ECON 0150 | MiniExam 07 Demo - Solution Guide

Question 1: Regression Diagnostics

1a) Plot A shows a violation of:

Answer: ✓ Homoskedasticity

Explanation:
Plot A shows a fan-shaped pattern where the spread of the residuals increases as the fitted values increase. This pattern indicates that the variance of the errors is not constant across all values of the independent variable(s), which is a violation of the homoskedasticity assumption. The homoskedasticity assumption requires that the variance of the error terms is constant across all levels of the independent variables.

1b) The appropriate fix for the issue in Plot A would be:

Answer: ✓ Use robust standard errors

Explanation:
When heteroskedasticity is present (as in Plot A), one of the best remedies is to use heteroskedasticity-robust standard errors. These adjusted standard errors account for the varying variance in the residuals and provide more accurate estimates of the uncertainty in the coefficient estimates. While transforming the dependent variable could sometimes help, robust standard errors are the most direct and commonly used solution for addressing heteroskedasticity without changing the model structure.

1c) If residuals show increasing spread with larger fitted values, this violates:

Answer: ✓ Homoskedasticity

Explanation:
Increasing spread in residuals as fitted values increase is the classic pattern of heteroskedasticity. The homoskedasticity assumption requires that the variance of errors remains constant across all values of the independent variables. When the spread increases (or decreases) systematically with fitted values, we have heteroskedasticity, which affects the accuracy of standard errors and hypothesis tests, though coefficients remain unbiased.

Question 2: Multiple Regression

2a) Write the estimated regression equation for this model.

Answer: ✓ wage = 5.427 + 1.238 × education + 0.452 × experience

Explanation:
The regression equation is written as $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, where Y is the dependent variable (wage), X_1 is education, X_2 is experience, β_0 is the intercept, and β_1 and β_2 are the coefficients for education and experience respectively. From the regression output, we see that the intercept (β_0) is 5.427, the coefficient for education (β_1) is 1.238, and the coefficient for experience (β_2) is 0.452.

2b) Interpret the coefficient for experience in this model.

Answer: ✓ For each additional year of experience, hourly wage increases by $0.452, holding education constant

Explanation:
In multiple regression, each coefficient represents the expected change in the dependent variable associated with a one-unit increase in the respective independent variable, holding all other variables constant. The coefficient for experience (0.452) indicates that, all else equal (particularly education level), each additional year of work experience is associated with a $0.452 increase in hourly wage. This interpretation highlights the "holding other variables constant" aspect, which is crucial in multiple regression.

Question 3: Regression Model Applications

3a) What regression model would be most appropriate to answer this question if the researcher wants to directly measure the relationship between funding levels and test scores?

Answer: ✓ TestScore = $\beta_0 + \beta_1$ × Funding + ε

Explanation:
To directly measure the relationship between funding levels and test scores, the most appropriate model places test scores as the dependent variable and funding as the independent variable. This allows the researcher to estimate how much test scores change with different levels of funding. The other options either reverse the relationship (option 3), create a binary classification of funding (option 2), or transform the variables into logarithms (option 4), which would measure percentage changes rather than absolute changes.

3b) The researcher fits a model and obtains a statistically significant positive coefficient for funding. After checking the residuals, they notice a curved pattern rather than a random cloud. This indicates:

Answer: ✓ The relationship between funding and test scores is non-linear

Explanation:
A curved pattern in the residuals (rather than a random cloud) suggests that the linear model is not capturing the true relationship between the variables. Specifically, it indicates that the relationship between funding and test scores is non-linear—there may be diminishing returns to additional funding, or there could be a threshold effect where funding has different impacts at different levels. This violation of the linearity assumption suggests the researcher should consider adding non-linear terms (like quadratic terms) or transforming the variables to better capture the relationship.