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

STAT 212: BUSINESS STATISTICS II

Semester 163
Second Major Exam
Sunday August 13, 2017
5:00 pm – 6:30 pm

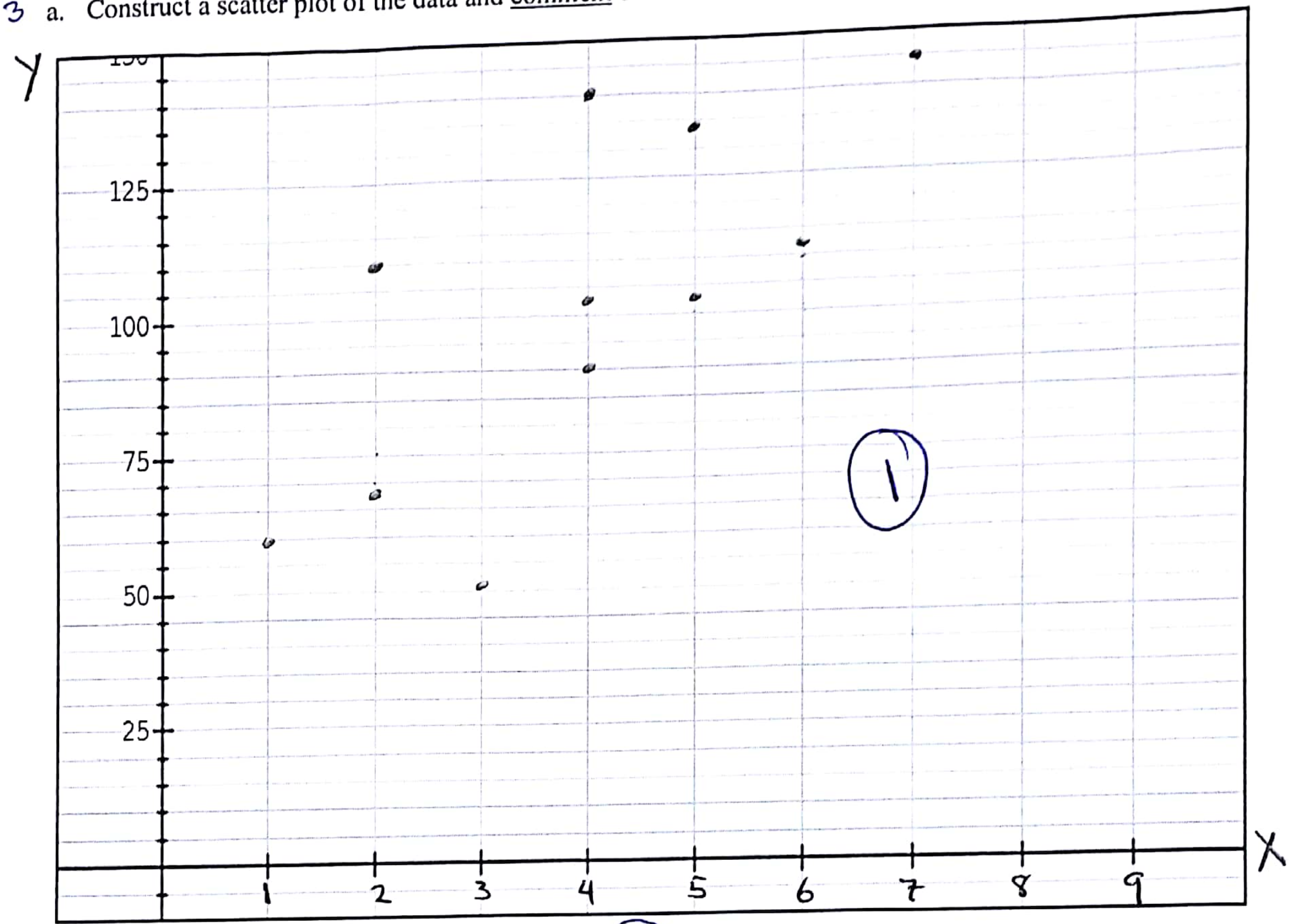
Name: KEY ID#: 0000 Section: 0 **Serial:** 00

Question No	Full Marks	Marks Obtained
1	20	
2	28	
3	32	
Total	**80**	

1. Accu-Copiers, Inc., sells and services the Accu-500 copying machine. As part of its standard service contract, the company agrees to perform routine service on this copier. To obtain information about the time it takes to perform routine service, Accu-Copiers has collected data for 11 service calls. The data are as follows:

Copiers serviced (C)	4	2	5	7	1	3	4	5	2	4	6	X
Minutes required (M)	140	68	103	145	60	51	103	134	110	90	112	Y

3 a. Construct a scatter plot of the data and comment on it.



1

There is a linear, positive (direct), moderate 1 relationship between X & Y.

$S_{CC} = S_{xx} = 32.9091$

$S_{MM} = S_{yy} = 10144.7273$

$S_{CM} = S_{xy} = 410.454545$

2

3

&5 b. The correlation coefficient =

$$r = \frac{S_{CM}}{\sqrt{S_{CC} \cdot S_{MM}}} = \frac{410.4545}{\sqrt{(32.9091)(10144.7273)}} = \boxed{0.7104}$$

(1) (1)

Interpretation: There is a linear, Positive, relatively Strong relationship between C & M.

(1) (1)

10 c. Do the data provide sufficient evidence to conclude that there is a direct correlation between the number of copiers serviced and the time it takes to be serviced? Use a significance level of 0.025.

Hypotheses are: H_0: $\rho \leq 0$ $\quad \alpha = 0.025$ (1)

H_1: $\rho > 0$

Assumptions are: * Both variables are random. (2)

* (C,M) has the Bivariate Normal dist.

Test statistic value: $t_0 = r\sqrt{\frac{n-2}{1-r^2}} = (0.7104)\sqrt{\frac{11-2}{1-(0.7104)^2}}$ (1)

$= \boxed{3.028}$ (1)

Critical value: $t_{\alpha, n-2} = t_{0.025, 9} = \boxed{2.2622}$ (1)

Decision Rule & Decision: If $t_0 > t_\alpha$ then Reject H_0 (1)

Since $3.028 > 2.2622$ then Reject H_0 (1)

Conclusion: There is (1) enough evidence that there is (1) a direct correlation between the number of Copies and the time in Minutes required at 2.5% level of significance.

2. Enterprise Industries produces FRESH, a brand of liquid laundry detergent. In order to study the relationship between the price and demand for FRESH, the company has gathered data concerning demand for FRESH over the last 30 sales periods where,

X: The *price* (in dollars) per bottle of FRESH and **Y**: The *demand* for FRESH (in 100,000's of bottles)

The following sums were obtained,

$n = 30$, $\sum x = 112.05$, $\sum x^2 = 418.742$, $\sum y = 251.48$, $\sum y^2 = 2121.53$, $\sum xy = 938.442$, and $SSE = 10.495$

Assuming that X is the independent variable and Y is the dependent variable then

2 a. Assumptions of a regression model:

* Linear relationship between X & Y.
* Independent Residuals.
* Normally dist. residuals
* Equal-variance residuals.

2

6 b. Fitted regression equation is:

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 418.742 - \frac{(112.05)^2}{30} = \boxed{0.23525}\ 1$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 2121.53 - \frac{(251.48)^2}{30} = \boxed{13.4569}\ 1$$

$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 938.442 - \frac{(112.05)(251.48)}{30} = \boxed{-0.8358}\ 1$$

$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{-0.8358}{0.23525} = \boxed{-3.5528}\ 1$$

$$b_0 = \bar{y} - b_1\bar{x} = \frac{251.48}{30} + (3.5528)\left(\frac{112.05}{30}\right) = \boxed{21.6524}\ 1$$

$$\Rightarrow \hat{Y} = 21.6524 - 3.5528X \quad 1$$

5 c. The standard error of the estimate is:

$$SST = S_{yy} = 13.4569 \ 1, \quad SSR = b_1 S_{xy} = (-3.5528)(-0.8358) = 2.9694 \ 1$$

$$SSE = SST - SSR = 13.4569 - 2.9694 = \boxed{10.4875}\ 1$$

$$MSE = \frac{SSE}{n-2} = \frac{10.4875}{30-2} = \boxed{0.3746}\ 1$$

$$\Rightarrow S_{Y\cdot X} = \sqrt{MSE} = \sqrt{0.3746} = \boxed{0.61201}\ 1$$

2 d. The predicted value of the *demand* if the *price* was $4.00 is:

$\hat{Y} = 21.6524 - 3.5528(4)$ (1) $= \$7.4412$ (in 100,000's)

$= \$744,120$ (1)

3 e. A 99% C.I. for the *demand* if the *price* of a bottle was $4.00 is:

$\hat{Y}_{|x=4} \pm t_{0.005,28}\, S_{YX}\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}$ (1)

$7.4412 \pm (2.7633)(0.61201)\sqrt{1+\frac{1}{30}+\frac{(4-3.735)^2}{0.23525}}$ (1)

$[5.4895, 9.3929]$ (1)

10 f. Do you think that the demand will increase by at most 300,000 bottles if the price was decreased by $1? Justify your answer using 10% significance level.

2 Hypotheses are: H_0: $\beta_1 \leq -3$ (1) $\alpha = 0.1$

H_1: $\beta_1 > -3$ (1)

Test statistic value:

$S_{b_1} = \frac{S_{YX}}{\sqrt{S_{xx}}} = \frac{0.61201}{\sqrt{0.23525}} = 1.2623$ (1)

3

$t_0 = \frac{-3.5528 + 3}{1.2623} = -0.4379$ (1) (1)

1 Critical value: $t_{\alpha, n-2} = t_{0.1, 28} = 1.3125$ (1)

Decision Rule & Decision: If $t_0 > t_\alpha$ then Reject H_0 (1)

2 Since $-0.4379 \not> 1.3125$ then we can NOT reject H_0. (1)

2 Conclusion: There is No evidence that the slope of the price is more than -3 at 10% level of significance. (2)

3. 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)
```

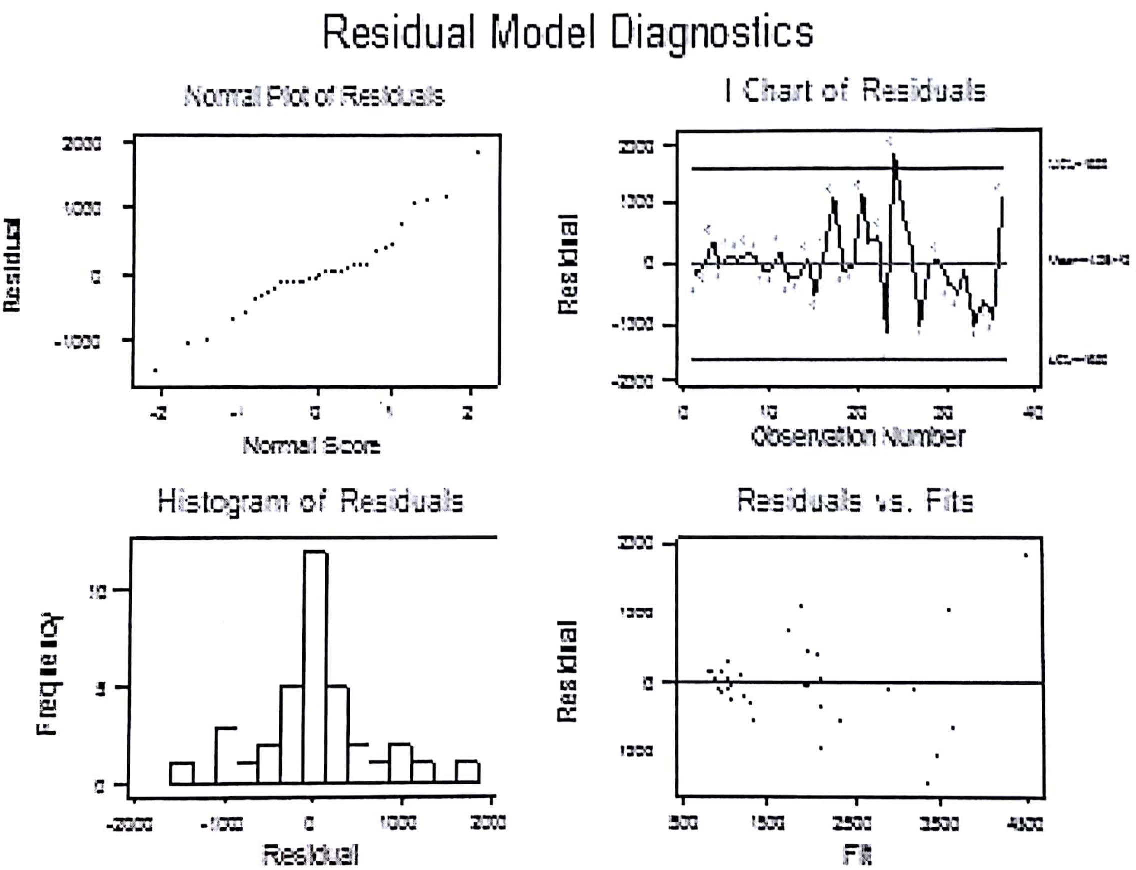
Values of Predictors for New Observations

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

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;

4 a. Interpret the slope of the predictor: *Monitor included*

If the monitor is included then the average (expected) retail price of a Computer will be $49 more than a one without a monitor if the other predictors kept constant. 1

3 b. Is the relationship between *RAM* and *Hard drive* significant? Why?

1 Yes, there is a significant relationship between X_2 & X_4. 1

The estimate of $\rho_{X_2, X_4} = 0.708 \Rightarrow$ p-value $= 0.000 < 0.05 = \alpha$ 1

4 c. Are the predictors significant in explaining the variation in the *Price*? Why?

(1) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_1: At least one $\beta_i \neq 0$ 1

(2) $F_0 = \frac{MSR}{MSE} = 14.31$

(3) p-value $= 0.000$ 1

(4) Since $0.000 < 0.05 = \alpha$ we can Reject H_0 1

(5) There is enough evidence that the overall model is sig. 1

4 d. Do you think that *Speed* and *Hard drive* interact on varying the value of the *Price*? Why?

Let $X_5 = X_2 X_4$

(1) $H_0: \beta_5 = 0$ (Interaction Not sig.)

$H_1: \beta_5 \neq 0$ (Interaction is sig.) 1

(2) p-value $= 0.363$ 1

(3) Since $0.363 \nless 0.05 = \alpha$ then we can NOT reject H_0 1

(4) No, there is NO enough evidence of an interaction between X_2 & X_4 1

4 e. Check the assumptions of the multiple regression From Residual Model Diagnostics

✓ * Linearity seems to be satisfied. 1

✓ * Independence " " " " . 1

✓ * Normality " " " " . 1

X * Equal-variance is NOT satisfied (Triangle shape) 1

2 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?

$$\hat{Y} = 1404 + 49(1) - 3.37(400) + 4.72(64) - 105(5) + 0.644(400 \times 5)$$ (1)

$$\cong \boxed{1170} = 1169.944$$ (1)

2 g. A 95% CI for the *Price* of a computer with the specs in (f) is

[640, 1700] = 95% P.I.

3 h. A 99% CI for the slope of the *Hard drive capacity* of the computer is

A $(1-\alpha)100\%$ C.I. for β_4

$b_4 \pm t_{\frac{\alpha}{2}, n-k-1} S_{b4}$ (1)

$-104.9 \pm t_{0.005,30} (304.6)$ (1)

$-104.9 \pm (2.75)(304.6)$

[−942.55, 732.73] (1)

2 i. The percentage of variation in *Price* explained by the variation in the predictors taking into account 4 predictors and the given sample size is

$R^2_{adj} = 65.5\%$

1 j. The estimated variance of the regression model is

MSE = 485789

3 k. Are the four predictors (as a whole) significant to the *Retail price*? If so, then which predictor(s) is/are the significant one(s)? Explain in detail.

From part c before we found that the overall model is significant to the Retail price.
On the other hand, None (1) of the predictors is significant (individually) to the Retail price which is a Contradiction (1).
This happened because of the VIF>5 (collinearity) (1)