SOC 30005:
Assignment 2: Analysis of Covariance
Due May 9, 2019

Our aim in this assignment is to study the impact of cognitive strategies on writing on writing (CSIW). Our analysis will be restricted to the following variables

CSIW (1=CSIW, 0=control)
Achievement Level (1=High, 2=Average, 3=Low, 4=Learning Disability)
Holistic pretest (pre-test on writing achievement)
Holistic posttest (posttest on writing achievement)
Grade (1=Grade 4, 2=Grade 5)

You will need to re-code Achievement Level into 4 dummy variables and grade into 1 dummy variable. Whenever you write down a model, make sure all terms are defined and all assumptions stated. If you want to discuss a table or plot, paste it directly into the text. Do not include any appendices. Do not include any tables or figures that you do not discuss in the text.

A. Naïve model

1.Write down a model in which the outcome is the holistic posttest and the predictor is CSIW.
2. Use your favorite package (R, SPSS, STATA, SAS, ETC) to estimate this model.
3. Say briefly what you have learned.

B. Analysis of covariance (ANCOVA) model

1. Write down the ANCOVA model with holistic pre-test as a covariate.
2. Estimate this model.
3. Tell us what you have learned (very briefly!)
4. Graph the predicted values as a function of pre-test and treatment group membership (that is CSIW=1 or 0). (Hint: recall the graph we had in class for ANCOVA).

C. Assessing Linearity

1. Make a scatter plot in which the vertical axis has the residuals from the ANCOVA model and the horizontal axis is the covariate.
2. Use a Loess Line and a Quadratic fit to check linearity.
3. Explain what you see.

D. Quadratic Model

Center the holistic pre-test around the mean, call it “pretest_c”
Square the centered holistic pre-test, call it “pretest_csq”

1. Write down a model that uses CSIW, pretest_c, and pretest_csq” are predictors

2. Estimate the model and tell us how to interpret all of the coefficients in the model (including the intercept)

E. Searching for confounders

1. Check to see if you have omitted any confounders. Tell us what you found.
2. Re-estimate the model now but add any confounders
3. Tell us briefly what you learned

F. Heterogeneity

1. Does the treatment effect depend on the grade level of the child?
2. Does the treatment effect depend on the child's prior achievement level?
Provide statistical evidence.

G. Checking assumptions about the random part of the model

1. Use a plot to check the normality of the residuals
2. Use a plot to check homoscedasticity
3. Tell us what you found

H. Conclusion

1. What is your best estimate of the impact of CSIW on writing achievement (provide a confidence interval).
2. Under what assumptions is this a valid estimate of the causal effect?