STAT 549
Fall 2014 **Homework 1**: Due Monday, September 22.

1. From the "Sampling Problems" on pp. 6-7 of the notes, give the sampling unit, population, and sampling plan for each stage of sampling for problem 1, all parts. Be specific in your answers.

2. Consider a population of size $N = 5$ divided into two strata where the response (y) values for the first stratum are 3, 7, and 8 and for the second stratum are 12 and 15. A stratified random sample consisting of one observation from each stratum will be taken. Let y_1 denote the sample observation from the first stratum and y_2 the sample observation from the second stratum.

   (a) Let $\overline{y} = \frac{1}{2}(y_1 + y_2)$. Derive the sampling distribution of $\overline{y}$ and show that it is a biased estimator of the population mean μ.

   (b) Let $\overline{y}_s = (3/5)y_1 + (2/5)y_2$. Derive the sampling distribution of $\overline{y}_s$ and show that it is an unbiased estimator of μ.

   (c) Compute the inclusion probability π_i for each observation in the population. [Note: The inclusion probability π_i for a unit i is defined in Chapter 6 as the probability that unit i is included in the sample. For an SRS of size n from a population of size N, $\pi_i = n/N$ for each unit i.]

3. In a square .1-acre section of a native hay field, ten 3 ft by 2 ft plots were randomly selected. Each was covered by a deer proof exclosure. At the end of the season, all vegetation in each plot was clipped at ground level and air dried. The air-dry weights in grams of the vegetation in the ten plots were: 68, 52, 87, 54, 39, 47, 37, 36, 42, 24.

   (a) Estimate the total production (air-dry weight in grams) for the entire .1-acre section if deer had been excluded. Obtain the standard error of the estimate and an approximate 90% confidence interval for the total.

   (b) On what assumptions or results is this confidence interval based, and how applicable is the use of the method here?

   (c) If you were charged with selecting a simple random sample of ten 3' x 2' plots from such a section, how would you do it? Be specific.

   (d) Estimate how big a sample of plots would be required to estimate the total biomass on the .1-acre section to within a margin of error of 3 kg. with 95% confidence.

4. Chap. 4, problem 1 (p. 56; 2nd ed: p. 38)

5. Chap. 4, problem 2

6. MATH students only: Do problem 4 of Chap. 2 (p. 36; 2nd ed: pp. 27-28), using <u>both</u> of the suggested methods. That is, prove the identity given, then take its expected value in two ways: 1) over all possible samples and 2) using indicator variables. You may use the expression for $\text{Var}(\overline{y})$ derived in Sec. 2.6 (pp. 21-22). Remember that σ^2 represents the finite-population variance.

7. MATH students only: Suppose you would like to take an SRS of size n from a list of N units, but do not know the population size N in advance. Consider the following procedure:

   (a) Set $S_0 = \{1, 2, \ldots, n\}$, so that the initial sample for consideration consists of the first n units on the list.

(b) For $k = 1, 2, \ldots$, generate a random number u_k between 0 and 1. If $u_k > n/(n+k)$, then set S_k equal to S_{k-1}. If $u_k \leq n/(n+k)$, then select one of the units in S_{k-1} at random, and replace it by unit $(n+k)$ to form S_k.

Show that S_{N-n} from this procedure is an SRS of size n. Hint: Use induction.

8. NON-MATH students only: From the "Sampling Problems" on pp. 6-7 of the notes, give the sampling unit, population, and sampling plan for each sampling stage for problem 7.

9. NON-MATH students only: Consider a population of $N = 5$ units with y-values 5, 0, 9, 6, 9. Consider simple random sampling with $n = 3$. List all the possible samples and show that the sample median is not unbiased for the population median. Calculate the bias, variance, and MSE of the sample median.