Name: ____

In-class midterm, EC421

Part 1: True or false (50 points)

Note: In this section, only select the correct answer. You do not need to explain your answer.

1. **[T/F]** (2pts) If any variables in the disturbance correlate with the explanatory variables in a regression, then exogeneity is violated.
2. **[T/F]** (2pts) Heteroskedasticity occurs when $E[u_i|X_i] = 0$.
3. **[T/F]** (2pts) An estimator that is biased can still be consistent.
4. **[T/F]** (2pts) Weighted least squares (WLS) downweights individuals with high-variance disturbances and upweights individuals with low-variance disturbances.
5. **[T/F]** (2pts) If the estimator $\hat{\alpha}$ is consistent for the parameter α, then $E[\hat{\alpha}] = \alpha$.
6. **[T/F]** (2pts) The *linearity* requirement (assumption) for OLS prohibits nonlinear transformations like x^2 (for example, $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$).
7. **[T/F]** (2pts) In the presence of heteroskedasticity, OLS produces unbiased standard errors.
8. **[T/F]** (2pts) Adding additional variables always increases R^2.
9. **[T/F]** (2pts) In the presence of omitted-variable bias, OLS's coefficient estimates are biased toward zero.
10. **[T/F]** (2pts) If u_i and u_j are correlated (for different individuals i and j), then we have a violation of exogeneity.
11. **[T/F]** (2pts) In the presence of heteroskedasticity, WLS is more efficient that OLS.
12. **[T/F]** (2pts) In the presence of homoskedasticity, WLS is more efficient that OLS.
13. **[T/F]** (2pts) If a Goldfeld-Quandt test finds that SSE_1 is equal to SSE_2, then it will conclude that there is no statistically significant evidence of heteroskedasticity.

14. **[T/F]** (2pts) If you use heteroskedasticity-based standard errors when the disturbances are actually homoskedastic, then your standard errors are biased.

15. **[T/F]** (2pts) The following regression model allows the effect of *Sunlight* on *Happiness* to vary by the individual's income level.

$$\text{Happiness}_i = \beta_0 + \beta_1 \text{Income}_i + \beta_2 \text{Sunlight}_i + u_i$$

16. **[T/F]** (2pts) In the following regression model, a one-percent increase in *Income* will change *Happiness* by β_1 percent (holding all else constant).

$$\log(\text{Happiness}_i) = \beta_0 + \beta_1 \text{Income}_i + u_i$$

17. **[T/F]** (2pts) Using a linear model rather than a log-log model can lead to heteroskedasticity.

18. **[T/F]** (2pts) Disturbances measure individuals' distances from the population regression line.

19. **[T/F]** (2pts) As the sum of squared error (SSE) increases, the standard error tends to increase.

20. **[T/F]** (2pts) In practice (in the "real world"), you should generally avoid standard errors that assume homoskedasticty.

21. **[T/F]** (2pts) The equation below shows that OLS will be consistent when the disturbance is homoskedastic.

$$\text{plim } \hat{\beta} = \beta + \frac{\text{Cov}(x_i,\ u_i)}{\text{Var}(x_i)}$$

22. **[T/F]** (2pts) Correlated disturbances can cause substantial bias in our standard errors.

23. **[T/F]** (2pts) Measurement error in an explanatory variable tends to bias OLS's estimates of the coefficients toward zero.

24. **[T/F]** (2pts) Disturbances are unobservable.

25. **[T/F]** (2pts) Static time-series models assume variables' effects happen immediately and do not affect future time periods.

Part 2: Short answer (50 points)

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

Infer from sample to population.
Allows us to account for uncertainty in estimates.

27. (5pts) Why do we typically prefer the White test to the Goldfeld-Quandt test?

White makes fewer assumptions and detects more "types" of heteroskedasticity.

28. (5pts) Compare and contrast consistency and unbiasedness.

Both describe how well an estimator performs.
Consistency describes behavior as $n \to \infty$.
Unbiasedness is mean bias = 0 when sample n is drawn repeatedly.

29. (5pts) For the regression model below, we estimate $\hat{\beta}_0 = 15.5$, $\hat{\beta}_1 = 12.3$, and $\beta_2 = -2.1$. Interpret the estimated coefficient for the effect of exercise (hours per week) on health (where health is an index between 1 and 100; 100 is best).

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

1-hour increase in exercise is expected to increase health score by 12.3 units (holding all else constant).

30. (5pts) Suppose income affects health and is correlated with exercise. Will the bias from omitting income cause us to over or under estimate β_1? Explain your answer.

Assume income is positively correlated w/. exercise and improves health. Then our $\hat{\beta}_1$ will overestimate β_1.

31. (5pts) We now are interested in the regression model

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

and the estimates are: $\hat{\beta}_0 = 12.3$, $\hat{\beta}_1 = 5.3$, $\hat{\beta}_2 = 2.6$, and $\hat{\beta}_3 = -1.1$.

What is the return to exercise for females in the sample?

The estimated return to exercise for females is $\hat{\beta}_1 + \hat{\beta}_3 = 4.2$.

32. (5pts) We are interested in the regression model

$$\log(\text{Health}_i) = \beta_0 + \beta_1 \log(\text{Exercise}_i) + \beta_2 \log(\text{Age}_i) + u_i$$

and the estimate for β_2 is -0.33. How do we interpret this coefficient?

It is an elasticity: 1% increase in age tends to reduce health by 0.33% (all else constant).

33. (5pts) Suppose your regression estimates a coefficient of $\hat{\beta} = 17.5$ with a standard error of 0.57. Will this relationship likely be 'statistically significant'? Explain.

Yes. Test statistic testing $\beta = 0$ would be $17.5/0.57$, which is a very large test stat. and would reject a null that $\beta = 0$.

34. (5pts) What is the difference between the residual and the disturbance?

Residual (e): Observation's distance to sample reg. line: $y - (\hat{\beta}_0 + \hat{\beta}_1 x)$

Disturbance (u): Observation's distance to pop. reg. line: $y - (\beta_0 + \beta_1 x)$

35. (5pts) Suppose an estimator is biased, and its bias is equal to $1/n$ (where n is the sample size). Do you think the estimator is consistent? Explain your answer.

As $n \to \infty$, $1/n \to 1/\infty \to 0$.

So the bias $\to 0$ as $n \to \infty$, which means we might expect the estimator to be consistent.

Part 3: Long-ish answer (10 points)

Note: Answer the prompt with 1–2 paragraphs.

36. (10 pts) Your friend recently downloaded data from 136 cities on the level of crime and policing in each of the cities. Using OLS regression, she regressed crime on policing. The results indicate a large, positive, and statistically significant relationship between the two variables.

    Should your friend be 'confident' in her results? Explain why or why not.