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 $\mu_x$ or $\tau_x$. 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 $\hat{\tau}_y = (\bar{y}/\bar{x})\tau_x$ is used instead of $\hat{\tau} = N\bar{y}$,

3. Even if $N$ is known, often the ratio estimator $\hat{\tau}_y$ has lower variance than $\hat{\tau} = N\bar{y}$,

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 $\bar{y}/\bar{x}$ as an estimator of the population ratio $R = \tau_y/\tau_x$, giving $\hat{R} = r = \bar{y}/\bar{x} = \sum y_i / \sum x_i$, with variance estimator:

$$\hat{V}(r) = \left( \frac{N-n}{nN} \right) \left( \frac{1}{\mu_x^2} \right) s_r^2,$$

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 $\tau_y$ and $\mu_y$:

We can use the same approach to obtain estimators of the total $\tau_y$ or the mean $\mu_y$ for $y$:

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

and

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

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