

Friday, Oct 18

Inclusion Probabilities — Revisited

Recall that an *inclusion probability* (π_i) of the i -th element is the probability that the element will be in the selected sample.

For *simple random sampling*, all elements have an inclusion probability of $\pi_i = n/N$.

For *stratified random sampling*, the inclusion probability of the i -th element is $\pi_i = n_j/N_j$ if the element is in stratum j .

What are the inclusion probabilities for a *cluster sampling design*? It depends on how the *clusters* are sampled. The inclusion probability of an element is the same as that of its cluster. Consider three ways to sample clusters.

1. Simple random sampling.
2. Stratified random sampling.
3. (Selection) probability proportional to (cluster) size (with replacement).

Question: Some people [define](#) simple random sampling as one where every element has the same inclusion probability. Is this correct? Why or why not?

Some Themes in Survey Sampling

Auxiliary Variables

An *auxiliary variable* is a variable that can be observed for every element (or sometimes the sampling unit) in the population.

Uses of auxiliary variables:

1. Improving a *sampling design*. What examples of this have we seen so far?
2. Improving an *estimator*. What examples of this have we seen so far?

Double Sampling

A *double sampling* can be used to facilitate the use of an auxiliary variable in compute an estimator.

1. How might we use double sampling to facilitate the use of estimators used in *stratified random sampling*?
2. How might we use double sampling to facilitate the use of a *ratio or regression estimator*?

Simple Random Sampling in Complex Sampling Designs

Some complex sampling designs using simple random sampling in some way.

1. How might *stratified random sampling* use simple random sampling?
2. How might *cluster sampling* use simple random sampling?

Sampling Design Decisions

Fundamentally, design decisions involve specifying the sample space of possible samples, and sometimes also changing their probabilities.

1. What design decisions are made in *simple random sampling*?
2. What design decisions are made in *stratified random sampling*?
3. What design decisions are made in *cluster sampling*?

Relative Precision of the Designs

1. How does *stratified random sampling* (usually) compare to simple random sampling?
2. How does *cluster sampling* compare (usually) to simple random sampling?