MAS6003(2)/MAS474 Extended Linear Models: Project

This project counts for 30% of the assessment on MAS474, and for 15% of the assessment on MAS6003.

1 Background

The dataframe `YouthSurvey` in the `YoutSurvey.RData` workspace (available on MOLE) contains data on the tolerance of 16 youths to deviant behaviour. Each year, when participants were ages 11, 12, 13, 14, and 15, they filled out a nine-item instrument designed to assess their tolerance of deviant behavior. Using a four-point scale (1 = very wrong, 2 = wrong, 3 = a little bit wrong, 4 = not wrong at all), they indicated whether it was wrong for someone their age to: (a) cheat on tests, (b) purposely destroy property of others, (c) use marijuana, (d) steal something worth less than five dollars, (e) hit or threaten someone without reason, (f) use alcohol, (g) break into a building or vehicle to steal, (h) sell hard drugs, or (i) steal something worth more than fifty dollars.

The data consists of 5 columns:

- `id` - an identification variable for each of the 16 youths included in the dataset
- `Sex` - an indicator variable taking the value 'M' if the respondent is male, and 'F' if female
- `Exposure` - this assesses the respondent's self-reported exposure to deviant behavior at age 11. To obtain values of this predictor, participants estimated the proportion of their close friends who were involved in each of the same nine activities on a five-point scale (ranging from 0 = none, to 4 = all). The value of `Exposure` is the average of his or her nine responses
- `Age` - the age of the respondent
- `Tolerance` - the respondents average across the nine tolerance questions.

2 Tasks

Use the data to investigate the tolerance level of youths as they progress through adolescence. Examples of the type of questions you may wish to consider, are whether there is a tendency for tolerance of deviant behaviour to either increase or decrease as they progress through adolescence, whether there is a significant difference between male and female youths (both in their average tolerance, and in the change in their tolerance through time), and whether previous exposure affects tolerance? There are of course many other questions you could investigate.

Predict the tolerance of a male youth chosen randomly from the same population, who reported an exposure of 1.8 to deviant behaviour. Finally, suppose that in addition, you are later told that this individual has a tolerance of 2.0 at age 11. Predict his tolerance at age 15.

It may be assumed that the respondents in this dataset were selected randomly from the population and that all respondents answered the questions independently of each other.

2.1 Advice

- To pass this assessment, you must investigate the use of random effects in your model.
- If you perform any hypothesis tests, be sure to clearly specify the hypotheses you are testing.
- You should include at least one exploratory data analysis plot, but be sure to only include graphs which show useful information.

3 Writing your report

Please use **Rmarkdown** (`http://rmarkdown.rstudio.com/authoring_quick_tour.html`) to produce your report. This will ensure that your work is reproducible, help you to avoid errors in your coding, and will ensure I can see all the steps you took in producing your solution.

Start your Rmarkdown document with the header

```
---
title: "MAS474 or MAS6003 coursework"
author: "Your student id number"
date: "7 March 2017"
output: pdf_document
---
```

Your report should be a maximum of **six pages** long. The page limit includes everything[1]. Please leave the important parts of the R code embedded in your document so that I can see what commands you used. You do not have to show the R output from your commands if it is unimportant. To do this, start your code chunks with

```
```{r results='hide'}
code
```
```

Unimportant code (preliminary data analysis, data manipulation, figure generation etc) can be included in code chunks which start

```
```{r echo=F}
code
```
```

Please be sure to show all the code used to fit models, generate confidence intervals, make predictions etc.

Marks will be deducted if you fail to write down a mathematical description of the model you have used. Latex commands can be included in Rmarkdown files in the usual way.

4 Plagiarism and collusion

Your submitted project must be entirely your own work. Do not show your report to anyone else. The University guidelines on plagiarism and collusion are available at

`http://www.shef.ac.uk/polopoly_fs/1.334794!/file/Use_of_Unfair_Means_Advice_to_Students.pdf`

5 Asking for help

As this is an assessed project, help will be limited (you can always ask, but I may say no!) and will typically be at the level of debugging R code if something strange is happening. If any aspect of the project is unclear, please ask questions via MOLE so that I can answer for the benefit of everyone.

[1]Every year, I get asked whether I really mean six pages, and whether it really includes figures etc. I do and it does. Six pages is deliberately tight - it forces you to think about what is important, and to be succinct. Anything beyond six pages won't be marked or given credit.

6 Deadlines and project submission

Submit your work on MOLE by Tuesday April 17th (12pm).

Please use a filename of the form `StudentNumber_Name.pdf`. So if your student number is 12345 and your name Aloysius, you would name your file

`12345_Aloysius.pdf`