Cluster estimation for proportions, sample size estimation

Cluster estimation for proportions: As with ratio estimation in the earlier chapter, this is the same as estimating a mean except that the data are just 0's and 1's. Let a_i be the total number of elements in cluster i that possess the desired characteristic ('yes', etc.). Then the estimator of p is:

$$\widehat{p} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i}, \text{ and } \widehat{V}(\widehat{p}) = \left(\frac{N-n}{N}\right)\left(\frac{1}{\overline{M}^2}\right)\frac{s_p^2}{n}, \text{ where}$$

$$s_p^2 = \frac{\sum_{i=1}^{n}(a_i - \widehat{p}m_i)^2}{n-1}.$$

See the ethiopian food example from class and the SAS program on the web for an example.

Sample size estimation: For single-stage cluster estimation, the sample size formulas are very much like those in the earlier chapter on ratio estimation. The sample size n is:

$$n = \frac{N\sigma_*^2}{ND + \sigma_*^2}, \text{ where } D = \begin{cases} B^2\overline{M}^2/4 & \text{for } \mu, \\ B^2/4N^2 & \text{for } \tau, \\ B^2\overline{M}^2/4 & \text{for } p, \end{cases}$$

where σ_*^2 equals $\sigma_r^2, \sigma_r^2, \sigma_t^2$, or σ_p^2, for μ, τ (using $M\overline{y}$), τ (using $N\overline{y}_t$), or p, respectively. Some examples are given in the text and in lecture.