KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICS & STATISTICS
DHAHRAN, SAUDI ARABIA

STAT 212: Statistics for Business II
Semester 161
Third Major Exam
Wednesday December 21, 2016
7:00 pm

Please circle your instructor's name:

Abbas Al-Sawi

Name: ID #: Section #: Serial #:

Question No	Full Marks	Marks Obtained
1	**10**	
2	**12**	
3	**06**	
4	**06**	
5	**08**	
6	**18**	
Total	**60**	

Statement 1: Predicting percentage of problem mortgages. Best's Review (June 1999) compared the mortgage loan portfolios for a sample of 25 life/health insurance companies. Suppose you want to model the percentage of problem mortgages (Y) of company as a function of total mortgage loan (X1), percentage of invested assets (X2), percentage of commercial mortgages (X3), and percentage of residential mortgages (X4).

Q.No.1:- (2+4+2+2 = 10 points) Use the computer outputs to answer the following questions:

Regression Analysis: Y(ProbMort) versus X1(MortLoan), X2(Assets), ...

```
The regression equation is
Y(ProbMort) = 28.9 - 0.000000 X1(MortLoan) + 0.844 X2(Assets)
              - 0.360 X3(CommMort) - 0.300 X4(ResMort)

Predictor            Coef     SE Coef      T      P    VIF
Constant            28.87       12.67   2.28  0.034
X1(MortLoan)  -0.00000011  0.00000028  -0.38  0.708  1.346
X2(Assets)         0.8440      0.2326   3.63  0.002  1.302
X3(CommMort)      -0.3600      0.1316  -2.74  0.013  1.586
X4(ResMort)       -0.3003      0.1834  -1.64  0.117  1.567

S = 5.98917   R-Sq = 51.2%   R-Sq(adj) = 41.5%

Analysis of Variance

Source          DF       SS      MS     F      P
Regression       4   753.76  188.44  5.25  0.005
Residual Error  20   717.40   35.87
Total           24  1471.17
```

(i) Write the first order model for E(y).

(ii) Test the overall usefulness of the model in part (a) using $\alpha=0.05$

H_0:

H_1:

Decision and conclusion:

(iii) Interpret the coefficients of X1 and X2.

(iv) Estimate the mean percentage of problem mortgages if total mortgage loan = 18803163, percentage of invested assets=20.5, percentage of commercial mortgages=98, and percentage of residential mortgages=0.2.

Q.No.2:- (2+6+2+2= 12 points) Referring to "Statement 1," the researcher decided that the variables X2 and X4 warrant inclusion in the model as a second-order (i.e. squared terms). Use the computer outputs to answer the following questions:

Regression Analysis: Y(ProbMort) versus X1(MortLoan), X2(Assets), ...

```
The regression equation is
Y(ProbMort) = 56.2 - 0.000000 X1(MortLoan) - 1.82 X2(Assets)
              - 0.449 X3(CommMort) + 0.223 X4(ResMort) + 0.0771 X2**2
              - 0.0189 X4**2

Predictor            Coef     SE Coef      T      P     VIF
Constant            56.17       13.81   4.07  0.001
X1(MortLoan)  -0.00000008  0.00000025  -0.31  0.760   1.559
X2(Assets)        -1.8177      0.9935  -1.83  0.084  34.692
X3(CommMort)      -0.4494      0.1127  -3.99  0.001   1.698
X4(ResMort)        0.2227      0.6079   0.37  0.718  25.152
X2**2             0.07707     0.02665   2.89  0.010  32.227
X4**2            -0.01887     0.02334  -0.81  0.429  27.113

S = 4.95611   R-Sq = 69.9%   R-Sq(adj) = 59.9%

Analysis of Variance

Source          DF       SS      MS     F      P
Regression       6  1029.03  171.51  6.98  0.001
Residual Error  18   442.13   24.56
Total           24  1471.17
```

(i) Which variable(s) contribute significant information for the prediction of the percentage of problem mortgages? Test using $\alpha=0.05$.

(ii) Using partial-F test, can we conclude that both second-order terms of the model contribute information for the prediction of the percentage of problem mortgages? Test using $\alpha=0.05$

H_0:

H_1:

Test Statistic: $F_0=$ with $v_1=$ and $v_2=$

Decision and conclusion:

(iii) Is there evidence of severe collinearity in the model? if yes, which variable should be removed first.

(iv) Interpret the coefficient of X2**2.

Q.No.3:- (2+2+2 = 06 points) Referring to "Statement 1," the researcher used the best subset technique for model building.

Best Subsets Regression: Y(ProbMort) versus X1(MortLoan), X3(CommMort), ...

Response is Y(ProbMort)

Vars	R-Sq	R-Sq(adj)	Mallows Cp	S	X1(MortLoan)	X3(CommMort)	X4(RMortMx2**2)
1	38.5	35.8	10.0	6.2721			X
1	8.8	4.9	24.9	7.6363	X		
2	53.9	49.7	4.2	5.5530		X	X
2	38.7	33.1	11.9	6.4022	X		X
3	59.8	54.0	3.3	5.3101		X	X X
3	53.9	47.3	6.2	5.6822	X	X	X
4	60.3	52.3	5.0	5.4063	X	X	X X

Based on the output of "**Best Subsets Regression",** What are the candidate models using:

(i) Adjusted R-square criterion?

Reason:

(ii) Mallows Cp criterion?

Reason:

(iii) Standard error?

Reason:

Q.No.4:- (2+2+2+= 06 points) Referring to "Statement 1," the researcher used the stepwise regression technique for model building.

Stepwise Regression: Y(ProbMort) versus X1(MortLoan), X3(CommMort), ...

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

Response is Y(ProbMort) on 4 predictors, with N = 25

Step                1        2        3
Constant       0.8450  24.5637  35.9328

x2**2          0.0225   0.0249   0.0250
T-Value          3.79     4.67     4.91
P-Value         0.001    0.000    0.000

X3(CommMort)            -0.266   -0.379
T-Value                  -2.71    -3.33
P-Value                  0.013    0.003

X4(ResMort)                       -0.28
T-Value                           -1.75
P-Value                           0.095

S                6.27     5.55     5.31
R-Sq            38.50    53.89    59.75
R-Sq(adj)       35.82    49.70    54.00
Mallows Cp       10.0      4.2      3.3
```

Based on the results of the stepwise regression,

(i) Which predictor will be chosen as the first variable entering in to the model?

(ii) Write the prediction equation for the best model.

(iii) Which of the predictor(s) of the best selected model are significant at 7% significance level?

Q.No.5:- (6+2 = 08 points) For a sample of vehicles described in Consumer Reports' New Car Preview 2001, data include highway fuel economy test results, vehicle weights, and vehicle types. Data describes 53 sports-utility vehicles in terms of their highway fuel economy (HwyMpg), weight and vehicle type (SUV). The variable SUV is coded as 1 for sports-utility vehicles and 0 for other vehicles. We obtained the following fitted regression equation and R^2:

HwyMpg = 52.3 – 0.00534Weight – 6.01SUV with R^2= 0.784

(i) Interpret all the three coefficients of the above equation.

(ii) Interpret the value of R^2.

Q.No.6:- (2×9 = 18 points) The computer output for the multiple regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \in$ is shown below. However, because of a printer malfunction some of the results are not shown. These are indicated by the boldface letters ***a*** to ***i***. Fill in the missing results (up to three decimal places).

Predictor	*Coef*	*StDev*	*T*
Constant	***a***	6.15	4.11
x_1	3.51	***b***	1.25
x_2	-0.71	0.30	***c***

S = ***d*** R-Sq = ***e***

Analysis of Variance

Source of Variation	*df*	*SS*	*MS*	*F*
Regression	2	412	***g***	***i***
Error	37	***f***	***h***	
Total	39	974		

a = _____

b = _____

c = _____

d = _____

e = _____

f = _____

g = _____

h = _____

i = _____

Some useful formulas

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}, \quad R^2_{adj} = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}, \quad S = \sqrt{\frac{SSE}{n-k-1}}$$

$$t_0 = \frac{\hat{\beta}_j - 0}{se(\hat{\beta}_j)}, \quad F_0 = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}, \quad F_0 = \frac{[\text{SSR(all)} - \text{SSR(all except new set of variables)}]/m}{\text{SSE(all)}/(n-k-1)}$$