Statistics 104 - Laboratory 1

1. Data collection is usually the first activity in a statistical study. For each of the following situations identify: the population, the variable of interest, type of variable categorical (qualitative) or numerical (quantitative), and the population parameter of interest.

   a. The proportion of registered voters in the U.S. who favor withdrawal of U.S. troops from Iraq.
   b. The median assessed value of residential homes in the state of Iowa.
   c. The average size of the high school graduating class for students attending Iowa State University.
   d. The proportion of students in the College of Agriculture who participated in Future Farmers of America (FFA).
   e. The proportion of Stat 104 students (all sections) who live on campus.
   f. The average number of college credits earned by students in this section of Stat 104.

2. When collecting sample data we want to avoid bias (under or over representing a portion of the population in the sample). Suppose in problem 1e we decide to use Section A of Stat 104 (which meets TTh 8:00 to 8:50 am and F 8:00 to 9:50 am) as our sample. What is a possible bias in this sample?

3. Calculate the average number of college credits earned by the 4 students in your group. Is this a population parameter or a sample statistic for 1f? Explain briefly.

4. Drawing samples from populations requires some method for random selection. In this activity you will use the table of random numbers (Table 1, Appendix B of your text) to select a simple random sample and also to look at another sampling method called transect sampling.

   Pine forests are periodically infested with beetles and the infestation often seems to jump from tree to tree in a random manner. How can a forester estimate the total number of trees infested with beetles in a large forest? By taking a sample of plots in the forest and counting the number of infested trees in the sample, a forester will be able to use the sample data to estimate the total. To simulate this activity, imagine that Figure 1 represents a forest and the black dots represent infested trees in that forest. The forest has been divided into 100 plots of equal area.

   a. You have a table of random numbers available for you to use. Devise a plan to produce a random sample from the forest using the random numbers in the table. Be sure to identify the population, variable of interest, and sample.
   b. Use your sampling plan to randomly select 10 plots from the forest. For each plot count the number of infested trees (dots) that are entirely within the plot. Do not count trees (dots) that overlap the edges of a plot. For example, the (Row 0, Column 0) plot has 2 infested trees.
c. Calculate the mean of the 10 counts. Using this mean, estimate the total number of infested trees in the forest.

d. Again use the random number table to select 10 plots. This time for each plot count the number of infested trees (dots) that have any part within the plot. For example, the (Row 0, Column 0) plot has 3 infested trees.

e. Calculate the mean of the 10 counts. Using this mean, estimate the total number of infested trees in the forest.

f. What would be the best estimate of the total number of infested trees in the forest? Briefly explain your choice.

g. An alternative to sampling the plots is to randomly pick a row (or column) of plots in the forest. This row (or column) represents a transect. It would be much easier for the forester to walk through the forest along a randomly selected transect than it would be to find 10 randomly selected plots in a real forest.

i. Use your random number table to select a row at random and count the number of infested trees entirely within the row (transect).

ii. Use your random number table to select a column at random and count the number of infested trees with some part within the column (transect).

iii. Come up with an estimate of the total number of infested trees based on the results from your transects. Show your work.

h. How well do your two estimates of the total number of infested trees (f and g) compare?
Statistics 104 - Laboratory 1
Group Answer Sheet

Names of Group Members: ____________________, ________________
____________________, ________________

1. Data collection
   a. Withdrawal from Iraq
      • Population
      • Variable
      • Type
      • Parameter
   b. Residential homes
      • Population
      • Variable
      • Type
      • Parameter
   c. High school graduating class
      • Population
      • Variable
      • Type
      • Parameter
   d. Future Farmers of America
      • Population
      • Variable
      • Type
      • Parameter
   e. Students living on campus
      • Population
      • Variable
      • Type
      • Parameter
   f. College credits earned
      • Population
      • Variable
      • Type
      • Parameter
2. What is the possible bias in this sample?

3. Credits earned for your group.
   Give individual data values and calculate the mean.
   Is this mean a population parameter or a sample statistic? Explain briefly.

4. Sampling trees in a forest
   a. Sampling Plan (indicate population, variable and sample and how you will use
      the table of random numbers to select your sample.)

   b. Identify plots selects and give counts for 10 selected plots, whole tree in plot.

   c. Mean number infested and estimate of the total number of infested trees in the
      forest.
d. Identify plots selected and give counts for 10 selected plots, any part of tree in plot.

e. Mean number infested and estimate of the total number of infested trees in the forest.

f. Best estimate of the total number of infested trees and explanation.

g. Transect sampling.
   i. Row (transect) selected: _______
      Number of infested trees, whole tree in transect: _______
   ii. Column (transect) selected: _______
      Number of infested trees, any part of tree in transect. _______
   iii. Estimate of total number of infested trees in the forest showing work.

h. How do the two estimates (f and g) compare?