King Fahd University of Petroleum & Minerals
Department of Mathematics & Statistics
STAT-212-Term122
Quiz #6

Name: ID: Serial:

A ten year study conducted by the American heart association provided on how age (X1), blood pressure (X2), and smoking (X3={1 if smoker, o if nonsmoker}) relate to the risk of strokes (Y). Risk interpreted as the probability (times 100) that a person will have a stroke over the next ten year period.

Using the **MINITAB OUTPUT** answer the following questions:

Best Subsets Regression: y versus x1, x2, x3, x1*X2, x2*x3, X1*x3

Response is y

					x	x	x	X	x	X
					1	2	3	2	3	1
					*	*	*			*
Vars	R-Sq	R-Sq(adj)	C-p	S	1	2	3	2	3	3
1	63.3	61.3	27.6	9.2430				X		
1	54.8	52.3	37.7	10.262					X	
1	54.7	52.2	37.8	10.265						X
1	46.3	43.3	47.8	11.182			X			
2	80.6	78.4	9.0	6.9083	X	X				
2	79.8	77.4	10.0	7.0538				X		X
2	79.5	77.1	10.3	7.1058	X			X		
2	77.3	74.6	13.0	7.4832		X		X		
3	87.9	85.6	2.4	5.6313	X	X			X	
3	87.3	85.0	3.0	5.7566	X	X	X			
3	87.1	84.7	3.3	5.8119	X	X				X
3	86.8	84.3	3.7	5.8905	X			X	X	
4	88.4	85.3	3.8	5.6908	X	X		X	X	
4	87.9	84.7	4.4	5.8111	X	X	X		X	
4	87.9	84.7	4.4	5.8156	X	X			X	X
4	87.6	84.3	4.7	5.8813	X	X	X	X		
5	88.5	84.4	5.6	5.8571	X	X	X	X	X	
5	88.4	84.3	5.8	5.8876	X	X		X	X	X
5	88.0	83.7	6.3	5.9994	X	X	X		X	X
5	87.6	83.2	6.7	6.0871	X	X	X	X		X
6	89.1	84.0	7.0	5.9391	X	X	X	X	X	X

Regression Analysis: y versus x1, x2, x2*x3

The regression equation is
y = - 85.5 + 1.05 x1 + 0.220 x2 + 0.0574 x2*x3

Predictor	Coef	SE Coef	T	P	VIF
Constant	-85.52	15.74	-5.43	0.000	
x1	1.0550	0.1643	6.42	0.000	1.5
x2	0.21982	0.04843	4.54	0.000	1.5
x2*x3	0.05743	0.01855	3.10	0.007	1.5

S = 5.631 R-Sq = 87.9% R-Sq(adj) = 85.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	3683.6	1227.9	38.72	0.000
Residual Error	16	507.4	31.7		
Total	19	4190.9			

a. Is the model **y = - 85.5 + 1.05 x1 + 0.220 x2 + 0.0574 x2*x3** the best model. Why?

b. At 5% level of significance test to see whether the addition of the **interaction term** between blood pressure and the smoking contribute significantly to the estimated regression equation developed in part (a)?

c. How much of the variation in the risk of strokes can be explained by the model in part (a), taking in account the number of predictors and the sample size.

d. What can you say about the multicollinearity between the independent variables? Explain.

e. A smoker has age 70 with blood pressure 173, what is the probability that the person will have a stroke over the next ten year period?

f. Would you conclude that the model is significant at 5% level of significance?