*Type up your answers in a Word document and then save it as PDF before uploading them to Canvas.*

1. (2 pts) I regressed wealth (dependent variable) on parents' wealth (independent variable) and found an $R^2 = 0.64$. What does this mean? Does a high $R^2$ imply that I am identifying the causal impact of parents' wealth on children's wealth?

2. (1 pt) Consider the following linear regression model:

$$Y = \beta_0 + \beta_1 X + u$$

   Show that $E(u|X) = 0$ implies $E(Y|X) = \beta_0 + \beta_1 X$.

3. (3 pts) A study was conducted to investigate the effects of short naps on memory. The study involved 200 participants who were allowed to nap for either 30 or 60 minutes. After waking up, each participant took a short test on short-term recall. The nap duration for each participant was randomly assigned based on a coin flip. Let $Y_i$ denote the score of the $i$'th participant on the test ($0 \leq Y_i \leq 100$), and let $X_i$ denote the length of time the participant slept before the test ($X_i = 30$ or $60$). Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

   (a) Explain what the term $u_i$ represents. Why might different participants have different values of $u_i$?

   (b) Do you think here $E(u_i|X_i) = 0$? Why or why not? Will the estimated coefficients be biased or unbiased?

   (c) The estimated regression is $\hat{Y}_i = 55 + 0.17X_i$. Compute the predicted score for a participant who napped for 30 minutes and for a participant who napped for 60 minutes. Compute the estimated gain/loss in score from a longer nap.

4. (3 pts) Output from three regression models is presented below.

```
=========================================================
                        Dependent variable:
                -----------------------------------------
                              incwage
                  (1)            (2)            (3)
---------------------------------------------------------
uhrswork                    2,276.766***   2,273.603***
                             (129.614)      (129.580)

female      -15,024.470*** -10,732.690*** -10,578.670***
             (1,551.198)    (1,522.257)    (1,523.637)

trantime                                     60.201**
                                            (29.645)

Constant     79,290.520*** -21,898.510*** -23,541.690***
             (1,015.570)    (5,843.996)    (5,897.813)

---------------------------------------------------------
Observations    4,673          4,673          4,673
R2              0.020          0.080          0.081
Adjusted R2     0.019          0.080          0.081
=========================================================
Note:                        *p<0.1; **p<0.05; ***p<0.01
```

The three models are as follows:

Model 1:   $incwage = \beta_0 + \beta_1 female + u$

Model 2:   $incwage = \beta_0 + \beta_1 female + \beta_2 uhrswork + u$

Model 3:   $incwage = \beta_0 + \beta_1 female + \beta_2 uhrswork + \beta_3 trantime + u$

The variables are defined as follows:

- *incwage*: annual wage and salary income (in \$)

- *female*: indicator variable that takes value 1 for females and 0 for males

- *uhrswork*: hours worked each week

- *trantime*: transit time to work (in minutes)

Answer the following questions (0.5 pts each):

(a) Interpret the coefficient on *female* in Model 1. Is this coefficient statistically significant at a 1% level of significance?

(b) Interpret the coefficient on *female* from Model 2.

(c) In the second model, we added the variable for hours worked and the coefficient on *female* went down considerably. Given this pattern, do you think women work more or fewer hours than men in our sample? Explain your reasoning. (Note: The coefficient on *uhrswork* is positive.)

(d) Interpret the coefficient on *trantime* in model three. If I wanted to test the null hypothesis that $\beta_3 = 0$, what would be the associated $t$-value for this test?

(e) Find the $p$-value for the $t$-statistic you found in (d).[1] What does this $p$-value tell us?

(f) Does the inclusion of *trantime* improve the explanatory power of the regression model for predicting wages? Why or why not?

Pay attention to the units of measurement of each variable while interpreting the coefficients.

5. (1 pt) Consider the following regression model:

$$wages_i = \beta_0 + \beta_1 experience_i + \beta_2 immigrant_i + \beta_3 experience_i \times immigrant_i + u$$

where $immigrant_i$ is a binary variable that takes the value 1 if an individual is an immigrant and 0 otherwise. $wages_i$ captures earnings and $experience_i$ captures work experience.

What is the interpretation of $\beta_3$? (Hint: Take the expectation of the equation conditional on $immigrant_i = 1$ and $immigrant_i = 0$.)

---

[1]You need to refer to the standard normal table or use `2 * (1 - pnorm(abs(t)))` in R.