Simple Regression

EPsy 8251

Assignment #8

You will use the data sets *NFL-FCI.csv* and *NFL-Meta-Data.csv* to examine whether age of a stadium predicts variation in FCI. In this assignment, the variable `fci` will be used as the outcome. Please submit your responses to each of the questions below in a printed document. All graphics should be resized so that they do not take up more room than necessary and all should have an appropriate caption. Any equations should be appropriately typeset within the document. There are 19 points possible for the assignment (each question is worth one point).

PREPARING THE DATA

1. Prepare the data by using the `merge()` function (or some other function) to merge the two data sets based on team name. Compute a new variable in this data set that provides the current age of the stadium. Use the `head()` function to print out the first six rows of this new data frame.

EXAMINING THE RELATIONSHIP BETWEEN AGE OF THE STADIUM AND FCI

1. Create a publication quality plot of the distribution of FCI conditioned on age of stadium (i.e., a scatterplot). Put FCI on the y-axis and age of the stadium on the x-axis. Be sure to format the figure according to APA (e.g., appropriate figure numbering and caption). Also be sure to appropriately size the figure so as not to take up too much page space. **(2pts.)**
2. Descibe the relationship between age of stadium and FCI. Be sure to comment on the form, direction and strength of the relationship. Also comment on any potential observations that deviate from following this relationship (unusual observations or clusters of observations). **(2pts.)**
3. Based on your answer to the previous question, is the Pearson correlation coefficient an appropriate summary measure of the relationship? Explain.
4. Compute and report the Pearson correlation coefficient between age of stadium and FCI.

REGRESSION MODEL: DESCRIPTIVE

5. Fit a regression model that regresses FCI on age of the stadium using the `lm()` function. Write the fitted regression equation using Equation Editor (or some other program that correctly types mathematical expressions). Be sure the equation is labeled and numbered according to the APA format.
6. Interpret the value of the intercept from the regression equation using the context of the data.
7. Interpret the value of the slope from the regression equation using the context of the data.
8. Compute the value for R^2 using the ANOVA output obtained from the `anova()` function. Show your work.
9. Interpret the value of R^2 using the context of the data.
10. Create a publication quality plot that displays the regression line based on the analysis. This plot should also include the data plotted as a scatterplot. The data should be semi-transparent, and the regression line should be completely opaque (non-transparent). Be sure to appropriately number, label, and re-size the plot for publication. **(2pts.)**

REGRESSION MODEL: INFERENCE

11. Using symbols, write the null hypothesis related to the "variation accounted for" parameter that is tested by the F-statistic in this analysis.

12. Write no more than two sentences (to be included in a publication) that provide a description of what is being tested by the F-test and the results. This description should report the F-statistic, df, and p-value. It should also indicate what the results suggest about the null hypothesis and what this means about the potential relationship between age of stadium and FCI. **(2pts.)**

13. Using symbols, write the null hypothesis that is tested by the t-statistic for the slope.

14. Based on what is being tested in each hypothesis, explain why the p-value for the t-test assoicated with the slope coefficient and the p-value associated with the model F-statistic are the same. **(2pts.)**