Mock Exam (Econometrics I)

- Time: **60 minutes**.
- All answers should be given in English.
- The theoretical results given on the slides can be used without proof unless providing proof is required in a problem.
- You may add assumptions to solve the problems if necessary. Full marks will be awarded as long as the discussion is reasonable.
- For problems requiring calculations, you should provide not only answers but also processes to obtain the answers.

Problem 1

We throw a fair dice n times, and X_i is a random variable that corresponds to a number generated by the dice $(i = 1, \ldots, n)$.

1. Answer the following questions.
   (a) Calculate $\mathbb{V}[X_1]$
   (b) Draw the CDF of $X_1 + X_2$.
   (c) Calculate $\mathrm{Cor}(X_1 + X_2, X_2 + X_3)$.
2. If we throw the dice many times, we can expect that the sample mean $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$ becomes closer to 3.5. Answer the following questions.
   (a) We can expect that if n is large, $\overline{X}_n$ is very likely between 3.49 and 3.51. Explain how this expectation is justified by the law of large numbers (LLN).
   (b) By using the central limit theorem (CLT), we can approximate the probability of $\overline{X}_n$ being inside of $(3.49, 3.51)$ when, for example, $n = 10^4$. Explain how we can calculate this probability. (You only need to describe the process and do not need to calculate the exact probability.)

Problem 2

We have a data set from a survey conducted in 1976 ($n = 526$), and the data record the information on the people's wages. In our data analysis, we mainly use the following variables:

- *wage*: Average hourly earnings (dollars)
- *educ*: Years of education

We are interested in how education affects wages. We first use two models:

Model 1: $wage_i = \beta_0 + \beta_1\ educ_i + \epsilon_i$

Model 2: $\log(wage_i) = \beta_0 + \beta_1\ educ_i + \epsilon_i$

When we use one of the models, we always assume that the error term ϵ_i has zero mean and its variance is σ^2 (unknown).

1. We first obtain the OLS estimators using the two models. Answer the following problems.
   (a) Provide an interpretation of β_1 in the two models, clarifying the difference.
   (b) If *educ* was recorded in months instead of years, how would it affect the OLS estimators?
   (c) For Model 1, suppose the realized value of the OLS estimator $\hat{\beta}_{n1}$ is 0.541, and the standard error $\hat{se}_\beta$ is 0.0533. Using the CLT, construct a 95% confidence interval for β_1.
2. In Models 1 and 2, *educ* can be endogenous and introduce a bias to the OLS estimators.
   (a) Explain the problem with the models in terms of the omitted variable problem.
   (b) Suppose we have an additional variable *college*, which is a dummy variable taking 1 for those who grew up near a college and 0 for the others. Explain why *college* is expected to work as a valid IV.
   (c) Using *college*, describe how we can obtain an IV estimator for Model 1.
3. Next, we add another variable IQ, which corresponds of IQ of each person:

   Model 3: $\log(wage_i) = \beta_0 + \beta_1\ educ_i + \beta_2\ IQ_i + \epsilon_i$

   (a) We can obtain the OLS estimators of $\beta_0, \beta_1, \beta_2$ by minimizing the sum of the squared loss as in the simple linear regression. Provide the first-order conditions that the estimators should satisfy. (You do not need to obtain the solutions.)
   (b) Suppose there was a human error in the process of calculating the people's IQ, and as a result, the recorded IQ tends to be lower than the actual values. If the gap is not constant and independent of the actual IQ, how would it affect the OLS estimation of β_2?
4. Explain what the type I error and the type II error in the hypothesis test would mean in the case of β_1 in Model 3.

Note: Problem 2 is based on data analyses discussed in the textbook

- Jeffrey M. Wooldridge (2020) Introductory Econometrics: A Modern Approach, 7th ed. Cengage Learning