KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICS & STATISTICS

Term 172

STAT 212: BUSINESS STATISTICS II

Exam III

18 April 2018 at 6:15 PM

Name:____

ID #:________________________________

Serial#:__________ **Section:** **1 2 (Al-Sawi) 3 (Abbas)**

Important Notes:

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	19	
Q2	13	
Q3	12	
Q4	8	
Q5	8	
Total	**60**	

1. In Hawaii, condemnation proceedings are under way to enable private citizens to own the property that their homes are built on. Until recently, only estates were permitted to own land, and homeowners leased the land from the estate. In order to comply with the new law, a large Hawaiian estate wants to use regression analysis to estimate the fair market value of the land. The following model was fit to data collected for n = 90 properties, 30 of which are located near a cove.

Model#1: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_1^2 X_2 + \epsilon$

where $Y=$ Sale price of property in thousands of dollars

X_1 = Size of property in thousands of square feet

X_2 = 1 if property located near cove, 0 if not

Using the data collected for the 90 properties, the following partial output is shown:

Model # 1 (Full Model)

```
The regression equation is
Sale price = - 199 + 0.391 Size + 403 Cove - 0.000037 Size^2
- 0.390 Size*Cove + 0.000080 Size^2 *Cove


Predictor            Coef     SE Coef      T      P      VIF
Constant           -198.6       130.9  -1.52  0.133
Size               0.3913      0.1188   3.30  0.051   41.058
Cove                403.5       285.2   1.41  0.161  110.679
Size^2        -0.00003718  0.00002445  -1.52  0.132   39.086
Size*Cove         -0.3900      0.2780  -1.40  0.164  457.894
Size^2 *Cove   0.00008046  0.00006458   1.25  0.216  144.662

S = 121.256   R-Sq = 60.6%   R-Sq(adj) = 58.3%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       5  1900039  380008  25.85  0.001
Residual Error  84  1235044   14703
Total           89  3135083
```

Model # 2

```
The regression equation is
Sale price = - 88 + 0.294 Size - 25.8 Cove - 0.000019
Size^2


Predictor         Coef     SE Coef      T      P
Constant         -87.5       111.0  -0.79  0.433
Size            0.2943      0.1025   2.87  0.050
Cove            -25.84       27.70  -0.93  0.354
Size^2     -0.00001879  0.00002168  -0.87  0.388


S = 121.667   R-Sq = 59.4%   R-Sq(adj) = 58.0%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3  1862045  620682  41.93  0.002
Residual Error  86  1273038   14803
Total           89  3135083
```

Model # 3

```
The regression equation is
Sale price = - 79 + 0.278 Size - 0.000015 Size^2


Predictor         Coef     SE Coef      T      P
Constant         -79.5       110.6  -0.72  0.474
Size            0.2776      0.1009   2.75  0.007
Size^2     -0.00001506  0.00002129  -0.71  0.481


S = 121.576   R-Sq = 59.0%   R-Sq(adj) = 58.0%


Analysis of Variance

Source          DF       SS      MS      F      P
Regression       2  1849165  924583  62.55  0.000
Residual Error  87  1285918   14781
Total           89  3135083
```

Model # 4.

```
The regression equation is
Sale price = - 11.8 + 0.214 Size + 43.5 Cove - 0.0326
Size*Cove


Predictor      Coef  SE Coef      T      P
Constant     -11.77    45.40  -0.26  0.796
Size        0.21358  0.02092  10.21  0.000
Cove          43.46    95.72   0.45  0.651
Size*Cove  -0.03259  0.04609  -0.71  0.481


S = 121.843   R-Sq = 59.3%   R-Sq(adj) = 57.9%


Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3  1858345  619448  41.73  0.000
Residual Error  86  1276738   14846
Total           89  3135083
```

I. (2+2 pts)

A. Using **Model # 4**, for a property located near cove, as the size increase by one sq.ft., the sale price will change, on average, by ……..

B. Interpret the coefficient of Size^2 in **Model # 2**:……….

II. Answer the following questions Using **Model #1**,

A. (4 pts) Is the overall model statistically adequate at $\alpha = 0.05$ for predicting sale price (Y)?

H_0:

H_1:

P-value:

Conclusion :

B. (2 pts) Which of the independent variables in the full model are significant at $\alpha = 0.05$? Explain.

C. (2 pts) Is there any contradiction between the answers of part A and part B? If Yes, Explain why.

D. (7 pts) Given a quadratic relationship between sale price (Y) and property size (X_1), what null hypothesis would you test to determine whether the curves differ from cove and non-cove properties?

H_0:
H_1:

Test Statistic:

Critical value:

Conclusion:

2. Suppose that the sales manager of a large automotive parts distributor wants to estimate the total annual sales for each of the company's regions. Five factors appear to be related to regional sales: the number of retail outlets in the region (x_1), the number of automobiles in the region registered as of April 1 (x_2), the total personal income recorded in the first quarter of the year (x_3), the average age of the automobiles (x_4), and the number of sales supervisors in the region (x_5). The data for each region were gathered for last year. For example, see the following table. In region 1 there were 1,739 retail outlets stocking the company's automotive parts, there were 9,270,000 registered automobiles in the region as of April 1, and so on. The region's sales for that year were $37,702,000.

Annual Sales ($ millions)	Number of Retail Outlets	Number of Automobiles Registered (millions)	Personal Income ($ billions)	Average Age of Automobiles (years)	Number of Supervisors
y	x_1	x_2	x_3	x_4	x_5
37.702	1,739	9.27	85.4	3.5	9
24.196	1,221	5.86	60.7	5	5
32.055	1,846	8.81	68.1	4.4	7
.	.	.	.	.	.

Use the following Minitab outputs to answer the questions on next page:

```
The regression equation is
Sales = −19.7 − 0.00063 outlets + 1.74 autos + 0.410 income
        + 2.04 age − 0.034 bosses

          Predictor          Coef      SE Coef        T        P
          Constant        −19.672        5.422    −3.63    0.022
          outlets       −0.000629     0.002638    −0.24    0.823
          automobiles      1.7399       0.5530     3.15    0.035
          income          0.40994      0.04385     9.35    0.001
          age              2.0357       0.8779     2.32    0.081
          bosses          −0.0344       0.1880    −0.18    0.864

Analysis of Variance
          SOURCE           DF        SS       MS        F        P
          Regression        5   1593.81   318.76   140.36    0.000
          Residual Error    4      9.08     2.27
          Total             9   1602.89
```

```
The regression equation is
Sales = −18.9 + 1.61 autos + 0.400 income + 1.96 age

          Predictor          Coef      SE Coef        T        P
          Constant        −18.924        3.636    −5.20    0.002
          automobiles      1.6129       0.1979     8.15    0.000
          income          0.40031      0.01569    25.52    0.000
          age              1.9637       0.5846     3.36    0.015

Analysis of Variance
          SOURCE           DF        SS       MS        F        P
          Regression        3   1593.66   531.22   345.25    0.000
          Residual Error    6      9.23     1.54
          Total             9   1602.89
```

(1) (3 pts) What percent of the variation in annual sales is explained by the regression equation including all 5 predictors, **taking into account** the **sample size** and **number of independent variables**?

(2) (6 pts) Using $\alpha = 0.05$, Test the following hypothesis:

$H_0 : \beta_1 = \beta_5 = 0$
$H_1 :$ *at least one of them is significantly different from zero*

Test Statistic:

Critical value:

Conclusion:

(3) (2 pts) Which variable is the most significant variable in explaining the annual sales? Explain.

(4) (2 pts) Which variable is the least significant variable in explaining the annual sales? Explain.

3. A human resource (HR) director wants to recruit sales managers. The company has the data for its 45 regions on 5 variables i.e.

Sales (**y**) —> Ratio of yearly sales divided by the target sales value for that region.

Wonder (**x_1**) —> Score from the Wonderlic Personnel Test.

SC (**x_2**) —> Score on the Strong-Campbell Interest Inventory Test.

Experience (**x_3**) —> Number of years of selling experience prior to becoming a sales manager.

Engineer (**x_4**) —> Dummy variable that equals 1 if the sales manager has a degree in electrical engineering and 0 otherwise.

The director wants to fit a linear regression for predicting sales of region and he feels that the sales can possibly depend on Wonder, SC, Experience and Engineer. To decide which variables should be included in the model, the director runs a stepwise regression and a best-subset regression. The Minitab outputs are given below:

<u>Stepwise Regression: Sales versus Wonder, SC, Experience, Engineer</u>

```
Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

Response is Sales on 4 predictors, with N = 45

Step            1       2
Constant    31.47   26.89

SC           1.36    1.34
T-Value      7.39    7.48
P-Value     0.000   0.000

Engineer              7.3
T-Value              1.83
P-Value             0.075

S            11.8    11.5
R-Sq        55.97   59.22
R-Sq(adj)   54.95   57.28
```

<u>Best Subsets Regression: Sales versus Wonder, SC, Experience, Engineer</u>

Vars	R-Sq	R-Sq(adj)	Mallows Cp	S	Wonder	SC	Experience	Engineer
1	56.0	55.0	2.3	11.782		X		
1	5.8	3.6	51.6	17.232	X			
1	4.9	2.6	52.6	17.320				X
2	59.2	57.3	1.1	11.474		X		X
2	56.1	54.0	4.2	11.904		X	X	
2	56.0	53.9	4.3	11.917	X	X		
3	59.3	56.3	3.0	11.598		X	X	X
3	59.2	56.2	3.1	11.613	X	X		X
3	56.1	52.9	6.1	12.045	X	X	X	
4	59.3	55.3	5.0	11.742	X	X	X	X

(A) (2 pts) Write down the final model through the stepwise regression.

(B) (6 pts) Based on the output "Best Subsets Regression",What are the best **two** candidate models based on:

1. Adjusted R-square criterion?

2. Mallows Cp criterion?

3. Standard error?

(C) (2 pts) Which variable is most strongly correlated with sales (*y*)? Justify your answer.

(D)(2 pts) What would be the final model through the stepwise regression if Alpha-to-Enter= 0.05 and Alpha-to-Remove = 0.05. Justify your answer.

4. (8 pts) Suppose we estimate the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$ using 36 months of data. From the regression results we calculate a Durbin-Watson test statistic of 1.03. **What can we conclude about the possibility of positive autocorrelation in this model at=0.05?**

H_0:
H_1:

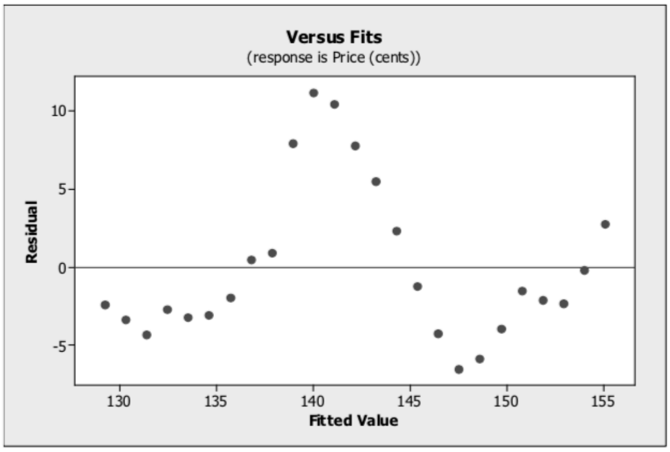
Test Statistic:

Critical values:

Decision Rules:

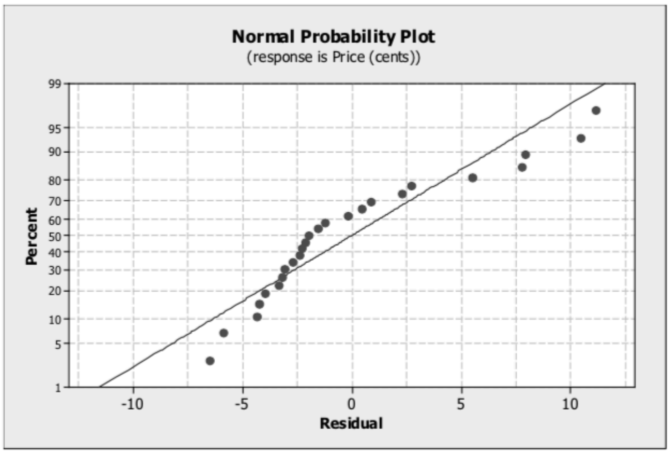
Conclusion:

5. (8 pts) For each of the following graphs, Which assumption appears to be violated? Explain



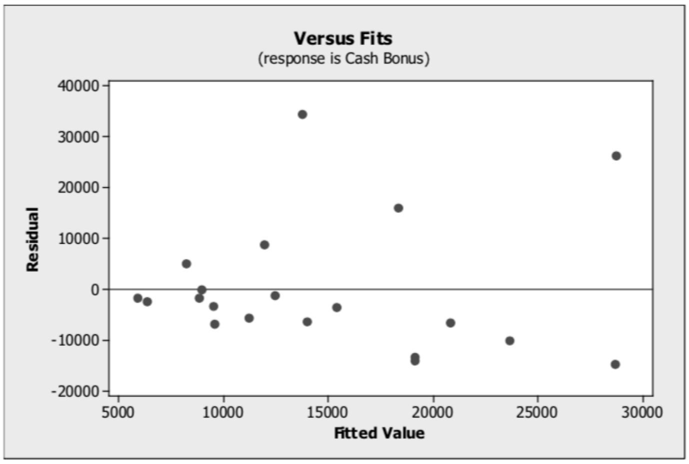
ANSWER:
Assumption appears to be violated:

Reason:



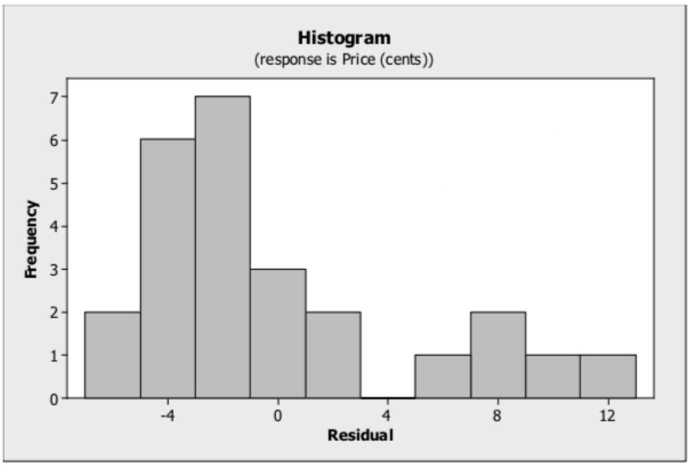
ANSWER:
Assumption appears to be violated:

Reason:



ANSWER:
Assumption appears to be violated:

Reason:



ANSWER:
Assumption appears to be violated:

Reason: