Multiple Regression and Modelling (Statistics for Management., Richard I. Levin., David S. Rubin., Prentice Hall ., Seventh Edition., p. 783-787)

A. True or False

1. The principal advantage of multiple regression over simple regression is that it allows us to use more of the information available to estimate the dependent variable (True)
2. Suppose, in the multiple regression equation $\hat{Y} = 24.4 + 5.6X_1 + 6.8X_2$, $\hat{Y}$ stands for weight (in pounds) and X_2stands for age (in years). For each additional year of age, then, it can be expected that weight will increase by 24.4 pounds (False – 6.8 pounds)
3. Although it is theoretically possible to do multiple regression calculations by hand, we seldom do so (True)
4. Suppose you are attempting to form a confidence interval for a value of Y from a multiple regression equation. If there are 20 elements in the sample and 4 independent variables are used in the regression, you should use 16 degrees of freedom when you get a value from t table (False – 15 degrees of freedom)
5. The standard error of the coefficient b_2 in a multiple regression is denoted s_2 (False – S_{b2})
6. Suppose we wish to test whether tha values of Y in multiple regression really depend on the values of X_1 . The null hypothesis for our test will be $\beta_1 = 0$ (True)
7. To determine whether a regression is significant as a whole, an observed value of F is calculated and compared to a value from a table (True)
8. If one knows the total sum of squares and regression sum of squares for a multiple regression, the error sum of square can always be quickly calculated (True)
9. Simple regressions of Y on X_1 and Y on X_2 are both significant explanatory variable for Y. But a multiple regression of Y on X1 and X2 says that neither X_1 nor X_2 is a significant explanatory variable for Y. Clearly, this is a case of multicollinearity (True)
10. Dummy variables are a technique that can be used to incorporate qualititative data into multiple regression (True)
11. When using a dummy variable with values of 0 and 1, it is very important to make sure that the 0's and 1's are used according to standard practice. Reversing the coding will completely destroy the results of the multiple regression (True)
12. Adding additional variables to a multiple regression will always reduce the standard error of estimate (False- not always)
13. Suppose a multiple regression yielded this equation $\hat{Y} = 5.6 + 2.8X_1 - 3.9X_2 + 5.6X_3$. If all X_1, X_2 and X_3 all had values of zero, then $\hat{Y}$ could be expected to have a value of 5.6 (True)
14. The analysis of residuals in a straight line regression model is done to determine the correct value for S_e (False – not the correct value for S_e)
15. If a regression includes all relevant explanatory factors, the residuals should be random (True)
16. A linear relationship between explanatory variables will always produce multicollinearity in the regression model (True)

B. Choose the right answer

17. Suppose a multiple regression yielded this equation $\hat{Y} = 51.21 + 6.88X_1 + 7.06X_2 - 3.71X_3$, the value of b_2 for this equation is
    a. 51.21
    b. 6.88
    c. 7.06
    d. -3.71
    e. Cannot be determined from the information given
    The answer (c)

18. We said that the standard error of estimate has n-k-1 degrees of freedom. What does the k stand for this expression?
    a. Number of elements in the sample
    b. Number of independent vatiables in the multiple regression
    c. Mean of the sample values of the dependent variable
    d. None of these
    The answer (b)

19. Suppose that you have run a multiple regression and have found that the value of b_1 is 1.66. Historical data, however, indicate that the value of β_1 should be 1.34. You wish to test, at 0.05 significant level, the null hypothesis that β_1 is still 1.34. Assuming that you have access to any table you may need, what other information is required for you to perform your test ?
    a. Degrees of freedom
    b. S_{b1}
    c. S_e
    d. (a) and (b) but not (c)
    e. (a) and (c) but not (b)
    The answer (d)

20. Suppose that a toy manufactures wishes to determine whether his red toys sell better than his blue toys. He gathered data regarding sales levels, color, price and average age levels for which the toys are intended. He entered this into a computer run. The resulting multiple regression equation was $\hat{Y} = 70.663 + 713X_1 - 59.6X_2 + 66.4X_3$, where $\hat{Y}$ refers to sale level in units, X1 refers to color (0=blue, 1 = red), Xe refers to retail price (in dollars), and X3 refers to average age level (in years). Which of the following is true if factors of price and level are held constant?
    a. Red toys should sell 713 more units than blue toys
    b. Red toys shouls sell 713 fewer units than blue toys
    c. Children will always choose a blue toy over a red one
    d. (b) and (c) but not (a)
    The answer (b)

21. In the equation $Y = b_0 + b_1X_1 + b_2X_2$, Y is independent of X_1 if,
    (a) $b_2 = 0$
    (b) $b_2 = -1$
    (c) $b_1 = 1$
    (d) None of these
    The answer (d)

22. Because $r^2 = 1 - \frac{\sum(Y-\hat{Y})^2}{\sum(Y-\bar{Y})^2}$, r^2 is equivalent to,
    (a) 1- SSR/SST
    (b) 1- SSE/SST
    (c) 1- SSE/SSR
    (d) 1- SST/SSR
    (e) 1- SST/SSE
    The answer (b)

23. For the multiple regression $Y = b_0 + b_1X_1 + b_2X_2$used to estimate $Y = \beta_0 + \beta_1X_1 + \beta_2X_2$, the form of a plausible confidence interval for $\beta_1$is,
    (a) $(\beta_1 - ts_{b_1}, \beta_1 + ts_{b_1})$
    (b) $(\beta_1 - ts_e, \beta_1 + ts_{e)}$
    (c) $(b_1 - ts_{b_1}, b_1 + ts_{b_1)}$
    (d) $(b_1 - ts_e, b_1 + ts_{e)}$
    The answer (c)

24. Sign of the possible presence of multicollinearity in a multiple regression are,
    (a) Significant t value for the coefficients
    (b) Low standard errors for the coefficients
    (c) A sharp increase in a t value for the coefficient of an explanatory variable when another variable is removed from the model
    (d) All of the above
    The answer (c)