ABSCHÄTZUNGEN FÜR ABWEICHUNGSMASZE

von A.V Boyd, Johannesburg

Originaltitel in "Teaching Statistics" Vol. 7 (1985):
The Standard Deviation and Absolute Deviations from the Mean

Übersetzung: Hans-Joachim Bentz

Zusammenfassung: Es gibt eine Reihe von Faustregeln, mit denen die Standardabweichung $\hat{\sigma}$ einer Stichprobe rasch abgeschätzt werden kann. Im vorliegenden Beitrag wird ein brauchbares Intervall für $\hat{\sigma}$ abgeleitet, dessen Grenzen sich aus einfachen Kenngrößen der Daten schnell ermitteln lassen.

1. Vorbemerkungen

In früheren Heften von "Teaching Statistics" haben Shiffler und Harsha (1980) sowie Macleod und Henderson (1984) Schranken für die Standardabweichung einer Stichprobe gegeben, und zwar in Abhängigkeit von der Spannweite und der mittleren linearen Abweichung.

$X_1, X_2, \ldots, X_n$, mit $X_1 \le X_2 \le \ldots \le X_n$, sei die der Größe nach geordnete Stichprobe. Weiter sei

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}, \quad (n \ge 2),$$ [1]

$$R = X_n - X_1$$

$$MLA = \frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}|$$

[1] Sowohl in den zitierten Artikeln als auch bei Boyd werden jeweils für $\hat{\sigma}$ als auch für die "eigentliche" Standardabweichung

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

Ungleichungen hergeleitet. Diese ergeben sich in trivialer Weise aus den hier angeführten Beziehungen, weil

$$s = \sqrt{\frac{n}{n-1}}\, \hat{\sigma}$$

gilt.

Stochastik in der Schule Heft 2, 6. Jg. (1986)

Shiffler und Harsha (1980) geben eine zweiseitige Abschätzung für $\hat{\sigma}$ mittels mittlerer linearer Abweichung und Spannweite:

$$\mathrm{MLA} \le \hat{\sigma} \le \frac{R}{2}. \tag{1}$$

Macleod und Henderson (1984) verwenden die Spannweite zu einer Abschätzung von $\hat{\sigma}$ nach unten:

$$\frac{R}{\sqrt{2n}} \le \hat{\sigma}. \tag{2}$$

In beiden Arbeiten wird der Nutzen solcher Ungleichungen für die praktische numerische Abschätzung der berechneten Standardabweichung von Daten diskutiert.

In der vorliegenden Arbeit werden mit Hilfe zweier anderer, einfacher Kenngrößen weitere Ungleichungen für die Standardabweichung abgeleitet.

2. Die Abschätzung für $\hat{\sigma}$

Definition: Die linearen Abweichungen vom Mittelwert sind $X_1-\bar{X}$, $X_2-\bar{X}$,... ..., $X_n-\bar{X}$. Bei der vorgegebenen Anordnung der Daten können $|X_1-\bar{X}|$ und $|X_n-\bar{X}|$ als Extremwerte der absoluten linearen Abweichung bezeichnet werden. Das Maximum bzw. Minimum der extremen absoluten Abweichungen wird dann definiert als:

$$\mathrm{Maxead}^{1)} := \max\{|X_1-\bar{X}|, |X_n-\bar{X}|\} \tag{3}$$

$$\mathrm{Minead} := \min\{|X_1-\bar{X}|, |X_n-\bar{X}|\} \tag{4}$$

Satz: Es gilt folgende Doppelungleichung für $\hat{\sigma}$:

$$\frac{\mathrm{Maxead}}{\sqrt{n-1}} \le \hat{\sigma} \le \sqrt{n-1}\ \mathrm{Minead} \tag{5}$$

Bemerkung: An einem Beispiel wird gezeigt werden, daß diese Ungleichung (5) bessere Abschätzungen für $\hat{\sigma}$ liefert als (1) und (2), falls die Verteilung der Daten schief ist.

1) Maxead: maximimum extreme absolute deviation.

Beweis:

(a) Erste Ungleichung

Für i = 1,...,n setze $x_i = X_i-\bar{X}$, demnach ist $\sum_{i=1}^{n} x_i = 0$.
Weiter ist:

$$\sum_{i=2}^{n} \left(x_i + \frac{x_1}{n-1}\right)^2 \ge 0$$

$$\Rightarrow \sum_{i=2}^{n} x_i^2 + \frac{2x_1}{n-1}\sum_{i=2}^{n} x_i + \left(\frac{x_1}{n-1}\right)^2 \sum_{i=2}^{n} 1 \ge 0$$

$$\Rightarrow \left\{\sum_{i=1}^{n} x_i^2 - x_1^2\right\} + \frac{2x_1}{n-1}(-x_1) + \frac{x_1^2}{n-1} \ge 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i^2 - \frac{n}{n-1} x_1^2 \ge 0.$$

Division durch n und Umordnen ergibt

$$\hat{\sigma}^2 \ge \frac{x_1^2}{n-1}.$$

Wegen der Anordnung der Stichprobenwerte ergibt sich $x_1 \le x_2 \le \dots \le x_n$; wegen $\sum x_i = 0$ folgt $x_1 \le 0$. Also hat man:

$$\hat{\sigma} \ge \sqrt{\frac{x_1^2}{n-1}} = \frac{-x_1}{\sqrt{n-1}} = \frac{\bar{X}-X_1}{\sqrt{n-1}}.$$

Geht man von

$$\sum_{i=1}^{n-1} \left(x_i + \frac{x_n}{n-1}\right)^2 \ge 0$$

aus, dann ergibt ein ähnliches Argument die Ungleichung

$$\hat{\sigma}^2 \ge \frac{x_n^2}{n-1},$$

sodaß wegen $x_n \ge 0$ gilt:

$$\hat{\sigma} \ge \frac{x_n}{\sqrt{n-1}} = \frac{X_n-\bar{X}}{\sqrt{n-1}}.$$

Weil ferner $\bar{X}-X_1 = |X_1-\bar{X}|$ und $X_n-\bar{X} = |X_n-\bar{X}|$ gilt, folgt weiters:

$$\hat{\sigma} \ge \frac{1}{\sqrt{n-1}} \max\{\bar{X}-X_1, X_n-\bar{X}\} = \frac{1}{\sqrt{n-1}} \text{ Maxead.}$$

(b) Zweite Ungleichung

Angenommen es seien k der Zahlen $x_1, x_2, \ldots, x_n$ größer als 0, dann folgt wegen $\sum x_i = 0$, daß k nicht größer sein kann als n-1, und die Zahlen $x_1, x_2, \ldots, x_{n-k}$ müssen alle ≤ 0 sein. Falls $k < n-1$ so enthält die Menge $L = \{1,2,\ldots,n-k\}$ mindestens zwei Elemente, und die Summe aller Produkte der Form $2x_ix_j$, wo x_i, x_j verschiedene Zahlen aus L sind, muß daher ≥ 0 sein.

Jetzt gilt:

$$n\hat{\sigma}^2 = \sum_{i=1}^{n} x_i^2$$

$$\Rightarrow n\hat{\sigma}^2 = \sum_{i=1}^{n-k} x_i^2 + \sum_{i=n-k+1}^{n} x_i^2$$

$$\Rightarrow n\hat{\sigma}^2 \le \left\{\sum_{i=1}^{n-k} x_i^2 + \sum_{\substack{i,j\in L\\ i\neq j}} 2x_ix_j\right\} + \sum_{i=n-k+1}^{n} x_i^2$$

$$= \left\{\sum_{i=1}^{n-k} x_i\right\}^2 + \sum_{i=n-k+1}^{n} x_i^2$$

$$= \left\{-\sum_{i=n-k+1}^{n} x_i\right\}^2 + \sum_{i=n-k+1}^{n} x_i^2$$

$$= \left\{\sum_{i=n-k+1}^{n} x_i\right\}^2 + \sum_{i=n-k+1}^{n} x_i^2$$

$$\Rightarrow n\hat{\sigma}^2 \le \left\{\sum_{i=n-k+1}^{n} x_n\right\}^2 + \sum_{i=n-k+1}^{n} x_n^2$$

$$= k^2x_n^2 + kx_n^2$$

$$= k(k+1)x_n^2$$

$$\Rightarrow n\hat{\sigma}^2 \le (n-1)nx_n^2.$$

Die gleiche Schlußkette ergibt sich auch bei $k = n-1$, der Term $\sum 2x_ix_j$ fehlt dabei einfach. Also gilt

$$\hat{\sigma} \le \sqrt{n-1}\, x_n = \sqrt{n-1}\,(X_n-\bar{X}).$$

In ähnlicher Weise zeigt man

$$\hat{\sigma} \le \sqrt{n-1}\,|x_1| = \sqrt{n-1}\,(\bar{X}-X_1).$$

Insgesamt hat man:

$$\hat{\sigma} \le \sqrt{n-1}\ \min\{\bar{X}-X_1, X_n-\bar{X}\} = \sqrt{n-1}\ \text{Minead.}$$

3. Ein numerisches Beispiel.

Im Beispiel von Macleod und Henderson (1984) waren die Daten

41, 42, 54, 42 und 51.

Demgemäß ist $\bar{X} = 46$ und $x_1 = -5$, $x_2 = x_3 = -4$, $x_4 = 5$, $x_5 = 8$.

Die Ungleichung (1) von Shiffler und Harsha (1980) ergibt

$$5{,}2 \le \hat{\sigma} \le 6{,}5,$$

die Ungleichung (2) von Macleod und Henderson (1984) liefert die untere Schranke

$$4{,}11 \le \hat{\sigma}.$$

Aus der Ungleichung (5) erhalten wir

$$4 \le \hat{\sigma} \le 10.$$

Die Beziehung (5) liefert in diesem Fall eine schlechte Abschätzung. Erhöht sich aber der größte Stichprobenwert von 54 auf 124, die Stichprobe erhält dadurch eine schiefe Verteilung, dann liefern die Ungleichungen folgende Abschätzungen:

(1) $25{,}6 \le \hat{\sigma} \le 41{,}3$,
(2) $26{,}25 \le \hat{\sigma}$,
(5) $32 \le \hat{\sigma} \le 38$.

Dies belegt, daß (5) nun bessere Abschätzungen für $\hat{\sigma}$ liefert als (1) und (2).

Falls die Stichprobe aus $X_1 = 1-n$, $X_2 = \ldots = X_n = 1$ besteht, vereinfacht sich (5) zu einer Gleichung.

Literatur

MACLEOD, A.J. u. G.R. HENDERSON: Bounds for the Sample Standard Deviation. Teaching Statistics 6 (1984), 72-76. Übersetzung in: Stochastik in der Schule 5 (1985), Heft 2, 32-36.

SHIFFLER, R.E u. P.D. HARSHA: Upper and Lower Bounds for the Sample Standard Deviation, Teaching Statistics 2 (1980), 84-86 .

Prof. Dr. A.V. Boyd
Statistics Department
University of the Witwatersrand
1 Jan Smuts Avenue
Johannesburg 2001
South Africa