STATISTICS 7 – SPRING 2008
Homework 7

Handed out: Friday May 16, 2008
Due: Friday May 23, 2008 (by 5pm to Bren Hall 2216)

Reading:	May 16	Statistical analysis with one sample - mean (Ch 18)
	May 19 - May 23	Statistical analysis with one sample - proportion (Ch 20)

NOTE: The 2nd midterm exam is scheduled for Wednesday May 21. It covers relevant material from Chapters 1-16. It will emphasize material since the last exam: Chapters 3, 10-16.

- The format will be similar to last exam (formula sheet, calculator, approx 3 problems).
- Questions from previous offerings of Stat 7 that are relevant are on the course web site.
- Homework questions 1, 2, 3 are relevant for the exam; questions 4 and 5 are not covered on the exam.
- Review sessions: Monday 5/19 4-5pm (Kameryn) and Tuesday 5/20 4-5:30pm (Hal), both in Bren 2011.

NOTE: A short (one paragraph) description of your plan for a course project is due Monday May 19. If you are working with other students, then only one description is required for the group. (For more details see Homework 6)

1. **Multiple choice questions on the basics of significance testing.** Problems 15.26-15.35 on pages 381-382. There are many questions but remember the answers are in the back! Please show your work.

2. **Significance testing - the effect of sample size.** (Loosely based on problem 17.29 on page 422.) A new apartment complex advertises that the mean apartment size is 1250 square feet. A tenant group is concerned that the apartments may be smaller than advertised. They hire an engineer to measure a random sample of 18 apartments. The data are:

   1267, 1275, 1250, 1275, 1208, 1275, 1283, 1166, 1208,
   1292, 1275, 1233, 1183, 1267, 1258, 1183, 1216, 1208

   Suppose that apartment sizes vary normally with standard deviation 40 square feet due to variation among models and variation introduced during construction. We analyze these data to determine if there is good reason to conclude that the apartment complex's advertisement is false.

   (a) Identify the population parameter that this study is concerned with.
   (b) What is the null hypothesis regarding this parameter? What is the alternative hypothesis regarding this parameter?
   (c) The sample mean across the 18 apartments is 1240.1. Find the P-value for evaluating the null hypothesis.
   (d) What conclusion do you reach regarding the null hypothesis?
   (e) Suppose the study were repeated with $n = 36$ apartments and the same mean was obtained. What would the P-value be then?
   (f) Suppose the study were repeated with $n = 72$ apartments and the same mean was obtained. What would the P-value be then?
   (g) Explain how the sample size is related to the conclusion of the significance test.
   (h) The complex owners complain because the tenant's engineer was lazy and only sampled apartments on the first two floors of the buildings. Explain why this is a concern for the data analysis.

3. **Significance testing - a two-sided alternative.** (Based on problem 17.30 on page 422.) It is commonly said that "normal" body temperature is 98.6 degrees (Fahrenheit). But is this true? The data below represent daily average body temperatures for 20 healthy adults:

   98.74, 98.83, 96.80, 98.12, 97.89, 98.09, 97.87, 97.42, 97.30, 97.84,
   100.27, 97.90, 99.64, 97.88, 98.54, 98.33, 97.87, 97.48, 98.92, 98.33

   Do these data give evidence that the mean body temperature for all healthy adults is not equal to the traditional 98.6 degrees? (Suppose that body temperature varies normally across the population with standard deviation 0.7 degree.)

(a) We are interested in testing the traditional claim that the mean body temperature of healthy adults is 98.6 degrees. The sample mean body temperature of the 20 individuals is 98.2. Perform a significance test to evaluate the traditional claim (identify population parameter of interest, state null/alternative hypothesis, find the P-value, and state your conclusion). (Hint: This question requires a two-sided alternative hypothesis which means that when we ask for the probability of a "more extreme" outcome we should ask about the probability of values less than 98.2 or greater than 99.0 (the latter would be more extreme in the opposite direction). See pages 368-370 for more details.)

(b) What conclusion do you reach regarding the null hypothesis?

(c) Find a 95% confidence interval for the mean body temperature in the population. Is 98.6 in the interval?

(d) Find a 99% confidence interval for the mean body temperature. Is 98.6 in the interval?

(e) Discuss the relationship between the two confidence intervals (and whether they contain 98.6) and the result of the two-sided significance test.

4. **t-procedures I.** Problem 18.26bc on page 452. Read part (a) but do not complete. (Helpful hint: The sample mean for these data is 13.15.)

5. **t-procedures II.** (Based on problems 18.29 and 18.31 on pages 453-454.) A study of unexcused absenteeism among a factory's workers looked at one year's records for the 668 workers in the factory. There is some concern that the factory has a problem with absenteeism. The industry standard is an average of 9 days absent per year. The mean number of days absent at this factory was 9.88 and the standard deviation was 17.85 days. In this case we have examined all of the workers at the factory for the given year so it is not obvious what population they represent – the owner wants to draw conclusions about "all workers at the factory in all years" (assuming the same factory/employment conditions). Thus we regard the present data as a random sample of all workers at the factory in all years (under the same conditions of employment).

   (a) The number of days absent can't be negative. The s.d. is much greater than the mean. This tells us the distribution of days absent can't be normally distributed. Explain and provide a hypothetical sketch of the distribution of number of days absent.

   (b) Explain why we can still use the t-procedures to learn about the population mean even though the population distribution (of days absent per worker) is not normal.

   (c) Test the hypothesis that mean absenteeism at this factory is the same as the industry standard. (What is an appropriate alternative hypothesis?) What do you conclude?