

Assignment 04

Multilevel Regression Models - Longitudinal Analysis

This assignment is intended to give you experience working with multilevel regression models to analyze longitudinal data. Use R Markdown to produce an HTML (or PDF) file that includes your responses to each of the questions on this assignment. Please adhere to the following guidelines for further formatting your assignment:

- All graphics should be resized so that they do not take up more room than necessary and should have an appropriate caption. Learn how to do this in a code chunk using [knitr syntax](#).
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document using Markdown's display equations. See [here](#) for some examples of how mathematics can be typeset in R Markdown.
- All syntax should be hidden (i.e., not displayed) unless specifically asked for. Any messages or warnings produced from loading packages should also be hidden.

For each question, specify the question number using a section header. Submit your HTML (or PDF) file and your RMD file via email that you send to both Jonathan and Andy.

For this assignment, you will use the file, *NHL-Wide.csv*. This file contains 9 years of data on 31 NHL teams. The variables in this file are:

- **team**: NHL team name
- **established**: Year the team was established
- **division**: Division the team plays in
- **X2002 – X2014**: Fan cost index (FCI) for the years 2002–2014. There are no data for 2012, since that year the NHL was locked out. The FCI comprises the prices of four (4) average-price tickets, two (2) small draft beers, four (4) small soft drinks, four (4) regular-size hot dogs, parking for one (1) car, two (2) game programs and two (2) least-expensive, adult-size adjustable caps. Costs were determined by telephone calls with representatives of the teams, venues and concessionaires. Identical questions were asked in all interviews.
- **location**: Location of the team
- **stadium**: Name of the arena where the team plays
- **yearOpened**: The year the arena opened
- **lat**: Latitude (north/south) of the arena
- **lon**: Longitude (east/west) of the arena
- **capacity**: Capacity of the arena

You will use these data to examine (1) the effect of time on FCI, and (2) the effect of latitude on FCI.

Preparation

To begin the assignment, first select only the following variables:

- The team name
- All the variables containing the yearly FCI values
- Latitude

This should result in a data frame with 31 rows and 11 columns.

Then, *gather* the data into the long format and assign it to another data frame for use in some of the analyses. This should result in a data frame with 279 rows and 4 columns. Assign the FCI values to a variable called **fci** and the years to a variable called **year** in this long data frame.

Finally, change the levels of the *year* variable to remove the **X** from each level (e.g., X2002 becomes 2002, etc.), and coerce the year variable into a numeric variable. Running `str()` on your data frame should show that the year variable is numeric or integer.

If all went well, the structure of your data should look like the following:

```
##           team      lat year    fci
## 1   Anaheim Ducks 33.80815 2002 212.28
## 2   Arizona Coyotes 33.53193 2002 214.28
## 3 Atlanta Thrashers 33.75753 2002 253.51
## 4    Boston Bruins 42.36644 2002 293.49
## 5    Buffalo Sabres 42.88380 2002 222.84
## 6    Calgary Flames 51.03743 2002 178.87
```

Data Exploration

1. Create a spaghetti plot of FCI over time for each team. In this plot, all teams should be in the same panel. Add the mean FCI over time as well. Make the teams' trajectories slightly transparent so that the mean trajectory is easily visible.
2. Create another spaghetti plot of FCI over time faceted by team. In this plot, each team should be in a separate panel.
3. Compute the mean and standard deviation of FCI conditioned on year. Present these in a table.
4. What do the plots and numerical summaries suggest about whether there differences in FCI over time? Explain.

Unconditional Models

Begin by mean centering the year predictor, and fitting the unconditional means model, the unconditional linear growth model, and the unconditional quadratic growth model, all with only a random-effect for intercept. Label these models, *Model A*, *Model B*, and *Model C* respectively.

5. Include the results from Models A–C in a table of model results. Be sure to include any goodness-of-fit and/or pseudo- R^2 measures for each model that you feel are appropriate. You will be using the results from these three models to determine the structure of the Level-1 model, so be sure to use an appropriate method of estimating the effects.
6. Based on the results of fitting the unconditional models, adopt a structure of the Level-1 model. Explain why you adopted this structure.
7. Using the adopted structure for the Level-1 model, examine which random-effects are needed. Report the pertinent results from any models you may have fitted in this investigation by writing prose that is suitable for publication. If you perform any other statistical analyses during this investigation that help you make your decision, report the results of these analyses as well. Remember to use an appropriate method of estimating the effects for examining random-effects.
8. Now that you have adopted the structures for both the fixed-effects (level-1) and random-effects (level-2) for the unconditional model, write the multilevel *and* composite equations for your adopted model.

Effect of Latitude

9. Using your adopted structure for the fixed- and random-effects, now also include an effect of latitude in (1) the Level-2 intercept model (*Model D*); (2) the Level-2 intercept and linear slope model (*Model E*); and (3) the level-2 intercept, linear slope, and quadratic slope model (*Model F*). Include the results from fitting these models in the table of model results (now should have six models included). Be sure to include any goodness-of-fit and/or pseudo- R^2 measures for the model that you feel are appropriate.
10. Which model should be adopted. Explain.

Plot of Model Results

12. Create a single display to compare the fitted equations for at least two teams (of your choice). The plot you create should highlight the effect of time on FCI. This plot should be suitable for publication. Also write 2–3 sentences that help a reader understand the effect of time and latitude on FCI. Use the teams you selected for your plot as examples to help readers better understand these effects. (If it is useful, feel free to include the average line in your plot as well.)