Stat 135, Fall 2006
HOMEWORK 1 (due Friday 9/8)

1. Let $x_1, x_2, \ldots, x_n$ be a list of numbers with mean μ and SD σ. Show that

$$\sigma^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i^2 \;-\; \mu^2$$

2. A class has two sections. Students in Section 1 have an average score of 75 with an SD of 10. Students in Section 2 have an average score of 60 with an SD of 10.

a) If possible, say whether the SD of the scores of all the students in the class is
(i) less than 10 (ii) equal to 10 (iii) greater than 10
Explain your choice. If it is not possible to make the choice with the information given, explain why not.

b) Suppose Section 1 has 30 students and Section 2 has 20. Find the SD of the scores of all the students in the class.

3. The mean as a "least squares" estimate. Suppose we play the following guessing game. You pick a number from the list $x_1, x_2, \ldots, x_n$, and ask me to guess what it is. My strategy is to guess that the value is some constant c no matter what you pick. Thus for every i in the range 1 through n, the amount of error that I make with my strategy is $x_i - c$. Define the *mean squared error* of my strategy to be

$$mse_c \;=\; \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2$$

Show that the value of c that minimizes mse_c is $c = \mu$, and that $mse_\mu = \sigma^2$.

4. A coin lands heads with probability p. It is tossed 400 times. Compute the boostrap 99% confidence interval for p and compare with the corresponding interval which uses a conservative estimate for the standard error, when the observed number of heads is

a) 210.

b) 110.

5. The true height of a mountain peak is unknown. Repeated measurements of the height are made using a laser altimeter. The measurements $X_1, X_2, \ldots, X_n$ satisfy the following conditions: for $1 \leq i \leq n$,

$$X_i \;=\; \mu \;+\; \epsilon_i$$

where μ is the unknown constant height of the mountain peak, and ϵ_i is a random error. The errors are i.i.d. with mean 0 and unknown SD σ.

The altimeter is used to make 100 measurements. The measurements have an average of 75,348 inches and an SD of 25 inches.

a) Construct an approximate 95% confidence interval for μ.

b) The interval in part a) is constructed

(i) to estimate the average of the 100 measurements and give ourselves some room for error in the estimate.

(ii) to estimate the height of the mountain and give ourselves some room for error in the estimate.

(iii) to provide a range in which 95 of the 100 measurements are likely to have fallen.

(iv) to provide a range in which 95% of all possible measurements are likely to fall.

Which of (i)-(iv) are false? Explain why they are false.

c) If possible, sketch the histogram of the 100 measurements, justify your choice of shape, and indicate the mean and the SD of the histogram as clearly as you can. If it is not possible to sketch the histogram with the information given, explain why not.

d) A new set of measurements will be made using this altimeter. Approximately how large should the set be to have about a 95% chance that its average is within 1 inch of μ?

6. Researchers studying blood pressure are working with a simple random sample of 625 people. The population is so large compared to the sample that you can assume that the blood pressures of the sampled people are essentially independent of each other. The distribution of blood pressures in the sample is very close to normal. The researchers have used the methods outlined in our class to construct an approximate 99% confidence interval for the mean blood pressure in the population. The interval runs from 126.45 mm to 128.55 mm.

Patients who have blood pressures over 140 mm will be considered at risk for various diseases. If possible, construct an approximate 90% confidence interval for the percent of "at risk" people in the population. If this is not possible with the information given, explain why not.

7. A deck consists of n cards, numbered 1 through n. The deck is going to be shuffled. Assume that all permutations of the n cards are equally likely. And let $1 \leq m_1, m_2, k_1, k_2 \leq n$.

a) How many permutations are possible?

b) In how many permutations does card number m_1 fall in place number k_1?

c) What is the chance that card number m_1 falls in place number k_1?

d) What is the chance that card number m_1 falls in place number k_1 and card number m_2 falls in place number k_2?

e) For $1 \leq i \leq n$, say that a match occurs at place i if card i falls in place i. Let M be the total number of matches. Find $E(M)$ and $Var(M)$.

[Hint: It is a good move to write M as the sum of indicators. If you have forgotten what those are, see Example D on page 129 of your text, or the section entitled *The Method of Indicators* starting on page 168 of Pitman's probability book. And don't forget the indicators, nor any dependence issues, when you calculate the variance.]

f) What happens to the distribution of M as $n \to \infty$? You don't have to give a mathematical proof, but you should provide a convincing heuristic argument. And you should show that your limit distribution is consistent with your answers to part e).

8. Let X have the distribution given below. Here $0 < \theta < 1/3$.

value	1	2	3
probability	θ	2θ	$1 - 3\theta$

Let $X_1, X_2, \ldots, X_n$ be i.i.d., drawn from the distribution above, independent of X.
Let $\bar{X}_{(n)} = (X_1 + X_2 + \ldots + X_n)/n$.

In each of parts (a), (b), and (c), classify the given quantity as a random variable or a real number. If you say "both," explain your answer. If possible, compute the value of the quantity. You can leave your answer in terms of θ if necessary.

a) X **b)** $E(X)$ **c)** $\bar{X}_{(n)}$

9. (Continuing Problem 8.) Say whether each of the following statements is true or false. Justify your answer.

a) $\bar{X}_{(n)} = E(X)$ for all n.

b) $E(\bar{X}_{(n)}) = E(X)$ for all n.

c) If n is large, $\bar{X}_{(n)}$ is likely to be equal to $E(X)$.

d) If n is large, $\bar{X}_{(n)}$ is likely to be close to $E(X)$.

10. (Continuing Problems 8 and 9.) Use the results of Problems 8 and 9 to come up with an estimate of θ based on $\bar{X}_{(n)}$. Is your estimate a random variable or a real number? Show that your estimate has at least one desirable property.