There are 5 questions. If you get stuck on one part, move on and do the rest. GOOD LUCK!

1. Using cross section data on individuals in a certain year, the following equation is estimated

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u$$

where the variables are
wage: average hourly earnings
educ: years of education
exper: years of experience.

The regression results are:

```
. reg wage educ exper expersq

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(3, 522)       =     64.11
       Model |  1927.87673         3  642.625576   Prob > F        =    0.0000
    Residual |  5232.53756       522  10.0240183   R-squared       =    0.2692
-------------+----------------------------------   Adj R-squared   =    0.2650
       Total |  7160.41429       525  13.6388844   Root MSE        =    3.1661

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .5953429   .0530251    11.23   0.000     .4911741    .6995118
       exper |    .268287   .0368969     7.27   0.000     .1958023    .3407717
     expersq |  -.0046123    .000822    -5.61   0.000     -.006227   -.0029975
       _cons |   -3.96489   .7521526    -5.27   0.000    -5.442508   -2.487272
------------------------------------------------------------------------------
```

a. Using the formula for the effect of experience on wage in this setup, what is the return to the fifth year of experience, i.e., when *exper* increases from 4 to 5? Provide the numerical value.

Answer: The return to experience is given by $\beta_2 + 2\beta_3 exper$. For the fifth year of experience, this is 0.268 + 2(-0.0046)4, i.e., 0.231.

b. At what value of *exper* does additional experience actually begin to lower predicted wage (i.e., the turning point)? Provide the numerical value.

Answer: The turning point is given by |0.268/2(-0.005)|, i.e., 26.8 years.

2. Consider a model where the return to education depends upon the amount of work experience (and vice versa):

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ \cdot exper + u$$

where the variables and data are as discussed in question 4.

The summary statistics are

```
. su wage educ exper
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	526	5.896103	3.693086	.53	24.98
educ	526	12.56274	2.769022	0	18
exper	526	17.01711	13.57216	1	51

The regression results are

```
. g educexper = educ*exper

. reg wage educ exper educexper
```

Source	SS	df	MS
Model	1615.96222	3	538.654074
Residual	5544.45207	522	10.6215557
Total	7160.41429	525	13.6388844

Number of obs =	526
F(3, 522) =	50.71
Prob > F =	0.0000
R-squared =	0.2257
Adj R-squared =	0.2212
Root MSE =	3.2591

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.6017355	.0899	6.69	0.000	.4251253	.7783457
exper	.0457689	.0426138	1.07	0.283	-.0379466	.1294844
educexper	.0020623	.0034906	0.59	0.555	-.004795	.0089197
_cons	-2.859916	1.18108	-2.42	0.016	-5.180169	-.5396623

a. Using the formula for the effect of experience on wage in this setup, what is the approximate return to experience for the average level of education? Provide the numerical value.

Answer: The return to experience is given by $\beta_2 + \beta_3 educ$. For the average level of education, this is 0.046 + 0.002 x 12.56, i.e., 0.071.

b. Using the formula for the effect of education on wage in this setup, what is the approximate return to education for the average level of experience? Provide the numerical value.

Answer: The return to education is given by $\beta_1 + \beta_3 exper$. For the average level of experience, this is 0.602 + 0.002 x 17.02, i.e., 0.636.

3. This question is inspired by a study published (by Andrew Rose) in Journal of International Economics in July 2004. The study is entitled “Do WTO Members have More Liberal Trade Policy?” Suppose using cross sectional data across countries in a certain year, the following equation is estimated:

$$open_i = \beta_0 + \beta_1 WTO_i + \beta_2 OECD_i + \beta_3 WTO_i \cdot OECD_i + \beta_4 population_i + u_i.$$

Here i denotes country, u denotes the unobserved factors, and the variables are
open: measure of openness, i.e., (exports + imports)/gross domestic product (GDP)
WTO: binary indicator defined as 1 for World Trade Organization (WTO) members (0 otherwise)
OECD: binary indicator defined as 1 for Organization for Economic Cooperation and Development (OECD) members (0 otherwise)
population: population size.
Note that WTO members do not necessarily belong to the OECD and vice versa.

a. Controlling for population size, which coefficient (or combination of coefficients) captures the difference in openness between non-OECD WTO members and non-OECD non-WTO members? Here, non-OECD WTO members means countries that are WTO members but not OECD members. Also, non-OECD non-WTO members means countries that are neither WTO nor OECD members.

Answer: β_1.

b. Controlling for population size, which coefficient (or combination of coefficients) captures the difference in openness between non-WTO OECD members and non-OECD non-WTO members? Here, non-WTO OECD members means countries that are OECD members but not WTO members. Also, non-OECD non-WTO members means countries that are neither WTO nor OECD members.

Answer: β_2.

c. Controlling for population size, which coefficient (or combination of coefficients) captures the difference in openness between countries belonging to both the WTO and OECD, and non-OECD non-WTO members? Here, non-OECD non-WTO members means countries that are neither WTO nor OECD members.

Answer: $\beta_1 + \beta_2 + \beta_3$.

d. In the above equation, can we add a non-WTO membership dummy defined as 1-WTO? Why or why not?

Answer: No, due to perfect collinearity or the dummy variable trap.

4. This question is inspired by a recent study published (by Kitae Sohn) in Economics & Human Biology in November 2017. The study is entitled "The Association between Height and Hypertension in Indonesia." Suppose, using cross sectional data on individuals in a certain year, the following linear probability model (LPM) is estimated by ordinary least squares (OLS):

$hypertension_i = \beta_0 + \beta_1 height_i + u_i$.

Here i denotes individual, u denotes the unobserved factors, and the variables are
hypertension: binary indicator defined as 1 for people with high blood pressure (0 otherwise)
height: height in cm.

a. Are the errors in a LPM homoskedastic?

Answer: No, the errors are necessarily heteroskedastic.

b. If β_1 is estimated as -0.461, how would we interpret this coefficient estimate corresponding to height?

Answer: An increase in height by 1 cm is associated with a decrease in the probability of hypertension by 0.46.

c. Specify one advantage of estimating the above relationship between hypertension and height using a logit or probit model instead of OLS.

Answer: One advantage is that the predicted probabilities are always between 0 and 1.

d. For a logit or probit model, the marginal effect of height on the probability of hypertension is not constant across all observations. Specify two approaches of calculating the marginal effect for such models?

Answer: One approach would be to calculate the effect at the average value of height. Alternatively, we can calculate the observation-specific marginal effects and then take the average.

5. This question is inspired by a recent study published (by Evelina Gavrilova, Takuma Kamada, and Floris Zoutman) in Economic Journal in January 2019. The study is entitled "Is Legal Pot Crippling Mexican Drug Trafficking Organizations? The Effect of Medical Marijuana Laws on US Crime." The state-level data from the U.S. is a panel with data across all states from two years, i.e., 1995 and 2010. The following equation is estimated:

$crime_{it} = \beta_0 + \delta_0 d10_t + \beta_1 MM_{it} + \beta_2 \log(income_{it}) + u_{it}$.

Here i denotes state, t denotes year, u denotes the unobserved factors, and the variables are
crime: crime rate per 100,000 inhabitants
d10: binary indicator defined as 1 for 2010 (0 otherwise)
MM: binary indicator defined as 1 for states that decriminalize production and consumption of (medical) marijuana (0 otherwise); note that no state decriminalized in 1995 but some did by 2010
income: median income.

a. Suppose the above regression is estimated by OLS, and the corresponding estimate of β_1 is 174.44. How would you interpret this coefficient estimate corresponding to MM?

Answer: Decriminalizing production and consumption of (medical) marijuana in a state is associated with an increase in crime rate of about 174.44 per 100,000 inhabitants.

b. Next, suppose $u_{it} = a_i + v_{it}$, such that MM_{it} and a_i are correlated. In other words, a_i is a time-invariant state characteristic that affects crime and is also correlated with MM. For example, think of inland states versus those bordering Mexico, or presence of local gangs. In such a scenario, what is a drawback of an OLS approach?

Answer: In this case, our OLS estimator is biased due to the correlation between MM_{it} and a_i.

c. Next, suppose $u_{it} = a_i + v_{it}$, such that MM_{it} and a_i are correlated. In other words, a_i is a time-invariant state characteristic that affects crime and is also correlated with MM. For example, think of inland states versus those bordering Mexico, or presence of local gangs. In such a scenario, given the drawback of an OLS approach, what estimation strategy would you pursue to control such for time-invariant state characteristics?

Answer: First-differencing.

d. Next, suppose $u_{it} = a_i + v_{it}$, such that MM_{it} and a_i are correlated. In other words, a_i is a time-invariant state characteristic that affects crime and is also correlated with MM. For example, think of inland states versus those bordering Mexico, or presence of local gangs. In such a scenario, say we use an estimation strategy to control such for time-invariant state characteristics. However, suppose v_{it} includes some measure of drug violence such that v_{it} is uncorrelated with MM_{it}, but v_{it-1} (i.e., past drug violence) is correlated with MM_{it}. In this case, would your strategy to control such for the time-invariant state characteristics be valid? Why or why not?

Answer: No. First-differencing relies on the assumption of strict exogeneity which is violated in this case.