

Name: _____

ECON 453
In-Class Exercise 3
September 14, 2023

Please download the file “In-Class 3.gdt”, a gretl data file. This is a dataset I discussed during the lecture on Monday. The data come from a study in the late 1980’s that examined factors determining the birthweight of a baby. Birthweight is meant to measure the health of the child at birth (higher weight is generally better). Please open the data file. The dataset contains basic descriptions of each of the variables.

1. Run a regression (**Model -> Ordinary Least Squares**) using birthweight (in ounces) as the dependent variable and two regressors: *cigs* and *income*.
 - a. Report your estimated equation. Provide a numerical interpretation of the coefficients. Do these make sense to you? Are they practically significant?
 - b. Examine the residual plots for both the *cigs* and the *income* variables (from the regression results window (Model 1), choose **Graphs -> Residual Plot**). Do these cause any concerns regarding the OLS assumptions? Do these raise any questions about the data? What is going on with that one baby?
2. Run another regression where you have the same setup as question 1 ($Y = \text{bweight (in ounces)}$, *cigs* and *income* as regressors), and add the mother’s education and father’s education as additional regressors.
 - a. Summarize what the results of the model indicate about the impact of parental education on the birthweight (health) of the baby. Does this seem reasonable?
 - b. Run two more regressions. The first should be the same as above but remove the *income* variable. The second should remove both the *income* and father’s education variables. Note: this can be done from the regression results window you have (select **Edit -> Modify Model**). Summarize your findings about how the education of the parents affects the birthweight of the baby.

3. Start with the same model as in question 1 ($Y = \text{bweight (in ounces)}$, cigs and income as regressors). Now add variables to your model to examine whether the gender of the baby or the birth order (parity) are related to birth weight. Consider how each variable should be specified.
 - a. Report your estimated equation and summarize what you have learned about whether/how these variables affect birthweight.

4. So far, we have been using 'cigarettes smoked per day' as a traditional quantitative variable. Let's try some alternative models where we use a series of categorical (dummy) variables we create. It would be a good idea to start by looking at the distribution of the variable (in the main gretl window, highlight cigs , then right-click and choose **Frequency distribution**).
 - a. Create a dummy variable for whether or not the person smoked cigarettes while pregnant. To do this, choose **Add -> Define new variable**. Run a regression using birthweight (in ounces) as the y-variable and income and your new dummy variable as regressors. Report your estimated equation and interpret the coefficient on your dummy variable.

 - b. Next, create several variables to characterize different levels of smoking. To do this: choose **Add -> Define new variable**. As an example, if I wanted to create a dummy variable for 1 to 5 cigs smoked per day, I would enter into the box: $\text{cigs_1to5} = \text{cigs} \geq 1 \ \&\& \ \text{cigs} \leq 5$. This should create a new variable called "cigs_1to5" that has a value of 1 if the person smoked 1 to 5 cigs per day while pregnant, and 0 if they smoked any other number. Note: you do not need to use this same range, this is just an example.
 - i. What categories did you choose to create? Briefly explain your reasoning.

 - ii. Run a regression with birthweight (in ounces) as the y-variable and income and your cigarette categories as the regressors. Report your estimated equation.

 - iii. Compare the models in Question 1, Question 4a and Question 4b. Which is your preferred method to model the relationship between smoking and birthweight?