Introduction to Bayes Regression

Exercises   
Full resource: https://www.ncrm.ac.uk/resources/online/all/?id=20843

You are going to use R to conduct some analyses. The dataset for these exercises is provided with the online material and it is called:

**private\_school.csv**

This is an extract from a dataset used in Chapter 12 “Methods Matter: Improving causal Inference in Educational and Social Science Research” by Richard J. Murnane and John B. Willett, Oxford University Press.

The main variables of interest in this extract are the following:

|  |
| --- |
| **id**: The adolescent ID in the study. |
| **read12**: Standardised reading scores in Grade 12 (M=50) |
| **math12**: Standardised math scores in Grade 12 (M=50) |
| **…**: |
| **catholic**: Whether the participant attended a private (in this dataset, Catholic Controlled) secondary school (1) or not (0) |
| **read8**: Standardised reading scores in Grade 8 (M=50) |
| **math8**: Standardised math scores in Grade 8 (M=50) |
| **female:** dummy to indicate female gender (1) or other (0) |
| **…**: |
| **faminc8**: Total annual family income in 8th grade categorised in 12 categories |
| … |
| **riskdrop8**: Number of risk factors for later drop out from none (0) to 5. |

Before you run the analyses

Before you can run the exercises, you will need to download and install the package “rethinking” and other packages related to “Stan”, as well as the “ggplot2” package. These also require the installation of Tool Chain ++. See this webpage for instructions:

<https://github.com/rmcelreath/rethinking#installation>

The installation of these packages may not always be straightforward. In my experience, if you are using Windows 10, one of the problems is the fact that Windows may automatically install R packages in a “One Drive” directory or in a “Temporary” directory. This means that R may not be able to find the installed packages in the *default* directory. If this is the case, either instructing R where to install a package, or copying the package across in the default directory, usually solves the problem.

While not necessary, using some R interface, e.g. *RStudio*, can greatly help familiarise and use R.

Exercises

Use the dataset provided to complete the following tasks:

1. Build a linear regression model where math scores in grade 12 are a function of math scores in grade 8.

1a. Consider the metric of outcome and predictors: would it be sensible to centre and/or standardise some of these variables?

1b. Create and check priors for the parameters in the model (intercept, slopes, SD).

1. Run the linear regression model specified above using the function “ulam”. Before running it, make sure that you use an ad-hoc list with the variables of interest. Run the model Inspect the output and consider algorithm diagnostics.
2. Run the linear regression model specified above again. Use the function “ulam” but specify 6 chains. Inspect the output and consider algorithm diagnostics.
3. After inspecting the poste marginal posterior distribution in tabular form, plot the observed results, the average slope, the 95% credibility interval around the mean, and the 95% credibility interval of math12 scores.