Project 4 *STAT 401: Fall 2016*

Due: 1:40pm on Tuesday, October 18

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, your justification for your answers, and by the organization and clarity of our write-up.

1. Do exercise 3.4(a)-(f) on p. 158. Make sure to:

   (b) Provide the formula and calculations that you use.

   (d)-(e) Indicate what probability calculation you are performing to answer these parts. If you use R to do the probability calculations, provide your code and output. If you use a table, indicate what table you are using.

2. Do exercise 3.12 on p. 160. If you use R to do the probability calculations, provide your code and output. If you use a table, indicate what table you are using. In addition to 3.12(a)-(d), also do:

   (e) 10 motorists are clocked by the California Highway Patrol driving down I-5. What is the probability that at least 1 is speeding? Recall from the 3.12(d) that the speed limit is 70 mph.

3. A random sample of 100 calls made to the customer service center of a small bank in a month was collected. The length of each call, in seconds, is given in the file ServiceCalls.txt, which can be downloaded from the Stat 401 website.

   (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 #3d, 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?