Problem Set 4: Effects of Educational Television

Due 4/9, adapted from an exercise developed by Will Lowe

Submit a short writeup of your answers as well as your R code (in a separate file) on Blackboard.

In this exercise we're going to look at the effect of a educational television program The Electric Company that ran from 1971-77 on children's reading scores. We will investigate what reading gains, if any, were made by the 1st through 4th grade classes as part of a randomized experiment.

This exercise is based on:

> Joan G. Cooney (1976) The Electric Company: Television and Reading, 1971-1980: A Mid-Experiment Appraisal. Children's Television Network Report.

The data comes from a two location trial in which treatment was randomized at the level of school classes.[1] Each class was either treated (to watch the program) or control (to not watch the program). The outcome of interest is the score on a reading test administered at the end of each year called `post.score`. Note that these are distinct classes from all four years. The variables are:

Name	Description
`pair`	The index of the treated and control pair (ignored here).
`city`	The city: Fresno ("F") or Youngstown ("Y")
`grade`	Grade (1 through 4)
`supp`	Whether the program replaced ("R") or supplemented ("S") a reading activity
`treatment`	"T" if the class was treated, "C" otherwise (randomized)
`pre.score`	Class reading score *before* treatment, at the beginning of the school year
`post.score`	Class reading score at the end of the school year

Read the data into an data frame named `electric`. Before we begin, let's create some variables we'll need later. First, create a dummy variable `grade.1` that takes the value of one if a student is in first grade, and zero otherwise. Repeat to create `grade.2` through `grade.4`. In addition, overwrite the existing treatment variable so that it is numerical: 1 when the class is treated and 0 when not.

Question 1

Let's consider the effect of the treatment. First, fit a linear model that predicts `post.score` with just `treatment`. Then fit a model that uses `grade` as well as `treatment`. Finally, estimate a model that uses `grade.2`, `grade.3`, and `grade.4` instead of `grade`. Note that we cannot include dummy indicators for all four grades, so we leave `grade.1` out. This means that the coefficients for `grade.2`, `grade.3`, and `grade.4` have to be interpreted relative to the first grade.

Summarize all three models in terms of how much of the variance in `post.score` they explain.

Then, consider each model's treatment coefficient. Are the estimates of this coefficient different in the three models? Why do you think that is?

[1] Classes were paired, but we will ignore that in the analysis

Question 2

Now estimate another model of the effect of `treatment` on `post.score` that in addition to controlling for `grade.2`, `grade.3`, and `grade.4` also controls for `pre.score` (the reading score before the year begins). Does the estimated effect of the treatment change? Why do you think that is? Compare the model to those in Question 1 in terms of how much of the variance of the dependent variable it explains.

Question 3

Now let's consider the effect of treatment *within* each grade. Estimate the effect of `treatment` on `post.score` (controlling for `pre.score`) *separately* for grades 1 to 4 (subsetting the data and estimating a separate regression for each). There are now *four* treatment effects. How do they differ as grade increases?

Question 4

Now let's try to learn about separate grade effects in a single model. One way to do this is to *interact* treatment with grade. Fit a model that includes `treatment`, `pre.score`, `grade.2`, `grade.3`, and `grade.4`, as well as all interactions between `treatment` and the three grade variables.

Estimate the treatment effects for grades 1 to 4 implied by this model. Do so as follows:

- Use the `predict()` command to predict the reading score for fourth-graders with a pre-treatment reading score of 80 who *received* the treatment.
- Use the `predict()` command to predict the reading score for fourth-graders with a pre-treatment reading score of 80 who *did not receive* the treatment.
- Subtract the second number from the first. This is the treatment effect for fourth-graders.
- Do the same with the other grades.

Report all four treatment effects. Compare them to the one's you found in Question 3.

Question 5

Overall, what have we learned about the effect of the educational television program on reading scores? Summarize in a paragraph.