Checking Assumptions

EPsy 8251

Assignment #7

You will again use the data set *FCI.csv* to examine whether there are mean differences in the cost of going to a game between the four major professional sports leagues in the United States. In these data, the variable `fci` is the outcome and `league` is the predictor. 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 17 points possible for the assignment (each question is worth one point).

CHECKING ASSUMPTIONS USING THE RESIDUALS

Begin by using R to fit the omnibus ANOVA model to the FCI data. After fitting the model, fortify the model using the `fortify()` function from **ggplot2**. Use these fortified data to help answer the following questions when applicable. For those questions that you do use the data, provide any supporting graphical or numerical (i.e., analysis) evidence that you used as evidence in support of your answer.

1. Evaluate the assumption of independent errors. Explain why you believe this assumption is, or is not, tenable. (Hint: There is no test or analysis needed to answer this question.)
2. Evaluate the assumption of normality. Explain why you believe this assumption is, or is not, tenable.
3. Evaluate the assumption of homogeneity of variance. Explain why you believe this assumption is, or is not, tenable.

CHECKING ASSUMPTIONS USING THE RAW DATA

Use the non-fortified data (the original FCI and league variables) to help answer the following questions. Provide any supporting graphical or numerical (i.e., analysis) evidence that you used as evidence in support of your answer.

1. Evaluate the assumption of normality. Explain why you believe this assumption is, or is not, tenable.
2. Evaluate the assumption of homogeneity of variance. Explain why you believe this assumption is, or is not, tenable.
3. How do the plots you created for examining the assumptions using the raw data compare to the plots you created for examining the assumptions using the residuals? Can you evaluate the assumptions forthe F-test by looking at plots of the raw data rather than the residuals?

RESIDUAL PLOT

4. Using the fortified data, create a scatterplot of the residuals (on the y-axis) versus the fitted values (on the x-axis). Add a horizontal line at $y = 0$ to this plot.
5. Explain why there are only four diffeent fitted values represent in this plot.
6. Explain how you could use this plot to evaluate the homogeneity of variance assumption.

7. Explain how adding the line $y = 0$ helps us evaluate any of the model assumptions.

8. Verify that the correlation between the fitted values and the residuals is zero using the fortified data. Show your syntax and output. (Note: The computed correlation might not be exactly zero because of rounding.)

9. What would you expect the scatterplot of the residuals versus the fitted value to look like if the correlation between them is zero.

10. Explain why we would expect the correlation between the fitted values and the residuals to be zero.