Project 6 - Estimation

Statistics 401: Fall 2006

Due Wednesday, October 25

1. Do problem 9.2 on page 367 of your textbook.
2. Do problem 9.8 on page 368.
3. Do problem 9.14 on page 379.
4. Give one advantage and one disadvantage of using a 99% confidence interval instead of a 90% confidence interval.
5. Do problem 9.16 on page 379.
   (a) Is this an experiment or an observational study? Explain.
   (b) Give the individuals being measured.
   (c) Give the variable being measured as well as the sample space of all possible outcomes.
   (d) Give a point estimate for the proportion of girls who would stop using Planned Parenthood if their parents were informed.
   (e) Is the sample size large enough to assume that the sample proportion p has an approximate normal distribution? If yes, and if the true value of $\pi = .4$, then give the sampling distribution of p, its mean and its standard deviation.
   (f) Construct a 95% CI.
   (g) Interpret the CI in terms of the problem.
   (h) To what population would it be reasonable to generalize the CI estimate?
6. Do problem 9.20 on page 379.
7. Do problem 9.42 on page 393. Put your results for part (a) in a table. In addition to parts (a) and (b), answer the following:
   (c) Is the sample size large enough to assume that the conclusions from the CI's are appropriate? Why or why not?
   (d) What must we assume about the data so that the conclusions from these CI's are valid.
   (e) Interpret the 95% CI for running in terms of the problem.
8. Do problem 9.46 on page 391.
9. In the Discover article *Malaria Parasite Makes Humans Smell More Attractive to Mosquitoes* on the STAT401 website, researchers observe 100 mosquitos and find that 67 of them are attracted to kids carrying gametocytes - the transmissible, reproductive stage of the parasite *P. falciparum* - in their bloodstream while only 33 mosquitos were attracted to other kids.

(a) Is there enough data to assume that the sample proportion p has an approximate normal distribution? Why or why not?

(b) Calculate a 99% CI for the true proportion of all mosquitos that are attracted to kids carrying gametocytes.

(c) Interpret the 99% CI in the context of this problem.

(d) Can we be confident that a majority of mosquitos prefer kids with gametocytes versus other kids? Why or why not?

(e) If the researchers want to cut the margin of error **in half**, how many mosquitos should they observe?

10. The Environmental Protection Agency has established an air quality standard for lead of 1.5 μg/m^3. Listed below are measured amounts of lead (in micrograms per cubic meter or μg/m^3) in the air recorded at Building 5 of the World Trade Center site on different days immediately following the destruction caused by the terrorist attacks of September 11, 2001. After the collapse of the two World Trade Center buildings, there was considerable concern about the quality of the air. The data file "lead.txt" can be found on the Stat 401 website.

5.40 1.10 0.42 0.73 0.51 1.10 0.66 1.02 0.45 0.69 0.72 0.55

(a) If the data is not normal, explain why the sample mean $\overline{X}$ may not be approximately normal.

(b) Does the evidence suggest that the data is not normal? Use the techniques, including appropriate graphs and the correlation test from Chapter 7 to answer this question.

(c) Transform the data. Use Box-Cox to determine the appropriate transform. Which value of λ are you using?

(d) Use plots and the correlation test to make sure that your transform worked.

(e) Create a 95% CI to estimate the mean amount of lead in the air at the World Trade Center on the days immediately following 9/11. Put the sample mean of the untransformed data, the sample standard deviation of the untransformed data, and the CI endpoints in a table.

NOTE: Since you transformed the response, find the CI on the transformed scale, then back-transform the endpoints of the CI. For example, if you use a lambda of $\frac{1}{2}$ (square root) as the optimal power to transform the data to a normal distribution and the CI calculated based on the transformed data is (2.05, 4.26), then you can back-transform the end points by raising each end point to $\frac{1}{\lambda}$, or in this example, squaring each end point, so the CI on the original scale is (4.20, 18.15).

(f) Confirm that the back-transformed CI is appropriate. *Hint*: Estimate $\frac{\sigma_x^2}{\mu_x^2}$ where μ_x and σ_x are parameters of the original lead measurements X.

(g) Interpret the 95% CI in the context of this problem.

(h) Does the EPA standard appear to be met? Why or why not?