Sullivan Section 1.3

Gene Quinn

Other Types of Sampling

In addition to simple random sampling, there are other types of sampling that may be used:

- stratified sampling
- systematic sampling
- cluster sampling

Stratified Sampling

In many cases, individuals in a population may be classified into subpopulations that are more homogenous than the population as a whole.

A technique called **stratified sampling** can be used to exploit this situation.

Stratified Sampling

In many cases, individuals in a population may be classified into subpopulations that are more homogenous than the population as a whole.

A technique called **stratified sampling** can be used to exploit this situation.

A **stratified sample** is obtained in two steps:

- divide the population into **strata** (nonoverlapping subgroups)
- obtain a simple random sample from each of the strata

Stratified Sampling

There are two decisions to be made with stratified sampling.

The first task is to decide how the population is to be divided into strata.

Stratified Sampling

There are two decisions to be made with stratified sampling.

The first task is to decide how the population is to be divided into strata.

Usually the strata are chosen so that each of them will be more homogenous than the population as a whole.

Stratified Sampling

Once the strata are defined, the second task is to decide the sample sizes for each of the strata.

Stratified Sampling

Once the strata are defined, the second task is to decide the sample sizes for each of the strata.

If no additional information is available, the sample sizes may be allocated in proportion to the size of the strata.

Stratified Sampling

Once the strata are defined, the second task is to decide the sample sizes for each of the strata.

If no additional information is available, the sample sizes may be allocated in proportion to the size of the strata.

Sometimes additional information about the homogeneity of individual strata may be used to determine a more efficient way of dividing the total sample among the strata.

Systematic Sampling

A **systematic sample** is obtained by choosing some whole number k, and then sampling every k^{th} individual in the population.

Systematic Sampling

A **systematic sample** is obtained by choosing some whole number k, and then sampling every k^{th} individual in the population.

The sampling process is started by choosing a whole number i at random between 1 and k, and sampling the i^{th} individual first.

After that, every k^{th} individual is sampled.

Systematic Sampling

A **systematic sample** is obtained by choosing some whole number k, and then sampling every k^{th} individual in the population.

The sampling process is started by choosing a whole number i at random between 1 and k, and sampling the i^{th} individual first.

After that, every k^{th} individual is sampled.

An advantage of this technique is that we do not have to have a frame, or list of the entire population, to use it.

Systematic Sampling

The main decision to be made with systematic sampling is the value of k.

How this is done depends on whether or not we have a frame.

We will assume that we have decided how large a sample is required.

Systematic Sampling

The main decision to be made with systematic sampling is the value of k.

How this is done depends on whether or not we have a frame.

We will assume that we have decided how large a sample is required.

If we have a frame, we know the size N of the population, and if the sample size we want is n, we choose:

$$k = \frac{N}{n}$$

Systematic Sampling

Example:

In a town with an adult population of 13,000, the town council wishes to evaluate the need for a library addition by a questionaire.

A consulting statistician has determined that a sample of size 300 will be adequate.

If systematic sampling is used, what is an appropriate value for k?

Systematic Sampling

Example:

In a town with an adult population of 13,000, the town council wishes to evaluate the need for a library addition by a questionaire.

A consulting statistician has determined that a sample of size 300 will be adequate.

If systematic sampling is used, what is an appropriate value for k?

Since we have a frame, we know that the population size N is $13,000$ and the desired sample size is 300. In this case,

$$\frac{N}{n} = \frac{13000}{300} = 43.33$$

The largest value of k that produces a sample of at least 300 is 43, so they should sample every 43^{rd} adult citizen.

Systematic Sampling

If no frame is available, some other information must be used to determine k.

This will be highly dependent on the nature of the individual situation.

Systematic Sampling

If no frame is available, some other information must be used to determine k.

This will be highly dependent on the nature of the individual situation.

If the sample is conducted on a particular day, for example, a store owner might have a good idea how many customers to expect that day.

In this case, N would be taken to be the expected number of customers for the entire day, and k would be computed as N/n.

Systematic Sampling

Example:

The turnpike authority wishes to measure the occupancy of vehicles passing through a certain toll plaza during the morning rush hour.

It is estimated that sample of size 180 will be adequate.

Based on cash receipts, it is expected that $9,200$ cars will pass through the toll plaza during a typical morning rush hour.

If systematic sampling is used, what is an appropriate value for k?

Systematic Sampling

Example:

The turnpike authority wishes to measure the occupancy of vehicles passing through a certain toll plaza during the morning rush hour.

It is estimated that sample of size 180 will be adequate.

Based on cash receipts, it is expected that $9,200$ cars will pass through the toll plaza during a typical morning rush hour.

If systematic sampling is used, what is an appropriate value for k?

In this case, we have no frame, but we expect $9,200$ cars, so

$$\frac{N}{n} = \frac{9200}{180} = 51.11$$

The largest value of k that produces a sample of at least 180 is 51, so they should sample every 51^{st} car.

Cluster Sampling

A **cluster sample** is obtained by selecting all individuals in some randomly chosen collection or group.

Cluster Sampling

A **cluster sample** is obtained by selecting all individuals in some randomly chosen collection or group.

The sampling process is started by first choosing a group.

Once a group is chosen, a sample is taken from the individuals in that group.

Cluster Sampling

A **cluster sample** is obtained by selecting all individuals in some randomly chosen collection or group.

The sampling process is started by first choosing a group.

Once a group is chosen, a sample is taken from the individuals in that group.

This technique is useful when it is easier to sample all members of the chosen group than it is to select a small number of individuals from many groups.

Cluster Sampling

Example:

A manufacturer of popcorn is interested in determining the relative frequency of blackened kernels in bags of its product.

Cluster Sampling

Example:

A manufacturer of popcorn is interested in determining the relative frequency of blackened kernels in bags of its product.

Since it is not practical to open every bag as it comes off the production line, it would make more sense to select bags at random, then open them and sample the kernels in the chosen bags.

Cluster Sampling

Example:

A manufacturer of popcorn is interested in determining the relative frequency of blackened kernels in bags of its product.

Since it is not practical to open every bag as it comes off the production line, it would make more sense to select bags at random, then open them and sample the kernels in the chosen bags.

Cluster Sampling

Three decisions need to be made when setting up a cluster sample:

- how is the population divided into clusters?
- how many clusters should be sampled?
- how many individuals in each cluster should be sampled?

Cluster Sampling

In general, if the individuals in each cluster tend to be very similar, it is more efficient to sample many clusters and only a few individuals from each cluster.

Cluster Sampling

In general, if the individuals in each cluster tend to be very similar, it is more efficient to sample many clusters and only a few individuals from each cluster.

If the clusters are not very homogeneous, it is more efficient to take fewer clusters and more individuals from each cluster.

Convenience Sampling

The term **convenience sampling** refers to a number of sampling techniques in which the individuals are easily obtained, often by non-random selection.

Convenience Sampling

The term **convenience sampling** refers to a number of sampling techniques in which the individuals are easily obtained, often by non-random selection.

An example of a convenience sample would be respondents to an opinion question presented on a news show. This is called a **self-selected** sample because each viewer determines whether or not they will be in the sample.

For this reason, results of this type of sample may be highly biased.