

Zahlenstrahl und Skalenniveau

Wilfried Mann,
Mettmann

Die einfachste Beurteilung einer Messreihe erfolgt auf einem Zahlenstrahl bspw. mit Hilfe eines Boxplots.

Das Skalenniveau gehört zu den wichtigen Eigenschaften von Merkmalen und ist bei der Durchführung von mathematisch-statistischen Analysen zu beachten.

1 Zahlenstrahl

Der Begriff Zahlenstrahl umfasst alle geordneten, positiven Zahlen, kleinstmöglich ist die Null. Dagegen liegen auf einer Zahlengerade auch negative Werte. Die Verteilung einer Messreihe kann mit einem **Boxplot** oder Box-Whisker-Plot zur Veranschaulichung von Lage- und Streuungsmaßen metrischer Daten graphisch dargestellt werden. Box und Whisker (Antenne) verdeutlichen folgende Kenngrößen:

- Unteres Quartil (Q1, 25 %-Quartil) mit Kennwert $\leq 25\%$ der Datenwerte, als Beginn der Box.
 - Oberes Quartil (Q3, 75 %-Quartil) mit Kennwert $\leq 75\%$ der Datenwerte, als Ende der Box.
 - Median (Q2, 50 %-Quartil).
 - Arithmetisches Mittel (MEAN); diese Kenngröße ist nicht in jeder Darstellung vorhanden.
 - Interquartilsrange/-abstand (IQR); Wertebereich in dem 50 % der Daten liegen, zwischen Q3 und Q1.
 - Unterer Whisker bei 2,5 %-Quartil, als Antennenanfang.
 - Oberer Whisker bei 97,5 %-Quartil, als Antennenende.
 - Spannweite (RANGE), Differenz zwischen größtem und kleinstem Wert (MAX – MIN).
 - Ausreißer: milde Ausreißer zwischen $\pm 1,5 \times \text{IQR}$ und $3,0 \times \text{IQR}$, darüber extreme Ausreißer.
- Diese Darstellung entspricht nicht der Identifikation von Ausreißern durch die Standardabweichung (s). Bei normalverteilten, großen Stichproben liegen extreme Ausreißer, dargestellt im Boxplot, aber i.d.R. über der 3-s-Grenze.

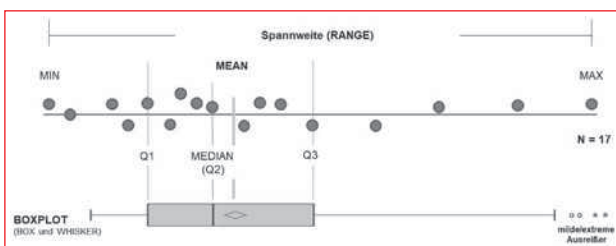


Abb. 1: Zahlenstrahl und Boxplot (typische Darstellung)

2 Skalenniveau

Je nach der Art eines Merkmals lassen sich verschiedene Stufen der Skalierbarkeit unterscheiden und systematisieren.

Kategorial skalierte Variable, auch gruppiert, diskret, qualitativ genannt, werden weiter gegliedert in:

- Nominalskaliert: Klassifikation und Identifikation (Beispiel Eckgrundstück: Ja oder Nein).
- Ordinalskaliert: Ordnung nach ihrem Rang, aber ohne Kenntnis zum Ausmaß der Unterschiede (Beispiel Wohnlagen: sehr gut bis einfach).

Kardinal skalierte Variable, auch kontinuierlich, metrisch, numerisch, stetig oder quantitativ genannt, werden weiter gegliedert in:

- Intervallskaliert: Reelle Zahlen in einem Intervall, diese können auch kategorial skaliert sein (Beispiel Datum: Jahrgänge als diskret, Dezimaljahr aus Jahr, Monat und Tag auch als stetig skalierbar).
- Verhältnisskaliert: Kann eine unendliche Anzahl von Werten beinhalten, auch stetig skaliert (Beispiele: Kaufpreise, Wohnflächen usw.).

3 Anwendung

Boxplots sind aufgrund des einfachen Aufbaus besonders zum schnellen Überblick und direkten Vergleich mehrerer metrischer Datensätze geeignet. Die Abbildung zeigt beispielhaft Sachwertfaktoren im Städtevergleich mit milden Ausreißern.

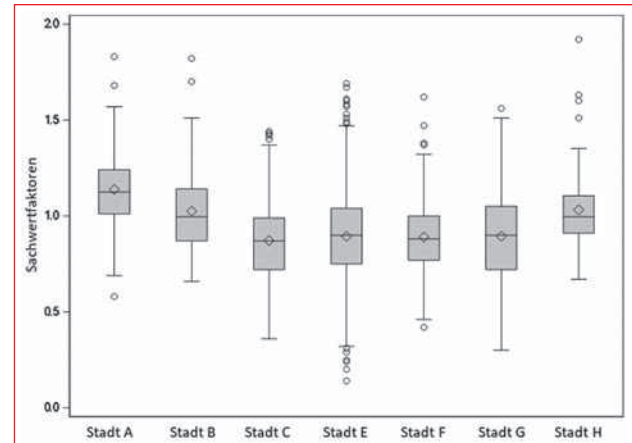


Abb. 2

An der Lage des Median innerhalb einer Box kann man erkennen, ob eine Verteilung symmetrisch oder schief ist.

Auch können eventuelle Ausreißer identifiziert werden oder Hinweise darauf geben, ob die Daten einer bestimmten Verteilung unterliegen. Wenn der Boxplot stark asymmetrisch ist und eine ungewöhnlich hohe Zahl von Ausreißern enthält, deutet das darauf hin, dass die Daten nicht normalverteilt sind.

Das **Skalenniveau** bestimmt die (mathematischen) Operationen, die mit einer entsprechend skalierten Variable zulässig sind. In der Regel gliedern Statistik-Programme die Variablen nach »metrische« und »nicht-metrische«. Besonderes Augenmerk ist hierbei auf intervallskalierte Variable zu legen, da diese ggf. beiden Haupttypen zugeordnet werden können.

Die mathematische Statistik hat für »metrische« und »nicht-metrische« Zufallsvariablen verschiedene »Verteilungsmodelle« entwickelt, die man für die statistische Analyse realer Phänomene benutzen kann.

Kardinal skalierte (metrische) Variablen führen in der Darstellung zu Kurven mit funktionalen Zusammenhängen. Kategorial skalierte (gruppierte) Variablen zu Häufigkeitstabellen.