Ratio Estimation

We have seen how to use additional information in designing a survey by using stratification. Ratio estimation is another way to improve estimates by using extra information, by using an additional variable. In using ratio estimation, in addition to having a variable of interest, y, we also have another variable, x, and often we have information on the population of x values, such as μ_x or τ_x. The oranges example in the text gives one example of the use of ratio estimation.

Common reasons for using ratio estimation:

1. to estimate a ratio, $R = \tau_y/\tau_x = \mu_y/\mu_x$,
2. to estimate a total when N is unknown, so $\widehat{\tau}_y = (\frac{\overline{y}}{\overline{x}})\tau_x$ is used instead of $\widehat{\tau} = N\overline{y}$,
3. Even if N is known, often the ratio estimator $\widehat{\tau}_y$ has lower variance than $\widehat{\tau} = N\overline{y}$,
4. ratio estimation is often used in other sampling contexts, such as to adjust for nonresponse.

Estimation of a ratio R **:**

We use the sample ratio $\overline{y}/\overline{x}$ as an estimator of the population ratio $R = \tau_y/\tau_x$, giving $\widehat{R} = r = \overline{y}/\overline{x} = \sum y_i/\sum x_i$, with variance estimator:

$$\widehat{V}(r) = \left(\frac{N-n}{nN}\right)\left(\frac{1}{\mu_x^2}\right)s_r^2, \text{ where } s_r^2 = \frac{\sum_{i=1}^{n}(y_i - rx_i)^2}{n-1}.$$

The expression for s_r^2 should look familiar from lectures you may have had on linear regression, because obtaining a ratio estimate is essentially the same as fitting a special kind of a linear regression model (specifically a regression model without an intercept and weighted by the $1/x_i$ values).

Estimation of τ_y and μ_y:

We can use the same approach to obtain estimators of the total τ_y or the mean μ_y for y:

$$\widehat{\tau}_y = r\tau_x, \text{ with } \widehat{V}(\widehat{\tau}_y) = (\tau_x)^2\widehat{V}(r) = N^2\left(\frac{N-n}{nN}\right)s_r^2$$

and

$$\widehat{\mu}_y = r\mu_x, \text{ with } \widehat{V}(\widehat{\mu}_y) = (\mu_x)^2\widehat{V}(r) = \left(\frac{N-n}{nN}\right)s_r^2 = \left(\frac{N-n}{N}\right)\frac{s_r^2}{n}.$$

Note that the variance formulas for $\widehat{\tau}_y$ and $\widehat{\mu}_y$ have the same form as in earlier chapters. For example, $\widehat{V}(\widehat{\mu}_y)$ is of the form $\left(\frac{N-n}{N}\right)\frac{\text{sample variance}}{n}$.