

Chapter 14 & 15 Homework Due Wednesday, May 9, 2012

Remark: No late submission will be accepted.

Q1:

From your textbook do the following problems:

14.43, 15.5, 15.9.

Note: All computer outputs must be included, but you need to clearly indicate the answer for each part, i.e. having the computer output alone will not be enough.

Q2:

A publishing company is attempting to develop a model that it can use to help predict textbook sales for books it is considering for future publication. The marketing department has collected data on six variables from a random sample of 15 books. These variables are as follows:

Y: Number of Volumes sold (1000's).

X1: Number of Pages.

X2: Number of competing books.

X3: Advertising Budget (\$1000's).

X4: Age of the Author.

X5: Production Expenditures (\$1000's).

X6: Number of Reviewers.

In addition to these variables, the type of book sold is incorporated in the model. The company produces only three types of books which are: Chemistry, Statistics, or Physics, as follows:

X7: 1=If a Biology book, 0 otherwise.

X8: 1=If a Statistics book, 0 otherwise.

Regression Analysis: Y versus X1, X2, X3, X4, X5, X6, X7, X8

The regression equation is

$$Y = -104 + 0.123 X1 - 0.55 X2 + 1.16 X3 + 1.34 X4 + 0.58 X5 + 1.61 X6 - 20.9 X7 - 29.6 X8$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	-103.69	37.57	-2.76	0.033	
X1	0.12302	0.09192	1.34	0.229	8.4
X2	-0.553	2.864	-0.19	0.853	4.0
X3	1.1649	0.7447	1.56	0.169	4.4
X4	1.3393	0.7707	1.74	0.133	2.0
X5	0.580	1.011	0.57	0.587	15.5
X6	1.613	6.837	0.24	0.821	5.8
X7	-20.95	19.09	-1.10	0.315	2.7
X8	-29.56	16.21	-1.82	0.118	1.6

S = 22.27

R-Sq = 91.9%

R-Sq(adj) = 81.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	33671.6	4209.0	8.48	0.009
Residual Error	6	2976.8	496.1		
Total	14	36648.4			

Durbin-Watson statistic = 2.14

Predicted Values for New Observations

New Obs	Fit	SE Fit	99.0% CI	99.0% PI
1	51.05	25.40	(-43.10, 145.21)	(-74.19, 176.29)X

X denotes a row with X values away from the center

Values of Predictors for New Observations

New Obs	X1	X2	X3	X4	X5	X6	X7	X8
1	204	3.00	30.0	50.0	34.0	6.00	0.0	0.0

From the Minitab outputs above, answer the following questions:

- How much of the total variation in the number of volumes sold is explained by the variation in these six explanatory variables?
- Would you conclude that the model is significant at 2.5% level of significance? Explain in detail.
- Which of the predictors can be concluded to be significant in explaining the variation in the number of the volumes sold? Explain. Using $\alpha = 0.05$.
- Find 99% C.I. for the regression coefficient of Advertising Budget, and interpret this C.I.
- Based on your solution in part (e), can you conclude that the Advertising Budget is significant? Why?
- Estimate the number of volumes sold for a book with each of the following given information and indicate which book the predicted value is for

1) $(X_1, X_2, \dots, X_8) = (204, 3, 30000, 50, 34000, 6, 1, 1)$	
Which book (Biology, Statistics, Physics, or not possible)?	
2) $(X_1, X_2, \dots, X_8) = (204, 3, 30000, 50, 34000, 6, 0.5, 1)$	
Which book (Biology, Statistics, Physics, or not possible)?	
3) $(X_1, X_2, \dots, X_8) = (204, 3, 30000, 50, 34000, 6, 0, 0)$	
Which book (Biology, Statistics, Physics, or not possible)?	

- Find a 99% C.I. for the number of volumes sold for a book with 204 pages, 3 competing books, \$30,000 advertising budget, 50 years old of the author, \$34,000 production expenses, 6 reviewers, and it is a Physics book.

Q2:

The following data ANOVA summary table is for a multiple regression model with two independent variables

Source	DF	SS	MS	F
Regression	2	60		
Residual Error	<u>18</u>			
Total	20	180		

- Determine the Sum of Square Error (SSE), Mean Square Regression (MSR) and the Mean Square Error (MSE).
- Compute the overall F_{stat} test statistic
- Determine whether there is a significant relationship between Y and the two independent variables at the 0.05 significance level.
- Compute the coefficient of multiple determination, R^2 , and interpret its meaning.
- Compute the adjusted R^2 .
- Compute the standard error of estimate S_e .

Q3:

The following regression equation is obtained for a sample of $n=25$:

$$\hat{y}_i = 5 + 3X_{1i} + 1.5X_{2i}^2$$

- Predict Y for $X_1 = 2$
 - Suppose that the computed t_{stat} test statistic for the quadratic regression coefficient is 2.35. At the 5 % level of significance, is there evidence that the quadratic model is better than the linear model?
 - Suppose that the computed t_{stat} test statistic for the quadratic regression coefficient is 1.17. At the 5 % level of significance, is there evidence that the quadratic model is better than the linear model?
-

Q4:

A publishing company is attempting to develop a model that it can use to help predict textbook sales for books it is considering for future publication. The marketing department has collected data on six variables from a random sample of 15 books. These variables are as follows:

- Y: Number of Volumes sold (1000's).
- X1: Number of Pages.
- X2: Number of competing books.
- X3: Advertising Budget (\$1000's).
- X4: Age of the Author.
- X5: Production Expenditures (\$1000's).
- X6: Number of Reviewers.

Correlations: Y, X1, X2, X3, X4, X5, X6

	Y	X1	X2	X3	X4	X5
X1	0.622					
	0.013					

X2	0.355	0.501					
	0.194	0.057					
X3	0.62	0.091	0.384				
	0.014	0.746	0.158				
X4	0.485	-0.019	-0.113	0.265			
	0.067	0.947	0.687	0.34			
X5	0.896	0.67	0.27	0.539	0.438		
	0	0.006	0.331	0.038	0.103		
X6	0.66	0.377	0.291	0.355	0.528	0.737	
	0.007	0.166	0.292	0.194	0.043	0.002	

Best Subsets Regression: Y versus X1, X2, X3, X4, X5, X6

Response is Y

Vars	R-Sq	R-Sq(adj)	C-p	S	X	X	X	X	X	X
					1	2	3	4	5	6
1	80.2	78.7	1.4	23.600						X
1	43.6	39.3	24.4	39.868						X
2	82.9	80.0	1.7	22.854			X			X
2	81.6	78.6	2.5	23.676		X				X
3	84.1	79.8	3.0	23.003	X		X			X
3	83.8	79.4	3.1	23.199			X	X		X
4	87.2	82.1	3.0	21.640	X		X	X		X
4	85.0	79.0	4.4	23.456		X	X	X		X
5	87.2	80.1	5.0	22.799	X	X	X	X		X
5	87.2	80.1	5.0	22.808	X		X	X	X	X
6	87.3	77.7	7.0	24.166	X	X	X	X	X	X

From the Minitab output above, answer the following:

- Do you think that Number of pages and Production expenditures are directly (positively) related? Test using 2.5 % level of significance.
- If you are going to fit a regression model using the forward selection method, what is the first predictor to be used? Why?
- What is the best model to be selected? Justify your selection.