KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICS AND STATISTICS
Term 151

STAT 212 BUSINESS STATISTICS II
Third Major Exam
Allowed time 75 minutes
Wednesday November 25, 2015

Please circle your instructor name:

R. Anabosi M. Saleh

Name:_____ ID #:_____ Srl #:___

Important Note:

1) You must **show all work** to obtain full credit for questions on this exam.
2) **DO NOT round** your answers at each step. Round answers only if necessary at your final step to **4 decimal places**.

Question No	Full Marks	Marks Obtained
Q1	29	
Q2	8	
Q3	14	
Q4	9	
Total	**60**	

Question One (16+3+2+2+6 points):

The Gilmore Accounting firm collected data in an effort to explain variation in client profitability. The variables used were:

Y = Net profit earned from the client
X_1 = Number of hours spent working with the client
X_2 = Type of client:
1 = manufacturing
2 = service
3 = governmental

A Minitab output is shown below.

Regression Analysis: Net profit versus Hours spent; Manufac; Service

The regression equation is
Net profit = - 586 + 22.9 Hours spent + 2302 Manufac + 1870 Service

Predictor	Coef	SE Coef	T	P	VIF
Constant		974.2		0.566	
Hours spent			0.78	0.461	1.481
Manufac		895.1		0.037	1.838
Service			2.45		1.564

S = R-Sq = 69.8% R-Sq(adj) = %

Analysis of Variance

Source	DF	SS	MS	F	P
Regression			5130297	5.39	0.031
Residual Error			951223		
Total	10	22049448			

Unusual Observations

Obs	Hours spent	Net profit	Fit	SE Fit	Residual	St Resid
2	56.0	4200	2564	677	1636	2.33R

R denotes an observation with a large standardized residual.

Durbin-Watson statistic = 2.46576

Using the output above, answer the following:

a. Fill in the gaps (shaded in gray) in the above output properly.
b. Interpret the regression coefficient of "Manufac".

c. Predict the Net profit from a governmental client if 34 hours spent working with him.

d. Is there a severe collinearity in the model above? Justify your answer.

e. Is the variable "Hours spent" significantly linearly related to "Net profit"? Explain.

Question Two (1+1+2+2+2 points):

A study, in India, on rice was conducted to examine the relationship between Y, the yield (kg/ha) of rice as a function of X, the number of days after flowering at which harvesting took place. Minitab output is given below.

Regression Analysis: y versus x

The regression equation is
y = 10.14 + 0.006353 x

S = 9.46438 R-Sq = 7.8% R-Sq(adj) = 1.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	105.96	105.957	1.18	0.295
Error	14	1254.04	89.574		
Total	15	1360.00			

Polynomial Regression Analysis: y versus x

The regression equation is
y = - 145.9 + 0.1062 x - 0.000016 x**2

S = 9.32427 R-Sq = 16.9% R-Sq(adj) = 4.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	229.75	114.877	1.32	0.300
Error	13	1130.25	86.942		
Total	15	1360.00			

Sequential Analysis of Variance

Source	DF	SS	F	P
Linear	1	105.957	1.18	0.295
Quadratic	1	123.796	1.42	0.254

Polynomial Regression Analysis: y versus x

The regression equation is
y = - 1925 + 1.818 x - 0.000557 x**2 + 0.000000 x**3

S = 9.01722 R-Sq = 28.3% R-Sq(adj) = 10.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	384.28	128.092	1.58	0.247
Error	12	975.72	81.310		
Total	15	1360.00			

Polynomial Regression Analysis: y versus x

The regression equation is

```
y = - 23736 + 29.6 x - 0.0137 x**2 + 0.000003 x**3 - 0.000000 x**4

S = 8.09441   R-Sq = 47.0%   R-Sq(adj) = 27.7%

Analysis of Variance

Source          DF       SS      MS     F      P
Regression       4   639.29  159.82  2.44  0.109
Residual Error  11   720.71   65.52
Total           15  1360.00
```

Polynomial Regression Analysis: y versus x

```
The regression equation is
y = - 48744 + 69 x - 0.039 x**2 + 0.000011 x**3 - 0.0000 x**4 +
0.0000 x**5

S = 8.47127   R-Sq = 47.2%   R-Sq(adj) = 20.9%

Analysis of Variance

Source          DF       SS      MS     F      P
Regression       5   642.38  128.48  1.79  0.203
Residual Error  10   717.62   71.76
Total           15  1360.00
```

Based on the output above, answer the following:

a. What is the size of the sample used in this study?
b. How many models are there in the output above?
c. Which model is significant in relating the Yield of rice to the Number of days after flowering, at 12% level of significance? Why?

d. Which model is the best model in predicting the variation in the Yield of rice using the Number of days after flowering? Why?

e. Based on your answer in part (d), predict the Yield of rice if was harvested 30 days after flowering.

Question Three (2+3+3+3+3 points):

A variety of sources uses simple body measurements to predict percentage Body fat. These simple body measurements include: Age, Weight (in pound), Height (in inch), Chest circumference (in cm), Abdomen circumference (in cm), Hip circumference (in cm), and Thigh circumference (in cm).

Given the following Minitab output

Regression Analysis: %Fat versus Age; Weight; ...

```
The regression equation is
%Fat = - 43.8 + 0.0937 Age - 0.221 Weight - 0.073 Height - 0.010 Chest
       + 0.921 Abdomen - 0.249 Hip + 0.742 Thigh

Predictor      Coef  SE Coef      T      P     VIF
Constant     -43.80    22.70  -1.93  0.060
Age         0.09370  0.08761   1.07  0.291   1.708
Weight     -0.22112  0.07028  -3.15  0.003  25.242
Height      -0.0733   0.1150  -0.64  0.527   1.674
Chest       -0.0103   0.1734  -0.06  0.953  10.680
Abdomen      0.9214   0.1631   5.65  0.000  17.650
Hip         -0.2486   0.2636  -0.94  0.351  23.991
Thigh        0.7416   0.2863   2.59  0.013  11.652

S = 3.94225   R-Sq = 83.8%   R-Sq(adj) = 81.2%

Analysis of Variance
Source          DF       SS      MS      F      P
Regression       7  3388.30  484.04  31.15  0.000
Residual Error  42   652.74   15.54
Total           49  4041.03

Unusual Observations
Obs   Age    %Fat     Fit  SE Fit  Residual  St Resid
 39  46.0  33.800  37.992   2.978    -4.192     -1.62 X
 42  44.0  31.700  31.483   3.786     0.217      0.20 X

X denotes an observation whose X value gives it large leverage.

Durbin-Watson statistic = 1.78689
```

Correlations: %Fat; Age; Weight; Height; Chest; Abdomen; Hip; Thigh

```
            %Fat      Age   Weight   Height    Chest  Abdomen      Hip
Age        0.517
           0.000

Weight     0.612    0.265
           0.000    0.063

Height    -0.266   -0.276    0.109
           0.061    0.052    0.451

Chest      0.723    0.376    0.912    0.014
           0.000    0.007    0.000    0.926
```

Abdomen	0.824	0.442	0.915	-0.052	0.942		
	0.000	0.001	0.000	0.717	0.000		
Hip	0.693	0.314	0.959	-0.045	0.911	0.942	
	0.000	0.026	0.000	0.754	0.000	0.000	
Thigh	0.682	0.219	0.937	-0.037	0.859	0.890	0.938
	0.000	0.127	0.000	0.801	0.000	0.000	0.000

Cell Contents: Pearson correlation
P-Value

Stepwise Regression: %Fat versus Age; Weight; ...

Backward elimination. Alpha-to-Remove: 0.05

Response is %Fat on 7 predictors, with N = 50

Step	1	2	3	4	5
Constant	-43.80	-44.37	-54.76	-61.69	-59.00
Age	0.094	0.094	0.110	0.112	
T-Value	1.07	1.09	1.34	1.37	
P-Value	0.291	0.284	0.186	0.179	
Weight	-0.221	-0.222	-0.247	-0.268	-0.277
T-Value	-3.15	-3.37	-4.63	-5.84	-6.04
P-Value	0.003	0.002	0.000	0.000	0.000
Height	-0.07	-0.07			
T-Value	-0.64	-0.64			
P-Value	0.527	0.523			
Chest	-0.01				
T-Value	-0.06				
P-Value	0.953				
Abdomen	0.921	0.916	0.911	0.863	0.950
T-Value	5.65	6.90	6.92	7.46	9.75
P-Value	0.000	0.000	0.000	0.000	0.000
Hip	-0.25	-0.25	-0.19		
T-Value	-0.94	-0.95	-0.77		
P-Value	0.351	0.346	0.443		
Thigh	0.74	0.74	0.80	0.74	0.65
T-Value	2.59	2.66	3.03	2.95	2.67
P-Value	0.013	0.011	0.004	0.005	0.010
S	3.94	3.90	3.87	3.85	3.89
R-Sq	83.85	83.85	83.69	83.47	82.78
R-Sq(adj)	81.16	81.59	81.84	82.00	81.66
Mallows Cp	8.0	6.0	4.4	3.0	2.8

Answer the following:

a. Comment, and do NOT interpret, on the regression coefficients of Weight, Height, Chest, and Hip. Explain in detail.

b. Estimate and interpret the percentage body fat of nobody and comment on it.

c. If one of the predictors is to be removed from the model, which one is that predictor? Why?

d. After some analysis, it was found that not all predictors are significantly useful to predict the percentage Body fat. Write the model with most useful predictors. Explain.

e. Use the model in d to predict the Body fat of a person whose age is 21, weight is 170, height is 70 inches, chest circumference is 100 cm, abdomen circumference is 90 cm, hip circumference is 105 cm, and whose thigh circumference is 60 cm?

Question Four (4+4+4 points):

Retail sales of electricity ($ Mil) are predicted using net generation by energy source. The different sources of energy included in the study are Coal, Petroleum, Natural Gas, Nuclear, and Hydroelectric. Given the Minitab output below, answer the questions that follow:

Correlations: Sales ($Mil); Coal; Petroleum; Natural Gas; Nuclear; Hydroelectric

	Sales ($Mil)	Coal	Petroleum	Natural Gas	Nuclear
Coal	0.881				
	0.000				
Petroleum	0.413	0.469			
	0.182	0.124			
Natural Gas	0.961	0.872	0.398		
	0.000	0.000	0.200		
Nuclear	0.912	0.853	0.348	0.960	
	0.000	0.000	0.268	0.000	
Hydroelectric	-0.480	-0.223	-0.461	-0.534	-0.527
	0.114	0.487	0.132	0.074	0.078

Cell Contents: Pearson correlation
P-Value

Regression Analysis: Sales ($Mil) versus Coal; Petroleum; ...

The regression equation is
Sales ($Mil) = 70612 + 0.0583 Coal - 0.041 Petroleum + 0.217 Natural Gas - 0.081 Nuclear - 0.035 Hydroelectric

Predictor	Coef	SE Coef	T	P
Constant	70612	75675	0.93	0.387
Coal	0.05834	0.07587	0.77	0.471
Petroleum	-0.0412	0.2140	-0.19	0.854
Natural Gas	0.2172	0.1032	2.11	0.080
Nuclear	-0.0806	0.1486	-0.54	0.607
Hydroelectric	-0.0354	0.1196	-0.30	0.777

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	5719764576	1143952915	17.37	0.002
Residual Error	6	395079130	65846522		
Total	11	6114843706			

Best Subsets Regression: Sales ($Mil) versus Coal; Petroleum; ...

Response is Sales ($Mil)

Vars	R-Sq	R-Sq(adj)	Mallows Cp	S	Coal	Petroleum	Natural Gas	Nuclear	Hydroelectric
1	92.4	91.7	-1.0	6799.5			X		
1	83.1	81.4	7.7	10158				X	
2	93.2	91.7	0.3	6797.2	X		X		
2	92.6	90.9	0.9	7093.8			X	X	
3	93.4	91.0	2.1	7078.5	X		X	X	
3	93.2	90.7	2.3	7197.6	X		X		X
4	93.5	89.8	4.0	7535.8	X		X	X	X
4	93.4	89.7	4.1	7567.3	X	X	X	X	
5	93.5	88.2	6.0	8114.6	X	X	X	X	X

a. What is the best group of predictors to fit the Retail sales of electricity? Why?

b. Is there a sever collinearity in the full regression model? Why?

c. Since the overall model is significant in predicting the Retail sales, which of the predictors is individually significant in predicting the Retail sales? Why?