Question 1 of 34

In the simple linear equation, the independent variable:

A. is random.
B. is nonrandom.
C. can be random or nonrandom.

Correct Answer: B

Explanation:

In the simple linear equation, the ***independent variable, X, is nonrandom.***

SIMPLE LINEAR REGRESSION

Regression analysis:

- is a tool used to examine whether a variable is useful to explain another variable.
- predicts the value of a dependent variable based on the value of at least one independent variable.

Simple Linear Regression (SLR)

A regression that summarizes the relation between the dependent variable and one independent variable through estimation of a linear relationship.

If more than one variable is used, it is called ***Multiple Regression***:

Dependent variable (a.k.a. explained variable) Y

The variable whose variation is being explained by the independent variable.

Independent variable (a.k.a. explanatory variable) X

The variable used to explain the dependent variable.

Question 2 of 34

The purpose of simple linear regression is to explain the variation in dependent variable in terms of the variation in a single independent variable. Here variation is referred to as:

A. sample variance.
B. population variance.
C. the degree to which a variable differs from its mean.

Correct Answer: C

Explanation:

The purpose of simple linear regression is to explain the variation in dependent variable in terms of the variation in a single independent variable. ***Here variation is not the same as variance. It is the degree to which a variable differs from its mean.***

SIMPLE LINEAR REGRESSION

Regression analysis:

- is a tool used to examine whether a variable is useful to explain another variable.
- predicts the value of a dependent variable based on the value of at least one independent variable.

Simple Linear Regression (SLR)

A regression that summarizes the relation between the dependent variable and one independent variable through estimation of a linear relationship.

If more than one variable is used, it is called ***Multiple Regression*:**

Dependent variable (a.k.a. explained variable) Y

The variable whose variation is being explained by the independent variable.

Independent variable (a.k.a. explanatory variable) X

The variable used to explain the dependent variable.

Question 3 of 34

The simple linear equation has:

A. one regression coefficient, the intercept, denoted as b_1.
B. one regression coefficient, the slope coefficient, denoted as b_1.
C. two regression coefficients: b_0 (intercept) and b_1 (the slope coefficient).

Correct Answer: C

Explanation:

The simple linear equation is as follows:

$$\boldsymbol{Y_i = b_0 + b_1 X_i + \varepsilon_i}$$

It has two regression coefficients: b_0(intercept) and b_1(the slope coefficient).

Interpreting the Regression Coefficients

How to calculate and interpret regression coefficient b_1 and b_0

1) ***Slope Coefficient (b_1):*** A change in the dependent variable for a one unit change in the independent variable.

Interpretation:
If slope is positive (negative), the change in the dependent and independent variable will be in the same(opposite) direction.

$$b_1 = \frac{cov(x,y)}{var(x)} = \frac{\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{n-1}}{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}} = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

2) ***Intercept (b_0):*** The predicted value of the dependent variable when the independent variable is set to zero.

$$b_0 = \bar{Y} - \hat{b}_1\bar{X}$$

where $\bar{Y}$ *and* $\bar{X}$ are mean values of X and Y.

Interpretation:

In some cases, zero intercept is meaningful but in some cases it does not make sense.

For example, if independent variable is:

- **money supply**, 0 intercept is meaningless becase zero money supply is not possible.
- **money supply growth**, 0 intercept is meaningful because zero money supply growth is possible.

According to the simple linear equation, the dependent variable, Y, is equal to the intercept:

A. times the slope coefficient plus the independent variable plus the error term.
B. plus the slope coefficient times the independent variable minus the error term.
C. plus the slope coefficient times the independent variable plus the disturbance term.

Correct Answer: C

Explanation:

In simple linear equation, the dependent variable, Y, is equal to the intercept, b_0, plus the slope coefficient, b_1, times the independent variable, X, plus the disturbance term, ε.

Interpreting the Regression Coefficients

How to calculate and interpret regression coefficient b_1 and b_0

1) ***Slope Coefficient (b_1):*** A change in the dependent variable for a one unit change in the independent variable.

Interpretation:
If slope is positive (negative), the change in the dependent and independent variable will be in the same(opposite) direction.

$$b_1 = \frac{cov(x,y)}{var(x)} = \frac{\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{n-1}}{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}} = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

2) ***Intercept (b_0):*** The predicted value of the dependent variable when the independent variable is set to zero.

$b_0 = \bar{Y} - \hat{b}_1\bar{X}$

where $\bar{Y}$ *and* $\bar{X}$ are mean values of X and Y.

Interpretation:
In some cases, zero intercept is meaningful but in some cases it does not make

sense.

For example, if independent variable is:

- **money supply**, 0 intercept is meaningless becase zero money supply is not possible.
- **money supply growth**, 0 intercept is meaningful because zero money supply growth is possible.

Question 5 of 34

Which of the following statements *best* characterizes the time-series data? Time-series data involves the observations of X and Y for:

A. the same company or asset class from different time periods.
B. the same company or asset class from the same time period.
C. different companies or asset classes from the same time period.

Correct Answer: A

Explanation:

Regression analysis can be used for both cross-sectional and time-series data.

Time-series data involves the observations of X and Y for the ***same*** company, asset class etc., from ***different*** time periods.

Cross-sectional data involves the observations of X and Y for ***different*** companies, asset classes etc., from the ***same*** time period.

Cross-Sectional vs. Time-Series Regressions

Two main types of data used in regression analysis are time series and cross-sectional:

Time-series:

It uses many observations from ***different time periods*** for the ***same*** company, asset class or country etc.

Cross-sectional

It uses many observations for the ***same time period***of ***different*** companies, asset classes or countries etc.

Panel Data: Mix of two types

It is a mix of time-series and cross-sectional data.

Question 6 of 34

Regression analysis can be used for:

A. time-series data only.
B. cross-sectional data only.
C. both cross-sectional and time-series data.

Correct Answer: C

Explanation:

Regression analysis can be used for ***both cross-sectional and time-series data.***

Cross-Sectional vs. Time-Series Regressions

Two main types of data used in regression analysis are time series and cross-sectional:

Time-series:

It uses many observations from ***different time periods*** for the ***same*** company, asset class or country etc.

Cross-sectional

It uses many observations for the ***same time period***of ***different*** companies, asset classes or countries etc.

Panel Data: Mix of two types

It is a mix of time-series and cross-sectional data.

Question 7 of 34

Cross-sectional data involves the observations of X and Y:

A. from the same company, asset class or country from the same time period.
B. for different companies, asset classes or countries from the same time period.
C. for different companies, asset classes or countries from different time periods.

Correct Answer: B

Explanation:

Regression analysis can be used for both cross-sectional and time-series data.

Cross-sectional data involves the observations of X and Y for ***different*** companies, asset classes etc., from the ***same*** time period.

Cross-Sectional vs. Time-Series Regressions

Two main types of data used in regression analysis are time series and cross-sectional:

Time-series:

It uses many observations from ***different time periods*** for the ***same*** company, asset class or country etc.

Cross-sectional

It uses many observations for the ***same time period***of ***different*** companies, asset classes or countries etc.

Panel Data: Mix of two types

It is a mix of time-series and cross-sectional data.

Question 8 of 34

In simple linear regression, the intercept $\hat{b}_0$ is calculated as:

A. $\hat{b}_1\bar{X} - \bar{Y}$.
B. $\bar{Y} - \hat{b}_1\bar{X}$.
C. $\text{Cov}\,(X,Y)/\sigma^2(X)$.

Correct Answer: B

Explanation:

In simple linear regression, the intercept $\hat{b}_0$ is calculated as $\boldsymbol{\bar{Y} - \hat{b}_1\bar{X}}$.

Interpreting the Regression Coefficients

How to calculate and interpret regression coefficient b_1 and b_0

1) ***Slope Coefficient (b_1):*** A change in the dependent variable for a one unit change in the independent variable.

Interpretation:
If slope is positive (negative), the change in the dependent and independent variable will be in the same(opposite) direction.

$$b_1 = \frac{cov(x,y)}{var(x)} = \frac{\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{n-1}}{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}} = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

2) ***Intercept (b_0):*** The predicted value of the dependent variable when the independent variable is set to zero.

$b_0 = \bar{Y} - \hat{b}_1\bar{X}$

where $\bar{Y}$ *and* $\bar{X}$ are mean values of X and Y.

Interpretation:
In some cases, zero intercept is meaningful but in some cases it does not make sense.

For example, if independent variable is:

- **money supply**, 0 intercept is meaningless becase zero money supply is not possible.
- **money supply growth**, 0 intercept is meaningful because zero money supply growth is possible.

Question 9 of 34

In simple linear regression, the estimated slope coefficient $\hat{b}_1$ is calculated as:

A. $\sigma^2(X)/\text{Cov}\,(X,Y)$.
B. $\text{Cov}\,(X,Y)/\sigma^2(X)$.
C. $\text{Cov}\,(X,Y)/\sigma(X)\sigma(Y)$.

Correct Answer: B

Explanation:

In simple linear regression, the estimated slope coefficient $\hat{b}_1$ describes the change in one unit of Y for one unit of X. **It is calculated as $\text{Cov}\,(X,Y)/\sigma^2(X)$.**

Interpreting the Regression Coefficients

How to calculate and interpret regression coefficient b_1 and b_0

1) ***Slope Coefficient (b_1):*** A change in the dependent variable for a one unit change in the independent variable.

Interpretation:
If slope is positive (negative), the change in the dependent and independent variable will be in the same(opposite) direction.

$$b_1 = \frac{cov(x,y)}{var(x)} = \frac{\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{n-1}}{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}} = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

2) ***Intercept (b_0):*** The predicted value of the dependent variable when the independent variable is set to zero.

$b_0 = \bar{Y} - \hat{b}_1\bar{X}$

where $\bar{Y}$ *and* $\bar{X}$ are mean values of X and Y.

Interpretation:
In some cases, zero intercept is meaningful but in some cases it does not make

sense.

For example, if independent variable is:

- **money supply**, 0 intercept is meaningless becase zero money supply is not possible.
- **money supply growth**, 0 intercept is meaningful because zero money supply growth is possible.

Question 10 of 34

If the intercept is –3%, what does it *most likely* indicate?

A. The dependent variable will decrease by 3 percent for each one-unit change in the intercept.
B. When the independent variable is zero, the dependent variable will increase by 3 percent.
C. When the independent variable is zero, the dependent variable will decrease by 3 percent.

Correct Answer: C

Explanation:

The intercept, $\hat{b}_0$, is the regression line's intercept with the Y-axis at $X = 0$. If the intercept is –3%, it indicates that ***when the independent variable is zero, the dependent variable will decrease by 3 percent.***

Interpreting the Regression Coefficients

How to calculate and interpret regression coefficient b_1 and b_0

1) ***Slope Coefficient (b_1):*** A change in the dependent variable for a one unit change in the independent variable.

Interpretation:

If slope is positive (negative), the change in the dependent and independent variable will be in the same(opposite) direction.

$$b_1 = \frac{cov(x,y)}{var(x)} = \frac{\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{n-1}}{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}} = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

2) ***Intercept (b_0):*** The predicted value of the dependent variable when the independent variable is set to zero.

$$b_0 = \bar{Y} - \hat{b}_1\bar{X}$$

where $\bar{Y}$ *and* $\bar{X}$ are mean values of X and Y.

Interpretation:

In some cases, zero intercept is meaningful but in some cases it does not make sense.

For example, if independent variable is:

- **money supply**, 0 intercept is meaningless becase zero money supply is not possible.
- **money supply growth**, 0 intercept is meaningful because zero money supply growth is possible.

Consider the following statements made by an analyst:

Statement A: "The standard error of estimate gauges the fit of the regression line."

Statement B: "The larger the standard error, the better the fit."

Which of the following *best* characterizes the analyst's statements?

A. Only Statement A is correct.
B. Both statements are correct.
C. Both statements are incorrect.

Correct Answer: A

Explanation:

Standard error of estimate (SEE) measures the vulnerability of the actual Y values relative to the estimated Y values from the regression line. It is the standard deviation of the error terms in the regression.

The standard error of estimate gauges the fit of the regression line; the ***smaller*** the standard error, the better the fit.

Measures of Goodness of Fit

Goodness of fit (i.e., how well the regression model fits the data) can be measured using several methods such as:

1. Coefficient of determination R^2
2. F-statistic
3. Standard error of regression S_e

The Standard Error of Estimate S_e

- Standard Error of Estimate (S_e) measures the degree of variability of the actual y-values relative to the predicted y-values from a regression equation.
- Smaller the S_E, better the fit.
- S_E is also called standard error of regression or root mean square error.
- $Standard\ Error\ of\ Estimate: S_E = \sqrt{MSE} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{\Sigma(y_i-\hat{y})^2}{n-k-1}}$

where n = Sample size
k = no. of independent variables = 1.

Example:
n = 100
SSE = 2,252,363
Thus,

$$S_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{2{,}252{,}363}{98}} = 151.60$$

Note

- Co-efficient of determination and F-statistic are relative measures of fit
- Standard error of the estimate is absolute measure.

Question 12 of 34

Standard error of estimate (SEE) is the:

A. variance of error terms in the regression.
B. mean deviation of error terms in the regression.
C. standard deviation of error terms in the regression.

Correct Answer: C
Explanation:

Standard error of estimate (SEE) measures the vulnerability of the actual Y values relative to the estimated Y values from the regression line. It is the standard deviation of the error terms in the regression.

The standard error of estimate gauges the fit of the regression line; the smaller the standard error, the better the fit.

It is the ***standard deviation*** of error terms in the regression.

Measures of Goodness of Fit

Goodness of fit (i.e., how well the regression model fits the data) can be measure using several methods such as:

1. Coefficient of determination R^2
2. F-statistic
3. Standard error of regression S_e

The Standard Error of Estimate S_e

- Standard Error of Estimate (S_e) measures the degree of variability of the actual y-values relative to the predicted y-values from a regression equation.
- Smaller the S_E, better the fit.
- S_E is also called standard error of regression or root mean square error.
- $Standard\ Error\ of\ Estimate: S_E = \sqrt{MSE} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{\Sigma(y_i-\hat{y})^2}{n-k-1}}$

where n = Sample size
k = no. of independent variables = 1.

Example:
n = 100
SSE = 2,252,363
Thus,

$$S_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{2,252,363}{98}} = 151.60$$

Note

- Co-efficient of determination and F-statistic are relative measures of fit
- Standard error of the estimate is absolute measure.

Standard error of estimate is:

A. low if the relation between dependent and independent variables is weak, and it is high if the relation is strong.
B. low if the relation between dependent and independent variables is strong, and it is high if the relation is weak.
C. high if the relation between dependent variable and slope coefficient is weak, and it is low if the relation is strong.

Correct Answer: B

Explanation:

Standard error of estimate is:

- low if the relation between dependent and independent variables is strong
- high if the relation between dependent and independent variables is weak

Standard error of estimate (SEE) measures the vulnerability of the actual Y values relative to the estimated Y values from the regression line. It is the standard deviation of the error terms in the regression.

The standard error of estimate gauges the fit of the regression line; the smaller the standard error, the better the fit.

It is the standard deviation of error terms in the regression. It is also called standard error of residual or standard error of the regression.

Measures of Goodness of Fit

Goodness of fit (i.e., how well the regression model fits the data) can be measured using several methods such as:

1. Coefficient of determination R^2
2. F-statistic
3. Standard error of regression S_e

The Standard Error of Estimate S_e

- Standard Error of Estimate (S_e) measures the degree of variability of the actual y-values relative to the predicted y-values from a regression equation.
- Smaller the S_E, better the fit.
- S_E is also called standard error of regression or root mean square error.
- *Standard Error of Estimate*: $S_E = \sqrt{MSE} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{\sum(y_i-\hat{y})^2}{n-k-1}}$

where n = Sample size
k = no. of independent variables = 1.

Example:
n = 100
SSE = 2,252,363
Thus,

$$S_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{2,252,363}{98}} = 151.60$$

Note

- Co-efficient of determination and F-statistic are relative measures of fit
- Standard error of the estimate is absolute measure.

The percentage of total variation in the dependent variable explained by the independent variable is called:

A. slope coefficient.
B. correlation coefficient.
C. coefficient of determination.

Correct Answer: C

Explanation:

The percentage of total variation in the dependent variable explained by the independent variable is called ***coefficient of determination*** **(R^2).**

Measures of Goodness of Fit

Goodness of fit (i.e., how well the regression model fits the data) can be measured using several methods such as:

1. Coefficient of determination R^2
2. F-statistic
3. Standard error of regression S_e

Coefficient of Determination R^2

- The coefficient of determination is the percentage of the total variation in the dependent variable that is explained by the independent variable.
- The coefficient of determination is also called R-squared and is denoted as R^2.
- It is descriptive measure.

$$Coefficient\ of\ determination\ (R^2) = \frac{Explained\ Variation (SSR)}{Total\ Variation\ (SST)}$$

$$= \frac{Total\ Variation\ (SST) - Unexplained\ Variation\ (SSE)}{Total\ Variation\ (SST)}$$

where, $0 \leq R^2 \leq 1$

In case of a single independent variable, the coefficient of determination is: $R^2 = r^2$

where,
R^2 = Coefficient of determination
r = Simple correlation coefficient

Example: Suppose coefficient of determination between returns of two assets is 0.64. This means that approximately 64 percent of the variability in the returns of one asset (or dependent variable) can be explained by the returns of the other asset (or indepepnent variable).

Question 15 of 34

In simple linear regression, what does coefficient of determination, R^2, of 55% mean?

A. 45% variation in the dependent variable is explained by the independent variable.
B. 55% variation in the dependent variable is explained by the independent variable.
C. 55% variation in the independent variable is explained by the dependent variable.

Correct Answer: B

Explanation:

The percentage of total variation in the dependent variable explained by the independent variable is called coefficient of determination (R^2).

In simple linear regression, coefficient of determination, R^2, of 55% means that 55% variation in the dependent variable is explained by the independent variable.

Measures of Goodness of Fit

Goodness of fit (i.e., how well the regression model fits the data) can be measured using several methods such as:

1. Coefficient of determination R^2
2. F-statistic
3. Standard error of regression S_e

Coefficient of Determination R^2

- The coefficient of determination is the percentage of the total variation in the dependent variable that is explained by the independent variable.
- The coefficient of determination is also called R-squared and is denoted as R^2.
- It is descriptive measure.

$$Coefficient\ of\ determination\ (R^2) = \frac{Explained\ Variation(SSR)}{Total\ Variation\ (SST)}$$

$$= \frac{Total\ Variation\ (SST) - Unexplained\ Variation\ (SSE)}{Total\ Variation\ (SST)}$$

where, $0 \leq R^2 \leq 1$

In case of a single independent variable, the coefficient of determination is: $R^2 = r^2$

where,
R^2 = Coefficient of determination
r = Simple correlation coefficient

Example: Suppose coefficient of determination between returns of two assets is 0.64. This means that approximately 64 percent of the variability in the returns of one asset (or dependent variable) can be explained by the returns of the other asset (or indepepnent variable).

Question 16 of 34

In simple linear regression, if coefficient of determination, R^2, is 0.85, then correlation coefficient, r, is *closest* to:

A. 0.82
B. 0.85
C. 0.92

Correct Answer: C

Explanation:

In simple linear regression, coefficient of determination $R^2 = r^2$.

Replacing R^2 with its given value in the above relation, we get:
$0.85 = r^2$
$r = 0.922$

Coefficient of Determination R^2

- The coefficient of determination is the percentage of the total variation in the dependent variable that is explained by the independent variable.
- The coefficient of determination is also called R-squared and is denoted as R^2.
- It is descriptive measure.

$$Coefficient\ of\ determination\ (R^2) = \frac{Explained\ Variation(SSR)}{Total\ Variation\ (SST)}$$

$$= \frac{Total\ Variation\ (SST) - Unexplained\ Variation\ (SSE)}{Total\ Variation\ (SST)}$$

where, $0 \le R^2 \le 1$

In case of a single independent variable, the coefficient of determination is: $R^2 = r^2$

where,
R^2 = Coefficient of determination
r = Simple correlation coefficient

Example: Suppose coefficient of determination between returns of two assets is 0.64. This means that approximately 64 percent of the variability in the returns of one asset (or dependent variable) can be explained by the returns of the other asset (or indepepnent variable).

Question 17 of 34

Linearity assumption of the simple linear regression model implies that:

A. independent variable must not be random.
B. variance of residuals is the same for all observations.
C. dependent or independent variables must be normally distributed.

Correct Answer: A

Explanation:

A is correct. Linearity assumption implies that independent variable must not be random (i.e., non-stochastic). Otherwise, there will be no linear relationship.

B is incorrect. According to homoscedasticity assumption, 'The variance of residuals is the same for all observations. It is known as Homoskedasticity (same scatter) assumption.'

C is incorrect. No assumption states that dependent or independent variables must be normally distributed.

ASSUMPTIONS OF THE SIMPLE LINEAR REGRESSION MODEL

The four key assumptions of the simple linear regression model are:

1) Linearity
2) Homoskedasticity
3) Independence
4) Normality

Assumption 1: Linearity

'Relation between the dependent variable and independent variable is linear.'

- If the relationship is nonlinear, the model will be biased (i.e., over or underestimate the dependent variable).
- Linearity assumption also implies that independent variable must not be random (i.e., non-stochastic). Otherwise, there will be no linear relationship.

Assumption 2: Homoskedasticity

'The variance of residuals is the same for all observations. It is known as Homoskedasticity (same scatter) assumption.'

- If the variance of the residuals differs across observations, this state is called heteroskedasticity (different scatter).
- In real-world data, structural changes (regime changes) often involve heteroskedasticity.

Assumption 3: Independence

'The observations (pairs of Xs and Ys) are independent of each other, which implies the residuals are uncorrelated across observations.'

- If variables are not independent, the residuals will be correlated (display a pattern). This is an indication of autocorrelation.

Assumption 4: Normality

'The regression residuals must be normally distributed.'

- In large sample sizes, dropping the normality assumption does not noticeably influence results.

Note: Normality assumption does not mean that dependent or independent variables must be normally distributed.

The F-statistic tests if the:

A. intercept in simple linear regression is equal to 0.
B. slope coefficient in simple linear regression is equal to 0.
C. slope coefficient and intercept in simple linear regression is equal to 0.

Correct Answer: B

Explanation:

In regression analysis, (ANOVA) determines the usefulness of the independent variable in explaining the variation in the dependent variable.

An important statistical test conducted in the analysis of variance is the F-test.

The F-statistic tests if the slope coefficient (b_1) in simple linear regression is equal to 0.

Measures of Goodness of Fit

Goodness of fit (i.e., how well the regression model fits the data) can be measured using several methods such as:

1. Coefficient of determination R^2
2. F-statistic
3. Standard error of regression S_e

F-distributed test Statistic or F-Test

F-distributed test statistic tests whether the slopes b_1 in regression are equal to zero, against the alternative hypothesis that at least one slope is not equal to zero.

H_0: $b_1 = 0$
H_1: $b_1 \neq 0$

The F statistic is calculated as the ratio of mean square regression (MSR) to mean squared errors (MSE).

$$F = \frac{MSR}{MSE} = \frac{\left(\frac{RSS}{k}\right)}{\left(\frac{SSE}{n-k-1}\right)} = \frac{MSR}{MSE}$$

$$MSR = \frac{SSR}{k} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{k} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

where df numerator = k = 1

$$MSE = \frac{SSE}{n-k-1} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}$$

where df denominator = n – k – 1 = n – 2 (in simple linear regression)

Note: F-test is always a one-tailed test (one sided).

Question 19 of 34

The F-statistic is a:

A. ratio of regression sum of squares to sum of squared errors.
B. ratio of average sum of squared errors to average regression sum of squares.
C. ratio of average regression sum of squares to average sum of squared errors.

Correct Answer: C

Explanation:

The F-statistic is a ratio of ***average regression sum of squares to average sum of squared errors.***

$$F = \frac{\text{RSS}/1}{\text{SSE}/(n-2)}$$

Measures of Goodness of Fit

Goodness of fit (i.e., how well the regression model fits the data) can be measured using several methods such as:

1. Coefficient of determination R^2
2. F-statistic
3. Standard error of regression S_e

F-distributed test Statistic or F-Test

F-distributed test statistic tests whether the slopes b_1 in regression are equal to zero, against the alternative hypothesis that at least one slope is not equal to zero.

H_0: $b_1 = 0$
H_1: $b_1 \neq 0$

The F statistic is calculated as the ratio of mean square regression (MSR) to mean squared errors (MSE).

$$F = \frac{MSR}{MSE} = \frac{(\frac{RSS}{k})}{(\frac{SSE}{n-k-1})} = \frac{MSR}{MSE}$$

$$MSR = \frac{SSR}{k} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{k} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

where df numerator = k = 1

$$MSE = \frac{SSE}{n-k-1} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}$$

where df denominator = n – k – 1 = n – 2 (in simple linear regression)

Note: F-test is always a one-tailed test (one sided).

Question 20 of 34

If the independent variable in the regression model explains none of the variation in dependent variable, then:

A. F-statistic will be 0.
B. F-statistic will be 1.
C. F-statistic will be ∞.

Correct Answer: A

Explanation:

If the independent variable in the regression model explains none of the variation in dependent variable, then:

$$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = 0$$

Then the F-statistic (given as follows) will also be 0

$$F = \frac{\text{RSS}/1}{\text{SSE}/(n-2)}$$

where the regression sum of squares RSS is calculated as:

$$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

F-distributed test Statistic or F-Test

F-distributed test statistic tests whether the slopes b_i in regression are equal to zero, against the alternative hypothesis that at least one slope is not equal to zero.

H_0: $b_i = 0$
H_1: $b_i \neq 0$

The F statistic is calculated as the ratio of mean square regression (MSR) to mean squared errors (MSE).

$$F = \frac{MSR}{MSE} = \frac{(\frac{RSS}{K})}{(\frac{SSE}{n-k-1})} = \frac{MSR}{MSE}$$

$$MSR = \frac{SSR}{k} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{k} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

where df numerator = k = 1

$$MSE = \frac{SSE}{n-k-1} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}$$

where df denominator = n – k – 1 = n – 2 (in simple linear regression)

Note: F-test is always a one-tailed test (one sided).

Question 21 of 34

Consider the following hypothetical data:

$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 363$ $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = 289$

Total number of observations is 35 and it is a simple regression.

The standard deviation of error (SEE) is *closest* to:

A. 1.49.
B. 3.32.
C. 2.24.

Correct Answer: A

Explanation:

$$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 363$$

$$SSR = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = 289$$

Sum of Squared Errors = SST – SSR
SSE = 363 – 289 = 74

Standard Error of Estimate = $(SSE/ n - 2)^{1/2}$
= $(74/33)^{1/2}$
= 1.4974

ANOVA and Standard Error of Estimate in Simple Linear Regression

Analysis of Variance (ANOVA) is a statistical method used to divide the total variance in a study into meaningful pieces that correspond to different sources.

Analysis of Variance Table for Simple Linear Regression				
ANOVA	df	SS	MS	F
Regression	1	$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/(n-2)}$
Error	n-2	$SSE = \sum_{i=1}^{n}(y_i - \hat{y})^2$	$MSE = \frac{SSE}{n-2}$	
Total	n–1	$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$		

Or

Source of Variability	df	Sum of Squares	Mean Sum of Squares
Regression (Explained)	1	SSR	MSR = RSS/1
Error (Unexplained)	n-2	SSE	MSE = SSE/n-2
Total	n-1	SST=SSR + SSE	

Question 22 of 34

Kane Robertson is a quantitative analyst at Pelican Investment Bank in London. He develops and enhances the pricing/risk management models for the Forex trading desk. He stated the following during a discussion with the new interns at the firm:

Statement 1:	"Analysis of variance (ANOVA) is a statistical procedure for dividing the total variation of a variable into components that can be attributed to a single source."
Statement 2:	"In regression analysis, (ANOVA) determines the usefulness of the dependent variable in explaining the variation in the independent variable and vice versa."

Which of the following *best* characterizes Kane's statements?

A. Both statements are incorrect.
B. Statement 1 is correct, but Statement 2 is incorrect.
C. Statement 1 is incorrect, but Statement 2 is correct.

Correct Answer: A
Explanation:

Kane's both statements are incorrect.

Analysis of variance (ANOVA) is a statistical procedure for dividing the total variation of a variable into components that can be attributed to ***different*** sources.

In regression analysis, (ANOVA) determines the usefulness of the ***independent*** variable in explaining the variation in the ***dependent*** variable.

ANOVA and Standard Error of Estimate in Simple Linear Regression

Analysis of Variance (ANOVA) is a statistical method used to divide the total variance in a study into meaningful pieces that correspond to different sources.

Analysis of Variance Table for Simple Linear Regression				
ANOVA	df	SS	MS	F
Regression	1	$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/(n-2)}$
Error	n-2	$SSE = \sum_{i=1}^{n}(y_i - \hat{y})^2$	$MSE = \frac{SSE}{n-2}$	
Total	n–1	$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$		

Or

Source of Variability	df	Sum of Squares	Mean Sum of Squares
Regression (Explained)	1	SSR	MSR = RSS/1
Error (Unexplained)	n-2	SSE	MSE = SSE/n-2
Total	n-1	SST=SSR + SSE	

Question 23 of 34

The value that gives the amount of total variation in Y that is explained in regression equation is:

A. sum of squares total, SST.
B. sum of squared errors, SSE.
C. sum of squares regression, RSS.

Correct Answer: C
Explanation:

The sum of squares regression is calculated as:

$$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

This value is the amount of total variation in Y that is explained in regression equation.

ANOVA and Standard Error of Estimate in Simple Linear Regression

Analysis of Variance (ANOVA) is a statistical method used to divide the total variance in a study into meaningful pieces that correspond to different sources.

Analysis of Variance Table for Simple Linear Regression

ANOVA	df	SS	MS	F
Regression	1	$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/(n-2)}$
Error	n-2	$SSE = \sum_{i=1}^{n}(y_i - \hat{y})^2$	$MSE = \frac{SSE}{n-2}$	
Total	n–1	$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$		

Or

Source of Variability	df	Sum of Squares	Mean Sum of Squares
Regression (Explained)	1	SSR	MSR = SSR/1
Error (Unexplained)	n-2	SSE	MSE = SSE/n-2
Total	n-1	SST=SSR + SSE	

Question 24 of 34

The degrees of freedom used for sum of squares total (SST) in simple regression is *most likely* given as:

A. n
B. n – 1.
C. n – 2.

Correct Answer: B

Explanation:

Total sum of squares (SST) = Explained variation + Unexplained variation

Total degrees of freedom = RSS df + SSE df
$= k + (n - k - 1)$
$= 1 + (n - 1 - 1)$
$= 1 + (n - 2)$
$= n - 1$

where k = number of slope coefficients = 1 (in case of one independent variable)

ANOVA and Standard Error of Estimate in Simple Linear Regression

Analysis of Variance (ANOVA) is a statistical method used to divide the total variance in a study into meaningful pieces that correspond to different sources.

Analysis of Variance Table for Simple Linear Regression				
ANOVA	**df**	**SS**	**MS**	**F**
Regression	1	$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/(n-2)}$
Error	n-2	$SSE = \sum_{i=1}^{n}(y_i - \hat{y})^2$	$MSE = \frac{SSE}{n-2}$	
Total	n–1	$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$		

Or

Source of Variability	df	Sum of Squares	Mean Sum of Squares
Regression (Explained)	1	SSR	MSR = SSR/1
Error (Unexplained)	n-2	SSE	MSE = SSE/n-2
Total	n-1	SST=SSR + SSE	

Question 25 of 34

Which of the following is *most likely* correct considering hypothesis testing for a regression coefficient? If the confidence interval at the desired level of significance:

A. includes zero, the null hypothesis is rejected, and the coefficient is said to be statistically different from zero.
B. does not include zero, the alternative hypothesis is rejected, and the coefficient is said to be statistically different from zero.
C. does not include zero, the null hypothesis is rejected, and the coefficient is said to be statistically different from zero.

Correct Answer: C
Explanation:

In hypothesis testing for a regression coefficient, it is tested if the estimated slope coefficient is statistically different from zero.

$H_0 = \text{null hypothesis} = \quad b_1 = 0$

$H_a = \text{alternative hypothesis} = \quad b_1 \neq 0$

If the confidence interval at the desired level of significance ***does not include zero, the null hypothesis is rejected***, and the coefficient is said to be statistically different from zero.

HYPOTHESIS TESTING OF LINEAR REGRESSION COEFFICIENTS

Hypothesis Tests of the Slope Coefficient

t-statistic is used to test the significance of the individual coefficients (e.g., slope) in a regression. It is used to test whether the population slope is different from a specific value.

Suppose we want to test a hypothesis about the slope.

Null and Alternative hypotheses
H_0: $b_1 = 0$ (no linear relationship)
H_1: $b_1 \neq 0$ (linear relationship does exist)

A t-test statistic is calculated by subtracting the hypothesized population slope B_1 from the estimated slope coefficient $\hat{b}_1$ and then dividing the difference by the standard error of the slope coefficient $s_{\hat{b}_1}$.

Test statistic $t = \frac{\hat{b}_1 - B1}{s_{\hat{b}_1}}$

where,

$\hat{b}_1$ = *Sample regression slope coefficient*

b_1 = *Hypothesized slope*

S_{b_1} = *Standard error of the slope*

df= n-k-1 = n–2

Decision Rule:

If test statistic is <– t-critical or > + t-critical with n-2 degrees of freedom, (if absolute value of $t > t_c$), Reject H_0; otherwise, do not Reject H_0.

Standard Error of Slope Coefficient $(s_{\hat{b}_1})$

It is the ratio of standard error of estimate S_e to the square root of variation of independent variable. for simple linear regression:

$$s_{\hat{b}_1} = \frac{s_e}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

The greater the variability of the independent variable the lower will be the standard error of slope and therefore greeter will be the calculated t-statistic.

Hypothesis Tests of the Intercept

This test is useful to determine wthether the population intercept is a specific value.

Intercept is the predicted value of the dependent variable when the independent variable is set to zero

Standard error of the intercept = $s_{\hat{b}_0} = \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$

The test is whether the intercept is different from the hypothesized value B_0 , using the following formula.

$$t_{intercept} = \frac{\hat{b}_0 - B_0}{s_{\hat{b}_0}} = \frac{\hat{b}_0 - B_0}{\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}}$$

Question 26 of 34

Indicator variable is a variable that takes on:

A. a value of either 0 or 1.
B. any value ranges from 0 to 1.
C. any value ranges from -1 to +1.

Correct Answer: A
Explanation:

Hypothesis Tests of Slope When Independent Variable Is an Indicator Variable

- Indicator variable (or dummy variable) is a variable that takes on a value of either 1 or 0.
- 1 if a particular condition is true and 0 if that condition is false.
- In simple linear regressions, the slope is the difference in the dependent variable for the two conditions.

Alpha measures the excess return and beta the riskiness of a fund (Y) relative to the market (X)

R-squared	0.825
Standard error of estimate	0.018
Number of observations	25

	Coefficients	**Standard error**
Alpha	0.0008	0.0034
Beta	0.821	0.0356

Two-tailed t-table at 5% level of significance

df	**t-critical**
25	2.060
24	2.064
23	2.069

Which of the following statements is *most likely* correct?

A. The fund has negative excess return.
B. The fund has significant excess return beyond the return associated with market risk of the fund.
C. The fund does not have significant excess return beyond the return associated with market risk of the fund.

Correct Answer: C

Explanation:

Here the intercept is Alpha, so we construct the null hypothesis as:

H_0: $b_1 = 0$
H_a: $b_1 \neq 0$

Degrees of freedom = n – 2 = 25 – 2 = 23
t-critical is 2.069 from the t-table

The decision rule is, reject H_0 if t-statistic > t-critical

For Alpha, t-statisitc $= \frac{b_0 - 0}{s_{b_0}}$
= 0.0008 – 0/ 0.0034
= 0.23529

But t-statistic is much less than t-critical, so we conclude that the intercept, alpha, is not statistically different from zero, i.e., the fund does not have significant excess return beyond the return associated with market risk of the fund.

For Beta, t-statisitc $= \frac{b_0 - 0}{s_{b_0}}$

$= 0.821 – 0/ 0.0356$

$= 23.06$

For Beta, t-statistic is much higher than t-critical, so we conclude that the beta, is statistically different from zero

HYPOTHESIS TESTING OF LINEAR REGRESSION COEFFICIENTS

Hypothesis Tests of the Slope Coefficient

t-statistic is used to test the significance of the individual coefficients (e.g., slope) in a regression. It is used to test whether the population slope is different from a specific value.

Suppose we want to test a hypothesis about the slope.

Null and Alternative hypotheses

H_0: $b_1 = 0$ (no linear relationship)

H_1: $b_1 \neq 0$ (linear relationship does exist)

A t-test statistic is calculated by subtracting the hypothesized population slope B_1 from the estimated slope coefficient $\hat{b}_1$ and then dividing the difference by the standard error of the slope coefficient $s_{\hat{b}_1}$.

Test statistic $t = \frac{\hat{b}_1 - B1}{s_{\hat{b}_1}}$

where,

$\hat{b}_1$ = *Sample regression slope coefficient*

b_1 = *Hypothesized slope*

S_{b_1} = *Standard error of the slope*

df= n-k-1 = n–2

Decision Rule:

If test statistic is <– t-critical or > + t-critical with n-2 degrees of freedom, (if absolute value of $t > t_c$), Reject H_0; otherwise, do not Reject H_0.

Standard Error of Slope Coefficient ($s_{\hat{b}_1}$)

It is the ratio of standard error of estimate S_e to the square root of variation of independent variable. for simple linear regression:

$$s_{\hat{b}_1} = \frac{S_e}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

The greater the variability of the independent variable the lower will be the standard error of slope and therefore greeter will be the calculated t-statistic.

Hypothesis Tests of the Intercept

This test is useful to determine wthether the population intercept is a specific value. Intercept is the predicted value of the dependent variable when the independent variable is set to zero

Standard error of the intercept = $s_{\hat{b}_0} = \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$

The test is whether the intercept is different from the hypothesized value B_0 , using the following formula.

$$t_{intercept} = \frac{\hat{b}_0 - B_0}{s_{\hat{b}_0}} = \frac{\hat{b}_0 - B_0}{\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}}$$

Question 28 of 34

In hypothesis testing for regression coefficient, as standard error of estimate increases the $s_{\hat{b}_1}$:

A. increases and the confidence interval widens.
B. increases and the confidence interval narrows.
C. decreases and the confidence interval narrows.

Correct Answer: A

Explanation:

In hypothesis testing for regression coefficient, as standard error of estimate increases, ***the $s_{\hat{b}_1}$ increases, and the confidence interval widens.***

Standard Error of Slope Coefficient ($s_{\hat{b}_1}$)

It is the ratio of standard error of estimate S_e to the square root of variation of independent variable. for simple linear regression:

$$s_{\hat{b}_1} = \frac{s_e}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

The greater the variability of the independent variable the lower will be the standard error of slope and therefore greeter will be the calculated t-statistic.

Question 29 of 34

Consider a confidence interval from 1.19 to 1.83, and an analyst is testing the following null hypothesis at 5% level of significance.

$H_0: \quad b_1 = 1 \qquad H_a: \quad b_1 \neq 1$

The analyst will *most likely*:

A. reject the null hypothesis because the confidence interval does not include 0.
B. accept the null hypothesis because the confidence interval does not include 1.
C. reject the null hypothesis because the confidence interval does not include 1.

Correct Answer: C

Explanation:

Since the confidence interval ranges from 1.19 to 1.83, ***the analyst should reject the null hypothesis because the confidence interval does not include 1.***

Had we been testing the null hypothesis of $b_1 = 0$, only then we would have rejected it on the basis that it does not include 0.

HYPOTHESIS TESTING OF LINEAR REGRESSION COEFFICIENTS

Hypothesis Tests of the Slope Coefficient

t-statistic is used to test the significance of the individual coefficients (e.g., slope) in a regression. It is used to test whether the population slope is different from a specific value.

Suppose we want to test a hypothesis about the slope.

Null and Alternative hypotheses

H_0: $b_1 = 0$ (no linear relationship)

H_1: $b_1 \neq 0$ (linear relationship does exist)

A t-test statistic is calculated by subtracting the hypothesized population slope B_1 from the estimated slope coefficient $\hat{b}_1$ and then dividing the difference by the standard error of the slope coefficient $s_{\hat{b}_1}$.

Test statistic $t = \frac{\hat{b}_1 - B1}{s_{\hat{b}_1}}$

where,

$\hat{b}_1$ = *Sample regression slope coefficient*

b_1 = *Hypothesized slope*

S_{b_1} = *Standard error of the slope*

df= n-k-1 = n–2

Decision Rule:

If test statistic is <– t-critical or > + t-critical with n-2 degrees of freedom, (if absolute value of t > t_c), Reject H_0; otherwise, do not Reject H_0.

Standard Error of Slope Coefficient $(s_{\hat{b}_1})$

It is the ratio of standard error of estimate S_e to the square root of variation of independent variable. for simple linear regression:

$$s_{\hat{b}_1} = \frac{S_e}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

The greater the variability of the independent variable the lower will be the standard error of slope and therefore greeter will be the calculated t-statistic.

Hypothesis Tests of the Intercept

This test is useful to determine wthether the population intercept is a specific value.

Intercept is the predicted value of the dependent variable when the independent variable is set to zero

Standard error of the intercept = $s_{\hat{b}_0} = \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$

The test is whether the intercept is different from the hypothesized value B_0 , using the following formula.

$$t_{intercept} = \frac{\hat{b}_0 - B_0}{s_{\hat{b}_0}} = \frac{\hat{b}_0 - B_0}{\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}}$$

Question 30 of 34

Decreasing level of significance from 5% to 1% *least likely*:

A. increases the probability of rejecting true null hypothesis.
B. decreases the probability of rejecting true null hypothesis.
C. decreases the probability of rejecting a false null hypothesis.

Correct Answer: A

Explanation:

A is correct. Reducing the level of significance decreases the chances of Type I and type II errors.

Rejecting null hypothesis when it is true is called a Type I error failing to reject null hypothesis when it is false is called type II error.

Test of Hypothesis: Level of Significance and p-Values

p-value

The p-value is the smallest level of significance at which the null hypothesis can be rejected.

For example, if the p-value is 0.05 i.e., 0.05 significance level, this indicates that there is a 5% chance of rejecting the hypothesis when actually it is true actually. This is a Type 1 error.

- The smaller the p-value, the smaller the chance of making type I error (i.e., the greater the probability of rejecting the null hypothesis).
- Type I error = False positive = rejecting the null hypothesis when it is true.
- Type II error = False negative = not rejecting the null hypothesis when it is wrong.

Question 31 of 34

Decreasing level of significance from 5% to 1%:

A. increases the probability of rejecting true null hypothesis.
B. increases the probability of failing to reject a false null hypothesis.
C. decreases the probability of failing to reject a false null hypothesis.

Correct Answer: B

Explanation:

Rejecting null hypothesis when it is true is called a *Type I* error. Reducing the level of significance decreases the chances of *Type I* error, but it also increases the probability of *Type II* error, which is failing to reject null hypothesis when it is false.

Test of Hypothesis: Level of Significance and p-Values

p-value

The p-value is the smallest level of significance at which the null hypothesis can be rejected.

For example, if the p-value is 0.05 i.e., 0.05 significance level, this indicates that there is a 5% chance of rejecting the hypothesis when actually it is true actually. This is a Type 1 error.

- The smaller the p-value, the smaller the chance of making type I error (i.e., the greater the probability of rejecting the null hypothesis).
- Type I error = False positive = rejecting the null hypothesis when it is true.
- Type II error = False negative = not rejecting the null hypothesis when it is wrong.

Question 32 of 34

p-value or probability value for a particular hypothesis is the smallest level of significance at which the:

A. null hypothesis can be rejected.
B. null hypothesis can be accepted.
C. alternative hypothesis can be rejected.

Correct Answer: A

Explanation:

p-value or probability value for a particular hypothesis is the smallest level of significance at which the null hypothesis can be rejected.

Test of Hypothesis: Level of Significance and p-Values

p-value

The p-value is the smallest level of significance at which the null hypothesis can be rejected.

For example, if the p-value is 0.05 i.e., 0.05 significance level, this indicates that there is a 5% chance of rejecting the hypothesis when actually it is true actually. This is a Type 1 error.

- The smaller the p-value, the smaller the chance of making type I error (i.e., the greater the probability of rejecting the null hypothesis).
- Type I error = False positive = rejecting the null hypothesis when it is true.

Question 33 of 34

Which of the following statements is *least likely* correct regarding the standard error of forecast s_f?

A. The closer the X_f is to $\bar{X}$, the smaller will be the s_f.
B. The better the fit of the regression, the smaller will be s_f.
C. The smaller the n (sample size), the smaller will be the s_f.

Correct Answer: C
Explanation:

PREDICTING USING SIMPLE LINEAR REGRESSION AND PREDICTION INTERVALS

- Regression results are used to make prediction about the dependent variable.
- Prediction intervals are used to realize how sure we are about the predicted results.
- Prediction interval for a forecasted value of a dependent variable is created by using equation $\hat{Y} \pm t_c s_f$ *where,*

$$s_f^2 = s_e{}^2\left[1+\frac{1}{n}+\frac{(X_f-\bar{X})^2}{(n-1)s_X^2}\right] = s_e{}^2\left[1+\frac{1}{n}+\frac{(X_f-\bar{X})^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right]$$

- Standard error of forecast = $s_f = s_e\sqrt{1+\frac{1}{n}+\frac{(X_f-\bar{X})^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}}$

as

$$s_f = \sqrt{s_f^2}$$

s_e = standard error of estimate
n = number of observations
X = value of independent variable
$\bar{X}$ = estimated mean of X
s^2_X= variance of independent variable
t_c = critical t-value for n −k −1 degrees of freedom.

The following can be taken from the equation:

1. The better the fit of the regressions, the smaller will be the s_e and therefore, the smaller will be the s_f.
2. The larger the n (sample size), the smaller will be the s_f.
3. The closer the X_f is to $\bar{X}$, the smaller will be the s_f.

where X_f is forecasted independent variable and $\bar{X}$ is forecasted mean of the independent variable.

Question 34 of 34

In which of the following model, the independent variable is in logarithmic form, but the dependent variable is in linear form?

A. Log-lin model
B. Lin-log model
C. Log-log model

Correct Answer: B
Explanation:

FUNCTIONAL FORMS FOR SIMPLE LINEAR REGRESSION

Economic and financial data often exhibit nonlinear relationships between two variables. To apply simple linear regression model on such data, we need to modify either the dependent or the independent variable.

Many functional forms can be used to transform data to enable their use in linear regression. Three commonly used functional forms that involve log transformation are as follows:

1. The log-lin model
2. The lin-log model
3. The log-log model

The Log-Lin Model

In this model, the dependent variable is in logarithmic form, but the independent variable is in linear form.

$$\ln Y_i = b_0 + b_i X_i$$

Slope coefficient → relative change in the dependent variable for an absolute change in the independent variable.

Note: Directly comparing the different model values is not possible because variables are not in the same form.

The Lin-Log Model

In this model, the independent variable is in logarithmic form, but the dependent variable is in linear

$Y_i = b_0 + b_i \ln X_i$

Slope coefficient → absolute change in the dependent variable for a relative change in the independent variable.

The Log-Log Model

In this model, both the dependent and independent variables are in logarithmic form. This model is also called double-log model.

$\ln Y_i = b_0 + b_i \ln X_i$

Slope coefficient → relative change in the dependent variable for a relative change in the independent variable.

This model is suitable in calculating elasticities.

Selecting the Correct Functional Form

Selection of the suitable functional form depends on examining the goodness of fit measures (R^2, F-statistic & S_e) as well as patterns in the residuals.

Many statistical software packages enable us to visually examine and inspect the distribution of the residuals