

Mehrdeutigkeit und Bedeutungsunterschiede: Ein Plädoyer gegen Breite einer Verteilung und für Variabilität¹

JENNIFER J. KAPLAN, NEAL T. ROGNESS, DIANE G. FISHER,
EAST LANSING, ALLENDALE, LAFAYETTE, USA

¹ Original ‚Lexical ambiguity: making a case against spread‘ in *Teaching Statistics* 34 (2012) 2, 56–60.
Übersetzung und Bearbeitung:
MANFRED BOROVCNIK, KLAGENFURT

Zusammenfassung: Wir argumentieren für die Vermeidung der Benennung „Breite“ einer Verteilung im Rahmen des einführenden Statistikkurses, wenn man eigentlich die Streuung oder Variabilität meint. „Breite“ sollte durch „Variabilität“ ersetzt werden. Die Empfehlung wird durch empirische Ergebnisse zum Verständnis und der Vieldeutigkeit der Worte gestützt.

1 Einleitung – Alltagssprache und wissenschaftliche Bezeichnungen

Margaret Rangelcroft hat in einem Artikel in *Teaching Statistics* die Frage thematisiert, dass Worte in der Alltagssprache und in der Statistik unterschiedliche Bedeutungen haben. Rangelcroft (2002) führt hierzu als Beispiel Mittel(wert) an.

[Englisch *mean*; die Deutungen beziehen sich auf das Wort im englischen Original.]

Mean (Mittelwert) wird in der Alltagssprache als Adjektiv oder Verb verwendet und kann u. a. folgendes bedeuten:

- (1) gemein oder kleinlich;
- (2) durchschnittlich oder dazwischen liegend;
- (3) bedeuten, meinen oder beabsichtigen.

Bereichsspezifische Worte, welche ähnlich zu gewöhnlichen Wörtern sind, werden auch polysemisch genannt [lexically ambiguous] (Barwell 2005). Wie Rangelcroft weiter ausführt, besteht bei gewöhnlichen Wörtern, die in einen technischen Bereich aufgenommen werden, folgende Gefahr: Wenn Studierende diese Worte zum ersten Mal im technischen Zusammenhang hören, so neigen sie dazu, diese als eine neue Facette des *schon* gelernten Begriffs zu interpretieren.

Der Gebrauch von Wörtern mit Überschneidung der Wortbedeutung im Alltag verführt zu Assoziationen zwischen den Wörtern, die sie schon kennen und Wörtern, die so ähnlich klingen, aber in der Statistik eine ganz andere Bedeutung haben.

Rangelcroft (2002, S. 36) schließt daraus Folgendes:

Zu wissen, dass ein Problem besteht, ist schon der erste Schritt dahin, es zu ‚lösen‘. Wenn wir uns als Lehrende besser darauf einstellen, welche Missverständnisse sich allein durch sprachliche Schwierigkeiten ergeben, so können wir sie vielleicht tatsächlich erkennen und die nötigen Erklärungen anschließen. Vielleicht können wir auch einen Schritt zurück machen und solche Schwierigkeiten durch sorgfältigen Gebrauch der Sprache vermeiden.

In diesem Beitrag wollen wir auf die Schwierigkeiten aufmerksam machen, die sich aus dem Wort *Breite* bzw. Ausbreitung [*spread*] ergeben, mit dem man den Begriff Variabilität oder Streuung vornehmlich anspricht. Wir regen – basierend auf unseren Untersuchungen – an, Breite durch Variabilität zu ersetzen, auch um gleichzeitig die Schwierigkeiten im Verständnis des Begriffs Variabilität besser ansprechen zu können.

2 Das Problem mit der *Breite* (Ausbreitung)

Die Bezeichnung *Breite* ist in einführenden Lehrbüchern und im Advanced-Placement-(AP)-Statistik-Programm in den USA zu einem Synonym für die Worte *Variabilität* und *Streuung* geworden.

Man findet häufig eine Anleitung wie ‚Vergiss nicht, die Gestalt, das Zentrum und die Breite der Verteilung zu erörtern.‘ (DeVeaux et al. 2009, S. 68). In der Tat, in diesem Lehrbuch findet man im Kapitel ‚Darstellen und Zusammenfassen quantitativer Daten‘ Breite exklusiv als Benennung für Streuung; das Wort Variabilität kommt gar nicht vor. ‚Elementare Methoden der Statistik‘ (Moore 2010), ein weiteres, sehr beliebtes Lehrbuch in den USA, verwendet im Kapitel, das Verteilungen beschreibt, das Wort Variabilität ein einziges Mal auf 30 Seiten.

Die Erstautorin hatte folgende nachhaltige Erfahrung: Bei der Beurteilungssitzung für das AP-Programm waren mehrere Hundert Statistik-Lehrende anwesend und beteten mantramäßig die Leier von ‚Gestalt, Zentrum und Breite‘ während sie darüber diskutierten, was als korrekte Antwort auf die Prüfungsfragen zur Beschreibung einer Verteilung gelten konnte.

Wir geben drei Gründe, *Breite* als primäre Bezeichnung für Variabilität oder Streuung zu vermeiden:

- (1) Breite hat schon eine Vielzahl von Bedeutungen; nur wenige haben mit Streuung zu tun.
- (2) Vorwissen zum Wort Breite ist näher zu Wortbedeutungen, welche Missverständnisse und Fehlvorstellungen geradezu heraufbeschwören.
- (3) Studierende haben nach einem Kurs in Statistik wenig Ahnung, was Breite bedeuten soll.

Das Oxford-English-Dictionary stellt mehr als 25 Definitionen von Breite [spread] bereit; nur eine tangiert den Begriff Streuung. Die anderen umfassen eine breite Palette von „jede Substanz, die man auf einem Stück Brot verstreichen kann“, „eine Werbung, die man an prominenter Stelle unterbringt“, „etwas entfalten, ausdehnen, dehnen“ bis zu „Dickleibigkeit“. Breite [spread] kann als Hauptwort, als Eigenschaftswort und als transitives bzw. intransitives Zeitwort verwendet werden.

[Die Vieldeutigkeit von Worten ist natürlich sprachabhängig; so besteht im Deutschen ein deutlicherer Unterschied zwischen Nomen und Verb; im Englischen kann man fast aus jedem Nomen ein Verb kreieren.]

Im folgenden Ausschnitt steht nur *eine* angesprochene Bedeutung mit dem statistischen Begriff Streuung in Verbindung:

Sagte der Statistiker aus West-Texas ...

Willkommen in meiner *Bleibe*; fühl dich wie zu Hause. Komm und *breite* deine Unterlagen aus. Bedien dich selbst, es gibt eine große *Auswahl* zu essen; vergiss nicht, den Schwarzbeeren-*Aufstrich*; *verteil* einfach ein bisschen auf dem Brot. Schau dir das tolle Tisch-tuch an; meine Oma hat das *Tuch* für mich angefertigt. Wenn du bereit bist, öffnen wir die *Tabellenkalkulation* und schauen, ob wir nicht doch die *Streuung* jener Daten bestimmen können.

Wenn wir sicher sein könnten, dass Studierende das Wort *Breite* im statistischen Zusammenhang mit seiner definitorischen Bedeutung von *Streuung* in Verbindung brächten, so würden wir dessen Gebrauch im einführenden Statistikkurs vorbehaltlos befürworten. Unsere Daten sprechen jedoch dagegen.

3 Eine qualitative Untersuchung der Wahrnehmung von *Breite*

Die Daten entstanden als Teil einer größeren Studie über Mehrdeutigkeit und Bedeutungsunterschiede in der Statistik. Wir benutzten ein Pre-Post-Test-Design.

Studierende im Anfängerkurs in Statistik schrieben in der ersten Woche Sätze und Definitionen für Breite, indem sie auf ihr Vorverständnis des Wortes zurückgriffen. In der letzten Woche wurde der Test wiederholt; sie sollten sich nun primär auf die Definition von Breite in der Statistik beziehen.

Die Ergebnisse stellen eine Zufallsauswahl von 160 Pre- und 123 Post-Tests dar; insgesamt waren drei Universitäten mit 14 Instruktoressen beteiligt. Das Projekt ist in Kaplan, Fisher & Rogness (2009, 2010) genauer beschrieben.

3.1 Pre-Test-Ergebnisse und Interaktionen zwischen den Bedeutungen

Entsprechend unseren Voranalysen gaben sehr viele an, dass Breite etwas mit *gleichmäßig überdecken* zu tun hat; eine etwas kleinere Gruppe bezog sich tatsächlich auf *Streuung*. Die weiteren Antworten sind nicht mehr so deutlich in starke Gruppen zu unterteilen.

Gleichmäßig überdecken, in einer dünnen Schicht überdecken	54	(34 %)
Verteilen, streuen, sich verteilen, sich über eine größere Fläche erstrecken	52	(33 %)
Reichweite, Spannweite zwischen den Daten, Ausbreitung von Punkten	19	(12 %)
7 weitere Kategorien:	3–7 Antworten	
Eine große Gruppe Ein Spreadsheet Marmelade, Butter, Soße Öffnen (Flügel) In einer ordentlichen Weise auslegen Eine große Fläche Mit einem Leintuch/einer Decke überziehen		

Tab. 1: Gruppierung der Pre-Test-Antworten auf Breite ($n = 160$)

Zusätzlich zur Vielfalt der Deutungen des Wortes *Breite* möchten wir noch auf mögliche Interaktionen zwischen den gegebenen Alltagsdefinitionen von Breite und dem statistischen Konzept von Variabilität hinweisen. Maßzahlen der Variabilität oder Streuung sagen etwas aus, wie stark Daten um ein Zentrum der Verteilung streuen.

Wenn wir ein Histogramm betrachten, so wird die Variabilität in der Verteilung größer sein, wenn es mehr hohe Balken weiter außerhalb vom Zentrum gibt, und kleiner, wenn die Balken weit draußen kleiner sind. Um die Variabilität zu erfassen, muss man also die Entfernung auf der horizontalen Achse betrachten.

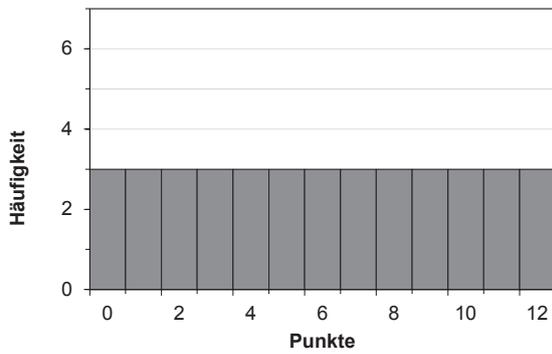


Abb. 1: Glattes Histogramm mit einer großen Spannweite

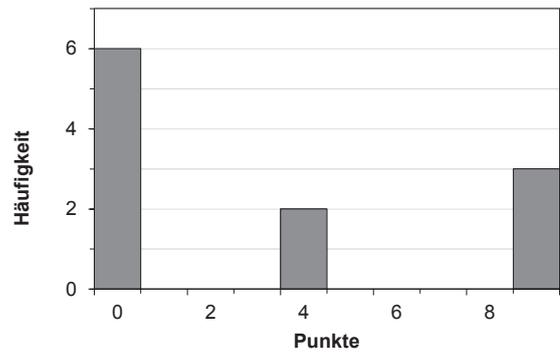


Abb. 3.: Gleichmäßig ausgebreitetes Histogramm

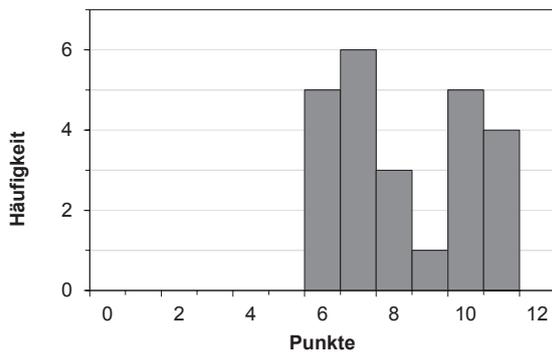


Abb. 2: Bauchiges Histogramm mit kleiner Spannweite

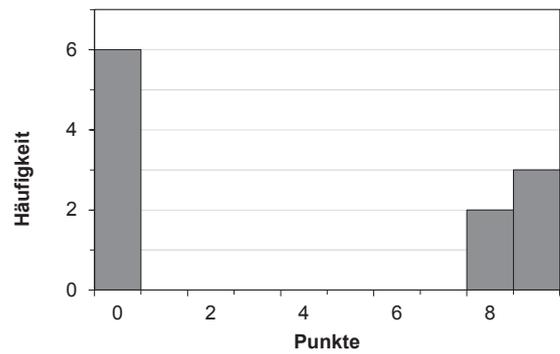


Abb. 4: Histogramm mit Daten an den Rändern

Studierende, die an *Breite* als Herstellen einer gleichmäßigen Schicht denken (etwa die Erdnussbutter auf dem Brot gleichmäßig verstreichen), mögen die „Variabilität der Verteilung auf der Basis der [...] sichtbaren Bauchigkeit des Graphen beurteilen“, wie delMas & Liu (2005, S. 55) feststellen. Solche Studierenden würden dann behaupten, dass ein Histogramm wie in Abb. 1, mit einem glatten oberen Verlauf aber einer großen Spannweite, weniger Breite (oder Variabilität) anzeigt als Abb. 2, welches ein bauchiges Histogramm mit einer kleinen Spannweite zeigt.

Weiters merkten delMas & Liu (2005) an, dass Studierende glauben, dass die Variabilität oder Breite in einer Verteilung größer wird, wenn die Balken eines Histogramms gleichmäßige Abstände über den ganzen horizontalen Bereich haben (Abb. 3). Statistisch gesprochen ist aber die Variabilität höher, wenn die Ränder stärkere Gruppierungen haben (wie in Abb. 4). Dieses Missverständnis könnte durch eine Wechselwirkung mit der Alltagsbedeutung von *Breite* (Ausbreitung) als gleichmäßig überdecken oder als ordentlich schön auslegen verursacht sein.

3.2 Post-Test-Ergebnisse

Als abschließendes Argument zur Vermeidung von *Breite* als Bezeichnung für Variabilität seien die Ergebnisse des Post-Tests angeführt.

Bezüge zur graphischen Darstellung	40	(32 %)
Streuung oder Variabilität	38	(31 %)
Bezüge zu Spreadsheets	26	(21 %)
Unklare Aussagen – nicht zuordenbar	15	(12 %)

Tab. 2: Post-Test-Antworten auf *Breite* ($n = 123$)

Die studentischen Definitionen fallen in drei Hauptkategorien (Tabelle 2): Antworten mit direktem Bezug zu einer graphischen Darstellung (32 %), mit Bezug zu statistischer Variabilität (31 %), sowie Spreadsheets (21 %). Ein Teil der Antworten (12 %) war unklar und konnte nicht klassifiziert werden.

In der Kategorie mit Bezug zu *Streuung* oder Variabilität identifizierten wir zwei Untergruppen: jene, die Variabilität, Varianz oder Standardabweichung, und jene, die Spannweite ansprechen (Tabelle 3).

Definitionskategorie	Exemplarischer Satz (S) und Definition (D) von Studierenden
<i>Definitionen, die mit Streuung zusammen hängen (31 %)</i>	
a) Variabilität, Varianz, Standardabweichung	S: Die Breite (Ausbreitung) der Daten ist sehr groß D: Bedeutet die Variabilität der Daten (klein oder groß)
b) Spannweite	S: Da ist eine große Breite in den Daten D: Spannweite
<i>Definitionen, die sich auf die Darstellung der Daten beziehen (32 %)</i>	
a) Wie sehen die Daten aus, wenn sie in einem Diagramm dargestellt werden	S: Die Daten sind gleichmäßig ausgebreitet D: Ausbreitung – Breite: wie sehen Daten in einem Graphen aus
b) Wie sind die Daten verteilt (ohne eine visuelle Darstellung zu erwähnen)	S: Die Gruppe von Daten ist gleichmäßig ausgebreitet D: Breite ist die Verteilung der Daten
c) Daten sind weit verstreut (kann, muss sich aber nicht auf eine visuelle Darstellung beziehen)	S: Die Punkte sind über einen großen Bereich der Punktwolke ausgebreitet D: Breite Fläche, nicht nahe beieinander

Tab. 3: Fünf exemplarische Antworten auf Breite bzw. Ausbreitung im Post-Test

In der Gruppe, die Bezüge zur graphischen Darstellung herstellten, machten wir drei Untergruppen aus: Antworten, die sich direkt auf das Aussehen der Verteilung in einem Diagramm beziehen; Antworten, die sich auf kein Diagramm beziehen; Antworten, in denen Formulierungen vorkommen wie, „die Daten sind weit auseinander“.

Unsere Ergebnisse zeigen insgesamt, dass nahezu ein Drittel das Wort Breite darauf beziehen, wie die Daten aussehen oder in einem Diagramm dargestellt werden. Und nicht, wie weit die Daten von einem Zentrum entfernt sind. Weiters sprechen zwei Drittel jener Aussagen, die sich auf Variabilität beziehen, direkt die Spannweite als Synonym für Breite an.

3 Vorschlag zur Ersetzung: Variabilität

Wenn nun die Aufgabe darin besteht, die Bezeichnung *Breite* für den Begriff *Variabilität* bzw. *Streuung* [variability and dispersion] zurückzunehmen, so ist die Frage, was wir stattdessen favorisieren sollten. Natürlich haben beide ihre Vorteile, aber wir meinen, dass für den Einführungskurs *Variabilität* am besten geeignet ist. Rangecroft (2002, S. 36) fragt

„Sollen wir auf Zugänglichkeit oder Präzision abzielen? Sollen wir über einen Term sprechen, der für die meisten Kinder vertraut ist, oder [...] das Wort nehmen [...], das] korrekt ist, aber [...] in der Alltagssprache eine alternative Bedeutung hat?“

Rangecroft beantwortet ihre eigene Frage und schlägt vor, unser professionelles Urteil als Lehrende zu nutzen, um die Entscheidung zu treffen. Im Folgenden wird erörtert, warum Dispersion (Streuung) für das Zielpublikum ungeeignet ist.

In der Umgangssprache deckt *Streuung* (Dispersion) das Wesentliche des Begriffs ab, der durch Breite definiert wird: Streuung, Streubereich, über einen größeren Raum verstreuen bzw. zerstreuen. Streuung erweckt flächige oder räumliche Bilder, welche gut geeignet sind, *Variabilität* in zweidimensionalen Daten, etwa in Punktwolken, zu beschreiben, aber für einzelne Merkmale irreführend sind.

Histogramme sind zweidimensionale Darstellungen von einzelnen Merkmalen. Wenn wir die *Variabilität* aus Histogrammen schätzen, so betrachten wir die durchschnittliche Entfernung des Hauptteils der Verteilung vom Zentrum. Wir trennen die Dimensionen und betrachten die Richtung, die mit der horizontalen Achse verbunden ist; das ist die Achse, welche das untersuchte Merkmal darstellt. Dann wenden wir uns der zweiten Dimension zu, der Höhe der Balken, um zu beurteilen, wo das „Gewicht“ der Verteilung liegt.

Die Bezeichnung „*Dispersion* bzw. *Streuung*“ mag die Studierenden dazu verleiten, beide Dimensionen gleichzeitig zu betrachten. Dann würden sie vielleicht die Spitzen der Balken als „Punkte“ in der Ebene ansehen, und die Fehlvorstellung liegt nahe, dass bauchige Graphen mehr Variabilität anzeigen als glatte. Weil das Konzept Variabilität zuerst für eindimensionale Merkmale eingeführt wird, sollte man Dispersion bzw. Streuung für die Beschreibung von Beziehungen zwischen zwei Merkmalen vorbehalten. Dann könnte es als Synonym für *Variabilität* eingeführt werden.

Damit verbleibt uns noch das Wort *Variabilität*. Während es klanglich nahe bei *Variable* und *Varianz* liegt, markiert es eine Ausgewogenheit zwischen nicht zu

vertraut und nicht zu technisch. Quantifizierung der Variabilität ist eine zentrale Idee der Statistik, aber ein Verständnis dafür und Urteile darüber aufzubauen ist keine leichte Aufgabe (Reading & Shaughnessy 2004). Wir schlagen daher die Bezeichnung *Variabilität* vor. Das Wort hat einen technischen Beigeschmack, der ihm eine Wichtigkeit verschafft, die notwendig ist, dass Studierende verstehen, dass es kein triviales Konzept ist.

In diesem Fall mag ein weniger bekanntes Wort auch dazu verhelfen, dass die Studierenden den Begriff, der damit abgedeckt wird, wichtig genug nehmen, sodass sie sich ernsthaft bemühen, ihn zu verstehen. Diese Wortwahl mag auch Lehrende an die Schwierigkeiten erinnern, die mit dem Begriff Variabilität verbunden sind. Schließlich verwendet auch David Moore (1985, S. 180) in der 5. Auflage seiner „Konzepte und Kontroversen“ die Formulierung „Maß der Breite oder Variabilität“, was andeutet, dass *Variabilität* die Bezeichnung der Wahl war, bevor *Breite* in Mode gekommen ist.

Wir schließen unser Plädoyer für *Variabilität* und gegen *Breite* mit zwei Caveats.

- Erstens, obwohl wir natürlich hoffen, dass der Gebrauch der Bezeichnung Variabilität für das Konzept der Variabilität dazu beiträgt, dass sich Lehrende an dessen Schwierigkeiten erinnern und dass Studierende weniger Fehlvorstellungen dazu entwickeln, sind wir uns dessen klar bewusst, dass ein Wechsel in der Bezeichnung kein Wundermittel darstellt, das alle Probleme löst.
- Zweitens, wir erwarten keineswegs, dass die Bezeichnung *Breite* überhaupt nie verwendet werden sollte, wenn man Variabilität von Daten bespricht, wir plädieren lediglich für einen vernünftigen Umgang damit. So etwa ist es durchaus sinnvoll, wenn Studierende folgenden Satz hören: „diese Daten weisen eine hohe Variabilität auf, weil sie so weit vom Zentrum ausgebreitet sind [spread out]“

Die Autoren dieses Beitrags sind in den vergangenen Semestern im Unterricht sehr sorgsam mit dem Gebrauch der Bezeichnung Variabilität anstelle von Breite umgegangen und möchten anregen, dass an-

dere, insbesondere auch Lehrbuchautoren, diesem Beispiel folgen.

Literatur

- Barwell, R. (2005): Ambiguity in the mathematics classroom. In: *Language and Education* 19(2), S. 118–126.
- delMas, R.; Liu, Y. (2005): Exploring students' conceptions of the standard deviation. In: *Statistics Education Research Journal* 4(1), S. 55–82. www.stat.auckland.ac.nz/serj (Zugriff: 8.8.2014).
- DeVeaux, R. D.; Velleman, P. F.; Bock, D. E. (2009): *Intro Stats*. 3rd edn. Boston: Pearson Education.
- Kaplan, J. J.; Fisher, D.; Rogness, N. (2010): Lexical ambiguity in statistics: how students use and define the words: association, average, confidence, random and spread. In: *Journal of Statistics Education* 18(2), S. 1–22. www.amstat.org/publications/jse/v18n2/kaplan.pdf (Zugriff: 8.8.2014).
- Kaplan, J. J.; Fisher, D.; Rogness, N. (2009): Lexical ambiguity in statistics: what do students know about the words: association, average, confidence, random and spread? In: *Journal of Statistics Education* 17(3). www.amstat.org/publications/jse/v17n3/kaplan.html (Zugriff: 8.8.2014).
- Moore, D. (2010): *The Basic Practice of Statistics*. 5th edn. New York: W. H. Freeman.
- Moore, D. (1985): *Concepts and Controversies*. 2nd edn. New York: W. H. Freeman.
- Rangecroft, M. (2002): The language of statistics. In: *Teaching Statistics* 24(2), S. 34–37.
- Reading, C.; Shaughnessy, M. (2004): Reasoning about variation. In: Ben-Zvi, D.; Garfield, J. (eds.): *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Dordrecht: Kluwer, S. 201–226.

Anschrift der Verfasser

Jennifer J. Kaplan
Michigan State University
East Lansing, MI 48823, USA
kaplan@stt.msu.edu

Neal Rogness
Grand Valley State University
Allendale, MI 49401, USA
rognessn@gvsu.edu

Diane G. Fisher
University of Louisiana
Lafayette, LA 70504, USA
dgf9042@louisiana.edu