ihre zentrale Lage schwanken, sondern daß die 'Streuung' auch von gewissem Grad oder Ausmaß ist.

Klar ist, daß die Summe der quadrierten Residuen $\sum (Y_i-\overline{Y})^2$ ein Maß für diese Streuung ist. Mit $\overline{Y}$ als Schätzwert für die zentrale Lage wird $\sum r_i^2$ minimal, und $\sum (Y_i-\overline{Y})^2$ 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 folgende Aussage möglich:

$$P(W \geq 90{,}7) \leq \frac{\mu}{90{,}7} = \frac{78{,}7}{90{,}7} = 0{,}87. \qquad (3)$$

Diese nicht sehr genaue Abschätzung kann mit Hilfe der Tschebyscheffschen Ungleichung verbessert werden. Insbesondere ermöglicht die Markowsche Ungleichung (1) auch einen einfachen Beweis des Ergebnisses von Tschebyscheff:
Hat eine Zufallsgröße X ($-\infty < x < \infty$) den bekannten Erwartungswert $E(X) = \mu$ und die Varianz $V(X) = \sigma^2$, so gilt

$$P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2} \qquad (4)$$

(für jede positive reelle Zahl λ).

Um die Markowsche Ungleichung anwenden zu können, betrachtet man die positive Zufallsgröße $Y = (X - \mu)^2$ mit dem Erwartungswert $E(Y) = E(X-\mu)^2 = \sigma^2$.
Mit (1) folgt dann

$$P(Y \geq \lambda^2 \cdot \sigma^2) \leq \frac{\sigma^2}{\lambda^2 \cdot \sigma^2} = \frac{1}{\lambda^2}, \qquad (5)$$

und damit (4), weil $(X-\mu)^2 \geq \lambda^2 \cdot \sigma^2$ und $|X - \mu| \geq \lambda\sigma$ äquivalent sind.

Dieser Weg zur Tschebyscheffschen Ungleichung dürfte manchem Lehrer einfacher erscheinen als der sonst übliche Beweis.

Nimmt man für die Standardabweichung σ des Körpergewichts bei Männern der Größe 177 cm den Wert 7 kg an, so liefert die Anwendung von (4)

$$P(|W - 78{,}7| \geq 7\lambda) \leq \frac{1}{\lambda^2}. \qquad (6)$$

Für die Wahrscheinlichkeit dafür, daß ein Mann dieser Größe mindestens 90,7 kg oder aber höchstens 66,7 kg wiegt, ergibt sich mit $\lambda = \frac{12}{7} = 1{,}714$ aus (6)

$$P(|W - 78{,}7| \geq 12) \leq \frac{49}{144} = 0{,}34, \qquad (7)$$

was weit unter dem in (3) aus der Markowschen Ungleichung gewonnenen Wert 0,87 liegt.

Je mehr also über die Verteilung der Zufallsgröße bekannt ist, desto besser wird auch die Abschätzung für $P(W \geq 90{,}7)$.

Geht man von einer Normalverteilung der Körpergröße mit arithmetischem Mittel 78,7 kg und Standardabweichung 7 kg aus, so ergibt sich der exakte Wert

$$P(W \geq 90{,}7) = 0{,}043,$$

der etwa ein Achtel der in (7) mit der Tschebyscheffschen und ein Zwanzigstel der in (3) mit der Markowschen Ungleichung erhaltenen Wahrscheinlichkeit beträgt. Eine offensichtliche Überschätzung der Wahrscheinlichkeiten bei der Berechnung nach Markow bzw. Tschebyscheff kann und muß angesichts der geringen Voraussetzungen in Kauf genommen werden.