Beispiel:

Die Messung der Durchmesser einer Stichprobe von Stahlkugeln ergibt folgende Häufigkeitstabelle:

Durchmesser (mm)	Häufigkeit
45,05-45,55	2
45,55-46,05	4
46,05-46,55	7
46,55-47,05	6
47,05-47,55	4
47,55-48,05	1
48,05-48,55	1

Für dieses Beispiel ergeben sich $\hat{\sigma} = 0{,}72$ und für die beiden Schranken die Werte 0,35 und 1,25.
Das entsprechende Ergebnis zu (5) ist

$$\frac{(k-2)d}{\sqrt{2(n-1)}} \leq s \leq \frac{kd}{2} \cdot \sqrt{\frac{n}{n-1}} , \qquad (7)$$

und für obige Daten ergibt sich $0{,}22 \leq 0{,}74 \leq 1{,}28$.

Die Autoren danken John HAYWARD und Alan HOOD für ihre hilfreichen Anmerkungen sowie dem verantwortlichen Redakteur für einige Vorschläge, die zur Verbesserung der ursprünglichen Fassung beitrugen.

LITERATUR

[1] SHIFFLER, R.E. und HARSHA, P.D. (1980)
Upper and Lower Bounds for the Sample Standard Deviation,
Teaching Statistics, 2.3, 84-86

Zur Einführung der Standardabweichung

von S.J. Wainwright, University College of Swansea

Originaltitel in "Teaching Statistics" Vol.6 (1984):
How Should We Teach the Standard Deviation? - Revisited

Übersetzung: Andreas Horn

Zwei Wege zur Einführung der Standardabweichung bei Studenten untersucht HART (1984) in [1] und kommt zu dem Schluß, daß eine geeignete Methode der Einführung der Standardabweichung bei der Berechnung der Abweichung vom Mittelwert ansetzen kann, wobei Vor- und Nachteile aufgezeigt werden. Doch kann auch ein Einstieg über den Schätzwert eines zentralen Lagemaßes sinnvoll sein.

Geht man von einem beliebigen zentralen Lagemaß F für die Stichprobenwerte Y_i aus, so kann zunächst das 'Residuum'

$$r_i = Y_i - F \qquad \text{eingeführt werden.}$$

Angewandt auf die häufigsten Lagemaße Zentralwert (Median) Y und Mittelwert $\bar{Y}$ erhält man für die Residuen

$$r_i = Y_i - Y \qquad \text{bzw.} \qquad r_i = Y_i - \bar{Y} .$$

Fragt man nun nach dem 'besten Schätzwert' für die zentrale Lage, so stellt sich heraus, daß dies im wesentlichen eine Frage der Definition ist. Denn je nachdem, ob man für die Summe der absoluten Residuen oder für die Summe der quadrierten Residuen das Minimum zu erreichen sucht, erweisen sich einmal der Zentralwert und im anderen Fall der Mittelwert als bestes Lagemaß.

Ein Beispiel mit kleinem Stichprobenumfang kann empirisch zeigen, daß im Vergleich mit jedem anderen willkürlichen Schätzwert für den Zentralwert $\sum |r_i|$ und für den Mittelwert $\sum r_i^2$ minimal wird. An dieser Stelle kann die Frage der Genauigkeit und Verläßlichkeit eines Schätzwertes auch gegenüber Ausreißern angesprochen werden.

Studenten begreifen im allgemeinen sehr schnell, daß ein Schätzwert, für den die Summe der quadrierten Residuen minimal wird, größere und damit bedeutendere Abweichungen stärker berücksichtigt als kleinere. Bei dieser Gelegenheit kann auch darauf hingewiesen werden, daß Daten nicht nur um ihre zentrale Lage schwanken, sondern daß die 'Streuung' auch von gewissem Grad oder Ausmaß ist.

Klar ist, daß die Summe der quadrierten Residuen $\sum (Y_i-\bar{Y})^2$ ein Maß für diese Streuung ist. Mit $\bar{Y}$ als Schätzwert für die zentrale Lage wird $\sum r_i^2$ minimal, und $\sum (Y_i-\bar{Y})^2$ kann als Maß für die Streuung dienen, doch hängt die Größe dieser Summe sowohl vom Umfang der Stichprobe als auch von der Streuung der Daten ab. Daher muß der Durchschnitt der quadrierten Residuen bestimmt werden, der nicht vom Stichprobenumfang, sondern nur von der Streuung der Daten abhängt. Die Summe der quadrierten Residuen ist also durch die Anzahl der unabhängigen Residuen zu teilen. (Die Tatsache, daß schließlich durch n-1 und nicht durch n dividiert wird, kann an dieser Stelle übergangen oder nur am Rande erwähnt werden, denn für $n \longrightarrow \infty$ strebt $n-1 \longrightarrow n$).

Jedenfalls steht nun als Maß für die Streuung ein Mittelwert von Quadraten zur Verfügung, der vom Stichprobenumfang unabhängig ist, wobei allerdings das Problem auftritt, daß die Einheit quadriert erscheint. Ist z.B. die Einheit des Mittelwerts cm, so treten jetzt cm^2 auf. Um dieselbe Einheit wie beim Mittelwert zu erhalten, ist nun noch die Wurzel zu ziehen, womit sich die Standardabweichung ergibt.

An dieser Stelle kann die Normalverteilung eingeführt und der Zusammenhang zwischen Standardabweichung und Normalverteilung angesprochen werden.

Ähnliche Überlegungen für den Zentralwert als Schätzwert für die zentrale Lage ergeben als geeignetes Maß für die Streuung die betragsmäßige Abweichung vom Zentralwert.

LITERATUR

[1] HART, A.E. (1984)
How should We Teach the Standard Deviation, Teaching Statistics, 6,24

Zwei Ungleichungen von Markow und Tschebyscheff

von J. Gani, University of Kentucky

Originaltitel in "Teaching Statistics" Vol.6 (1984):
Markov to Chebyshev - Some Useful Inequalities

Übersetzung: Andreas Horn

Viele Zufallsgrößen des täglichen Lebens sind positiv, unter anderem die Körpergröße oder das Körpergewicht von Personen. Beispielsweise kann es nützlich sein, auch ohne Kenntnis der zugrunde liegenden Wahrscheinlichkeitsverteilung Aussagen über die Wahrscheinlichkeit darüber zu machen, daß die Werte von einzelnen Personen oberhalb einer gewissen Grenze liegen.

Betrachtet man die immer positive Zufallsgröße 'Körpergewicht' und bezeichnet diese mit W, so lautet die Ungleichung von Markow für einen speziellen Wert w_0 des Körpergewichts

$$P(W \geq w_0) \leq \frac{\mu}{w_0} \qquad (1)$$

wobei $E(W) = \mu$ der Erwartungswert dieser Zufallsgröße ist. Natürlich hat dieses Ergebnis nur für Werte $w_0 > \mu$ Bedeutung. Der Beweis dieser Ungleichung ist sehr einfach.
Angenommen, W ist eine stetige Zufallsgröße mit der Wahrscheinlichkeits-Dichtefunktion f(w), so gilt

$$\mu = E(W) = \int_0^\infty wf(w)dw \geq \int_{w_0}^\infty wf(w)dw$$
$$\geq w_0 \int_{w_0}^\infty f(w)dw = w_0 P(W \geq w_0) \qquad (2)$$

und damit auch (1). Ein ähnlicher Beweis kann für diskrete Zufallsgrößen geführt werden (Übungsaufgabe).

Legt man z.B. die Verteilung der Körpergewichte bei Männern der Größe 177 cm mit dem arithmetischen Mittel 78,7 kg zugrunde, wie sie eine Untersuchung der Universität von Kentucky 1981 ergab, so können 90,7 kg eines Mannes dieser Größe bereits als beträchtliches Übergewicht angesehen werden. Ohne Kenntnis der Verteilung der Körpergewichte ist nun