Varianzschätzung bei korrelierten Beobachtungen

- Vorsicht σ Falle -

von **Sonja Michels**, Dortmund

Zusammenfassung: Dieser Beitrag gibt einfache untere und obere Schranken für den Erwartungswert der Stichprobenvarianz bei korrelierten Beobachtungen.

Einfache Stichprobe

Bekanntlich ist die Stichprobenvarianz $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ bei unabhängigen Zufallsvariablen $Y_1, \ldots, Y_n$ eine erwartungstreue Schätzung für $\sigma^2 = E(Y - EY)^2$. Gilt das gleiche auch bei korrelierten Y_i?

Die Antwort ist nein. Wie man leicht sieht, haben wir im allgemeinen Fall:

$$\begin{aligned} Es^2 &= \frac{1}{n-1}\left\{E(\sum_{i=1}^{n}(y_i - \bar{y})^2)\right\} \\ &= \frac{1}{n-1}\left\{E(\sum_{i=1}^{n}(y_i - \mu)^2 - n(\bar{y} - \mu)^2)\right\} \\ &= \frac{1}{n-1}\left\{n\sigma^2 - n\ Var(\bar{y})\right\} \\ &\leq \frac{n}{n-1}\sigma^2. \end{aligned}$$

Dabei bezeichnet μ den Erwartungswert der Y_i. Wir erhalten diese Abschätzung, da die Varianz von $\bar{y}$ zwischen 0 und σ^2 liegen muß. Der erste Extremfall tritt ein, wenn die Varianz der zugrundeliegenden Zufallsvariablen 0 ist, und der zweite, wenn die Korrelation zwischen den Y_i gleich 1 ist, d.h. wenn sie Kopien voneinander sind.

Es gilt somit

$$0 \leq Es^2 \leq \frac{n}{n-1}\sigma^2.$$

Aus dieser Abschätzung geht hervor, daß die empirische Varianz in diesem Fall kein geeigneter Indikator ist. Hinter einer kleinen Varianzschätzung verbirgt sich möglicherweise eine Korrelation der zugrundeliegenden Zufallsvariablen.

Lineare Regression

Auch für kompliziertere Fälle gibt es ähnliche Schranken für s^2. Betrachten wir eine lineare Regression mit k Regressoren und n Beobachtungen. Im Modell $y = X\beta + u$ bezeichnen y die abhängige Variable, X die Regressormatrix, β den Parametervektor sowie u den Störgrößenvektor. Die Komponenten des Störvektors sollen den gleichen Erwartungswert 0 und die gleiche Varianz σ^2 besitzen. Falls sie unkorreliert sind, ist

$$s^2 = \frac{\hat{u}'\hat{u}}{n-k}$$

mit $\hat{u} = y - X\hat{\beta}$ ein unverzerrter Schätzer für σ^2. Mit einem kleinen Umweg über die Matrixalgebra erhalten wir hier die folgende Ungleichungskette

$$0 \leq Es^2 \leq \frac{n}{n-p}\sigma^2.$$

Auf den Beweis will ich an dieser Stelle verzichten. Nachzulesen ist er z.B. bei Dufour(1986). Die Schranken sind von derselben Struktur wie im Fall der einfachen Stichprobe und so bleiben auch die oben genannten Konsequenzen. Hier sollte allerdings noch ergänzt werden, daß auch das Bestimmtheitsmaß – in das die Fehlerquadratsumme direkt eingeht – mit dem entsprechenden Mißtrauen behandelt werden sollte.

Literatur

David, H.A. (1985): Bias of s^2 under dependence, *The American Statistician, 39, S. 201.*

Dufour, J.-M. (1986): Bias of s^2 in linear regression with dependent errors, *The American Statistician, 40, S. 284-285.*