Name: _____

In-class midterm, EC421

145 points possible

1 True or false (75 points; 30 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 correlates with a regressor, then exogeneity is violated.
2. (2.5 points) **[T/F]** When OLS is biased for the standard errors, it is also biased for the coefficients.
3. (2.5 points) **[T/F]** The *linearity* assumption in OLS prohibits models from including squared terms like x_i^2.
4. (2.5 points) **[T/F]** Omitted-variable bias occurs when an omitted variable correlates with an included regressor.
5. (2.5 points) **[T/F]** Adding more explanatory variables will always increase R^2.
6. (2.5 points) **[T/F]** Heteroskedasticity makes OLS estimates of coefficients biased.
7. (2.5 points) **[T/F]** Weighted least squares gives more weight to observations with smaller variances in their disturbances.
8. (2.5 points) **[T/F]** Residuals are observable; disturbances are not.
9. (2.5 points) **[T/F]** If $\widehat{\beta}_1 = 0.3$ in a log-log model, a 1% increase in x increases y by 30%.
10. (2.5 points) **[T/F]** A p-value of 0.99 indicates a high degree of statistical significance.
11. (2.5 points) **[T/F]** Consistency describes the behavior of an estimator as sample size grows.
12. (2.5 points) **[T/F]** Heteroskedasticity-robust standard errors are biased in the presence of homoskedasticity.
13. (2.5 points) **[T/F]** Correlated disturbances violate the exogeneity assumption.
14. (2.5 points) **[T/F]** The Goldfeld-Quandt test compares error variances across two subsets of the data.

15. (2.5 points) **[T/F]** A coefficient on an interaction term captures how the effect of one variable depends on the level of another.

16. (2.5 points) **[T/F]** Functional-form misspecification can lead to heteroskedasticity.

17. (2.5 points) **[T/F]** Homoskedasticity means Var$(u_i) \neq$ Var(u_j) for two individuals i and j.

18. (2.5 points) **[T/F]** The White test is better than the Goldfeld-Quandt test at detecting general heteroskedasticity.

19. (2.5 points) **[T/F]** The sum of squared residuals directly influences the magnitude of the standard errors.

20. (2.5 points) **[T/F]** If Cov$(x_i, u_i) \neq 0$, then OLS is still unbiased but inconsistent.

21. (2.5 points) **[T/F]** In the regression $\log(y_i) = \beta_0 + \beta_1 x_i + u_i$, the effect of x on y is percent-based.

22. (2.5 points) **[T/F]** A violation of exogeneity implies biased OLS coefficient estimates.

23. (2.5 points) **[T/F]** OLS minimizes the $\sum_i e_i$, where e_i is the residual.

24. (2.5 points) **[T/F]** The presence of heteroskedasticity invalidates standard OLS hypothesis tests.

25. (2.5 points) **[T/F]** For an estimator with bias $1/n$, the estimator may still be consistent.

26. (2.5 points) **[T/F]** The assumption $\mathbb{E}[u_i^2|X_i] = \sigma^2$ is necessary for OLS to be unbiased for the coefficients.

27. (2.5 points) **[T/F]** The p-value tells us the probability that our hypothesis is true.

28. (2.5 points) **[T/F]** WLS is less efficient that OLS in the presence of heteroskedasticity.

29. (2.5 points) **[T/F]** For the regression model

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

the White test for heteroskedasticity would run the following regression:

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

30. (2.5 points) **[T/F]** Heteroskedasticity-robust standard errors and 'plain' OLS standard errors will always use the same coefficient estimates.

2 Multiple choice (20 points; 5 questions)

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

31. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Which of the following assumptions are "classic" regression assumptions?

    ○ $E[u_i|X_i] = 0$ ○ $\text{Var}(u_i) = 0$ for all i ○ $\text{Cov}(u_i, u_j) = \sigma$ ○ $\text{Var}(X) > 0$

32. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Which of the relationships imply OLS is biased for estimating the coefficients in a regression model?

    ○ $E[u_i|X_i] = 0$ ○ $\text{Var}(u_i) \neq \text{Var}(u_j)$ ○ $E\left[\hat{\beta}\right] \neq \beta$ ○ $\text{Cov}(u_i, u_j) = 0$

33. (4 points) **[Multiple choice]** Choose *all* correct answers:

    In the presence of heteroskedasticity, which of the following statements are true?

    ○ OLS is unbiased for the coefficients. ○ Standard OLS confidence intervals are biased.

    ○ OLS is unbiased for the standard errors. ○ OLS is less efficient than WLS.

34. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Imagine you are in a setting where you believe the disturbance is heteroskedastic. What are your 'options' for estimating the model that will yield believable estimates?

    ○ Use het.-robust standard errors. ○ Estimate the model using WLS.

    ○ Check the specification. ○ Look for omitted variables.

35. (4 points) **[Multiple choice]** Choose *all* correct answers:

    Which of the following will generally make our standard errors smaller?

    ○ Adding additional regressors to the model. ○ Omitted variable bias.

    ○ Ignoring correlated disturbances. ○ Subtracting the mean from the outcome variable.

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.

36. (5 points) Define the concept of a *standard error*.

37. (5 points) Explain how OLS defines the “best-fit” line.

38. (5 points) Explain the difference between a *residual* and a *disturbance*.

39. (5 points) We showed in class that the probability limit of the OLS estimator (for the slope coefficient in a simple linear regression) is

$$\text{plim}\,\hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x,u)}{\text{Var}(x)}$$

Using this formula, explain how our OLS assumptions imply that the OLS estimator is consistent (when the assumptions are satisfied).

40. (5 points) Define the two requirements for a variable to cause *omitted-variable bias*.

41. (5 points) How do we quantify the uncertainty behind our OLS estimates?

The questions on this page refer to the regression model below.

$$\text{Health}_i = \beta_0 + \beta_1 \text{Income}_i + \beta_2 \text{Hispanic}_i + \beta_3 \text{Income}_i \times \text{Hispanic}_i + u_i$$

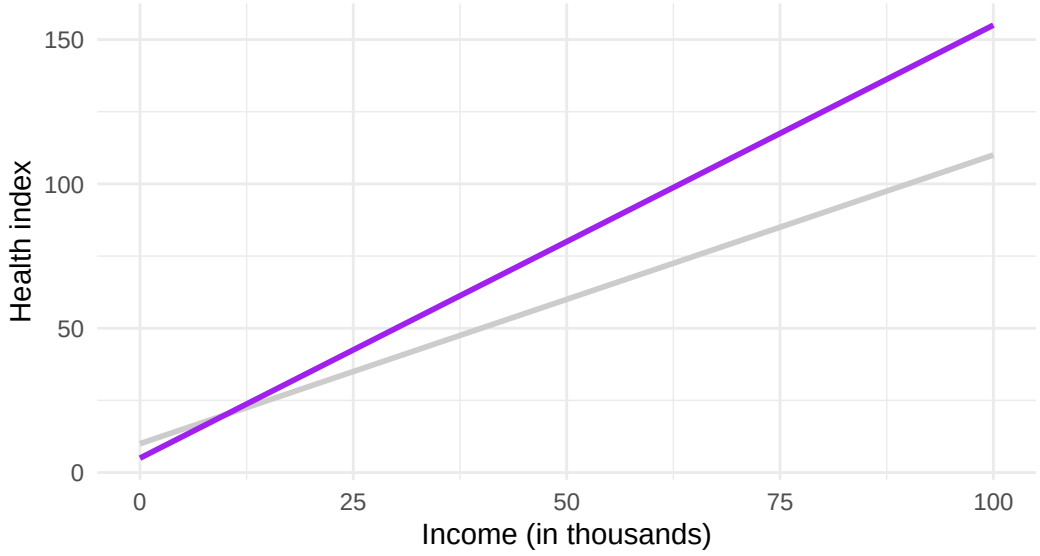
where Health_i is an index that runs between 0 and 100; Income_i measured in thousands of dollars; Hispanic_i is a binary (indicator) variable that equals 1 if individual i is Hispanic and 0 otherwise.

Suppose $\beta_0 = 10$, $\beta_1 = 1$, $\beta_2 = -5$, and $\beta_3 = 0.5$.

42. (5 points) Interpret β_1 and β_3 in the context of the model above.

43. (5 points) What is the expected value of health for a Hispanic individual with an income of $50,000?

44. (5 points) Draw (approximately) the graph of the two lines implied by the preceding regression model. Don't worry about making it perfect—but do make sure to label the axes and get the general idea right.



45. (5 points) Draw a scatter plot that illustrates heteroskedasticity. Label the axes.