The following information applies to questions. The data set is the result of an experiment to test the strength of cement concrete mix. The experiment consisted of applying a compressive force on the top of different sample specimens. Two responses were collected: the stress and stain at which a sample specimen failed. The factors relate to mixture proportions, rates of speed at which the force was applied, and ambient temperature. Higher values of response variables indicate stronger materials.

The variables are:

X1: percent binder (the amount of cement in the mixture);

X2: loading rate (the speed at which the force is applied);

X3: the ambient temperature;

Y1: the stress at which the sample specimen failed;

Y2: the strain at which the specimen failed.

Perform a multiple regression analysis using Y2 as dependent variable and X1, X2 and X3 as independent variables and answer the following questions.

Obs x1 x2 x3 y1 y2

1 5.3 0.02 77 42 3.2

2 5.3 0.02 32 481 0.73

3 5.3 0.02 0 543 0.16

4 6 2 77 609 1.44

5 7.8 0.2 77 444 3.68

6 8 2 104 194 3.11

7 8 2 32 977 0.19

8 8 2 0 872 0

9 8 0.02 104 35 5.86

10 7 0.02 77 96 5.97

11 7 0.02 32 663 0.29

12 7 0.02 0 702 0.04

13 10 2 77 518 2.72

14 12 0.01 77 40 7.35

15 12 0.02 0 683 0.14

16 12 0.02 104 22 15

17 14 0.01 77 35 11.8

For questions 1 to 3, please select all that applies.

1. Based on the DFFITS, which of the observations below will be considered an outlier?
2. The 5th observation is an outlier
3. The 6th observation is an outlier
4. The 10th observation is an outlier
5. The 11th observation is an outlier
6. The 12th observation is not an outlier
7. Based on the Studentized Residuals, which of the following is true?
8. Observation 15 is an outlier
9. Observation 16 will not be considered as an outlier
10. Observation 3 is an outlier
11. Observation 16 should be deleted to allow for better inferences
12. Based on the DFBETAS, select all that are true.
13. It is possible there was an error in recording the values during the experiment
14. The experiment must be conducted again to resolve the outlier problems
15. We will need to rerun the analysis because of the outliers
16. Any observation does not appear to have significant influence on any estimated regression coefficient or intercept (b0, b1, b2, b3).
17. Which of the following statements is/are true?
18. The homogeneity of variance assumption is not satisfied based the SAS output.
19. The percent binder and the ambient temperature are significant, however, the loading rate is not significant at alpha=0.05 level.
20. The Correlation matrix, Variance Inflation Factor and Conditional Index do not provide evidence of collinearity between the predictor variables.
21. a and c only
22. Based on the cook’s distance, the 5th and 6th observations are outliers
23. True
24. False
25. Which of following is true based on normal Q-Q Plot and normality tests.
26. Variable x3 (ambient temperature) follows a normal distribution
27. Variable x2 (loading rate )follows a normal distribution
28. Variable x1 does not follow a normal distribution
29. a, b and c
30. Using backward stepwise selection method, find the best subset of predictor variables to predict Y1: the stress at which the sample specimen failed. Set slstay at 0.5. What is the final model?
31. Y= 670.28+ 176.33X1 +6.62X2
32. Y= 767.58-6.43X3
33. Y= 670.28+ 176.33X2 -6.62X3
34. Y= 670.28+ 176.33X1 -6.62X2
35. Based on the Adjusted R-Square values, select the best set for the regression model.
36. X1
37. X1 X3
38. X1 X2 X3
39. X2 X3
40. X1 X2
41. Based on the Malow's Cp, which one is the best set?
42. X1
43. X1 X3
44. X2 X3
45. X1 X2 X3
46. X1 X2
47. Based on questions 7, 8 and 9, the best set selected by each method above is the same.
48. True
49. False