Comparison of Estimates

For any two random variables, y_1 and y_2, we have $E(y_1 - y_2) = E(y_1) - E(y_2)$ and $V(y_1 - y_2) = V(y_1) + V(y_2) - 2\text{cov}(y_1, y_2)$. If y_1 and y_2 are independent then the variance simplifies to: $V(y_1 - y_2) = V(y_1) + V(y_2)$.

For comparing sample means, we will only consider the simplified case where the estimates are independent. In that case we have:

$$\widehat{\mu_1 - \mu_2} = \overline{y}_1 - \overline{y}_2 \text{ and } \widehat{V}(\overline{y}_1 - \overline{y}_2) = \widehat{V}(\overline{y}_1) + \widehat{V}(\overline{y}_2).$$

If the population sizes are large we often disregard the finite population correction terms.

For comparing sample proportions, we will consider both the case where the samples are independent and also the case where the two proportion estimates are dependent because of multinomial sampling. When the estimates are independent, we have:

$$\widehat{p_1 - p_2} = \widehat{p}_1 - \widehat{p}_2 \text{ and}$$

$$\widehat{V}(\widehat{p}_1 - \widehat{p}_2) = \widehat{V}(\widehat{p}_1) + \widehat{V}(\widehat{p}_2) = \frac{\widehat{p}_1\widehat{q}_1}{n_1 - 1}\left(\frac{N_1 - n_1}{N_1}\right) + \frac{\widehat{p}_2\widehat{q}_2}{n_2 - 1}\left(\frac{N_2 - n_2}{N_2}\right),$$

but notice that in the text they use n in the denominator of the variance instead of $n - 1$, and they ignore the fpc terms. When the estimates are from a multinomial sample (like 'yes', 'no', or 'maybe') then the proportion estimates are dependent and the appropriate estimators are:

$$\widehat{p_1 - p_2} = \widehat{p}_1 - \widehat{p}_2 \text{ and}$$

$$\widehat{V}(\widehat{p}_1 - \widehat{p}_2) = \widehat{V}(\widehat{p}_1) + \widehat{V}(\widehat{p}_2) - 2\widehat{\text{cov}}(\widehat{p}_1, \widehat{p}_2) = \left[\frac{\widehat{p}_1\widehat{q}_1}{n - 1} + \frac{\widehat{p}_2\widehat{q}_2}{n - 1} + \frac{2\widehat{p}_1\widehat{p}_2}{n - 1}\right]\left(\frac{N - n}{N}\right).$$

Notice that the third term in the variance expression has a '+' sign instead of a '-' sign. This is because estimates of proportions from multinomial sampling are negatively correlated. In other words, if you know that more people said 'yes', then less of them would have said 'no', for example.

Advantages and Disadvantages of SRS designs