STATISTICS 7 – SPRING 2008
Homework 6

Handed out: Friday May 9, 2008
Due: Friday May 16, 2008 (by 5pm to Bren Hall 2216)

Reading:	May 9 – May 12	Introduction to inference (Ch 14-16)
	May 12 – May 16	Statistical analysis with one sample (Ch 18, 20)

NOTE: The 2nd midterm exam is scheduled for Wednesday May 21. It covers relevant material from Chapters 1-18 and Chapter 20. It will emphasize material since the last exam: Chapters 3, 10-16, 18, 20. The format will be similar to last exam (formula sheet, calculator, approx 3 problems).

NOTE: No STATA problem this week. (Hooray!) For those that enjoy Stata (and have it on their computer), you can use it to help with some of the probability calculations. The command "`display normal(z)`" will show the probability that a standard normal is less than z (i.e., the same information that is in Table A); "`display invnormal(p)`" will show the z value that goes with a particular probability. For example `display normal(1)` will yield .8413; `display invnormal(.75)` will yield .6745. In the same vein "`display binomial(n,p,k)`" will show the probability that a binomial distribution with n trials and probability of success p yields k or fewer successes.

1. **Binomial distribution and normal approximation.** (*This is Problem 13.32 on page 341 worked out in some detail so that I can provide some additional instruction along the way.*) Leakage from underground gasoline tanks at service stations can damage the environment. It is estimated that 25% of these tanks leak. You examine 15 tanks chosen at random, independently of each other.

   (a) What is the distribution of the number of leaking tanks in a sample of 15 tanks?

   (b) What is the mean number of leaking tanks in such samples of 15? What is the standard deviation of the number of leaking tanks in such samples of 15?

   (c) What is the probability that 10 or more of the 15 tanks leak? (For small n you should calculate binomial probabilities directly as on page 330.)

   (d) Now you do a larger study, examining a random sample of 1000 tanks nationally. What is the probability that at least 275 of these tanks are leaking? (HINT: It is possible to calculate this using the binomial distribution as on page 330 but that is a real pain. Instead we should use the normal distribution to approximate the binomial distribution as on pages 334-336. Just find the mean and standard deviation (as in (b)) and then use the normal distribution with this mean and s.d..)

2. **Sampling distributions ... one more time.** (Based on Problem 11.42 on pages 298-299.) For airline planning purposes the Federal Aviation Agency tells airlines to assume that passengers' average weight is 190 pounds (including clothing and carry-on baggage). The FAA did not specify a standard deviation but a reasonable standard deviation is 35 pounds. (The distribution of weights is not normal but is probably fairly close to normal.)

   (a) A commuter plane carries 19 passengers. What is the approximate probability that the average weight of the 19 passengers exceeds 210 pounds. (You may assume 19 passengers is enough for the central limit theorem to be used.)

   (b) What is the approximate probability that the TOTAL weight of the 19 passengers exceeds 4000 pounds? (Hint: How does this relate to the previous part?)

3. **Confidence intervals I.** Problem 14.6 (part b only) on page 352 (helpful fact ... $\bar{x} = 105.84$)

4. **Confidence intervals II.** Problem 14.7 on page 354.

5. **Confidence intervals III.** Problem 14.8 on pages 354-355.

6. **Confidence intervals and testing ... calibrating a scale.** (Based on Problem 14.33 on page 360.) To assess the accuracy of a laboratory scale a standard weight that is known to weigh 10 grams is weighed repeatedly on the laboratory scale. The purpose is to see if the scale is free of bias or in other words to see if the mean scale reading over many trials is equal to the known true weight (10 grams). The repeated scale readings are assumed to be approximately normal with standard deviation .0021 grams.

   (a) What is the population parameter that we are interested in estimating?

(b) The standard weight is weighed 10 times and the mean of the resulting measurements is 10.0022 grams. Give a 95% confidence interval for the mean of repeated measurements of the standard weight on this scale.

(c) Based on this confidence interval do you believe the scale is free of bias? Explain your answer.

(d) A testing approach ... assume for the moment that the scale is free of bias (so that $\mu = 10$ and $\sigma = .0021$). Use the sampling distribution of $\bar{X}$ to find the probability that the mean of 10 weighings is greater than or equal to 10.0022 grams. What does this probability calculation suggest about the plausibility that the scale is free of bias?

7. **Project - planning.** The purpose of the Statistics 7 course project is to have each student get some experience applying the ideas of the course to a question or topic that interests them. There are several options: these include designing and carrying out a small experiment or data analysis, reading and discussing a published study/article, and reporting on a statistical topic not covered in class. These are described in more detail below. You are encouraged to work in groups of size 2-4 (although individual projects are permitted). The end result should be a brief report (2-5 pgs). The project will be graded based on appropriateness of statistical methods, clarity of the report, and originality. The project will be due Friday June 6.

   A project proposal is due Monday May 19. You will lose some credit on the project if you do not turn in a proposal on this date. The proposal should identify the group members working on the project and provide a one paragraph description of the intended project. Please feel free to consult with me concerning ideas for projects before you prepare your proposals.

   Data collecton and analysis: You are especially encouraged to collect data of your own to answer a particular question. This would mean planning an experiment, carrying it out and then analyzing the data. Your report should describe the question that motivated the study, your methods for collecting data, the analysis, and your conclusions. It is important to think carefully about and provide an adequate description of the methods used to design the study and collect the data (was it a randomized comparative study? was it a simple random sample?) along with any possible biases that might affect the conclusions. The data analysis could rely only on exploratory tools (graphs and numerical summaries) or make use of formal inference procedures from the second half of the course. Possible topics: — Vary the design of paper airplanes and see which design flies better. — Examine the distribution of chips in chocolate chip cookies (how many per cookie)?

   Does the mean differ from brand to brand? — Discover whether attendance in class affected by the weather?

   Analysis of a data set that is available to you: You might find a data set in another class, in a book, or in a magazine that you are interested in investigating. For example, data could be obtained from a computer magazine to investigate the relationship between price and memory or price and machine speed. Or as another example, you might explore baseball player salaries and their relationship to performance. As with the preceding type of project, you should pay critical attention to issues concerning how the data were collected as well as to the statistical analysis.

   Review of an article: If you would rather not collect and analyze data, you may read an article in which statistics or probability is applied to a problem that interests you. You should describe what was done in the application, and then discuss strengths and weaknesses in the use of statistical reasoning. Appropriate articles can be found in many scientific journals or in books like the following: Quantitative Analysis of Social Problems - Tufte
   Statistics and the Law - DeGroot, Fienberg and Kadane
   Pygmalion in the Classroom - Rosenthal, Jacobsen
   Mismeasure of Man - Gould

   Report on a statistical method: Read about and describe a technique that we don't discuss in class, e.g., control charts in Chapter 11, conditional probability in Chapter 12, Chapter 23. If you use a section from a textbook, you should show your understanding by doing a couple of problems from the end of the section. The challenge with this type of project is to be original and make your topic interesting to the reader (that's me!).