

# BT2101 Week 9 Homework

**Due date: March 16 2022**

A problem of interest to health officials (and others) is to determine the effects of smoking during pregnancy on infant health. One measure of infant health is birth weight; a birth weight that is too low can put an infant at risk for contracting various illnesses. In this project, you will use a dataset named “BT2101\_hw\_week9.csv” to analyze the causal effect of smoking on infant health. The data comes from “Introductory Econometrics” written by Wooldridge. The variables in the dataset indicate the following.

1. faminc: 1988 family income, \$1000s
2. bwght: birth weight, pound
3. fatheduc: father’s years of education
4. motheduc: mother’s years of education
5. male: male=1 if male child
6. white: white=1 if white
7. cigs: cigs smoked per day while pregnant

## **I. Exploring the nature of your data**

Before diving into the regressions, you want to first analyze the data in the following ways.

1. Get some summary statistics about the two variables “bwght” and “cigs” including their means and standard deviations. Does the variable “bwght” look like a normal random variable? Hint: a normal random variable is symmetric (skewness=0), and its kurtosis is equal to 3.
2. Get the correlation between the variables “bwght” and “cigs”. Is there a negative relationship between the two variables?
3. Divide the data into two groups. In the first group, mothers don’t smoke during pregnancy, and in the second group, mothers smoke during pregnancy. Conduct a difference-in-mean test on these

two groups. Investigate whether the birth weight of infants from the first group is statistically different from that of the second group.

## **II. A simple regression model**

Now you want to run a simple regression to investigate how smoking during pregnancy affects infant health by regressing “bwght” on “cigs”.

4. In this simple regression, what is the estimated effect of smoking on infant health? Is this effect large or small? Is it statistically significant the 5% level?

5. You further investigate how much variation in “bwght” can be explained by “cigs”. The R-squared turns out to be very small. Comment on this statement: since the R-squared from this simple regression is very small, smoking during pregnancy seems to be irrelevant to infant health; in other words, health officials (and others) shouldn’t devote any attention to this issue at all.

6. You suspect that the simple regression suffers an omitted variable bias problem. Based on the data, what are potential omitted variables? Why do they qualify as omitted variables?

## **III. Multiple regression model**

7. Run a multiple regression of “bwght” on “cigs”, “faminc”, “fatheduc”, “motheduc”, “male”, and “white”. What is the new estimated effect of smoking on infant health? What do the coefficients of “male” and “white” tell you?

8. Since “faminc”, “fatheduc”, “motheduc” are not individually significant at the 5% level, you decide to conduct a test with the null that these three variables jointly have no effect on infant health. What kind of test should you use? Can you reject the null at the 5% significant level?

9. Although “famine” is not significant, you still decide to include it in the regression to avoid the omitted variable bias. Higher income families tend to smoke higher quality cigarettes, which tend to have a less severe impact on the health of an infant, your further conjecture that the effects of smoking on infant health should be decreasing with the family income. Run a regression to incorporate this new conjecture, including all variables in question 8. Do you see such a decreasing effect in the regression? Is it statistically significant at the 5% level?

10. You suspect that the effect of smoking on birth weight is stronger for infant boys ( $\text{male}=1$ ) than for infant girls ( $\text{male}=0$ ). Run a multiple regression to incorporate this effect, including all variables in question 8. Does the regression result suggest such an asymmetric effect? Is it statistically significant at the 5% level?