Name: _____

In-class midterm, EC421

140 points possible

1 True or false (50 points; 20 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]** When disturbances' means differ across groups, we have heteroskedasticity.
2. (2.5 points) **[T/F]** Adding additional variables mechanically increases R^2.
3. (2.5 points) **[T/F]** If you omit a variable from a regression, it will cause omitted-variable bias.
4. (2.5 points) **[T/F]** Heteroskedastic disturbances make OLS biased for estimating coefficients.
5. (2.5 points) **[T/F]** If we estimate the econometric model below via regression, $\widehat{\beta}_1$ will equal average(education for females) minus average(education for non-females).

$$\text{Education}_i = \beta_0 + \beta_1 \text{Female}_i + u_i$$

6. (2.5 points) **[T/F]** Exogeneity essentially says that the disturbance must be independent of your explanatory variables.
7. (2.5 points) **[T/F]** A *p*-value larger than 0.05 suggests the data do not support the null hypothesis.
8. (2.5 points) **[T/F]** The econometric model below allows the effect of gender on education to depend upon the individual's age.

$$\text{Education}_i = \beta_0 + \beta_1 \text{Female}_i + \beta_2 \text{Age}_i + u_i$$

9. (2.5 points) **[T/F]** Correlated disturbances make OLS biased when estimating standard errors.
10. (2.5 points) **[T/F]** Weighted least squares (WLS) upweights individuals with high-variance disturbances and downweights individuals with low-variance disturbances.

11. (2.5 points) **[T/F]** A variable will cause omitted-variable bias in OLS estimates for coefficients when the following things are true:

    1. A variable is omitted from the regression.
    2. The omitted variable correlates with one of the included regressors.

12. (2.5 points) **[T/F]** The heteroskedasticity-robust standard error estimator is unbiased when the disturbance is homoskedastic.

13. (2.5 points) **[T/F]** If your model of interest is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

    then the regression for the corresponding White test for heteroskedasticity should estimate

$$e_i^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1^2 + \alpha_5 x_2^2 + \alpha_6 x_3^2 + v_i$$

14. (2.5 points) **[T/F]** In the model

$$\log \text{Income}_i = \beta_0 + \beta_1 \text{Education}_i + \beta_2 \log \text{Age}_i + u_i$$

    If $\hat{\beta}_1 = 0.57$, then, on average, a one-percent increase in education generates a 57-percent increase in income (all else equal).

15. (2.5 points) **[T/F]** Suppose you observe that whenever the variable X increases, the variable Y increases. **True or false:** Based on this observation, either X causes Y or Y causes X.

16. (2.5 points) **[T/F]** If u_i correlates with u_j, then exogeneity is violated.

17. (2.5 points) **[T/F]** In the presence of heteroskedasticity, WLS can be more efficient than OLS.

18. (2.5 points) **[T/F]** If a Goldfeld-Quandt test finds that SSE_1 equals SSE_2, then it will conclude that there is statistically significant evidence of heteroskedasticity.

19. (2.5 points) **[T/F]** Using a linear model rather than a log-log model could cause heteroskedasticity.

20. (2.5 points) **[T/F]** While disturbances are unobservable, the residuals *are* observable.

2 Multiple choice (20 points; 5 questions)

Note In this section, check (✓ or ×) **all** correct answers. You do not need to explain your answer.

21. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Which of the following statements are part of the "standard" assumptions for OLS regression?

    ○ $E[u_i|X_i] = 0$ ○ $Var(u_i) = 0$ ○ $Var(X_i) > 0$ ○ $Cov(u_i, u_j) = 0$

22. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Which statements must be true for the OLS estimator $\hat{\beta}$ to be unbiased?

    ○ $E[u_i|X_i] = 0$ ○ $E[\hat{\beta}] = \beta$ ○ $Var(u_i) = \sigma^2$ ○ *p*-value < 0.05

23. (4 points) **[Multiple choice]** Choose *all* correct answers:

    What will happen if the sum of squared errors (SSE) decreases (holding everything else constant)?

    ○ R^2 increases ○ *p*-value increases ○ confidence interval widens ○ t statistic increases

24. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Suppose you have evidence that heteroskedasticity is present. What are your options?

    ○ WLS ○ Stick with OLS standard errors ○ Fix your inference ○ Check your specification

25. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Which specifications are "allowed" under the OLS's requirement of *linearity*?

    ○ $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$

    ○ $y = \beta_0 + \beta_1 x_1 + \beta_2 (x_2 \times u)$

    ○ $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1)^2 + u$

    ○ $y = \beta_0 + \beta_1 x_1 + (x_2)^{\beta_3} + u$

3 Short answer (50 points; 10 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.

26. (5 points) Define the concept of a ‘standard error’.

27. (5 points) Explain why standard errors are important in econometrics.

28. (5 points) Why is weighted least squares often ‘infeasible’?

29. (5 points) We've all heard that *correlation is not causation*. Using terms from our course, explain why.

30. (5 points) Where does the "least squares" part of the name in OLS come from?

31. (5 points) Compare and contrast the disturbance and the residual.

32. (5 points) For the model below, suppose our estimates of the parameters are $\hat{\beta}_0 = 7.4$, $\hat{\beta}_1 = -0.2$, and $\hat{\beta}_2 = 1.4$. Interpret each of the coefficient estimates (you can ignore the intercept).

$$\text{log Crime}_i = \beta_0 + \beta_1 \text{ log Police}_i + \beta_2 \text{ log Population}_i + u_i$$

33. (5 points) For the model below, assume the variable Youth_i is an indicator that equals one when the individual is less than 16 years old. Interpret the coefficients β_2 and β_3 and then explain the potential importance of including $\text{Income}_i \times \text{Pollution}_i$.

$$\text{Health}_i = \beta_0 + \beta_1 \text{Youth}_i + \beta_2 \text{Pollution}_i + \beta_3 \text{Youth}_i \times \text{Pollution}_i + u_i$$

34. (5 points) In the following model, *Income* is measured in dollars, *Female* is an indicator, and *Age* is reported in years (since birth).

$$\text{Income}_i = \beta_0 + \beta_1 \text{Female}_i + \beta_2 \text{Age}_i + \beta_3 \text{Female}_i \times \text{Age}_i + u_i \tag{1}$$

Answer the following questions (you will have βs in your answers).

- What is the expected income for a 50-year-old female?
- What is the average difference between a 20-year-old female and 20-year-old non-female?

35. (5 points) Suppose we are analyzing the (very familiar) model

$$\text{Income}_i = \beta_0 + \beta_1 \text{Education}_i + u_i$$

using a sample of individuals from across the state of Oregon.

Explain why we should be concerned about correlated disturbances. Your answer should include a specific example of why the disturbances could correlate.

4 Short-answer and a graph (20 points; 4 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.

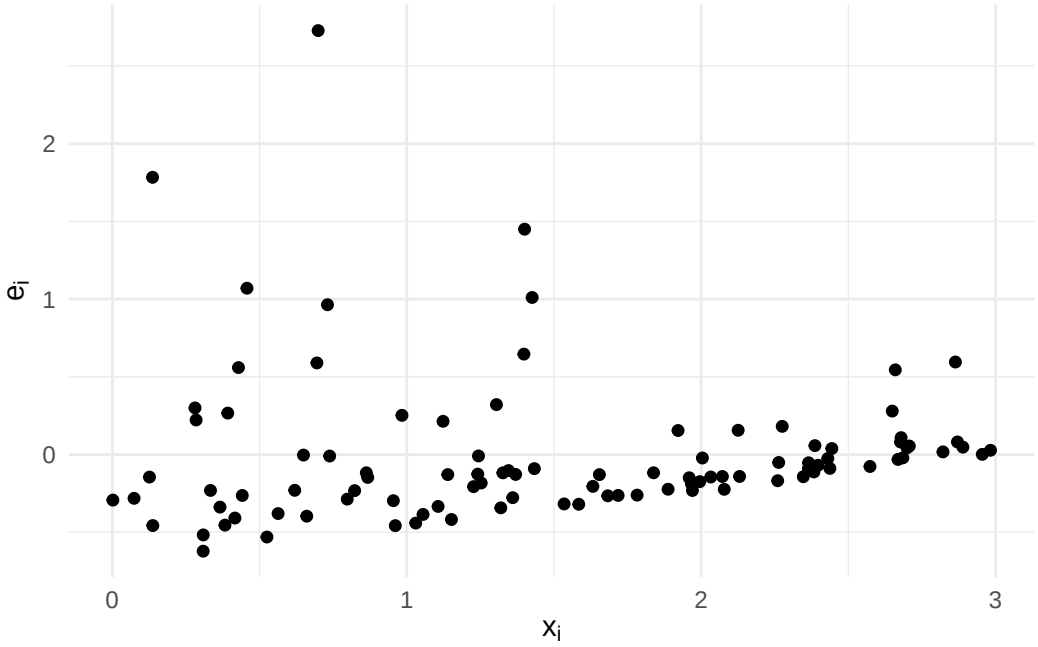

Figure 1: Assumptions

Description In Figure 1 (above), I estimated a simple linear regression by regressing y_i on x_i. I then plotted the resulting residuals (e_i) against x_i.

36. (5 points) What—if anything—does the plot suggest about heteroskedasticity? Explain your answer.

37. (5 points) Explain what you would expect to find if you conducted a Goldfeld-Quandt test for heteroskedasticity.

38. (5 points) What—if anything—does the plot suggest about exogeneity? Explain your answer.

39. (5 points) What would you do if you saw a residual plot like the one in Figure 1? Explain.