***Business Analytics, 2e, GE* (Evans)**

**Chapter 8Trendlines and Regression Analysis**

1) A regression model that involves a single independent variable is called \_\_\_\_\_\_\_\_.

A) single regression

B) unit regression

C) simple regression

D) individual regression

Answer: C

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Explain the purpose of regression analysis and provide examples in business.

LO2: Identify the components of simple linear regression models and discuss their applications

2) Regression models of \_\_\_\_\_\_\_\_ data focus on predicting the future.

A) missing

B) time-series

C) panel

D) cross-sectional

Answer: B

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Explain the purpose of regression analysis and provide examples in business.

LO2: Identify the components of simple linear regression models and discuss their applications

The following table exhibits the age of antique furniture and the corresponding prices. Use the table to answer the following question(s). (Hint: Use scatter diagram and the Excel Trendline tool where necessary).

|  |  |
| --- | --- |
| **Number of years** | **Value**  **($)** |
| 78 | 930 |
| 91 | 1010 |
| 83 | 970 |
| 159 | 1950 |
| 134 | 1610 |
| 210 | 2880 |
| 88 | 980 |
| 178 | 2010 |
| 124 | 1370 |
| 72 | 900 |

3) What is the relationship between the age of the furniture and their values?

A) Nonlinear

B) Linear

C) Curvilinear

D) No relationship

Answer: B

Diff: 2

Blooms: Apply

AACSB: Analytic Skills

Topic: Simple Linear Regression

LO1: Use a scatter chart to identify the type of relationship between two variables.

LO2: Identify the components of simple linear regression models and discuss their applications

4) Which of the following is true of linear functions used in predictive analytical models?

A) It is used when the rate of change in a variable decreases or increases quickly and then levels out.

B) It is used when there is a steady decrease or increase over a range of a variable.

C) It is used when there is increase at a specific rate.

D) It is used when there is a rise or fall at a constantly increasing rate.

Answer: B

Diff: 2

Blooms: Remember

Topic: Data-Driven Modeling

LO1: List the common types of mathematical functions used in predictive modeling.

LO2: Identify different business uses for statistics and the major statistical tools businesses use

5) \_\_\_\_\_\_\_\_ are mathematical functions used in predictive analytical models which define phenomena that increase at a specific rate, and is represented by the formula*y = axb*

A) Exponential functions

B) Power functions

C) Polynomial functions

D) Logarithmic functions

Answer: B

Diff: 1

Blooms: Remember

Topic: Data-Driven Modeling

LO1: List the common types of mathematical functions used in predictive modeling.

LO2: Identify different business uses for statistics and the major statistical tools businesses use

6) Which of the following mathematical functions, used in predictive analytical models, is represented by the formula *y* = *ax3* + *bx*2+ c*x* + *d*?

A) exponential functions

B) power functions

C) logarithmic functions

D) polynomial functions

Answer: D

Diff: 1

Blooms: Remember

Topic: Data-Driven Modeling

LO1: List the common types of mathematical functions used in predictive modeling.

LO2: Identify different business uses for statistics and the major statistical tools businesses use

7) In \_\_\_\_\_\_\_\_ functions, represented by *y* = *abx*, *y* rises or falls at constantly increasing rates.

A) logarithmic

B) power

C) exponential

D) polynomial

Answer: C

Diff: 1

Blooms: Remember

Topic: Data-Driven Modeling

LO1: List the common types of mathematical functions used in predictive modeling.

LO2: Identify different business uses for statistics and the major statistical tools businesses use

8) In Excel's *Trendline* tool, the value of the\_\_\_\_\_\_\_\_ gives the measure of fit of the line to the data.

A) linear function

B) *R*-squared

C) moving average

D) set intercept

Answer: B

Diff: 1

Blooms: Remember

Topic: Data-Driven Modeling

LO1: Apply the Excel Trendline tool to fit models to data.

LO2: Identify different business uses for statistics and the major statistical tools businesses use

9) Which of the following is true of the *R*-squared (*R*2) value in Excel's *Trendline* function?

A) A value of 1.0 for *R*2 indicates maximum deviation of the data from the line.

B) If the value of *R*2 is above 1.0, the line will be at a perfect fit for the data.

C) The value of *R*2 will always be between -1 and 1.

D) As the value of *R*2 gets higher, the line will be a better fit for the data.

Answer: D

Diff: 1

Blooms: Remember

Topic: Data-Driven Modeling

LO1: Apply the Excel Trendline tool to fit models to data.

LO2: Identify different business uses for statistics and the major statistical tools businesses use

10) Which of the following equations correctly expresses the relationship between the two variables?

A) *Value = (-181.16) + 13.493 × Number of years*

B) *Number of years = Value / 12.537*

C) *Value = (459.34 / Number of years) × 4.536*

D) *Number of years = (17.538 × Value) / (-157.49)*

Answer: A

Diff: 3

Blooms: Apply

AACSB: Analytic Skills

Topic: Simple Linear Regression

LO1: Use the Excel Trendline tool to find the best-fitting simple linear regression model.

LO2: Identify the components of simple linear regression models and discuss their applications

11) What is the expected value for a 90 year-old piece of furniture?

A) $1002.45

B) $997.98

C) $934.56

D) $1033.21

Answer: D

Diff: 3

Blooms: Apply

AACSB: Analytic Skills

Topic: Simple Linear Regression

LO1: Use the Excel Trendline tool to find the best-fitting simple linear regression model.

LO2: Identify the components of simple linear regression models and discuss their applications

12) In a linear relationship, which of the following accounts for the many possible values of the dependent variable that vary around the mean?

A) the coefficient of the dependent variable X

B) the value of the intercept *ß0*

C) the random error term ε

D) the standard error SYX

Answer: C

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Explain how least-squares regression finds the best-fitting regression model.

LO2: Identify the components of simple linear regression models and discuss their applications

13) Which of the following is true about the observed errors associated with estimating the value of the dependent variable using the regression line?

A) They are the horizontal distances between slopes and *y*-intercepts.

B) The errors are also referred to as critical values.

C) They are always maximized by the regression lines.

D) The errors can be negative or positive.

Answer: D

Diff: 2

Blooms: Understand

Topic: Simple Linear Regression

LO1: Explain how least-squares regression finds the best-fitting regression model.

LO2: Identify the components of simple linear regression models and discuss their applications

14) For an independent variable Y, the error associated with the *i*th observation is:

A) *ei* = *Yi* - *Ŷi*

B) *Yi*= (*ei)2* - *Ŷi*

C) *(Ŷi)2 ei* = *Yi*

D) *ei* = (*Yi* + *Ŷi)2*

Answer: A

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Explain how least-squares regression finds the best-fitting regression model.

LO2: Identify the components of simple linear regression models and discuss their applications

Use the data given below to answer the following question(s).

Following is an extract from the database of a construction company. The table shows the height of walls in feet and the cost of raising them. The estimated simple linear regression equation is given as Ŷ = b0 + b1X. (Hint: Use Excel functions).

|  |  |
| --- | --- |
| **Height (ft)** | **Cost ($)** |
| 4 | 670 |
| 3 | 430 |
| 7 | 810 |
| 9 | 1100 |
| 6 | 790 |
| 8 | 880 |
| 5 | 760 |
| 11 | 1200 |

15) What is the value of the coefficient *b*0?

A) -2.25321

B) 0.010697

C) 254.8371

D) 86.81704

Answer: C

Diff: 3

Blooms: Apply

AACSB: Analytic Skills

Topic: Simple Linear Regression

LO1: Use Excel functions to find least-squares regression coefficients.

LO2: Identify the components of simple linear regression models and discuss their applications

16) What is the value of the coefficient *b*1?

A) 86.81704

B) 254.8371

C) 0.010697

D) -2.14625

Answer: A

Diff: 3

Blooms: Apply

AACSB: Analytic Skills

Topic: Simple Linear Regression

LO1: Use Excel functions to find least-squares regression coefficients.

LO2: Identify the components of simple linear regression models and discuss their applications

17) What is the estimated cost of raising a 10-inch wall?

A) 1505.786

B) 1103.578

C) 968.6109

D) 1123.008

Answer: D

Diff: 3

Blooms: Apply

AACSB: Analytic Skills

Topic: Simple Linear Regression

LO1: Use Excel functions to find least-squares regression coefficients.

LO2: Identify the components of simple linear regression models and discuss their applications

18) Which of the following statements is true when using the Excel *Regression* tool?

A) The range for the independent variable values must be specified in the box for the *Input Y Range.*

B) Checking the option *Constant is Zero* forces the intercept to zero.

C) The *Regression* tool can be found in the *Tools* tab under *Insert* group.

D) Adding an intercept term reduces the analysis' fit to the data.

Answer: B

Diff: 2

Blooms: Understand

Topic: Simple Linear Regression

LO1: Use the Excel Regression tool for both single and multiple linear regressions.

LO2: Identify the components of simple linear regression models and discuss their applications

19) Which of the following generates a scatter chart in Excel with the values predicted by the regression model included?

A) *Trendline*

B) *Residual Plots*

C) *R Square*

D) *Line Fit Plots*

Answer: D

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Use the Excel Regression tool for both single and multiple linear regressions.

LO2: Identify the components of simple linear regression models and discuss their applications

20) Which of the following is true about Excel outputs *Multiple R*?

A) It is often referred to as the coefficient of determination.

B) A value of 0 indicates positive correlation.

C) A negative slope of the regression line denotes a positive *Multiple R*.

D) It is another name for the sample correlation coefficient, *r*.

Answer: D

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Interpret the regression statistics of the Excel Regression tool.

LO2: Identify the components of simple linear regression models and discuss their applications

21) The *R2*value:

A) is the variability of the observed *Y*-values from the predicted values.

B) indicates that as the independent variable increases, the intercept term does too.

C) gives the proportion of variation in the dependent variable that is explained by the independent variable.

D) transforms the cumulative probability scale (vertical axis) so that the graph of the cumulative normal distribution is a straight line.

Answer: C

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Interpret the regression statistics of the Excel Regression tool.

LO2: Identify the components of simple linear regression models and discuss their applications

22) For a simple linear regression model, significance of regression is:

A) a measure of how well the regression line fits the data.

B) a hypothesis test of whether the true regression coefficient ß1 is zero.

C) a statistic that modifies the value of *R*2 by incorporating the sample size and the number of explanatory variables in the model.

D) the variability of the observed Y-values from the predicted values.

Answer: B

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Interpret significance of regression from the Excel Regression tool output.

LO2: Identify the components of simple linear regression models and discuss their applications

23) Which of the following Excel functions is applied to test for significance of regression?

A) COVAR

B) ANOVA

C) SINH

D) TREND

Answer: B

Diff: 1

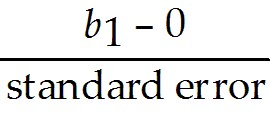
Blooms: Remember

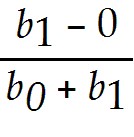
Topic: Simple Linear Regression

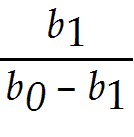
LO1: Interpret significance of regression from the Excel Regression tool output.

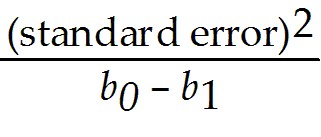
LO2: Identify the components of simple linear regression models and discuss their applications

24) While testing hypotheses for regression coefficients, the *t*-test for the slope is expressed as:

A) *t*= 

B) *t*= 

C) *t*= 

D) *t*= 

Answer: A

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Draw conclusions for tests of hypotheses about regression coefficients.

LO2: Identify the components of simple linear regression models and discuss their applications

25) \_\_\_\_\_\_\_\_ provide information about the unknown values of the true regression coefficients, accounting for sampling error.

A) Standard errors

B) Confidence intervals

C) Adjusted R Squares

D) *P-*values

Answer: B

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Interpret confidence intervals for regression coefficients

LO2: Identify the components of simple linear regression models and discuss their applications

26) Standard residuals:

A) help detect outliers that may bias the results of a regression analysis.

B) cause differences in the regression equation by changing the slope and intercept.

C) point out the ranges for the population intercept and slope at a 95% confidence level.

D) provide information for testing hypothesis associated with the intercept and slope.

Answer: A

Diff: 2

Blooms: Remember

Topic: Residual Analysis and Regression Assumptions

LO1: Calculate standard residuals.

LO2: Identify the components of simple linear regression models and discuss their applications

27) A(n) \_\_\_\_\_\_\_\_ is an extreme value that is different from the rest of the data.

A) critical value

B) standard error

C) expected value

D) outlier

Answer: D

Diff: 1

Blooms: Remember

Topic: Residual Analysis and Regression Assumptions

LO1: Calculate standard residuals.

LO2: Identify the components of simple linear regression models and discuss their applications

28) While checking for linearity by examining the residual plot, the residuals must:

A) exhibit a linear trend.

B) form a parabolic shape.

C) be randomly scattered.

D) be below the x-axis.

Answer: C

Diff: 1

Blooms: Remember

Topic: Residual Analysis and Regression Assumptions

LO1: Calculate standard residuals.

LO2: Identify the components of simple linear regression models and discuss their applications

29) Which of the following is true when testing for normality of errors?

A) Normality is verified by inspecting for a bell-shaped distribution.

B) It is easier to evaluate normality with small sample sizes.

C) A scatter diagram of the whole data is always used to verify normality.

D) Errors are normally distributed when the scatter diagram shows a straight-line distribution.

Answer: A

Diff: 1

Blooms: Remember

Topic: Residual Analysis and Regression Assumptions

LO1: List the assumptions of regression analysis and describe methods to verify them.

LO2: Identify the components of simple linear regression models and discuss their applications

30) \_\_\_\_\_\_\_\_ means that the variation about the regression line is constant for all values of the independent variable.

A) Autocorrelation

B) Normality of errors

C) Homoscedasticity

D) Linearity

Answer: C

Diff: 1

Blooms: Remember

Topic: Residual Analysis and Regression Assumptions

LO1: List the assumptions of regression analysis and describe methods to verify them.

LO2: Identify the components of simple linear regression models and discuss their applications

31) Which of the following helps in evaluation of autocorrelation?

A) Breusch-Pagan test

B) Durbin-Watson statistic

C) Hosmer-Lemeshow test

D) Cochran-Mantel-Haenszel statistics

Answer: B

Diff: 1

Blooms: Remember

Topic: Residual Analysis and Regression Assumptions

LO1: List the assumptions of regression analysis and describe methods to verify them.

LO2: Identify the components of simple linear regression models and discuss their applications

32) In multiple regression, *R Square* is referred to as the:

A) multiple correlation coefficient.

B) coefficient of autocorrelation.

C) coefficient of multiple determination.

D) multiple significance coefficient.

Answer: C

Diff: 1

Blooms: Remember

Topic: Multiple Linear Regression

LO1: Explain the differences in the Excel Regression tool output for simple and multiple linear regression models.

LO2: Develop and test a multiple regression model

33) Which of the following is true about multiple linear regression?

A) It is a linear regression model with more than one dependent variable.

B) The regression coefficients are called fractional regression coefficients.

C) It uses least squares to estimate the intercept and slope coefficients.

D) The ANOVA tests for the significance of each variable separately.

Answer: C

Diff: 1

Blooms: Understand

Topic: Multiple Linear Regression

LO1: Explain the differences in the Excel Regression tool output for simple and multiple linear regression models.

LO2: Develop and test a multiple regression model

34) When using the *t*-statistic in multiple regression to determine if a variable should be removed:

A) *R*2 will increase if the variable is removed.

B) if |t| > 1, the standard error will decrease.

C) a large number of independent variables is convenient.

D) if |t| < 1, the standard error will increase.

Answer: A

Diff: 2

Blooms: Understand

Topic: Building Good Regression Models

LO1: Apply a systematic approach to build good regression models.

LO2: Develop and test a multiple regression model

35) When two or more independent variables in the same regression model can predict each other better than the dependent variable, the condition is referred to as \_\_\_\_\_\_\_\_.

A) autocorrelation

B) heteroscedasticity

C) multicollinearity

D) homoscedasticity

Answer: C

Diff: 1

Blooms: Remember

Topic: Building Good Regression Models

LO1: Explain the importance of understanding multicollinearity in regression models.

LO2: Develop and test a multiple regression model

36) Which of the following is true about multicollinearity?

A) The effect of a dependent variable on another becomes difficult to isolate.

B) Regression coefficients become clearer and are easier to interpret.

C) *P-*values reduce significantly leading to rejection of null hypothesis.

D) It is best measured using the statistic variance inflation factor (VIF).

Answer: D

Diff: 1

Blooms: Remember

Topic: Building Good Regression Models

LO1: Explain the importance of understanding multicollinearity in regression models.

LO2: Develop and test a multiple regression model

37) Categorical variables that have been coded are called \_\_\_\_\_\_\_\_.

A) limited dependent variables

B) dummy variables

C) instrumental variables

D) observable variables

Answer: B

Diff: 1

Blooms: Remember

Topic: Regression with Categorical Independent Variables

LO1: Build regression models for categorical data using dummy variables.

LO2: Develop and test a multiple regression model

38) Interaction is:

A) the principle of having a model with maximum explanatory variables.

B) the process of coding categorical variables.

C) a measure to determine the correlation between dependent variables.

D) the dependence between two independent variables.

Answer: D

Diff: 1

Blooms: Remember

Topic: Regression with Categorical Independent Variables

LO1: Test for interactions in regression models with categorical variables.

LO2: Develop and test a multiple regression model

39) How many additional dummy variables are required if a categorical variable has 4 levels?

A) 2

B) 3

C) 1

D) 4

Answer: B

Diff: 1

Blooms: Understand

Topic: Regression with Categorical Independent Variables

LO1: Build regression models for categorical data using dummy variables.

LO2: Develop and test a multiple regression model

40) When a scatter chart of data shows a nonlinear relationship, the nonlinear model can be expressed as:

A) *Y* = *β0*+ *β1X* + *β2X2*+ *ε*

B) *Y* = *β0*+ *β1X* + (*β2X)2*+ *ε*

C) *Y* = *β0*+ *β1X* + *β2X*

D) *Y* = *β*0 + *β1X2*+ *β2X2*+ *ε*

Answer: A

Diff: 1

Blooms: Remember

Topic: Regression Models with Nonlinear Terms

LO1: Identify when curvilinear regression models are more appropriate than linear models.

LO2: Develop and test a multiple regression model

41) In a curvilinear regression model, the \_\_\_\_\_\_\_\_ represents the curvilinear effect.

A) intercept

B) error term

C) slope

D) *R Square*

Answer: C

Diff: 1

Blooms: Remember

Topic: Regression Models with Nonlinear Terms

LO1: Identify when curvilinear regression models are more appropriate than linear models.

LO2: Develop and test a multiple regression model

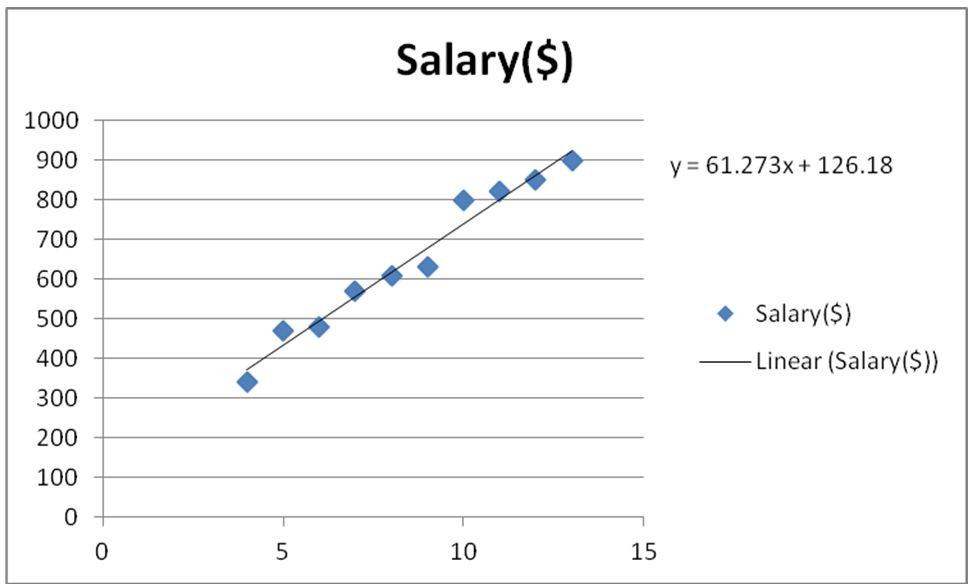
Use the data given below to answer the following question(s).

Following is an extract from a firm's database detailing the number of hours spent on the job by employees and their corresponding pay. (Note: Assume a level of significance of 0.05 wherever necessary.)

|  |  |
| --- | --- |
| **Hours spent**  **on the job** | **Salary ($)** |
| 4 | 340 |
| 12 | 850 |
| 7 | 570 |
| 5 | 470 |
| 11 | 820 |
| 8 | 610 |
| 9 | 630 |
| 13 | 900 |
| 10 | 800 |
| 6 | 480 |

42) Construct a scatter diagram and use the Excel Trendline tool to find the best-fitting simple linear regression model.

Answer:



Diff: 2

Blooms: Apply

AACSB: Analytic Skills

Topic: Simple Linear Regression

LO1: Use a scatter chart to identify the type of relationship between two variables. Use the Excel Trendline tool to find the best-fitting simple linear regression model.

LO2: Identify the components of simple linear regression models and discuss their applications

43) Is the hours spent on the job a statistically significant variable in explaining the variation in pay of employees? (Hint: Use Regression tool).

Answer: Hypothesis test for significance of regression:

*H*0 :*β*1 = 0

*H*1 :*β*1 ≠ 0

where, *β*1 is the slope of the intercept.

*Significance F*, that is, the *p*-value associated with the hypothesis test is essentially zero (2.525 × 10-7). Therefore, assuming a level of significance of 0.05, the null hypothesis must be rejected and conclude that the slope–the coefficient for number of hours spent on the job–is not zero. This means that work hours is a statistically significant variable in explaining the variation in employee pay.

Diff: 3

Blooms: Apply

AACSB: Analytic Skills

Topic: Simple Linear Regression

LO1: Interpret significance of regression from the Excel Regression tool output.

LO2: Identify the components of simple linear regression models and discuss their applications

44) Draw conclusions for test of hypothesis for regression coefficients.

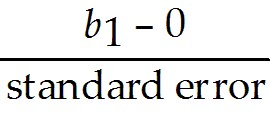
Answer: Hypothesis test for regression coefficients:

*H*0 :*β1* = 0

*H*1 :*β1* ≠ 0

where, β1 is the slope of the intercept.

The *t*-test for the slope:

*t =*  = 61.27272727 / 3.868357187 = 15.83946991

Because the *P-value* for both coefficients is essentially zero, it can be concluded that neither coefficient is statistically equal to zero.

Diff: 3

Blooms: Apply

AACSB: Analytic Skills

Topic: Simple Linear Regression

LO1: Draw conclusions for tests of hypotheses about regression coefficients.

LO2: Identify the components of simple linear regression models and discuss their applications

45) Interpret the confidence intervals.

Answer: A 95% confidence interval for the intercept is [46.15, 206.22]. Similarly, a 95% confidence interval for the slope is [52.35, 70.19]. Although the regression model is Ŷ = 126.18 + 61.273X, the confidence intervals suggest a bit of uncertainty about predictions using the model. Although, it can be estimated that an employee who spent 10.5 hours on the job will receive a pay of 126.18 + (61.273 × 10.5) = $769.5465, if the true parameters are at the extremes of the confidence intervals, the estimate might be as low as 46.15 + (52.35 × 10.5) = $595.825 or as high as 206.22 + (70.19 × 10.5) = $943.215.

Diff: 3

Blooms: Apply

AACSB: Analytic Skills

Topic: Simple Linear Regression

LO1: Interpret confidence intervals for regression coefficients

LO2: Identify the components of simple linear regression models and discuss their applications

46) Interpret residual output.

Answer: From the data, the first observation has a salary of $340 and the regression model predicts $371.27. Thus the residual is 340 - 371.27 = - $31.27. By dividing this residual by its standard deviation, the standardized residual for the first observation is - 0.944. This means that the first observation is about 0.94 standard deviations below the regression line. There are no standardized residuals outside of ± 3 standard deviations. It can be concluded that there are no outliers.

Diff: 3

Blooms: Apply

AACSB: Analytic Skills

Topic: Residual Analysis and Regression Assumptions

LO1: Calculate standard residuals.

LO2: Identify the components of simple linear regression models and discuss their applications

47) In predictive analysis models, a second-order polynomial has only one hill or valley.

Answer: TRUE

Diff: 1

Blooms: Remember

Topic: Data-Driven Modeling

LO1: List the common types of mathematical functions used in predictive modeling.

LO2: Identify different business uses for statistics and the major statistical tools businesses use

48) Excel's *Trendline* feature cannot be used in modeling trends which include time variables.

Answer: FALSE

Diff: 1

Blooms: Understand

Topic: Data-Driven Modeling

LO1: Apply the Excel Trendline tool to fit models to data.

LO2: Identify different business uses for statistics and the major statistical tools businesses use

49) The best-fitting line maximizes the residuals.

Answer: FALSE

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Explain how least-squares regression finds the best-fitting regression model.

LO2: Identify the components of simple linear regression models and discuss their applications

50) Creating a scatter chart with an added trendline is visually superior to the scatter chart generated by line fit plots.

Answer: TRUE

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Use a scatter chart to identify the type of relationship between two variables.

LO2: Identify the components of simple linear regression models and discuss their applications

51) The standard error may be assumed to be large if the data are clustered close to the regression line.

Answer: FALSE

Diff: 1

Blooms: Remember

Topic: Simple Linear Regression

LO1: Interpret the regression statistics of the Excel Regression tool.

LO2: Identify the components of simple linear regression models and discuss their applications

52) An increase in adjusted *R2* indicates that the regression model has improved.

Answer: TRUE

Diff: 1

Blooms: Remember

Topic: Building Good Regression Models

LO1: Interpret the regression statistics of the Excel Regression tool.

LO2: Develop and test a multiple regression model

53) A good regression model has the fewest number of explanatory variables providing an adequate interpretation of the dependent variable.

Answer: TRUE

Diff: 1

Blooms: Remember

Topic: Building Good Regression Models

LO1: Apply a systematic approach to build good regression models.

LO2: Develop and test a multiple regression model

54) Why is regression analysis necessary in business? What categories of regression models are used?

Answer: Regression analysis is a tool for building statistical models that characterize relationships among a dependent variable and one or more independent variables, all of which are numerical. Two broad categories of regression models are used often in business settings: (1) regression models of cross-sectional data and (2) regression models of time-series data, in which the independent variables are time, or some function of time, and the focus is on predicting the future. Time-series regression is an important tool in forecasting.

Diff: 2

Blooms: Remember

Topic: Simple Linear Regression

LO1: Explain the purpose of regression analysis and provide examples in business.

LO2: Identify the components of simple linear regression models and discuss their applications

55) When are logarithmic functions used in predictive analysis?

Answer: Logarithmic functions are used when the rate of change in a variable increases or decreases quickly and then levels out, such as with diminishing returns to scale. Logarithmic functions are often used in marketing models where constant percentage increases in advertising, for instance, result in constant, absolute increases in sales.

Diff: 1

Blooms: Remember

Topic: Data-Driven Modeling

LO1: List the common types of mathematical functions used in predictive modeling.

LO2: Identify different business uses for statistics and the major statistical tools businesses use

56) While conducting regression analysis, how is constructing a normal probability plot useful?

Answer: While conducting regressions analysis using the Excel *Regression* tool, a normal probability plot for the dependent variable may be chosen to be constructed. It transforms the cumulative probability scale (vertical axis) so that the graph of the cumulative normal distribution is a straight line. The closer the points are to a straight line, the better the fit to a normal distribution.

Diff: 2

Blooms: Understand

Topic: Simple Linear Regression

LO1: Use the Excel Regression tool for both single and multiple linear regressions.

LO2: Identify the components of simple linear regression models and discuss their applications

57) Briefly explain the assumptions on which the statistical hypothesis tests associated with regression analysis are predicated.

Answer: The statistical hypothesis tests associated with regression analysis are predicated on some key assumptions about the data.

(1) Linearity: This is usually checked by examining a scatter diagram of the data or examining the residual plot. If the model is appropriate, then the residuals should appear to be randomly scattered about zero, with no apparent pattern.

(2) Normality of errors: Regression analysis assumes that the errors for each individual value of *X* are normally distributed, with a mean of zero.

(3) Homoscedasticity: This means that the variation about the regression line is constant for all values of the independent variable.

(4) Independence of errors: Residuals should be independent for each value of the independent variable.

Diff: 1

Blooms: Remember

Topic: Residual Analysis and Regression Assumptions

LO1: List the assumptions of regression analysis and describe methods to verify them.

LO2: Identify the components of simple linear regression models and discuss their applications

58) List the systematic approach to build good multiple regression models.

Answer:

(1) Construct a model with all available independent variables. Check for significance of the independent variables by examining the *p*-values.

(2) Identify the independent variable having the largest *p*-value that exceeds the chosen level of significance.

(3) Remove the variable identified in step 2 from the model and evaluate adjusted *R2*. (Don't remove all variables with *p*-values that exceed *α* at the same time, but remove only one at a time.)

(4) Continue until all variables left in the model are significant.

Diff: 1

Blooms: Remember

Topic: Building Good Regression Models

LO1: Apply a systematic approach to build good regression models.

LO2: Develop and test a multiple regression model

59) Explain the concept of curvilinear regression model.

Answer: Linear regression models are not appropriate for every situation. A scatter chart of the data might show a nonlinear relationship, or the residuals for a linear fit might result in a nonlinear pattern. In such cases, a nonlinear model is proposed to explain the relationship. For instance, a second-order polynomial model would be:

*Y = β0 + β1X + β2X2 + ε*

Sometimes this is called a curvilinear regression model. In this model, *β*1 represents the linear effect of *X* on *Y*, and *β*2 represents the curvilinear effect. However, although this model appears to be quite different from ordinary linear regression models, it is still *linear in the parameters*. In other words, all terms are a product of a beta coefficient and some function of the data, which are simply numerical values. In such cases least squares can still be applied to estimate the regression coefficients.

Diff: 2

Blooms: Understand

Topic: Regression Models with Nonlinear Terms

LO1: Identify when curvilinear regression models are more appropriate than linear models.

LO2: Develop and test a multiple regression model.