Schichten mit geschätzten Schichtgrößen II

(Eine Weiterentwicklung und eine Korrektur zum Artikel des gleichen Autors im Heft 1/1996 der *Österreichischen Zeitschrift für Statistik*)

Andreas Quatember
Institut für Angewandte Statistik, Universität Linz

Quatember (1996) untersucht die statistischen Eigenschaften der Schätzer $\bar{x}$ in geschichteten Zufallsauswahlen. Die Varianz des Schätzers $\bar{x}$ beträgt bei proportionaler Allokation des Stichprobenumfanges auf die Schichten und Ziehen mit Zurücklegen (für die Notationen siehe: Quatember, 1996)

a) bei Verwendung der wahren Schichtanteile π_i (i=1,2,...,r) des Schichtmerkmals für eine geschichtete Zufallsauswahl vom Umfang n_S:

$$\operatorname{Var} \bar{x}_S = \frac{1}{n_S} \sum_{i=1}^{r} \pi_i \, \sigma_i^2 \quad , \tag{1}$$

b) bei Verwendung geschätzter Schichtanteile für eine geschichtete Zufallsauswahl vom Umfang n_{gS}, wenn die Schätzung der r Schichtgrößen auf Basis einer uneingeschränkten Zufallsauswahl vom Umfang n_Z erfolgt ist:

$$\operatorname{Var} \bar{x}_{gS,Z} = \frac{1}{n_{gS}} \sum_{i=1}^{r} \pi_i \, \sigma_i^2 + \frac{1}{n_Z} \sum_{i=1}^{r} \pi_i \, (\mu_i - \mu)^2 \quad , \tag{2}$$

c) bei Verwendung geschätzter Schichtanteile für eine geschichtete Zufallsauswahl vom Umfang n_{gS}, wenn die Schätzung der r Schichtgrößen auf Basis einer geschichteten Zufallsauswahl vom Umfang n_S erfolgt ist:

$$\operatorname{Var} \bar{x}_{gS,S} = \frac{1}{n_{gS}} \sum_{i=1}^{r} \pi_i \, \sigma_i^2 + \frac{1}{n_S} \left(\sum_{i=1}^{r} \pi_i \mu_i^2 - \sum_{k=1}^{s} \left(\sum_{i=1}^{r} \mu_i \pi_{i|k} \right)^2 \frac{N_k}{N} \right) . \tag{3}$$

(3) ist dabei eine Weiterentwicklung der Varianzformel aus Quatember (1996), denn der dort in (9) auf S.51 angegebene Subtrahend

$$\sum_i \sum_j \mu_i \mu_j \left(\sum_k \frac{N_k}{N} \pi_{i|k} \pi_{j|k} \right)$$

läßt sich darstellen als

$$\begin{aligned} \sum_i \sum_j \mu_i \mu_j \left(\sum_k \frac{N_k}{N} \pi_{i|k} \pi_{j|k} \right) &= \sum_k \frac{N_k}{N} \sum_i \mu_i \pi_{i|k} \sum_j \mu_j \pi_{j|k} \\ &= \sum_k \left(\sum_i \mu_i \pi_{i|k} \right)^2 \frac{N_k}{N} \quad , \end{aligned}$$

woraus sich (3) ergibt.

Der durch diese Vorgangsweise entstehende Schichtungseffekt $\Delta(\mathrm{Var})$ läßt sich mit (3) nun folgendermaßen darstellen:

$$\Delta(\mathrm{Var}) = \mathrm{Var}\ \overline{x}_Z - \mathrm{Var}\ \overline{x}_{gS,S}$$

$$= \frac{1}{n_{gS}} \sum_i \pi_i (\mu_i - \mu)^2 - \frac{1}{n_S} \left(\sum_i \pi_i \mu_i^2 - \sum_k \left(\sum_i \mu_i \pi_{i|k} \right)^2 \frac{N_k}{N} \right) . \tag{4}$$

Der Schichtungseffekt (4) ist bei identischen Umfängen n_S und n_{gS} dann gleich null, wenn in der Basisstichprobe kein Schichtungseffekt erzielt wird, wenn also für alle k Schichten $\pi_{i|k} = \pi_i$ ist. Für diesen Fall gilt nun:

$$\sum_k \left(\sum_i \mu_i \pi_{i|k} \right)^2 \frac{N_k}{N} = \mu^2 , \tag{5}$$

und damit wird die Varianz (3) wieder zu:

$$\mathrm{Var}\ \overline{x}_{gS,S} = \frac{1}{n_{gS}} \sum_i \pi_i\ \sigma_i^2 + \frac{1}{n_{gs}} \sum_i \pi_i (\mu_i - \mu)^2 ,$$

der Varianz des Mittelwertschätzers einer uneingeschränkten Zufallsauswahl des gleichen Umfanges n_{gS} wie die Auswahl nach geschätzten Schichtgrößen. Sind jedoch nicht alle $\pi_{i|k} = \pi_i$, so besitzt die geschichtete Basisstichprobe einen positiven Schichtungseffekt. Denn es gilt in diesem Fall (**hier ist Quatember, 1996, auf S. 55 oben zu korrigieren**):

$$\mathrm{Var}\left(\sum_i \mu_i \pi_{i|k} \right) = E\left(\sum_i \mu_i \pi_{i|k} \right)^2 - E^2\left(\sum_i \mu_i \pi_{i|k} \right) > 0$$

und somit

$$E\left(\sum_i \mu_i \pi_{i|k} \right)^2 > \mu^2 .$$

Daraus folgt nun:

$$\sum_k \left(\sum_i \mu_i \pi_{i|k} \right)^2 \frac{N_k}{N} > \mu^2 \tag{5'}$$

und schließlich:

$$\mathrm{Var}\ \overline{x}_{gS,S} < \mathrm{Var}\ \overline{x}_Z .$$

Literatur

QUATEMBER A. (1996). Schichten mit geschätzten Schichtgrößen. *Österreichische Zeitschrift für Statistik*, 25. Jahrgang, Heft 1, 43-58.