Regression Estimation

We have seen in our study of ratio estimation that it is appropriate for data where y and x are linearly related, with a line going through the origin. If the linear relationship between y and x does not go through the origin, a regression estimator may be more appropriate. Recall that a least-squares linear regression line is of the form: $\hat{y}_i = a + bx_i$, where a is the intercept and b is the slope. Since the least-squares estimates satisfy $a = \bar{y} - b\overline{x}$, we can substitute to get: $\hat{y}_i = \overline{y} + b(x_i - \bar{x})$. In a sampling problem we are interested in estimating μ_y, which we obtain from the regression line by substituting μ_x for x_i: $\widehat{\mu}_{yL} = \overline{y} + b(\mu_x - \bar{x})$, where $b = \sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x}) \,/\, \sum_{i=1}^{n}(x_i - \overline{x})^2$ and

$$\begin{aligned}\widehat{V}(\widehat{\mu}_{yL}) &= \left(\frac{N-n}{N}\right)\left(\frac{1}{n}\right)\left(\frac{1}{n-2}\right)\left[\sum_{i=1}^{n}(y_i - \overline{y})^2 - b^2\sum_{i=1}^{n}(x_i - \bar{x})^2\right] \\ &= \left(\frac{N-n}{N}\right)\left(\frac{1}{n}\right) MSE,\end{aligned}$$

where MSE is the mean-squared error of a simple linear regression of y on x.

See Example 6.9 in the text and the SAS code for it on the web.

Difference Estimation

Difference estimation is a special case of regression estimation where the slope b is set to 1. This gives $\widehat{\mu}_{yD} = \overline{y} + (\mu_x - \bar{x}) = \mu_x + \overline{d}$, where $\bar{d} = \bar{y} - \bar{x}$, with estimated variance of:

$$\widehat{V}(\widehat{\mu}_{yD}) = \left(\frac{N-n}{N}\right)\left(\frac{1}{n}\right)\frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1}, \text{ with } d_i = y_i - x_i.$$

See Example 6.10 in the text and SAS code for it on the web.