ST 432 Unequal Probability Sampling (Chapter 6)

This is an important topic for later in the semester. Unequal probability sampling is very important to many topics including cluster sampling and two-stage sampling, network sampling, and sampling of animals where detection is an issue. I emphasize that these notes do not replace careful reading of text as they are too brief for that.

Why do we use Unequal Probability Sampling?

1. **Unequal Importance of Sampling Units .**

I am reminded of George Orwell's famous quote from "Animal Farm".

" All animals are created equal but some animals are more equal than others" (☺).

Sampling units are often not of equal importance.

Example-If we were surveying the state then regions may not be equally important. In a survey on coastal development the surveyor may want to give more weight to the Coast vs the Piedmont and the Mountains.

2. **Unequal Sizes of Sampling Units**

Some units may be bigger and therefore should be sampled at a higher rate.

Example 1- If the sampling unit is a family then you might want sample proportional to the size of the family.

Example 2- Hospital survey where individual hospital is the unit. Here it might make sense to sample proportional to the number of beds in the hospital. The No. of beds is a measure of size of the hospital.

Example 3- Business survey where an individual business is the sampling unit. Sample proportional to some know measure of size (no of employees, gross sales etc).

Sampling With or Without Replacement.

With replacement- ***Hansen Hurwitz Theorem***

In this case we assume that we are sampling with replacement from a population $\{y_1, y_2, \ldots, y_N\}$and we define p_i as the probability of drawing unit i on each draw. We then have a sample $\{y_1, y_2, \ldots, y_n\}$. It is reasonably simple to show that

$$\hat{\tau}_p = \frac{1}{n}\sum_{1}^{n}(\frac{y_i}{p_i}) ,$$

is an unbiased estimate of the population total and I will go through this proof in class. The detailed equations are given in the text.

Simple Example_ Suppose we had a srs with replacement sampling design then all the pi would be 1/N and hence

$$\hat{\tau}_p = \frac{1}{n}\sum_{1}^{n}\left(\frac{y_i}{1/N}\right) = N\bar{y},$$

which is the standard result for srs.

The selections are independent on each draw because the unit is replaced and could be drawn again. This greatly simplifies derivations and greatly simplifies drawing the sample but some elements may be sampled more than once. When sampling without replacement we don't have this issue but there are a lot of other issues to consider which become quite complex.

With or Without Replacement-*Horvitz Thompson Theorem*

This is one of the foundation equations of sampling. We may sometimes use sampling with replacement because sampling without replacement has some practical difficulties when we are using general unequal probability sampling methods. In this case we assume that we are sampling *with or without* replacement from a population $\{y_1, y_2, \ldots, y_N\}$and we define π_i as the probability of including unit i in the sample. We then have a sample $\{y_1, y_2, \ldots, y_n\}$. It is reasonably simple to show that

$$\hat{\tau}_\pi = \sum_{1}^{v}\left(\frac{y_i}{\pi_i}\right),$$

where v is the effective sample size, that is the number of distinct units in the sample.

Example 1_ Suppose we had a srs without replacement sampling design then all the π_i would be n/N and hence

$$\hat{\tau}_\pi = \sum_{1}^{n}\left(\frac{y_i}{n/N}\right) = N\bar{y},$$

which is the standard result for srs without replacement.

Example 2 –Unequal probability sampling with replacement. P55 in Text. I will show you how we can derive the inclusion probabilities in this more complex example.