

Homework 9

Here are the problems for the ninth homework assignment.

Multiple choice questions

1. Which one of the following is correct?
 - A** the residuals form a basis for assessing the standard deviation of the error term in a well fitted multiple regression;
 - B** the residuals are defined as the fitted values less the mean observed value;
 - C** the residuals are only important for diagnostic purposes and are not important for prediction purposes;
 - D** the residuals are always within 2 standard deviations of the predicted values;
 - E** none of statements A, B, C, D above is correct.
2. t -ratios in multiple regression
 - A** are expected to vary between +2 and -2;
 - B** test the significance of differences between regression coefficients;
 - C** test the statistical significance of individual regression coefficients;
 - D** exceed 2 only if the corresponding X variables are not contributing to variability of the response variable;
 - E** may be plotted against the fitted values to assess the goodness of fit.
3. Which one of the following is correct?
 - A** the predicted values are the values of the X variables multiplied by the corresponding coefficients;
 - B** the predicted values can be defined only for the X -values that have been used to fit the regression model;
 - C** in order to evaluate residuals the predicted values have to be evaluated at the X -values used to fit the regression model;
 - D** the average of predicted values is always zero;
 - E** the predicted values are not affected by deleting outliers from the data.
4. Which of the following is *not* a standard graphical tool in the diagnostic of multi-variate regression models?
 - A** a plot of deleted residuals against fitted values;
 - B** a plot of the fitted values against the observed values;
 - C** a Normal plot of the deleted residuals;
 - D** dotplots of the response and explanatory variables;
 - E** a time series plot of the deleted residuals.

5. Given a set of explanatory variables, the parameters of a multiple regression model
 - A emerge from the initial diagnostic stage of the model fitting;
 - B include the regression coefficients which determine the prediction formula and the error standard deviation which is the key parameter in determining prediction error;
 - C determine the observed values of the response variable;
 - D determine the exceptional values of the response variable.
 - E should be first tested for significance before any diagnostic of the model is performed.

Problems

1. In the textbook and lecture, the dependence of 'Jobtime' variable on the explanatory variables: 'Units', 'Operations per Unit', 'Total Operations' and 'Rushed' has been studied through the multivariate regression analysis. After the diagnostic process, three values have been excluded corresponding to the order numbers: 9, 11, and 16. Here is the computer printout representing the multivariate regression fit to all the data

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	77.24147	44.75814	1.726	0.105
Units	-0.15074	0.11215	-1.344	0.199
OperU	7.15154	4.30466	1.661	0.117
TotOper	0.11460	0.01322	8.668	<0.00001 ***
Rush	-24.94413	19.11023	-1.305	0.211

Residual standard error: 37.46 on 15 degrees of freedom
Multiple R-Squared: 0.9729

and the one for the data from which 9, 11, and 16 orders have been excluded

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	44.216498	9.079992	4.870	0.000385 ***
Units	-0.069307	0.028526	-2.430	0.031758 *
OperU	9.828585	0.887341	11.076	<0.00001 ***
TotOper	0.107795	0.004114	26.202	<0.00001 ***
Rush	-37.960103	3.857205	-9.841	<0.00001 ***

Residual standard error: 7.413 on 12 degrees of freedom
Multiple R-Squared: 0.9978

- Write down the models corresponding to these two fits involving the two sigma prediction band.

- Give the arguments based on the computer printouts for deletion of 9, 11, and 16 orders when fitting the model.
- Evaluate the residuals to the 9th, 11th, and 16th orders for both the fits. Compare them with two sigma prediction band and comment.
- According to the second fit how much time is saved when the requested job is rushed?
- Suppose that the company has received an order for a job that has 85 units and 14 operations per unit. The client would like to know the rushed and non-rushed jobtimes that will be required for completion of the order. Provide with the appropriate information based on the fitted regression model.
- Suppose that the client opted for the rushed order and it took 272 hours to complete the job. Is it in line with the prediction given to the client prior to taking the order?