

Statistik I

Einheit 2: Skalenniveaus und statistische Kennwerte

24.10.2024 | Prof. Dr. Stephan Goerigk

Wiederholung Einheit 1 - Was können wir bereits?

- Wir kennen unterschiedliche Variablenarten
 - stetig vs. diskret
 - manifest vs. latent
 - AV vs. UV
 - Drittvariablen: Stör- und Kontrollvariable, Moderator, Mediator
- Wir kennen Kriterien wissenschaftlicher Hypothesen (ungerichtet vs. gerichtet)

Kompetenzen:

- Erfassung **einer** Variable in eine **Urliste/Vektor**
- Kombination **mehrer** Variablen in einer **Datenmatrix**
- Berechnen von **absoluten/relativen Häufigkeiten** und Darstellung in **Häufigkeitstabelle**
- Berechnen von Summen und Notation mit **Summenzeichen**.

Agenda für Heute:

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

Was wir danach wissen werden:

- Kenntnis der **Skalenniveaus**
- Kenntnis univariater **statistischer Kennwerte**

Was wir danach können werden:

- **Darstellung** einer Variable
- Zuordnung von Variablen zu einem **Skalenniveau**
- Zusammenfassung/Beschreibung einer Variablen mit **statistischen Kennwerten** (Deskriptivstatistik)

Skalenniveaus und statistische Kennwerte

Darstellung einer Variable

- Mit der **Häufigkeitstabelle** kennen wir bereits eine Möglichkeit, die Ausprägungen einer Variable darzustellen

Urliste ($N = 30$): Anzahl korrekt gelöster Aufgaben in einem Intelligenztest (10 Fragen)

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
IQ	4	5	9	6	6	10	7	2	4	4	9	6	7	6	4	10	7	1	7	4	3	5	3	4	4	1	8	6	3	9

Häufigkeitstabelle:

1	2	3	4	5	6	7	8	9	10
2 (6.67)	1 (3.33)	3 (10)	7 (23.33)	2 (6.67)	5 (16.67)	4 (13.33)	1 (3.33)	3 (10)	2 (6.67)

→ Wie wir sehen, schaffen es die meisten Probanden eine mittlere Anzahl von Fragen zu lösen [ca. 4-7]

Skalenniveaus und statistische Kennwerte

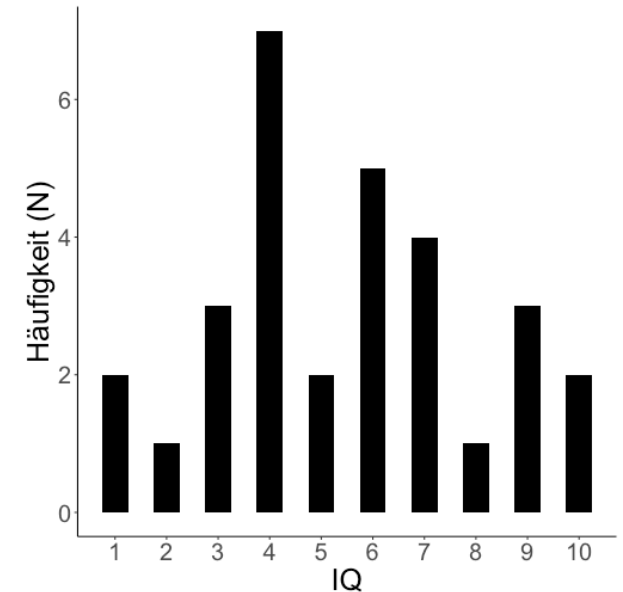
Darstellung einer Variable

Häufigkeiten lassen sich auch graphisch darstellen, mit einem **Histogramm** (Säulendiagramm):

Häufigkeitstabelle:

1	2	3	4	5	6	7	8	9	10
2 (6.67)	1 (3.33)	3 (10)	7 (23.33)	2 (6.67)	5 (16.67)	4 (13.33)	1 (3.33)	3 (10)	2 (6.67)

- **X-Achse:** Ausprägungen der Variablen
- **Y-Achse:** Absolute Häufigkeiten dieser Ausprägungen
- Ziel: Visualisierung der Datenstruktur (aka der **Verteilung**) der Variable



Skalenniveaus und statistische Kennwerte

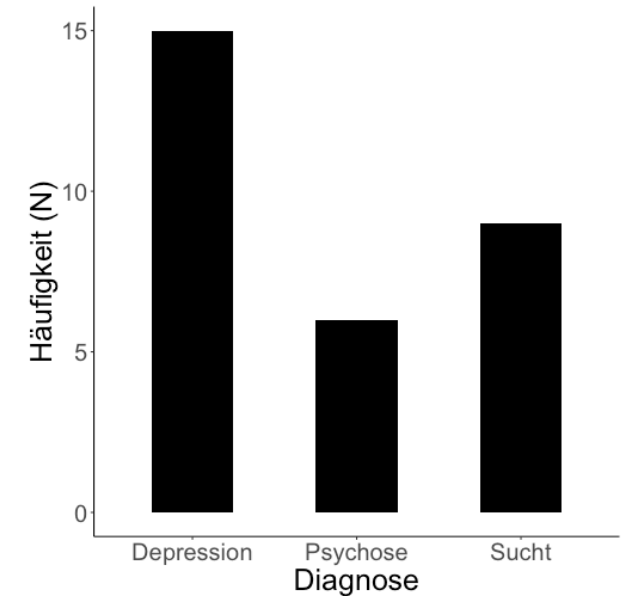
Darstellung einer Variable

Auch **Kategorien** (Wörter) lassen sich so darstellen (dann heißt das Diagramm **Balkendiagramm**):

Häufigkeitstabelle: z.B. Diagnosen

Depression	Psychose	Sucht
15 (50)	6 (20)	9 (30)

- **X-Achse:** Ausprägungen der Variablen
- **Y-Achse:** Absolute Häufigkeiten dieser Ausprägungen
- Ziel: Visualisierung der Datenstruktur (aka der **Verteilung**) der Variable



Skalenniveaus und statistische Kennwerte

Darstellung einer Variable

Noch einmal zusammengefasst:

- Histogramm oder Blockdiagramm - Grafische Darstellung der Häufigkeitstabelle
- Beobachtete Ausprägungen geordnet auf der X-Achse
- Relative (r_j) oder absolute Häufigkeiten (f_j) auf der Y-Achse
- Rechtecksflächen sind gleich den r_j oder f_j
- Gesamtfläche des Histogramms ist gleich 1 (relative Häufigkeiten) bzw. N (absolute Häufigkeiten)
- Maßstab auf der X-Achse beliebig und wird so gewählt, dass die Verteilung möglichst anschaulich wird (wichtig bei stetigen Variable mit ggf. tausenden Ausprägungen)
- Histogramm für numerische Variablen; Balkendiagramm für kategoriale Variablen

Skalenniveaus und statistische Kennwerte

Darstellung einer Variable

Vorteile des Histogramms

- Visualisiert **Häufigkeitsverteilung** einer Variable
- Zeigt, welche Merkmalsausprägungen besonders **typisch/häufig** sind
- Zeigt, ob es exotische/untypische Werte in der Verteilung gibt (**Ausreißer**)

Aufgabe: Zeichnen Sie ein Histogramm für die folgende Variable $N = 20$: **Alkoholkonsum** (Gläser/Woche)

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Gläser	1	2	4	2	2	4	3	0	1	1	3	2	2	2	1	4	3	0	3	1

Tipp:

- **Schritt 1:** Häufigkeitstabelle erstellen (absolute Häufigkeiten)
- **Schritt 2:** Histogramm zeichnen

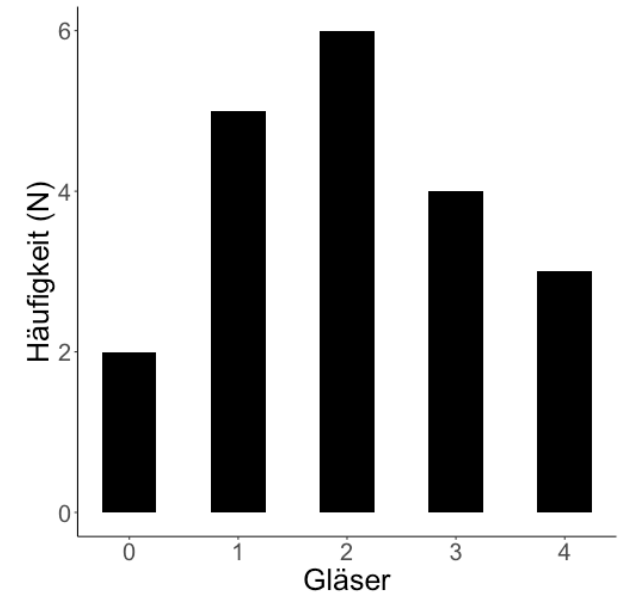
Skalenniveaus und statistische Kennwerte

Darstellung einer Variable

Lösung:

Häufigkeitstabelle: Alkoholkonsum (Gläser/Woche)

0	1	2	3	4
2 (10)	5 (25)	6 (30)	4 (20)	3 (15)



Skalenniveaus und statistische Kennwerte

Skalenniveaus

- Um mit Variablen rechnen zu können, müssen wir sie zunächst in Zahlen darstellen (**quantifizieren**)
- Dieser Vorgang heißt in der Wissenschaft "**Messen**"
- **Definition:** Messen = homomorphe Abbildung eines empirischen Relativs in ein numerisches Relativ

Wie funktioniert Messen?

- einzelnen Ausprägungen werden Messwerte (Zahlen) auf einer **Skala** zugeordnet
- **Skala** = Vorschrift, die jeder Person der Stichprobe einen Beobachtungswert zuordnet
- Variablen lassen sich unterschiedlich differenziert in Zahlen abbilden (**Messniveau aka Skalenniveau**)



Skalenniveaus und statistische Kennwerte

Skalenniveaus

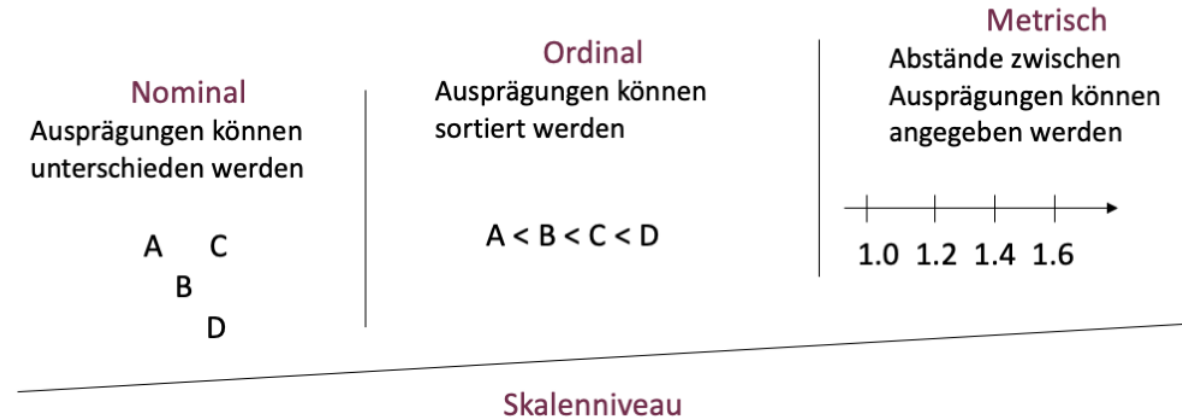
- **Skalenniveau:** eine der wichtigsten statistischen Eigenschaften einer Variable
- bestimmt welche Rechenoperationen (statistische Tests) mit der Variable **zulässig** sind
- Je höher das Skalenniveau, desto mehr Vergleichsaussagen und Rechenoperationen sind möglich

Es gibt 3 hierarchisch geordnete Skalenniveaus:

1. **Nominal** (kategorial)
2. **Ordinal** (kategorial)
3. **Metrisch**
 - 3.1. **Intervallskala**
 - 3.2. **Verhältnisskala**

Skalenniveaus und statistische Kennwerte

Skalenniveaus



Einordnung des Skalenniveaus abhängig von:

1. **Eigenschaften** des zu messenden Merkmals selbst
2. Art der Abbildung durch das **Messinstrument**

Skalenniveaus und statistische Kennwerte

Nominalskala

- **niedrigstes** Skalenniveau in der Statistik (niedrigster Informationsgehalt)
- Ausprägungen der Variablen können **unterschieden** werden (nur Beziehungen "gleich", "ungleich" möglich)
- eine logische Reihenfolge ist nicht möglich (Reihenfolge der Ausprägungen ist **austauschbar**)
- Erlauben **eindeutige Transformationen**
- **Codierung**: Zuweisung von Zahlen zu verbalen Kategorien

Beispiele:

Geschlecht	Beruf	Geburtsort
1 = männlich	1 = Dachdecker	1 = Berlin
2 = weiblich	2 = Psychologe	2 = Frankfurt
3 = divers	3 = Polizist	3 = Paris
	4 = Koch	4 = New York

Skalenniveaus und statistische Kennwerte

Ordinalskala

- Kann auch nominale Aussagen abbilden (Gleichheit/Verschiedenheit)
- Zusätzlich: **Größer-Kleiner-Relationen** (eine Rangfolge kann gebildet werden)
- Wird auch **Rangskala** genannt
- Aussage über den absoluten Abstand zwischen zwei Werten ist nicht möglich
- Erlauben **monotone Transformationen**

Beispiele:

Schulnote	Wettlauf	Alkoholkonsum
1 = Note 1	1 = Erster	1 = täglich
2 = Note 2	2 = Zweiter	2 = einmal pro Woche
3 = Note 3	3 = Dritter	3 = einmal pro Monat
...	...	4 = nie

Skalenniveaus und statistische Kennwerte

Metrische Variablen

- Merkmalsausprägungen können verglichen und sortiert werden
- Zusätzlich: **Abstände** zwischen den Ausprägungen können berechnet werden
- Gleich große Abstände zwischen zugeordneten Zahlen repräsentieren gleich große Einheiten des Kontrukts (Differenzen und Summen können sinnvoll gebildet werden)
- Es werden **Intervallskala** und **Verhältnisskala** unterschieden

Beispiele:

Einkommen	Körpergröße	IQ
3522 €	175 cm	102
4225 €	156 cm	98
8327 €	192 cm	121
2174 €	181 cm	106

Skalenniveaus und statistische Kennwerte

Metrische Variablen

Intervallskala vs. Verhältnisskala

Intervallskala

- **Wichtigste** Skala in den Sozialwissenschaften
- Macht Aussagen über Größe der Unterschiede zwischen Merkmalsausprägungen
- Erlaubt **lineare Transformationen** $y_i = a + b \cdot x_i$

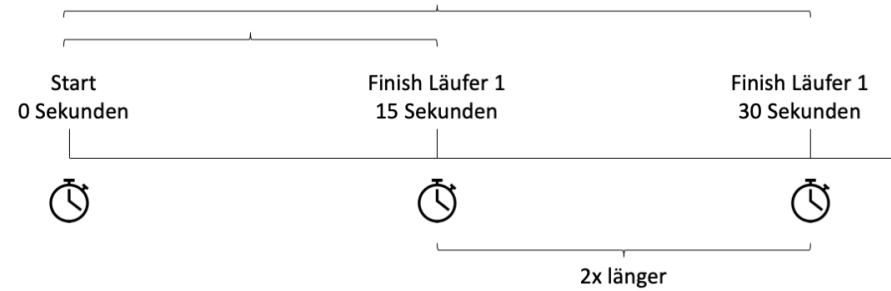
Verhältnisskala

- eine Annahme mehr: Skala hat **absoluten Nullpunkt**
- Nullpunkt lokalisiert, wo die Variable aufhört zu **existieren**
- **Selten** in Sozialwissenschaften (Variablen wie Intelligenz, Neurotizismus... haben keine sinnvolle 0)
- macht somit Aussagen über das Verhältnis von Merkmalsausprägungen
- Erlaubt **Ähnlichkeitstransformationen** $y_i = b \cdot x_i$

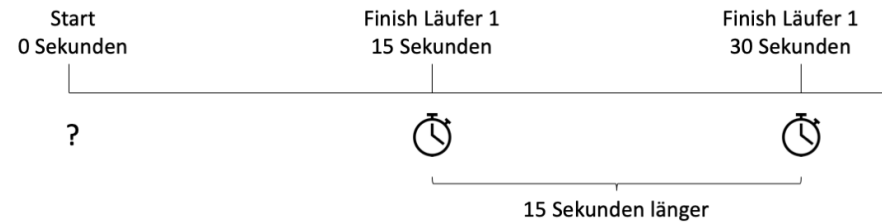
Skalenniveaus und statistische Kennwerte

Metrische Variablen

Beispiel Verhältnisskala:



Beispiel Intervallskala:



Skalenniveaus und statistische Kennwerte

Zusammenfassung (Rasch, 2008)

Skalenniveau	Aussagen	Beispiele	Transformationen
Nominalskala	Gleichheit/Verschiedenheit	Diagnosen, Nationalität, Gruppenzugehörigkeit	wenn x_1 ungleich x_2 dann y_1 ungleich y_2
Ordinalskala	Größer-Kleiner Relation	Schulabschlüsse, Bundesligatabelle, Medaillenfarbe	wenn $x_1 > x_2 > x_3$ dann $y_1 > y_2 > y_3$ oder y_1
Intervallskala	Gleichheit von Differenzen	IQ, Persönlichkeit	$y_i = a + b \cdot x_i$
Verhältnisskala	Gleichheit von Verhältnissen	Gewicht, Länge	$y_i = b \cdot x_i$

Skalenniveaus und statistische Kennwerte

Veränderung des Skalenniveaus:

- Variablen lassen sich abwärts der Hierarchie des Skalenniveaus **runterskalieren**
- **Aggregation:** Zusammenfassen auf das nächst-gröbere Level
- von grob nach fein, also nominal → ordinal → intervall ist jedoch **nicht möglich**
- Beispiel Körpergröße in cm ($N = 10$):

ID	1	2	3	4	5	6	7	8	9	10
metrisch	151	170	161	192	182	201	188	162	174	180
ordinal	10.	7.	9.	2.	4.	1.	3.	8.	6.	5.
nominal	klein	klein	klein	groß	groß	groß	groß	klein	klein	groß

Skalenniveaus und statistische Kennwerte

Festlegung des Skalenniveaus:

- **Praxis:** es kommt vor, dass eingesetzte Messinstrumente Daten auf niedrigerem Skalenniveau erfassen, als theoretisch möglich (z.B. aus Sparsamkeit)
- **Statistischer Nachweis** des Messniveaus (z.B. eines Intelligenztests) oft sehr aufwändig (eigene Disziplin: **Messtheorie**)
- Psychologie: **Likertskala** wird häufig für Fragebogenitems (Einzelfragen) eingesetzt (z.B. 1 = stimmt nicht zu, 2 = stimme etwas zu...)
- **Einzelwert** des Items auf der Likerskala i.d.R. ordinalskaliert
- **Gesamtfragebogenwert** (Summe der der Einzelitems) wird als metrisch behandelt
- Scores gut konstruierter **psychologischer Tests**: Intervallskalenniveau wird angenommen (Steier & Eid, 1999)

Skalenniveaus und statistische Kennwerte

Beispiele:

Beispiel	Skalenniveau
Wohnorte in Deutschland	?
Hotelbewertung auf einer Skala von 1 bis 5	?
Religionsbekenntnis	?
CO2-Ausstoss im Jahr	?
Motivationsscore von Arbeitnehmern	?
Zeugnisnoten von 1 bis 6	?
Telefonnummern von Befragten	?
Pflegestufe eines Patienten	?
Wohnfläche in m2	?
Arbeitszufriedenheit auf einer Skala von 1 bis 4	?

Skalenniveaus und statistische Kennwerte

Beispiele:

Beispiel	Skalenniveau
Wohnorte in Deutschland	Nominalskala
Hotelbewertung auf einer Skala von 1 bis 5	Ordinalskala
Religionsbekenntnis	Nominalskala
CO2-Ausstoss im Jahr	metrisch, Verhältnisskala
Motivationsscore von Arbeitnehmern	metrisch, Intervallskala
Zeugnisnoten von 1 bis 6	Ordinalskala
Telefonnummern von Befragten	Nominalskala
Pflegestufe eines Patienten	Ordinalskala
Wohnfläche in m2	metrisch, Verhältnisskala
Arbeitszufriedenheit auf einer Skala von 1 bis 4	Ordinalskala

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:

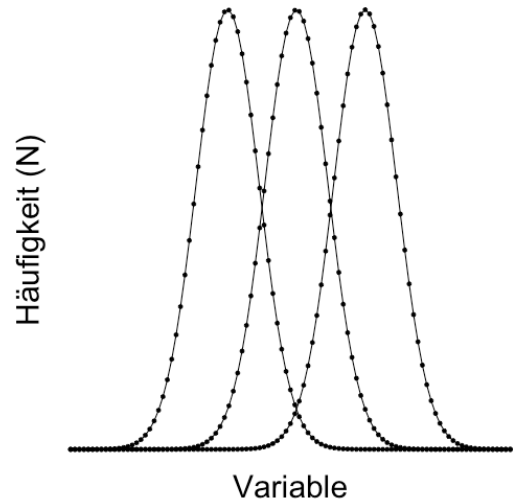
1. **Maße der zentralen Tendenz** (aka. Lagemaße)
 - repräsentieren alle Einzelwerte der Verteilung zusammenfassend
2. **Streuungsmaße** (aka. Dispersionsmaße)
 - geben Auskunft über Variation der Messwerte

Skalenniveaus und statistische Kennwerte

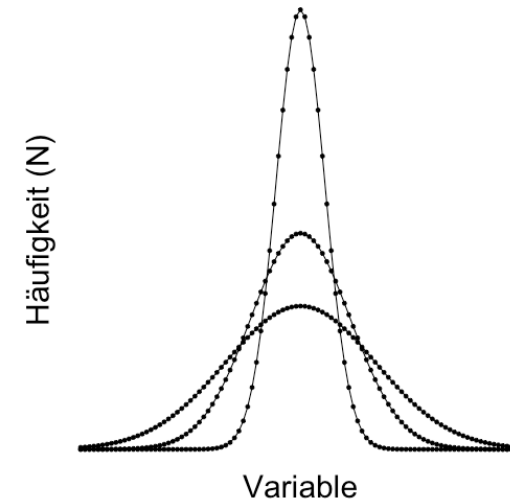
Statistische Kennwerte

Zentrale Tendenz und Streuung

Unterschiedliche Lagemaße, gleiche Streuung



Gleiche Lagemaße, unterschiedliche Streuung



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Gängige Maße der zentralen Tendenz:

- Modalwert
- Arithmetisches Mittel (Mittelwert)
- Median

Gängige Streuungsmaße:

- Spannweite
- Varianz und Standardabweichung
- Quartilabstand

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Glossar Symbole:

- Auf Stichprobenebene werden **lateinische Buchstaben** verwendet
- Auf Populationsebene werden **griechische Buchstaben** verwendet

Statistik	Stichprobenebene	Populationsebene	Populationsschätzer
Mittelwert	\bar{x}	μ	$\hat{\mu}$
Median	Md	η	$\hat{\eta}$
Varianz	s^2	σ^2	$\hat{\sigma}^2$
Standardabweichung	s	σ	$\hat{\sigma}$
Korrelationskoeffizient	r	ρ	$\hat{\rho}$
Regressionskoeffizient	b	β	$\hat{\beta}$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Maße der zentralen Tendenz:

Modalwert (aka Modus)

- Der am **häufigsten vorkommende** Wert einer Verteilung
- folglich auch der **wahrscheinlichste** Wert (bei zufälligem Ziehen aus der Verteilung)
- Berechnung erfordert lediglich **Nominalskalenniveau**
- Wert mit höchster absoluter Häufigkeit (f_j)
- Graphen mit nur einem Modus heißen **unimodal** (aka eingipfelig)
- 2 Modi = **bimodale** Verteilung; mehrere Maximalwerte nebeneinander = **breitgipfelige** Verteilung

Skalenniveaus und statistische Kennwerte

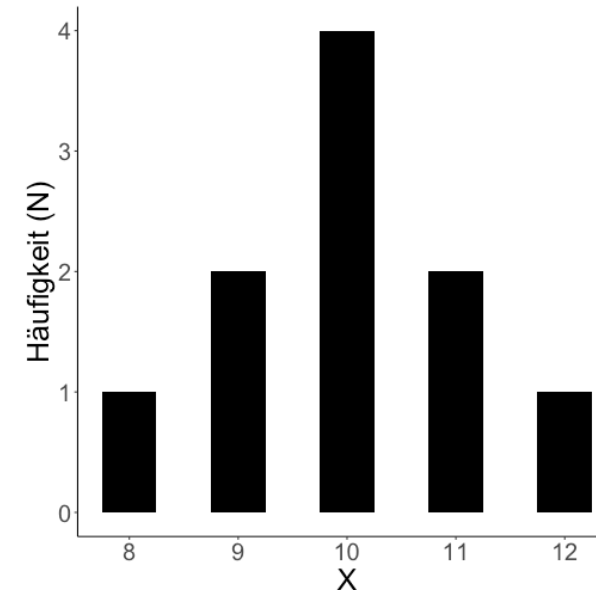
Statistische Kennwerte

Maße der zentralen Tendenz:

Modalwert (aka Modus)

Beispiel, Verteilung mit Modalwert = 10

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



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Maße der zentralen Tendenz:

Mittelwert (aka arithmetisches Mittel)

- gebräuchlichstes Maß der zentralen Tendenz
- **Durchschnittswert** einer Verteilung
- nur für metrische Variablen sinnvoll (**mindestens Intervallskalenniveau**)
- Der Mittelwert einer Variable x wird geschrieben als \bar{x}
- **Berechnung:** Summe aller Werte dividiert durch den Stichprobenumfang N

Formel:

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

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Rechenbeispiel Mittelwert: Mittelwert aus Variable X ($N = 10$)

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

Formel:

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

Lösungsweg:

$$\bar{x} = \frac{10 + 11 + 9 + 10 + 12 + 10 + 10 + 11 + 8 + 9}{10} = \frac{100}{10} = 10$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Aufgabe Mittelwert: Berechnen Sie das Durchschnittsalter folgender Stichprobe (in Jahren) ($N = 12$)

ID	1	2	3	4	5	6	7	8	9	10	11	12
Alter	21	44	52	15	27	52	63	16	99	24	56	40

Formel:

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

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Aufgabe Mittelwert: Berechnen Sie das Durchschnittsalter folgender Stichprobe (in Jahren) ($N = 12$)

ID	1	2	3	4	5	6	7	8	9	10	11	12
Alter	21	44	52	15	27	52	63	16	99	24	56	40

Formel:

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

Lösungsweg:

$$\bar{x} = \frac{21 + 44 + 52 + 15 + 27 + 52 + 63 + 16 + 99 + 24 + 56 + 40}{12} = \frac{509}{12} = 42.42$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Mathematische Eigenschaften des Mittelwerts:

- Die Summe der Differenzen aller Werte vom Mittelwert ist 0

→ positive bzw. negative Abweichungen vom Mittelwert heben sich auf

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

- Die Summe der quadrierten Differenzen aller Werte zum Mittelwert ist ein Minimum

→ Minimum = Wert ist kleiner als die Summe der quadrierten Differenzen aller Werte zu einem anderen Wert

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

Aufgabe: Überprüfen Sie diese Eigenschaften für sich (z.B. mit dem leichten Beispiel der Werte: 1, 2, 3, 4)

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Berechnung des Mittelwerts auf Grundlage der Häufigkeitstabelle:

- Liegt bereits eine Häufigkeitstabelle vor, sparen wir uns Tipparbeit (nicht alle Messwerte müssen einzeln eingegeben werden)
- jede Merkmalsausprägung mit ihrer absoluten Häufigkeit multiplizieren und die Summe über alle Merkmalsausprägungen bilden:

Formel:

$$\bar{x} = \frac{\sum_{j=1}^k x'_j \cdot f_j}{n}$$

- x'_j = mögliche Merkmalsausprägungen; f_j = absolute Häufigkeit der jeweiligen Merkmalsausprägung

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Berechnung des Mittelwerts auf Grundlage der Häufigkeitstabelle

Beispiel von vorher: **Häufigkeitstabelle:** korrekt gelöste Aufgaben in einem Intelligenztest (10 Fragen) $N = 30$

1	2	3	4	5	6	7	8	9	10
2 (6.67)	1 (3.33)	3 (10)	7 (23.33)	2 (6.67)	5 (16.67)	4 (13.33)	1 (3.33)	3 (10)	2 (6.67)

Formel:

$$\bar{x} = \frac{\sum_{j=1}^k x'_j \cdot f_j}{n}$$

Lösungsweg:

$$\bar{x} = \frac{1 \cdot 2 + 2 \cdot 1 + 3 \cdot 3 + 4 \cdot 7 + 4 \cdot 2 + 6 \cdot 5 + 7 \cdot 4 + 8 \cdot 1 + 9 \cdot 3 + 10 \cdot 2}{30} = \frac{162}{30} = 5.47$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Aufgabe Mittelwert: Eine **Häufigkeitstabelle** für die Anzahl an unternommenen Reisen pro Jahr in einer Stichprobe liegt vor ($N = 40$). Berechnen Sie die durchschnittliche Zahl von Reisen pro Jahr:

1	2	3
6 (15)	29 (72.5)	5 (12.5)

Formel:

$$\bar{x} = \frac{\sum_{j=1}^k x'_j \cdot f_j}{n}$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Aufgabe Mittelwert: Eine **Häufigkeitstabelle** für die Anzahl an unternommenen Reisen pro Jahr in einer Stichprobe liegt vor ($N = 40$). Berechnen Sie die durchschnittliche Zahl von Reisen pro Jahr:

1	2	3
6 (15)	29 (72.5)	5 (12.5)

Formel:

$$\bar{x} = \frac{\sum_{j=1}^k x'_j \cdot f_j}{n}$$

Lösungsweg:

$$\bar{x} = \frac{1 \cdot 6 + 2 \cdot 29 + 3 \cdot 5}{40} = \frac{79}{40} = 1.98$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Berechnung des Mittelwerts für 2 oder mehr Datensätze (Gruppen):

- **Voraussetzungen:** es liegen für die selbe Variable Daten aus 2 Gruppen/Stichproben vor
- Die Mittelwerte der beiden Gruppen sind **bereits berechnet** worden

Berechnung des gemeinsamen Mittelwerts $\bar{\bar{x}}$:

- \bar{x}_1 und \bar{x}_2 addieren
- Mittelwerte zusätzlich mit den Stichprobengrößen der beiden Gruppen gewichten

$$\bar{\bar{x}} = \frac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2}{n_1 + n_2}$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Maße der zentralen Tendenz:

Median

- Wert, der in der Mitte der Verteilung liegt (halbiert die Verteilung)
- es liegen genau so viele Messwerte über wie unter dem Median
- Bestimmung bei ungerader Anzahl von Werten:
 - ordnen der Werte der Größe nach → mittlerer Wert = Median
 - z.B. [1, 3, 4, 6, 9] → Median = 4
- Bestimmung bei gerader Anzahl von Werten:
 - einfach: ordnen der Werte der Größe nach → numerische Mitte aus beiden mittleren Werten = Median
 - z.B. [1, 3, 4, 5, 6, 9] → Median = $(4+5) / 2 = 4.5$
- Bildet nur größer-kleiner Relationen ab → erfordert lediglich Ordinalskalenniveau

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Median - Rechenbeispiel: Median aus Variable X ($N = 10$)

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

Formel:

$$Md = \begin{cases} \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{falls } n \text{ gerade} \\ x_{(\frac{n+1}{2})} & \text{falls } n \text{ ungerade} \end{cases}$$

Lösungsweg:

$$x_{\text{sortiert}} = 8; 9; 9; 10; 10; 10; 10; 11; 11; 12$$

$$Md = \frac{x_{(\frac{10}{2})} + x_{(\frac{10}{2}+1)}}{2} = 10$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Median - Aufgabe: Berechnen Sie den Median für die Variable Monatseinkommen ($N = 8$)

ID	1	2	3	4	5	6	7	8
Einkommen	2300	5332	4272	982	1048	8261	2037	3000

Formel:

$$Md = \begin{cases} \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{falls } n \text{ gerade} \\ x_{(\frac{n+1}{2})} & \text{falls } n \text{ ungerade} \end{cases}$$

Skalenniveaus und statistische Kennwerte

Statistische Kennwerte - Maße der zentralen Tendenz:

Aufgabe Median: Berechnen Sie den Median für die Variable Monatseinkommen ($N = 8$)

ID	1	2	3	4	5	6	7	8
Einkommen	2300	5332	4272	982	1048	8261	2037	3000

Formel:

$$Md = \begin{cases} \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{falls } n \text{ gerade} \\ x_{(\frac{n+1}{2})} & \text{falls } n \text{ ungerade} \end{cases}$$

Lösungsweg:

$$x_{sortiert} = 982; 1048; 2037; 2300; 3000; 4272; 5332; 8261$$

$$Md = \frac{x_{(\frac{8}{2})} + x_{(\frac{8}{2}+1)}}{2} = 2650$$

Skalenniveaus und statistische Kennwerte

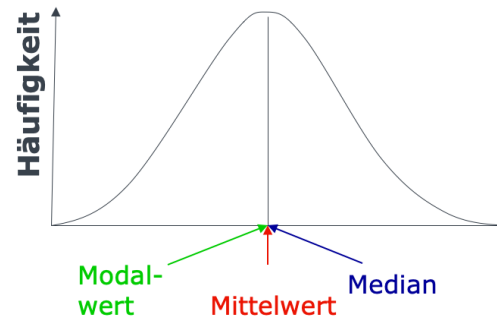
Statistische Kennwerte

Maße der zentralen Tendenz:

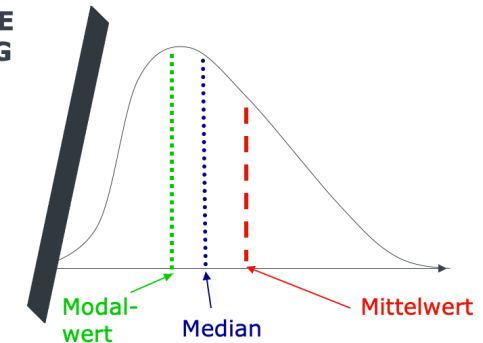
Vorteile Median gegenüber Mittelwert:

- kann auch bei rangskalierten Merkmalen verwendet werden
- Werte, die weit von allen übrigen entfernt liegen (**Ausreißer**), beeinflussen den Median kaum
- Bei schiefen Verteilungen (hohe und niedrige Werte ungleich häufig) bildet Median zentrale Tendenz besser ab

SYMMETRISCHE VERTEILUNG



LINKSSTEILE VERTEILUNG



Skalenniveaus und statistische Kennwerte

Statistische Kennwerte

Maße der zentralen Tendenz:

Vorteile Median gegenüber Mittelwert:

Beispiel für Robustheit gegenüber Extremwerten: 10. Wert der Variable X ist ein Ausreißer

ID	1	2	3	4	5	6	7	8	9	10
X	10	11	9	10	12	10	10	11	8	500

Mittelwert:

$$\bar{x} = \frac{10 + 11 + 9 + 10 + 12 + 10 + 10 + 11 + 8 + 500}{10} = \frac{591}{10} = 59.1$$

Median:

$$Md = \frac{x_{(\frac{10}{2})} + x_{(\frac{10}{2}+1)}}{2} = 10$$

Take-aways

- **Histogramm** und **Balkendiagramm** eignen sich zur Darstellung der Verteilung einer Variable.
- Wir unterscheiden 3 Arten von **Skalenniveaus**: Nominal-, Ordinal-, Intervall- und Verhältnisskalenniveau.
- Die Skalenniveaus **entscheiden**, welche Statistiken wir rechnen dürfen.
- Der **Modalwert** ist der häufigste Wert einer Verteilung und erfordert mind. Nominalskalenniveau.
- Der **Mittelwert** ist der Durchschnitt einer Verteilung und erfordert mind. Intervallskalenniveau.
- Der **Median** trennt die Verteilung in der Mitte und erfordert mind. Ordinalskalenniveau.
- Der Median ist im Gegensatz zum Mittelwert **robust** gegenüber schiefen Verteilungen und Ausreißern.

