STAT 549
Fall 2010 **Homework 1**: Due Friday, September 17.

1. Consider a population of size $N = 5$ divided into two strata where the response (y) values for the first stratum are 2, 5, and 8 and for the second stratum are 10 and 13. 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.]

2. With stratified random sampling, we discovered through an example in class that a stratified random sample was better for estimating the population mean (in the sense of having a smaller variance) than a simple random sample of the same size, when the variability between strata is high compared to the variability within strata.

   (a) What do you think will be the case for cluster sampling in terms of the variability between clusters as compared to the variability within clusters? Why?

   (b) Construct a simple numerical example to illustrate your conclusions in part (a). You will be graded not just on the example you use, but how well you illustrate your findings.

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. From the handout titled "Sampling Problems," give the sampling unit, population, and sampling plan for each stage of sampling for problems 4 & 6. Be specific in your answers.

5. Problem 1, page 38

6. Problem 2, page 38

7. MATH students only: Problem 4, pages 27-28, using the indicator variable method
   NON-MATH students only: From the "Sampling Problems" handout, give the sampling unit, population, and sampling plan for each sampling stage for problems 7 and 8.