STATISTICS 500 – Fall 2009 – Homework 2

DUE DATE:
  on campus   Friday, 11 Sept 2009, in lecture (11 am)
               or by e-mailto Chuanlong, dclong@iastate.edu, no later than noon.
  off campus   Monday, 14 Sept 2009, by 4 pm to Nicole Rembert, email: rembeall@iastate.edu or
               FAX: 515-294-4040 (please include cover page with Stat 500 / Nicole Rembert).

1. Inference based on t-statistics – Financial statisticians are interested in the volatility of a market. When
   volatility is low, prices change very little over time. When volatility is high, prices change a lot. The direction
   of the change (up or down) is not important. All that matters is the magnitude of the change. The following
   data are a small part of a larger study on volatilty of stock prices.

   There are three major stock exchanges in the US. This small study concerns the volatility on two exchanges
   (NYSE and NASDAQ) during the week of 12-16 Feb 2007 On each exchange, 40 stocks were randomly selected
   from all stocks listed on the exchange that week. The volatility was calculated for each stock as the absolute
   value of the proportional change in stock price. A value of 0% indicates no change in stock price. An exchange
   with an average change of 5% is less volatile than one with an average change of 10%.

   Summary statistics for each exchange are:

Exchange	n	average	s.d.
NYSE	40	2.91%	2.23%
NASDAQ	40	4.39%	3.37%

   (a) Is this study a randomized experiment? Explain briefly.

   (b) Can these data be used to make inferences about the mean volatility of all stocks on the NYSE during
       the week of 12-16 Feb 2007? Explain briefly.

   (c) Estimate the pooled standard deviation, s_p. How many degress of freedom does it have?

   (d) For this and subsequent parts, use a two-sample t-test and assume that all assumptions are reasonable.
       Find the P-value for testing the hypothesis that volatility is the same on the two exchanges versus a
       two-sided alternative. Write an appropriate one sentence conclusion.

   (e) Your company is currently listed on the NASDAQ. A consultant recommends that you apply for listing
       on the NYSE because 'this study shows that switching from the NASDAQ to the NYSE will reduce the
       volatility of your companies stock by an average of 1.48%'. Do you agree? Explain briefly.

   (f) Assume that the sample size was 20 in each group, but the observed mean and s.d. are unchanged.
       Calculate the P-value.
       Hint to save some computation: the pooled s.d. does not change.

   (g) Assume that the sample size was 100 in each group, but the observed mean and s.d. are unchanged.
       Calculate the P-value.

   (h) How does the conclusion of your t-test change as the sample size increases? Explain (briefly) why this is
       reasonable.

   (i) Calculate a 95% confidence interval for the difference in mean volatility (NASDAQ - NYSE).

   (j) Repeat part 1i assuming that $n = 20$ in each group (again assume that the mean and s.d. are the same).
       Estimate the difference and give a 95% confidence interval . How does your answer compare to the answer
       in (1i)? Explain (briefly) why this is reasonable.

2. Sample size calculations – The investigators are planning on repeating the study and they come to you for
   advice. They want to know how many stocks to sample in each exchange (remember, n was 40 in the previous
   experiment).

   (a) To start with, the investigators ask you to compute the power if they repeat the study with the same
       sample sizes. Approximately, what is the power of a $\alpha = 0.05$ 2-sided t-test when $\delta = 1.48\%$, $\sigma = S_p = 2.85\%$, and $n = 40$ in each group?
       HINT: look at the equation for δ on p 33 of the notes. Solve for the term involving the power. You

should have values for everything in the equation except for the term involving the power, so you can calculate that. Remember that a t distribution is symmetrical. This should give you enough information to calculate power $= 1 - \beta$.

(b) Now, you are asked to suggest a sample size (n) for the new study. The investigators will use a 2-sided t-test with size $(\alpha) = 0.05$. They want to make sure that their test has a power of 80% to detect a difference of 1.48%. They believe that a reasonable value for $\sigma = S_p$ is 2.85%. Please calculate the required sample size for them (n in each group). If the t-table in the text does not include the d.f. you need, use the closest d.f. that is in the table.

(c) What if the investigators chose to use the s.e. approach? Again, assume that $\sigma = S_p = 2.85\%$. What n would you recommend if the investigators want the s.e. of the difference to be smaller than 1.48%?

(d) And, what if the investigators chose the ci width approach? What n would you recommend if the investigators want the width of the 95% c.i. to be less than 1.48%? Again, $\sigma = 2.85\%$.

3. Effect of speed limits on traffic fatalities –
Most of you have driven a car only in the era of high freeway speedlimits (e.g. 65 mph, 70 mph, 80 mph). For many years until 1995, the US national speed limit was 55 mph. A federal law in late 1995 abolished the national limit and allowed states to set their own limits. Many, but not all, states promptly increased the freeway speed limit. The following data on change in freeway traffic fatalities were collected as part of a study on the effects of changing the speed limit.

Thirty-two states increased the speed limit either at the beginning of or during 1996. Nineteen states left the speed limit unchanged. The data include the District of Columbia, which is why there are 51 observations. The data file, speedlimit.txt, contains the state, whether or not it increased the speedlimit, and the % change in traffic fatalities from 1995 to 1996. The % change is used as a response because states with large populations will have more fatalities, both in 1995 and in 1996.

Imagine you work in the Department of Transportation in one of the states that did not change its speed limit. Your boss has asked to analyze the data and write a short report summarizing your findings. Your boss is specifically interested in:
Was the change in traffic fatalities the same in the two groups of states?
How large is the difference and how precisely is that estimated?
Your state legislature is considering increasing the speed limit. They ask you to predict the effect on traffic fatalities in your state if they were to do that. What do you tell them?

You may assume that the standard model (independent, normally distributed errors with common variance) is reasonable.

Your report should include one sentence (or more if necessary) describing your analysis and an answer to each of your boss's questions, using appropriate statistical summaries. Your boss is a busy person so she only wants to see results from a single analysis. If you analyze the data many ways, present only the one that you believe is the most appropriate.

Note: We'll talk about using SAS to do these computations in Tuesday's lab. The code is in your lab packet. There are many other ways to do the computations if you don't want to wait until Tuesday.

4. Speed Limits - minor detail, very few points —
None of your boss's questions involve (at least directly) the s.e. of the mean for one group of states. Think carefully about the standard model (p. 27 of the lecture notes). Why is the usual s.e. of the mean (or a confidence interval for the mean) the wrong thing to calculate for these data?

Note: The stuff I ask for in problem 3 is appropriate, in spite of my concern about the s.e. of the mean.