# Handout: Proof that $s^2_{y,S}$ is unbiased for $s^2_{y,U}$

First, definitions

$$\bar{y}_U \;=\; \frac{1}{N}\sum_{i=1}^{N} y_i \;, \quad \bar{y}_S \;=\; \frac{1}{n}\sum_{i\in S} y_i \;=\; \frac{1}{n}\sum_{i=1}^{N} I_{[i\in S]}\, y_i$$

$$s^2_{y,U} \;=\; \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y}_U)^2 \;=\; \frac{1}{N-1}\sum_{i=1}^{N} y_i^2 \;-\; \frac{N}{N-1}\,\bar{y}_U^2$$

$$s^2_{y,S} \;=\; \frac{1}{n-1}\sum_{i\in S}(y_i-\bar{y}_S)^2 \;=\; \frac{1}{n-1}\sum_{i\in S} y_i^2 \;-\; \frac{n}{n-1}\,\bar{y}_S^2\}$$

Next, today's verification, under SRS, $n$ out of $N$ :

$$E(s^2_{y,S}) \;=\; E[\frac{1}{n-1}\sum_{i=1}^{N} I_{[i\in S]}\, y_i^2 \;-\; \frac{n}{n-1}\,\bar{y}_S^2]$$

$$= \frac{n}{N(n-1)}\sum_{i=1}^{N} y_i^2 \;-\; \frac{n}{n-1}\,[E(\{\bar{y}_S\}^2) \;+\; \mathrm{Var}(\bar{y}_S^2)]$$

and substituting our formula for theoretical variance $(1-n/N)\,s^2_{y,U}/n$ of $\bar{y}_S$:

$$= \frac{n}{N(n-1)}\,[\sum_{i=1}^{N} y_i^2 \;-\; N\,\bar{y}_U^2] \;-\; \frac{n}{n-1}\cdot\frac{(N-n)}{nN}\,s^2_{y,U}$$

$$= s^2_{y,U}\,\Big[\frac{n(N-1)}{N(n-1)} \;-\; \frac{(N-n)}{(n-1)N}\Big] \;=\; s^2_{y,U}$$

All this was under SRS sampling. If sampling was done equiprobably **with** replacement, then the estimator $\bar{W} \;=\; n^{-1}\sum_{j=1}^{n} W_j$, where the $W_j$'s are *iid* random variables equal to $y_i$ with probability $1/N$, for all $i = 1,\ldots,N$. So the mean of $W_1$ is $\bar{y}_U$, and the variance of $W_1$ is $N^{-1}\sum_{i=1}^{N}(y_i-\bar{y}_U)^2 \;=\; (1-1/N)\,s^2_{y,U}$, and

$$\text{in \textbf{with}-replacement sampling,} \qquad \mathrm{Var}(\bar{W}) \;=\; n^{-1}\,\mathrm{Var}W_1 \;=\; \frac{N-1}{nN}\,s^2_{y,U}$$

which is estimated unbiasedly by $1/n$ times the sample variance of the observed $W_j$ random variables.