Simple Random Sampling for Proportions

Often we are interested in estimating a population proportion p for some characteristic, such as the proportion of voters favoring some proposal, or the proportion of an animal species having a particular genetic condition. To estimate a proportion for a particular characteristic, we define the variable y_i for each sampled element to be equal to 1 if the element has the characteristic, and 0 otherwise. Then our estimator $\widehat{p}$ for the proportion p is just the sample mean of y:

$$\widehat{p} = \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n},$$

and since $\widehat{p} = \overline{y}$, the variance estimator can be obtained by using our expression for $\widehat{V}(\overline{y})$ and expressing it in terms of $\widehat{p}$ and $\widehat{q} = 1 - \widehat{p}$:

$$\widehat{V}(\widehat{p}) = \frac{\widehat{p}\widehat{q}}{n-1}\left(\frac{N-n}{N}\right).$$

Sample Size selection for Proportions

Again, we can use the same approach that we used earlier to obtain a sample size expression for estimating proportions, by using the expression for the bound B and solving for the sample size n :

$$n = \frac{Npq}{(N-1)(B^2/4) + pq}.$$

The question also arises as to what value to use for p, since we are trying to estimate it. Here, however, it is easier, because if we do not have information from previous studies, we can set $p = .5$ as a conservative value.

Examples