

TRUE/FALSE. Write 'T' if the statement is true and 'F' if the statement is false.

- 1) In regression, an independent variable is sometimes called a response variable. 1) _____
- 2) One purpose of regression is to understand the relationship between variables. 2) _____
- 3) One purpose of regression is to predict the value of one variable based on the other variable. 3) _____
- 4) The variable to be predicted is the dependent variable. 4) _____
- 5) The dependent variable is also called the response variable. 5) _____
- 6) A scatter diagram is a graphical depiction of the relationship between the dependent and independent variables. 6) _____
- 7) In a scatter diagram, the dependent variable is typically plotted on the horizontal axis. 7) _____
- 8) There is no relationship between variables unless the data points lie in a straight line. 8) _____
- 9) In any regression model, there is an implicit assumption that a relationship exists between the variables. 9) _____
- 10) In regression, there is random error that can be predicted. 10) _____
- 11) Estimates of the slope, intercept, and error of a regression model are found from sample data. 11) _____
- 12) Error is the difference in the actual value and the predicted value. 12) _____
- 13) The regression line minimizes the sum of the squared errors. 13) _____
- 14) In regression, a dependent variable is sometimes called a predictor variable. 14) _____
- 15) Summing the error values in a regression model is misleading because negative errors cancel out positive errors. 15) _____
- 16) The SST measures the total variability in the dependent variable about the regression line. 16) _____
- 17) The SSE measures the total variability in the independent variable about the regression line. 17) _____
- 18) The SSR indicates how much of the total variability in the dependent variable is explained by the regression model. 18) _____
- 19) The coefficient of determination takes on values between -1 and +1. 19) _____

- 20) The coefficient of determination gives the proportion of the variability in the dependent variable that is explained by the regression equation. 20) _____
- 21) The correlation coefficient has values between -1 and $+1$. 21) _____
- 22) Errors are also called residuals. 22) _____
- 23) The regression model assumes the error terms are dependent. 23) _____
- 24) The regression model assumes the errors are normally distributed. 24) _____
- 25) The errors in a regression model are assumed to have an increasing mean. 25) _____
- 26) The errors in a regression model are assumed to have zero variance. 26) _____
- 27) If the assumptions of regression have been met, errors plotted against the independent variable will typically show patterns. 27) _____
- 28) Often, a plot of the residuals will highlight any glaring violations of the assumptions. 28) _____
- 29) The error standard deviation is estimated by MSE. 29) _____
- 30) The standard error of the estimate is also called the variance of the regression. 30) _____
- 31) An F -test is used to determine if there is a relationship between the dependent and independent variables. 31) _____
- 32) The null hypothesis in the F -test is that there is a linear relationship between the X and Y variables. 32) _____
- 33) If the significance level for the F -test is high enough, there is a relationship between the dependent and independent variables. 33) _____
- 34) When the significance level is small enough in the F -test, we can reject the null hypothesis that there is no linear relationship. 34) _____
- 35) The coefficients of each independent variable in a multiple regression model represent slopes. 35) _____
- 36) For statistical tests of significance about the coefficients, the null hypothesis is that the slope is 1. 36) _____
- 37) Both the p -value for the F -test and r^2 can be interpreted the same with multiple regression models as they are with simple linear models. 37) _____
- 38) The multiple regression model includes several dependent variables. 38) _____
- 39) In regression, a binary variable is also called an indicator variable. 39) _____
- 40) Another name for a dummy variable is a binary variable. 40) _____

- 41) The best model is a statistically significant model with a high r -square and few variables. 41) _____
- 42) The adjusted r^2 will always increase as additional variables are added to the model. 42) _____
- 43) The value of r^2 can never decrease when more variables are added to the model. 43) _____
- 44) A variable should be added to the model regardless of the impact (increase or decrease) on the adjusted r^2 value. 44) _____
- 45) Transformations may be used when nonlinear relationships exist between variables. 45) _____
- 46) A high correlation always implies that one variable is causing a change in the other variable. 46) _____
- 47) A dummy variable can be assigned up to three values. 47) _____

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

- 48) Which of the following statements is true regarding a scatter diagram? 48) _____
- A) It gives the percent of variation in the dependent variable that is explained by the independent variable.
 - B) It provides very little information about the relationship between the regression variables.
 - C) It has a value between -1 and +1.
 - D) It is a line chart of the independent and dependent variables.
 - E) It is a plot of the independent and dependent variables.
- 49) The random error in a regression equation 49) _____
- A) is the predicted error.
 - B) includes both positive and negative terms.
 - C) is maximized in a least squares regression model.
 - D) will sum to a large positive number.
 - E) is used to estimate the accuracy of the slope.
- 50) Which of the following equalities is correct? 50) _____
- A) $SSR = SST + SSE$
 - B) $SST = SSC + SSR$
 - C) $SST = SSR + SSE$
 - D) $SSE = SSR + SST$
 - E) $SSE = \text{Actual Value} - \text{Predicted Value}$
- 51) The sum of squared error (SSE) is 51) _____
- A) a measure of the total variation in Y about the mean.
 - B) a measure in the variation of Y about the regression line.
 - C) a measure in the variation of X about the regression line.
 - D) a measure of the total variation in X about the mean.
 - E) None of the above

52) If computing a causal linear regression model of $Y = a + bX$ and the resultant r^2 is very near zero, then one would be able to conclude that 52) _____

A) a multiple linear regression model is a good forecasting method for the data.
 B) $Y = a + bX$ is not a good forecasting method.
 C) $Y = a + bX$ is a good forecasting method.
 D) a multiple linear regression model is not a good forecasting method for the data.
 E) None of the above

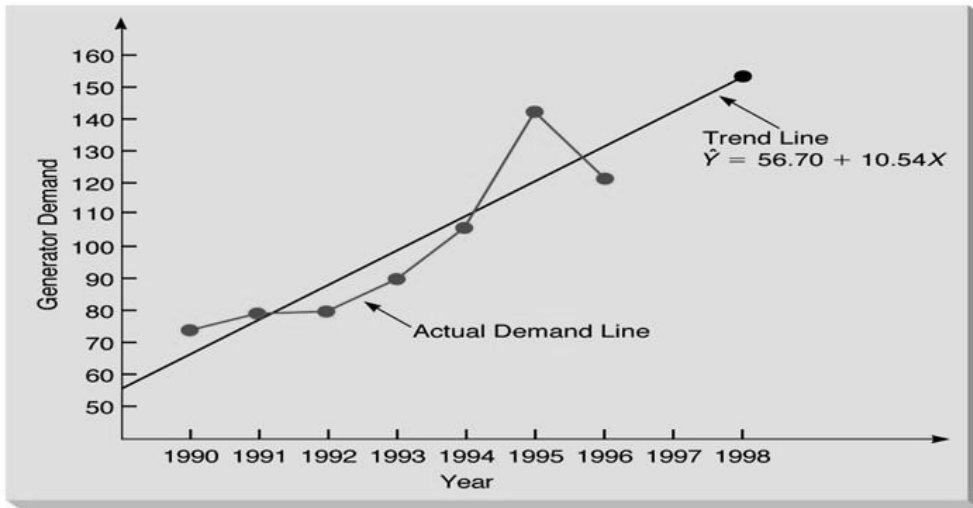
53) Which of the following statements is true about r^2 ? 53) _____

A) It represents the percent of variation in X that is explained by Y .
 B) It ranges in value from -1 to $+1$.
 C) It is also called the coefficient of determination.
 D) It is also called the coefficient of correlation.
 E) It represents the percent of variation in the error that is explained by Y .

54) The coefficient of determination resulting from a particular regression analysis was 0.85 . What was the slope of the regression line? 54) _____

A) 0.922
 B) -0.85
 C) 0.85
 D) There is insufficient information to answer the question.
 E) None of the above

55) The diagram below illustrates data with a 55) _____



- A) negative correlation coefficient.
 B) positive correlation coefficient.
 C) zero correlation coefficient.
 D) correlation coefficient equal to $+1$.
 E) None of the above

- 56) The correlation coefficient resulting from a particular regression analysis was 0.25. What was the coefficient of determination? 56) _____
- A) -0.5
 - B) 0.0625
 - C) 0.5
 - D) There is insufficient information to answer the question.
 - E) None of the above
- 57) The coefficient of determination resulting from a particular regression analysis was 0.85. What was the correlation coefficient, assuming a positive linear relationship? 57) _____
- A) 0.922
 - B) 0.5
 - C) -0.5
 - D) There is insufficient information to answer the question.
 - E) None of the above
- 58) In a good regression model the residual plot shows 58) _____
- A) a cone pattern.
 - B) an arched pattern.
 - C) an increasing pattern.
 - D) a decreasing pattern.
 - E) a random pattern.
- 59) Which of the following represents the underlying linear model for hypothesis testing? 59) _____
- A) $Y = b_0 + b_1 X + \varepsilon$
 - B) $Y = b_0 + b_1 X$
 - C) $Y = \beta_0 + \beta_1 X$
 - D) $Y = \beta_0 + \beta_1 X + \varepsilon$
 - E) None of the above
- 60) Which of the following statements is false concerning the hypothesis testing procedure for a regression model? 60) _____
- A) The null hypothesis is rejected if the adjusted r^2 is above the critical value.
 - B) The alternative hypothesis is that the true slope coefficient is not equal to zero.
 - C) The null hypothesis is that the true slope coefficient is equal to zero.
 - D) An α level must be selected.
 - E) The F -test statistic is used.
- 61) Suppose that you believe that a cubic relationship exists between the independent variable (of time) and the dependent variable Y . Which of the following would represent a valid linear regression model? 61) _____
- A) $Y = b_0 + 3b_1 X$, where $X = \text{time}$
 - B) $Y = b_0 + b_1 X$, where $X = \text{time}^{1/3}$
 - C) $Y = b_0 + b_1 X^3$, where $X = \text{time}$
 - D) $Y = b_0 + 3b_1 X$, where $X = \text{time}^3$
 - E) $Y = b_0 + b_1 X$, where $X = \text{time}^3$

- 62) A prediction equation for starting salaries (in \$1,000s) and SAT scores was performed using simple linear regression. In the regression printout shown below, what can be said about the level of significance for the overall model? _____

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.935018123
R Square	0.87425889
Adjusted R Square	0.860287655
Standard Error	3.307295949
Observations	11

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	684.4652324	684.4652	62.57564	2.42144E-05
Residual	9	98.44385847	10.93821		
Total	10	782.9090909			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>
Intercept	-29.14060468	8.660080997	-3.36493	0.008324	-48.73108387
SAT	0.065443841	0.008273059	7.910476	2.42E-05	0.046728866

- A) The significance level for SAT indicates the slope is equal to zero.
- B) SAT is not a good predictor for starting salary.
- C) The significance level for SAT indicates the slope is not equal to zero.
- D) The significance level for the intercept indicates the model is not valid.
- E) None of the above

ESSAY. Write your answer in the space provided or on a separate sheet of paper.

- 63) An air conditioning and heating repair firm conducted a study to determine if the average outside temperature could be used to predict the cost of an electric bill for homes during the winter months in Houston, Texas. The resulting regression equation was:

$$Y = 227.19 - 1.45X, \text{ where } Y = \text{monthly cost, } X = \text{average outside air temperature}$$

- (a) If the temperature averaged 48 degrees during December, what is the forecasted cost of December's electric bill?
- (b) If the temperature averaged 38 degrees during January, what is the forecasted cost of January's electric bill?

- 64) Bob White is conducting research on monthly expenses for medical care, including over-the-counter medicine. His dependent variable is monthly expenses for medical care while his independent variable is number of family members. Below is his Excel output.

SUMMARY OUTPUT					
<i>Regression Statistics</i>					
Multiple R		0.695			
R Square		0.483			
Adjusted R Square		0.474			
Standard Error		55.278			
Observations		21			
<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	99929.47	99929.47	32.70311	0.00016
Residual	19	58057.48	3055.657		
Total	20	157987			
<i>Coefficients</i>					
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	
Intercept	110.47	35.53407	3.109087	0.0057	
X	16.83	0.819024	5.718663	0.00016	

- (a) What is the prediction equation?
 (b) Based on his model, each additional family member increases the predicted costs by how much?
 (c) Based on the significance F -test, is this model a good prediction equation?
 (d) What percent of the variation in medical expenses is explained by the size of the family?
 (e) Can the null hypothesis that the slope is zero be rejected? Why or why not?
 (f) What is the value of the correlation coefficient?
- 65) Consider the regression model $Y = 389.10 - 14.6X$. If the r^2 value is 0.657, what is the correlation coefficient?

66) Bob White is conducting research on monthly expenses for medical care, including over the counter medicine. His dependent variable is monthly expenses for medical care while his independent variables are number of family members and insurance type (government funded, private insurance and other). He has coded insurance type as the following:

$X_2 = 1$ if government funded, $X_3 = 1$ if private insurance

Below is his Excel output.

<i>Regression Statistics</i>					
Multiple R		0.8590			
R Square		0.7379			
Adjusted R Square		0.7182			
Standard Error	952.3605				
Observations	18				

ANOVA	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	1.32E+08	43977657	48.48745	1.21E-07
Residual	14	12697866	906990.4		
Total	17	1.45E+08			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	144.91	1025.911	0.141246	0.889688
X1	11.63	1.247247	9.330762	2.19E-07
X2	-13.70	8.786907	-1.55916	0.141272
X3	-9.11	1.166068	7.810781	1.81E-06

- What is the prediction equation?
- Based on the significance F -test, is this model a good prediction equation?
- What percent of the variation in medical expenses is explained by the independent variables?
- Based on his model, what are the predicted monthly expenses for a family of four with private insurance?
- Based on his model, what are the predicted monthly expenses for a family of two with government funded insurance?
- Based on his model, what are the predicted monthly expenses for a family of five with no insurance?

Answer Key

Testname: CHAPTER 4

- 1) FALSE
- 2) TRUE
- 3) TRUE
- 4) TRUE
- 5) TRUE
- 6) TRUE
- 7) FALSE
- 8) FALSE
- 9) TRUE
- 10) FALSE
- 11) FALSE
- 12) TRUE
- 13) TRUE
- 14) FALSE
- 15) TRUE
- 16) FALSE
- 17) FALSE
- 18) TRUE
- 19) FALSE
- 20) TRUE
- 21) TRUE
- 22) TRUE
- 23) FALSE
- 24) TRUE
- 25) FALSE
- 26) FALSE
- 27) FALSE
- 28) TRUE
- 29) FALSE
- 30) FALSE
- 31) TRUE
- 32) FALSE
- 33) FALSE
- 34) TRUE
- 35) TRUE
- 36) FALSE
- 37) TRUE
- 38) FALSE
- 39) TRUE
- 40) TRUE
- 41) TRUE
- 42) FALSE
- 43) TRUE
- 44) FALSE
- 45) TRUE
- 46) FALSE
- 47) FALSE
- 48) E
- 49) B
- 50) C

Answer Key

Testname: CHAPTER 4

- 51) B
- 52) B
- 53) C
- 54) D
- 55) B
- 56) B
- 57) A
- 58) E
- 59) D
- 60) A
- 61) E
- 62) C
- 63) (a) $\$227.19 - \$1.45(48) = \$157.59$
(b) $\$227.19 - \$1.45(38) = \$172.09$
- 64) (a) $Y = 110.47 + 16.83X$
(b) \$16.83
(c) Yes, because the p -value for the F -test is low.
(d) 48.3% of the variation in medical expenses is explained by family size.
(e) The null hypothesis can be rejected, the slope is not equal to zero based on the low p -value.
(f) 0.695
- 65) $-(0.657)^{1/2} = -0.811$
- 66) (a) $Y = 144.91 + 11.63X_1 - 13.70 X_2 - 9.11X_3$
(b) The model is a good prediction equation because the significance level for the F -test is low.
(c) 73.79 percent of the variation in medical expenses is explained by family size and insurance type.
(d) \$182.32
(e) \$154.47
(f) \$203.06