Name _____ **Surname** _____ **Student ID.** _____
DUE DATE : Thursday 27, April 2017. Before 3.00 pm. at B.E. office.

Assignment 5: (130 points)

I pledge to the Honor Code and to obey all rules for taking and performing homework assignments as specified by the course instructor.

Student Signature: _____

1. (30 points) Consider the following regression equation:

$$\begin{array}{rcccccc} \widehat{\text{sleep}}_i & = & 3,840.83 - 0.163\text{totwrk}_i - 11.71\text{educ}_i - 8.70\text{age}_i + 0.128\text{age}_i^2 + 87.75\text{male}_i \\ & & (235.11) \quad (0.018) \quad (5.86) \quad (11.21) \quad (0.134) \quad (34.33) \end{array} \tag{Eq.1}$$

$$n = 706, \quad R^2 = 0.123, \quad \bar{R}^2 = 0.117$$

The variable sleep is total minutes per week spent sleeping at night, totwrk is total weekly minutes spent working, educ and age are measured in years, and male is a gender dummy. (1.1) All other factors being equal, is there evidence that men sleep more than women? How strong is the evidence?

(1.2) Is there a statistically significant tradeoff between working and sleeping? What is the estimated tradeoff?

(1.3) What other regression do you need to run to test the null hypothesis that, holding other factors fixed, age has no effect on sleeping?

2. (40 points) Consider the following regression models to explain child birth weight in terms of various factors:

Model 1

$$\begin{array}{rcl} \widehat{\ln(\text{bwght})}_i & = & 4.66 - 0.0044\text{cigs}_i + 0.0093\ln(\text{faminc})_i + 0.016\text{parity}_i + 0.027\text{male}_i + 0.055\text{white}_i \\ & & (0.22) \quad (0.0009) \quad (0.0059) \quad (0.006) \quad (0.010) \quad (0.013) \end{array} \tag{Eq.2}$$

$$n = 1388, \quad R^2 = 0.0472$$

Model 2

$$\begin{array}{rcl} \widehat{\ln(\text{bwght})}_i & = & 4.65 - 0.0052\text{cigs}_i + 0.0110\ln(\text{faminc})_i + 0.017\text{parity}_i + 0.034\text{male}_i + 0.045\text{white}_i \\ & & (0.38) \quad (0.0010) \quad (0.0085) \quad (0.006) \quad (0.011) \quad (0.015) \\ & & -0.0030\text{motheduc}_i + 0.0032\text{fatheduc}_i \\ & & (0.0030) \quad (0.0026) \end{array} \tag{Eq.3}$$

$$n = 1191, \quad R^2 = 0.0493$$

The variables are defined as:

bwght = birth weight, in pounds.
cigs = average number of cigarettes the mother smoked per day during pregnancy.
parity = the birth order of this child.
faminc = annual family income.
motheduc = years of schooling for the mother.
fatheduc = years of schooling for the father.
We have added a dummy variable for whether the child is male (1 if male and 0 if female) and a dummy variable indicating whether the child is classified as white (1 if white and 0 if non-white).

(2.1) In the first equation, interpret the coefficient on the variable cigs. In particular, what is the effect on birth weight from smoking 10 more cigarettes per day?

(2.2) How much more is a white child predicted to weight than a nonwhite child, holding the other factors in the first equation fixed? Is the difference statistically significant?

(2.3) Comment on the estimated effect and statistical significance of motheduc.

(2.4) From the given information, why are you unable to compute the F statistic for joint significance of motheduc and fatheduc? What would you have to do to compute the F statistic?

Empirical work: STATA program

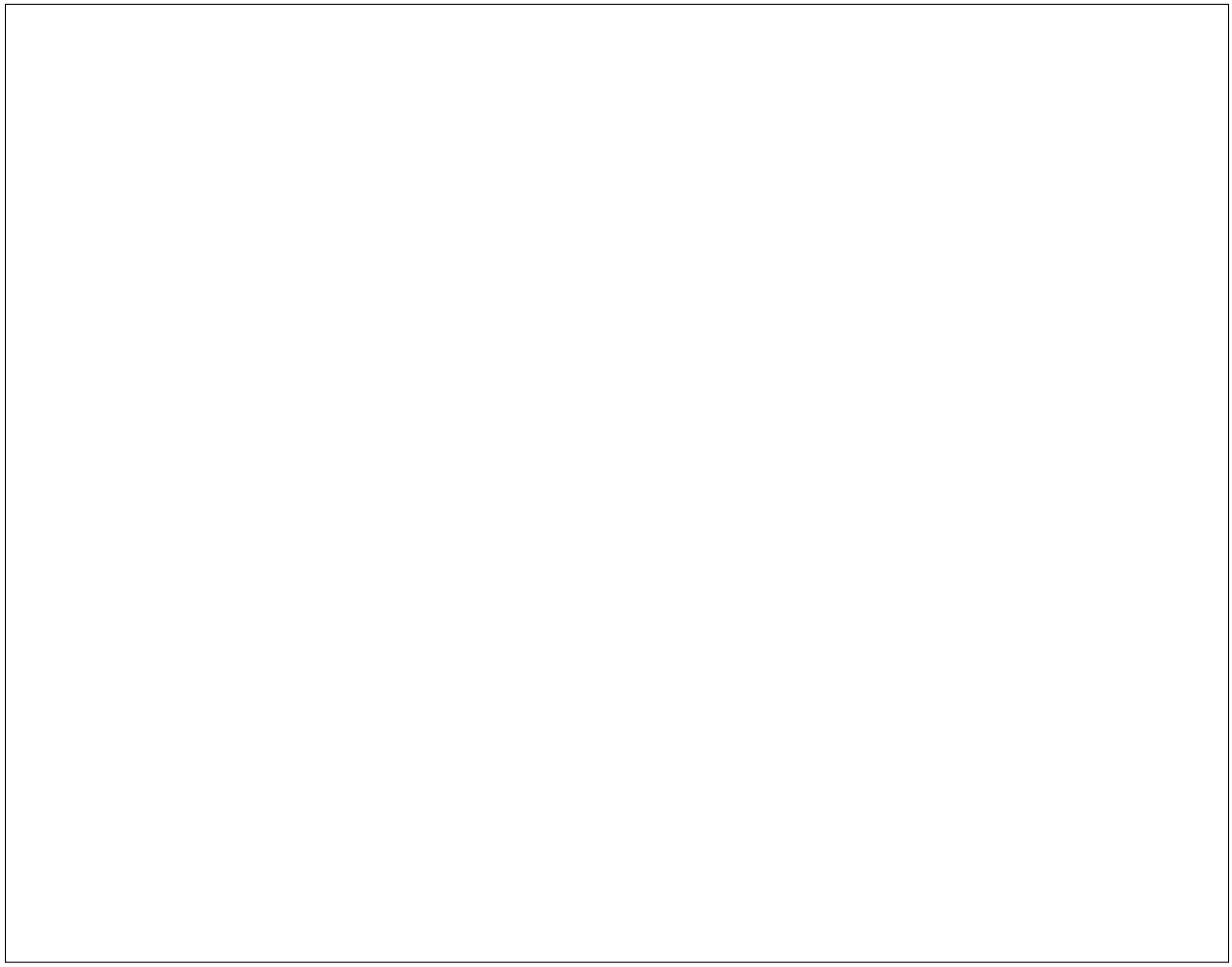
3. (60 points) Use the data in DISCRIM.DTA to answer this question. These are ZIP code?level data on prices for various items at fast-food restaurants, along with characteristics of the zip code population, in New Jersey and Pennsylvania. The idea is to see whether fast-food restaurants charge higher prices in areas with a larger concentration of blacks.

(3.1) Find the average values of prpblck and income in the sample, along with their standard deviations. What are the units of measurement of prpblck and income?

(3.2) Consider a model to explain the price of soda, psoda, in terms of the proportion of the population that is black and median income:

$$\text{psoda}_i = \beta_0 + \beta_1 \text{prpblck}_i + \beta_2 \text{income}_i + u_i \quad \text{(Eq.4)}$$

Estimate this model by OLS and report the results in equation form, including the sample size and R-squared. Interpret the coefficient on prpblck. Do you think it is economically large?



(3.3) Compare the estimate from part (3.2) with the simple regression estimate from psoda on prpblck. Is the discrimination effect larger or smaller when you control for income?

(3.4) A model with a constant price elasticity with respect to income may be more appropriate. Report estimates of the model

$$\log(\text{psoda})_i = \beta_0 + \beta_1 \text{prpblck}_i + \beta_2 \log(\text{income})_i + u_i \quad \text{(Eq.5)}$$

If prpblck increases by .20 (20 percentage points), what is the estimated percentage change in psoda? (Hint: The answer is 2.xx, where you fill in the "xx.")

(3.5) Now add the variable prppov to the regression in part (3.5). What happens to β_{prpblck}?

(3.6) Find the correlation between log(income) and prppov. Is it roughly what you expected?