Variabilitätsmaße

St. Lange[1], R. Bender[2]

[1] Abteilung für Medizinische Informatik, Biometrie und Epidemiologie der Ruhr-Universität Bochum
[2] Fakultät für Gesundheitswissenschaften, AG Epidemiologie und medizinische Statistik, Universität Bielefeld

Die Lagemaße einer Verteilung können nur dann sinnvoll interpretiert werden, wenn eine Vorstellung von der Variabilität (Streuung) der Beobachtungen innerhalb der Stichprobe besteht. Wenn beispielsweise der Altersdurchschnitt einer Gruppe von Patienten 35 Jahre beträgt, ist es für Vergleichszwecke – und damit für eine adäquate Interpretation – wichtig zu wissen, ob es sich hierbei ausschließlich um Patienten in einem Alter zwischen 32 und 37 Jahren handelt, oder ob sich auch 20-jährige oder gar Kinder bzw. 60-jährige oder Greise in der Stichprobe befinden.

Das einfachste Streuungsmaß ist der Abstand zwischen dem größten und kleinsten Wert der Stichprobe, auch als Spannweite oder Range bezeichnet. Ein großer Nachteil dieses sehr einfachen und zunächst vernünftig erscheinenden Variabilitätsmaßes ist, dass es vom Stichprobenumfang abhängig ist, eine sehr ungünstige Eigenschaft, die den Nutzen sehr stark einschränkt (3).

Ein anderes Streuungsmaß, der Interquartilsabstand, das heißt die Differenz von 75%- und 25%-Quantil (die »box« aus dem Box-and-Whisker Plot), hat nicht diesen Nachteil und lässt sich anschaulich interpretieren: Zwischen dem 75%- und dem 25%-Quantil liegt die Hälfte der beobachteten Werte (2).

Sehr häufig wird zur Beschreibung der Variabilität innerhalb einer Stichprobe die Varianz bzw. die Standardabweichung benutzt. Hierbei wird zunächst von der Differenz jedes einzelnen Beobachtungswertes vom Mittelwert ausgegangen. Da sich positive und negative Abweichungen bei diesem Vorgehen gegenseitig aufheben (der Mittelwert besitzt gerade die Eigenschaft, dass die Summe der Abweichungsdifferenzen Null ergibt), werden die Differenzen quadriert, anschließend aufsummiert, und diese Summe schließlich durch den um eins verminderten Stichprobenumfang (n-1) dividiert. Diese mittlere quadratische Abweichung wird als empirische Varianz bezeichnet.

Warum wird bei der Berechnung der empirischen Varianz die Summe der Abweichungsquadrate durch n-1 und nicht durch n – was ja zunächst viel einleuchtender erscheint – geteilt? Im Allgemeinen ist es das Ziel, von einer Stichprobe Schlussfolgerungen auf eine (zumeist fiktive) »Grundgesamtheit« zu ziehen. Das heißt, mit dem Mittelwert und der empirischen Varianz aus der Stichprobe sollen der Erwartungswert und die Varianz der Grundgesamtheit geschätzt werden. Es lässt sich zeigen, dass mit der Division durch n-1 eine bessere Schätzung für die Varianz der Grundgesamtheit erzielt werden kann als mit der Division durch n.

Dtsch. Med. Wschr. 2001; 126: T 29–T 30

Tab. 1 Übersetzungen (deutsch – englisch)

Spannweite	range
Standardabweichung	standard deviation
Varianz	variance
Standardfehler des Mittelwertes	standard error of the mean
Variabilitätsmaß	measure of variability
Spannweite	range
Interquartilsabstand	interquartile range
Summe der Abweichungsquadrate	sum of squares

Da die Varianz aufgrund der Quadrierung eine andere Dimension als der Mittelwert hat, wird im Zusammenhang mit dem Mittelwert in der Regel die Standardabweichung angegeben, die sich als Wurzel aus der Varianz berechnet. Daten für quantitative Merkmale aus medizinischen Untersuchungen werden gerne als Mittelwert ± Standardabweichung, häufig auch in Form einer Graphik, ausgewiesen. Dies rührt vermutlich daher, dass für den Fall einer bestimmten theoretischen Verteilung, nämlich der Normalverteilung, sich etwa zwei Drittel aller Werte in einem solchen Bereich befinden.

Für eine Beschreibung von Daten ist dies aber nicht immer sinnvoll: Einerseits wird durch die Abtragung der Standardabweichung zu beiden Seiten des Mittelwertes eine Symmetrie suggeriert, die oft nicht besteht (1), andererseits existiert mit dem Box-and-Whisker Plot eine Darstellungsmöglichkeit mit mehr Information (5).

Problematischer ist es allerdings, anstelle der Standardabweichung den Standardfehler des Mittelwertes (»standard error of the mean«, SEM) zu verwenden. Der SEM beschreibt die Variabilität von Mittelwerten aus Stichproben mit dem gleichen Stichprobenumfang (n) und ist ein Maß für die Präzision der Schätzung des Erwartungswertes durch den Mittelwert (4). Er berechnet sich aus der Standardabweichung nach Division durch die Wurzel aus n. Somit ist der SEM immer kleiner als die Standardabweichung, was vermutlich zu seiner »Beliebtheit« beiträgt. Er hat eine große Bedeutung bei der schließenden Statistik, lässt aber für die Beschreibung von Daten aus einer Stichprobe – im Gegensatz zu den Quantilen oder der Standardabweichung – keine unmittelbare Interpretation zu.

Tab. 1 zeigt die englischsprachigen Übersetzungen der hier besprochenen Begriffe.

kurzgefasst: **Die sinnvolle Beschreibung von Daten erfordert neben der Angabe von zentralen Lagemaßen (Mittelwert, Median) auch noch die Darstellung von Variabilitätsmaßen (Standardabweichung, Interquartilsabstand). Variabilitätsmaße vermitteln eine Vorstellung davon, wie stark die einzelnen Werte um die zentralen Lagemaße streuen, das heißt, wie repräsentativ die zentralen Lagemaße für die Stichprobe (bzw. Grundgesamtheit) sind.**

Literatur

1 Altman DG, Gardner MJ. Presentation of variability. Lancet 1986; 332: 639
2 Altman DG, Bland JM. Quartiles, quintiles, centiles, and other quantiles. Brit med J 1994; 309: 996
3 Altman DG, Bland JM. Detcting skewness from summary informations. Brit Med J 1996; 313: 1200–1201
4 Gardner MJ, Altman DG. Confidence intervals rather than p-values: Estimation rather than hypothesis testing. Brit Med J 1986; 292: 746–750
5 Wilson APR. Box-plots for microbiologists? Lancet 1993; 341: 282

Korrespondenz
Dr. Stefan Lange
Abteilung für Medizinische Informatik, Biometrie und Epidemiologie
Ruhr-Universität
Universitätsstraße 150
44780 Bochum
E-Mail: stefan.f.lange@ruhr-uni-bochum.de