ADMS2320.com

We Make Stats Easy.

Chapter 17

Tutorial Length
1 Hour 55 Mins

Chapter 17 Note

This chapter extends on simple linear regression to *multiple* regression analysis.

- Multiple regression model & equation
- Testing Validity of the Model
- Testing of Coefficients
- Adjusted Coefficient of Determination
- ANOVA table and Excel Output Analysis
- Multicollinearity
- Error conditions

Please download the formula sheet before starting. This formula sheet is from the previous semester. It is recommended you check your course page for an updated posting on the formula sheet so that you can study with the most up to date version in case the one on this website is outdated.

How to use ADMS3330.com

The goal of this website is to quickly and efficiently make you understand all the **major concepts**. We have many years of experience tutoring students for ADMS 3330 and we have a pretty good idea of what you need to concentrate on in order to do well.

Each chapter, we will go through theory and then apply that theory through multiple choice and short answer questions. We attempt to make our videos quick and to the point. As such, it is a good idea to always *review your class notes and textbook readings* to go over additional topics and examples.

Note: Not all theory is covered prior to answering practice questions. Some theory is taught while answering questions in order to put the theory into immediate practice. Make sure to go through all the videos in a given chapter.

Multiple Regression

Multiple Regression vs Simple Linear Regression

In simple linear regression, we try to determine if a relationship exists between a single independent variable x and the dependent variable y. If a relationship exists, we can create a regression equation to make predictions, $\hat{y}$=$b_0 + b_1 x$.

In multiple linear regression, we try to determine if a relationship exists between multiple independent variables ($x_1, x_2, x_3, \ldots$) and the dependent variable.

Multiple Regression Model

If there are k independent variables related to the dependent variable, then the model is:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + \epsilon$

β_0 is the y-intercept (population parameter)
$\beta_1, \beta_2, \ldots \beta_k$ are the coefficients of the independent variables (population parameter)
ϵ is the error variable

The required conditions for the error variable discussed in the simple linear regression section still apply here.

Multiple Regression Equation

This equation is created from sample data and used to make predictions.

$\hat{y}$=$b_0 + b_1 x_1 + b_2 x_2 + \cdots b_k x_k$

b_0: y-intercept, the predicted value when all independent variables are 0. However, we usually don't try to interpret the intercept. Having all the independent variables be 0 is typically outside the range of the sample data.

$b_\#$ interpretation: For every unit increase in $x_\#$, y will increase/decrease on average by $|b_\#|$, if all other variables are held constant.

Note: We were able to do manual calculations for the slope and intercept for simple linear regression. Due to the complexity of the calculations for multiple regression, we will be interpreting data outputs (usually created in Excel), and relatively simple calculations based on that output.

Regression Example

Example

Life insurance companies are keenly interested in predicting how long their customers will live because their premiums and profitability depend on such numbers. An actuary for one insurance company gathered data from 100 recently deceased male customers. He recorded the age at death of the customer plus the ages at death of his mother and father, the mean ages at death of his grandmothers, and the mean ages at death of his grandfathers. A regression analysis was performed and the output is provided below.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.86084275
R Square	0.74105023
Adjusted R Square	0.73014708
Standard Error	2.66407493
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1929.516954	482.379238	67.96663	4.85786E-27
Residual	95	674.2430463	7.09729522		
Total	99	2603.76			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	3.24382122	5.423411981	0.59811448	0.55118683
Mother	0.45085829	0.054501503	8.27240099	8.0028E-13
Father	0.41118348	0.049788304	8.258636	8.5583E-13
Gmothers	0.016553	0.066107372	0.25039567	0.80282209
Gfathers	0.08685832	0.065656685	1.32291664	0.18903865

Regression Example

A) What is the population regression model

B) What is the sample regression model

C) What is the predicted age of death for a male whose father died at age 64, mother died at age 72, grandmothers died at a mean age of 68, and grandfathers died at a mean age of 70.

D) Interpret the coefficient for Mother

Regression Statistics

Standard Error of Estimate

$$s_\epsilon = \sqrt{\frac{SSE}{n-k-1}}$$

n is sample size
k is number of independent variables

s_ϵ can be thought of as the average error in a predicted value. To determine if a standard error of estimate is large or small, we compare against $\bar{y}$, that is, the average of the dependent variable. s_ϵ of 1.25 versus a mean of 60 is small, but versus a mean of 3 might be large.

Coefficient of Determination

$$R^2 = 1 - \frac{SSE}{\sum(y_i - \bar{y})^2} = 1 - \frac{SSE}{SST} \qquad R^2 = \frac{SSR}{SST}$$

Recall: $SST = \sum(y_i - \bar{y})^2$

This number represents the amount (%) of variation in the dependent variable (y) that can be explained by the variation in the independent variables ($x_1, x_2, \ldots x_k$).

Adjusted Coefficient of Determination

$$Adjusted\ R^2 = 1 - \frac{\left(\frac{SSE}{n-k-1}\right)}{\left(\frac{\sum(y_i - \bar{y})^2}{n-1}\right)} = 1 - \left(\frac{MSE}{s_y^2}\right)$$

$$Adjusted\ R^2 = 1 - \left(\frac{n-1}{n-k-1}\right)\left(1 - R^2\right)$$

This number represents the amount (%) of variation in the dependent variable (y) that can be explained by the variation in the independent variable (x), *adjusted to take into account sample size and number of independent variables..*

Note

1) Adjusted R^2 < R^2
2) Adjusted R^2 has no bounds, it can be positive or negative

Regression Example

Example

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.86084275
R Square	0.74105023
Adjusted R Square	0.73014708
Standard Error	2.66407493
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1929.516954	482.379238	67.96663	4.85786E-27
Residual	95	674.2430463	7.09729522		
Total	99	2603.76			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	3.24382122	5.423411981	0.59811448	0.55118683
Mother	0.45085829	0.054501503	8.27240099	8.0028E-13
Father	0.41118348	0.049788304	8.258636	8.5583E-13
Gmothers	0.016553	0.066107372	0.25039567	0.80282209
Gfathers	0.08685832	0.065656685	1.32291664	0.18903865

E) What is the coefficient of determination and what does it mean?

F) Why is the adjusted R^2 different from R^2?

Testing Validity of the Model

In simple linear regression, there was a t-test for the slope to determine if a linear relationship existed between the independent and dependent variable. This same test determined if the model was valid.

For multiple regression, we have multiple independent variables however, so we use the F-test for validity, which incorporates techniques from ANOVA (Analysis of Variance). This test checks if there is an **overall** relationship between the independent variables and the dependent variable.

Hypothesis

$H_0: \beta_1 = \beta_2 = \cdots \beta_k = 0$
H_1: At least one β_i is not equal to 0

Test Statistic

$F = MSR/MSE$

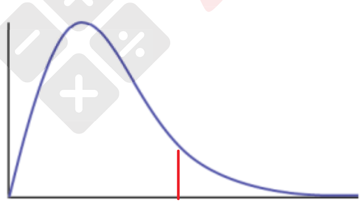
ANOVA TABLE FOR REGRESSION

Source Of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F-Statistic
Regression	k	SSR	$MSR = \dfrac{SSR}{k}$	$F = \dfrac{MSR}{MSE}$
Residual	$n-k-1$	SSE	$MSE = \dfrac{SSE}{n-k-1}$	
Total	$n-1$	$SSTotal = \sum (y_i - \bar{y})^2$		

The null hypothesis is stating that none of the independent variables are linearly related to the dependent variable, thus the model is not valid. The alternative hypothesis is saying that if at least one of the coefficients is not 0, then there is some validity of the model.

This is always a right-tailed F-test. In general, if F is large, that means most of the variation in y is explained by the regression equation, and therefore the model is valid. Alternatively, if F is small, most of the variation in y is not explained by the regression equation, and therefore the model is invalid.

Rejection Regions *For F* Distribution



$v_1 = k$
$v_2 = n - k - 1$

$\alpha = 0.05$
(unless another level of significance is stated in the question)

$F > F_{\alpha, v_1, v_2}$
$F > F_{\alpha, k, n-k-1}$

Example

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.86084275
R Square	0.74105023
Adjusted R Square	0.73014708
Standard Error	2.66407493
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1929.516954	482.379238	67.96663	4.85786E-27
Residual	95	674.2430463	7.09729522		
Total	99	2603.76			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	3.24382122	5.423411981	0.59811448	0.55118683
Mother	0.45085829	0.054501503	8.27240099	8.0028E-13
Father	0.41118348	0.049788304	8.258636	8.5583E-13
Gmothers	0.016553	0.066107372	0.25039567	0.80282209
Gfathers	0.08685832	0.065656685	1.32291664	0.18903865

G) Test to determine the validity of the model

Hypothesis	
Rejection Region	
Test Statistic	
Conclusion	

Test For Coefficients

We can also check individual independent variables using the t-test to see if a linear relationship exists. Why not just do multiple t-tests to start with? Each time you do a t-test, there is a probability of a type 1 error (α). So do an F-test to check overall validity, and then a t-test to check if the individual independent variables have a linear relationship with the dependent variable.

Test For Coefficients

Hypothesis

H_0: $\beta_i = 0$
H_1: $\beta_i < 0$ **OR**
$\beta_i \neq 0$ **OR**
$\beta_i > 0$

For $i = 1,2, \ldots, k$

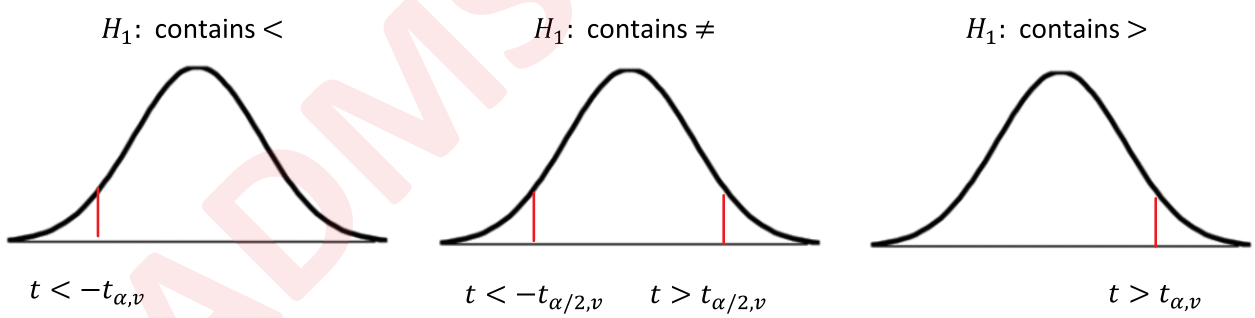
Test Statistic

$$t = \frac{(b_i - \beta_i)}{s_{b_i}}$$

$$= \frac{b_i}{s_{b_i}}$$

The null hypothesis is that there is no linear relationship, thus the line of best fit is horizontal with a slope of 0. The alternative hypothesis is there is some linear relationship, whereby the slope must therefore not be 0. Both 2 tailed and 1 tailed tests are possible. $\beta_i > 0$ if we want to see if there's a positive relationship, $\beta_i < 0$ if we want to see if there's a negative relationship, or $\beta_i \neq 0$ if we just want to see if any linear relationship exists.

Rejection Regions *For t* Distribution



$v = n - k - 1$
v represents degrees of freedom (also referred to as df)
n represents the number of pairs of sample data
k is the number if independent variables

Regression Example

Example

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.86084275
R Square	0.74105023
Adjusted R Square	0.73014708
Standard Error	2.66407493
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1929.516954	482.379238	67.96663	4.85786E-27
Residual	95	674.2430463	7.09729522		
Total	99	2603.76			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	3.24382122	5.423411981	0.59811448	0.55118683
Mother	0.45085829	0.054501503	8.27240099	8.0028E-13
Father	0.41118348	0.049788304	8.258636	8.5583E-13
Gmothers	0.016553	0.066107372	0.25039567	0.80282209
Gfathers	0.08685832	0.065656685	1.32291664	0.18903865

H) Test to determine if the mothers age has a positive relationship with age at death.

1.	Hypothesis	
2.	Rejection Region	
3.	Test Statistic	
4.	Conclusion	

Regression Example

Example

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.86084275
R Square	0.74105023
Adjusted R Square	0.73014708
Standard Error	2.66407493
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1929.516954	482.379238	67.96663	4.85786E-27
Residual	95	674.2430463	7.09729522		
Total	99	2603.76			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	3.24382122	5.423411981	0.59811448	0.55118683
Mother	0.45085829	0.054501503	8.27240099	8.0028E-13
Father	0.41118348	0.049788304	8.258636	8.5583E-13
Gmothers	0.016553	0.066107372	0.25039567	0.80282209
Gfathers	0.08685832	0.065656685	1.32291664	0.18903865

I) Is there evidence to infer that fathers age is related to age at death?

1.	Hypothesis	
2.	Rejection Region	
3.	Test Statistic	
4.	Conclusion	

Regression Output - Fill In The Blanks

Source Of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F-Statistic
Regression	k	SSR	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE}$
Residual	$n-k-1$	SSE	$MSE = \frac{SSE}{n-k-1}$	
Total	$n-1$	$SSTotal = \sum(y_i - \bar{y})^2$		

$$R^2 = 1 - \frac{SSE}{\sum(y_i - \bar{y})^2} = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}$$

$$Adjusted\ R^2 = 1 - \frac{\left(\frac{SSE}{n-k-1}\right)}{\left(\frac{\sum(y_i - \bar{y})^2}{n-1}\right)} = 1 - \left(\frac{MSE}{s_y^2}\right)$$

$$Adjusted\ R^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1 - R^2)$$

$$s_\epsilon = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{MSE}$$

$$t = \frac{b_i - \beta_i}{s_{b_i}} = \frac{b_i}{s_{bi}} \quad because\ \beta_i = 0$$

Regression Output - Fill In The Blanks

Example Determine A though I

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.86084275
R Square	**A**
Adjusted R Square	**B**
Standard Error	**C**
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	**D**	1929.516954	**G**	**H**	4.85786E-27
Residual	**E**	**F**	7.09729522		
Total	99	2603.76			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	3.24382122	5.423411981	0.59811448	0.55118683
Mother	0.45085829	0.054501503	**I**	8.0028E-13
Father	**J**	0.049788304	8.258636	8.5583E-13
Gmothers	0.016553	**K**	0.25039567	0.80282209
Gfathers	0.08685832	0.065656685	1.32291664	0.18903865

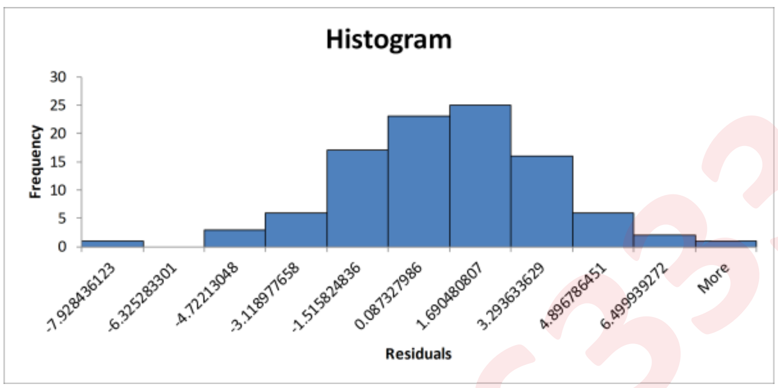
Error Variable - Required Conditions

Error Variable – Required Conditions

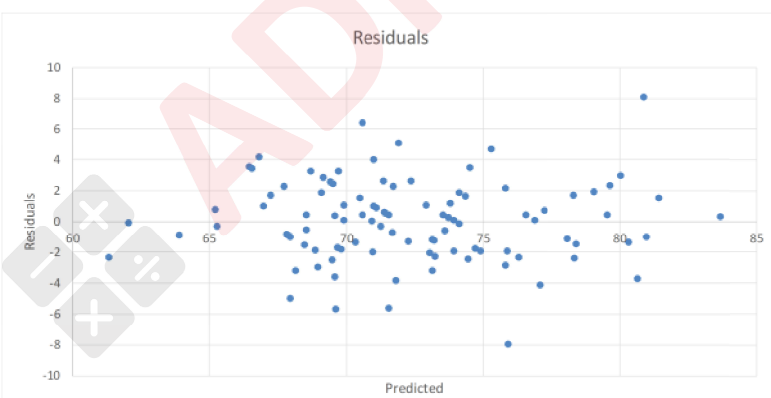
1. ϵ is normally distributed
2. The mean is 0 (expected value is 0) of the distribution of the error variable
3. The standard deviation of ϵ is σ_ϵ, which is always a constant for each x.
   This condition can also be thought of as the error variance being constant.
4. The value of ϵ for any particular value of y is independent from ϵ for any other value of y

Condition 1 can be checked by plotting the residuals and checking if the residual plot is approximately normal. If there is an extreme deviation from normality, then the condition has been violated. (This indicates minor non-normality is fine). The plot would be residuals on the x-axis and frequency on the y-axis.

Condition 2 can also be checked from this plot as well by seeing if the data is centered around 0.



Condition 3 can be checked by plotting residuals (y-axis) versus predicted values (x-axis). If there are no patterns, then we're good and we have homoscedasticity. If there is a pattern, then there is heteroscedasticity, which means the condition is violated.



Condition 4 can be checked by plotting the residuals versus time periods. If there is a pattern, then the errors are not independent.

Multicollinearity

Multicollinearity/Collinearity/Intercorrelation

Multicollinearity exists when the independent variables are **highly** correlated (values > 0.6, or 60%) with one another. When multicollinearity exists, T-tests might indicate some linearly independent variables are not linearly related to the dependent variable when in reality they actually are. (The F-test however is not affected by multicollinearity). If Multicollinearity exists, the coefficient interpretations are meaningless since the coefficients require all other variables to be held constant, but if there is collinearity, then changing one will probably affect the other. This implies the coefficients are quite not good representations for the population parameters.

Multicollinearity almost always exists to some extent. All we can do is to try and make sure we include variables that are independent of each other when doing regression analysis. Multicollinearity is only a problem in multiple regression models, not in simple regression.

Multicollinearity can be resolved by redeveloping the model or by applying transformations. Nonnormality and heteroscedasticity can also be resolved to by applying transformations.

Example

Consider the following correlation matrix. Is there evidence suggesting multicollinearity exists?

	X1	*X2*	*X3*	*X4*
X1	1			
X2	0.14	1		
X3	0.43	0.19	1	
X4	0.02	0.12	0.21	1

Example

Consider the following correlation matrix. Is there evidence suggesting multicollinearity exists?

	X1	*X2*	*X3*
X1	1		
X2	0.87	1	
X3	0.92	0.05	1

Multiple Choice

The following three questions (Q1-3) are based on the following partial multiple regression results.

Source	Degrees of freedom	Sum of Squares
Regression	5	750
Error	30	500

1. The total sample size is:

   a. 35
   b. 36
   c. 30
   d. Not enough information

2. The number of independent variables is:

   a. 6
   b. 36
   c. 5
   d. Not enough information

3. The total variation in y is:

   a. 500
   b. 1250
   c. 250
   d. Not enough information

4. Which of the following is an assumption of the regression model?

   a. The errors are independent.
   b. The errors are not normally distributed.
   c. The errors have a standard deviation of zero.
   d. The errors have an irregular variance.
   e. The errors follow a cone pattern.

5. The standardized residual is defined as:

   a. residual divided by the standard error of estimate
   b. residual multiplied by the square root of the standard error of estimate
   c. residual divided by the square of the standard error of estimate
   d. residual multiplied by the standard error of estimate

6. When the variance of the error variable is a constant no matter what the value of *x* is, this condition is called:

   a. nonnormality
   b. heteroscedasticity
   c. homoscedasticity
   d. autocorrelation

7. In a regression model involving 30 observations, the following estimated regression model is: $\hat{y} = 60 + 2.8x_1 + 1.2x_2 - x_3$ For this model, total variation in Y = SST = 800 and SSE = 200. The value of the *F* statistic for testing the validity of this model is:

a. 26.00
b. 3.38
c. 7.69
d. 0.039
e. None of the above

8. In a regression model involving 50 observations, the estimated regression model is $\hat{y} = 10.5 + 3.2x_1 + 5.8x_2 + 6.5x_3$ For this model, SSR = 450 and SSE = 175. The value of MSE is:

a. 12.5035
b. 150
c. 275
d. 3.804
e. None of the above

9. Which of the following measures can be used to assess the multiple regression model's fit?

a. sum of squares for error
b. sum of squares for regression
c. single t-test
d. none of the above

10. In a multiple regression model, the standard deviation of the error variable ϵ is assumed to be:

a. constant.
b. 0.
c. 1.0.
d. None of these choices.

11. In a multiple regression analysis involving 40 observations and 5 independent variables, the following statistics are given: Total variation in y = 350 and SSE = 50. Then, the coefficient of determination is:

a. 0.8408
b. 0.8571
c. 0.8469
d. 0.8529
e. None of the above

12. In order to test the validity of a multiple regression model involving 5 independent variables and 30 observations, the numerator and denominator degrees of freedom for the critical value of F are, respectively,

a. 5 and 30
b. 6 and 29
c. 5 and 24
d. 6 and 25

Multiple Choice

13. In a multiple regression analysis, there are 20 data points and 5 independent variables, and the sum of the squared differences between observed and predicted values of y is 450. The standard error of estimate will be:

a. 4.830
b. 5.669
c. 5.477
d. 3.464

14. A regression model involves 8 independent variables and 110 observations. The critical value for testing the significance of each of the independent variable's coefficients will have

a. 101 degrees of freedom
b. 102 degrees of freedom
c. 109 degrees of freedom
d. 8 degrees of freedom

15. In a multiple regression analysis involving 10 independent variables and 200 observations, SST= 500 and SSE = 150. The coefficient of determination is

a. 0.500
b. 0.700
c. 0.300
d. 0.192

16. A multiple regression model has:

a. only one independent variable.
b. only two independent variables.
c. more than one independent variable.
d. more than one dependent variable.

17. The model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$ is referred to as a:

a. simple linear regression model
b. first-order model with one predictor variable
c. second-order model with one predictor variable
d. third order model with two predictor variables

18. A multiple regression analysis involving three independent variables and 25 data points results in a value of 0.769 for the unadjusted multiple coefficient of determination. Then, the adjusted multiple coefficient of determination is:

a. 0.385
b. 0.877
c. 0.591
d. 0.736
e. None of the above

19. In a regression model involving 50 observations, the following estimated regression model was obtained: $\hat{y} = 10.5 + 3.2x_1 + 5.8x_2 + 6.5x_3$. For this model, the following statistics are given: SSR = 450 and SSE = 175. Then, the value of MSR is:

a. 12.50
b. 275
c. 150
d. 3.804

20. For the multiple regression model: $\hat{y} = 45 + 25x_1 - 15x_2 + 10x_3$, if x_2 were to increase by 5, holding other variables constant, the value of y will:

a. Increase by 75
b. Increase by 45
c. Decrease on average by 75
d. Decrease on average by 45

21. In a multiple regression model, the adjusted R^2

a. Cannot be negative
b. Can sometimes be negative
c. Can sometimes be greater than +1
d. Has to fall between 0 and + 1

22. A human resources analyst is developing a regression model to predict electricity production plant manager compensation as a function of production capacity of the plant, number of employees at the plant, and plant technology (coal, oil, and nuclear). The dependent variable in this model is ______.

a. plant manager compensation
b. plant capacity
c. number of employees
d. plant technology
e. nuclear

23. A regression model involves two independent variables and 30 observations. The critical value of the test statistic for testing if the coefficient β_i of the independent variable x_i is positive at the 1% level of significance will be

a. 2.763
b. 2.771
c. 2.467
d. 2.473
e. None of the above

24. A multiple regression analysis includes 20 data points and 4 independent variables produced the following statistics: Total variation in Y = SST = 200 and SSR = 160. Then, the multiple standard error of estimate will be:

a. 0.80
b. 3.266
c. 3.651
d. 1.633

Multiple Choice

25. In a multiple regression model, the error variable ϵ is assumed to have a mean of:
   a. –1.0
   b. 0.0
   c. 1.0
   d. Any value smaller than –1.0
   e. None of the above

26. Which of the following statements regarding multicollinearity is not true?

   a. It exists in virtually all multiple regression models.
   b. It is also called collinearity and intercorrelation.
   c. It is a condition that exists when the independent variables are highly correlated with one another
   d. It does not affect the f-test of the analysis of variance.
   e. None of the above

27. An appropriate method to identify multicollinearity in a regression model is to ____.

   a. examine a residual plot
   b. examine the ANOVA table
   c. examine a correlation matrix
   d. examine the partial regression coefficients
   e. examine the *R*2 of the regression model

28. A regression model involves 9 independent variables and 110 observations. The test statistic for testing the significance of the coefficient β_i of the independent variable x_i at the 1% level of significance will have

   a. 101 degrees of freedom
   b. 100 degrees of freedom
   c. 109 degrees of freedom
   d. 9 degrees of freedom
   e. None of the above

29. A regression model involves two independent variables and 30 observations. The critical value of the test statistic for testing if the coefficient β_i of the independent variable x_i is positive at the 1% level of significance will be

   a. 2.763
   b. 2.771
   c. 2.467
   d. 2.473
   e. None of the above

Multiple Choice

<u>Questions #30 through #33</u>. Given the following multiple regression output based on a sample of 50 observations involving the severance pay of employees laid off by North York Manufacturing Company (NYMC). The model under consideration is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$ where y =number of weeks of severance pay, x_1=age of employee, x_2 =employee's number of years with NYMC, and x_3 =employee's annual pay (in thousand dollars). Regression analysis output follows:

SUMMARY OUTPUT					
Regression Statistics					
Multiple R	0.8378				
R Square	0.7020				
Adjusted R Square	0.6825				
Standard Error	1.9210				
Observations	50				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	399.8602	133.2867	36.1169	0.0000
Residual	46	169.7598	3.6904		
Total	49	569.62			
	Coefficients	*Standard Error*	*t Stat*	*P-value*	
Intercept	6.0611	2.6040	2.3276	0.0244	
Age	-0.0078	0.0664	-0.1175	0.9069	
Years	0.6035	0.0966	6.2498	0.0000	
Pay	-0.0702	0.0524	-1.3413	0.1864	

30. Which of the independent variables is positively related to the dependent variable at the 1% level of significance?
   a. x_1
   b. x_2
   c. Both x_1 and x_3
   d. All of the independent variables
   e. None of the independent variables

31. The critical value of the test statistic for testing the validity of the model at the 5% level of significance is approximately
   a. 2.01
   b. 2.81
   c. 1.62
   d. 2.79
   e. None of the above

32. Which of the following is an appropriate hypothesis that should be rejected in support of the validity of the above multiple regression model?
   a. $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
   b. $\beta_i \neq 0$ for at least one i=1,2,3
   c. $\beta_i \neq 0$ for every i=1,2,3
   d. $\beta_i \neq 0$ for every i=0,1,2,3
   e. None of the above

33. If the sample regression equation is accepted as valid and accordingly used, what will be the predicted severance pay, approximately, for a 36-year-old NYMC employee who has worked for the company for 10 years and receives an annual pay of $32,000
   a. $9,570
   b. $2241
   c. 9.6 weeks' pay
   d. 22 weeks' pay
   e. None of the above

Next Steps

Proceed to practice section for this chapter OR proceed to past test questions and see if you can pick out which questions belong to this section. We recommend doing a few problems from the practice section first before moving on to practice test questions.