Econ 322 Practice Final, Spring 2025

Name: _____

Instructions: This is a closed-book exam. However, you are allowed to bring four 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 120 minutes.

1. The following simple model is used to determine the annual savings of an individual on the basis of his annual income and education. $Savings = \beta_0 + \delta_0 Educ + \beta Inc + u$ The variable *Edu* takes a value of 1 if the person is educated and the variable *Inc* measures the income of the individual.

   Refer to the model above. The inclusion of another binary variable in this model that takes a value of 1 if a person is uneducated, will give rise to the problem of

   A. omitted variable bias

   B. self-selection

   C. dummy variable trap

   D. heteroskedasticity

2. The significance level of a test is:

   A. the probability of rejecting the null hypothesis when it is false.

   B. one minus the probability of rejecting the null hypothesis when it is false.

   C. the probability of rejecting the null hypothesis when it is true.

   D. one minus the probability of rejecting the null hypothesis when it is true.

3. The linear probability model contains heteroskedasticity unless

   A. the intercept parameter is zero

   B. the independent variables are binary

   C. all the slope parameters are positive

   D. all the slope parameters are zero

4. Consider the population regression model

$$wage = \beta_0 + \beta_1\ urban + \beta_2\ married + \beta_3\ urban \times married + u,$$

where *wage* is monthly earnings measured in dollars, *married* is a binary variable equal to one if the individual is married and zero otherwise, *urban* is another binary variable equal to one if the individual lives in urban area and zero otherwise, and *urban* × *married* is the interaction between *urban* and *married*. The Stata output is given below.

```
. reg wage urban married urban_married, r

Linear regression                               Number of obs     =        935
                                                F(3, 931)         =      23.50
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0606
                                                Root MSE          =     392.54
```

wage	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
urban	133.2084	73.16928	1.82	0.069	-10.38747	276.8042
married	147.0516	65.4751	2.25	0.025	18.55571	275.5475
urban_married	55.33681	78.50774	0.70	0.481	-98.73583	209.4095
_cons	695.8696	61.38277	11.34	0.000	575.4049	816.3342

(a) Using the above estimates, what is the average wage for unmarried people living in non-urban area?

Solution: 695.870 dollars

(b) What is the average wage for unmarried people living in urban area?

Solution: $695.8696 + 133.2084 = 829.078$ dollars

(c) What is the average wage for married people living in non-urban area?

Solution: $695.8696 + 147.0516 = 842.921$ dollars

(d) What is the average wage for married people living in urban area?

Solution: $695.8696 + 133.2084 + 147.0516 + 55.33681 = 1031.466$ dollars

(e) What is the difference between the "urban" wage premium for married people and the "urban" wage premium for unmarried people?

Solution: 55.337 dollars

(f) Is the difference in part(e) statistically significant at the 10% level? Why? (Hint: You can find the answer from the Stata output table.)

Solution: No, because the p-value is greater than 0.1.

5. Using data from the national longitudinal study of youth, we find the following results for a regression of log weekly wage on years of education, experience, experience squared, and an intercept:

$$\widehat{log(wage)} = \underset{(0.222)}{4.016} + \underset{(0.008)}{0.092}\, educ + \underset{(0.025)}{0.079}\, exper - \underset{(0.001)}{0.002}\, exper^2$$
$$N = 15,310 \quad R^2 = 0.12$$

(a) Construct a 95% confidence interval for the effect of years of education on log weekly earnings.

Solution:

$$[0.092 - 1.96 \times 0.008, 0.092 + 1.96 \times 0.008] = [0.076, 0.108]$$

(b) Consider an individual with 10 years of experience. What would you expect to be the return to an additional year of experience for such an individual (the effect on log weekly earnings)?

Solution:

$$0.079 - 2 \times 0.002 \times 10 = 0.039$$

Or more accurately,

$$0.079 \times 11 - 0.002 \times 11^2 - (0.079 \times 10 - 0.002 \times 10^2) = 0.037$$

(c) How would you test the hypothesis that experience does not impact salary? Specifically, what would be the null and alternative hypotheses? Do you have enough information from the sample regression function to test this hypothesis? Explain your answer carefully.

Solution:

$$H_0 : \beta_{exper} = 0 \text{ and } \beta_{exper^2} = 0$$

$$H_1 : H_0 \text{ not true}$$

Because we need a joint F-test to test this null hypothesis, we do not have enough information from the sample regression function.

(d) Labor economists studying the relation between education and earnings are often concerned about what they call "ability bias." Suppose that individuals differ in ability, and that the correct specification of the regression function is one that includes ability (To simplify the problem, let us ignore *exper* for now.):

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 ability + u$$

In this regression, what do you expect the sign of β_2 (the coefficient on ability) to be?

Solution: Expect it to be positive.

(e) What do you think is the sign of the correlation between ability and years of education?

Solution: Positive correlation between ability and years of education.

(f) If we estimate the regression function with ability included, do you think that the estimated value of β_1 will be greater or less than what it was in the regression without ability? Explain.

Solution: The regression without *ability* will suffer from the omitted variable bias. The bias is $\hat{\beta}_2 \times \tilde{\delta}_1$, where $\tilde{\delta}_1$ is the slope of *educ* in the regression of *ability* on *educ*. Since both $\hat{\beta}_2$ and $\tilde{\delta}_1$ are expected to be positive, the omitted bias is positive. That is, the estimated value of β_1 without *ability* will be larger than the current one with *ability*.

6. This dataset provides information on the purchase of ecologically friendly apples. There are several factors involved in the purchase decision, including

- *ecobuy* = 1 if at the prices given, an individual would buy any ecologically friendly apples and = 0 otherwise;
- *ecoprc* denotes the price of ecolabeled apples per pound in dollars;

- *regprc* denotes the price of regular apples per pound in dollars;
- *educ* indicates years of schooling;
- *male* = 1 if the individual is male and = 0 otherwise.

```
. sum ecobuy

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
      ecobuy |        660    .6242424    .4846852          0          1


. gen male_educ = male*educ

. reg ecobuy ecoprc regprc educ male male_educ, r

Linear regression                               Number of obs     =        660
                                                F(5, 654)         =      18.75
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1112
                                                Root MSE          =     .45867

------------------------------------------------------------------------------
             |               Robust
      ecobuy | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      ecoprc |  -.8240013   .1063145    -7.75   0.000     -1.03276   -.6152423
      regprc |     .73431   .1300799     5.65   0.000     .4788854    .9897346
        educ |   .0203179   .0093772     2.17   0.031     .0019048    .0387309
        male |  -.4651065    .263034    -1.77   0.077    -.9815996    .0513866
   male_educ |   .0255286   .0176815     1.44   0.149    -.0091908     .060248
       _cons |   .6003401   .1559916     3.85   0.000     .2940353     .906645
------------------------------------------------------------------------------

. test male male_educ

 ( 1)  male = 0
 ( 2)  male_educ = 0

       F(  2,   654) =    3.49
            Prob > F =    0.0310
```

(a) What fraction of individuals claim they would buy ecolabeled apples?

Solution: 62.424%

(b) What is the estimated coefficient on *ecoprc*? How do you interpret it?

> **Solution:** -.824. Holding other explanatory variables fixed, if the price per pound for ecolabeled apples increases by one dollar, the probability of buying any ecologically friendly apples falls by 82.4 percentage points on average.

(c) How would you test the hypothesis that gender is not a factor in purchase decisions? Specifically, what would be the null and alternative hypotheses? Do you reject or fail to reject the null hypothesis at a significance level of 0.05? Be sure to explain how you came to this conclusion.

> **Solution:**
>
> $$H_0 : \beta_{male} = 0 \text{ and } \beta_{male_educ} = 0$$
>
> $$H_1 : \text{not } H_0$$
>
> As the output from the Stata code indicates, the resulting F-stat is 3.49, with a p-value of $0.031 < 0.05$, so we reject the null hypothesis in favor of the alternative at the 5% level.

7. Use the following Stata output to answer this question. The model of interest is

$$math12 = \beta_0 + \beta_1\ cathhs + \beta_2\ lfaminc + u,$$

where $math12$ is the mathematics standardized score of the student measured in points, $cathhs$ is a binary variable equal to one if the student attends a Catholic high school, and $lfaminc$ is log of family income.

```
. reg math12 cathhs lfaminc, robust

Linear regression                               Number of obs     =      7,430
                                                F(2, 7427)        =     465.85
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1028
                                                Root MSE          =     8.9608

------------------------------------------------------------------------------
             |               Robust
      math12 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      cathhs |   2.102619    .408851     5.14   0.000     1.301156    2.904083
     lfaminc |   3.695732   .1267883    29.15   0.000     3.447191    3.944273
       _cons |    13.7426   1.312541    10.47   0.000     11.16964    16.31555
------------------------------------------------------------------------------

. ivregress 2sls math12 lfaminc (cathhs= parcath), robust

Instrumental variables 2SLS regression          Number of obs   =      7,430
                                                Wald chi2(2)    =     894.55
                                                Prob > chi2     =     0.0000
                                                R-squared       =     0.1023
                                                Root MSE        =     8.9615

------------------------------------------------------------------------------
             |               Robust
      math12 | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      cathhs |   1.207227   1.542194     0.78   0.434    -1.815417    4.229871
     lfaminc |   3.725553    .135551    27.48   0.000     3.459878    3.991228
       _cons |   13.48832   1.373757     9.82   0.000     10.79581    16.18084
------------------------------------------------------------------------------
Instrumented: cathhs
 Instruments: lfaminc parcath
```

(a) How many students are in the sample?

Solution: 7,430 observations

(b) If *cathhs* is correlated with u, what does this mean for the linear regression (OLS) estimator of β_1?

Solution: biased and inconsistent

(c) Some researchers have suggested using an instrumental variables (IV) approach

to estimate β_1. Specifically, a recent study suggested using the variable *parcath*, which is equal to one if a parent of the student reports being Catholic and zero if not, as an instrumental variable for *cathhs*. Write down the first stage of the two-stage least squares.

Solution:

$$cathhs = \pi_0 + \pi_1\ parcath + \pi_2\ lfaminc + v$$

(d) What two conditions does *parcath* have to satisfy to serve as a suitable IV for *cathhs*? Which of these can be formally tested, and how? Be specific. (Hint: You need to use the first stage in part (c).)

Solution: $Cov(parcath, u) = 0$ or exogeneity

$\pi_1 \neq 0$ or *parcath* is partially correlated with *cathhs*

Partial correlation can be tested.

Test whether $\pi_1 = 0$ using either t statistic or F statistic in the first stage. Need to include $lfaminc$ as well.

(e) What are the OLS and IV estimates of β_1? How do you interpret these estimates?

Solution: OLS: 2.10. Everything else equal, students who attend Catholic high schools score 2.10 points higher in math than students who do not attend Catholic high schools on average.

IV: 1.21. Everything else equal, students who attend Catholic high schools score 1.21 points higher in math than students who do not attend Catholic high schools on average.