Stat 257 Fall 2005
Cluster Sampling

Example: The City of Regina chief engineer is considering a zoning change for the Wascana Park subdivision to allow a new shopping complex to be built. In order to assess the opinion of residents, a cluster sample is used. The subdivision map is marked into $M = 170$ blocks, and a random sample of $m = 15$ blocks is surveyed. For each of these blocks (labelled i, for $i = 1, \ldots 15$), the number of adult residents (n_i) and the number of adult residents opposing the zoning change (y_i) are recorded. The data are summarized below:

$$\sum_{i=1}^{15} n_i = 546, \quad \sum_{i=1}^{15} n_i^2 = 9981, \quad \sum_{i=1}^{15} n_i y_{iT} = 4035, \quad \sum_{i=1}^{15} y_{iT} = 182, \quad \sum_{i=1}^{15} y_{iT}^2 = 1819,$$

$$\sum_{i=1}^{15} \overline{y}_i^2 = 1.54, \quad \sum_{i=1}^{15} n_i^2 \overline{y}_i^2 = 2103, \quad \sum_{i=1}^{15} n_i^2 \overline{y}_i = 4571.$$

Use an appropriate estimator to construct an approximate 95% confidence interval for $\overline{Y}$, the proportion of Wascana Park adult residents opposed to the zoning change.

Solution: Since we do not know the population size N, and we do not suspect that the cluster sizes N_i are the same, we use the estimator $\overline{y}_{c(a)}$. Thus,

$$\overline{y}_{c(a)} = \frac{\sum_{i=1}^{15} y_{iT}}{\sum_{i=1}^{15} n_i} = \frac{182}{546} = \frac{1}{3}$$

and

$$\begin{aligned} s^2(\overline{y}_{c(a)}) = \frac{(M-m)m}{M(m-1)} \sum_{i=1}^{15} \left(\frac{n_i}{n}\right)^2 \left(\overline{y}_i - \overline{y}_{c(a)}\right)^2 &= \frac{(M-m)m}{M(m-1)n^2} \sum_{i=1}^{15} n_i^2 \left(\overline{y}_i^2 - 2\overline{y}_i \overline{y}_{c(a)} + \overline{y}_{c(a)}^2\right) \\ &= \frac{(M-m)m}{M(m-1)n^2} \left(\sum_{i=1}^{15} n_i^2 \overline{y}_i^2 - 2\overline{y}_{c(a)} \sum_{i=1}^{15} n_i^2 \overline{y}_i + \overline{y}_{c(a)}^2 \sum_{i=1}^{15} n_i^2\right) \\ &= \frac{(170-15)\cdot 15}{170 \cdot (15-1) \cdot 546^2} \left(2103 - 2 \cdot \frac{1}{3} \cdot 4571 + \left(\frac{1}{3}\right)^2 \cdot 9981\right) \\ &= 0.000540. \end{aligned}$$

Hence, an approximate 95% confidence interval for

$$0.333 \pm 1.96 \cdot 0.023 \quad \text{or} \quad (0.288, 0.379).$$