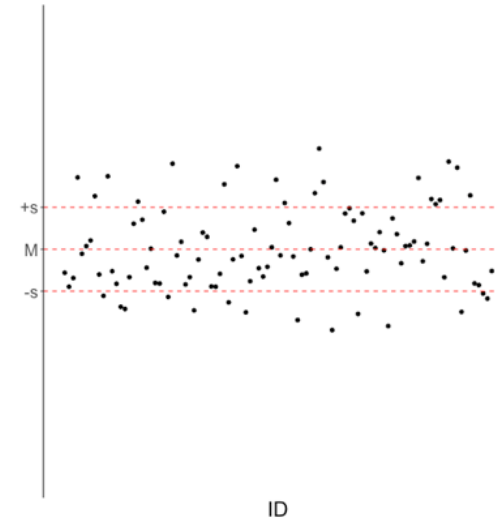
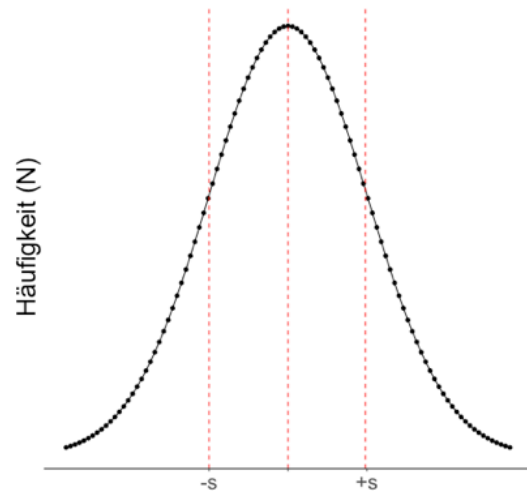
- positive Wurzel aus der Varianz
- gibt Streuung in Einheit der Variable an (Vorteil gegenüber Varianz)
- Abstand des Mittelwerts zum **Wendepunkt** einer Normalverteilung (für NV siehe Folgevorlesung)
- zeigt ebenfalls wie stark Messwerte um Mittelwert streuen
- (wie Varianz und Mittelwert) nur für **metrische** Variablen
- **Ausreißer** beeinflussen die Varianz stark, da Bezugsgröße der Mittelwert ist
- $s = 0$ bedeutet, dass es keinerlei Unterschiede zwischen Messwerten gibt

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Standardabweichung (s):



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Standardabweichung (s):

68 - 95 - 99.7 Faustregel (gilt bei annähernd normalverteilten Daten):

- 68% aller Daten liegen innerhalb 1 Standardabweichung vom Mittelwert
- 95% aller Daten liegen innerhalb 2 Standardabweichungen vom Mittelwert
- 99.7% aller Daten liegen innerhalb 3 Standardabweichungen vom Mittelwert

→ Bei normalverteilten Daten reicht also ein kurzer Blick auf den Mittelwert und die Standardabweichung, um eine Vorstellung davon zu erhalten, in welchen Bereich sich die meisten Daten bewegen

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Beispiel - Standardabweichung:

Standardabweichung einer Variable **erinnerte Items** in Gedächtnisaufgabe ($N = 5$)

ID	1	2	3	4	5
Items	9	10	13	10	10

Formel:

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

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

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Beispiel - Standardabweichung:

Standardabweichung einer Variable **erinnerte Items** in Gedächtnisaufgabe ($N = 5$)

ID	1	2	3	4	5
Items	9	10	13	10	10

Lösungsweg:

$$\bar{x} = \frac{9 + 10 + 13 + 10 + 10}{5} = \frac{52}{5} = 10.4$$

$$\sigma^2 = \frac{(9 - 10.4)^2 + (10 - 10.4)^2 + (13 - 10.4)^2 + (10 - 10.4)^2 + (10 - 10.4)^2}{5 - 1} = \frac{9.2}{4} = 2.3$$

$$\sigma = \sqrt{2.3} = 1.52$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Quartilabstand:

- so wie der Median **robust** gegenüber Ausreißern (dann geeignetes Streuungsmaß)
- Berechnung des Quartilabstandes ist sinnvoll für mindestens **rangskalierte** Variablen
- Zwischen dem unteren und oberen Quartil liegen **50% aller Werte**

Formel:

$$q_A = \tilde{x}_{0.75} - \tilde{x}_{0.25}$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Quartilabstand:

Berechnung der Quantile:

- Der Median ist das Quantil mit $\alpha = 0.50$ (50% Quantil)
- Die Quantile mit $\alpha = 0.25$ und $\alpha = 0.75$ heißen unteres bzw. oberes Quartil

Formel:

$$\tilde{x}_{\alpha} = \begin{cases} x_{(l)} & \text{falls } n \cdot \alpha \text{ keine ganze Zahl ist;} \\ & l = \text{die auf } n \cdot \alpha \text{ folgende ganze Zahl} \\ \frac{x_{(l)} + x_{(l+1)}}{2} & \text{falls } n \cdot \alpha \text{ eine ganze Zahl ist;} \\ & l = n \cdot \alpha \end{cases}$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Beispiel - Quartilabstand: Stunden Smartphonenuutzung pro Woche ($N = 15$)

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Stunden	9	10	17	11	12	18	13	6	8	9	16	12	13	11	9	

Lösungsweg:

$$x_{\text{sortiert}} = 6; 8; 9; 9; 9; 10; 11; 11; 12; 12; 13; 13; 16; 17; 18$$

$$n \cdot \alpha = 15 \cdot 0.25 = 3.75 \rightarrow I_{0.25} = 4$$

$$n \cdot \alpha = 15 \cdot 0.75 = 11.25 \rightarrow I_{0.75} = 12$$

$$q_{0.25} = x_{(I_{0.25})} = x_{(4)} = 9$$

$$q_{0.75} = x_{(I_{0.75})} = x_{(12)} = 13$$

$$q_A = \tilde{x}_{0.75} - \tilde{x}_{0.25} = 13 - 9 = 4$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Streuungsmaße:

Beispiel: Quartilabstand Stunden Smartphonenuutzung pro Woche ($N = 20$)

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Stunden	9	10	17	11	12	18	13	6	8	9	16	12	13	11	9	18	13	3	14	9

Lösungsweg:

$$x_{\text{sortiert}} = 3; 6; 8; 9; 9; 9; 9; 10; 11; 11; 12; 12; 13; 13; 13; 14; 16; 17; 18; 18$$

$$n \cdot \alpha = 20 \cdot 0.25 = 5 \rightarrow I_{0.25} = 5$$

$$n \cdot \alpha = 20 \cdot 0.75 = 15 \rightarrow I_{0.75} = 15$$

$$q_{0.25} = \frac{x_{(I_{0.25})} + x_{(I_{0.25}+1)}}{2} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{9 + 9}{2} = 9$$

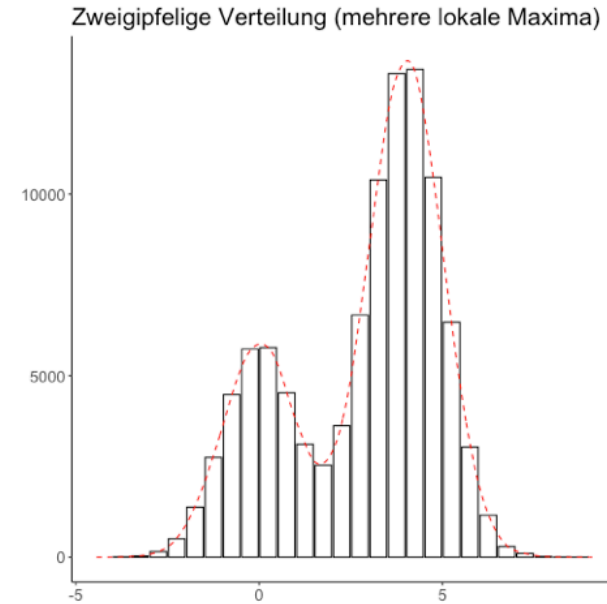
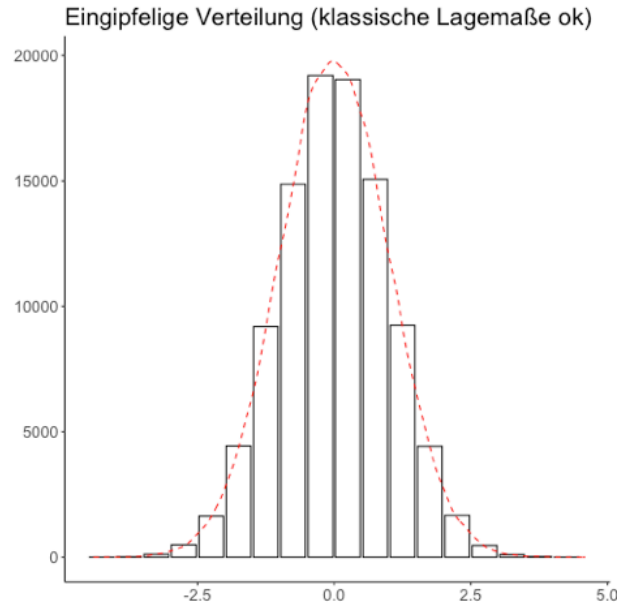
$$q_{0.75} = \frac{x_{(I_{0.75})} + x_{(I_{0.75}+1)}}{2} = \frac{x_{(15)} + x_{(16)}}{2} = \frac{13 + 14}{2} = 13.5$$

$$q_A = \tilde{x}_{0.75} - \tilde{x}_{0.25} = 13.5 - 9 = 4.5$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Eingipfelige vs. mehrgipfelige Verteilung



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Schiefe (engl. *curtosis*):

- Die Schiefe g_1 ist eine Maßzahl, die uns angibt, in welche Richtung eine Häufigkeitsverteilung schief ist.
- Ist $g_1 = 0$, so ist die Verteilung symmetrisch.
- Je stärker negativ/positiv g_1 ist, desto linksschiefer/rechtsschiefer ist die Verteilung.
- Die Schiefe ist sinnvoll für eingipfelige Verteilungen.
- ACHTUNG: Muss in Prüfung nicht berechnet werden.

Formel:

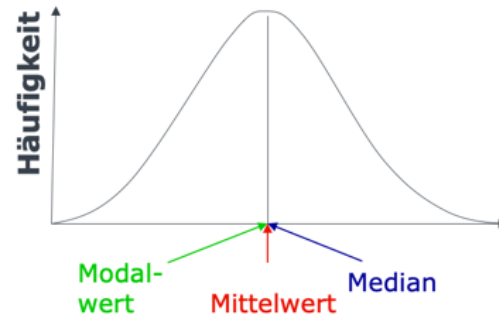
$$g_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^3}}$$

Skalenniveaus und statistische Kennwerte

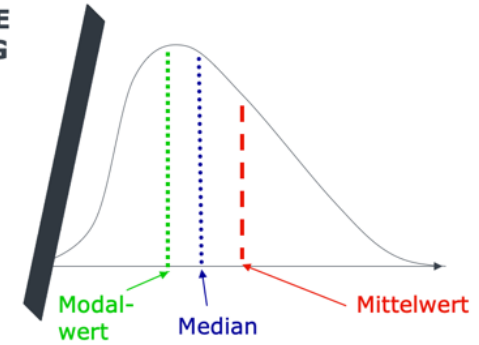
Statistische Kennwerte

Schiefte (engl. *curtosis*):

SYMMETRISCHE VERTEILUNG



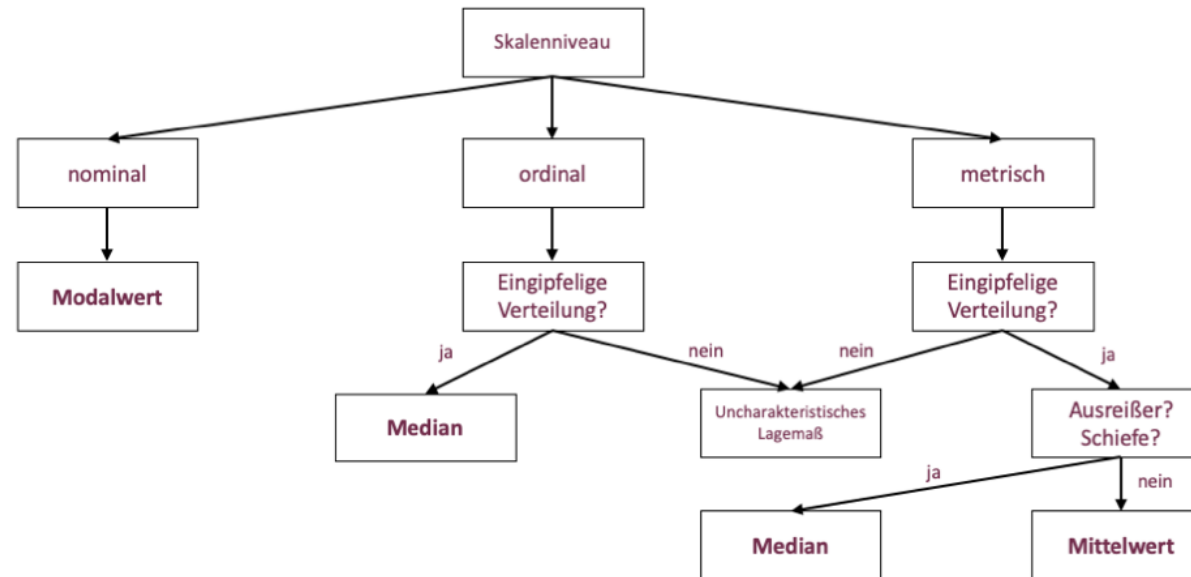
LINKSSTEILE VERTEILUNG



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

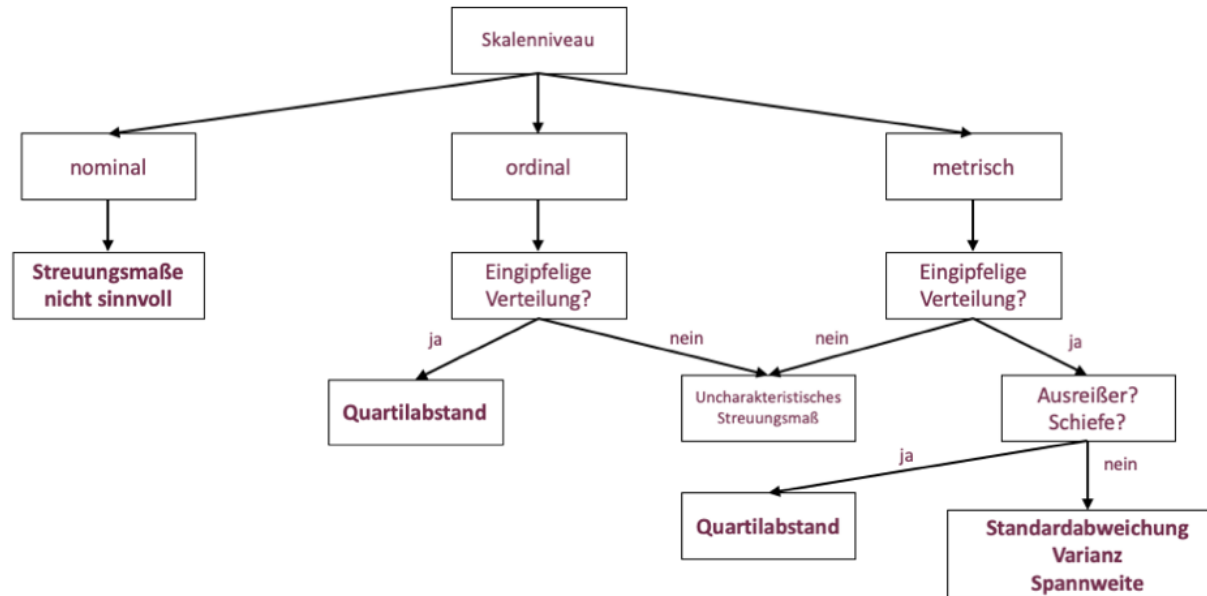
Wegweiser Maße der zentralen Tendenz:



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Wegweiser Streuungsmaße:



Skalenniveaus und statistische Kennwerte

Übung: Erstellen einer Stichprobenbeschreibungstabelle (Tabelle 1 eines Papers)

Muster Datenmatrix:

	1	2	3	4	5	6
ID						
Geschlecht	weiblich	weiblich	weiblich	weiblich	männlich	weiblich
Alter	42	55	66	19	23	41
BDI	8	12	41	20	14	2
Behandlungsgruppe	A	A	A	B	B	B

Muster Stichprobenbeschreibungstabelle (zusammengefasste Daten):

Charakteristik	A (N=3)	B (N=3)	Total (N=6)
Geschlecht, w, n (%)	3 (100)	2 (66.67)	5 (83.33)
Alter, m (SD)	54.33 (12.01)	27.67 (11.72)	41 (18.06)
BDI, m (SD)	20.33 (18.01)	12 (9.17)	16.17 (13.57)

Skalenniveaus und statistische Kennwerte

Übung: Erstellen einer Stichprobenbeschreibungstabelle (Tabelle 1 eines Papers)

Datenmatrix:

	1	2	3	4	5	6
ID						
Geschlecht	männlich	männlich	weiblich	weiblich	männlich	weiblich
Alter	63	52	19	31	54	32
BDI	44	2	3	55	12	41
Behandlungsgruppe	B	B	B	A	A	B

Aufgabe: Erstellen Sie für gegebene Datenmatrix eine Stichprobenbeschreibungstabelle!

(BDI = Beck Depression Inventory; Fragebogen zur Messung depressiver Symptomatik)

Take-aways

- **Streuungsmaße** geben Variation der Messwerte um Maß der zentralen Tendenz an.
- **Varianz** wichtigstes aber unpraktisches Maß der Streuung um den Mittelwert.
- **Standardabweichung** ist Wurzel aus Varianz und kann in Einheit der Variable interpretiert werden.
- Varianz und Standardabweichung nur bei **metrischen Daten ohne Ausreißer** sinnvoll.
- Bei Vorliegen von ordinal skalierten Daten und Ausreißern berechnen wir den **Quartilabstand**.
- Die erste Tabelle eines Papers ist i.d.R. eine **Stichprobenbeschreibung**. Diese enthält für kategoriale Variablen absolute/relative Häufigkeiten und für numerische Variablen Lagemaße/Streuungsmaße.

