Assignment 1

1. Consider a population of 31 felled black cherry trees (see file trees.csv on Canvas). This data set provides measurements of the girth, height and volume of timber in 31 felled black cherry trees.
    a. The variable height is the height of the cherry tree in inches. Take a SRS of size n = 10 heights from this dataset. Using your sample, estimate the population mean tree height and the standard error of the estimator.
    b. Report a 95% confidence interval for the mean tree height.
    c. Repeat part a. 10,000 times. Construct a histogram of the 10,000 sample means. Comment on the histogram.
    d. Using your samples in part c., compute the mean of your sample means. How does this compare to the population mean?
    e. For each of the samples in part c., compute a 95% confidence interval. What proportion of the intervals contain the true population mean? Is this what you expected (explain)?

2. A researcher from Resource and Environmental Management at SFU wishes to estimate the total number of cedar trees in a study area on Vancouver Island. The study area has been divided into 10,000 non-overlapping units or plots of land. From a pilot survey, the researcher approximates the standard deviation of the number of trees per plot of land to be about 9 trees. If a simple random sample of plots is to be taken, what is the sample size required to estimate the total number of trees in the study area within (i) 200 trees with 95% confidence; (ii) 500 trees with 95% confidence; and (iii) 1,000 trees with 95% confidence?

3. Suppose we wish to estimate the average number of pages in a book in the SFU's W.A.C. Bennett Library. Also, suppose we do not have a list of books from which we can sample.

    Propose a sampling method to achieve this goal (it is early in the course, so nothing too fancy is required). In presenting your solution, consider the following:

    i. What practical problems arise in establishing a frame from which to sample? How will you do this?
    ii. Is everything in the library a book?
    iii. What is the sampling unit?
    iv. What is the observational unit?
    v. How is the sample selection carried out?
    vi. How would you estimate the mean and variance for the number of pages in a book?
    vii. Are there any obvious problems with the proposed method?

4. Suppose the City of Coquitlam has 100,000 homes, of which 35,000 are houses, 35,000 are condominiums and 30,000 are apartments. Researchers wish to investigate energy usage in the city. Suppose that the standard deviations of energy usage for houses, condominiums, and apartments are known from a pilot study to be 500, 300 and 350 kilowatt-hours (kWh), respectively.
    a. The researchers want to take a stratified sample of n = 1,000 dwelling units in order to estimate the mean electricity consumption per unit in the population. How would you allocate a stratified sample of 1,000 observations if we wanted to use proportional allocation? What would the margin of error be in this case?
    b. If proportional allocation is to be used, what sample size would have to be taken for the margin of error to be at most 30 kWh?