Name: _____

In-class final, EC421

180 points possible

1 True or false (105 points; 35 questions)

Note: In this section, select the correct answer (true or false). You do not need to explain your answer.

1. (2.5 points) **[T/F]** If the disturbance in an OLS regression correlates with a regressor, exogeneity is violated.
2. (2.5 points) **[T/F]** Homoskedasticity means that the variance of the disturbances differs across observations.
3. (2.5 points) **[T/F]** In the potential outcomes framework, $\mathbb{E}[\mathsf{Y}_1|D_i = 0]$ represents the average untreated outcome for the treated group.
4. (2.5 points) **[T/F]** The *fundamental problem of causal inference* is that we cannot observe the both the treated and untreated outcomes for the same individual.
5. (2.5 points) **[T/F]** The variable $x_t = 0.3t + \varepsilon_t$ is stationary (assume ε_t is mean zero).
6. (2.5 points) **[T/F]** *Selection bias* occurs when the individuals *select* into treatment or control groups.
7. (2.5 points) **[T/F]** Measurement error in an explanatory variable tends to bias OLS's coefficient estimates toward zero.
8. (2.5 points) **[T/F]** There are situations where OLS is consistent but biased for estimating coefficients.
9. (2.5 points) **[T/F]** The instrumental variable (IV) estimator requires that the regressor of interest is exogenous.
10. (2.5 points) **[T/F]** The variance of the disturbance affects whether OLS is biased for estimating coefficients.

11. (2.5 points) **[T/F]** Adding additional variables always increases R^2.

12. (2.5 points) **[T/F]** The Goldfeld-Quandt test is a test for autocorrelation.

13. (2.5 points) **[T/F]** OLS is biased when estimating coefficients in models with lagged outcome variables.

14. (2.5 points) **[T/F]** The presence of omitted-variable bias implies that OLS is also inconsistent.

15. (2.5 points) **[T/F]** If the disturbance is heteroskedastic, conventional OLS-based confidence intervals will be invalid.

16. (2.5 points) **[T/F]** The first stage of two-stage least squares (2SLS) involves regressing the outcome variable on the instrumental variable.

17. (2.5 points) **[T/F]** Two-stage least squres can address bias arising from omitted variables.

18. (2.5 points) **[T/F]** If OLS is biased for estimating β_2 below, then it is also inconsistent for estimating β_2.

$$\text{Health}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Health}_{t-1} + u_t$$

19. (2.5 points) **[T/F]** With a *p*-value of 0.997, you would strongly reject the null hypothesis.

20. (2.5 points) **[T/F]** Random walks are non-stationary because their means grow over time.

21. (2.5 points) **[T/F]** In the Rubin causal model, Y_{1i} refers to the outcome for individual i when they receive treatment.

22. (2.5 points) **[T/F]** If u_i correlates with u_j, then exogeneity is violated.

23. (2.5 points) **[T/F]** The OLS estimator estimates the coefficients by minimizing $\sum_i (y_i \hat{y}_i)^2$, where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ (for a simple linear regression).

24. (2.5 points) **[T/F]** Consistency describes the behavior of an estimator as the sample size grows.

25. (2.5 points) **[T/F]** If a Goldfeld-Quandt test finds $\text{SSE}_1/\text{SSE}_2 > 100$, then it will likely conclude that there is no statistically significant evidence of heteroskedasticity.

26. (2.5 points) **[T/F]** In the model $\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + u_t$, the parameter β_1 gives the effect of income in the previous time period on births in the current period.

27. (2.5 points) **[T/F]** *Cluster-robust* standard errors account for the possibility that u_i may correlate with u_j for some individuals.

28. (2.5 points) **[T/F]** Residuals are observable; disturbances are not.

29. (2.5 points) **[T/F]** The econometric model below allows the effect of age on health to depend upon the individual's gender.

$$\text{Health}_i = \beta_0 + \beta_1 \text{Female}_i + \beta_2 \text{Age}_i + \beta_3 \text{Female}_i \times \text{Age}_i + u_i$$

30. (2.5 points) **[T/F]** In the econometric model below, the effect of age can depend on the individual's age.

$$\text{Health}_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Age}_i^2 + u_i$$

31. (2.5 points) **[T/F]** Correlated disturbances make OLS biased when estimating coefficients.

32. (2.5 points) **[T/F]** Randomizing the explanatory variable helps avoid selection bias.

33. (2.5 points) **[T/F]** In the model below, if the disturbance is *not* autocorrelated, then OLS consistently estimates the coefficients (assume the 'standard' assumptions hold).

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 y_{t-1} + u_t$$

34. (2.5 points) **[T/F]** In the model

$$\text{log Quantity}_t = \beta_0 + \beta_1 \text{ log Price}_t + u_t$$

if $\widehat{\beta}_1 = -0.3$, then, on average, a one-percent increase in price generates a 30-percent reduction in quantity (all else equal).

35. (2.5 points) **[T/F]** If the variable x_t is a random walk, then $\Delta x_t = x_t - x_{t-1}$ is stationary.

36. (2.5 points) **[T/F]** The definition of causality discussed in our class assumes that the treatment effect for and individual does not depend upon the treatment status of other individuals.

2 Short answer (75 points; 16 questions)

Note: In this section, briefly answer the questions/prompts in 1–3 short (and complete) sentences. We will deduct points for excessively long answers.

36. Suppose you estimate the following model with OLS

$$\text{Income}_i = \beta_0 + \beta_1 \text{Education}_i + \beta_2 \text{Female}_i + \beta_3 \text{Education}_i \times \text{Female}_i + u_i$$

and find parameter estimates $\hat{\beta}_0 = 30$, $\hat{\beta}_1 = 1$, $\hat{\beta}_2 = -5$, and $\hat{\beta}_3 = 0.5$.

(a) (5 points) Interpret $\hat{\beta}_3$ in the context of the model.

(b) (5 points) What is the expected income for a non-female individual with 12 years of education?

(c) (5 points) Explain how the regression above estimates two lines—and provide the slope of each line.

(d) (5 points) Explain how you could use this model to test for gender-based discrimination.

37. People commonly debate the impact of the minimum wage on unemployment. Suppose you have data on unemployment and minimum wage for 100 cities in the United States.

(a) (3 points) Write down the econometric (regression) model you would use to estimate the effect of minimum wage on unemployment.

(b) (2 points) Which parameter in your model gives the effect of minimum wage on unemployment?

(c) (5 points) Suppose you estimate your model (in (a) above) with OLS. Will these OLS estimates provide causal evidence about the effect of minimum wage on unemployment? Explain your answer.

(d) (5 points) Could instrumental variables help here? Explain.

38. (5 points) Define the concept of a *standard error*.

39. (5 points) Using mathematical notation, explain the concept of *exogeneity*. Then explain why exogeneity is important for OLS.

40. (5 points) Suppose you are analyzing time-series data for your job. Why should you be concerned about non-stationarity?

41. (5 points) The probability limit of the instrumental variables (IV) estimator (with instrument z, endogenous regressor x, and disturbance u) is

$$\text{plim } \widehat{\beta}_{\text{IV}} = \beta + \frac{\text{Cov}(z, u)}{\text{Cov}(z, x)}$$

Use the equation above to explain how a valid instrument makes the IV estimator consistent.

42. (5 points) Explain how correlated disturbances can affect OLS.

43. (5 points) Explain how randomized experiments avoid omitted-variable bias.

44. (5 points) Suppose you want to know how a recent advertising campaign affected sales. You have monthly data on sales (in dollars) and advertising spending (in dollars) for several years. Write down the econometric model you would use to estimate the effect of advertising on sales.

45. (5 points) What are the key assumptions required to interpret your coefficient estimates from the model in the previous question as causal? Are these assumptions plausible in this context? Explain.