<p style="text-align:center"><strong style="color:red">SOLUTIONS
King Fahd University of Petroleum & Minerals
Department of Mathematics & Statistics
<strong style="color:red">STAT-212-Lab-Exam</p>

Name: ID: Serial:

An economist from a major east coast bank has collected on major cities in the US. You are to develop an estimate of a multiple regression model that would allow you to predict the Labor Market Stress Index (Y), based on the other variables

X1: 1995 populations
X2: Growth 1990-1995
X3: 1998 unemployed
X4: SAT City
X5: SAT suburb
X6: 1998 incomes
X7: 998 White-collar

Use the following correlation matrix, answer 1 and 2:
Correlations: Y, X1, X2, X3, X4, X5, X6, X7

```
            Y        X1        X2        X3        X4        X5        X6
X1     0.099
       0.582

X2    -0.133     0.072
       0.462     0.692

X3    -0.528     0.228    -0.098
       0.002     0.201     0.588

X4    -0.022     0.011     0.395    -0.142
       0.905     0.954     0.023     0.429

X5    -0.065     0.003     0.231    -0.085     0.903
       0.721     0.986     0.197     0.639     0.000

X6     0.188     0.272    -0.353     0.116    -0.026     0.053
       0.295     0.125     0.044     0.520     0.884     0.768

X7     0.193     0.643     0.080     0.082     0.030     0.055     0.191
       0.282     0.000     0.659     0.650     0.870     0.760     0.288
```

1. The number of independent variables that are significantly correlated with Y = **1**

(Which is x3)

2. The number of independent variables that are significantly correlated with X5 = **1**

(Which is x4)

2

Using the best subset regression, answer questions 3 and 4:
Best Subsets Regression: Y versus X1, X2, X3, X4, X5, X6, X7

```
Response is Y

                      Mallows        X X X X X X X
Vars  R-Sq  R-Sq(adj)     Cp      S  1 2 3 4 5 6 7
  1   27.9      25.6     2.3  9.7577      X
  1    3.7       0.6    12.8  11.277                  X
  2   34.2      29.8     1.6  9.4762      X       X
  2   33.5      29.1     1.9  9.5250      X           X
  3   38.0      31.6     1.9  9.3575      X       X X
  3   37.8      31.4     2.0  9.3700    X X           X
  4   39.9      31.3     3.1  9.3737    X X       X X
  4   39.8      31.2     3.1  9.3792      X     X X X
  5   41.0      30.1     4.6  9.4594    X X     X X X
  5   40.5      29.5     4.8  9.4978  X X X       X X
  6   42.0      28.6     6.2  9.5549      X X X X X X
  6   41.5      28.0     6.4  9.5988  X X X     X X X
  7   42.4      26.3     8.0  9.7111  X X X X X X X
```

3. Find the number of independent variables in the best model.**3**

(Which are: X$_3$, X$_6$, and X$_7$)

Justify your answer: **Because it has the highest R^2- adjusted and smallest Standard error Sε**

4. the coefficient of determination for the best model

R^2= 38.0%

Using the multiple regression model with all the independent variables and answer 5 to 14

Regression Analysis: Y versus X1, X2, X3, X4, X5, X6, X7

```
Predictor         Coef      SE Coef      T       P     VIF
Constant        104.74        25.58    4.10   0.000
X1         0.00000056   0.00000137    0.41   0.684   1.915
X2            -0.4360       0.4539   -0.96   0.346   1.556
X3            -7.088        1.915    -3.70   0.001   1.103
X4             0.00687      0.01084   0.63   0.532   7.022
X5            -0.008158     0.009617 -0.85   0.404   6.244
X6             0.0005357    0.0005879 0.91   0.371   1.308
X7             0.0007415    0.0008266 0.90   0.378   1.750

S = 9.71114   R-Sq = 42.4%   R-Sq(adj) = 26.3%

Analysis of Variance

Source            DF        SS       MS      F       P
Regression         7   1736.40   248.06   2.63   0.035
Residual Error    25   2357.66    94.31
Total             32
```

5. The intercept of the model = **104.74**
6. The slope of X_3 = **-7.088**
7. S_{b5} = **0.009617**
8. The test statistic for testing the significance of X_3 = **-3.70**
9. The p-value for the significance test of X_7 = **0.378**
10. The variance inflation factor for X_4 = **7.022**
11. S_e = **9.71114**
12. R^2 = **42.4%**
13. SST = **SSR+SSE=1736.40+2357.66=4094.06**
14. The p-value for the significance test of the overall model : **0.035**
15. Do you think that there is any colinearity between any of the independent variables? Explain.

Yes, because the VIF value for X_4 and X_5 is more than 5