LECTURE 4: SIMPLE RANDOM SAMPLING 3.1 – 3.2

3.1 What is a Simple Random Sample?

Suppose we have a population of size N and we would like to take a sample of size n. The number T of possible samples of n elements from a population of N elements is

$$T = \binom{N}{n} = \frac{N!}{n!(N-n)!} .$$

Give examples.

Definition. A simple random sample (SRS) of n elements from a population of size N is one in which each of the $\binom{N}{n}$ possible samples has the same probability $1/\binom{N}{n}$ of being chosen.

Note that we implicitly assume that we draw a sample without replacement. A sample is said to be drawn without replacement if each element in a population can be chosen at most one time. A sample is said to be drawn with replacement if each element in a population can be chosen more than once.

Note that a there are a total of N^n simple random samples with replacement, each having probability N^{-n} of being chosen.

We will study sampling without replacement.

3.1.2 Probability of an Element Being Selected

The probability that a particular element of a population is chosen for a simple random sample is

$$\frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \ldots = \frac{n}{N} .$$

Next time we will use Excel and SAS to see how a simple random sample may be drawn.

3.2 Estimation of Population Parameters Under Simple Random Sampling

Lemma Suppose we draw a simple random sample of size n from a population of size N, and obtain observations $x_1, \ldots, x_n$. The sample mean

$\bar{x} = \sum_{i=1}^{n} x_i$ is (i) an unbiased estimator of the population mean $\bar{X}$, and (ii) has the variance equal to

$$\mathbb{V}ar(\bar{x}) = \frac{1}{n}\Big(1 - \frac{n}{N}\Big)\Big(\frac{N\sigma_X^2}{N-1}\Big)\,.$$

Proof: We write

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{N} X_i\, Z_i\,,$$

where $Z_i = 1$ if element i is in the sample, and 0, otherwise. Z_i's are identically distributed Bernoulli random variables with the probability of success $\mathbb{P}(Z_i = 1) = \frac{n}{N}$. The random variables Z_i's are correlated.

(i) We compute

$$\mathbb{E}(\bar{x}) = \frac{1}{n}\sum_{i=1}^{N} X_i \mathbb{E}(Z_i) = \frac{1}{n}\sum_{i=1}^{N} X_i\,\frac{n}{N} = \bar{X}\,.$$

(ii) Note that

$$\mathbb{V}ar(Z_i) = \frac{n}{N}\Big(1 - \frac{n}{N}\Big),$$

and

$$\mathbb{E}(Z_i Z_j) = \mathbb{P}(Z_i = Z_j = 1)$$
$$= \mathbb{P}(Z_i = 1|Z_j = 1)\mathbb{P}(Z_j = 1) = \Big(\frac{n-1}{N-1}\Big)\Big(\frac{n}{N}\Big)\,,\quad i \neq j\,.$$

Thus, for $i \neq j$,

$$\mathbb{C}ov(Z_i, Z_j) = \mathbb{E}(Z_i Z_j) - \mathbb{E}(Z_i)\mathbb{E}(Z_j)$$
$$= \Big(\frac{n-1}{N-1}\Big)\Big(\frac{n}{N}\Big) - \Big(\frac{n}{N}\Big)^2 = -\frac{1}{N-1}\Big(1 - \frac{n}{N}\Big)\Big(\frac{n}{N}\Big)\,.$$

Finally,

$$\mathbb{V}ar(\bar{x}) = \frac{1}{n^2}\,\mathbb{V}ar\Big(\sum_{i=1}^{N} Z_i\, X_i\Big)$$
$$= \frac{1}{n^2}\Big[\sum_{i=1}^{N} X_i^2 \mathbb{V}ar(Z_i) + \sum_{i=1}^{N}\sum_{j\neq i}^{N} X_i\, X_j \mathbb{C}ov(Z_i, Z_j)\Big]$$
$$= \frac{1}{n^2}\Big[\frac{n}{N}\Big(1 - \frac{n}{N}\Big)\sum_{i=1}^{N} X_i^2 - \sum_{i=1}^{N}\sum_{j\neq i}^{N} X_i\, X_j \frac{1}{N-1}\Big(1 - \frac{n}{N}\Big)\Big(\frac{n}{N}\Big)\Big]$$
$$= \frac{1}{n^2}\frac{n}{N}\Big(1 - \frac{n}{N}\Big)\Big[\sum_{i=1}^{N} X_i^2 - \frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N} X_i\, X_j\Big]$$
$$= \frac{1}{n}\Big(1 - \frac{n}{N}\Big)\frac{1}{N(N-1)}\Big[(N-1)\sum_{i=1}^{N} X_i^2 - \Big(\big(\sum_{i=1}^{N} X_i\big)^2 - \sum_{i=1}^{N} X_i^2\Big)\Big]$$
$$= \frac{1}{n}\Big(1 - \frac{n}{N}\Big)\frac{1}{N(N-1)}\Big[N\sum_{i=1}^{N} X_i^2 - \big(\sum_{i=1}^{N} X_i\big)^2\Big] = \frac{1}{n}\Big(1 - \frac{n}{N}\Big)\Big(\frac{N\sigma_X^2}{N-1}\Big)\,. \quad \square$$