



Kantonsschule Im Lee

Wahlpflichtmodule Maturitätsarbeit

Statistik

Skript

Dr. Stephan Schönenberger, Cyril Blum

© Winterthur, 15. März 2026

Inhaltsverzeichnis

1	Statistik: Daten verstehen und beschreiben	2
1.1	Wozu Statistik?	2
1.2	Daten und Skalenniveaus	2
1.2.1	Was sind Daten?	2
1.2.2	Skalenniveaus (Skalenwerte)	2
1.2.3	Übersicht der Skalenniveaus	3
1.3	Univariate deskriptive Statistik	4
1.3.1	Zentrale Lage	4
1.3.2	Streuung	6
1.4	Darstellungen: Tabelle, Text, Grafik	8
1.4.1	Welche Darstellung wofür?	8
1.4.2	Formen von Verteilungen	12
1.4.3	Kritisches Lesen von Grafiken	12
1.5	Arbeiten mit CODAP und Datensätzen	13
1.5.1	Datensatz bereitstellen	13
1.5.2	Datensatz 1: Schlaf und Fokus	13
1.5.3	Aufgabenpaket für die SuS	14
1.6	Mini-Checkliste: Statistik in Texten und Grafiken	16

Kapitel 1

Statistik: Daten verstehen und beschreiben

1.1 Wozu Statistik?

Statistik hilft uns, aus Daten sinnvolle Aussagen zu gewinnen, ohne uns von Einzelfällen, Ausreißern oder geschickt gestalteten Grafiken täuschen zu lassen. In diesem Kapitel arbeiten wir mit kleinen, realistischen Datensätzen und üben das Zusammenspiel von

Daten → **Kennzahlen** → **Visualisierung** → **Interpretation**.

Achtung

Wichtiger Hinweis 1.1:

Statistik ist keine Rechenakrobatik, sondern schult das **rigorose Denken**: Welche Daten liegen vor? Was darf man damit berechnen? Und was bedeutet das Ergebnis in Worten?

1.2 Daten und Skalenniveaus

1.2.1 Was sind Daten?

Daten sind beobachtete oder erhobene Werte zu einer Fragestellung. Oft stammen sie aus Umfragen, Messungen, Protokollen oder Apps (z.B. Schlaftracking).

Definition 1.1:

Eine **Variable** ist eine Eigenschaft, die bei verschiedenen Personen oder Objekten unterschiedliche Werte annehmen kann.

Beispiel 1.1:

Die Variable *Schlafdauer* kann für verschiedene Personen unterschiedliche Werte (z.B. 6.5 Stunden, 7.2 Stunden, ...) annehmen.

1.2.2 Skalenniveaus (Skalenwerte)

Das **Skalenniveau** (auch: Messniveau) bestimmt, welche Aussagen und Rechnungen sinnvoll sind.

Definition 1.2:

Ein **Skalenniveau** beschreibt, welche Struktur eine Variable besitzt (nur Kategorien? Rangordnung? Abstände? echter Nullpunkt?) und legt damit fest, welche Kennzahlen und Diagramme sinnvoll sind.

Bemerkung 1.1:

Wenn das Skalenniveau unklar ist, helfen zwei Fragen:

- Gibt es eine **Reihenfolge** (mehr/weniger)?
- Sind **Abstände** oder sogar **Verhältnisse** (doppelt so viel) sinnvoll?

Definition 1.3:

Nominalskala: Kategorien ohne Ordnung.

Beispiele: Lieblingsfach, Wohnort, Musikgenre.

Definition 1.4:

Ordinalskala: Kategorien mit Rangordnung, aber die Abstände zwischen Werten sind nicht sinnvoll interpretierbar.

Beispiele: Zufriedenheit (1–5), Schulnoten, Ranglisten.

Definition 1.5:

Intervallskala: Gleiche Abstände sind sinnvoll, aber es gibt keinen echten Nullpunkt.

Beispiel: Temperatur in °C (0°C bedeutet nicht “keine Temperatur”).

Definition 1.6:

Verhältnisskala (Ratioskala): Gleiche Abstände **und** echter Nullpunkt. **Beispiele:** Länge, Gewicht, Zeitdauer, Temperatur in Kelvin.

1.2.3 Übersicht der Skalenniveaus

Skalenniveau	Ordnung	Abstände	Nullpunkt	Lagemasse	Streuungsmaße
Nominal	✗	✗	✗	Modus	—
Ordinal	✓	✗	✗	Median, Modus	IQR
Intervall	✓	✓	✗	Mittelwert, Median	SD, IQR
Verhältnis	✓	✓	✓	Mittelwert, Median	SD, IQR

Tabelle 1.1: Übersicht der Skalenniveaus und geeignete statistische Kennzahlen

Bemerkung 1.2:

Je höher das Skalenniveau, desto mehr Rechenoperationen und statistische Verfahren sind möglich. Bei Unsicherheit: Lieber ein niedrigeres Niveau wählen (konservativer Ansatz).

 Aufgabe 1.1

Ordnen Sie den folgenden Variablen ein Skalenniveau zu (Nominal / Ordinal / Intervall / Verhältnis):

- Schlafdauer in Stunden
- Temperatur in °C
- Anzahl Nachrichten pro Tag
- Zufriedenheit (1–5)
- Lieblingsmusikgenre

 Lösungsvorschlag zu Aufgabe 1.1

- Schlafdauer: Verhältnisskala
- Temperatur in °C: Intervallskala
- Anzahl Nachrichten: Verhältnisskala
- Zufriedenheit (1–5): Ordinalskala
- Lieblingsmusikgenre: Nominalskala

1.3 Univariate deskriptive Statistik

1.3.1 Zentrale Lage

Wir beschreiben, wo “die Mitte” der Daten liegt.

Definition 1.7:

Der **Modus** ist der Wert, der in den Daten am häufigsten vorkommt.

Definition 1.8:

Der **Median** ist der mittlere Wert der sortierten Daten. Bei gerader Anzahl ist er der Mittelwert der beiden mittleren Werte.

Definition 1.9:

Der **Mittelwert** (arithmetisches Mittel) einer Datenreihe $x_1, \dots, x_n$ ist

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

 Achtung**Wichtiger Hinweis 1.2:**

Der Mittelwert reagiert stark auf Ausreisser. Der Median ist robust.

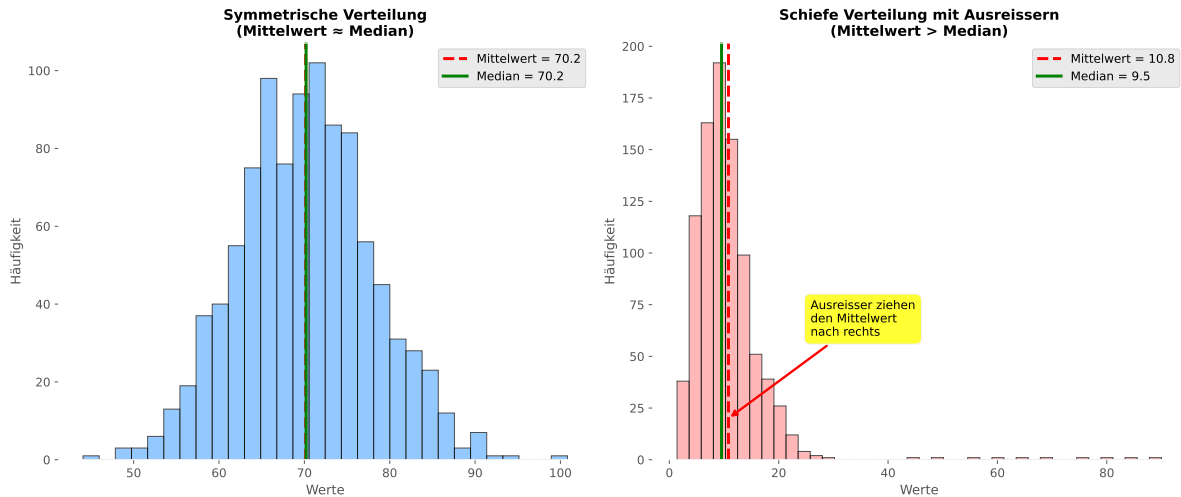


Abbildung 1.1: Vergleich: Mittelwert und Median bei symmetrischer vs. schiefer Verteilung

Aufgabe 1.2

Betrachten Sie die Daten (z.B. monatliches Taschengeld in CHF):

20, 25, 25, 30, 30, 35, 40, 400.

1. Bestimmen Sie den Median.
2. Schätzen Sie den Mittelwert (grob reicht).
3. Welche Kennzahl beschreibt das “typische” Taschengeld besser? Begründen Sie.

✓ Lösungsvorschlag zu Aufgabe 1.2

1. Sortiert sind die Werte bereits. Median ist der Mittelwert des 4. und 5. Wertes:
 $(30 + 30)/2 = 30$.
2. Mittelwert: Summe = 605, geteilt durch 8 ergibt ca. 75.6.
3. Median, weil der Ausreisser 400 den Mittelwert stark nach oben zieht.

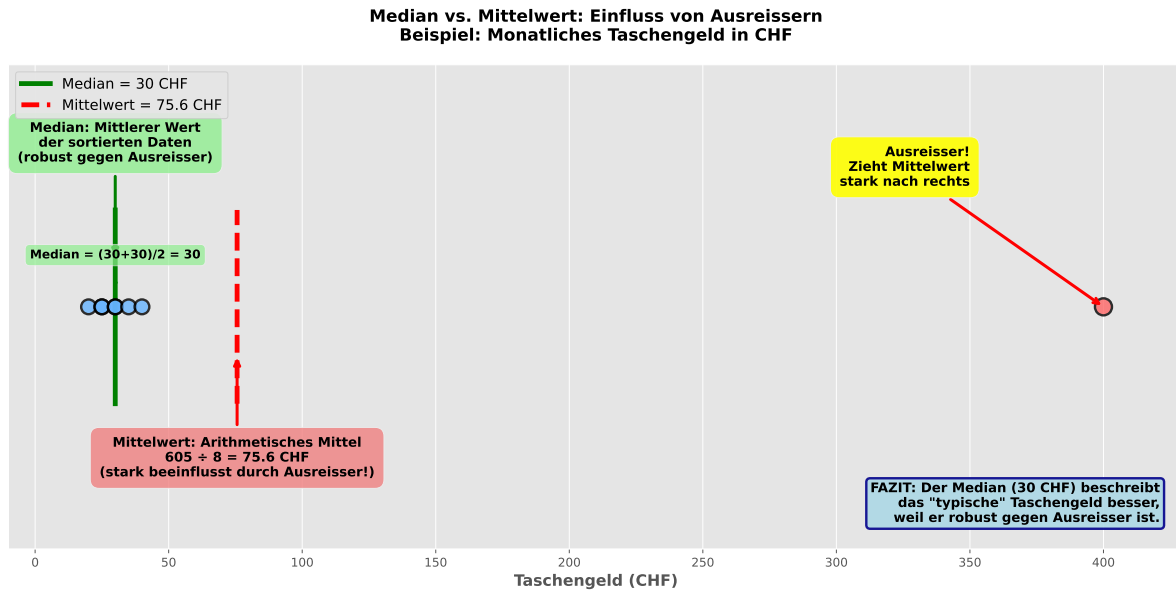


Abbildung 1.2: Visualisierung der Median-Berechnung und des Ausreisser-Effekts

1.3.2 Streuung

Streuung beschreibt, wie stark Daten um die zentrale Lage variieren.

Definition 1.10:

Die **Spannweite** ist $\max(x) - \min(x)$. Diese Metrik ist sehr anfällig für Ausreisser, da sie nur die Extremwerte betrachtet.

Definition 1.11:

Quartile: Q_1 ist das 25%-Quantil, Q_3 das 75%-Quantil. Quartile teilen die Daten in vier gleich grosse Teile. Damit dies möglich ist, müssen die Daten sortiert werden. Damit Daten überhaupt sortierbar sind, müssen sie mindestens auf einer Ordinalskala vorliegen. Der **Interquartilsabstand (IQR)** ist $IQR = Q_3 - Q_1$. Dieser wird häufig verwendet, um die Streuung zu beschreiben, da er robust gegenüber Ausreissern ist.

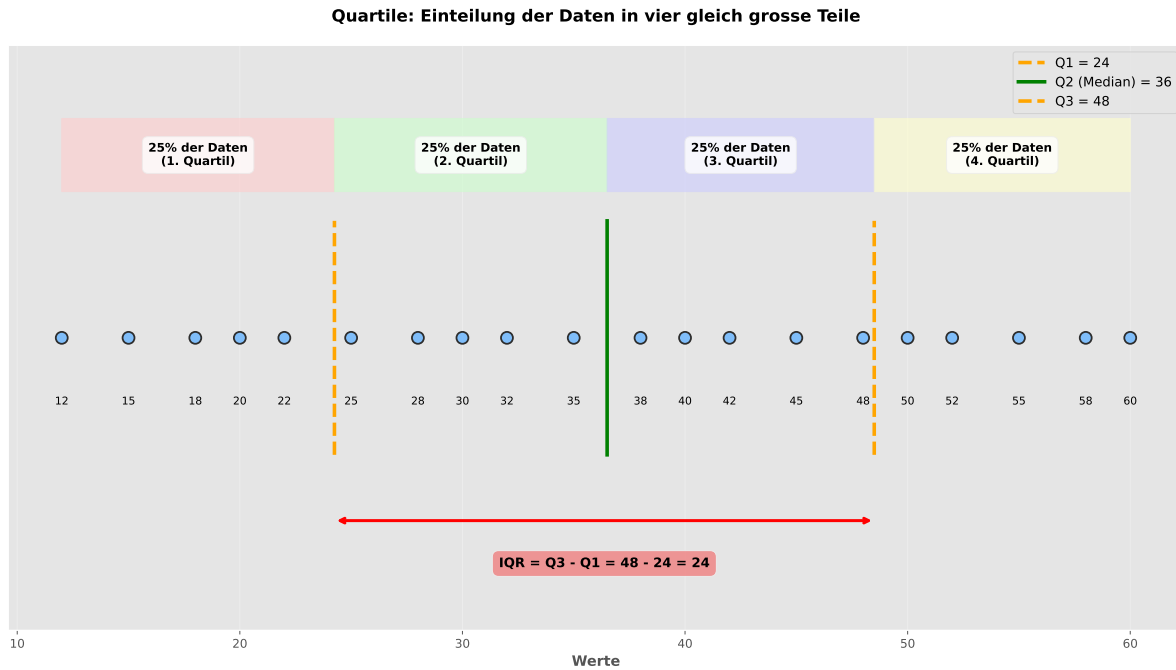


Abbildung 1.3: Quartile teilen die sortierten Daten in vier gleich grosse Teile

Definition 1.12:

Die **Standardabweichung** (Stichprobe) ist

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

Wobei $\bar{x}$ der Mittelwert der Daten ist. Die Standardabweichung misst also die durchschnittliche Abweichung der Daten vom Mittelwert. Sie ist jedoch empfindlich gegenüber Ausreissern, da diese den Mittelwert stark beeinflussen können.

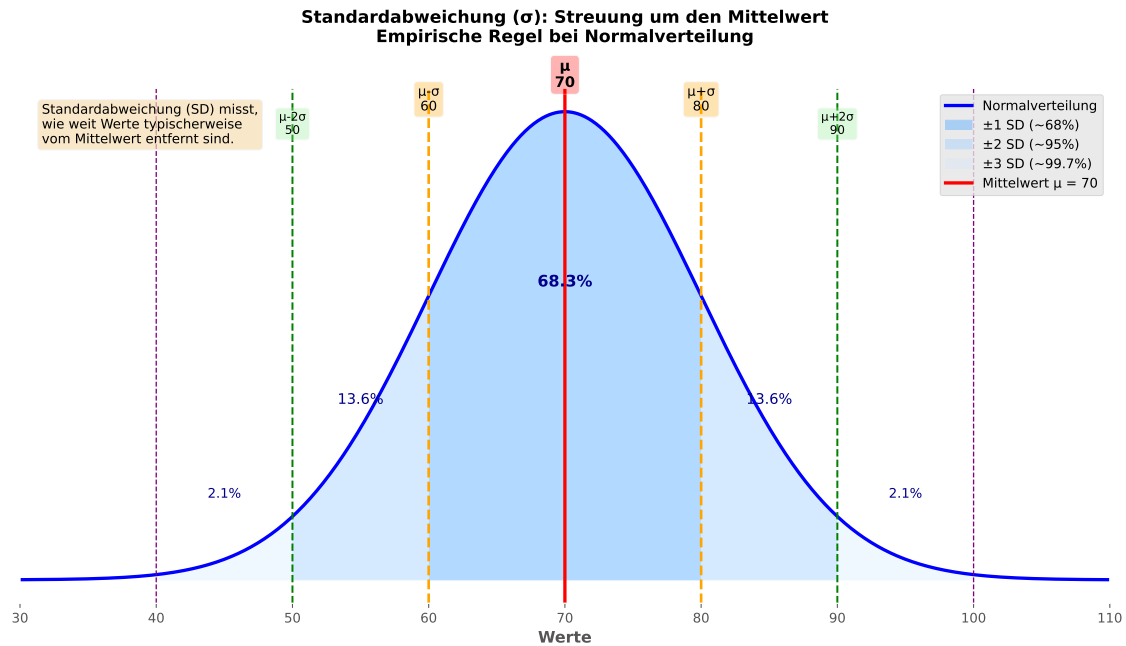


Abbildung 1.4: Standardabweichung: Streuung um den Mittelwert bei Normalverteilung

Bemerkung 1.3:

Typische Paarung:

- **Median** zusammen mit **IQR**
- **Mittelwert** zusammen mit **Standardabweichung**

1.4 Darstellungen: Tabelle, Text, Grafik

1.4.1 Welche Darstellung wofür?

Definition 1.13 (Text und Tabellen):

Text und Tabellen sind ideal, um genaue Werte zu kommunizieren (z.B. Mittelwert = 7.2 Stunden) oder um kleine Stichproben übersichtlich darzustellen.

Beispiel: Ein Artikel berichtet: “Die durchschnittliche Schlafdauer beträgt 7.2 Stunden (SD = 1.5).” Hier werden genaue Werte kommuniziert, die in Textform gut verständlich sind.

Ein anderes Beispiel: Eine Tabelle zeigt die Schlafdauer von 3 Personen. Hier ist die tabellarische Darstellung übersichtlicher als eine Grafik, da es sich um wenige Datenpunkte handelt.

Tabelle 1.2: Beispiel einer Tabelle mit Schlafdaten von 3 Personen

Person	Schlafdauer (Stunden)
A	7.0
B	6.5
C	8.0

Definition 1.14 (Balken- und Kreisdiagramme):

Balken- und Kreisdiagramme sind geeignet, um Verhältnisse zwischen Kategorien darzustellen. Sie eignen sich besonders gut, wenn die Kategorien nominal skaliert sind.

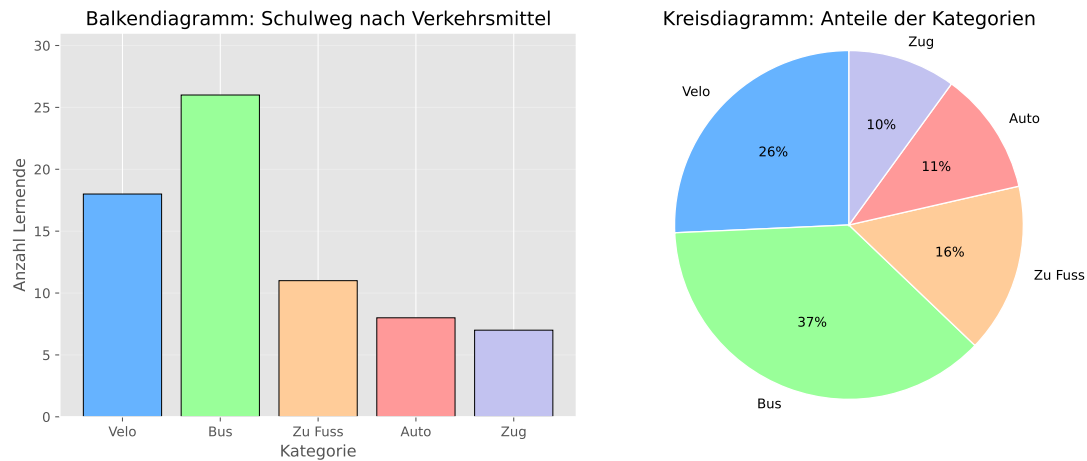


Abbildung 1.5: Beispiel eines Balkendiagramms und eines Kreisdiagramms

Balkendiagramme können auch für ordinal oder intervallskalierte Daten verwendet werden und eignen sich besser, wenn die Kategorien viele Werte haben oder wenn genaue Werte kommuniziert werden sollen. Kreisdiagramme sollten nur bei wenigen Kategorien verwendet werden, da sie bei vielen Kategorien unübersichtlich werden können.

Definition 1.15 (Scatterplot):

Ein **Scatterplot** zeigt die Beziehung zwischen zwei metrischen Variablen (z.B. Schlafdauer und Fokus-Score), indem jeder Datenpunkt als Punkt in einem Koordinatensystem dargestellt wird. Er eignet sich besonders gut, um Korrelationen oder Muster in den Daten zu erkennen.

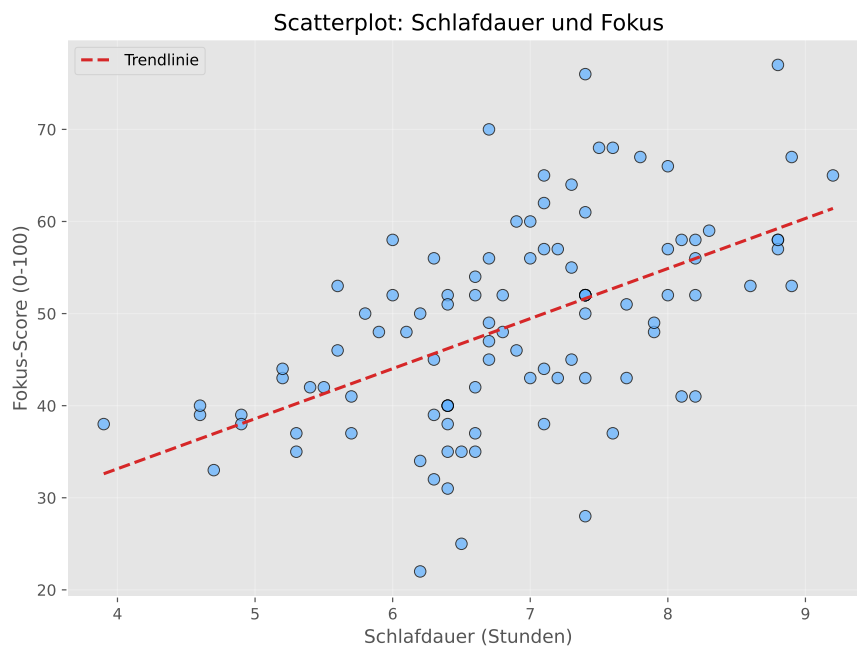


Abbildung 1.6: Beispiel eines Scatterplots: Schlafdauer und Fokus-Score

Definition 1.16 (Histogramm):

Ein **Histogramm** zeigt die Verteilung einer metrischen Variable (z.B. Schlafdauer), indem es die Häufigkeit (oder relative Häufigkeit) von Werten in bestimmten Intervallen (Bins) darstellt. Es eignet sich besonders gut, um die Form der Verteilung zu erkennen (z.B. Schiefe, Modus, Ausreisser).

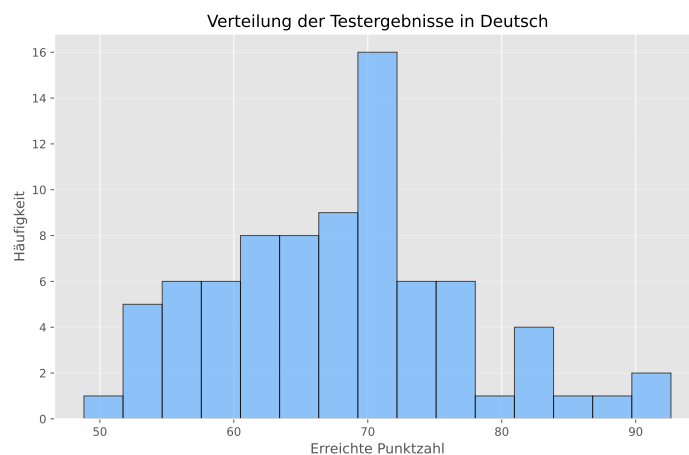


Abbildung 1.7: Beispiel eines Histogramms

Definition 1.17 (Boxplot):

Ein **Boxplot** (Kastendiagramm) visualisiert die Verteilung einer metrischen Variable durch:

- Eine Box von Q_1 bis Q_3 (Interquartilsabstand)
- Eine Linie im Inneren der Box für den Median

- “Whiskers” (Antennen) bis zum letzten Wert innerhalb von $1.5 \times IQR$ von den Quartilen
- Punkte für Ausreisser, die ausserhalb der Whiskers liegen

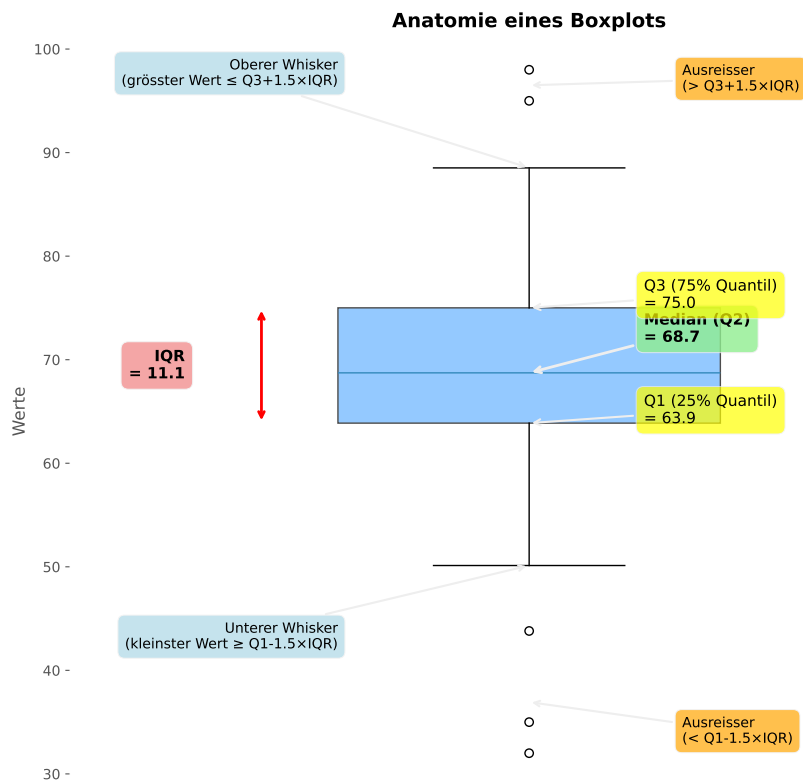


Abbildung 1.8: Anatomie eines Boxplots: Alle Komponenten erklärt

1.4.2 Formen von Verteilungen

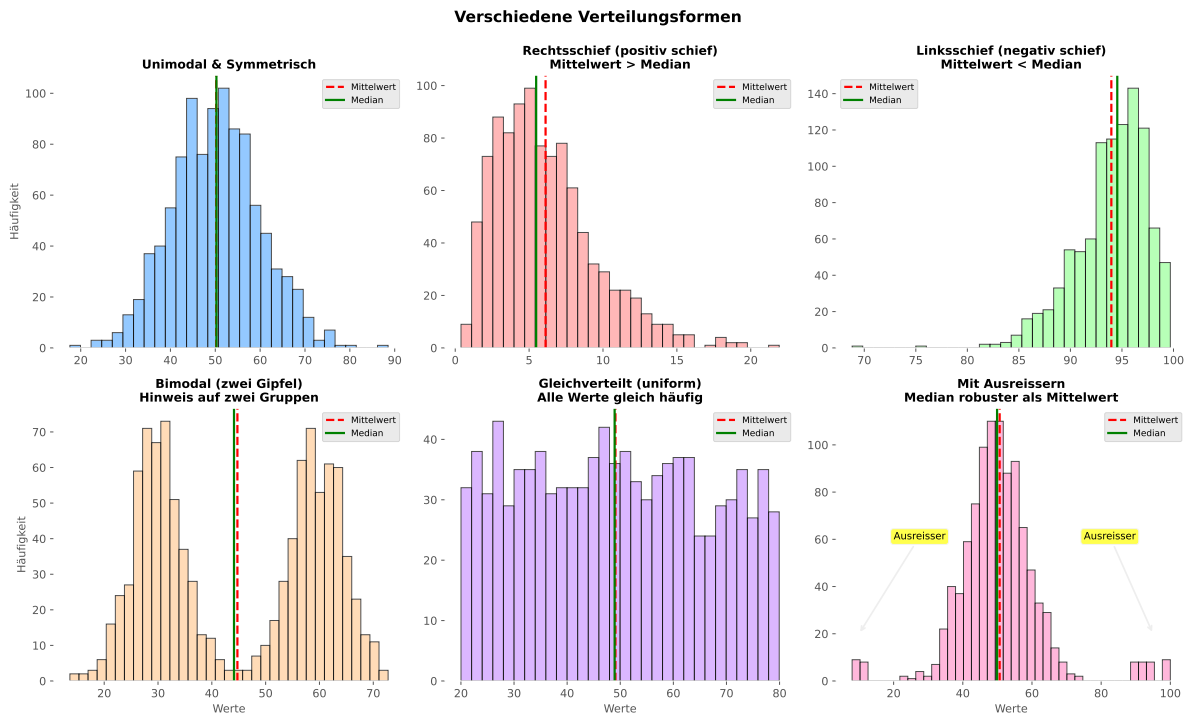


Abbildung 1.9: Verschiedene Verteilungsformen und ihre Eigenschaften

Bemerkung 1.4:

Die Form einer Verteilung gibt wichtige Hinweise:

- **Symmetrisch:** Mittelwert $\approx$ Median
- **Rechtsschief:** Wenige hohe Ausreisser, Mittelwert $>$ Median
- **Linksschief:** Wenige niedrige Ausreisser, Mittelwert $<$ Median
- **Bimodal:** Hinweis auf zwei unterschiedliche Gruppen in den Daten

1.4.3 Kritisches Lesen von Grafiken

! Achtung

Wichtiger Hinweis 1.3:

Beim Lesen von Grafiken immer prüfen:

- Startet die Achse bei 0 (Gefahr von Verzerrung)?
- Sind Kategorien gleich breit (Gefahr von visueller Überinterpretation einzelner Kategorien)?
- Fehlen Werte oder Gruppen (Gefahr von inhaltlichen Interpretationen)?

Aufgabe 1.3

Eine Zeitung zeigt ein Balkendiagramm: Die y-Achse startet bei 60 statt bei 0.

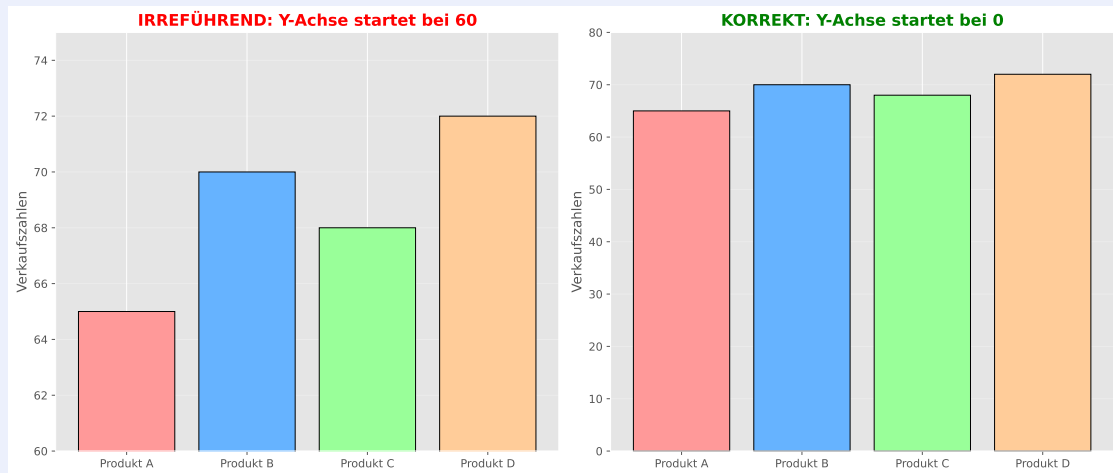


Abbildung 1.10: Vergleich: Irreführende vs. korrekte Darstellung der Y-Achse

1. Welchen Effekt hat das auf die Wirkung der Grafik?
2. Formulieren Sie eine sachliche Kritik in einem Satz.

✓ Lösungsvorschlag zu Aufgabe 1.3

1. Unterschiede wirken viel grösser und dramatischer als sie tatsächlich sind.
2. "Da die y-Achse nicht bei 0 startet, werden die Unterschiede optisch übertrieben dargestellt."

1.5 Arbeiten mit CODAP und Datensätzen

1.5.1 Datensatz bereitstellen

Sie erhalten zwei Datensätze als CSV-Dateien, die Sie (oder Ihre Lehrperson) bereitstellen:

- `sleep_focus.csv` (Schlaf, Smartphone, Stress, Fokus)
- `music_study.csv` (optional: Kontingenztabelle, siehe Challenge)

⚠ Achtung

Wichtiger Hinweis 1.4:

Import in CODAP: codap.concord.org → Oben links → Öffnen → Lokale Datei.

1.5.2 Datensatz 1: Schlaf und Fokus

Der Datensatz `sleep_focus.csv` enthält pro Person:

- `SleepHours` (Stunden Schlaf)
- `PhoneMinutes` (Smartphone-Minuten)
- `Coffee` (0–4, ordinal)
- `Stress` (1–5, ordinal)

- **FocusScore** (Punktzahl 0–100)

Tabelle 1.3: Beispiel: Erste Zeilen des Datensatzes `sleep_focus.csv`

Person	SleepHours	PhoneMinutes	Coffee	Stress	FocusScore
1	7.5	120	2	3	78
2	6.0	200	3	4	62
3	8.2	45	1	2	85
4	5.5	310	4	5	54
5	7.0	150	2	3	72
...	...	...	...	...	...

1.5.3 Aufgabenpaket für die SuS

Aufgabe 1.4

Teil A: Kennzahlen (Einzelarbeit)

Arbeiten Sie mit `sleep_focus.csv`.

1. Bestimmen Sie für **SleepHours** Median, Mittelwert, Minimum und Maximum.
2. Entscheiden Sie: Würden Sie für **PhoneMinutes** eher Median oder Mittelwert berichten? Begründen Sie.
3. Nennen Sie eine Situation, in der der Mittelwert sinnvoll ist, und eine, in der der Median sinnvoller ist.

✓ Lösungsvorschlag zu Aufgabe 1.4

1. Werte in CODAP ablesen (je nach Anzeige). Wichtig ist die Interpretation: typische Schlafdauer liegt im Bereich um 7 Stunden; Minimum/Maximum zeigen Spannweite.
2. Eher Median, weil einzelne sehr hohe Nutzungszeiten den Mittelwert stark beeinflussen können.
3. Mittelwert: symmetrische, ausreisserarme Daten (z.B. Körpergrößen). Median: schiefe Verteilungen oder Ausreisser (z.B. Einkommen).

Aufgabe 1.5

Teil B: Visualisierung und Interpretation (Partnerarbeit)

1. Erstellen Sie ein Histogramm von **SleepHours**. Beschreiben Sie die Verteilung (unimodal? schief? Ausreisser?).
2. Erstellen Sie einen Boxplot von **PhoneMinutes**. Notieren Sie Q1, Median und Q3 (oder lesen Sie sie ab).
3. Erstellen Sie einen Scatterplot **SleepHours** vs **FocusScore**. Formulieren Sie eine vorsichtige Aussage in einem Satz.

✓ Lösungsvorschlag zu Aufgabe 1.5

1. Unimodal, die meisten Werte zwischen etwa 6 und 8 Stunden, einzelne kurze Nächte als Randwerte.
2. Q1/Median/Q3 je nach CODAP-Ablesung; wichtig: IQR beschreibt typische Streuung ohne Ausreisser.
3. "Im Datensatz zeigt sich tendenziell: Wer mehr schläft, hat häufig einen höheren Fokus-Score; der Zusammenhang ist aber nicht perfekt."

Beispiel 1.2:

Beispiel einer Kontingenztabelle: Musik beim Lernen vs. Lernmodus

	Alleine	Gruppe	Summe
Mit Musik	45 (56%)	25 (31%)	70
Ohne Musik	35 (44%)	55 (69%)	90
Summe	80	80	160

Tabelle 1.4: Beispiel einer Kontingenztabelle: Musik beim Lernen vs. Lernmodus

Aus dieser Tabelle darf man z.B. beschreiben, welche Kombinationen in der Stichprobe häufiger vorkommen. Man darf daraus aber noch keine Ursache-Wirkung-Aussage ableiten.

🏆 Aufgabe (Challenge) 1.6

Challenge (optional): Kontingenztabelle

Arbeiten Sie mit der Kontingenztabelle aus dem Beispiel oben (ohne Computer).

1. Formulieren Sie zwei Aussagen, die durch die Tabelle gestützt werden.
2. Formulieren Sie eine Aussage, die **nicht** zulässig ist, und begründen Sie kurz, warum.

✓ Lösungsvorschlag zu Aufgabe 1.6

Zulässig sind beschreibende Aussagen zur Stichprobe, z.B. "In dieser Stichprobe lernen Personen ohne Musik häufiger in Gruppen als Personen mit Musik.". Eine weitere zulässige Aussage könnte sein: "Insgesamt lernen 70 Personen mit Musik, 90 Personen ohne Musik."

Nicht zulässig sind kausale Aussagen wie "Musik verbessert das Lernen."

1.6 Mini-Checkliste: Statistik in Texten und Grafiken

Aufgabe 1.7

Finden Sie einen Online-Artikel, der eine Statistik-Grafik enthält. Überprüfen Sie die Grafik kritisch anhand der folgenden Checkliste:

- ☐ Woher stammen die Daten (Quelle)?
- ☐ Wie gross ist die Stichprobe?
- ☐ Wird Median oder Mittelwert verwendet und passt das?
- ☐ Startet die Achse bei 0?
- ☐ Werden Prozente mit absoluten Zahlen verwechselt?

Statistik Cheatsheet

Skalenniveaus

Nominalskala

Kategorien ohne Ordnung.

Beispiele: 1. Geschlecht 2. Nationalität
3. Lieblingsmusik

Sinnvolle Kennzahlen: 1. Häufigkeit
2. Modus

Ordinalskala

Kategorien mit Rangordnung.

Beispiele: 1. Schulnoten 2. Ranglisten
3. Zufriedenheitsskala

Sinnvolle Kennzahlen: 1. Median 2. Quartile

Intervallskala

Gleiche Abstände, aber kein echter Nullpunkt.

Beispiel: Temperatur in °C

Erlaubt: 1. Mittelwert 2. Standardabweichung

Verhältnisskala

Numerische Werte mit echtem Nullpunkt.

Beispiele: 1. Gewicht 2. Länge 3. Einkommen

Verhältnisse sind sinnvoll.

Lagekennzahlen

Minimum und Maximum

Kleinster und grösster Wert im Datensatz.

Spannweite:

$$\text{Spannweite} = \text{Max} - \text{Min}$$

Modus

Häufigster Wert im Datensatz.

Geeignet für alle Skalenniveaus, besonders Nominalskala.

Beispiel: 1, 2, 2, 3, 4 Modus = 2

Mittelwert

Durchschnitt aller Werte.

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

Empfindlich gegenüber Ausreissern.

Median

Mittlerer Wert nach Sortierung.

Robust gegenüber Ausreissern.

Beispiel:

1, 2, 3, 4, 100

Median = 3

Streuungsmaße

Quartile

Datensatz wird in vier Teile geteilt.

- $Q1 = 25\%$
- Median = 50%
- $Q3 = 75\%$

Quartilsabstand

Interquartilsabstand:

$$IQR = Q3 - Q1$$

Robust gegenüber Ausreissern.

Standardabweichung

Misst Streuung um den Mittelwert.

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

Empirische Regel:

- ± 1 SD: ca. 68%
- ± 2 SD: ca. 95%
- ± 3 SD: ca. 99.7%

Ausreisser und Verteilungsformen

Ausreisser

Wert, der deutlich vom Rest der Daten abweicht.

Faustregel (Boxplot):

- $< Q1 - 1.5 \cdot IQR$
- $> Q3 + 1.5 \cdot IQR$

Behandlung: 1. Median statt Mittelwert
2. IQR statt SD

Verteilungsformen

Symmetrisch (unimodal): Mittelwert
Median

Rechtsschief: Wenige hohe Ausreisser,
Mittelwert > Median

Linksschief: Wenige niedrige Ausreisser,
Mittelwert < Median

Bimodal: Zwei Häufigkeitsgipfel → Hinweis auf zwei Gruppen

Darstellungen

Histogramm

Zeigt Verteilung numerischer Daten.

Gut für:

- Form der Verteilung
- Schiefe
- Ausreisser

Boxplot

Zeigt:

- Median
- Quartile
- Ausreisser

Sehr kompakte Darstellung.

Scatterplot

Zeigt Zusammenhang zwischen zwei metrischen Variablen.

Gut zum Erkennen von:

- Positiven Korrelationen
- Negativen Korrelationen
- Ausreißern