Econ 322 Practice 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. In the equation $C = \beta_0 + \beta_1 I + u$, where C denotes consumption and I denotes income. What is the residual of the 5-th observation if $C_5 = \$500$ and $\hat{C}_5 = \$475$?

   A. \$975
   B. \$300
   C. \$25
   D. \$50

2. If the total sum of squares (SST) in a regression equation is 100, and the residual sum of squares (SSR) is 36, what is the explained sum of squares (SSE)?

   A. 64
   B. 56
   C. 32
   D. 18

3. Which of the following is a nonlinear regression model?

   A. $Y = \beta_0 + \beta_1 X^{1/2} + u$
   B. $\log(Y) = \beta_0 + \beta_1 \log(X) + u$
   C. $Y = 1/(\beta_0 + \beta_1 X) + u$
   D. $Y = \beta_0 + \beta_1 X + u$

4. 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 $n = 40$ Sundays. The sample average of those attendances is $\bar{y} = 231.7$ and the sample variance is $s^2 = 5948.4$. Consider the null hypothesis that the population mean, β_0, of attendance for a Sunday is 250 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) Write the null and alternative hypothesis as a mathematical statement about β_0.

Solution: $H_0: \ \beta_0 = 250$ v.s. $H_1: \ \beta_0 \neq 250$

(b) Carry out your test at the 10% significance level ($\alpha = 0.1$). Please compute the t-statistic, state the rejection rule, obtain the critical value, and find the result of your test.

Solution:

$$t = \frac{\bar{y} - 250}{se(\bar{y})} = \frac{\bar{y} - 250}{\sqrt{\frac{s^2}{n}}} = \frac{231.7 - 250}{\sqrt{\frac{5948.4}{40}}} = -1.501$$

Because this is a two-tailed test, the rejection rule is to reject H_0 if $|t| > cv_{\alpha/2}$ where α is the significance level. With $\alpha = 0.1$ and taking the critical value from the table, the rejection rule is: reject H_0 if $|t| > 1.64$. Hence, the null cannot be rejected at the 10% level.

(c) What is the p-value of the test in part (a)? Interpret your p-value.

Solution: Let Z be the standard normal random variable. Because we are doing a two-tailed test, the p-value is $2P(Z > |t|) = 2P(Z > 1.501) = 0.133$. The largest significance level such that we don't reject is, by definition of the p-value, the significance level 0.133.

(d) Compute a 99% non-rejection region (confidence interval) for β_0. Interpret your non-rejection region.

Solution: The formula for the confidence interval for a two-tailed test was derived in class:

$$\bar{y} - se(\bar{y})cv_{\alpha/2} \leq \alpha_0 \leq \bar{y} + se(\bar{y})cv_{\alpha/2}$$

Setting $\alpha = 0.01$, using $cv_{0.005} = 2.58$ and plugging in the data gives

$$231.7 - \sqrt{\frac{5948.4}{40}} * 2.58 \leq \alpha_0 \leq 231.7 + \sqrt{\frac{5948.4}{40}} * 2.58$$
$$200.238 \leq \alpha_0 \leq 263.162$$

We cannot reject at the 1% level any null for β_0 in the range 200.238 to 263.162.

5. To study the effects of the incinerator location on housing price, you run a regression of *lprice* on *ldist* in Stata and get the following results, where *lprice* is the log of the housing price and *ldist* is the log of the distance from the house to the incinerator.

```
. reg lprice ldist
```

Source	SS	df	MS	Number of obs	=	321
				F(1, 319)	=	43.48
Model	7.36920667	1	7.36920667	Prob > F	=	0.0000
Residual	54.0697787	319	.169497739	R-squared	=	0.1199
				Adj R-squared	=	0.1172
Total	61.4389853	320	.191996829	Root MSE	=	.4117

lprice	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ldist	.3172192	.0481095	6.59	0.000	.2225672	.4118713
_cons	8.257501	.4738309	17.43	0.000	7.325272	9.189729

(a) Write down the population regression model, being as specific as possible.

Solution:

$$ln(price) = \beta_0 + \beta_1 ln(dist) + u$$

(b) What is the meaning of the error term u in this regression? Provide an example of what u represents.

Solution: The error term is a variable in the regression equation that contains unobserved factors which affect the dependent variable, housing price. In this regression, u can represent the square footage of the house, the number of bedrooms, the number of bathrooms, etc.

(c) What is the sample size? What are the estimates of the intercept and slope parameters?

Solution: The sample size is 321. The estimated intercept is 8.258 and the estimated slope is 0.317.

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

Solution: A 1% increase in distance from the house to the incinerator results in a 0.32% increase in housing price, on average.

6. The following question is based on a real data set. Do students that attend lecture more regularly have different GPAs than students that attend lecture less often? To provide some evidence on this question, 674 students were randomly sampled at an elite university. Each student was asked to provide their GPA, percentage of classes they attended, and percentage of problem sets handed in. These variables are denoted by GPA_i, $attendpct_i$ and $pspct_i$. The following regression was estimated by OLS:

$$\widehat{GPA}_i \;=\; 0.67 \;+\; 0.0237\; attendpct_i, \qquad R^2 = 0.29$$

Assume for now that Assumptions SLR.1-SLR.4 hold.

(a) What does the estimated coefficient for the slope parameter suggest about the magnitude of the relationship between GPA and class attendance? On average, what is the estimated difference in GPA for a student that attends all lectures ($attendpct = 100$) and a student that always skips the Friday lecture (in a class that meets Monday, Wednesday, and Friday)?

Solution: Increasing attendance rate by one percentage point on average is associated with a GPA increase of .0237 points holding all else constant. The marginal difference is given by $(100 - 66.66) * .0237 = 0.79$ which is like going from a 4.0 to about a 3.0 grade level (a big difference).

(b) How much of the variation in GPA is explained by lecture attendance? Does this regression model suggest that GPA is primarily explained by lecture attendance? Why or why not?

Solution: $R^2 = 0.29$ and hence 29% of the variation in GPA is explained by attendance whereas 71% of the variation in GPA is explained by unobserved factors. Attendance explains a nontrivial portion of the variation in GPA.

(c) A critic might argue that there are other important factors besides class attendance that affect GPA. One obvious variable that might affect GPA is the problem set hand-in rate. A regression, estimated by OLS, of *attendpct* on *pspct* yielded the following.

$$\widehat{attendpct}_i \;=\; 35 \;+\; 0.536\; pspct_i, \qquad R^2 = 0.39$$

Given this information, how sure are you that Assumption SLR.4 holds for the GPA-attendpct regression? Give a well thought-out and logical reason for your answer.

Solution: There is evidence that *attendpct* and *pspct* are correlated. Because *pspct* is part of u in the original regression, we have evidence that *attendpct* is correlated with part of u and SLR.4 is violated.

(d) How does your answer to part (c) affect the validity of your answers to parts (a)-(b)?

Solution: With SLR.4 violated, the OLS estimates in the original regression will be biased and our analysis in part (a) won't be valid. At least the R^2 is still valid so we still learn that 29% of the variation in GPA is explained by *attendpct*.