STA 101 Spring 2017 Discussion 03

Full Detail

1. 113 hospitals were measured, and the goal was to build a model to predict the average length of stay for patients. Let Y = the average length of stay of a patient, X_1 = the age of the patient, X_2 = the average number of nurses at the hospital, and X_3 = `Y` if the hospital is associated with a medical school, `N` otherwise.
Two models were fit, with the following values of SSE:

Interaction Model:	$\hat{y} = 3.861 + 0.093X_1 + 0.004X_2 - 6.518X_3 + 0.171X_1X_3 - 0.004X_2X_3$	SSE = 327.73
Simple Model:	$\hat{y} = 3.511 + 0.101X_1 + 0.003X_2 + 0.973X_3$	SSE = 333.566

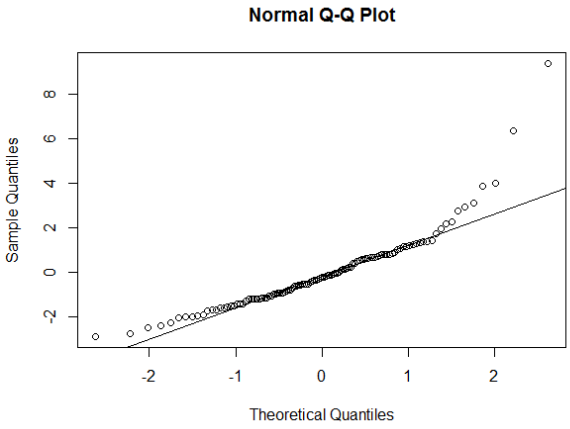
(a) State the null and alternative hypothesis for dropping both interaction terms from the model.

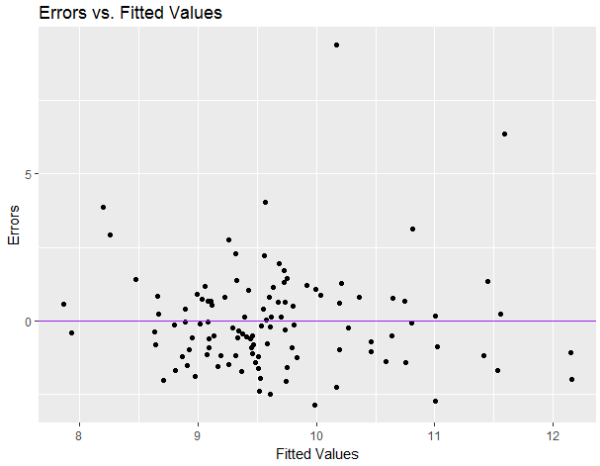
(b) Find the test-statistic associated with the hypothesis in (a).

(c) If the p-value for the test-statistic in (b) is: 0.3886816, state the conclusion of your test at $\alpha = 0.05$.

(d) Does your final model suggest that there are different affects of the age of a patient on the length of the stay, depending on if th hospital is ssociated with a medical school or not? Explain your answer.

2. The final model was selected by a statistician to be the simplier model. Some diagnostic plots follow:





(a) Based on one of the above plots, does your errors appear to be normally distributed? Explain.

(b) There are potential outliers in the plots above. Should we remove them? Explain why or why not.

(c) The final model chosen had the following 90% confidence intervals:

	Parameter	5 %	95 %
(Intercept)	5.551	2.884	8.218
β_1	0.061	0.012	0.110
β_2	0.004	0.001	0.006
β_3	0.734	-0.023	1.490

Interpret the confidence interval for β_2 in terms of the problem.

(d) If the AIC for the model that includes X_1, X_2, and X_3 has an AIC of: 382.638, and the model that includes only X_1 and X_2 have an AIC of : 383.295, what would you choose for the best model and why?