Continuing Education Course #391
Regression Analyses

1. In the Least Squares method of Regression Analysis, the X^TX matrix resulting from the normal equations is a (an)
- a. Square matrix
- b. Orthogonal matrix
- c. Diagonal matrix
- d. All of the above

2. The X^TX matrix resulting from Least Squares method of Regression Analysis is always symmetric
- a. True
- b. False

3. In the Least Squares method of Regression Analysis, the number of normal equations is:
- a. More than the number of parameters in the model
- b. Less than the number of parameters in the model
- c. Equal to the number of parameters in the model
- d. Has nothing to do with number of parameters in the model

4. R^2 can literarily (or forcibly) become 100% (or 1) by the following means
- a. Overfitting
- b. Replication
- c. Orthogonalization
- d. Diagonalization

5. What experimental design approach will most likely prevent R^2 from becoming 100% (or 1)
- a. Overfitting
- b. Replication
- c. Orthogonalization
- d. Diagonalization

6. Linear models are linear in the parameters, but they may or may not be linear in the variables
- a. True
- b. False

7. A model that is linear in both the parameters as well as the variables is a linear regression model
- a. True
- b. False

8. A model that is nonlinear in the parameters but which by suitable transformation can be made linear-in-the-parameter is
- a. Intrinsically nonlinear
- b. Intrinsically linear

○ c. Indeterminate
○ d. None of the above

9. If a model is nonlinear in the parameters and such a model cannot be linearized in the parameters, it is called intrinsically nonlinear regression model
○ a. Only if the variables of the models are linear
○ b. Only if the variables are nonlinear
○ c. Whether or not the variables are linear or nonlinear
○ d. None of the above

10. Because of its complex nature, a method that is used to solve nonlinear regression model is:
○ a. Least Squares method
○ b. Numerical simulation
○ c. Successive approximation
○ d. b and c
○ e. None of the above

11. The correction CF is the of Sum of Squares for β_0 and is used to adjust for the Sum of squares for the various component of the ANOVA table. Which of the following is equivalent to CF or the SS(β_0)?
○ a. $\frac{\sum y^2}{n}$
○ b. $\frac{(\sum y)^2}{n}$
○ c. $Y^T 1 1^T Y$
○ d. $\frac{Y^T 1 1^T Y}{n}$
○ e. b and d

Questions (12-14)
A model produced the following preliminary calculations for an ANOVA table after adjusting for the Correction Factor CF
Sum of Squares Regression (SSR)=170
Total Sum of Squares (SST)= 185
Total number of data point n= 25
Total number of predictors (including the constant term) k (=p+1) 4

12. What is your estimate of the Error or Residual Sum of Squares (SSE)
○ a. It is unknown
○ b. 15
○ c. 18
○ d. 25

13. What is the estimate of R-Square (R^2)
○ a. 8%
○ b. 28%
○ c. 92%
○ d. 82%

14. What is your estimate of Adjusted R-Square (Adjusted R^2)?
○ a. 89%
○ b. 92%

○ c. 91%
○ d. 11%

Questions (15-18)
A model produced the following preliminary calculations for an ANOVA table after adjusting for the Correction Factor CF gave:

Sum of Squares Error (SSE)=30, Total Sum of Squared (SST)= 195
Total number of data point n= 20
Total number of predictors (including the constant term β_0) p= 5

15. What is the estimate of error variance (Se^2)?
○ a. 3
○ b. 2
○ c. 5
○ d. 4

16. What is the estimate of error std. deviation (S_e)?
○ a. 2
○ b. $\sqrt{2}$
○ c. 3
○ d. $\sqrt{5}$

17. What is your estimate of the coefficient of Variation (CV) where CV is the residual estimate of the error standard deviation, measured as a percent of the average response Y?
○ a. 71%
○ b. 39%
○ c. 65%
○ d. All of the above

18. What is the estimate of R-Square (R^2)
○ a. 50%
○ b. 15%
○ c. 85%
○ d. None of the above

Questions (19-23)
The following data was obtained for a dependent variable Y and an independent variable X, with their relationship defined as: $Y = \beta_0 + \beta_1 X + \varepsilon_{ij}$. Preliminary calculation for an ANOVA table gave:

X	**10**	**15**	**20**	**25**	**30**	**35**	**40**
Y	**27**	**24**	**17**	**15**	**11**	**10**	**8**

$$\sum X = 175, \sum X^2 = 5075, \dot{X} = 25, \sum Y = 112, \sum Y^2 = 2104, \sum XY = 2345, \dot{Y} = 16.00$$

19. Compute S_{XX}
○ a. S_{XX} = 750
○ b. S_{XX} = 500
○ c. S_{XX} = 700
○ d. S_{XX} = 450

20. Compute S_{XY}

- a. S_{XY} = -455
- b. S_{XY} = -550
- c. S_{XY} = 455
- d. S_{XY} = 225

21. Compute S_{YY}

- a. S_{YY} = 322
- b. S_{YY} = 320
- c. S_{YY} = -344
- d. S_{YY} = 312

22. Compute β_1

- a. β_1 = -0.35
- b. β_1 = -1.35
- c. β_1 = -3.35
- d. β_1 = -0.65

23. Compute the estimate of β_0

- a. β_0 = 35
- b. β_0 = 32
- c. β_0 = 22
- d. β_0 = 28

Questions (24-25)

An $\mathbf{X^TX}$ matrix is given for a model defined as $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon_{ij}$

$(X^TX) = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$**, the $\mathbf{X^TY}$ matrix is given by** $X^TY = \begin{bmatrix} 20 \\ 5 \\ 4 \end{bmatrix}$

24. What is the inverse matrix $(X^TX)^{-1}$

- a. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
- b. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}$
- c. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 10 \end{bmatrix}$
- d. $\begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

25. What are the estimates of the parameters, $\beta_0, \beta_1, \beta_2$? $\rightarrow \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} ? \\ ? \\ ? \end{bmatrix}$

- a. β_0=0.2, β_1 =5, β_2 =0.4
- b. β_0=20, β_1 =5, β_2 =2

○ c. β_0=2, β_1 =10, β_2 = 4
○ d. β_0 =20, β_1= 25 β_2 =4

26. Which of the following about multicollinearity is true?
○ a. Multicollinearity occurs when there are high correlations between two or more predictor variables in a regression model
○ b. Multicollinearity creates redundant information because the variables that exhibit multicollinearity work together moving the response jointly in one direction or in the opposite direction, thus skewing the results in the model.
○ c. Because of multicollinearity of the independent variables, the independent are truly not independent since they are dependent on each other.
○ d. None of the above is true
○ e. All of the above is true

Question (27-30)
The following data was obtained from a system with 3 regressor variables (X1, X2, X3). It is desired to determine if there is multicollinearity among the regressor and if so, the severity of the effect of multicollinearity using the Variance Inflation Factor (VIF) as a metric.

Variables			
	X1 vs X2X3	**X2 vs X1X3**	**X3 vs X1X2**
R2=	**0.433099**	**0.4895625**	**0.1489575**
VIF=			

27. What is the VIF for X1 vs (X2, X3)
○ a. 1.95
○ b. 10.5
○ c. 1.76
○ d. 1.17

28. What is the VIF for X2 vs (X1, X3)
○ a. 1.95
○ b. 15.68
○ c. 1.76
○ d. 1.17

29. What is the VIF for X3 vs (X1, X2)
○ a. 1.95
○ b. 12.3
○ c. 1.76
○ d. 1.17

30. Based on the values of VIF for these three regressor variables, how would you classify their multicollinearity severity
○ a. $VIF = 1 \Rightarrow$ no correlation
○ b. $1 \leq VIF \leq 5 \Rightarrow$ moderate correlation
○ c. $VIF > 5 \Rightarrow$ Critical level of multicollinearity
○ d. None of the above