ECON 4650–001: Principles of Econometrics

Marcio Santetti | Spring 2021

Midterm Exam

Instructions: This Midterm exam is worth 40 points, and it covers the contents of *Weeks 1–6*. Carefully read the following problems, and prepare your answers in an R script. That is, you will submit a *single* .R file with all your answers.

This assignment's page has a **template** for your answer script. Please consider using it.

This exam is **due 03/07** (Sunday). As per our class syllabus, late submissions **will not be accepted**. Upload your R script with your name (mine would be `marcio.R`) to this assignment's page.

Contents

Problem 1

Does a mother's *smoking* affect the *birthweight* of her child? Using data in the file `bweight.csv` (available at this assignment's page) taken from Cattaneo (2010), answer the following questions:

(a) Estimate a simple regression model for *infant birthweight*, controlling for *whether or not the mother has smoked during pregnancy*.

(b) What is the estimated *effect* of smoking during pregnancy on infant birthweight? Is this effect statistically significant at $\alpha = 0.05$? Explain your answer.

(c) Now, add the *number of prenatal care visits* to (a)'s model. Interpret its effect on the dependent variable.

(d) Now, include the *square* of the number of prenatal care visits to (c)'s model, also *log-transforming* the dependent variable. Interpret the effect of this variable once again.

(e) Lastly, store the residuals from (d)'s model and plot them against $nprenatal_i$. From a *visual* perspective, do you believe that the variance of the residual term is constant, in accordance with CLRM Assumption V? Explain. *Hint*: after storing the residuals, add them as a new column to the original data set. It will make your plotting life a lot easier.

Problem 2

The file `cocaine.csv` (available at this assignment's page) contains 56 observations on variables related to sales of cocaine powder in northeastern California over the 1984–1991 period. The data are a subset of those used in Caulkins and Padman (1993).

(a) Estimate the following model:

$$price_t = \beta_0 + \beta_1 quant_t + \beta_2 qual_t + \beta_3 trend_t + u_t$$

(b) Interpret the estimated coefficients for $quant_t$ and $qual_t$.

(c) Are all slope coefficients *jointly* significant? Explain your answer.

(d) Are $quant_t$ and $qual_t$ *jointly* significant? Run the appropriate F-test and assume $\alpha = 5\%$.

(e) Is $qual_t$ statistically significant at $\alpha = 5\%$? Explain.

Problem 3

From the `wooldridge` package, load the `htv` data set. It includes information on wages, education, parents' education, and several other variables for 1,230 working men in 1991.

(a) Estimate the following model:

$$educ_i = \beta_0 + \beta_1 motheduc_i + \beta_3 fatheduc_i + \beta_3 abil_i + \beta_4 abil_i^2 + u_i$$

(b) Interpret this model's *adjusted* R^2.

(c) Verify *CLRM Assumption II*, on the expected value of the residual term. Is this assumption satisfied?

(d) Verify *CLRM Assumption III*, on the correlation between the independent variables and the residual term. Is this assumption satisfied? *Hint*: do not forget the squared term!

(e) Verify *CLRM Assumption VII*, on the normality of the residual term. Run the appropriate Shapiro-Wilk test at $\alpha = 0.05$. Is this assumption satisfied?

Problem 4

Use the data in `sleep75` for this exercise. These data are also from the `wooldridge` package.

(a) Regress *sleep* on *totwrk*, *educ*, *age*, age^2, and *yngkid*.

(b) All else constant, how do minutes worked per week influence sleeping time?

(c) Now, evaluate whether being in the labor force *negatively* influences sleeping time. Include the *inlf* variable to (a)'s model and interpret its coefficient.

(d) Now, add an *interaction* between *age* and *male* to (c)'s model. What is the *gender differential* in sleeping time, if we compare one male and one female individual who are the same age?

(e) Is an individual's *marital status* statistically significant to determine one's weekly sleeping time? Add the appropriate variable to (d)'s model.