Statistics 104 — Fall, 2004 — Assignment 6
Due Wednesday, November 17th, 2004.

Readings (Moore and McCabe)

- Chapter 6.

Against All Odds videotape

The relevant tapes for this week are numbers 17 (Binomial Distribution), 18 (The sample mean and control charts), 19 (Confidence Intervals) and 20 (Significance Tests). This is completely optional supplementary viewing!

Written Assignment (Moore and McCabe)

- **MM:** 5.2, 5.6, 5.12, 5.32, 5.40, 5.46, 6.6, 6.22, 6.26, 6.34, 6.48, 6.64, 6.66

Bonus Problem

1. Sensitive survey questions and randomized response

   In surveys with sensitive or potentially embarassing questions (such as "Have you ever shoplifted?"), people will often lie. Assuming that the true fraction of the population should answer yes is p, most surveys with questions like this will have sample proportions of people answering yes less than p. One approach around this problem, is a randomized response question. While there are different ways of implementing this procedure, the following is probably the easiest. Have each subject preform these steps

   - Flip a fair coin, but don't tell the interviewer what happened.
   - If the coin comes up heads, answer yes to the question, "Have you ever shoplifted?", regardless to whether they actually did or not.
   - If the coin comes up tails, answer the question, "Have you ever shoplifted?" truthfully.

   Since the interviewer doesn't know the result of the coin flip, the interviewer doesn't know whether a person who answers yes to the question is actually giving a true answer to the question.

   (a) What is the probability that a randomly chosen person will answer yes to the question?

   (b) Suppose n people were asked the question using this procedure. What is the expected number that will answer yes?

   (c) One way of estimating the value of p is to set the expected number of people answering yes to the observed number (call it X) and to solve for p. What is the estimate of p using this procedure (call it $\tilde{p}$).

   (d) Is $\tilde{p}$ an unbiased estimator of p? (Hint: write $\tilde{p}$ as $a + bX$.)

(e) Show that the variance of $\tilde{p}$ is $\frac{1-p^2}{n}$? (Hint: use above hint and note that X is Binomial).

(f) Assume that you could get everybody to answer truthfully so you don't need use the above procedure. Lets suppose you did another survey with m subjects and let Y be the number of who actually shoplifted. Then as we have seen in class $Y \sim Bin(m, p)$. As we will see, the standard estimate of p is this setting is $\hat{p} = \frac{Y}{m}$, the sample proportion. Show that $\hat{p}$ is unbiased and find its variance.

(g) Suppose you want to get the same variance for both estimation procedures. How much bigger than m does n need to be to get the same variance?

2. The above procedure usually isn't the one used. Instead what is usually done is closer to the following.

   - Roll a die, but don't tell the interviewer what happened.
   - If the die comes up 1 or 2 , lie when answering the question, "Have you ever shoplifted?".
   - If the die comes up 3 to 6 , answer the question, "Have you ever shoplifted?" truthfully.

   The advantage to this scheme is that it is impossible to tell with certainty whether any person in particular shoplifted or not. In the scheme used in the problem, if somebody answered no, you knew that they didn't shoplift, which might cause some subjects not to follow the rules. Even though you don't know with certainty who is telling the truth, there is enough of a pattern in this new setup to come up with an unbiased estimate for p, which happens to be

$$\bar{p} = 3\left(\frac{X}{n} - \frac{1}{3}\right).$$

   (a) Verify that $\bar{p}$ above is an unbiased estimate of p when the improved randomized response scheme is used.

   (b) Find the variance of $\bar{p}$.