Assignment 2 – Relevant SAS Output

(a) Segmentation Analysis: Output from performing cluster analysis using consumers' response to survey questions (*x1-x12*) in *Calibration.xls*

```
                                   The FASTCLUS Procedure
                    Replace=FULL  Radius=0  Maxclusters=2 Maxiter=20  Converge=0.02

                                       Cluster Summary

                                         Maximum Distance
                            RMS Std             from Seed      Radius      Nearest      Distance Between
Cluster      Frequency     Deviation       to Observation      Exceeded     Cluster     Cluster Centroids
─────────────────────────────────────────────────────────────────────────────────────────────────────────
   1               192        0.9210               6.5667                         2                8.2252
   2               101        0.9871               7.3471                         1                8.2252


                        7 Observation(s) were omitted due to missing values.


                                   Statistics for Variables

                   Variable      Total STD     Within STD       R-Square     RSQ/(1-RSQ)
                   ──────────────────────────────────────────────────────────────────────
                   x1              1.38836        1.36327       0.039144        0.040738
                   x2              1.34963        1.28276       0.099774        0.110832
                   x3              1.31256        1.14084       0.247167        0.328316
                   x4              1.36519        1.05113       0.409231        0.692708
                   x5              1.38784        0.91479       0.567038        1.309669
                   x6              1.43532        0.82030       0.674507        2.072267
                   x7              1.49185        0.75526       0.744593        2.915325
                   x8              1.53485        0.70698       0.788570        3.729707
                   x9              1.57357        0.66992       0.819382        4.536535
                   x10             1.57250        0.68533       0.810714        4.283008
                   x11             1.59060        0.74935       0.778827        3.521356
                   x12             1.60994        0.85310       0.720185        2.573793
                   OVER-ALL        1.47117        0.94439       0.589349        1.435160


                                Pseudo F Statistic =   417.63


                        Approximate Expected Over-All R-Squared =   0.08070


                            Cubic Clustering Criterion =  197.008

                  WARNING: The two values above are invalid for correlated variables.
```

```
                         Cluster Means

Cluster             x1             x2             x3             x4             x5             x6
-------------------------------------------------------------------------------------------------
   1       2.497382199    2.694736842    3.078947368    3.484210526    3.784210526    4.015789474
   2       1.920000000    1.797979798    1.704081633    1.650000000    1.585858586    1.540000000

                                          Cluster Means

Cluster             x7             x8             x9            x10            x11            x12
-------------------------------------------------------------------------------------------------
   1       4.273684211    4.481481481    4.631578947    4.773684211    4.831578947    4.841269841
   2       1.570000000    1.612244898    1.640000000    1.800000000    1.878787879    1.969696970


              Distance Between Cluster Centroids

              Distance Between Cluster Centroids

                           Nearest
                           Cluster                1                 2
                           -------------------------------------------
                                     1           .                  8.225210474
                                     2          8.225210474         .
```

(b) Response Analysis: Output from performing discriminant analysis using demographic variables

(*dist, Age, Gender, Married, License, Adults, Children, Cars, Education, SpouseEd, Years, Workers, Income, Ethnic*) in *Calibration.xls*

```
The DISCRIM Procedure
```

I am only reporting the overall correlations to keep the output concise – the correlations within each group are similar and not very high – if you feel the need to use the correlations to respond to any question in the assignment assume that they are low OR run the code on the website to confirm this yourself.

```
The DISCRIM Procedure

                   Pooled Within-Class Correlation Coefficients  /  Pr > |r|

  Variable          Cars   Education    SpouseEd       Years     Workers      Income      Ethnic

  dist          -0.03382    -0.05022    -0.05917     0.04690     0.06458     0.14403    -0.06730
  dist            0.6504      0.5008      0.4275      0.5295      0.3864      0.0524      0.3667

  Age            0.02816     0.01583    -0.06468     0.46772    -0.07814     0.15111    -0.01316
  Age             0.7059      0.8320      0.3857      <.0001      0.2944      0.0417      0.8601

  Gender        -0.06123    -0.17514    -0.06616    -0.11083     0.22325    -0.21160     0.08037
  Gender          0.4116      0.0180      0.3748      0.1363      0.0025      0.0041      0.2808

  Married        0.12386    -0.03019     0.05808    -0.00024    -0.22154    -0.12816    -0.03279
  Married         0.0957      0.6858      0.4361      0.9974      0.0027      0.0847      0.6604

  License        0.09788    -0.02170    -0.04038    -0.12285    -0.08033    -0.17653     0.16014
  License         0.1887      0.7712      0.5884      0.0985      0.2810      0.0171      0.0308

  Adults         0.02660     0.12036     0.01639     0.19785     0.31200     0.06398     0.12483
  Adults          0.7215      0.1056      0.8262      0.0074      <.0001      0.3908      0.0931

  Children      -0.02620     0.08256     0.14766    -0.18040    -0.07230     0.09220    -0.01895
  Children        0.7256      0.2679      0.0467      0.0148      0.3321      0.2158      0.7996

  Cars           1.00000     0.04817     0.08886    -0.01755     0.02814    -0.03502    -0.00625
  Cars                        0.5184      0.2329      0.8141      0.7062      0.6388      0.9333

  Education      0.04817     1.00000     0.46016    -0.12411    -0.02069     0.36969     0.05321
  Education       0.5184                  <.0001      0.0951      0.7816      <.0001      0.4756

  SpouseEd       0.08886     0.46016     1.00000    -0.09187    -0.01023     0.29211    -0.00870
  SpouseEd        0.2329      <.0001                  0.2174      0.8910      <.0001      0.9072

  Years         -0.01755    -0.12411    -0.09187     1.00000     0.02489     0.07811    -0.02576
  Years           0.8141      0.0951      0.2174                  0.7387      0.2946      0.7299

  Workers        0.02814    -0.02069    -0.01023     0.02489     1.00000     0.22929     0.02402
  Workers         0.7062      0.7816      0.8910      0.7387                  0.0018      0.7476

  Income        -0.03502     0.36969     0.29211     0.07811     0.22929     1.00000    -0.18987
```

```
 Income         0.6388       <.0001       <.0001       0.2946       0.0018                    0.0103

 Ethnic       -0.00625      0.05321     -0.00870     -0.02576      0.02402     -0.18987      1.00000
 Ethnic         0.9333       0.4756       0.9072       0.7299       0.7476       0.0103
```

```
The DISCRIM Procedure
Simple Statistics

                                        Total-Sample

                                                                                  Standard
      Variable     Label            N           Sum          Mean      Variance   Deviation

      dist         dist           183     724.00000       3.95628       2.14094      1.4632
      Age          Age            183     786.00000       4.29508       1.01135      1.0057
      Gender       Gender         183     217.00000       1.18579       0.15210      0.3900
      Married      Married        183     186.00000       1.01639       0.01621      0.1273
      License      License        183     185.00000       1.01093       0.01087      0.1043
      Adults       Adults         183     431.00000       2.35519       0.53798      0.7335
      Children     Children       183     153.00000       0.83607       1.00594      1.0030
      Cars         Cars           183     310.52498       1.69686       1.31872      1.1484
      Education    Education      183     796.00000       4.34973       2.31658      1.5220
      SpouseEd     SpouseEd       183     664.00000       3.62842       2.45457      1.5667
      Years        Years          183     732.00000       4.00000       1.61538      1.2710
      Workers      Workers        183     288.00000       1.57377       0.53162      0.7291
      Income       Income         183          1122       6.13115       3.38930      1.8410
      Ethnic       Ethnic         183     196.00000       1.07104       0.07734      0.2781
---------------------------------------------------------------------------------------------
```

```
                                   The DISCRIM Procedure
                                     Simple Statistics

                                        CLUSTER = 1

                                                                                  Standard
      Variable     Label            N           Sum          Mean      Variance   Deviation

      dist         dist           119     479.00000       4.02521       2.10953      1.4524
      Age          Age            119     514.00000       4.31933       0.96496      0.9823
      Gender       Gender         119     141.00000       1.18487       0.15197      0.3898
      Married      Married        119     121.00000       1.01681       0.01666      0.1291
      License      License        119     120.00000       1.00840       0.00840      0.0917
      Adults       Adults         119     284.00000       2.38655       0.56117      0.7491
      Children     Children       119      98.00000       0.82353       0.96012      0.9799
      Cars         Cars           119     247.06035       2.07614       1.45433      1.2060
      Education    Education      119     528.00000       4.43697       2.36676      1.5384
      SpouseEd     SpouseEd       119     440.00000       3.69748       2.36533      1.5380
      Years        Years          119     482.00000       4.05042       1.62455      1.2746
      Workers      Workers        119     193.00000       1.62185       0.61003      0.7810
      Income       Income         119     751.00000       6.31092       2.99573      1.7308
      Ethnic       Ethnic         119     125.00000       1.05042       0.04828      0.2197
```

```
                                The DISCRIM Procedure
                                  Simple Statistics

                                     CLUSTER = 2

                                                                                 Standard
Variable     Label            N           Sum          Mean      Variance       Deviation

dist         dist            64     245.00000       3.82813       2.20809          1.4860
Age          Age             64     272.00000       4.25000       1.11111          1.0541
Gender       Gender          64      76.00000       1.18750       0.15476          0.3934
Married      Married         64      65.00000       1.01563       0.01563          0.1250
License      License         64      65.00000       1.01563       0.01562          0.1250
Adults       Adults          64     147.00000       2.29688       0.49777          0.7055
Children     Children        64      55.00000       0.85938       1.10689          1.0521
Cars         Cars            64      63.46464       0.99163       0.30870          0.5556
Education    Education       64     268.00000       4.18750       2.21825          1.4894
SpouseEd     SpouseEd        64     224.00000       3.50000       2.63492          1.6232
Years        Years           64     250.00000       3.90625       1.61012          1.2689
Workers      Workers         64      95.00000       1.48438       0.38070          0.6170
Income       Income          64     371.00000       5.79688       4.00570          2.0014
Ethnic       Ethnic          64      71.00000       1.10938       0.13070          0.3615
```

```
The DISCRIM Procedure

                                   Univariate Test Statistics

                             F Statistics,    Num DF=1,   Den DF=181

                                  Total     Pooled    Between
                               Standard   Standard   Standard             R-Square
   Variable   Label           Deviation  Deviation  Deviation  R-Square  / (1-RSq)  F Value  Pr > F

   dist       dist               1.4632     1.4642     0.1329    0.0041     0.0042     0.75  0.3864
   Age        Age                1.0057     1.0079     0.0468    0.0011     0.0011     0.20  0.6578
   Gender     Gender             0.3900     0.3911   0.001771    0.0000     0.0000     0.00  0.9655
   Married    Married            0.1273     0.1277   0.000797    0.0000     0.0000     0.00  0.9525
   License    License            0.1043     0.1045   0.004870    0.0011     0.0011     0.20  0.6562
   Adults     Adults             0.7335     0.7342     0.0605    0.0034     0.0034     0.62  0.4318
   Children   Children           1.0030     1.0056     0.0242    0.0003     0.0003     0.05  0.8184
   Cars       Cars               1.1484     1.0274     0.7314    0.2039     0.2562    46.37  <.0001
   Education  Education          1.5220     1.5215     0.1682    0.0061     0.0062     1.12  0.2916
   SpouseEd   SpouseEd           1.5667     1.5682     0.1332    0.0036     0.0036     0.66  0.4176
   Years      Years              1.2710     1.2726     0.0972    0.0029     0.0030     0.53  0.4658
   Workers    Workers            0.7291     0.7282     0.0927    0.0081     0.0082     1.48  0.2248
   Income     Income             1.8410     1.8296     0.3467    0.0178     0.0182     3.29  0.0716
   Ethnic     Ethnic             0.2781     0.2774     0.0398    0.0103     0.0104     1.88  0.1721


                                        Average R-Square

                                Unweighted               0.0187836
                                Weighted by Variance       0.02259


                         Multivariate Statistics and Exact F Statistics

                                       S=1    M=6    N=83

         Statistic                         Value    F Value    Num DF    Den DF    Pr > F

         Wilks' Lambda                0.76656082       3.65        14       168    <.0001
         Pillai's Trace               0.23343918       3.65        14       168    <.0001
         Hotelling-Lawley Trace       0.30452793       3.65        14       168    <.0001
         Roy's Greatest Root          0.30452793       3.65        14
```

```
Pooled Within Canonical Structure

                            Variable       Label                 Can1

                            dist           dist              0.116962
                            Age            Age               0.059770
                            Gender         Gender           -0.005835
                            Married        Married           0.008042
                            License        License          -0.060058
                            Adults         Adults            0.106131
                            Children       Children         -0.030974
                            Cars           Cars              0.917216
                            Education      Education         0.142472
                            SpouseEd       SpouseEd          0.109424
                            Years          Years             0.098439
                            Workers        Workers           0.164052
                            Income         Income            0.244143
                            Ethnic         Ethnic           -0.184646



                                    Raw Canonical Coefficients

                            Variable       Label                 Can1

                            dist           dist           0.093648408
                            Age            Age            0.003238556
                            Gender         Gender         0.458645532
                            Married        Married       -0.415270188
                            License        License       -1.176406055
                            Adults         Adults         0.079397249
                            Children       Children       0.003462073
                            Cars           Cars           0.933057070
                            Education      Education      0.057273684
                            SpouseEd       SpouseEd      -0.034822458
                            Years          Years          0.068154708
                            Workers        Workers        0.003074142
                            Income         Income         0.116757556
                            Ethnic         Ethnic        -0.488172569


                                Class Means on Canonical Variables

                                    CLUSTER             Can1

                                      1            0.4024795433
                                      2             -.7483604009
                                      The DISCRIM Procedure

                                   Linear Discriminant Function

                                  _     -1 _                              -1 _
                   Constant = -.5 X' COV   X      Coefficient Vector = COV   X
                                   j        j                                 j


                             Linear Discriminant Function for CLUSTER

                        Variable     Label                  1             2

                        Constant                   -110.83072    -109.09252
                        dist         dist             1.50264       1.39487
                        Age          Age              7.04979       7.04606
                        Gender       Gender          11.55074      11.02291
                        Married      Married         49.54991      50.02782
                        License      License         70.06406      71.41791
                        Adults       Adults          -1.14490      -1.23628
                        Children     Children         4.82479       4.82080
                        Cars         Cars             0.72052      -0.35328
                        Education    Education        1.55045       1.48454
                        SpouseEd     SpouseEd         0.38405       0.42412
                        Years        Years            1.80667       1.72824
                        Workers      Workers          4.60073       4.59719
                        Income       Income           1.85893       1.72456
                        Ethnic       Ethnic          12.27488      12.83669

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.TMP
Resubstitution Summary using Linear Discriminant Function


Number of Observations and Percent Classified into CLUSTER

                     From CLUSTER            1            2        Total

                                1           81           38          119
                                         68.07        31.93       100.00

                                2           14           50           64
                                         21.88        78.13       100.00

                            Total           95           88          183
                                         51.91        48.09       100.00

                           Priors          0.5          0.5
```

```
The DISCRIM Procedure
Classification Results for Calibration Data: WORK.TMP
Resubstitution Results using Linear Discriminant Function


Posterior Probability of Membership in Each CLUSTER

                                             2                    2
                          Pr(j|X) = exp(-.5 D (X)) / SUM exp(-.5 D (X))
                                             j        k           k


        Number of Observations and Average Posterior Probabilities Classified into CLUSTER

                           From CLUSTER            1            2

                                      1           81           38
                                              0.7544       0.6758

                                      2           14           50
                                              0.6101       0.7152

                                  Total           95           88
                                              0.7332       0.6982

                                 Priors          0.5          0.5

  Posterior Probability of Membership in CLUSTER

                                           Classified
                                   From          into
                         Obs    CLUSTER       CLUSTER         1         2

                           2          2           1 *    0.6593    0.3407
                           3          2           2      0.3328    0.6672
                           5          2           2      0.2114    0.7886
                           6          2           2      0.2534    0.7466
                          10          2           1 *    0.5619    0.4381
                          12          2           2      0.2831    0.7169
                          13          2           2      0.3283    0.6717
                          14          2           2      0.4372    0.5628
                          15          2           2      0.2851    0.7149
                          17          2           2      0.3424    0.6576
                          20          2           2      0.2688    0.7312
                          21          2           1 *    0.6947    0.3053
                          23          2           1 *    0.6315    0.3685
                          25          2           2      0.3573    0.6427
                          27          2           2      0.2068    0.7932
                          28          2           1 *    0.6761    0.3239
                          30          2           1 *    0.5273    0.4727
                          32          2           2      0.1920    0.8080
                          34          2           2      0.4806    0.5194
                          35          1           2 *    0.2292    0.7708
* Misclassified observation
```

I have only printed out 20 of 300 consumers in the calibration.xls to illustrate how these consumers would be classified given the estimates of the Discriminant Analysis. This is the same sample that is used to estimate the discriminant function.

Classification of consumers in prospect.xls

```
The DISCRIM Procedure
Classification Summary for Test Data: WORK.PROS
Classification Summary using Linear Discriminant Function


        Posterior Probability of Membership in Each CLUSTER

                                             2                    2
                          Pr(j|X) = exp(-.5 D (X)) / SUM exp(-.5 D (X))
                                             j        k           k


                    Number of Observations and Percent Classified into CLUSTER

                     From CLUSTER            1            2        Total

                                1           73           46          119
                                         61.34        38.66       100.00

                                2           19           45           64
                                         29.69        70.31       100.00

                            Total           92           91          183
                                         50.27        49.73       100.00

                           Priors          0.5          0.5
```

```
The DISCRIM Procedure
Classification Results for Test Data: WORK.PROS
Classification Results using Linear Discriminant Function


                              Generalized Squared Distance Function

                                2         _          -1     _
                               D (X) = (X-X    )' COV    (X-X    )
                                j          (X)j      (X)     (X)j

                       Posterior Probability of Membership in Each CLUSTER

                                             2                    2
                          Pr(j|X) = exp(-.5 D (X)) / SUM exp(-.5 D (X))
                                             j        k           k


        Number of Observations and Average Posterior Probabilities Classified into CLUSTER

                           From CLUSTER            1            2
```

```
     1           73           46
             0.7632       0.6935

     2           19           45
             0.6564       0.6900

 Total           92           91
             0.7411       0.6918

Priors          0.5          0.5
```

(c) Validation: Output from comparing the classification of prospective consumers in *prospect.xls* with their actual behavior tracked in *validation.xls*

```
                  The FREQ Procedure

               Table of used by _INTO_

         used(used)     _INTO_(Cluster)

         Frequency|
         Percent  |
         Row Pct  |
         Col Pct  |       1|       2|  Total
         ─────────┼────────┼────────┼
                1 |     33 |     20 |     53
                  |  22.60 |  13.70 |  36.30
                  |  62.26 |  37.74 |
                  |  50.00 |  25.00 |
         ─────────┼────────┼────────┼
                2 |     33 |     60 |     93
                  |  22.60 |  41.10 |  63.70
                  |  35.48 |  64.52 |
                  |  50.00 |  75.00 |
         ─────────┼────────┼────────┼
         Total          66       80      146
                     45.21    54.79   100.00
```

As discussed in class the first column labeled *USED* denotes whether or not the consumer actually used Mass Transit: *INTO* denotes the classification performed by Discriminant Analysis.