Assignment 2: Linear Regression Review

EDP 380C.16: Spring 2021

Fall 2021 – Unique: 11585

Brian Keller, Ph.D.
bk@utexas.edu
University of Texas at Austin
September 14, 2021

Due September 23, 2021 by 11:59pm CDT

Submission and Grading

You may collaborate in **groups of two** for this assignment. If you choose to work in a group, you still must submit an individual assignment with answers supplied in your own words. You will submit the assignment via the course Canvas under Assignments.

https://utexas.instructure.com/courses/1323167

File Upload:
- Submit a PDF with your answers to the questions.
- If you worked in a group, include the other group member names as a comment on the assignment turn-in.

The assignment is worth 110 points. Each question is worth ten points. For each question, partial credit will be awarded based on the following criteria:

No Credit (0 pts.): The response did not correctly answer the question.

Insufficient (3 pts.): The response correctly answered some parts of the question, but incorrectly answered or provided information for more than one part of the question.

Sufficient (7 pts.): The response correctly answered most of the question, but incorrectly answered or provided information for one part of the question.

Full Credit (10 pts.): The response provided correct answers for all parts of the question and provided all information.

Linear Regression Review

The data are from a randomized experiment investigating the impact of a curriculum designed to improve math problem-solving skills in middle school. Students were randomly assigned to participate in an intervention or control condition. They subsequently took a math problem-solving test (the primary outcome) on four occasions (the repeated measures variables were collected in roughly two-month intervals). The primary analysis goal is to estimate the treatment (`condition`) effect for problem solving ability at wave 4 (`ability4`) while controlling for baseline ability scores (`ability1`) and baseline math self-efficacy scores (`mathse`). The order of variables, their descriptions, and their missing data rates are given in Table 1.

Table 1: `probsolv.txt` Data Description

Variable	Description
`condition`	Intervention (0 = control, 1 = intervention)
`age`	Student age in years
`female`	Biological Sex (0 = male, 1 = female)
`hispanic`	Hispanic dummy code (0 = other, 1 = Hispanic)
`lunch`	Lunch assistance enrollment (0 = none, 1 = enrolled)
`esollevel`	Ordinal Variable (1 to 6) denoting English speaking level
`stanmath`	Scores from standardized math test
`stanread`	Scores from standardized reading test
`learndis`	Learning disability indicator (0 = none, 1 = disability)
`ability1`	Math problem solving score at wave 1
`ability2`	Math problem solving score at wave 2
`ability3`	Math problem solving score at wave 3
`ability4`	Math problem solving score at wave 4
`mathse`	Math self-efficacy measurement from wave 1

Please briefly answer **all questions** below. Each question is worth 10 points.

1. Obtain the descriptive statistics for all variables in the analysis (i.e., `condition`, `ability4`, `ability1`, and `mathse`). Include a screenshot of the output.

2. Write out the regression equation for a single predictor model where ability at time four is predicted by treatment assignment. Estimate this model (either with `lm` or `rblimp`). Include the resulting output (or a screenshot of it).

3. In your own words, provide a substantive interpretation of the intercept and the slope. Note, because the treatment was randomly assigned, you can refer to these as effects.

4. Write out the regression equation for the full analysis model: ability at time four regressed on treatment assignment, and the two pretest covariates. In line with a standard ANCOVA analysis, make sure to center the pretest covariates at their means. Estimate this regression

model (either with `lm` or `rblimp`). Include the resulting output (or a screenshot of it).

5. Compute the adjusted group means for both treatment and control. You can do this either by hand or using R. Show the calculation (either by hand or the R syntax involving the computation).

6. What is the variance explained by the treatment **above and beyond** the pretest covariates? Include any calculations you did to obtain this value.

   Hint: You must estimate a model with just the covariates and compare that R-squared to the model with all predictors in it.

7. Perform a three degree of freedom F test that evaluates the joint significance of the three predictor variables (i.e., perform a test that evaluates the null hypothesis that the regression coefficients are equal to zero). Include a screenshot of the output

8. Provide a brief (one or two sentences max) interpretation of the `condition` coefficient that is suitable for the results section of a published research report. Report the significance test/p-value or provide the credible intervals.

9. Write out the regression equation for a model that includes an interaction between the pretest ability score (`ability1`) and the treatment indicator. Estimate this regression model (either with `lm` or `rblimp`). Include the resulting output (or a screenshot of it).

10. Write out the two conditional regression equations(one for the treatment group and one for the control group). Create a scatter plot with `ability4` on the y-axis and `ability1` on the x-axis. On the scatter plot, graph the two conditional regression equations. Make sure to clearly identify the treatment line and the control line clearly. You can produce this plot with whatever software/method you like.

11. Do you feel that there is justification for differential treatment effects as a function of a student's pretest ability score (i.e., the difference between treatment and control differs as a function of `ability1`)? Why or why not? Make sure to reference various statistical indices or statistical tests to justify your answer.