## CASE HANDOUT:

# Conglomerate Inc.'s New PDA: A Segmentation Study


## By

## Prof. P.V. (Sundar) Balakrishnan


## Cluster Analysis

This note has been prepared in a manner that it should be possible for you to replicate the results shown below by looking at the screen shots shown in the appropriate places.

According to the information in this case, involving exploratory research, 160 people were surveyed using two questionnaires:

- a "needs questionnaire" (Segmenation data - Basis Variable) and
- a "demographic questionnaire" (Classification variables – Discrimination data).

The Case and the data can be found from the location where you loaded the software:

Go to the folder containing the case and data:
**~\Cases and Exercises\ConneCtor PDA 2001 (Segmentation)**

- The Case is in a PDF file called:
  - ConneCtor PDA 2001 Case (Segmentation).pdf
- The Data is in an Excel Spreadsheet, called:
  - ConneCtor PDA 2001 Data (Segmentation).xls

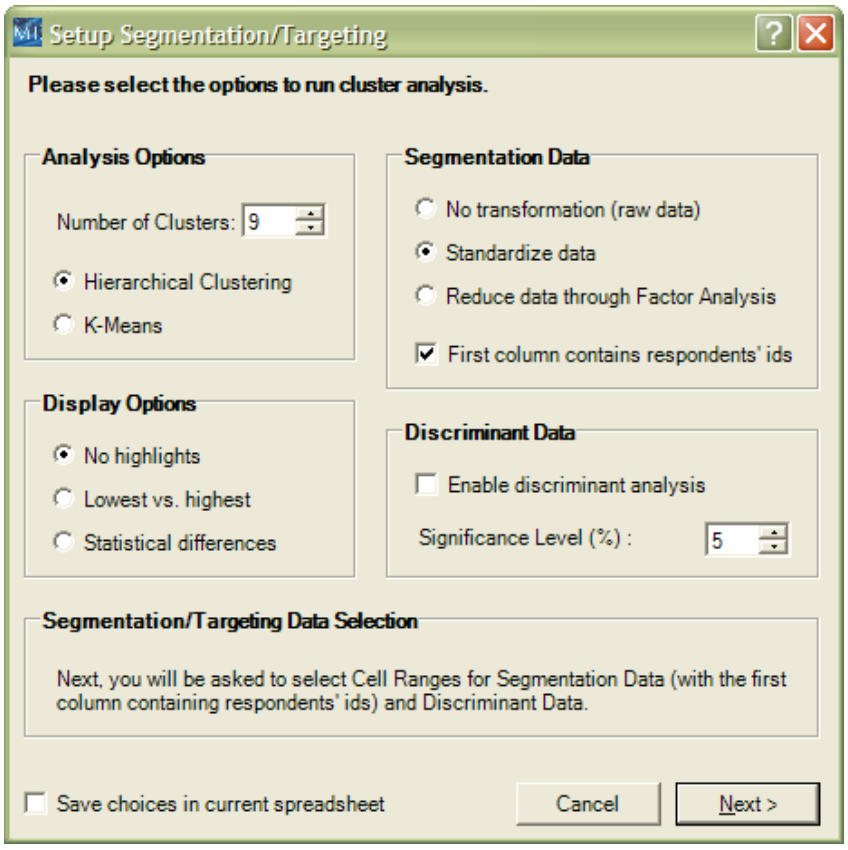
After you have read the case, launch the Excel spreadsheet containing the data.

## *Question 1*

Run only cluster analysis (without Discrimination) on the data to try to identify the number of distinct segments present in this market. Consider both the distances separating the segments and the characteristics of the resulting segments. (Note the need to standardize the data!)

*Solution:*

Within the Excel Toolbar, select ME->XL ➔ then select
SEGMENTATION AND CLASSIFICATION ➔ RUN SEGMENTATION
This should bring up a dialog box as follows. Make the appropriate selections.

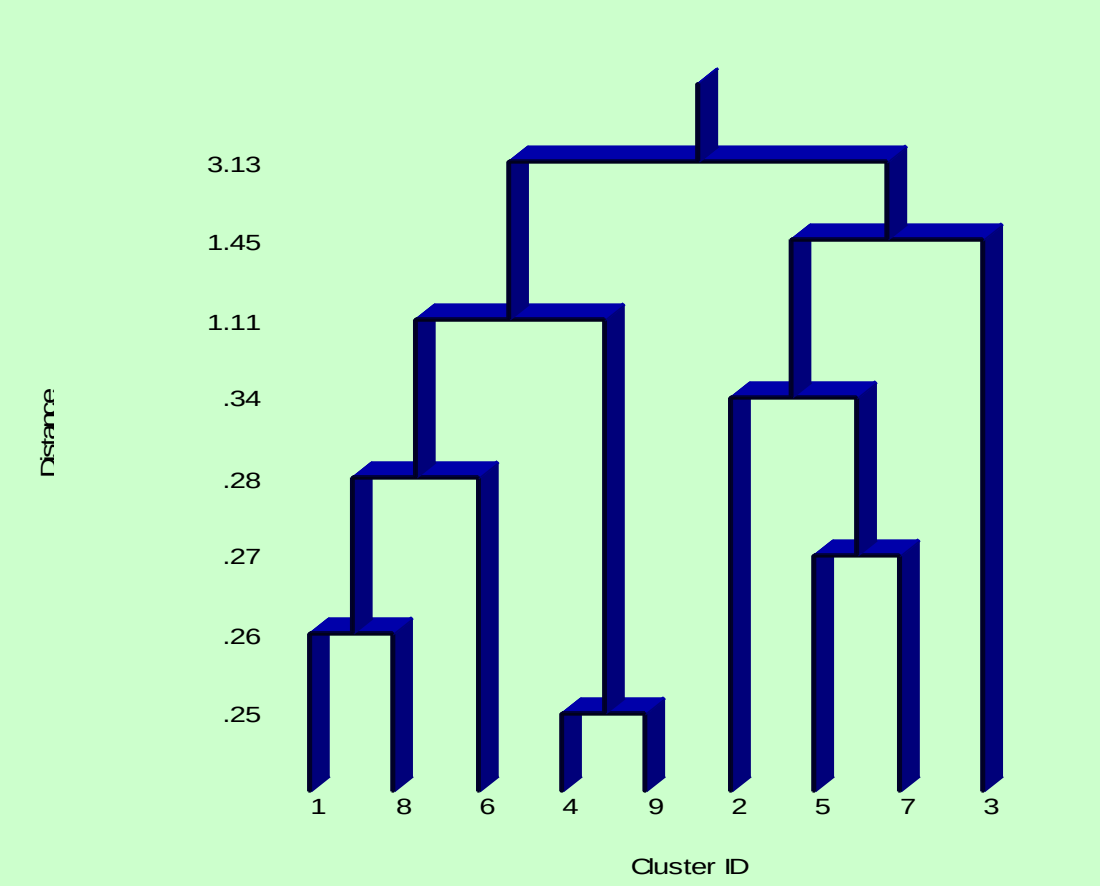



Among the major decisions that you will hav to make are whether or not to **standardize the data**.

By running the default *Hierarchical Clustering (*Ward's method) for the default nine clusters (i.e, segments) we get the dendogram as shown next:

## Dendogram:

Dendograms provide graphical representations of the loss of information generated by grouping different clusters (or customers) together.




Look the vertical axis -- distance measure (Error Sum of Squares) -- on the dendogram. This shows the following clusters are quite close together and can be combined with a small loss in consumer grouping information:
A) clusters 4 and 9 at 0.25, ii) clusters 1 and 8 at 0.26, ii) cluster 7-5 at 0.27.
B) fused clusters 1,8 and 6 at 0.28, ii) fused cluster 7-5 and cluster 2 (0.34).
However, note that when going from a four-cluster solution to a three-cluster solution, the distance to be bridged is much larger (1.11); thus, the four-cluster solution is indicated by the ESS...

Diagnostics for the cluster analysis
---

The following table lists the cluster number to which a row element will belong for varying specifications of the number of required clusters.

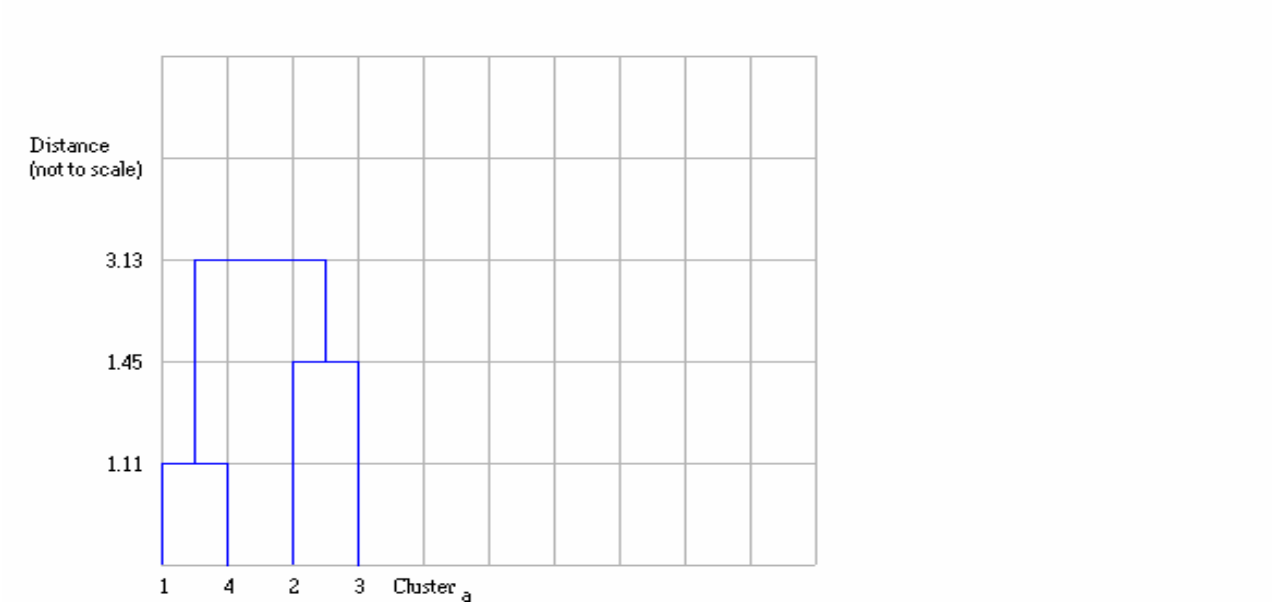
## *Cluster Members*

**The following table lists the cluster number to which each observation belongs for varying cluster solutions. For example, the column "for 2 clusters" gives the cluster number of each observation in a 2-cluster solution. The cluster solution you have selected is in bold with a yellow background.**

| **Observation / Cluster solution** | With 2 clusters | With 3 clusters | With 4 clusters | With 5 clusters | With 6 clusters | With 7 clusters | With 8 clusters | With 9 clusters |
|---|---|---|---|---|---|---|---|---|
| 1 | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** |
| 2 | **1** | **1** | **1** | **1** | **1** | **1** | **8** | **8** |
| 3 | **1** | **1** | **1** | **1** | **1** | **1** | **8** | **8** |
| 4 | **1** | **1** | **1** | **1** | **1** | **1** | **8** | **8** |
| 5 | **1** | **1** | **1** | **1** | **1** | **1** | **8** | **8** |
| 6 | **1** | **1** | **1** | **1** | **1** | **1** | **8** | **8** |
| 7 | **1** | **1** | **1** | **1** | **6** | **6** | **6** | **6** |
| 8 | **1** | **1** | **1** | **1** | **1** | **1** | **8** | **8** |
| 9 | **1** | **1** | **1** | **1** | **1** | **1** | **8** | **8** |

| ROW NOS | CL2 | CL3 | CL4 | CL5 | CL6 | CL7 | CL8 | CL9 |
|---|---|---|---|---|---|---|---|---|
| | | | <SNIP> | | | | | |
| 105 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 4 |
| 106 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 9 |
| 107 | 1 | 1 | 1 | 1 | 1 | 1 | 8 | 8 |
| 108 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 4 |
| 109 | 1 | 1 | 1 | 1 | 1 | 1 | 8 | 8 |
| 110 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 9 |
| 111 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 9 |
| 112 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 9 |
| 113 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 4 |
| 114 | 1 | 1 | 1 | 1 | 1 | 1 | 8 | 8 |
| 115 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 9 |
| | | | <SNIP> | | | | | |
| 158 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 159 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 160 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |

If we run the analysis again, now set to four segments, the program will perform the agglomeration for us. The first paragraph below tells us about cluster membership.




NOTE:. Another Cluster Analysis option is the **K-Means** procedure that attempts to identify relatively homogeneous groups of cases based on selected characteristics, using an algorithm that can handle large numbers of cases better than hierarchical methods.

If a K-Means procedure is chosen, its four-cluster solution is very similar to the four-cluster solution based on the Ward's hierarchical clustering procedure (See Appendix I & II at end of document for the K-Means results).

### *Question 2*

Identify and profile (name) the clusters that you select. Given the attributes of ConneCtor, which cluster would you target for your marketing campaign?

*Solution*

## *Cluster Sizes*

**The following table lists the size of the population and of each segment, in both absolute and relative terms.**

| Size / Cluster | Overall | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|---|
| Number of observations | 160 | 56 | 51 | 16 | 37 |
| Proportion | 1 | 0.35 | 0.319 | 0.1 | 0.231 |

From the Results, we get the mean for each variable in each cluster:

## *Segmentation Variables*

**Means of each segmentation variable for each segment.**

| Segmentation variable / Cluster | Overall | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|---|
| Innovator | 3.44 | 3.62 | 2.43 | **2.19** | **5.11** |
| Use Message | 5.2 | **6.52** | 5.63 | **3.19** | 3.49 |
| Use Cell | 5.62 | **6.02** | 5.43 | **4.31** | 5.84 |
| Use PIM | 3.99 | **5.84** | **2.33** | 3.06 | 3.86 |
| Inf Passive | 4.45 | 5.02 | 3.88 | **6.12** | **3.65** |
| Inf Active | 4.47 | 5.11 | 3.9 | **6.25** | **3.51** |
| Remote Access | 4 | 3.89 | 5.04 | **5.31** | **2.16** |
| Share Inf | 3.75 | 3.5 | 3.73 | **6.12** | **3.14** |
| Monitor | 4.79 | **4.29** | **5.55** | 5 | 4.43 |
| Email | 4.73 | **5.98** | 3.31 | **2.88** | 5.59 |
| Web | 4.46 | 5.62 | 3.04 | **1.44** | **5.97** |
| Mmedia | 3.98 | 5.11 | 2.45 | **1.94** | **5.27** |
| Ergonomic | 4.63 | **3.95** | 4.16 | 5.5 | **5.95** |
| Monthly | 28.7 | **24.5** | 25.3 | **45.3** | 32.6 |
| Price | 331 | 285 | **273** | **488** | 411 |

To characterize clusters (obtained with the default Ward method) we look for means that are either well above or well below the Overall mean.

*To see the results using the **K-Means** option, please turn to Appendix I.*

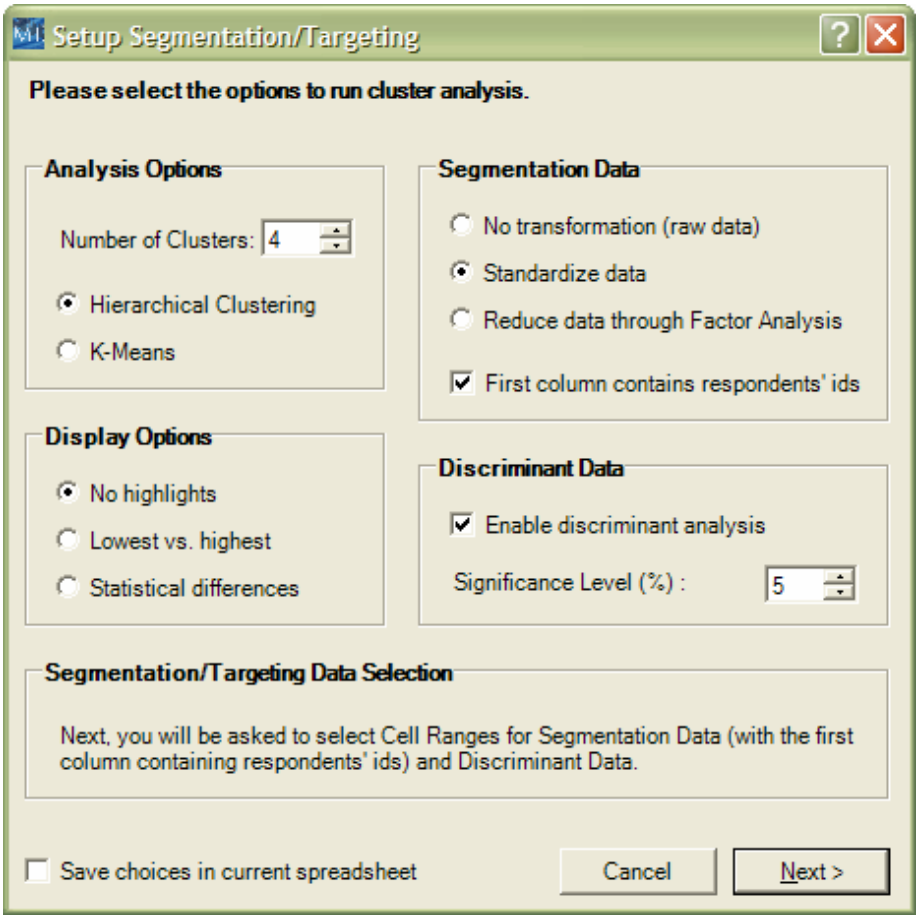
It is difficult to decide which segment(s) to **target** based on the above information. We need some means of discrimination of these segments. To this end, we employ the classification data collected from these respondents.

### *Question 3*

Go back to MENU, check Enable Discrimination and rerun the analysis. How would you go about targeting the segment(s) you picked in question 2?

*Solution*

By checking the Discrimination option in the Setup, for the 4 cluster solution, we get:



## *Confusion Matrix*

**Comparison of cluster membership predictions based on discriminant data, and actual cluster memberships. High values in the diagonal of the confusion matrix (in bold) indicates that discriminant data is good at predicting cluster membership.**

| Actual / Predicted cluster | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Cluster 1 | **34** | 11 | 4 | 7 |
| Cluster 2 | 10 | **35** | 5 | 1 |
| Cluster 3 | 1 | 3 | **12** | 0 |
| Cluster 4 | 3 | 0 | 0 | **34** |

| Actual / Predicted cluster | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Cluster 1 | **60.70%** | 19.60% | 07.10% | 12.50% |
| Cluster 2 | 19.60% | **68.60%** | 09.80% | 02.00% |
| Cluster 3 | 06.20% | 18.80% | **75.00%** | 00.00% |
| Cluster 4 | 08.10% | 00.00% | 00.00% | **91.90%** |

**Hit Rate (percent of total cases correctly classified) 71.88%**

**Other Diagnostics for Discriminant Analysis**

## *Discriminant Function*

**Correlation of variables with each significant discriminant function (significance level < 0.05).**

| Discriminant variable / Function | Function 1 | Function 2 | Function 3 |
|---|---|---|---|
| Away | -0.705 | -0.132 | 0.116 |
| Education | 0.704 | 0.098 | -0.035 |
| PDA | 0.669 | 0.219 | 0.114 |
| Income | 0.629 | 0.138 | -0.266 |
| Business Week | 0.405 | -0.062 | 0.055 |
| Mgourmet | 0.276 | 0.15 | -0.164 |
| PC | 0.28 | -0.549 | -0.073 |
| Construction | -0.197 | 0.37 | 0.036 |
| Emergency | -0.161 | 0.363 | 0.027 |
| Cell | 0.156 | -0.348 | -0.011 |
| Computers | 0.211 | 0.297 | -0.061 |
| Sales | -0.014 | -0.386 | 0.652 |
| Service | -0.308 | -0.308 | -0.468 |
| Age | -0.002 | 0.069 | 0.409 |
| Field & Stream | -0.347 | 0.103 | -0.379 |
| PC Magazine | 0.048 | 0.075 | -0.354 |
| Professional | 0.327 | 0.014 | -0.338 |
| **Variance explained** | **50.48** | **31.68** | **17.84** |
| **Cumulative variance explained** | **50.48** | **82.16** | **100** |

Next, we examine as to how the Classification (such as demographics) Variables describe the different segments so as to better target them.

## Means for Discrimination Variables in each Cluster

Results Using Ward's (default) Method

### *Discriminant Variables*

**Means of each discriminant variable for each segment.**

| Discriminant variable / Cluster | Overall | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|---|
| Away | 4.21 | 4.36 | 4.84 | **5.38** | **2.60** |
| Education | 2.51 | 2.48 | 2.20 | **1.94** | **3.22** |
| PDA | 0.44 | 0.45 | **0.18** | 0.19 | **0.89** |
| Income | 66.89 | 62.59 | 60.53 | **52.44** | **88.43** |
| Business Week | 0.28 | 0.30 | 0.18 | **0.00** | **0.49** |
| Mgourmet | 0.02 | **0.00** | **0.00** | **0.00** | **0.08** |
| PC | 0.98 | **1.00** | **1.00** | **0.81** | **1.00** |
| Construction | 0.08 | **0.05** | 0.06 | **0.31** | **0.05** |
| Emergency | 0.04 | **0.02** | 0.02 | **0.19** | 0.03 |
| Cell | 0.88 | **0.91** | 0.90 | **0.63** | 0.89 |
| Computers | 0.23 | 0.18 | **0.14** | 0.31 | **0.41** |
| Sales | 0.30 | **0.54** | 0.24 | **0.06** | 0.14 |
| Service | 0.18 | 0.11 | **0.39** | 0.13 | **0.03** |
| Age | 40.01 | **43.07** | **36.77** | 42.19 | 38.89 |
| Field & Stream | 0.13 | 0.04 | 0.24 | **0.31** | **0.03** |
| PC Magazine | 0.24 | **0.14** | 0.29 | 0.25 | **0.32** |
| Professional | 0.16 | 0.09 | 0.16 | **0.00** | **0.35** |

To help us characterize clusters we look for variable means that are either well above or well below the Overall mean.

### *Question 4*

How has this analysis helped you to segment the market for ConneCtor?

### *Question 5*

What other data and analysis would you do to develop a targeted marketing program for ConneCtor?

# APPENDIX I :

## Diagnostics for the K-means Cluster Analysis

### *Cluster Sizes*

**The following table lists the size of the population and of each segment, in both absolute and relative terms.**

| Size / Cluster | Overall | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|---|
| Number of observations | 160 | 58 | 48 | 16 | 38 |
| Proportion | 1 | 0.363 | 0.3 | 0.1 | 0.237 |

Means for each basis variable in each cluster:

### *Segmentation Variables*

**Means of each segmentation variable for each segment.**

| Segmentation variable / Cluster | Overall | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|---|
| Innovator | 3.44 | 3.55 | 2.4 | 2.19 | 5.13 |
| Use Message | 5.2 | 6.52 | 5.58 | 3.19 | 3.55 |
| Use Cell | 5.62 | 5.98 | 5.42 | 4.31 | 5.87 |
| Use PIM | 3.99 | 5.78 | 2.21 | 3.06 | 3.89 |
| Inf Passive | 4.45 | 5.02 | 3.85 | 6.12 | 3.63 |
| Inf Active | 4.47 | 5.14 | 3.81 | 6.25 | 3.53 |
| Remote Access | 4 | 3.78 | 5.21 | 5.31 | 2.26 |
| Share Inf | 3.75 | 3.5 | 3.71 | 6.12 | 3.18 |
| Monitor | 4.79 | 4.29 | 5.62 | 5 | 4.42 |
| Email | 4.73 | 6 | 3.17 | 2.88 | 5.55 |
| Web | 4.46 | 5.47 | 3.04 | 1.44 | 6 |
| Mmedia | 3.98 | 5.1 | 2.31 | 1.94 | 5.24 |
| Ergonomic | 4.63 | 4.03 | 4.06 | 5.5 | 5.89 |
| Monthly | 28.7 | 24.5 | 25.1 | 45.3 | 32.6 |
| Price | 331 | 285 | 269 | 488 | 411 |

## *Cluster Members*

**The following table lists the probabilities of each observations belonging to one of the 4 clusters. The probabilities are based on the inverse of the distance between an observation and each cluster centroid. The last column lists the cluster with the highest probability (discrete cluster membership).**

| **Observation / Cluster solution** | Probabilility to belong to cluster 1 | Probabilility to belong to cluster 2 | Probabilility to belong to cluster 3 | Probabilility to belong to cluster 4 | Cluster membership |
|---|---|---|---|---|---|
| 1 | **0.511** | 0.148 | 0.095 | 0.247 | 1 |
| 2 | **0.457** | 0.216 | 0.127 | 0.200 | 1 |
| 3 | **0.316** | 0.283 | 0.188 | 0.212 | 1 |
| 4 | **0.482** | 0.145 | 0.126 | 0.247 | 1 |
| 5 | **0.469** | 0.138 | 0.152 | 0.242 | 1 |
| 6 | **0.434** | 0.160 | 0.143 | 0.264 | 1 |
| 7 | **0.578** | 0.117 | 0.064 | 0.241 | 1 |
| 8 | **0.499** | 0.184 | 0.118 | 0.199 | 1 |
| 9 | **0.448** | 0.174 | 0.110 | 0.269 | 1 |
| 10 | **0.446** | 0.239 | 0.112 | 0.203 | 1 |

<SNIP>

| | | | | | |
|---|---|---|---|---|---|
| 150 | 0.142 | 0.230 | **0.495** | 0.133 | 3 |
| 151 | 0.147 | 0.153 | **0.566** | 0.135 | 3 |
| 152 | 0.114 | 0.153 | **0.629** | 0.105 | 3 |
| 153 | 0.083 | 0.109 | **0.734** | 0.073 | 3 |
| 154 | 0.136 | 0.193 | **0.528** | 0.142 | 3 |
| 155 | 0.148 | 0.205 | **0.509** | 0.139 | 3 |
| 156 | 0.116 | 0.150 | **0.613** | 0.122 | 3 |
| 157 | 0.159 | 0.187 | **0.494** | 0.160 | 3 |
| 158 | 0.113 | 0.164 | **0.609** | 0.114 | 3 |
| 159 | 0.078 | 0.107 | **0.752** | 0.063 | 3 |
| 160 | 0.060 | 0.087 | **0.805** | 0.048 | 3 |

## APPENDIX II :

Discrimination Means of the Classification Variables

Results Using K-Means Method

### *Discriminant Variables*

**Means of each discriminant variable for each segment.**

| Discriminant variable / Cluster | Overall | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|---|
| **Away** | 4.21 | 4.31 | 4.98 | **5.38** | **2.58** |
| **Education** | 2.51 | 2.50 | 2.15 | **1.94** | **3.21** |
| **PDA** | 0.44 | 0.45 | **0.15** | 0.19 | **0.90** |
| **Income** | 66.89 | 62.81 | 59.65 | **52.44** | **88.37** |
| **Business Week** | 0.28 | 0.29 | 0.17 | **0.00** | **0.50** |
| **Field & Stream** | 0.13 | 0.03 | 0.25 | **0.31** | **0.03** |
| **Mgourmet** | 0.02 | **0.00** | **0.00** | **0.00** | **0.08** |
| **PC** | 0.98 | **1.00** | **1.00** | **0.81** | **1.00** |
| **Construction** | 0.08 | **0.05** | 0.06 | **0.31** | 0.05 |
| **Emergency** | 0.04 | **0.02** | 0.02 | **0.19** | 0.03 |
| **Cell** | 0.88 | 0.90 | **0.92** | **0.63** | 0.90 |
| **Computers** | 0.23 | 0.19 | **0.13** | 0.31 | **0.40** |
| **Sales** | 0.30 | **0.55** | 0.21 | **0.06** | 0.13 |
| **Age** | 40.01 | **43.72** | **36.08** | 42.19 | 38.37 |
| **Professional** | 0.16 | 0.07 | 0.17 | **0.00** | **0.37** |
| **Service** | 0.18 | 0.10 | **0.42** | 0.13 | **0.03** |
| **PC Magazine** | 0.24 | **0.16** | 0.29 | 0.25 | **0.32** |

## *Confusion Matrix*

**Comparison of cluster membership predictions based on discriminant data, and actual cluster memberships. High values in the diagonal of the confusion matrix (in bold) indicates that discriminant data is good at predicting cluster membership.**

| Actual / Predicted cluster | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Cluster 1 | **37** | 10 | 4 | 7 |
| Cluster 2 | 10 | **33** | 5 | 0 |
| Cluster 3 | 2 | 2 | **12** | 0 |
| Cluster 4 | 3 | 0 | 0 | **35** |

| Actual / Predicted cluster | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Cluster 1 | **63.80%** | 17.20% | 06.90% | 12.10% |
| Cluster 2 | 20.80% | **68.80%** | 10.40% | 00.00% |
| Cluster 3 | 12.50% | 12.50% | **75.00%** | 00.00% |
| Cluster 4 | 07.90% | 00.00% | 00.00% | **92.10%** |

**Hit Rate (percent of total cases correctly classified) 73.12%**