



Vorlesung Statistik I - Kapitel 2: Deskriptive Statistik

(Beispielskript, B.Sc. Psychologie, 1. Semester)

2.1 Merkmale und Skalenniveaus

Eine Untersuchungseinheit (z. B. eine Versuchsperson) besitzt Merkmale (Variablen), die verschiedene Ausprägungen annehmen können. Welche Rechenoperationen sinnvoll sind, hängt vom Skalenniveau ab:

- Nominalskala: Ausprägungen lassen sich nur unterscheiden, nicht ordnen (Beispiele: Geschlecht, Studienfach, Haarfarbe). Sinnvoll: Häufigkeiten, Modus.
- Ordinalskala: Ausprägungen lassen sich in eine Rangfolge bringen, die Abstände sind aber nicht interpretierbar (Beispiele: Schulnoten, Rangplätze, Zustimmung „gar nicht – etwas – stark“). Sinnvoll zusätzlich: Median, Quantile.
- Intervallskala: gleiche Abstände bedeuten gleiche Unterschiede, es gibt aber keinen natürlichen Nullpunkt (Beispiele: Temperatur in °C, IQ-Werte). Sinnvoll zusätzlich: arithmetisches Mittel, Varianz.
- Verhältnisskala (Ratioskala): zusätzlich ein natürlicher Nullpunkt, Verhältnisse sind interpretierbar (Beispiele: Reaktionszeit in ms, Körpergröße, Einkommen). „Doppelt so lang“ ist hier eine sinnvolle Aussage.

Intervall- und Verhältnisskala werden zusammen als metrische Skalen (Kardinalskalen) bezeichnet.

2.2 Häufigkeiten

Die absolute Häufigkeit $h(a)$ gibt an, wie oft die Ausprägung a vorkommt. Die relative Häufigkeit ist $f(a) = h(a) / n$, wobei n der Stichprobenumfang ist. Die Summe aller relativen Häufigkeiten ergibt 1. Ab ordinalem Skalenniveau kann man kumulierte Häufigkeiten bilden: $F(a)$ ist der Anteil der Beobachtungen, die kleiner oder gleich a sind.

2.3 Lagemaße

Lagemaße beschreiben, wo das „Zentrum“ einer Verteilung liegt.

- Modus (Modalwert): die am häufigsten vorkommende Ausprägung. Ab Nominalskala zulässig.
- Median: der Wert, der die geordnete Datenreihe in zwei gleich große Hälften teilt. Bei ungeradem n ist es der mittlere Wert, bei geradem n der Mittelwert der beiden mittleren Werte. Ab Ordinalskala zulässig. Der Median ist robust gegenüber Ausreißern.
- Arithmetisches Mittel: $\bar{x} = (x_1 + x_2 + \dots + x_n) / n$. Ab Intervallskala zulässig. Das Mittel reagiert empfindlich auf Ausreißer.

Beispiel: Reaktionszeiten in ms: 310, 290, 350, 300, 900. Geordnet: 290, 300, 310, 350, 900. Median = 310 ms. Mittelwert = $2150 / 5 = 430$ ms. Der Ausreißer 900 ms zieht den Mittelwert stark nach oben, den Median kaum.

Bei einer rechtsschiefen (linkssteilen) Verteilung gilt typischerweise $\text{Modus} < \text{Median} < \text{Mittelwert}$.

2.4 Streuungsmaße





Streuungsmaße beschreiben, wie stark die Werte um das Zentrum schwanken.

- Spannweite (Range): $R = \text{Maximum} - \text{Minimum}$. Sehr anfällig für Ausreißer.
- Quartile und Interquartilsabstand: Das erste Quartil $Q1$ trennt die unteren 25 % der Werte ab, das dritte Quartil $Q3$ die unteren 75 %. Der Interquartilsabstand $IQR = Q3 - Q1$ umfasst die mittleren 50 % der Werte und ist robust gegenüber Ausreißern. Ab Ordinalskala sinnvoll.
- Varianz: mittlere quadrierte Abweichung vom Mittelwert. In der deskriptiven Statistik: $s^2 = \sum (x_i - \bar{x})^2 / n$. Als Schätzer für die Populationsvarianz wird durch $n - 1$ geteilt (korrigierte Stichprobenvarianz). Welche Variante gemeint ist, muss immer angegeben werden.
- Standardabweichung: $s = \sqrt{s^2}$. Sie hat dieselbe Einheit wie die Daten und ist daher leichter zu interpretieren als die Varianz.

Beispiel: Werte 2, 4, 6. Mittelwert = 4. Quadrierte Abweichungen: 4, 0, 4, Summe 8. Varianz (Division durch n) = $8/3 \approx 2,67$; Standardabweichung $\approx 1,63$. Mit Division durch $n - 1$: Varianz = 4, Standardabweichung = 2.

2.5 z-Standardisierung

Um Werte aus Verteilungen mit unterschiedlichem Mittelwert und unterschiedlicher Streuung vergleichbar zu machen, werden sie z-standardisiert: $z = (x - \bar{x}) / s$. Ein z-Wert gibt an, um wie viele Standardabweichungen ein Wert über ($z > 0$) oder unter ($z < 0$) dem Mittelwert liegt. Die z-Werte einer Stichprobe haben den Mittelwert 0 und die Standardabweichung 1.

Beispiel: Anna erreicht in Test A 60 Punkte (Mittelwert 50, $s = 5$), in Test B 75 Punkte (Mittelwert 70, $s = 10$). $z_A = (60 - 50)/5 = 2$, $z_B = (75 - 70)/10 = 0,5$. Relativ zur Vergleichsgruppe war Anna in Test A deutlich besser.

2.6 Grafische Darstellung

Für nominale und ordinale Daten eignen sich Balken- und Kreisdiagramme. Metrische Daten werden mit Histogrammen dargestellt; beim Histogramm entspricht die Fläche eines Balkens der Häufigkeit der Klasse. Der Boxplot zeigt Median, Quartile, Whisker und Ausreißer auf einen Blick und eignet sich gut zum Vergleich mehrerer Gruppen.

Merke für die Klausur: Erst das Skalenniveau bestimmen, dann das passende Lage- und Streuungsmaß wählen. Bei Ausreißern Median und IQR berichten, bei symmetrischen metrischen Daten Mittelwert und Standardabweichung.

