Project 4

STAT 401: Fall 2006 *Due Wednesday, October 11*

For the probability calculations in this project, you must show your work.

For the last problem, you will use R to check for normality of a data set and you will perform a normalizing transformation. You will need to assemble the tables, plots, and answers into a coherent write-up. Remember to label all of your graphs, and reference them from the body of your report as outlined in Chapter 4 of the Course Notes. Put ALL R code in an Appendix.

Your write-up must be typed. Please number your answers as the questions were numbered. Your grade will be determined by how well you answer the questions and by the professionalism and clarity of our write-up.

1. Use the January 2006 *Discover* article "Inhaled 'Cuddle' Hormone Promotes Trust" from the STAT401 web-site to answer the following questions:

   (a) Assuming that the experiment was a completely randomized design, how must the experiment have been conducted?

   (b) Assume that the individuals in the experiment are **not** from a random sample. Second, assume that that the experiment was a CRD. What type of conclusion can be drawn from this experiment?

   (c) Suppose that the percentages given in the article are truly representative of all U.S. adults. Further suppose that a random sample of four U.S. adults is taken.

      i. If none of these four adults whiffs oxytocin, what is the probability that exactly one adult decides to transfer their entire bankroll to a trustee?

      ii. If two of the adults whiff oxytocin and two do not, what is the probability that exactly one adult transfers their entire bankroll to a trustee?

2. Use the January 2006 *Discover* article "Malaria Parasite Makes Humans Smell More Attractive to Mosquitoes" from the STAT401 web-site to answer the following questions:

   (a) What is the probability that out of the 100 mosquitos in the experiment, that a randomly selected mosquito flies towards kids carrying gametocytes?

   (b) Sixty seven of the 100 mosquitos in the experiment fly towards kids with gametocytes. When randomly selecting (**without replacement**) two mosquitos out of these 100 mosquitos, what is the probability that BOTH mosquitos go to the tent with kids carrying gametocytes?

   (c) Suppose that the probability given in problem #2a is true for all mosquitos. Two friends are out camping in separate tents. One is carrying gametocytes and the other is not. Two mosquitos find the two friends in the middle of the night. Assuming independence, what is the probability that BOTH of the mosquitos fly towards the infected kid?

   (d) Your answers for problems #2b and 2c should be almost identical. Why is this case? What property about simple random samples from a finite population support your answer?

   (e) Note that there are NO treatments in this "experiment."

      i. Who are the individuals?

      ii. What is the response?

3. Do Exercise 7.28 on page 316 of your textbook.

4. To determine if the results from the mosquito experiment in problem #2 are repeatable, many more experiments will be performed. For each experiment, the sample proportion of the 100 mosquitos who fly towards the tent which holds the kids infected with gametocytes will be calculated. So the sample proportion p has a **distribution**, and the distribution is approximately normal

$$p \sim N(\mu = \frac{2}{3}, \sigma = \frac{\sqrt{2}}{30})$$

when the data is from a large random sample. (This fact follows from the Central Limit Theorem).

   (a) What is the probability that in an experiment, the sample proportion of mosquitos who fly towards the infected tent is over 75%? *Hint:* Take a square root in R by using **sqrt( )**.

   (b) What is the probability that in an experiment, the sample proportion of mosquitos who fly towards the infected tent is exactly $\frac{2}{3}$?

5. Moore and McCabe's Introduction to the Practice of Statistics present data from 31492 calls made to the customer service center of a small bank in a month. The file ServiceCalls.txt, which can be downloaded from the Stat 401 website, contains 100 measurements from this data set.

   (a) Construct a density plot, a boxplot, and a normal probability plot to display the distribution of the calls data. Include these plots in your report.

   (b) Provide a **summary table** displaying the sample mean, sample standard deviation, and and five number summary for the data.

   (c) Compute the correlation between the normal scores and the calls data. Since Table 7.1 on page 320 of your textbook does not have a critical r value for $n = 100$, use critical r of .98 to check for normality of the data.

   (d) Does the distribution of calls data appear to depart from a normal distribution? If so, why? Use more than a single output from R to justify your answer.

   (e) Regardless of your answer to problem #5d, use the boxcox function to estimate the optimal lambda value to use in the transformation. Include the plot in your report. Your chosen λ value should be in the confidence interval, preferably near the center. It is best to choose a simple value such as $-1, -\frac{1}{4}, \frac{-1}{2}, 0, \frac{1}{4}, \frac{1}{2}, 1$, or 2 rather than a value such as -0.47352.

   (f) Power transform the data using the λ you just chose. For example, if you chose $\lambda = -\frac{1}{2}$, then in R execute:

```
> calls.transformed = calls^(-1/2)
```

   (g) Construct a density plot, a boxplot, and a normal probability plot to display the distribution of the transformed calls data. Include these plots in your report.

   (h) Provide a summary table displaying the sample mean, sample standard deviation, and five number summary for the transformed calls data.

   (i) Compute the correlation between the normal scores and the transformed calls data. Use .98 as the critical r value to check for normality.

   (j) Does the distribution of the transformed calls data appear to depart from a normal distribution? If so, why or why not? Use more than a single output from R to justify your answer. Did the transformation work?