Multiple Regression and Transformations

EPsy 8251

Assignment #9

You will use the data set *FCI2.csv* to examine whether a set of predictors predicts variation in FCI. In this assignment, the variable `fci`, or a transformation of `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 25 points possible for the assignment (each question is worth one point).

PREPARING THE DATA

There are no points for anything in this section. However, this preparation is necessary to complete the remainder of the assignment.

- Prepare the data by using the `merge()` function (or some other function) to merge the three data sets.

- Compute a new variable in this data set that provides the current age of the stadium.

TRANSFORMING THE OUTCOME

1. Create a box-and-whiskers plot of FCI conditioned on league. What does this plot foreshadow about the assumptions of normality and homogeneity of variance? Explain.

2. Re-express FCI by computing the natural logarithm of FCI. Re-create the box-and-whiskers plot of the transformed FCI conditioned on league. Do the assumptions of normality and homogeneity of variance seem better satisfied? Explain.

3. Regress the transformed FCI variable on league. 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.

4. Based on the output from the regression summary, are there differences in FCI between leagues? Explain.

5. Use the `predict()` function to compute the predicted FCI (both transformed and untransformed) for each league. Show your syntax and output.**(2pts.)**

AGE OF THE STADIUM

6. Create the density plot of the age predictor. Explain why we might want to log-transform this predictor?

7. Re-express age by computing the natural logarithm of age. BE CAREFUL...before you transform, look at the `summary()` of age.

8. Create a scatterplot of the transformed FCI values versus the transformed ages. Put the transformed FCI values on the *y*-axis and the transformed ages on the *x*-axis. Add the loess smoother to this plot.

9. Does the scatterplot suggest any further transformations. Explain.

10. Regress the transformed FCI variable on the linear and quadratic effects of the transformed age variable. 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.

11. Is there statistical evidence to support the quadratic effect? Explain.

12. Describe the effect of age of stadium on FCI.

COST OF LIVING

13. Create a scatterplot of the transformed FCI values versus cost of living. Put the transformed FCI values on the y-axis and cost of living (COL) on the x-axis. Add the loess smoother to this plot.

14. Does the scatterplot suggest any further transformations. Explain.

CONTROL MODEL

If you felt that you needed to further transform the cost of living predictor, make those transformations now.

15. Fit a control model that includes the effects of age of the stadium (linear and quadratic) and cost of living (all you believe should be included) to predict FCI (transformed). 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.

16. Interpret the value of R^2 for the model using the context of the data.

17. Based on the statistical evidence, which predictors are important covariates in explaining variation in FCI? Explain.

CONTROL MODEL + LEAGUE

18. Fit a "final" prediction model. This main-effects model should include he effects of age of the stadium (linear and quadratic) and cost of living (all you believe should be included), and league. 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.

19. Based on the output from the regression summary, are there still differences in FCI between leagues? Explain.

20. Use the `predict()` function to compute the predicted FCI (both transformed and untransformed) for each league (i.e., the adjusted means controlling for all other predictors in the model). Show your syntax and output. **(2pts.)**

21. Create a publication quality plot that displays the results from you "final" model. For this plot, put the predictor age of the stadium (transformed) on the x-axis. Show the predicted lines for each league. Control out the effects for cost of living. Make sure the predicted lines are labelled, either in the plot or via a legend. Also, be sure to appropriately number, label, and re-size the plot for publication. **(3pts.)**