Interaction Models II

EPsy 8251

Assignment #11

You will use the data set *FCI-2014.csv*, *FCI-Team-Meta-Data.csv*, and *FCI-Cost-of-Living.csv* to examine whether a set of predictors predicts variation in FCI. In this assignment, the natural logarithm of 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 13 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.

- Merge the three data sets together.
- Compute the natural logarithm of FCI as a new variable (i.e., another column in the data). Name it `Lfci`.
- Compute the current age of each stadium as a new variable (i.e., another column in the data). Name it `ageStadium`.

Use the `summary()` function to examine the `Lfci`, `league`, `cost`, and `ageStadium` variables. You should have the following output. If not, re-do the preparaton work.

```
##       Lfci      league        cost       ageStadium
##  Min.   :4.84   MLB:30   Min.   : 78   Min.   :  0.0
##  1st Qu.:5.51   NBA:30   1st Qu.: 87   1st Qu.: 12.0
##  Median :5.82   NFL:32   Median : 97   Median : 16.0
##  Mean   :5.79   NHL:30   Mean   :100   Mean   : 20.8
##  3rd Qu.:6.07            3rd Qu.:113   3rd Qu.: 22.0
##  Max.   :6.52            Max.   :146   Max.   :102.0
```

FITTING MODELS

Fit the following regression models using R. There are no points for anything in this section. You will use the output from the fitted models to answer the questions in the assignment.

Model 1: `Lfci ~ ageStadium + ageStadium`^{2}

Model 2: `Lfci ~ ageStadium + ageStadium`^{2} `+ league`

Model 3: `Lfci ~ ageStadium + ageStadium`^{2} `+ league + cost`

Model 4: `Lfci ~ ageStadium + ageStadium`^{2} `+ league + cost + ageStadium:league + + ageStadium:cost + ageStadium`^{2}`:league + ageStadium`^{2}`:cost`

SECOND-ORDER INTERACTION EFFECTS

1. Use the `summary()` function to examine the output from Model 4. Based on this output, and taking an exploratory approach to model-building, should the second-order interactions (between the quadratic effect of age of stadium and cost and league) be retained? Explain.

2. Fit a model (called Model 5) that drops both second-order interactions from Model 4. Use the `anova()` function to compare Model 5 to Model 4. Is there evidence that the second-order interaction terms need to be retained? Explain.

FINDING A "FINAL" MODEL

3. Examine the output from Model 5 using the `summary()` function. Should all of the predictors in this model be retained? Explain.

4. Drop any predictors you believe should be removed from Model 5 and fit a "final" model. (This model will be referred to as Model 6. If you believe all of the predictors from Model 5 should be retained, Model 6 = Model 5.)

5. Create a plot of the fitted regression equations for Model 6. In this plot, show the effect of stadium age on the x-axis and the effect of league through different lines. Also show the effect of cost-of-living through different lines.

6. Use the plot to explain the effect of stadium age on FCI. (If there are interactions, be careful.)

7. Based on the results from Model 6, would it be appropriate to make the pairwise comparisons between each league? Explain.

SOME STUFF[1] FOR PUBLICATION

8. Present all six (or five) models in a regression table. This table should include the estimates of the regression coefficients and standard errors, as well as some indication of statistical significance. It should also include quantitative summary information (e.g., F, R^2, etc.) for each model. Be sure your table has an appropriate caption. This table should be presented on a page that is landscape oriented (not portrait oriented). If necessary, the table can be presented on multiple pages. (If you use multiple pages, be sure to have the appropriate caption, etc. on the extra pages). (Note: For those of you using R Markdown, check out the Stargazer package to create regression tables.) **(2pts.)**

[1] A technical term

ONE MORE ANALYSIS

9. Based on Model 3, carry out the analysis to examine all of the pairwise league comparisons controlling for the linear and quadratic effects of stadium age, and the effect of cost-of-living. Report the estimated differences, the unadjusted p-values for those differences, and the Benjamani–Hochberg adjusted p-values for the differences in a table. Be sure your table has an appropriate caption. **(2pts.)**

10. Write 2–3 sentences explaining what the results of these analyses suggest about differences in the FCI (log) between leagues after controlling for the linear and quadratic effects of stadium age, and the effect of cost-of-living? **(2pts.)**