Welche Spannweite hat eine Zahlenmenge ?

David Harding,
Solihull School, England
Übersetzung: Agnes Richard, Kronberg im Taunus

Zusammenfassung: Sind nur arithmetisches Mittel und Standardabweichung einer monoton steigenden Zahlenfolge des Umfangs n bekannt, so kann die n-te Zahl nicht beliebig groß sein. Die Schranken für die Spannweite werden angegeben.

Neulich unterlief mit beim Unterrichten in einer 11. Klasse ein Fehler in der Aufgabenstellung, die wie folgt lautete:

„Für 20 Zahlen errechnete ich den Mittelwert und die Standardabweichung zu 10 resp. 4. Unglücklicherweise ging der Zettel, auf dem ich die Zahlen aufgeschrieben hatte, verloren, aber meiner Erinnerung nach war die größte Zahl die 30. Aufgabenstellung: Eliminiert diesen größten Wert aus der Zahlenmenge, und errechnet den neuen Mittelwert und die neue Standardabweichung."

Nennen wir die unbekannten Zahlen $x_1, x_2, \ldots, x_{20}$, dann ergibt sich

$\frac{\sum x_j}{20} = 10$, also $\sum x_j = 200$.

Für den Mittelwert der 19 Zahlen gilt dann $\frac{200 - 30}{19} = 8{,}947$.

Für die zugehörige Varianz erhält man

$4^2 = \frac{\sum x_i^2}{20} - 10^2 \Leftrightarrow \sum x_j^2 = 2320$,

woraus für die Varianz der 19 ersten Zahlen $\frac{2320 - 30^2}{19} - \left(\frac{170}{19}\right)^2 = -5{,}31$ folgt,

was wegen des negativen Vorzeichens unmöglich ist!

Daraus schloß ich, daß der Wert 30 für die größte Zahl zu extrem angesetzt ist, ich mußte mich also geirrt haben.

Ich stellte mir die Frage: „Wenn (nur) das arithmetische Mittel und die Standardabweichung einer Zahlenmenge gegeben sind, welcher größte und welcher kleinste Wert ergibt sich daraus?"

Angenommen, eine Population bestehe aus n Zahlen ($x_1, x_2, \ldots, x_n$) mit Mittelwert m und Standardabweichung s. Ein bestimmter Wert von X sei x_i. Dann ist der Mittelwert des 'Restes' $\frac{n \cdot m - X}{n - 1}$. Also haben wir $s^2 = \frac{\sum x^2}{n} - m^2$. Das heißt,

Stochastik in der Schule 17(1997), Heft 2, S. 25-26
Übersetzung aus Teaching Statistics 18(Autumn 1996)3, S. 81

die neue Varianz wird durch $\frac{(m^2 + s^2)n - X^2}{n - 1} - \left(\frac{nm - X}{n - 1}\right)^2$ bestimmt, dies ein Term, der nicht negativ sein kann.

Äquivalenzumformungen ergeben

$[(m^2 + s^2)n - X^2](n - 1) \geq (nm - X)^2$
$X^2 - 2mX + s^2 - s^2n + m^2 \leq 0$
$(X - m)^2 \leq s^2(n - 1)$
$-\sqrt{n - 1} \leq \frac{X - m}{s} \leq \sqrt{n - 1}$,

welches die möglichen Extrema der Zahlenreihe sind.

Für $n = 20$, $m = 10$, $s = 4$ erhalten wir $-\sqrt{19} \leq \frac{X - 10}{4} \leq \sqrt{19}$,

woraus $-7{,}4 \leq X \leq 27{,}4$ folgt
(was mein oben angenommenes Maximum von 30 als zu hoch entlarvt).

Für $n = 10$ ergäbe sich $\left|\frac{X - m}{s}\right| \leq 3$,

was bedeutet, daß für Datenmengen des Umfangs 10 die Spannweite höchstens 6 Standardabweichungen beträgt.

Obwohl es offensichtlich ist, daß die Spannweite, die Standardabweichung und das arithmetische Mittel miteinander in Beziehung stehen, hatte ich diese Konsequenz zuvor nicht gesehen und war zudem von der einfachen Form überrascht - speziell, was den Normalisierungs-Term angeht.

So zeigte der Beweis nicht nur, daß ich mich irrte, sondern es ergab sich zusätzlich eine äußerst interessante Erkenntnis.

Anmerkung der Herausgeberin

In Lothar Sachs: *Angewandte Statistik*, Springer-Verlag [7]1992, S. 164f findet sich auszugsweise folgendes:

Die Spannweite gestattet nach $\frac{R}{2} \cdot \sqrt{\frac{n}{n - 1}} \geq s$ ... eine grobe Abschätzung der Standardabweichung, wenn nur die Spannweite bekannt ist und über die Form der Verteilung nichts ausgesagt werden kann.

Grobschätzung der Standardabweichung aus den Extremwerten hypothetischer Stichproben sehr großer Umfänge: Ist anzunehmen, daß die den Werten zugrunde liegende Verteilung durch eine Normalverteilung approximiert wird, dann läßt sich die Standardabweichung der Grundgesamtheit überschlagsmäßig nach $\hat{s} \approx \frac{R}{6}$ schätzen, da beim Vorliegen einer *Normalverteilung* der Variationsbereich 6σ bekanntlich 99,7% aller Werte umfaßt. Für **angenähert normalverteilte Beobachtungen** läßt sich anhand der Spannweite R schnell überschlagsmäßig feststellen, ob die berechnete Standardabweichung s die richtige Größenordnung aufweist:

n	$\leqq 12$	20-40	≈ 100	> 400
$\hat{s}$	$R/\sqrt{n}$	$R/4$	$R/5$	$R/6$