STAT 403 Tutorial · Week 13 (April 1st, 2020)

Topics:

- Proportion estimator in stratified random sampling;
- Sample allocation methods (proportional allocation; Neyman allocation; optimal allocation).

Again, these concepts will be reviewed with an example.

Example (from the book of Sharon Lohr (2009)). Suppose that a city has 90,000 dwelling units, of which 35,000 ($\uparrow N_1$) are houses (h=1), 45,000 ($\uparrow N_2$) are apartments (h=2), and 10,000 ($\uparrow N_3$) are condominiums (h=3).

(a). You believe that the mean electricity usage is about twice as much for houses as for apartments or condominiums, and that the standard deviation is proportional to the mean so that $S_1 = 2S_2 = 2S_3$. How would you allocate a stratified sample of 900 observations if you wanted to estimate the mean electricity consumption for all households in the city?

Solution: Given the information on the number of units N_h and population variance S_h^2 of different strata, we can apply the Neyman allocation where the sample sizes for each stratum is proportional to $N_h S_h$ for $h = 1, 2, 3$. Therefore

$$n_1 = n \cdot \frac{N_1 S_1}{N_1 S_1 + N_2 S_2 + N_3 S_3} = 900 \times \frac{35 \times 2S_2}{35 \times 2S_2 + 45 \times S_2 + 10 \times S_2} = 504$$

$$n_2 = n \cdot \frac{N_2 S_2}{N_1 S_1 + N_2 S_2 + N_3 S_3} = 900 \times \frac{45 \times S_2}{35 \times 2S_2 + 45 \times S_2 + 10 \times S_2} = 324$$

$$n_3 = n \cdot \frac{N_3 S_3}{N_1 S_1 + N_2 S_2 + N_3 S_3} = 900 \times \frac{10 \times S_2}{35 \times 2S_2 + 45 \times S_2 + 10 \times S_2} = 72.$$

The sample sizes for the three strata are 504, 324 and 72, respectively.

(b). If, in addition to information in (a), it is known that the cost of investigating the electricity assumption of a house is also twice as much as that of an apartment or a condominium. How would you allocate the 900 observations now?

Solution : Let c_h be the cost of investigating a single unit in stratum h. Then we can use the optimal allocation where the sample sizes are proportional to $N_h S_h/\sqrt{c_h}$. Thus,

$$n_1 = 900 \times \frac{35\times 2/\sqrt{2}}{35\times 2/\sqrt{2} + 45\times 1/\sqrt{1} + 10\times 1/\sqrt{1}} = 426.30$$

$$n_2 = 900 \times \frac{45\times 1/\sqrt{1}}{35\times 2/\sqrt{2} + 45\times 1/\sqrt{1} + 10\times 1/\sqrt{1}} = 387.57$$

$$n_3 = 900 \times \frac{10\times 1/\sqrt{1}}{35\times 2/\sqrt{2} + 45\times 1/\sqrt{1} + 10\times 1/\sqrt{1}} = 86.13 .$$

By rounding to the nearest integer, we have

$$n_1 = 426 , \quad n_2 = 388 , \quad n_3 = 86 .$$

(c). Suppose that you take a stratified random sample with proportional allocation and want to estimate the overall proportion $\hat{p}$ of households in which energy conservation is practiced. If 45% of house dwellers, 25% of apartment dwellers, and 3% of condominium residents practice energy conservation, what is p for the population? What gain would the stratified sample with proportional allocation offer over an SRS, that is, what is $Var(\hat{p}_{str})/Var(\hat{p}_{SRS})$?

Solution: Let p_h be the proportion of households that practice energy conservation in stratum h. Then $p_1 = 45\%$, $p_2 = 25\%$, $p_3 = 3\%$. The total number of households that practice energy conservation is

$$T = p_1 N_1 + p_2 N_2 + p_3 N_3 = 45\% \times 35{,}000 + 25\% \times 45{,}000 + 3\% \times 10{,}000 = 27{,}300.$$

Then the population proportion is

$$p = T/N = 27{,}300/90{,}000 = 30.33\%.$$

(i) If we use an SRS of size n and estimate p by $\hat{p}_{SRS}$, then

$$Var(\hat{p}_{SRS}) = \left(1 - \frac{n}{N}\right) \cdot \frac{S^2}{n}, \quad S^2 = \frac{N}{N-1} p(1-p). \quad \text{(See tutorial notes of Week 11)}.$$

(ii) If we use a stratified sample of size n with proportional allocation, and estimate p by $\hat{p}_{str}$, we have

$$Var(\hat{p}_{str}) = \frac{1}{N^2} \sum_{h=1}^{3} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \cdot \frac{S_h^2}{n_h}, \quad S_h^2 = \frac{N_h}{N_h - 1} p_h (1-p_h).$$

Since we use proportional allocation, $\frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} = \frac{n}{N}$.

Therefore,

$$\frac{Var(\hat{p}_{str})}{Var(\hat{p}_{SRS})} = \frac{\sum_{h=1}^{3} \frac{N_h^2}{N^2} \cdot \frac{S_h^2}{n_h}}{\frac{1}{n} S^2}$$

$$= \frac{\sum_{h=1}^{3} \frac{N_h^2}{N^2} \cdot \frac{n}{n_h} \cdot S_h^2}{S^2}$$

$$= \frac{\sum_{h=1}^{3} \frac{N_h^2}{N^2} \cdot \frac{N}{N_h} \cdot S_h^2}{S^2}, \quad \because N/N_h = n/n_h \text{ by proportional allocation}$$

$$= \sum_{h=1}^{3} \frac{N_h}{N} \cdot S_h^2 \Big/ S^2 \approx \sum_{h=1}^{3} \frac{N_h}{N} p_h(1-p_h) \Big/ p(1-p), \quad \text{assume } \frac{N_h}{N_h - 1} \approx 1, \frac{N}{N-1} \approx 1$$

$$= \frac{\frac{35}{90} \times 45\% \times (1-45\%) + \frac{45}{90} \times 25\% \times (1-25\%) + \frac{10}{90} \times 3\% \times (1-3\%)}{30.3\% \times (1-30.3\%)}$$

$$\approx 91.5\%.$$