

# Fachbegriffe

**Merkmalsträger** Alle Schülerinnen und Schüler, die an der JIM -Studie teilgenommen haben, fassen wir als Merkmalsträger auf.

**Fall** Jede einzelne Zeile in der Tabelle entspricht einer Schülerin oder einem Schüler, der/die an der Umfrage teilgenommen hat. Man sagt dazu auch Fall. Hier entspricht ein Merkmalsträger einem Fall.

**Umfang** Die Anzahl aller Fälle bezeichnet den Umfang des Datensatzes, dieser wird häufig mit **n** beschrieben.

**Merkmal/Variable** In Merkmalen werden die Angaben der Befragten notiert. Bei einer Umfrage steht ein Merkmalsname symbolhaft für die dahinterstehende Frage. Man unterscheidet kategoriale und numerische Merkmale. Merkmale sind zum Beispiel Geschlecht, Alter, Laptop\_nutzen, etc.

**Merkmalsausprägung/Ausprägung** Die Werte, die eine statistische Variable annehmen kann, nennt man Merkmalsausprägung.

**Absolute Häufigkeit** Mit der absoluten Häufigkeit wird angegeben, wie oft eine Merkmalsausprägung im Datensatz vorkommt.

**Relative Häufigkeit** Mit der relativen Häufigkeit wird der Anteil der absoluten Häufigkeit einer Merkmalsausprägung an der Gesamtanzahl **n** der Merkmalsträger angegeben (wobei nur diejenigen berücksichtigt werden, die eine Angabe zu einem Merkmal gemacht haben). Die relative Häufigkeit kann als Kommazahl (zwischen 0 und 1) oder als Prozentzahl (zwischen 0% und 100%) notiert werden.

**Kategoriales Merkmal** Bei einem kategorialen Merkmal lassen sich die Ausprägungen (=Antworten der Befragten) in bestimmte Kategorien einteilen. Zum Beispiel bei dem Merkmal *Geschlecht* liegt eine einfache Form eines kategorialen Merkmals vor, denn es hat nur zwei mögliche Ausprägungen im Datensatz: männlich und weiblich. Kategoriale Merkmale können auch viel mehr verschiedene Ausprägungen haben, zum Beispiel hat das Merkmal Laptop\_nutzen sieben Ausprägungen, die angeben, wie häufig der Laptop durch den Merkmalsträger genutzt wird: täglich, mehrmals pro Woche, einmal pro Woche, einmal in 14 Tagen, mehrmals im Monat, seltener, nie.

**Prozent** Da häufig Gruppen miteinander verglichen werden, ist es wichtig zu unterscheiden, wie groß die jeweiligen Gruppen sind (z. B. Anzahl der Jungen und Anzahl der Mädchen). Wenn die Gruppen unterschiedlich groß sind, sollte anhand von Prozenten Vergleiche gemacht werden.

**Zusammenfassen** Wenn ein Merkmal mehrere Ausprägungen hat, kann es hilfreich sein, diese zunächst auf zwei Gruppen zu reduzieren beziehungsweise mehrere Ausprägungen zusammenzufassen, so dass zwei Gruppen entstehen. Hier gibt es jedoch kein richtig und falsch. Man könnte zum Beispiel alle, die angeben, niemals Online-Zeitschriften zu lesen, in die Gruppe der „Nie-Leser“ ordnen und alle anderen in die „Überhaupt-Leser“. Anders könnte man aber auch alle, die zwischen täglich und mindestens einmal pro Woche Online-Zeitschriften lesen als „Viel-Leser“ gruppieren und alle anderen als „Wenig-Leser“. Auch andere Gruppenbildungen sind möglich. Gruppierungen sollten immer (zum Beispiel inhaltlich) begründet werden.

**Numerisches Merkmal** Bei einem numerischen Merkmal sind die Ausprägungen Zahlen (und nichts anderes!), lassen sich also insbesondere der Größe nach ordnen. Beispiele für numerische Merkmale sind zum Beispiel Körpergröße oder Anzahl\_Accounts.

**Spannweite** Die Differenz vom größten Wert (Maximum) zum kleinsten Wert (Minimum) einer Verteilung nennt man Spannweite.

**Maximum** Ein Maximum kann man nur bei einem numerischen Merkmal angeben. Es bezeichnet den größten vorkommenden Wert einer Verteilung.

**Minimum** Ein Minimum kann nur bei einem numerischen Merkmal angegeben werden. Es bezeichnet den kleinsten vorkommenden Wert einer Verteilung.

**Arithmetisches Mittel** Mit dem Durchschnitt wird umgangssprachlich meistens das arithmetische Mittel bezeichnet. Das arithmetische Mittel berechnet sich aus der Summe aller Werte eines Merkmals geteilt durch die Anzahl der Fälle.

**Median** Alternativ zum arithmetischen Mittel kann man den Median betrachten, dies ist der Wert in der Mitte einer geordneten Datenmenge. Ein anderes Wort für den Median ist „Zentralwert“. Der Median ist robust gegen Ausreißer, das arithmetische Mittel eher nicht. Vor allem bei schiefen Verteilungen ist der Median eher geeignet, um den „Durchschnitt“ zu beschreiben.