Stat414/614; Fall 2024; Worksheet 02; 20 Points; NAME:

1. Use the `DownloadFestival` data set for this part. The dataset and its analysis is available in DSUR book.

   (a) Plot histograms and boxplots for the hygene scores (day1,day2,day3) for the three days of the Download Festival data set.

   (b) Add normal curves (with appropriate mean and standard deviations) to the above histograms and comment on the appropriateness of Normal distribution to model each data set.

   (c) Create Q-Q plots for the variables day1, day2 and day3. Comment on the appropriateness of Normal distribution to model each data set.

2. The dataset halibut (available in the R package EnvStats) has two variables: Annual Catch Per Unit Effort (CPUE), and biomass and exploitable biomass of Pacific halibut for the years 1935 through 1989. Dataset is analyzed in the Millard and Neerchal book. (Warning: It uses SPLUS, a precursor to the R package. Syntax is very similar, but not identical.)

   (a) Compute summary statistics for each variable.

   (b) For each variable, create a histogram, boxplot, strip plot, quantile plot, normal Q-Q plot, and Tukey mean-difference Q-Q plot.

   (c) Do both variables appear to come from a normal popula-tion?

   (d) For each of the variable, look at a set of Box-Cox trans-formations by varying the transform power λ between –2 and 2 in increments of 0.5. (This can be done using the function `boccox` of the `EnvStats` package.)That is, follow the steps shown in the section Box-Cox Data Transformations and Q-Q Plots (See p. 109, Chapter 3 of Millard and Neerchal book.) Based on the plot of PPCC vs. λ and the Q-Q plots, which transformation appears to be best?

3. Use the function `stat.desc()` of R (install the relevant R package as necessary) and obtain the output of page 57 of the `Exploring Data, Exploring Assumptions, Graphs.pdf` for the variables CPUE and biomass. Provide the definition/formula for each of the statistic computed in this output. (These formulas, for the most part, are available in Millard and Neerchal book. You should be able to look up the index and find the formula.) Draw conclusions on the normality (or lack of) of the variables.