

Standardabweichung und Standardfehler: der kleine, aber feine Unterschied

Standard Deviation and Standard Error: the Small But Important Difference

Autor

J. Koschack

Institut

Abteilung Allgemeinmedizin, Universität Göttingen

Schlüsselwörter

- ◉ beschreibende und schließende Statistik
- ◉ Standardabweichung
- ◉ Standardfehler

Key words

- ◉ descriptive and inferential statistics
- ◉ standard deviation
- ◉ standard error

Einleitung

Immer wieder stößt der aufmerksame Leser wissenschaftlicher Publikationen auf Artikel, in denen statt des Mittelwerts und der Standardabweichung (engl. standard deviation, SD) Mittelwert und Standardfehler (engl. standard error of the mean, SEM) angegeben werden [1]. In manchen Publikationen fehlt gar die Angabe, um welchen dieser beiden Werte es sich handelt; ein Umstand, der nicht nur verwirrt, sondern sogar in die Irre führen kann [2]. Wenn man kein Biometriker oder Biostatistiker ist, denkt man sich zunächst nichts Böses dabei, sondern geht davon aus, dass beide Maße die Variabilität des untersuchten Merkmals in der Patientengruppe beschreiben, also die Verteilung der erhobenen Werte um den Mittelwert. Dies ist jedoch nicht der Fall! Für die eigene Beurteilung der präsentierten Daten einer Publikation ist es daher wichtig, den Unterschied zu kennen. Im Folgenden werden Standardabweichung und Standardfehler beschrieben und miteinander verglichen. Ein Rechenbeispiel veranschaulicht, welcher kleiner, aber feiner Unterschied zwischen diesen beiden Maßen besteht.

Die Standardabweichung

Die Standardabweichung zeigt die reale Streubreite, die der erhobene Wert, z. B. ein Laborwert, innerhalb der untersuchten Gruppe von Patienten aufweist. ◉ **Abb. 1** veranschaulicht, was die Standardabweichung inhaltlich bedeutet: Dargestellt ist die schematisierte Dichtefunktion einer Normalverteilung, die sich ergibt, wenn sich die erhobenen Meßwerte normal um den Gruppenmittelwert \bar{x} verteilen. In dem Bereich der Kurve jeweils eine Standardabweichung rechts und links vom Mittelwert befinden sich 68% aller erhobenen Meßwerte.

Die Standardabweichung wird berechnet, indem die Summe der Abweichungen (in der Formel „ Σ “) des erhobenen Wertes x des Patienten i (in der Formel „ x_i “) vom Gruppenmittelwert \bar{x} gebildet und durch die Anzahl der untersuchten Patienten (in der Formel „ n “) geteilt wird:

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

Drei Besonderheiten sollen noch kurz erklärt werden: Erstens werden die Abweichungen, also der Ausdruck „ $x_i - \bar{x}$ “, zum Quadrat erhoben. Das sorgt dafür, dass große Abweichungen noch mehr Gewicht erhalten. Zweitens wird nicht durch n , sondern durch $n - 1$ geteilt; das sind die Freiheitsgrade, die die Anzahl der bei der Berechnung eines Kennwerts frei variierbaren Werte angeben¹. Drittens wird die Wurzel gezogen; es ergäbe sich sonst nicht die Standardabweichung, sondern die Streuung s (= Varianz). Die Standardabweichung ist aber anschaulicher, da sie die gleiche Einheit wie die ursprünglichen Meßwerte aufweist.

Der Standardfehler

Der Standardfehler hingegen gibt die theoretische Streubreite der Gruppenmittelwerte an, die sich ergeben würde, wenn theoretisch unendlich viele Stichproben (und eben nicht nur die eine) aus der sogenannten Grundgesamtheit gezogen werden würde. Eine Stichprobe ist eine

¹ Ein Beispiel: Die Summe der Differenzen aller Werte von ihrem Mittelwert ergibt 0. Sind von $n=10$ Werten 9 ($=n-1$) bereits zufällig gewählt, steht fest, wie groß die 10. Differenz sein muß. Die Standardabweichung – deren Formel diese Differenzen vom Mittelwert beinhaltet – hat daher $n-1$ Freiheitsgrade. Bei [3] wird dies auf Seite 22 näher erläutert.

Peer reviewed article

eingereicht: 09.03.2008

akzeptiert: 26.03.2008

Bibliografie

DOI 10.1055/s-2008-1073146
 Online-Publikation: 2008
 Z Allg Med 2008; 84: 258–260
 © Georg Thieme Verlag KG
 Stuttgart · New York
 ISSN 1433-6251

Korrespondenzadresse

Dr. rer. nat. J. Koschack
 Abteilung Allgemeinmedizin,
 Universität Göttingen
 Humboldtallee 38
 37073 Göttingen
 jkoscha@gwdg.de

nach bestimmten Kriterien zusammengestellte Teilgruppe der eigentlichen Grundgesamtheit, z. B. 20 allgemeinmedizinische Praxen aus dem Raum Göttingen als Stichprobe aus der Grundgesamtheit aller allgemeinmedizinischen Praxen in Deutschland. Mit anderen Worten: Der Standardfehler zeigt die *theoretische* Streubreite des Stichprobenmittelwerts, im Gegensatz zur Standardabweichung, die die *reale* Streubreite aller Werte der Stichprobe beschreibt. Der Standardfehler wird berechnet, indem die Varianz s der untersuchten Patientengruppe durch die Wurzel des Stichprobenumfangs n geteilt wird:

$$SEM = \frac{s}{\sqrt{n}}$$

Es wird nun gerne behauptet, die Angabe des Standardfehlers sei sinnvoll, weil damit die Genauigkeit der Messung des Mittelwerts ersichtlich wird. In einer nicht natürlichen Untersuchungssituation, wie z. B. bei Experimenten mit genetisch identischen Labortieren, ist dieses Argument nachvollziehbar: Da alle Tiere identisch sind, gehen die Unterschiede zwischen der Kontrollgruppe und der Experimentalgruppe allein auf die Versuchsanordnung zurück. Ein Beispiel wäre ein Laborwert, in dem sich die beiden Gruppen unterscheiden, weil die Tiere unterschiedliches Futter bekommen haben. Dass nicht alle Tiere einer Gruppe den absolut identischen Laborwert aufweisen, liegt daran, dass die Bestimmung des Laborwerts einer gewissen Ungenauigkeit unterworfen ist. Es ist nicht auf die biologische Variabilität zurückzuführen, da diese durch den Einsatz genetisch identischer Tiere konstant gehalten wurde. In klinischen Studien mit Patienten hingegen interessiert genau diese biologische Variabilität, die durch die Standardabweichung beschrieben wird! Die Angabe des Standardfehlers anstelle der Standardabweichung verschleiert also gelegentlich die sehr große (biologisch bedingte) Streubreite des erhobenen Werts. Im Gegensatz zur Standardabweichung, die nur in einem geringen Ausmaß durch die Größe der Stichprobe beeinflussbar ist, wird der Standardfehler des Mittelwerts kleiner, je größer die Stichprobe ist.

Die Angabe der Standardabweichung ist also auf jeden Fall der des Standardfehlers vorzuziehen. Im Hinterkopf sollte aber behalten werden, dass statistische Aussagen, z. B. über einen signifikanten Unterschied zwischen zwei Patientengruppen, mathematisch auf den Standardfehler zurückgehen. Die Statistik trifft eine Aussage darüber, ob der Unterschied zwischen zwei Patientengruppen nicht nur zufällig, sondern unter Berücksichti-

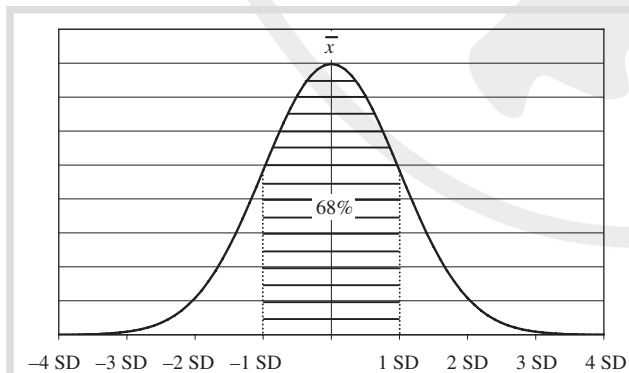


Abb. 1 Normalverteilungskurve mit \bar{x} als Mittelwert und SD als Standardabweichung.

Tab. 1 Vergleich zwischen Standardabweichung SD und Standardfehler SEM

Standardabweichung (SD)	Standardfehler (SEM)
– ist eine Aussage über die Streuung der erhobenen Werte in einer Stichprobe	– ist eine Aussage über die „Genauigkeit“ des Mittelwerts in einer Stichprobe
– hängt von der biologischen Variabilität ab	– hängt von der Meßgenauigkeit ab
– ist ein beschreibendes Maß	– ist ein statistisches Maß
– ist nur wenig durch die Größe der Stichprobe beeinflussbar	– steht in direktem Verhältnis zur Größe der Stichprobe

gung der Meßgenauigkeit (Standardfehler!) statistisch signifikant ist (◉ **Tab. 1**).

Ein kleines Rechenbeispiel

Ein kleines Rechenbeispiel, basierend auf Daten einer realen Studie [4], veranschaulicht die Zusammenhänge zwischen Stichprobengröße und Standardfehler und damit zusammenhängend die Signifikanz von Gruppenunterschieden. Der zuvor beschriebene Unterschied zwischen Standardabweichung und Standardfehler wird ebenfalls deutlich.

Gegeben seien zwei Gruppen mit jeweils 15 Patienten. Beide Gruppen setzen sich zusammen aus Patienten mit kardiovaskulären Risikofaktoren; Patienten der Gruppe A1 zeigen eine eingeschränkte Ejektionsfraktion, Patienten der Gruppe B1 hingegen haben unauffällige echokardiographische Ergebnisse. Ein Mittelwertsvergleich soll nun prüfen, ob sich die beiden Gruppen bezüglich eines Laborwerts, dem sogenannten NT-proBNP (N-terminal pro-brain natriuretic peptide), signifikant unterscheiden. Dahinter steht die Idee, diesen Laborparameter zukünftig zur Diagnostik einer eingeschränkten linksventrikulären Funktion und somit als Ersatz für die Echokardiographie einzusetzen. Als statistische Spielerei werden zwei weitere Gruppen A2 und B2 mit jeweils 45 Patienten dargestellt. Deren Daten ergeben sich durch eine simple Verdreifung der realen Werte der Gruppen A1 und B1. ◉ **Tab. 2** zeigt nun die Mittelwerte, Standardfehler und Standardabweichungen sowie das Ergebnis des t-Tests für unabhängige Stichproben, mit dem überprüft wurde, ob sich die beiden Gruppen bezüglich des NT-proBNP statistisch signifikant unterscheiden.

Die in ◉ **Tab. 2** abgebildeten Zahlen verdeutlichen, in welchem Ausmaß der Standardfehler kleiner wird, wenn die Stichprobengröße verdreifacht wird, nämlich um 44% (Gruppe A2 zu A1 = 378 zu 670 bzw. Gruppe B2 zu B1 = 25 zu 45). Die Standardabweichung hingegen verringert sich lediglich um 2% (Gruppe A2 zu A1 = 2535 zu 2595 bzw. Gruppe B2 zu B1 = 170 zu 174). Obwohl die Mittelwerte identisch bleiben und die biologische Variabilität (= Standardabweichung) der Laborwerte sich nur unwesentlich verringert, wird die Überprüfung der Gruppenunterschiede signifikant. Jeder Kliniker würde selbst bei diesem signifikanten Ergebnis zögern, den Laborwert für eine diagnostische Entscheidung heranzuziehen, wenn er die sehr großen Standardabweichungen in den Gruppen sieht. Es ist ihm klar, dass der Überlappungsbereich der Laborwerte der beiden Gruppen sehr groß ist, also sehr viele Werte keine klare Entscheidung zulassen, ob eine verminderte Ejektionsfraktion vorliegt oder nicht. Allenfalls extrem hohe oder sehr niedrige Werte sind aussagekräftig; eine Vielzahl an Werten bleibt jedoch uneindeutig. Die Angabe der Standardfehler statt der Standardabweichungen verschleiert

	Gruppe A1 (n = 15)			Gruppe B1 (n = 15)			Signifikanz ¹ p-Wert
	MW	SEM	SD	MW	SEM	SD	
NT-proBNP, pg/ml	1421	± 670	± 2595	153	± 45	± 174	0,069

	Gruppe A2 (n = 45)			Gruppe B2 (n = 45)			Signifikanz p-Wert
	MW	SEM	SD	MW	SEM	SD	
NT-proBNP, pg/ml	1421	± 378	± 2535	153	± 25	± 170	0,001

Tab. 2 Vergleich der Gruppen bezüglich NT-proBNP; Gruppe A1 und B1 basieren auf realen Werten; Gruppe A2 und B2 basieren auf der Verdreifachung der realen Werte von A1 und B1

MW Mittelwert, SEM Standardfehler des Mittelwerts, SD Standardabweichung

¹t-Test für unabhängige Stichproben (zweiseitig getestet; $\alpha = 0.05$)

diese große biologische Variabilität. Deswegen kann folgende Daumenregel gelten:

„So, when you see SEMs in a publication, smile, then mentally convert them into SDs to see how big the differences are between the groups. For example, if there are 25 subjects in a group, increase the size of the SEM by a factor of 5 (=square root of 25) to turn it into an SD².“ (zitiert nach: <http://www.sportsci.org/resource/stats/meansd.html>).

Interessenskonflikte: keine angegeben.

Literatur

- 1 Lubien E, DeMaria A, Krishnaswamy P, Copton P, Koon J, et al. Utility of B-natriuretic peptide in detecting diastolic dysfunction. *Circulation* 2002; 105: 595–601
- 2 Altman DG, Bland JM. Standard deviations and standard errors. *BMJ* 2005; 331: 903
- 3 Bortz J. Statistik für Sozialwissenschaftler. Berlin: Springer 1993
- 4 Koschack J, Scherer M, Luers C, Kochen MM, Wetzel D, et al. Natriuretic peptide vs. clinical information for diagnosis of left ventricular systolic dysfunction in primary care. *BMC Fam Prac* 2008; 9: 14 [Epub ahead of print]

Zur Person



Dipl.-Psych. Dr. rer. nat. Janka Koschack, Jahrgang 1972, Diplom 1998, Promotion in der Neuropsychologie 2003, seit 2004 in der Abteilung Allgemeinmedizin der Universität Göttingen. Arbeitsschwerpunkte: Methoden der allgemeinmedizinischen Versorgungsforschung, Medikamentenadhärenz und Krankheitsbewältigung bei Menschen mit chronischen Erkrankungen.

²„Wenn Sie in einer Publikation die Angabe von Standardfehlern entdecken, dann schmunzeln Sie und wandeln die angegebenen Standardfehler in Standardabweichungen um, um die tatsächlichen Unterschiede zwischen den Gruppen einschätzen zu können. Wenn z.B. eine Gruppe aus 25 Probanden besteht, nehmen Sie den Standardfehler mal Faktor 5 (Wurzel aus 25 ist 5) und erhalten so die Standardabweichung.“