IST DIE STANDARDABWEICHUNG AN DEN MITTELWERT GEBUNDEN ?

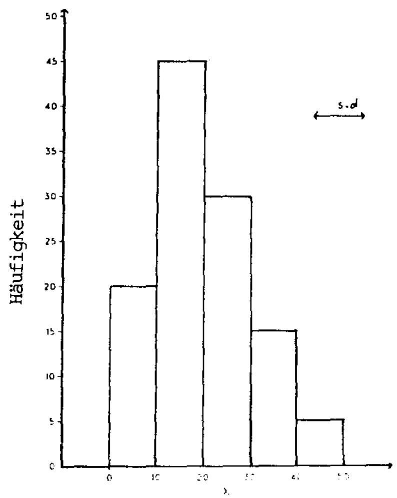
nach T. GORDON, Loughborough University

Originaltitel in "Teaching Statistics" Vol. 8 (1986), Nr. 2:
Is the Standard Deviation Tied to the Mean?
Übersetzung und Bearbeitung: K. Röttel

Zusammenfassung: Den Wert für die Varianz als arithmetisches Mittel der Abweichungsquadrate vom Mittelwert erhält man überraschenderweise auch, wenn man sämtliche Differenzen der Verteilung als Abweichungen verwendet, d. h. das arithmetische Mittel ihrer Quadrate berechnet. Es wird jedoch empfohlen, die Standardabweichung, die Wurzel aus der Varianz, trotz der "demokratischen" Möglichkeit für die Berechnung bei Darstellungen rechts und links vom Mittelwert zu zeichnen.

ZDM-Klassifikation: K40

Jüngst tauchte in einer Diskussion mit einem Kollegen die Frage auf, wie man die Standardabweichung graphisch darstellen könne. Nehmen Sie an, es läge ein Histogramm wie in Figur 1 vor und man möchte dort die Standardabweichung, die den Wert 10.6 besitze, einzeichnen.



Figur 1 Wohin mit der Standardabweichung?

Die Standardabweichung muß als Entfernung oder Intervall dargestellt werden, und ich war der Meinung, daß ein horizontaler "Vektor" mit Spitzen an beiden Enden, der die Länge von 10.6 Einheiten hat, richtig wäre. Das Problem ist, wohin dieser "Doppelvektor" zu setzen sei.

Wir waren uns einig, daß eine Plazierung wie in Figur 2 passend wäre; nichtsdestoweniger blieb die Frage: Bezeichnet die Standardabweichung notwendig die Entfernung vom Mittelwert aus, oder ist die Standardabweichung eine selbständige Größe. Mein Kollege Neville war der ersten Meinung, und ich ein Widerspruchsgeist, nahm den zweiten Standpunkt ein.

Neville führte an, man müsse ganz deutlich von einer Standardabweichung vom Mittelwert sprechen; denn bei der Berechnung aus den Rohwerten anhand der Formel

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

ist die Standardabweichung die Quadratwurzel aus dem arithmetischen Mittel der Abweichungsquadrate vom Mittelwert. Es sei deshalb logisch, die Standardabweichung als Bereich zu beiden Seiten des Mittelwertes zu deuten.

Ich begann, nach einer "demokratischeren" Definition der Standardabweichung zu suchen, d.h. einer, die den Mittelwert nicht verwendet und alle Meßwerte gleich behandelt. Naheliegend war, die Wurzel aus dem mittleren Differenzenquadrat aller n Daten zu bilden. Und ich fand, daß dieser Ansatz eine interessante Anregung bot.

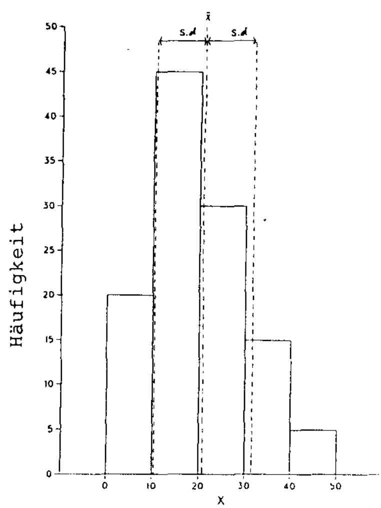
Da sowohl i als auch j von 1 bis n laufen, muß die Summe aus n^2 Summanden bestehen, von denen n automatisch null werden (wenn $i=j$). Die Wurzel aus dem Mittelwert der Abweichungsquadrate ist dann

$$\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2}{n^2}} \; . \qquad (2)$$

Wenn man die Klammern ausrechnet, sieht man, daß dieser Wert das $\sqrt{2}$ -fache der Standardabweichung ist in Übereinstimmung mit (1) (bis auf den Faktor $\sqrt{2}$). Diese "demokratische" Definition der Standardabweichung ist vom Ansatz her nicht an den Mittelwert gebunden - das gleiche gilt für den "Pfeil" in Figur 1.

Das Besondere ist die Methode der Mittelwertbildung in Gleichung (2). Weil n Terme automatisch null wurden, ist es sinnvoll, den Nenner durch n^2-n zu ersetzen; leicht überprüft man, daß der Ausdruck dann $s\cdot\sqrt{2}$ ist, wobei s^2 der übliche Zufallsschätzwert der Populationsvarianz ist. Bekanntlich ist die "übliche Varianz" nicht als Wurzel aus dem mittleren Abweichungsquadrat wie in Gleichung (1) definiert, sondern mit dem Nenner n-1. Für beide Arten haben sich auch die Bezeichnungen (s_n und s_{n-1}) eingebürgert, die manche Taschenrechner gleichfalls ausweisen.

Figur 2
Die Standardabweichung vom Mittelwert aus



Ich bin nun überzeugt, daß die Standardabweichung tatsächlich nicht an den Mittelwert gebunden ist, aber ich bin mir dessen bei Neville nicht sicher. Was die graphische Darstellung betrifft, so muß man den Pfeil irgendwohin zeichnen; ich werde bei der in Figur 2 eingezeichneten Art bleiben.

Ich danke Neville Hunt, mich zu den obigen Gedanken angeregt zu haben.

Anmerkung des Übersetzers

Ein kleines Beispiel zur Veranschaulichung, das der Übersetzer mit seinem Freund Hans Trenz durchrechnete, wohl wissend, daß Schulnoten laut Definition (sehr gut, gut,..) nur eine Rangskala darstellen:

Verteilung der Noten:

x					
x			x	x	
x		x	x	x	
1	2	3	4	5	6

Mittelwert $\bar{x}$ = 3.00 .

Berechnung nach Gleichung (1) liefert $\sqrt{\frac{22}{8}}$ = 1.6583 . Die Berechnung nach der "üblichen" Definition der Varianz gibt $\sqrt{\frac{22}{7}}$ = 1.7728. Gleichung (2) führt zum Wert

$$\sqrt{\frac{352}{64}} = \sqrt{\frac{44}{8}} = \sqrt{2}\sqrt{\frac{22}{8}} = \sqrt{2}\cdot 1.6583 \; .$$

Bei Verwndung von n^2-n (d.h. 64 - 8 = 56) erhält man schließlich

$$s = \sqrt{\frac{352}{56}} = \sqrt{\frac{44}{7}} = \sqrt{2}\sqrt{\frac{22}{7}} = \sqrt{2}\cdot 1.7728 \; .$$

Als übergeordneten Gedanke, der die Ansätze von Neville und Tim friedlich vereint, scheint uns die Tatsache, daß ja der Mittelwert aus jenen Daten künstlich gebildet wurde, die der "demokratische" Ansatz als einzige verwertet. Eine allgemeine Beweisführung, bei welcher man $\bar{x}$ als

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

einbringt, zeigt die "Richtigkeit" beider Ansätze sofort.