Stat 453 - Rossi

Lab/HW #4: Due Thursday 11/8/18

1. Use the data in `pgatour200.xlsx` to perform the analysis outlined below.
   (a) Summarize the quantitative variables in this dataframe.
   (b) Create a boxplot of each variable. Use as few plots as possible.
   (c) Create a density histogram of `AveDrivingDistance` with density line.
   (d) Create a normal plot of `AveDrivingDistance` with the line.
   (e) Find a probability model that fits this reasonably well. Fit this model and report the estimates of the model parameters and their standard errors.
   (f) Estimate the 75th percentile of `AveDrivingDistance` from the observed data using the `quantile` command.
   (g) Estimate the 75th percentile of `AveDrivingDistance` from the fitted model using the `qmodelname` command (i.e., `qexp` if exponential model).
   (h) Bootstrap the 75th percentile computed by the `quantile` command. Report a histogram, boxplot, and normal plot of the bootstrap sampling distribution along with the estimated standard error of the sample 75th percentile estimate.
   (i) Describe the shape of the sampling distribution of the sample 75th percentile.
   (j) Determine the 5th and 95th percentiles of the sampling distribution of the sample 75th percentile.

2. Use the data in `pgatour200.xlsx` to perform the analysis outlined below.
   (a) Compute the pairwise correlations between the quantitative variables.
   (b) Create the variable `lnPrize` which is the natural logarithm of `PrizeMoney`. Add this variable to the dataframe.
   (c) Compute the correlation of `lnPrize` with all variables except for `PrizeMoney`.
   (d) Create a scatterplot of `lnPrize` versus `GIR` (greens in regulation) with LOESS line. Identify any outliers in the plot and summarize the scatterplot.
   (e) Bootstrap the sampling distribution of the correlation between `lnPrize` and `GIR`. Report a histogram and normal plot of the sampling distribution along with the estimated standard error of the sample correlation coefficient.
   (f) Fit the model `lnPrize`$\sim$`GIR`. Report the model diagnostic plots, the fitted equation of the line, the estimated coefficients and their standard errors.
   (g) Estimate the mean value of `lnPrize` when `GIR=65` using the command `predict.lm`. You will need to create a dataframe `new` using the command

$$\texttt{new} = \texttt{data.frame}(\texttt{GIR} = 65)$$

   and use the option `se.fit=T` so that the standard error of the estimated mean is reported.
   (h) Transform the fitted value in part 2g back to $ units.

3. Use the data in `Titanic-Survival-Data.xlsx` to perform the analysis outlined below.
   (a) Summarize the variable `Age`.
   (b) Use the `table` command to tabulate the number of survivors and non-survivors.
   (c) Use the `table` command to tabulate the number of males and females on board.
   (d) Use the `table` command to tabulate the number in each passenger class.
   (e) Use the `table` command to cross-tabulate the number of survivors and non-survivors by gender.

(f) Use the `table` command to cross-tabulate the number of survivors and non-survivors by passenger class.

(g) Use the `table` command to cross-tabulate the number of survivors and non-survivors by whether they were a child or an adult.

(h) Estimate the proportion of males that survived and its standard error.

(i) Estimate the proportion of females that survived and its standard error.

(j) Estimate the proportion of coach class passengers that survived and its standard error.

(k) For the children on the Titanic, estimate the probability of survival and report its standard error.

4. Use the data in `Titanic-Survival-Data.xlsx` to perform the analysis outlined below.

(a) Estimate the median age of the males that survived.

(b) Estimate the median age of the females that survived.

(c) Bootstrap the sampling distribution of the median age of the males that survived. Report a histogram, normal plot, and boxplot of the sampling distribution of the sample median along with the estimated standard error.

(d) Bootstrap the sampling distribution of the median age of the females that survived. Report a histogram, normal plot, and boxplot of the sampling distribution of the sample median along with the estimated standard error.

(e) Estimate the median age of the first class passengers that survived.

(f) Estimate the median age of the coach class passengers that survived.

(g) Bootstrap the sampling distribution of the median age of the first class passengers that survived. Report the sample median along with deciles of the sampling distribution and the estimated standard error.

(h) Bootstrap the sampling distribution of the median age of the coach class passengers that survived. Report the sample median along with the quartiles of the sampling distribution and the estimated standard error.