Quiz 3
MSAN 601

September 14, 2016

Question 1 (2 pts)

If a histogram of residuals is generated, which assumption would one be attempting to verify? What would one expect as the center / mean value of the residuals of the histogram?

Answer

Normality of the error terms. We would expect the mean to be zero.

Question 2 (8 pts)

1. For what purpose would one use a Brown-Forsythe test?
2. What is it good at detecting (be specific).
3. Name a weakness of the Brown-Forsythe test?
4. How does the Brown-Forsythe test differ from the Breush-Pagan test?

Answer

1. To verify the homoscedasticity of the error terms.
2. Megaphone or inverted megaphone-shaped heteroskedasticity.
3. Grouping of residuals is subjective.
4. BP test assumes that the error terms are independent and normally distributed. BF test uses a two-sample t-test to verify if the two groups have equal median whereas BP uses chi-square test if variance of an error term is a function of X.

Question 3 (2 pts)

When generating a residual plot, give **two** different examples of what can be on the x and y axes. E.g., on the x-axis we plot ... and on the y-axis we plot ....

Answer

1. residuals (y) on fitted values ($\hat{y}$)
2. residuals (y) on predictor (x)

Question 4 (3 pts)

What is the difference between a standardized and studentized residual? Which plots—standardized or studentized residual plot—would you prefer to employ to detect outliers? Why?

Answer

A standardized residual is calculated as $\frac{e_i}{\sqrt{MSE}}$ while a studentized residual is calculated as $\frac{e_i}{\sqrt{MSE_{(-i)}}}$. Thus, a studentized residual measures the residual of the data point when its influence has been removed.

If the point is an outlier, MSE will decrease considerably making the residual more prominent. Thus, I would use studentized residual method to detect outliers.

Question 5 (2 pts)

Draw a sketch of what a QQ-plot would look like if the data were left-skewed.

Answer

See SLR slides.

Question 6 (3 pts)

How do outliers affect the regression line? What are some of the cases when you can ignore them?

Answer

Outliers can alter the regression line significantly by affecting its slope, for example. Outliers can be ignored when they are a result of an error in recording, reporting, a miscalculation or equipment malfunctioning.

Question 7 (4 pts)

1. What is the purpose of transforming a variable?
2. When trying to correct for heteroskedasticity **only** in an SLR model, which variable do you transform? What issues can you run into in making such a transformation, and how would you correct for them?

Answer

1. Transformation helps when SLR model doesn't fit the given data or the residuals violate an assumption.
2. Transform the response variable first. This may change the linear relation into a curvilinear one. Transforming the predictor will help resolve this.

Question 8 (4 pts)

Given the model

$$\log(y) = 2.19 + 0.17 \log(x)$$

if x is decreases by 1%, y would be expected to change by how much?

Answer

$y = e^{2.19+0.17log(x)}$
$y = e^{2.19}x^{0.17}$
x is decreased by 1%, $\therefore y_1 = e^{2.19}(0.99)x^{0.17}$

Thus, $y_1 = 0.99829y$, i.e., y decreases by $(1 - 0.99^{0.17}) \sim 0.17\%$, therefore $\sim 2\%$

Question 9 (2 pts)

Given a compelling argument both for and against using linear regression as a statistical learning tool.

Answer

Linear regression models are very simple to construct and interpret. However, they are restrictive in their assumptions, and may not be the most accurate when predicting in comparison to other machine learning techniques.

Question 10 (3 pts)

Given an MLR model with $n = 492$ and eight predictor variables, and $SSE = 4,000$, compute MSE.

Asnwer

MSE $= \frac{4000}{492-9} = 8.28$