ECON 3161 Midterm Exam

Instruction:

- There are 100 points in total. You can either type or handwrite your answer.
- Please clearly number your answers. To receive full credits, your answers should be legible and include all the necessary steps unless otherwise specified. Incomplete or non-legible answers will only receive partial or no credits.

Good Luck!

---

1. (22 points) Let $Y_1, Y_2, \cdots Y_n$ be n independent, identically distributed random variables from a population with mean μ and variance σ^2, and $n \geq 2$. Let $\widehat{\mu} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ denote the average of these n random variables. Answer the following questions:

(a) What are the expected value and variance of $\widehat{\mu}$? (8 points)

(b) Now consider a different estimator of μ: $\widetilde{\mu} = Y_1$. Is $\widetilde{\mu}$ is an unbiased estimator of μ? Why? (6 points)

(c) For the two estimators $\widehat{\mu}$ and $\widetilde{\mu}$, which one do you prefer? Why? (8 points)

2. (35 points) Suppose a researcher is interested in studying the gender wage gap. She considered the following model:

$$wage_i = \beta_1 + \beta_2 female_i + u_i, \tag{1}$$

where $female$ is a dummy variable equal to 1 if the person is female, and 0 if the person is male. Answer the following questions:

(a) Why do we need the condition $E(u_i) = 0$ to identify the model (1)? (6 points)

(b) If you have a dataset $\{female_i, wage_i\}_{i=1}^{100}$, and $female_1 = female_2 = \cdots = female_{100}$ (i.e., all the observations are from female). Are you able to estimate model (1) with this dataset? Give the reason for your answer. (6 points)

(c) For β_2 to have a causal interpretation, what condition or assumption do we need? How do you interpret this assumption intuitively? (You need to explain what does this assumption mean without using any math) (6 points)

(d) Do you think the assumption in (c) is likely to hold in practice? (3 points)

(e) Based on your assumption in (c), calculate the two quantities $E(wage|female = 1)$ and $E(wage|female = 0)$. Express β_2 in terms of these two quantities. (8 points)

(f) What do the parameters β_1 and β_2 represent? (Or how would you interpret them?) (6 points)

3. (13 points) Suppose a researcher studied the effect of woman's income on her child's birth weight (which is an important health indicator for newborns). She used the following model:

$$log(bwght_i) = \beta_1 + \beta_2 income_i + \epsilon_i, \tag{2}$$

where *bwght* denotes the infant birth weight in ounces, and the explanatory variable *income* denotes the mother's monthly income (measured in thousand dollars) during pregnancy.

Using data on $n = 2,000$ births, she estimated model (2) by OLS and got the following results:

$$\widehat{log(bwght_i)} = 2.5 + 1.2 income_i. \tag{3}$$

Answer the following questions:

(a) What is the predicted birth weight when $income = 3,000$? (You don't have to calculate the final number, just plugging the number is also okay). (3 points)

(b) Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's income? Provide your thought. (5 points)

(c) Suppose your model is able to capture a causal relationship between the child's birth weight and the mother's income, how would you interpret the estimated coefficient of *income* in model (3)? (5 points)

4. (15 points) Suppose that at Georgia Tech, college grade point average, GPA, and ACT score, ACT are related by the conditional expectation $E(GPA|ACT) = 2.5 + 0.01ACT + 0.0005ACT^2$. Answer the following questions (**Note: For questions (a)–(c) below, You don't have to calculate the final number. You will receive full credits as long as you plug in the right numbers.)**:

(a) Find the expected GPA when $ACT = 30$. (5 points)

(b) If a student's ACT score is 30, does this mean he or she will have the GPA found in part (a)? Why? (5 points)

(c) If the distribution of Georgia Tech students' ACT score is $N(32, 1)$, what is the average GPA? (5 points)

5. (15 points) Consider the household monthly food consumption and income (measured in thousand dollars) function

$$consumption = \beta_0 + \beta_1 income + u,$$

where $u = \epsilon\sqrt{0.01income^2 - 5income + 100}$, and ϵ is a random variable with $E(\epsilon) = 0$ and $var(\epsilon) = 1$. Assume that ϵ is independent of *income*.

(a) Calculate $var(u|income)$. (5 points)

(b) Is the conditional variance in (b) homoskedastic? (3 points)

(c) How does the variance of consumption change with family income? At what level of income will the variance be smallest? (7 points)