Quiz 1
MSAN 601

September 01, 2016

Question 1 **(2 pts)**

In the SLR model, the probability distribution of Y (i.e., Y_i) has the same mean and variance for all levels of X (i.e., X_i). True or False? Explain.

Answer

False. The variance remains the same, however the mean depends on the level of X.

Question 2 **(2 pts)**

The number of points above the fitted regression line is always equal to the number of points below it. True or False? Explain.

Answer

False. For the sum of residuals to be zero, there need not be equal number of points on either side of the regression line, but there has to be at least one.

Question 3 **(1 pt)**

In SLR, what does β_1 measure?

Answer

In SLR, β_1 is the slope of the regression line. It measures the average change in Y that the model predicts for a unit change in X.

Question 4 **(4 pts)**

In the context of an SLR model, prove the following:

1. $E[Y_i] = \beta_0 + \beta_1 X_i$

   $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
   $\therefore E[Y_i] = E[\beta_0 + \beta_1 X_i + \epsilon_i]$
   $\therefore E[Y_i] = E[\beta_0] + E[\beta_1 X_i] + E[\epsilon_i]$
   $\therefore E[Y_i] = E[\beta_0] + E[\beta_1 X_i] + E[\epsilon_i]$
   $\therefore E[Y_i] = \beta_0 + \beta_1 X_i + 0$
   $\because \beta_0, \beta_1, X_i$ are constants, $E[constant] = constant$ and $E[\epsilon_i] = 0$ (SLR model assumption)
   $\therefore E[Y_i] = \beta_0 + \beta_1 X_i$

2. $V(Y_i) = constant \ \forall \ i$

   $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
   $\therefore V[Y_i] = V[\beta_0 + \beta_1 X_i + \epsilon_i]$
   $\therefore V[Y_i] = V[\beta_0] + V[\beta_1 X_i] + V[\epsilon_i]$

$\therefore V[Y_i] = 0 + 0 + V[\epsilon_i]]$
$\because \beta_0, \beta_1, X_i$ are constants, $V[constant] = 0$ and $V[\epsilon_i] = \sigma^2 = constant$ (SLR model assumption)
$\therefore V[Y_i] = constant$

Question 5 **(2 pts)**

For the SLR model, $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, how many random variables are there. Explain.

Answer

For the SLR model, there are two random variables Y_i and ϵ_i. β_0, β_1 *and* X_i have fixed values whereas the values ϵ_i and Y_i take depend on their distributions. $\epsilon_i \sim N(0, \sigma^2)$ whereas the for the distribution of Y, mean depends on the level of X and variance is equal to that of ϵ_i.

Question 6 **(4 pts)**

Write out the normal error regression model and its assumptions (in English and math).

Answer

Model: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

Assumptions:

1. $E[\epsilon_i|X_i] = 0$, the expected value of the error terms is zero (exogeneity).
2. $\sigma^2[\epsilon_i|X_i] = \sigma^2$ $(constant)$, the variance of the error terms is constant (homoscedasticity).
3. $\sigma[\epsilon_i, \epsilon_j|X_i] = 0 \forall i \neq j$, error terms are independent and identically distributed and follow the normal distribution $(N \sim (0, \sigma^2))$.
4. $\sigma[X_i, \epsilon_i] = 0$, there is no correlation between the error terms and predictors.

Question 7 **(3 pts)**

Write in English and mathematically (using correct symbols) how a residual is computed. Given the regression line $\hat{Y} = 3 + 4.5X$, compute the residual for $(2, 11)$.

Answer

A residual is the difference between the actual and fitted value of the response variable.
$e_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X$
$e_i = 11 - (3 + 4.5(2)) = -1$

Question 8 **(1)**

What does a negative value of β_1 indicate about the relation between X and Y?

Answer

It indicates that X and Y are inversely proportional or there is a negative correlation between the two.

Question 9 **(6 pts)**

Prove that $\bar{X}$ and $\bar{Y}$ always lie on the regression line.

Answer: SLR slides 32 - 36.