INTRODUCTION TO SAMPLING

This handout provides a brief overview of some of the main ideas in applied sampling.

What is sampling?

• Sampling is typically done to estimate some unknown population parameter.

• From a population of sampling units, we select a sample and record information concerning the parameter of interest on each sampled unit. How the units are selected is the primary subject of *sampling*.

Parameters

• *A* parameter is a numerical attribute of the population.

• E.g., if we want to estimate the number of grizzly bears in a region, then the population is the collection of all grizzly bears in the region and the parameter is the number of bears in region. Usually, we assume that the number of sampling units, N, or population size, is finite.

• If your objective is to estimate the amount of rangeland within some area of interest that is infested with leafy spurge, how should you define the population and parameter of interest? What are the sampling units? How can you draw a representative sample? Is access important - are easy-to-access locations different from locations that are distant from travel routes?

• Suppose that your objective is to monitor the amount of rangeland within some area of interest that is infested with leafy spurge. How often should you sample; should you return to the same locations every year, or sample new ones?

Basic Sampling Designs

1. Census - sample the entire population. A census allows you to compute the parameter without error, so most statistical methods (confidence intervals for example) are not necessary.

2. Simple Random Sampling - random selection of n sampling units without replacement (n is the sample size). Every set of n sampling units has the same probability of being the sample, and every observation has the same probability of being sampled.

3. Systematic Random Sampling - select every mth member of the population. Can we introduce randomness into this design?

4. Unequal Probability Sampling - any sampling plan where sampling units have different probabilities of being selected. Does this introduce bias into the estimate of the parameter?

    • One example is sampling proportional to size (PPS sampling). For example, if sampling units are lakes, PPS sampling can be accomplished by randomly locating points on a map. If a point falls on a lake, then the lake is included in the sample.

5. Stratified Random Sampling - sampling where the population is divided into different strata, and a SRS is drawn from each stratum.

    • This is a two-stage sampling plan, where the stratification is one stage, and the SRS is the second stage. Any of the designs 1-4 could be used at the second stage.

- Stratified random sampling is appropriate when the strata are different, and there is homogeneity among strata.

6. Cluster Sampling - sampling where the population is first divided into clusters, and then a SRS of *clusters* is selected. Typically, every sampling unit within a selected cluster is then sampled.

- Cluster sampling is appropriate when clusters are similar, and there is heterogeneity among clusters.

7. Convenience or haphazard sampling - observations are selected as they become available. A major problem is that we do not know how well the sample represents the population. However, the sample might be a useful for exploratory analyses aimed at developing hypotheses and research questions.

Examples of Sampling Problems

For the following situations, identify the population, the sampling unit (or population unit), and the sampling plan. If it is a multistage sampling plan, identify the population, sampling unit and sampling plan at each stage.

1. A researcher is to estimate the proportion of bare ground on a 40-acre parcel of land.

(a) She stands at a central location and randomly chooses 5 integers from 1 to 360. These represent the directions, in degrees from north, of five transects through a center point which extend to the edges of the parcel. On each transect, she uses a measuring wheel to determine how much of the transect lies in bare ground.

(b) As in (a) except that she chooses ten points at random along each transect and centers a 1 m radius circle at each of the points. She then determines the proportion of bare ground in each of the circles.

(c) As in (a) except that for each transect she chooses a random distance from 0 to 5 meters from the center. At 5 meter intervals along the transect starting at the random point, she centers a circle with a radius of 1 m and determines the proportion of bare ground in the circle.

2 A geologist is interested in the surface geology of an area. He partitions the area as a lattice consisting of 20 equal-area parcels. Within each parcel, he randomly selects 4 points and obtains measurements at each of these points.

3. A fire researcher is interested in estimating the average fuel moisture of the leaves on bushes in a small area. She randomly selects 10 bushes from the area and then randomly removes 2 branches from each bush. She removes all leaves from the branches and measures the moisture content of the leaves.

4. A sociologist is interested in the demographics of Missoula bar patrons. He randomly selects 5 bars within the city limits and visits them in random order, on each of 5 consecutive nights. He observes all people entering the bar from 8 to 12 pm, recording the sex and estimated age of each.

5. A researcher is interested in the size and habitat preferences of black bears in a certain geographic region.

(a) He sets up traps at 5 accessible locations scattered throughout the region. He traps until 10 bears have been caught. He measures various physical characteristics of the trapped bears.

(b) As in (a), but he radio-collars the bears. Each bear is located once each week during the summer at the same time of day and day of the week. These observations are used to estimate the proportion of time spent in each of the available habitats in the region during the summer

(c) As in (b), but each bear is located at a random time and random day of the week.