

# HW 3, SDGB 7840: Modeling Literacy Rate

Instructor: Prof. Nagaraja

Due: 3/28 in class

Submit **THREE** files through Blackboard: (a) .Rmd R Markdown file with answers and code, (b) Word document of knitted R Markdown file, and (c) your data file. Your code/Word files should be named as follows: “HW[X]-[Full Name]-[Class Time]” and include those details in the body of those files.

Complete your work individually and comment your code for full credit. For an example of how to format your homework see the files posted with Lecture 1 on Blackboard. **Show all of your code in the knitted Word document.**

In this assignment you will use multiple regression to model the literacy rate across countries; the goal is to understand which factors might be related to the literacy rate.

Use the data provided by the World Bank to determine which 10 explanatory variables to consider (<http://data.worldbank.org/indicator>; and click here for the link to literacy rate data: [link](#)). (This data requires a lot of cleaning.)

Write your paper as a research report. It should be no longer than 8 pages (this includes graphs and tables but not references) and should include the following six sections. Any pages beyond the 8th page will not be graded. You can use the posted paper, “The regional dimension of MNEs’ foreign subsidiary localization” by Arregle, Beamish, and Hebert (2009) as an example of how to determine what information is important to include in a report.

1. Executive Summary: short paragraph summarizing your paper (this is like the business version of the abstract in the Arregle, et al paper).
2. Introduction: define the literacy rate and the purpose of the study

3. Data: source of your data; discuss which 10 explanatory variables you considered and why (not just the ones you ended up including in your final model); relevant summary information about the explanatory and response variables; which countries are included in your data set; which year(s) are included in your data set; how you cleaned your data.
4. Methods: relevant plots; model building summary (transformations, variable selection, collinearity, etc.); check regression assumptions; model evaluation; relevant hypothesis tests (include hypotheses, test statistic, degrees of freedom,  $p$ -value,  $\alpha$  value and conclusion)
5. Discussion: final regression model; interpretation of model; discussion of usefulness of model; any ideas you have for improvement
6. References: cite the World Bank as your data source and any other publications you may have used to learn about the response variable or which explanatory variables may be helpful, etc. (Note: you do not have to use other sources, but if you do, cite them.) Your reference list is not included in the 8 pages. Finally, DO NOT COPY TEXT from sources; write your report in your own words and add citations.