

Comparison of estimators

Here we discuss the choice of an estimator of the mean μ_y , comparing the ratio, regression, and difference estimators of the current chapter, and the SRS estimator from Chapter 4. We first consider the bias of the four methods, then we introduce the concept of relative efficiency as a way to compare variances of estimators.

Bias

The sample mean from Chapter 4 is unbiased with SRS, and the difference estimator is also unbiased with SRS. The regression estimator is biased in sampling from finite populations, but the bias is usually small if the relationship between y and x is linear. The ratio estimator $r = \bar{y}/\bar{x}$ is a biased estimator of $R = \mu_y/\mu_x$, and an approximation to its relative bias is:

$$\frac{E(r) - R}{R} \approx \left(\frac{N-n}{N}\right) \left(\frac{1}{n}\right) \left(\frac{s_x^2}{\bar{x}^2} - \hat{\rho} \frac{s_y}{\bar{y}} \frac{s_x}{\bar{x}}\right),$$

where $\hat{\rho}$ is the sample correlation between x and y . This quantity can be calculated for a given data set, and simulations can also be used to understand the bias of both the ratio and regression estimators. See the example in Table 6.12 in the book and the SAS example on the web.

Comparing variances: the concept of relative efficiency

If two estimators are unbiased or have sufficiently small bias, we then rely on variance comparisons to choose the best estimator. One way to make a variance comparison is through the concept of relative efficiency. For two estimators denoted by E_1 and E_2 , based on the same sample size, the relative efficiency of E_1 to E_2 is defined as:

$$RE\left(\frac{E_1}{E_2}\right) = \frac{V(E_2)}{V(E_1)}.$$

Thus if $RE(E_1/E_2) > 1$, estimator E_1 is favored. We will focus on estimated variances and use an estimated relative efficiency of:

$$\widehat{RE}\left(\frac{E_1}{E_2}\right) = \frac{\hat{V}(E_2)}{\hat{V}(E_1)}.$$

The comparisons in the text are based on the following estimated variances:

$$\begin{aligned} \hat{V}(\bar{y}) &= s_y^2 \left(\frac{N-n}{Nn}\right), \\ \hat{V}(\hat{\mu}_y) &= s_r^2 \left(\frac{N-n}{Nn}\right) = (s_y^2 + r^2 s_x^2 - 2r\hat{\rho}s_x s_y) \left(\frac{N-n}{Nn}\right), \\ \hat{V}(\hat{\mu}_{yL}) &\approx (s_y^2 - b^2 s_x^2) \left(\frac{N-n}{Nn}\right), \text{ and} \end{aligned}$$

$$\hat{V}(\hat{\mu}_{yD}) = (s_y^2 + s_x^2 - 2\hat{\rho}s_x s_y) \left(\frac{N-n}{Nn} \right),$$

for the sample mean of y , the ratio estimator, the regression estimator, and the difference estimator, respectively. After a few algebraic manipulations, the following expressions emerge:

$$\widehat{RE} \left(\frac{\hat{\mu}_{yL}}{\bar{y}} \right) = \frac{1}{1 - \hat{\rho}^2} > 1 \text{ unless } \hat{\rho} = 0.$$

$$\widehat{RE} \left(\frac{\hat{\mu}_{yL}}{\hat{\mu}_y} \right) = \frac{(s_y^2 + r^2 s_x^2 - 2r\hat{\rho}s_x s_y)}{s_y^2(1 - \hat{\rho}^2)} > 1 \text{ unless } b = r.$$

$$\widehat{RE} \left(\frac{\hat{\mu}_{yL}}{\hat{\mu}_{yD}} \right) = \frac{(s_y^2 + s_x^2 - 2\hat{\rho}s_x s_y)}{s_y^2(1 - \hat{\rho}^2)} > 1 \text{ unless } b = 1.$$

Conclusions: