Econ 322 Midterm, Spring 2025

Name: _____

Instructions: This is a closed-book exam. However, you are allowed to bring two sheets of notes on standard paper, with writing on both sides. A calculator is also permitted. Write the solutions to the problems directly on the exam. Be sure to show your work as partial credits will be given. If you run out of room for an answer, continue on the back of the page. Round all your answers to three decimal places. You have 80 minutes.

1. (3 points) In the equation $C = \beta_0 + \beta_1 I + u$, where C denotes consumption and I denotes income. What is the fitted value of the 5th observation if $C_5 = \$100$ and $\hat{u}_5 = \$\text{-}20$?

   A. \$100

   B. \$120

   C. \$-20

   D. \$80

2. (3 points) What is the interpretation of the intercept β_0 in a regression model?

   A. The expected value of Y given $X = 0$.

   B. The change in Y for one-unit increase in X.

   C. The correlation between X and Y.

   D. The standard deviation of the error term.

3. (3 points) What does the unbiasedness of OLS estimators tell us?

   A. The error term is normally distributed.

   B. The expected value of the OLS estimators equals the true population parameters.

   C. The sample size is large.

   D. The OLS estimates are equal to the population parameters.

4. (3 points) Consider a simple linear regression model, $Y = \beta_0 + \beta_1 X + u$. What does the homoskedasticity assumption imply?

   A. The expected value of the error term, u, is zero, regardless of what the value of the explanatory variable, X, is.

   B. The estimated average value of β_1 is zero.

C. The variance of the error term, u, is a constant, regardless of what the value of the explanatory variable, X, is.

D. The variance of the explained variable, Y, is zero, regardless of what the value of the explanatory variable, X, is.

5. Suppose we wish to know the mean attendance at a swimming pool on a Sunday. We are interested in this number because we need to know how many lifeguards to have on duty for this coming Sunday. We observe attendances at our swimming pool from a randomly selected set of $n = 50$ Sundays. The sample average of those attendances is $\bar{y} = 221.7$ and the sample variance is $s^2 = 6948.4$. Consider the null hypothesis that the population mean, β_0, of attendance for a Sunday is no less than 200 people. Assume you do not know the population distribution of attendance. You will have to rely on the central limit theorem and approximate the null distribution of your t-statistic using the $N(0, 1)$ distribution.

   (a) (6 points) Write the null and alternative hypotheses as a mathematical statement about β_0.

   Solution: $H_0 : \ \beta_0 \geq 200$ v.s. $H_1 : \ \beta_0 < 200$

   (b) (12 points) Carry out your test at the 5% significance level ($\alpha = 0.05$). Please compute the t-statistic, state the rejection rule, obtain the critical value, and find the result of your test.

   Solution:

$$t = \frac{\bar{y} - 200}{se(\bar{y})} = \frac{\bar{y} - 200}{\sqrt{\frac{s^2}{n}}} = \frac{221.7 - 200}{\sqrt{\frac{6948.4}{50}}} = 1.841$$

   Because this is a left-tailed test, the rejection rule is to reject H_0 if $t < -cv_\alpha$ where α is the significance level. With $\alpha = 0.05$, the rejection rule is: reject H_0 if $t < -1.64$. Hence, the null cannot be rejected at the 5% level.

   (c) (10 points) What is the p-value of the test in part (a)? Interpret your p-value. Can you reject the null hypothesis at the 10% level based on your p-value?

   Solution: Let Z be the standard normal random variable. Because we are doing a left-tailed test, the p-value is $P(Z < t) = P(Z < 1.84) = 0.967$. The largest significance level such that we don't reject is, by definition of the p-value, the significance level 0.967. Since $0.1 < 0.967$, we cannot reject the null at the 10% level.

6. You are given the definition of the variables and the following regression output from Stata.

```
. des price lprice lotsize

Variable      Storage   Display    Value
    name         type    format    label      Variable label
------------------------------------------------------------------------------------------
price           float   %9.0g                 house price, $1000s
lprice          float   %9.0g                 log(price)
lotsize         float   %9.0g                 size of lot in square feet

. sum price lprice lotsize

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       price |         87     293.265    103.2748        111        725
      lprice |         87    5.631699    .3050125    4.70953   6.586172
     lotsize |         87    8.058241    4.732402          1         31

. reg lprice lotsize

      Source |       SS           df       MS      Number of obs   =        87
-------------+----------------------------------   F(1, 85)        =     51.62
       Model |  3.02301524         1  3.02301524   Prob > F        =    0.0000
    Residual |  4.97778952        85   .05856223   R-squared       =    0.3778
-------------+----------------------------------   Adj R-squared   =    0.3705
       Total |  8.00080476        86  .093032614   Root MSE        =      .242

------------------------------------------------------------------------------
      lprice | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     lotsize |   .0396177   .0055141     7.18   0.000     .0286541    .0505813
       _cons |    5.31245   .0514542   103.25   0.000     5.210145    5.414755
------------------------------------------------------------------------------
```

(a) (4 points) What is the maximum house price and average lot size in this sample?

Solution: The maximum house price is 725,000 dollars. The average lot size is 8,058 square feet. (There is a typo in the exam. The lot size is measured in 1,000 square feet. If your answer is 8.058 square feet, you get full credit.)

(b) (6 points) Write down the population regression model, being as specific as possible.

Solution:

$$log(price) = \beta_0 + \beta_1 lotsize + u$$

(c) (6 points) Write down the sample regression function, being as specific as possible.

Solution:

$$\widehat{log(price)} = 5.312 + .040\ lotsize$$

(d) (6 points) How do you interpret the slope coefficient, being as specific as possible?

Solution: Increasing 1,000 square feet of the lot size results in a 3.962% increase in housing price, on average. (Because of the typo, you get full credit if your answer is associated with 1 square foot.)

(e) (8 points) Find standard errors for the slope and intercept estimators respectively. Are these standard errors valid under SLR.1-SLR.4? Please explain why.

Solution: The standard error for the slope estimator is .006 and the standard error for the intercept estimator is .051. Stata reports homoskedastic standard errors by default. These standard errors are only valid under SLR.1-SLR.5.

(f) (7 points) Find the R^2 in this regression. How do you interpret the R^2, being as specific as possible?

Solution: The R^2 is 0.3778. 37.78% of the variation in log(price) is explained by lot size.

7. All else equal, we would expect people with greater ability to earn higher college GPAs. A common but crude measure of ability is results of IQ tests. This question explores the relationship between IQ test scores and college GPAs. Let $colgpa$ denote the college GPAs. Let IQ denote the scores of IQ tests. Using a random sample of $n = 500$ people, the following regression model estimates were obtained using OLS.

$$\widehat{colgpa}_i = 0.701 + 0.031\ IQ_i$$

(a) (8 points) What does the estimated coefficient for the slope parameter suggest about the magnitude of the relationship between IQ and college GPA? On average, what is the estimated difference in college GPA for a person with an IQ of 110 versus a person with an IQ of 100?

Solution: Increasing IQ by one point on average is associated with a college GPA increase of 0.031 points. The marginal difference is given by $(110-100)*0.031 = 0.31$.

(b) (5 points) Provide an example of what the error term u in this College GPA-IQ regression represents.

Solution: There is no single answer. One example could be the number of hours spent studying.

(c) (10 points) Write down SLR.4, the zero conditional mean assumption, using the variables in this example. How sure are you that Assumption SLR.4 holds? Give a well thought-out and logical reason for your answer.

Solution: SLR.4, $E(u|IQ) = 0$. Assumption SLR.4 is unlikely to hold. Suppose the hours spent studying are contained in u. Since study hours are expected to be correlated with IQ, u is correlated with IQ, which violates SLR.4.