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

STAT 212: BUSINESS STATISTICS II

Semester 133
Second Major Exam
Wednesday July 16, 2014
4:00 − 5:30 pm

Name: ID #: Serial #:

Question No	Full Marks	Marks Obtained
1	**11**	
2	**16**	
3	**23**	
Total	**50**	

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

Q1. (11 marks) A manufacturing company is interested in predicting the number of defects that will be produced each hour on the assembly line. The managers believe that there is a relationship between the defect rate and the production rate per hour. The managers believe that they can use production rate to predict the number of defects. The following data were collected for 10 randomly selected hours.

Defects (X)	20	30	10	20	30	25	30	20	10	40
Production rate (Y)	400	450	350	375	400	400	450	300	300	300

Given the following

$\sum x = 235$, $\sum y = 3725$, $\sum x^2 = 6325$, $\sum y^2 = 1418125$ and $\sum xy = 89000$.

a. Find the correlation coefficient between the two variables

b. Is there a significant direct correlation between the two variables? Explain.

H_0:

H_1:

Assumptions:

Test statistic:

Critical value:

Decision:

Conclusion:

Q2. (16 marks) A study was done in which the high daily temperature and the number of traffic accidents within the city were recorded. These sample data are shown as follows:

Temperature (X)	91	56	75	68	50	39	98
# of Accidents (Y)	7	4	9	11	3	5	8

Given the following

$$\sum x = 477,\quad \sum y = 47,\quad \sum x^2 = 35291,\quad \sum y^2 = 365,\quad \sum xy = 3413$$

$$\sum (x-\bar{x})^2 = 2786.86,\quad \sum (y-\bar{y})^2 = 49.4286 \text{ and } \sum (x-\bar{x})(y-\bar{y}) = 210.286$$

a. Find the equation for predicting the number of accidents using the temperature:

b. Construct a 95% confidence interval for the slope of the regression line.

c. Is there a significant linear relationship between the two variables? Explain.

H_0:

H_1:

Assumptions:

Decision Rule:

Decision:

Conclusion:

Q4. (23 marks) The following Minitab output is the result of a multiple regression analysis in which we are interested in explaining the variation in retail price (**Y**) of personal computers based on four independent variables, monitor included (1=Yes, 0=No) (**X1**), CPU *Speed* in Mhz (**X2**), *RAM* in MB's (**X3**), and *Hard drive* capacity in GB's (**X4**).

Regression Analysis: Y versus X1; X2; X3; X4; X2X4

```
The regression equation is
Y = 1404 + 49 X1 - 3.37 X2 + 4.72 X3 - 105 X4 + 0.644 X2X4

Predictor        Coef     SE Coef          T        P       VIF
Constant         1404        1765       0.80    0.433
X1               48.7       240.5       0.20    0.841       1.0
X2             -3.372       4.689      -0.72    0.478       8.3
X3              4.721       3.005       1.57    0.127       2.2
X4             -104.9       304.6      -0.34    0.733     133.3
X2X4           0.6442      0.6967       0.92    0.363     176.2

S = 697.0       R-Sq = 70.5%     R-Sq(adj) = 65.5%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	34753583	6950717	14.31	0.000
Residual Error	30	14573666	485789		
Total	35	49327250			

Source	DF	Seq SS
X1	1	252592
X2	1	21234267
X3	1	5713693
X4	1	7137818
X2X4	1	415213

Unusual Observations

Obs	X1	Y	Fit	SE Fit	Residual	St Resid
23	1.00	1900	3364	441	-1464	-2.71R
24	1.00	6360	4511	440	1849	3.42R

R denotes an observation with a large standardized residual

Durbin-Watson statistic = 2.07

Predicted Values for New Observations

New Obs	Fit	SE Fit	95.0% CI	95.0% PI
1	1170	259	(640; 1700)	(-349; 2689)

Values of Predictors for New Observations

New Obs	X1	X2	X3	X4	X2X4
1	1.00	400	64.0	5.00	2000

Best Subsets Regression: Y versus X1; X2; X2X4

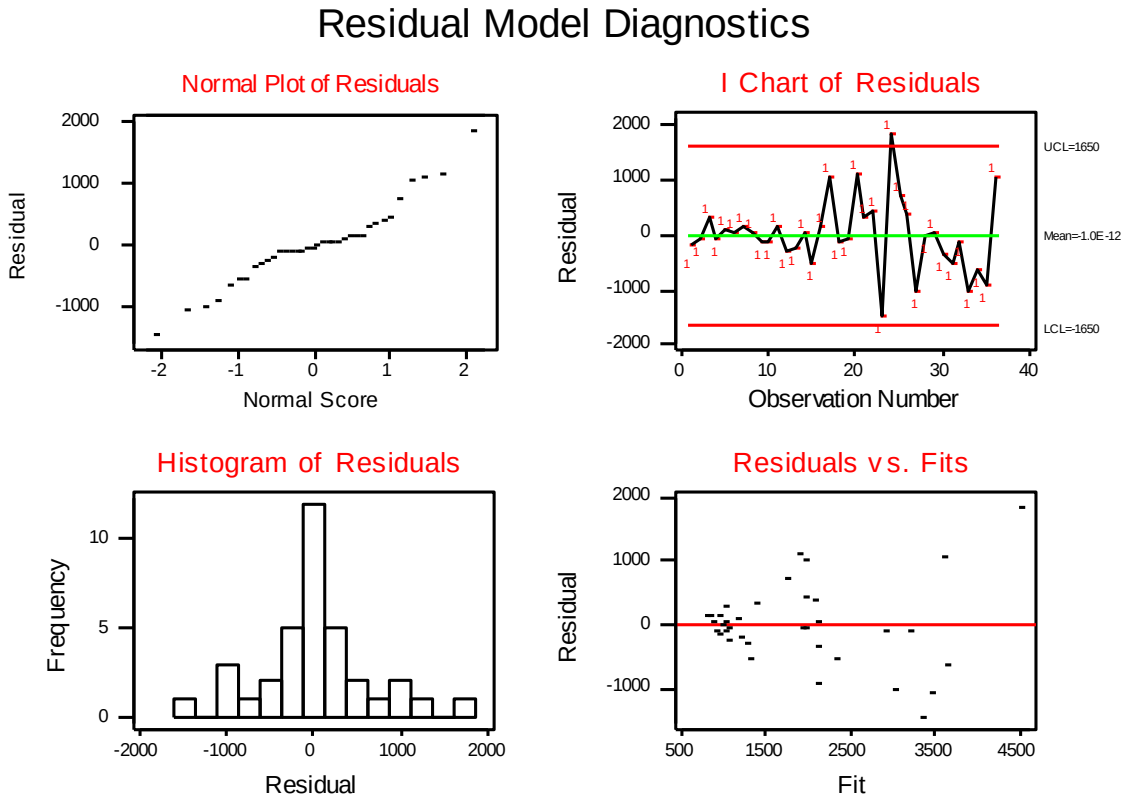
Response is Y

Vars	R-Sq	R-Sq(adj)	C-p	S	X1	X2	X2X4
1	67.7	66.8	0.3	684.26			X
1	42.8	41.2	25.2	910.58		X	
2	68.0	66.0	2.0	691.92		X	X
2	67.8	65.8	2.2	693.89	X		X
3	68.0	65.0	4.0	702.25	X	X	X

Correlations: Y; X1; X2; X3; X4

	Y	X1	X2	X3
X1	0.072			
	0.678			
X2	0.655	-0.020		
	0.000	0.910		
X3	0.691	0.045	0.658	
	0.000	0.795	0.000	
X4	0.819	0.083	0.761	0.708
	0.000	0.632	0.000	0.000

```
Cell Contents: Pearson correlation
               P-Value
```



Given this output and your knowledge of multiple regression, answer the following;

a. The slope of the *Speed* variable is	
b. Is the relationship between *RAM* and *Hard drive* significant? Why?	H_o: H_A:
	Decision:
c. Are the independent variables significant in explaining the variation in the *Price*? Why?	H_o: H_A:
	Decision:
d. Do *Speed* and *Hard drive* interact on varying the value of the *Price*? Why?	H_o: H_A:
	Decision:

e. Check the assumptions of the multiple regression	
f. What will be the *Price* of a computer *including* the monitor, has a *Speed* of 400 Mhz, a *RAM* of 64 MB's and a *Hard drive* capacity of 5 GB's?	
g. A 95% CI for the *Price* of a computers with the specs in (f) is	
h. The two variables that have no severe multicolinearity are	
i. A 99% CI for the slope of the *Hard drive capacity* of the computer is	
j. The percentage of variation in *Price* explained by the indep. variables is	
k. The estimated variance of the regression model is	
l. Using the best subset regression option, what is the best group of indep. variables that explain the variation in the *Price*? And has a C-p value of	

With My Best Wishes