KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICS AND STATISTICS
Term 121

STAT 310: Regression Analysis

Final Exam Tuesday January 1, 2013

Name: _____ ID #: _____

☺ Instructions:

1. Read the questions carefully and answer accordingly.
2. Interpretations, discussions and conclusions should be clear and thorough.
3. In any testing problem, write down the hypotheses, the test statistic, the decision and the conclusion. Use a significance level of 5%.
4. In modeling problems, write the model explicitly.
5. Label your graphs, and make your references to the output clear.
6. Save your Minitab work in a word file "yourname.docx" and email it to me.

.

Question No	Full Marks	Marks Obtained
1	28	
2	20	
3	22	
4	30	
Total	100	

1) A researcher is conducting a study about tuition at American colleges and universities. So far, he has collected data from 20 schools about their tuition costs, Y (in thousands of dollars), their score on an independent rating scale, X1 (in points out of 100), their size, X2 (in thousands of undergraduates), and whether they are a public or privateschool. A printout for the multiple regression of Y on the three X variables is shown

The regression equation is

Tuition = - 2.41 + 0.0967 Rating - 0.0192 Size + 16.9 Type

Predictor	Coef	StDev	T	p-value
Constant	-2.4053	0.9257	-2.60	0.019
Rating	0.09671	0.01172	8.25	0.000
Size	-0.01923	0.01606	-1.20	0.249
Type	16.8581	0.3357	50.21	0.00

S = 0.5869 R-Sq = 99.7% R-Sq(adj) = 99.6%

Analysis of Variance

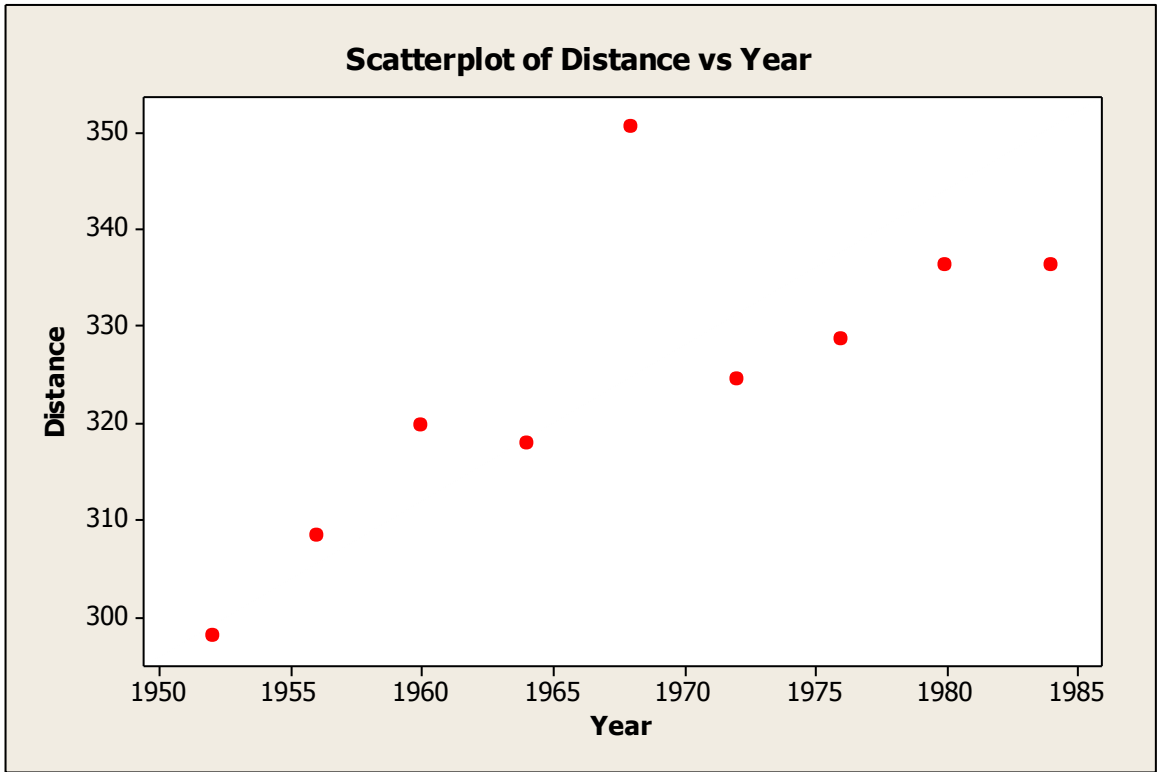
Source	DF	SS	MS	F	p-value
Regression	3	1742.29	580.76	1686.01	0.000
Residual	16	5.51	0.34		
Total	19	1747.8		1747.80	

Fit	StDev Fit	95.0% CI	95.0% PI
20.236	0.402	(19.384, 21.088)	(18.728, 21.744)

a) Interpret the intercept and the variable coefficients and comment on them.
b) Test whether the combination of rating, size and type of school is useful in modeling tuition.
c) Should you use R-sq or R-sq(adjusted)? And Why? What does R-sq(adjusted) guard against?
d) Find a 95% confidence interval for the coefficient of size. Interpret that interval, and what it means with respect to the importance of size.
e) Test whether a school is public or private has a significant impact on tuition? What conclusion do you draw?
f) Your friend is interested in attending ABC private university, which has 1000 students and a rating of 60. Help him find a range of values in which he can be 95% certain that the true tuition of ABC University lies. Explain why you chose the interval that you did.
g) Another friend suggested adding the number of faculty members in the university as a variable to improve the model. What do you think and why?

2) The following table gives the winning distance (in inches) for the Olympic long jump from 1952 to 1984.

Year	Distance
1952	298.00
1956	308.25
1960	319.75
1964	317.75
1968	350.50
1972	324.50
1976	328.50
1980	336.25
1984	336.25



a) Comment on the scatter plot.
b) Fit a simple linear regression, perform the usual diagnostics, and discuss.
c) In 1968 the Olympics were held in Mexico city and many records were broken, probably due to the high altitude. What kind of observation is that? Would you remove it? What will happen if you do? Answer those questions ***without*** the help of any refitting?
d) Now refit the model after removing the 1968 observation, and perform the appropriate analysis.
e) Because of the slight curvilinear nature of the data, a colleague suggested adding a second order term. Critique this idea.
f) Now test whether a second order term contributes significantly to the model.

3) A pharmaceutical firm wanted to obtain information on the relationship between the dose level of a drug and its potency. The designed experiment yielded the following data:

Dose Level	Potency (Response)
2	5, 7, 3
4	10, 12, 14
8	15, 17, 18
16	20, 21, 19
32	23, 24, 29

a) Plot the data.
b) Fit a linear and a quadratic model.
c) Examine the residual plots vs the independent variable and the predicted response in each model.
d) Based on the above, which model is more appropriate.
e) Can you think of a better model? If yes, fit it and compare it to the linear and the quadratic models by looking at the same criteria as above.

4) Consider the data problem 4 in the file final_exam_data.xls
a) Find the best model using stepwise regression, doing the step by step analysis ***NOT*** using the stepwise regression procedure in Minitab. At each step show the appropriate sums of squares and update the model using α to enter = α to remove = 0.05.
b) Find the best subset regression using Minitab, and judge the best model by the most possible criteria.
c) Is this the same model you found above? Explain.