ECONOMETRICS I: ECO 221
Assignment 1
Deadline: February 25, 2022
Instructor: Gaurav Arora
(150 points)

Q1.

Specify the following simple regression models and provide an interpretation of in each case

1. level-level

2. log-level

3. level-log

4. log-log

Q2.

Show for a simple linear regression model $y_i = \beta_0 + \beta_1 x_i + u_i$, where $i = 1, 2, ..., n$ that the ordinary least squares estimator is given as

$$\hat{\beta}_1 = \frac{\sum_{i=1,...,n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1,...,n}(x_i - \bar{x})^2} .$$

Suppose that in the above simple regression model *y* represent test-scores for each student *i* in the mathematics course and *x* represents the number of hours invested by student *i* to study math during the semester. Show how the above $\hat{\beta}_0$ and $\hat{\beta}_1$ will change if an alternative regressor is included in the regression as

(a) Minutes invested in studying math during the semester

(b) Seconds invested in studying math during the semester.

(c) Logarithm of hours invested in studying math during the semester.

Regression analysis:

Q3. In a study of wage differences between native and non-native workers of similar age and similar training the following equation is estimated

$$W_i = \alpha + \beta D_i + u_i \qquad (1)$$

where W_i is wage of worker *i* and D_i is a dummy variable that takes value 1 only if the worker is non-native or zero otherwise, and u_i is the stochastic error term. Let $\overline{W}_{nat}$ and $\overline{W}_{non}$, and n_{nat} and n_{non} the average wage and number of natives and non-natives in the sample. Also, let $\bar{W}$ and $\bar{D}$ be average of W_i and D_i.

(a) Show that the following relationships are true.

$$(i) \qquad \bar{W} = \frac{n_{non}\bar{W}_{non} + n_{nat}\bar{W}_{nat}}{n}$$

$$(ii) \qquad \bar{D} = \frac{n_{non}}{n_{non} + n_{nat}}$$

$$(iii) \qquad \sum_{i=1}^{n} \left(D_i - \bar{D}\right)^2 = \frac{n_{non} \cdot n_{nat}}{n_{non} + n_{nat}}$$

(b) Evaluate the OLS estimates of α and β for equation (1). Show $\hat{\alpha} = \overline{W}_{nat}$ and interpret.

Q4. Consider two random variables *A* and *B*. *A* is the response variable that is assumed to be related to the predictor *B* through a function *f* such that $f(B)$ approximates *A*. In the regression form we specify this relationship as follows

$$A = f(B) + u = \gamma_1 B + u, \qquad (1)$$

where we assume that $Eu = 0$ and γ_1 is the regression parameter (a constant). To understand how closely *A* and *B* might be related we utilize two performance metrics

a) Coefficient of correlation, defined as (recall from MTH201)

$$\rho_{A,\hat{A}} = \frac{\sum_{i=1}^{n}(A_i - \bar{A})(\hat{A}_i - \bar{A})}{\sqrt{\sum_{i=1}^{n}(A_i - \bar{A})^2 \sum_{i=1}^{n}(\hat{A}_i - \bar{A})^2}}, \text{ and}$$

b) Goodness-of-fit parameter, defined as (refer class notes)

$$R^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST} = \frac{\sum_{i=1}^{n}\left(\hat{A}_i - \bar{A}\right)^2}{\sum_{i=1}^{n}\left(A_i - \bar{A}\right)^2}$$

Show that $R^2 = \left(\rho_{A,\hat{A}}\right)^2$.

Q5. Consider a variable X that records the duration of a phone call in minutes. Suppose X is a random variable with probability density function $f(x) = c\exp\left(-\frac{x}{10}\right)$, where c is a constant and $x \geq 0$.

- Find c.
- What is the probability that a call lasts exactly seven minutes.

Q6. I roll a fair die repeatedly until a number larger than 4 is observed. Let N be the total number of times until a number larger than 4 is observed. Find $Pr\left(N = k\right)$ for $k = 1, 2, 3$.

Probability & Statistics Revision (Ungraded)

1. You are given the following information regarding the joint distribution of X (the age of a person) and Y (the number of days they choose to spend at Saylorville Lake).

		Values of Y			
		0	1	2	3
Values of X	20	0.25	0.04	0.01	0.00
	40	0.15	0.12	0.08	0.05
	60	0.25	0.04	0.01	0.00

   a. What are the marginal distributions of X and Y?

   b. Compute $E(X)$ and $E(Y)$.

   c. Compute σ_X^2 and σ_Y^2.

   d. Compute σ_{XY} and $Corr(X, Y)$.

   e. Are X and Y independent?

   f. What are the conditional means $E(Y|X = 20)$, $E(Y|X = 40)$, and $E(Y|X = 60)$?

   g. A randomly selected person reports that they have spent 2 days at Saylorville Lake. What is the probability that they are 40?

   h. Finally, suppose that time spent at Saylorville Lake costs \$100 plus \$25 per day. That is, if Z denotes the total travel expenditure of an individual, then $Z = 100 + 25 \times Y$. What is the mean expenditure of individuals visiting Saylorville Lake and the standard deviation of these expenditures?

2. Compute the following probabilities:

   a. If $Y \sim N(2, 25)$, then what is $Pr(Y > 4)$?

   b. If $Y \sim N(7, 49)$, then what is $Pr(Y < 0)$?

   c. If $Y \sim N(5, 4)$, then what is $Pr(3 < Y \leq 7)$?

   d. If $Y \sim N(5, 16)$, then what is $Pr(3 < Y \leq 11)$?

3. Compute the following probabilities:

   a. If $Y \sim \chi^2_{11}$, then what is $Pr(Y > 19.68)$?

   b. If $Y \sim \chi^2_3$, then what is $Pr(Y > 11.34)$?

   c. If $Y \sim F_{4,20}$, then what is $Pr(Y > 2.25)$?

   d. If $Y \sim F_{3,7}$, then what is $Pr(Y > 8.45)$?