

**KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**  
**DEPARTMENT OF MATHEMATICAL SCIENCES**  
**DHAHRAN, SAUDI ARABIA**

**STAT 212: BUSINESS STATISTICS II**

Semester 042

Final Exam

Monday 6, 2005

12:30pm – 2:30pm

Please **circle** your instructor's name:

Prof. Hassen A .Muttlak

Raid Anabosi

Name:

ID#:

Section:

Serial:

Question No	Full Marks	Marks Obtained
1	20	
2	10	
3	10	
4	10	
5	6	
6	4	
7	8	
8	12	
<b>Total</b>	80	

**Note:** For each question: Clearly state your hypotheses, assumptions and your conclusions. Use 5% level of significance unless specified other wise in the problem.

1. A publishing company is attempting to develop a multiple regression model that can be used to help predict text book sales for books it is considering for future publication. The marketing department has collected data on several variables from a random sample of 15 books, these variables are:  $Y$ =Volumes Sold,  $X_1$ =Pages,  $X_2$  = Competing Books,  $X_3$ =Advertising Budget,  $X_4$ =Age of Author.

**Use the MINITAB output given below to answer the following questions:**

- a. How much of the total variation in the book sales can be explained by these four independent variables? Would you conclude that the model is significant at the 5% level? Explain.
- b. Develop a 95% confidence interval for the regression coefficient of the variable Age of Author and interpret this confidence interval. Based on your finding can you conclude that the Age of Author is playing a significant role? Explain.
- c. Which of the independent variables can be concluded to be significant in explaining the variation in the book sales? Use  $\alpha = 0.05$ .
- d. Predict the book sales if,  $X_1=400$ ,  $X_2=10$ ,  $X_3=10000$  and  $X_4=35$ .

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

f. Select the best model using the MINITAB output. Clearly justify your selection.

**The regression equation is**

$$Y = -125308 + 176 X1 - 1574 X2 + 1.59 X3 + 1614 X4$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	-125308	31082	-4.03	0.002	
X1	175.90	39.77	4.42	0.001	1.4
X2	-1574	1996	-0.79	0.449	1.7
X3	1.5917	0.4445	3.58	0.005	1.4
X4	1613.7	625.0	2.58	0.027	1.2

S = 23850      R-Sq = 84.5%      R-Sq(adj) = 78.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	30960327043	7740081761	13.61	0.000
Residual Error	10	5688072957	568807296		
Total	14	36648400000			

**Best Subsets Regression: Y versus X1; X2; X3; X4**

Response is Y

Vars	R-Sq	R-Sq(adj)	C-p	S	X 1	X 2	X 3	X 4
1	38.6	33.9	28.5	41593	X			
1	38.5	33.8	28.6	41644		X		
2	70.7	65.8	9.9	29930	X	X		
2	63.3	57.2	14.6	33463	X		X	
3	83.5	79.0	3.6	23436	X	X	X	
3	74.1	67.1	9.7	29357	X	X	X	
4	84.5	78.3	5.0	23850	X	X	X	X

2. A study was recently performed in which it attempted to develop a regression model to explain variation in the mileage ratings of new cars. At one stage of the analysis, the estimated regression took the following form:

$$\hat{y} = 34.2 - .003X_1 + 4.56X_2$$

where  $X_1$  = Vehicle weight, and  $X_2$  = 0 if standard transmission and  $X_2$  = 1 if automatic transmission.

- Interpret the regression coefficient for the variables  $X_1$  and  $X_2$ .
- Cadillac's 2000 with automatic transmission weights 4012 pounds. Provide an estimate of the average highway mileage you would expect to obtain from this model.

3. Recently a survey was conducted involving customers of a fitness center. Participants were asked to indicate how often they use the club by checking one of the following categories: 0-1 time per week; 2-3 times per week; 4-5 times per week. The following data show how males and females responded to this question.

	0-1	2-3	4-5
Males	41	61	50
Females	109	89	60

One of the purposes of the survey was to determine whether there is a relationship between the gender of the customer and the number of visits made each week.

- State the appropriate null and alternative hypothesis.
- What test procedure is appropriate to use to conduct this test?
- Conduct the hypothesis test using  $\alpha=0.05$ .

4. The following data set represent the price in dollars and demand in thousand KG for a commodity:

Price	20	60	100	140	180	220	260	300	340	380
Demand	0.18	0.37	0.35	0.78	0.56	0.75	1.18	1.36	1.17	1.65

Use the MINITAB output given below to answer the following questions:

- Report the estimated regression equation, is the model significant? And comment on the value of  $R^2$ .
- Test the hypothesis that the slope of the regression is more than 8.
- Using the output only, what can you say about the assumptions of the model? Clearly comment on the assumptions.

**MINITAB OUTPUT**

The regression equation is  
 Demand = 0.069 + 0.00383 Price

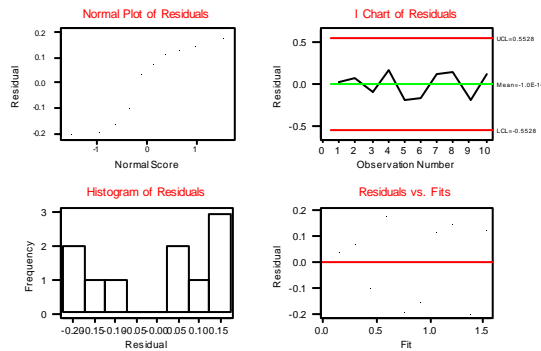
Predictor	Coef	SE Coef	T	P
Constant	0.0692	0.1010	0.69	0.512
Price	0.0038288	0.0004378	8.75	0.000

S = 0.1591      R-Sq = 90.5%      R-Sq(adj) = 89.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1.9351	1.9351	76.49	0.000
Residual Error	8	0.2024	0.0253		
Total	9	2.1374			

Residual Model Diagnostics



5. A company claims that its bulbs are **superior** to those of its main competitor. If a study showed that a sample of 16 of its bulbs had a mean lifetime of 647 hours of continuous use with a standard deviation of 27 hours, while a sample of 16 bulbs made by its main competitor had a mean lifetime of 638 hours of continuous use with standard deviation of 31 hours. Do you think that the company claim is justified? Use 5% level of significance.

6. A company's market share is very sensitive to both its level of advertising and the level of its competitors advertising. A firm known to have a 56% market share wants to test whether this value is still valid in view of recent advertising campaigns of its competitors and its own increased level of advertising. A random sample of 500 consumers reveals 298 use the company's product. Is there evidence to conclude that the company's market share is no longer 56% at the 0.01 level of significance?
7. The following data are average 1992-1997 (based on the first quarter) market shares of Japanese cars in the market. Do trend analysis to predict next year value.
- 27, 25.5, 26.4, 27.3, 28.2, 30.1**



8. A major developer of housing communities in a city kept a record of the relative cost of labor and materials in its market areas for the past three years. These data are as follows:

Years	1998	1999	2000
Average Labor cost	SR 49000	57000	63000
Average Material Cost	SR 95000	104000	110000
% Material Cost	67	68	66

- Determine the simple index for each component in the construction of the house using 1999 as the base year.
- Find the unweighted aggregate index for the two components in the construction of the house using 1999 as the base year.
- Construct a Paasche index number using 1999 as the base year.