Elements of Statistics
Lecture Notes # 10

1 Systematic Sampling

Things to know

- The convenience of systematic sample

- Its disadvantage, when it occurs, and variation(Repeated systematic sampling).

- The probability of every unit in the population to be selected is equal.

Definition 1.1 *For a **1-in-k systematic sample**, you order the units of the population in some way and randomly select one of the first k units in the ordered list. This selected unit is the first unit to be included in the sample. You continue through the list selecting every k^{th} unit from then on.*

The advantage of systematic sample is that it really fast and easy to convenient when you already have a list of the units in your population. But the disadvantage is that systematic sample might lead to bias. Suppose the population is arranged in a pattern and if the periodicity of sampling might exactly match the periodicity of that pattern. However, if we can assume that the population list is randomly shuffled, then systematic sampling is equivalent to simple random sample, where there is no bias.

Repeated systematic sampling is a variation of the systematic sampling that seeks to avoid the systematic bias due to periodicity. It is done by taking several smaller systematic samples, each with a different random starting point, rather than one pass through the whole population.

Example Say we have a roster below and we want to take a 1-in-4 systematic sample.
   A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T.
   And we do perform a systematic sampling. What is the probability that A is selected? What is the probability that F is selected? In a 1-to-k systematic sample, every unit has a probability of $1/k$ to be selected.

Exercise Page 120 Let's Do It! 2.10
   (a) No. Some presentations had less than 50 registered. The smallest number of programmers registered is 9 for presentation 13, so the maximum value of k is 9.
   (b) Using a $k = 9$ and the Table, we have $6, 15, 24, 33, 42, 51, 60, 69, 78, 87$.
   (c) The sample size is
$$\left\lfloor \frac{1268 - 6}{9} \right\rfloor + 1 = 141.$$
   (d) Do a stratified random sample.

2 Cluster Sampling

Things to know

- Definition & Examples

- Difference between cluster and stratum

- Advantages and disadvantages.

This is a sampling method in some sense similar to stratified random sampling. The motivation would be suppose you want to investigate the television-viewing habits of households by face-to-face interview, in particular for children, adult females, and adult males. We can do a stratified sampling on each of the three groups. But think about if one child selected live in Alaska, an adult male lives in New York, and an adult female lives in Alabama. Traveling to each of the places would be incredibly expensive. And in this case, a cluster sampling does help. We can combine every family as a new unit, called **cluster**, and do a random sampling on the clusters.

Definition 2.1 *In **cluster sampling**, the units of the population are grouped into clusters. One or more clusters are selected at random. If a cluster is selected, all of the units form that cluster are included in the sample.*

Don't be confused with cluster and strata.

Q: What is the difference between cluster and strata?

(1) In strata, everybody is homogeneous. But the composition of cluster is more involved in that it looks like a "minipopulation".
(2) When sampling from strata, only a portion is selected from each sample. But when sampling from cluster, complete clusters will be selected.

The advantage of cluster sampling is saving in time and money. Less cost is incurred if the interviewer stays within a specific location rather than traveling around.

However, it might be the case that the naturally clusters, like families, hospitals, do not reflect the variation that is present in the population. They may not look like a minipopulation and the sample might be biased.