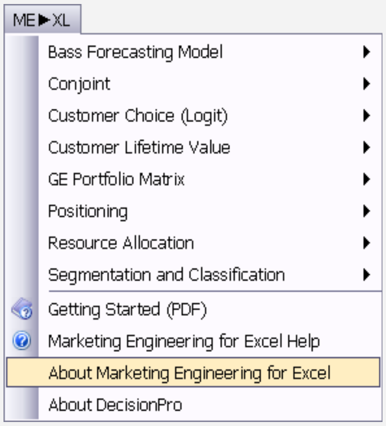
Tutorial

Segmentation and Classification

Marketing Engineering for Excel is a *Microsoft Excel* add-in. The software runs from within Microsoft Excel and only with data contained in an Excel spreadsheet.

After installing the software, simply open *Microsoft Excel*. A new menu appears, called "*ME ▸XL.*" This tutorial refers to the "*ME ▸XL/Segmentation and Classification*" submenu.



Overview

Segmentation and classification is an analytic technique that helps firms compare and group customers who share common characteristics (i.e., segmentation variables) into homogeneous segments and identify those particular customers in a market on the basis of external variables (i.e., discriminant variables).

Segmentation refers to the process of classifying customers into homogenous groups (segments), such that each group of customers shares enough characteristics in common to make it viable for the firm to design specific offerings or products for it. This application identifies customer segments using needs-based variables called basis variables. Cluster analysis helps firms:

- ✓ Better understand their customers.
- ✓ Identify different segments in a market.
- ✓ Choose attractive customer segments for classification with its marketing programs.

Getting Started

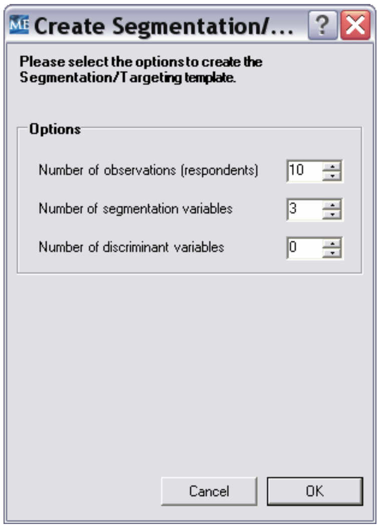
To apply segmentation and classification analysis, you can use your own data directly or a template preformatted by the ME►XL software.

The next section explains how to create an easy-to-use template to enter your own data.

If you want to run a segmentation and classification analysis immediately, open the example file “*OfficeStar (Segmentation).xls*” and jump to “Step 3: Running analysis” (p. 4). By default, the example files install in “*My Documents/My Marketing Engineering/.*”

Step 1 Creating a template

In Excel, if you click on ME►XL → Segmentation and Classification → Create Template, a dialog box appears. This box represents the first step in creating a template to run the segmentation and classification analysis software.



Create Segmentation/...

Please select the options to create the Segmentation/Targeting template.

Options

Number of observations (respondents) 10

Number of segmentation variables 3

Number of discriminant variables 0

The dialog box requests three pieces of information to design the template:

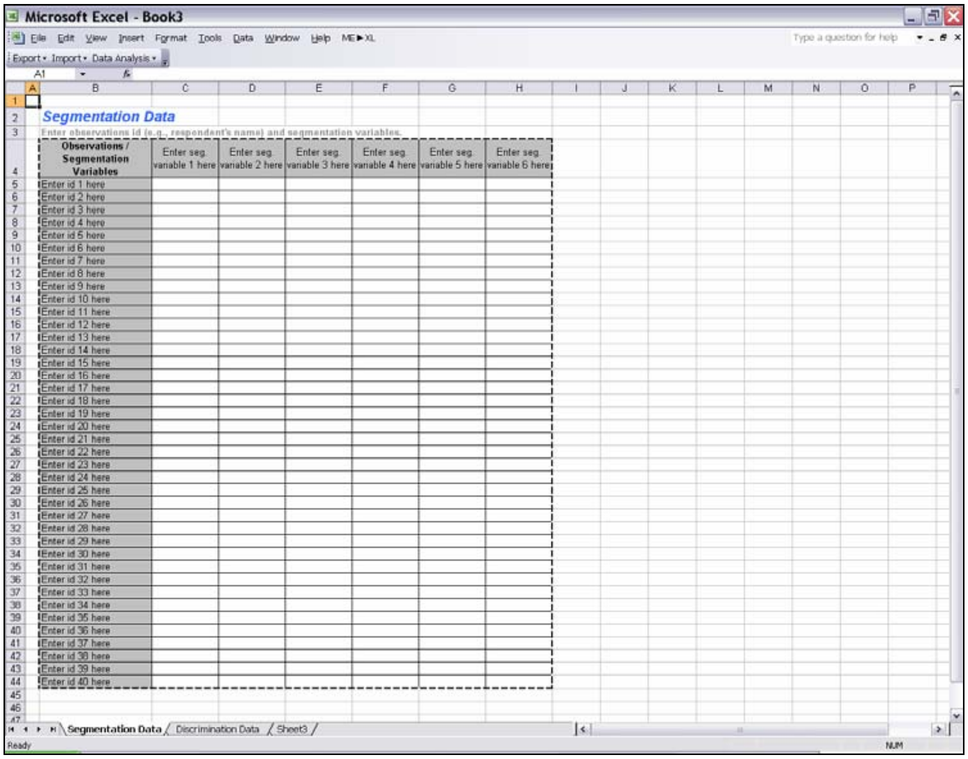
- **Observations** (respondents) indicate the number of customers or respondents in the data that need to be clustered.
- **Segmentation variables** help us assess the similarity between two respondents. These variables serve as the basis for segmentation and are often called **basis variables**. They might include customer's needs, wants, expectations, or preferences.
- **Discriminant variables**, also called **descriptors**, are optional variables that can describe the segments formed on the basis of the segmentation variables. These include demographic variables, such as educational level, gender, income, media consumption, and the like.

F.Y.I.

It is not always clear whether a specific variable should be treated as segmentation variable or discriminant variable. This choice might depend on the context, the managerial question, or the product category.

When in doubt, ask yourself the following questions: (1) Would this piece of information tell me what that customer wants, in which case it should be treated as segmentation variable, or (2) does this piece of information tell me who that customer is and therefore should be treated as discriminant variable? For example, "gender" would fall in the second category most of the time, whereas "need for timely information" usually falls in the former category.

After specifying the number of observations and variables, click OK to proceed. The software generates a template that contains either one or two sheets, depending on whether you have included discriminant data.

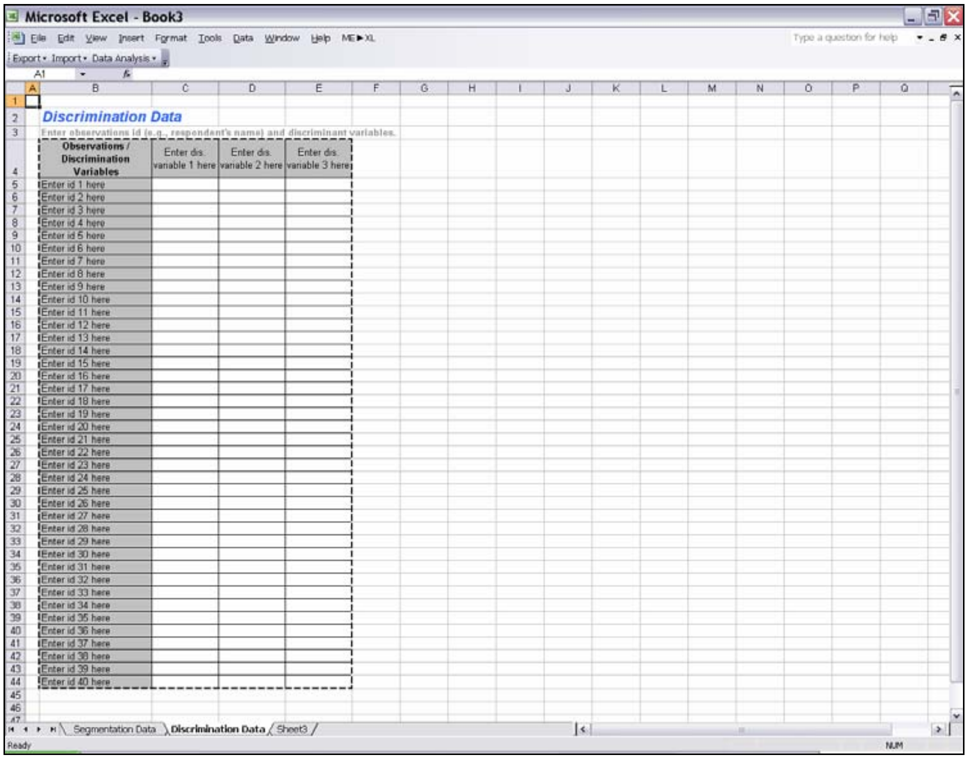


Microsoft Excel - Book3
File Edit View Insert Format Tools Data Window Help ME▶XL
Type a question for help
Export ▾ Import ▾ Data Analysis ▾
A1
Segmentation Data | Discrimination Data | Sheet3
Ready
NUM

Segmentation Data

Enter observations id (e.g., respondent's name) and segmentation variables.

Observations / Segmentation Variables	Enter seg. variable 1 here	Enter seg. variable 2 here	Enter seg. variable 3 here	Enter seg. variable 4 here	Enter seg. variable 5 here	Enter seg. variable 6 here
Enter id 1 here						
Enter id 2 here						
Enter id 3 here						
Enter id 4 here						
Enter id 5 here						
Enter id 6 here						
Enter id 7 here						
Enter id 8 here						
Enter id 9 here						
Enter id 10 here						
Enter id 11 here						
Enter id 12 here						
Enter id 13 here						
Enter id 14 here						
Enter id 15 here						
Enter id 16 here						
Enter id 17 here						
Enter id 18 here						
Enter id 19 here						
Enter id 20 here						
Enter id 21 here						
Enter id 22 here						
Enter id 23 here						
Enter id 24 here						
Enter id 25 here						
Enter id 26 here						
Enter id 27 here						
Enter id 28 here						
Enter id 29 here						
Enter id 30 here						
Enter id 31 here						
Enter id 32 here						
Enter id 33 here						
Enter id 34 here						
Enter id 35 here						
Enter id 36 here						
Enter id 37 here						
Enter id 38 here						
Enter id 39 here						
Enter id 40 here						



Discrimination Data

Enter observations id (e.g., respondent's name) and discriminant variables.

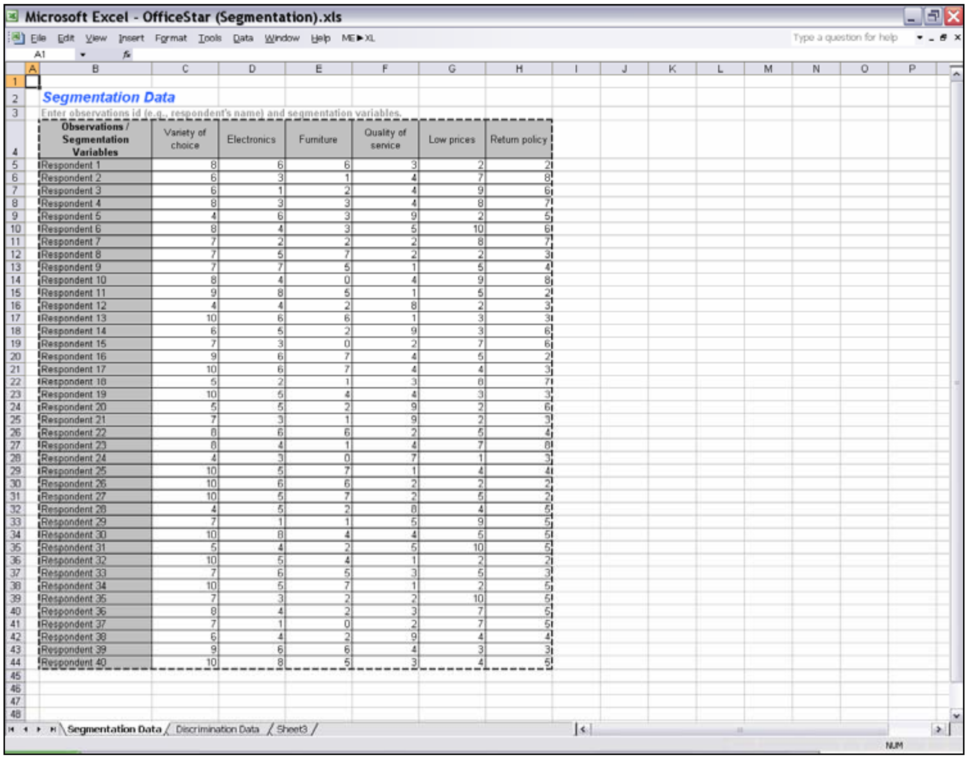
Observations / Discrimination Variables	Enter dis. variable 1 here	Enter dis. variable 2 here	Enter dis. variable 3 here
Enter id 1 here			
Enter id 2 here			
Enter id 3 here			
Enter id 4 here			
Enter id 5 here			
Enter id 6 here			
Enter id 7 here			
Enter id 8 here			
Enter id 9 here			
Enter id 10 here			
Enter id 11 here			
Enter id 12 here			
Enter id 13 here			
Enter id 14 here			
Enter id 15 here			
Enter id 16 here			
Enter id 17 here			
Enter id 18 here			
Enter id 19 here			
Enter id 20 here			
Enter id 21 here			
Enter id 22 here			
Enter id 23 here			
Enter id 24 here			
Enter id 25 here			
Enter id 26 here			
Enter id 27 here			
Enter id 28 here			
Enter id 29 here			
Enter id 30 here			
Enter id 31 here			
Enter id 32 here			
Enter id 33 here			
Enter id 34 here			
Enter id 35 here			
Enter id 36 here			
Enter id 37 here			
Enter id 38 here			
Enter id 39 here			
Enter id 40 here			

Step 2 Entering your data

F.Y.I.

In this tutorial, we use the example file "*OfficeStar (Segmentation).xls,*" which appears by default in "*My Documents/My Marketing Engineering/.*"

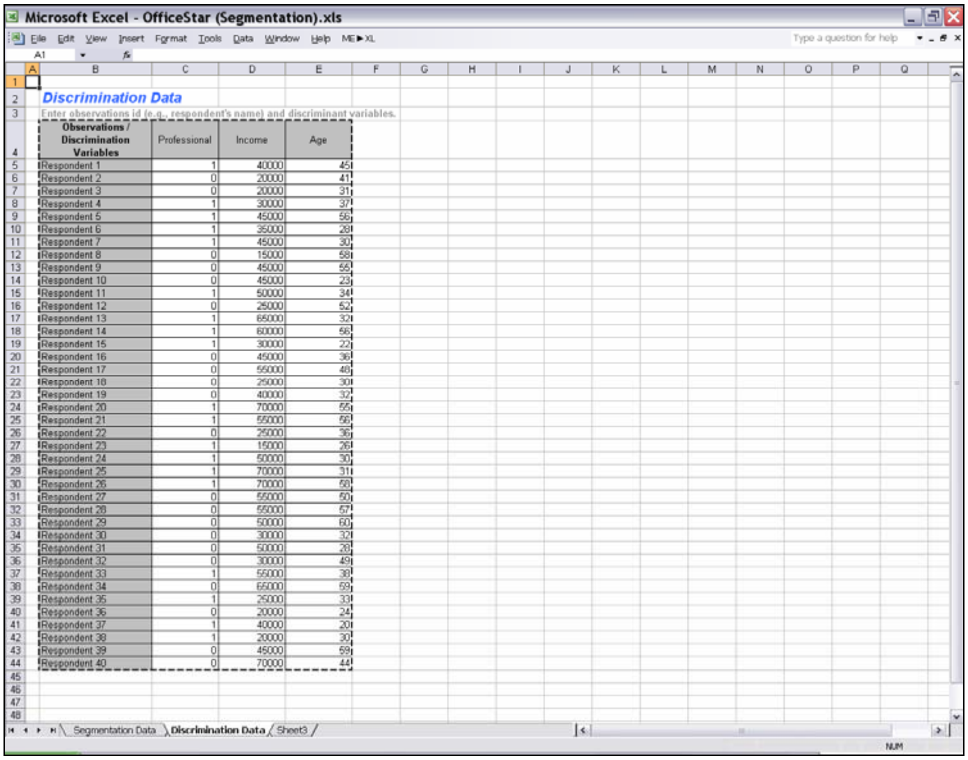
To view a proper data format, open that spreadsheet in Excel. A snapshot is shown below.


Microsoft Excel - OfficeStar (Segmentation).xls
File Edit View Insert Format Tools Data Window Help ME▶XL
Type a question for help
A1
Segmentation Data / Discrimination Data / Sheet3
NUM

Segmentation Data

Enter observations id (e.g., respondent's name) and segmentation variables.

Observations / Segmentation Variables	Variety of choice	Electronics	Furniture	Quality of service	Low prices	Return policy
Respondent 1	8	6	6	3	2	2
Respondent 2	6	3	1	4	7	8
Respondent 3	6	1	2	4	9	6
Respondent 4	8	3	3	4	8	7
Respondent 5	4	6	3	9	2	5
Respondent 6	8	4	3	5	10	6
Respondent 7	7	2	2	2	8	7
Respondent 8	7	5	7	2	2	3
Respondent 9	7	7	5	1	5	4
Respondent 10	8	4	0	4	9	8
Respondent 11	9	8	5	1	5	2
Respondent 12	4	4	2	8	2	3
Respondent 13	10	6	6	1	3	3
Respondent 14	6	5	2	9	3	6
Respondent 15	7	3	0	2	7	6
Respondent 16	9	6	7	4	5	2
Respondent 17	10	6	7	4	4	3
Respondent 18	5	2	1	3	8	7
Respondent 19	10	5	4	4	3	3
Respondent 20	5	5	2	9	2	6
Respondent 21	7	3	1	9	2	3
Respondent 22	8	6	6	2	5	4
Respondent 23	8	4	1	4	7	8
Respondent 24	4	3	0	7	1	3
Respondent 25	10	5	7	1	4	4
Respondent 26	10	6	6	2	2	2
Respondent 27	10	5	7	2	5	2
Respondent 28	4	5	2	8	4	5
Respondent 29	7	1	1	5	9	5
Respondent 30	10	8	4	4	5	5
Respondent 31	5	4	2	5	10	5
Respondent 32	10	5	4	1	2	2
Respondent 33	7	6	5	3	5	3
Respondent 34	10	5	7	1	2	5
Respondent 35	7	3	2	2	10	5
Respondent 36	8	4	2	3	7	5
Respondent 37	7	1	0	2	7	5
Respondent 38	6	4	2	9	4	4
Respondent 39	9	6	6	4	3	3
Respondent 40	10	8	5	3	4	5


Microsoft Excel - OfficeStar (Segmentation).xls
File Edit View Insert Format Tools Data Window Help ME▶XL
Type a question for help
A1
Segmentation Data / Discrimination Data / Sheet3
NUM

Discrimination Data

Enter observations id (e.g., respondent's name) and discriminant variables.

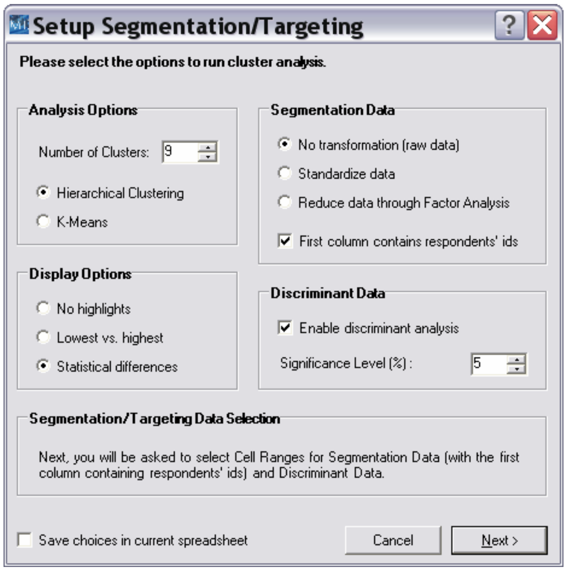
Observations / Discrimination Variables	Professional	Income	Age
Respondent 1	1	40000	45
Respondent 2	0	20000	41
Respondent 3	0	20000	31
Respondent 4	1	30000	37
Respondent 5	1	45000	56
Respondent 6	1	35000	28
Respondent 7	1	45000	30
Respondent 8	0	15000	58
Respondent 9	0	45000	55
Respondent 10	0	45000	23
Respondent 11	1	50000	34
Respondent 12	0	25000	52
Respondent 13	1	65000	32
Respondent 14	1	60000	56
Respondent 15	1	30000	22
Respondent 16	0	45000	36
Respondent 17	0	55000	48
Respondent 18	0	25000	30
Respondent 19	0	40000	32
Respondent 20	1	70000	55
Respondent 21	1	55000	56
Respondent 22	0	25000	36
Respondent 23	1	15000	26
Respondent 24	1	50000	30
Respondent 25	1	70000	31
Respondent 26	1	70000	58
Respondent 27	0	55000	50
Respondent 28	0	55000	57
Respondent 29	0	50000	60
Respondent 30	0	30000	32
Respondent 31	0	50000	28
Respondent 32	0	30000	49
Respondent 33	1	55000	38
Respondent 34	0	65000	59
Respondent 35	1	25000	33
Respondent 36	0	20000	24
Respondent 37	1	40000	20
Respondent 38	1	20000	30
Respondent 39	0	45000	59
Respondent 40	0	70000	44

A typical segmentation spreadsheet contains one or two spreadsheets that contain segmentation and/or discrimination data.

- **Segmentation data** are required for the segmentation model. This data set contains the respondent identifier and a column for each segmentation variable collected in the study. The data within each column must be scaled using the same scale (e.g., 1–10), but each column can have a different scale (e.g., 1–10 for satisfaction, 1–5 for convenience). Typically, segmentation variables are numerical values (interval or ratio scale). The data set contains one row per respondent in your study.
- **Discriminant data** constitute an optional data set, depending on whether your study has collected discrimination data. Recall that discrimination data enables you to differentiate one customer from another (e.g., age, income, gender). Again, data within a column must be scaled using the same scale, but different columns may use different scales. Typically, discriminant variables are numerical (interval or ratio scale) or nominal (“male”, “female”). Each respondent in your study appears in a separate row.

Step 3 Running segmentation analyses

After you enter your data in an Excel spreadsheet with the appropriate format, click on ME▸XL → SEGMENTATION AND CLASSIFICATION → RUN SEGMENTATION. The dialog box that appears indicates the next steps required to perform a segmentation analysis of your data.



Setup Segmentation/Targeting

Please select the options to run cluster analysis.

Analysis Options

Number of Clusters: 9

● Hierarchical Clustering

○ K-Means

Segmentation Data

● No transformation (raw data)

○ Standardize data

○ Reduce data through Factor Analysis

☑ First column contains respondents' ids

Display Options

○ No highlights

○ Lowest vs. highest

● Statistical differences

Discriminant Data

☑ Enable discriminant analysis

Significance Level (%) : 5

Segmentation/Targeting Data Selection

Next, you will be asked to select Cell Ranges for Segmentation Data (with the first column containing respondents' ids) and Discriminant Data.

□ Save choices in current spreadsheet

Analysis options

You may specify the number of segments (clusters) to develop during the analysis. For the segmentation method, you can choose either K-means or hierarchical clustering.

- **Hierarchical clustering** builds up or breaks down the data, customer by customer (row by row).
- **K-means** partitioning breaks the data into a prespecified number of segments and then reallocates or swaps customers to improve some measure of effectiveness.

> Usually, a segmentation analysis consists of two steps. First, you run the analysis with a large number of segments (up to 9). Second, on the basis of a dendogram analysis (discussed subsequently), you determine the optimal number of segments to retain.

Segmentation data

This section enables you to specify how to treat the data and whether a first column of respondent identifiers exists.

- **No transformation.** This button indicates you want to use the original data.
- **Standardize data.** This option scales all variables to 0 mean and unit variance before the analysis. Choosing this option is a good idea if you have measured the variables on different scales.
- **Reduce data through Factor Analysis.** This button combines related variables into unique factors.

Display options

In this section, you specify how you want the cluster data presented.

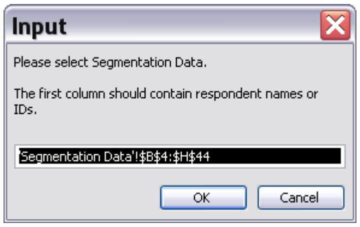
- **No highlights.** The data are unformatted.
- **Lowest vs. highest.** For each variable, colors highlight the value of the cluster with the highest (green) and lowest (red) values.
- **Statistical differences.** For each variable, colors highlight clusters whose values are statistically different from the overall mean at a 95% confidence level. Those that are different from the mean at a 99% confidence level appear in italics.

Discriminant data

Decide whether you want the analysis to include a discriminant analysis. Check this button if you wish to perform discriminant analysis, and indicate the level of statistical significance you wish to use.

> The *Save choices in current worksheet* option allows you to save cell range selections when you perform Run Analysis. If you are using your own data or have modified a Marketing Engineering for Excel template, you should choose this check box to save your selections.

After selecting all the options, you must select the cells containing the data. When you click Next, the following dialog box appears:



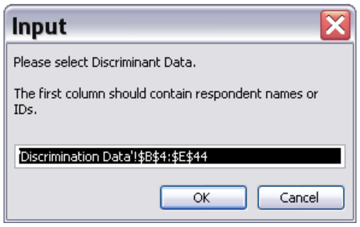
Please select Segmentation Data.

The first column should contain respondent names or IDs.

Segmentation Data'!$B$4:$H$44

The software requests a range for the segmentation data. If you are using a *Marketing Engineering for Excel* template, the software preselects the cell ranges.

If you have specified the inclusion of discriminant data, the following dialog box appears, which allows to select your discrimination data. The cell ranges might be preselected.



Input

Please select Discriminant Data.

The first column should contain respondent names or IDs.

Discrimination Data'!$B$4:$E$44

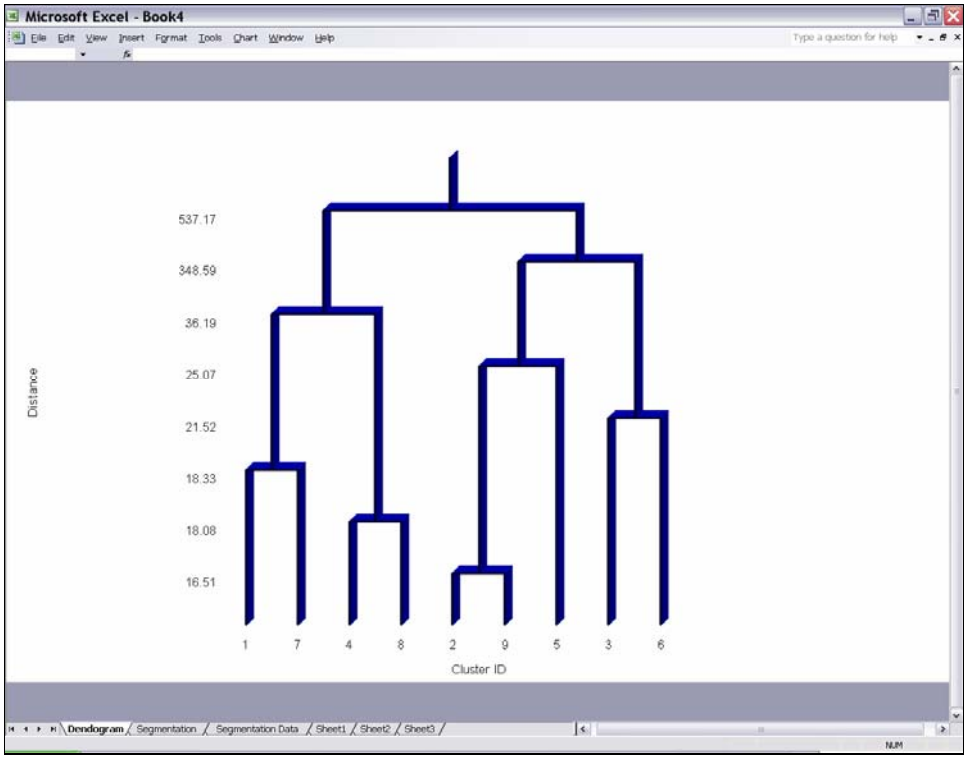
The newly generated workbook contains the results of your segmentation analysis.

Step 4 Interpreting the segmentation results

The workbook generated by segmentation analysis may contain several worksheets, depending on whether your study has included discriminant data.

Dendogram

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



At one extreme (upper part of the dendogram), all customers group into one cluster, and the loss of information is maximum, because they all receive undifferentiated treatment, regardless of their characteristics.

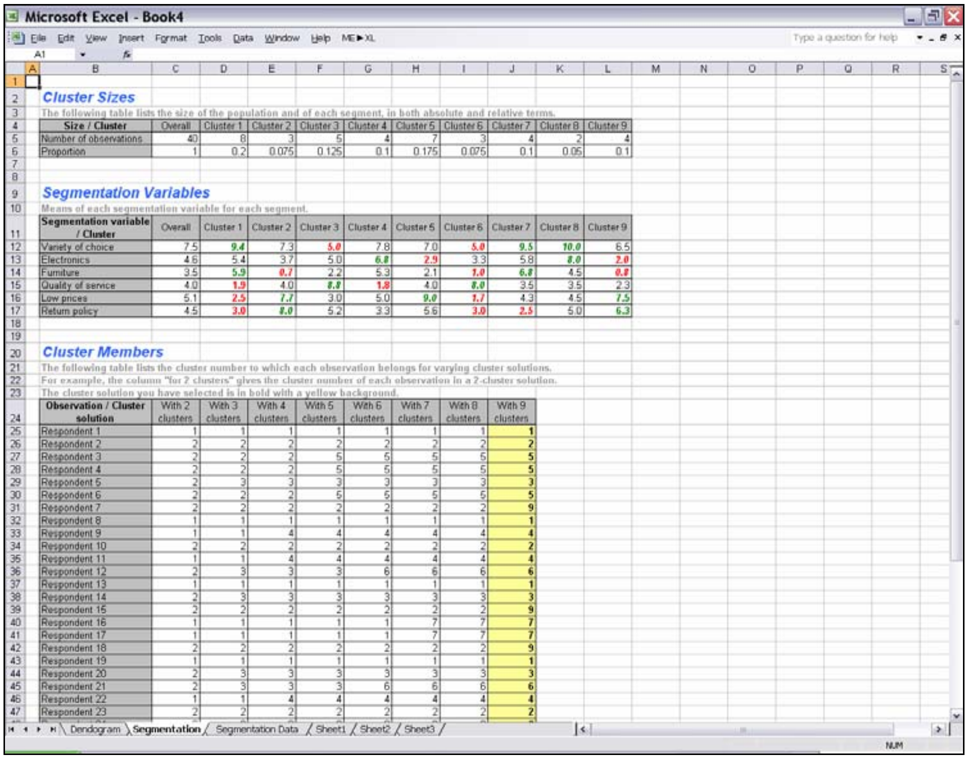
At the other extreme (lower part of the dendogram), customers appear in separate, small clusters, and only those customers very similar to one another group together ("similar" or "close" in this context refers to the distance between two customers in terms of the segmentation variables).

When reviewing a dendogram, look for significant distances or "jumps" in the distances. For example, the *OfficeMax* example contains a very large jump when moving from three to two clusters. Grouping these three clusters into two generates a significant loss of information; in other words, it results in grouping within the same cluster customers who are very dissimilar. In the preceding example, a three-cluster solution seems to be the best approach.

A dendogram is simply a graphical representation of the clustering output. For a more detailed understanding of cluster members and attributes, you must analyze the other tabs in the segmentation output as well.

Segmentation

The tab contains the statistical output of the cluster process and shows cluster sizes (number of members), cluster means, and the placement of each member in clusters (highlighted in yellow). This tab also provides columns that represent individual members and where they would be clustered in a 2–9 cluster 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	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Number of observations	40	8	3	5	4	7	3	4	2	4
Proportion	1	0.2	0.075	0.125	0.1	0.175	0.075	0.1	0.05	0.1

Segmentation Variables

Means of each segmentation variable for each segment.

Segmentation variable / Cluster	Overall	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Variety of choice	7.5	**9.4**	7.3	**5.0**	7.8	7.0	**5.0**	**9.5**	**10.0**	6.5
Electronics	4.6	5.4	3.7	5.0	**6.8**	**2.9**	3.3	5.8	**8.0**	**2.0**
Furniture	3.5	**5.9**	**0.7**	2.2	5.3	2.1	**1.0**	**6.8**	4.5	**0.8**
Quality of service	4.0	**1.9**	4.0	**8.8**	**1.8**	4.0	**8.0**	3.5	3.5	2.3
Low prices	5.1	**2.5**	**7.7**	3.0	5.0	**9.0**	**1.7**	4.3	4.5	**7.5**
Return policy	4.5	**3.0**	**8.0**	5.2	3.3	5.6	**3.0**	**2.5**	5.0	**6.3**

Cluster Members

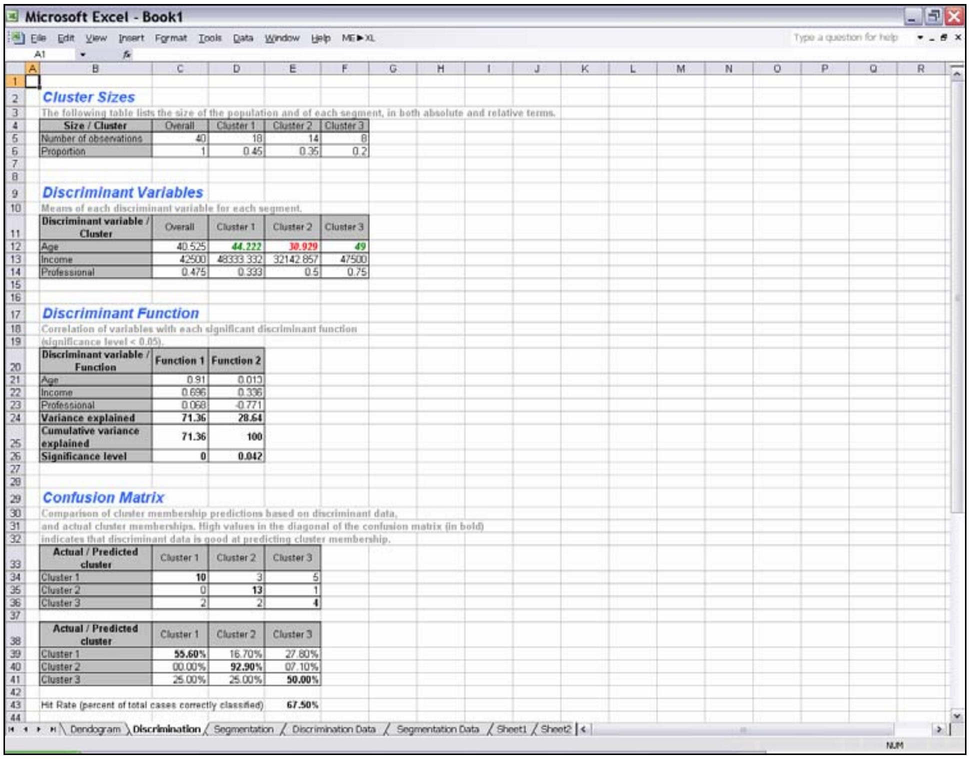
The following table lists the cluster number to which each observation belongs for varying cluster solutions.
For example, the column "for 2 clusters" gives the cluster number of each observation in a 2-cluster solution.
The cluster solution you have selected is in bold with a yellow background.

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
Respondent 1	1	1	1	1	1	1	1	**1**
Respondent 2	2	2	2	2	2	2	2	**2**
Respondent 3	2	2	2	5	5	5	5	**5**
Respondent 4	2	2	2	5	5	5	5	**5**
Respondent 5	2	3	3	3	3	3	3	**3**
Respondent 6	2	2	2	5	5	5	5	**5**
Respondent 7	2	2	2	2	2	2	2	**9**
Respondent 8	1	1	1	1	1	1	1	**1**
Respondent 9	1	1	4	4	4	4	4	**4**
Respondent 10	2	2	2	2	2	2	2	**2**
Respondent 11	1	1	4	4	4	4	4	**4**
Respondent 12	2	3	3	3	6	6	6	**6**
Respondent 13	1	1	1	1	1	1	1	**1**
Respondent 14	2	3	3	3	3	3	3	**3**
Respondent 15	2	2	2	2	2	2	2	**9**
Respondent 16	1	1	1	1	1	7	7	**7**
Respondent 17	1	1	1	1	1	7	7	**7**
Respondent 18	2	2	2	2	2	2	2	**9**
Respondent 19	1	1	1	1	1	1	1	**1**
Respondent 20	2	3	3	3	3	3	3	**3**
Respondent 21	2	3	3	3	6	6	6	**6**
Respondent 22	1	1	4	4	4	4	4	**4**
Respondent 23	2	2	2	2	2	2	2	**2**

Discrimination

This optional spreadsheet reflects the output of the discrimination analysis. The matrices included on this sheet are as follows:

- **Cluster sizes** depicts the number of respondents who appear in each cluster, along with the proportion of the whole population that each cluster represents.
- **Discriminant variables** depict the means of each discriminant variable for each cluster.
- **Discriminant function** reflects the correlation of the variables with each significant discriminant function and thus indicates the predictive ability of each discriminant function.
- **Confusion matrix** depicts how well the discriminant data predict correct clusters. Two matrices are available, one showing the actual data counts and the other showing percentages for these same data.
- **Classification weights** and **classification coefficients** are intermediary results required to run further classification analyses on external data. These matrices are of no particular interest as is, and cannot be easily interpreted, but are necessary to carry over further classification analyses.



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
Number of observations	40	18	14	8
Proportion	1	0.45	0.35	0.2

Discriminant Variables

Means of each discriminant variable for each segment.

Discriminant variable / Cluster	Overall	Cluster 1	Cluster 2	Cluster 3
Age	40.525	44.222	30.929	49
Income	42500	48333.332	32142.857	47500
Professional	0.475	0.333	0.5	0.75

Discriminant Function

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

Discriminant variable / Function	Function 1	Function 2
Age	0.91	0.013
Income	0.696	0.336
Professional	0.068	-0.771
Variance explained	**71.36**	**28.64**
Cumulative variance explained	**71.36**	**100**
Significance level	**0**	**0.042**

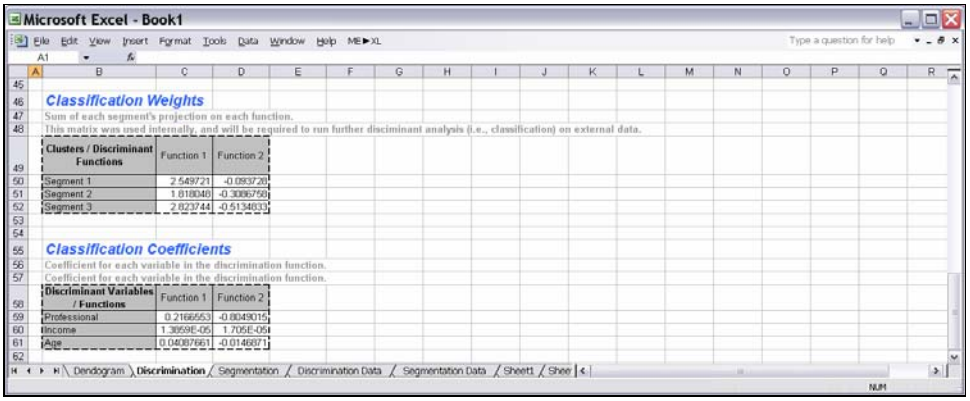
Confusion Matrix

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 1	**10**	3	5
Cluster 2	0	**13**	1
Cluster 3	2	2	**4**

Actual / Predicted cluster	Cluster 1	Cluster 2	Cluster 3
Cluster 1	**55.60%**	16.70%	27.80%
Cluster 2	00.00%	**92.90%**	07.10%
Cluster 3	25.00%	25.00%	**50.00%**

Hit Rate (percent of total cases correctly classified) **67.50%**



Classification Weights

Sum of each segment's projection on each function.

This matrix was used internally, and will be required to run further disciminant analysis (i.e., classification) on external data.

Clusters / Discriminant Functions	Function 1	Function 2
Segment 1	2.549721	-0.093728
Segment 2	1.818048	-0.3086758
Segment 3	2.823744	-0.5134833

Classification Coefficients

Coefficient for each variable in the discrimination function.

Coefficient for each variable in the discrimination function.

Discriminant Variables / Functions	Function 1	Function 2
Professional	0.2166553	-0.8049015
Income	1.3859E-05	1.705E-05
Age	0.04087661	-0.0146871

Segmentation and discriminant data

These tabs contain the original segmentation and discriminant data used for the segmentation analysis, included in the output for your convenience. The original spreadsheet used for the analysis remains intact, so you can modify it for subsequent analysis runs. The data preserved with this tab always reflect the data represented in the dendogram and segmentation tabs.

Step 5 Running classification analyses

Introduction

If you ran segmentation analysis with discriminant data, the software estimated the best way to predict to which cluster an individual is most likely to belong based solely on discriminant data. This is very useful to predict whether young people (age as a discriminant factor) are more likely to be more price sensitive (price sensitivity as a segmentation variable); or if businesses in certain industries require more support than others.

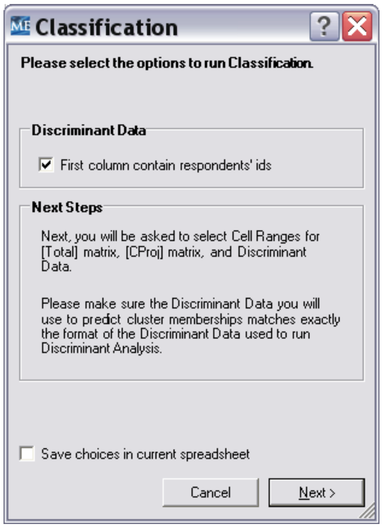
The ability of recouping segment membership based on discriminant variables is best summarized by the confusion matrix and hit rate (see above).

Once this discriminant analysis has been applied to the original dataset, it can be applied again to external customers for whom discriminant data –but no segmentation data- is available. The process of classifying customers among segments, based on a preceding segmentation analysis, but using discriminant data only, is called **classification analysis**.

Classification analysis is usually applied to new customers, for whom segmentation data is not available. For learning purpose, you can also apply it to discriminant data of customers for whom segmentation data is available, and see how well segment memberships are recouped. This analysis is automatically done when you run a segmentation analysis, and its results are summarized by the confusion matrix.

Selecting data

Click on ME►XL → SEGMENTATION AND CLASSIFICATION → RUN CLASSIFICATION. The dialog box that appears indicates the next steps required to perform a classification analysis of your data.



Please select the options to run Classification.

Discriminant Data

☑ First column contain respondents' ids

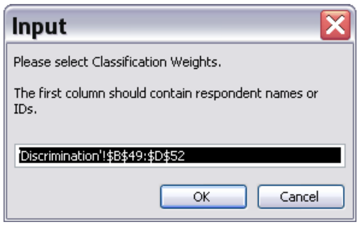
Next Steps

Next, you will be asked to select Cell Ranges for [Total] matrix, [CProj] matrix, and Discriminant Data.

Please make sure the Discriminant Data you will use to predict cluster memberships matches exactly the format of the Discriminant Data used to run Discriminant Analysis.

☐ Save choices in current spreadsheet

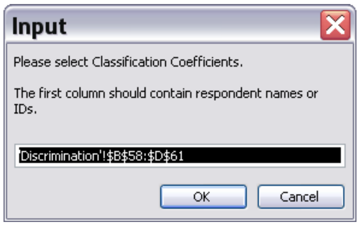
The first steps of the classification analysis consist of selecting two cell ranges: classification weights and classification coefficients. You can find that data at the bottom of the “discriminant” sheet in the analysis workbook generated by segmentation analysis. The cell ranges might be preselected.



Please select Classification Weights.

The first column should contain respondent names or IDs.

'Discrimination'!$B$49:$D$52

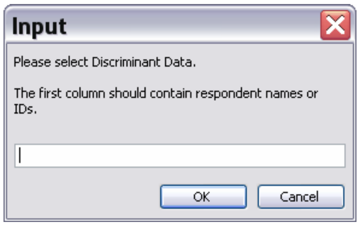


Please select Classification Coefficients.

The first column should contain respondent names or IDs.

'Discrimination'!$B$58:$D$61

The last step is to select discriminant data. In most cases, that consists of data about new customers for whom no segmentation data is available. It is important that formatting of the discriminant data matches exactly the format of the discriminant data (both variables, orders and ranges) used for the original segmentation analysis.

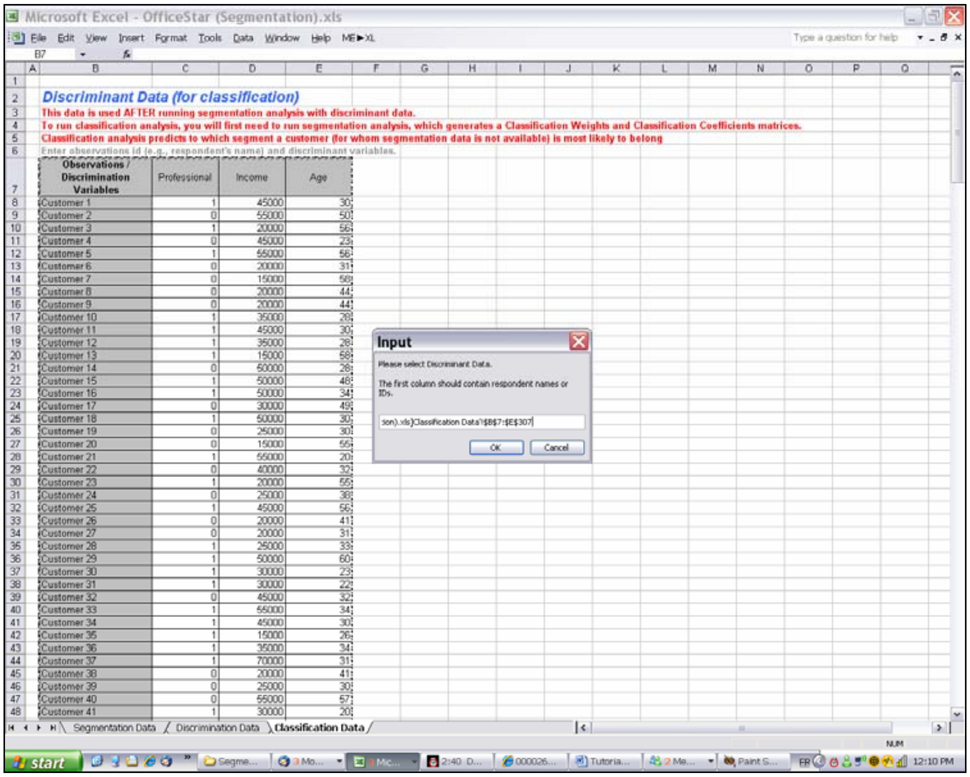


Input

Please select Discriminant Data.

The first column should contain respondent names or IDs.

Discriminant data of "new" customers is available on the original OfficeStars workbook, in the last sheet. Go back to the OfficeStar workbook, and manually select the discriminant data available for the 300 additional customers (the last sheet of the workbook, named classification data).



Microsoft Excel - OfficeStar (Segmentation).xls
File Edit View Insert Format Tools Data Window Help ME►XL
Type a question for help
B7

Discriminant Data (for classification)

This data is used AFTER running segmentation analysis with discriminant data.

To run classification analysis, you will first need to run segmentation analysis, which generates a Classification Weights and Classification Coefficients matrices.

Classification analysis predicts to which segment a customer (for whom segmentation data is not available) is most likely to belong

Enter observations id (e.g., respondent's name) and discriminant variables.

Observations / Discrimination Variables	Professional	Income	Age
Customer 1	1	45000	30
Customer 2	0	55000	50
Customer 3	1	20000	56
Customer 4	0	45000	23
Customer 5	1	55000	56
Customer 6	0	20000	31
Customer 7	0	15000	58
Customer 8	0	20000	44
Customer 9	0	20000	44
Customer 10	1	35000	28
Customer 11	1	45000	30
Customer 12	1	35000	28
Customer 13	1	15000	58
Customer 14	0	60000	28
Customer 15	1	50000	48
Customer 16	1	50000	34
Customer 17	0	30000	49
Customer 18	1	60000	30
Customer 19	0	25000	30
Customer 20	0	15000	55
Customer 21	1	55000	20
Customer 22	0	40000	32
Customer 23	1	20000	55
Customer 24	0	25000	38
Customer 25	1	45000	56
Customer 26	0	20000	41
Customer 27	0	20000	31
Customer 28	1	25000	33
Customer 29	1	50000	60
Customer 30	1	30000	23
Customer 31	1	30000	22
Customer 32	0	45000	32
Customer 33	1	55000	34
Customer 34	1	45000	30
Customer 35	1	15000	26
Customer 36	1	35000	34
Customer 37	1	70000	31
Customer 38	0	20000	41
Customer 39	0	25000	30
Customer 40	0	55000	57
Customer 41	1	30000	20

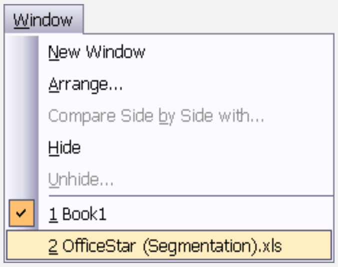
Input

Please select Discriminant Data.

The first column should contain respondent names or IDs.

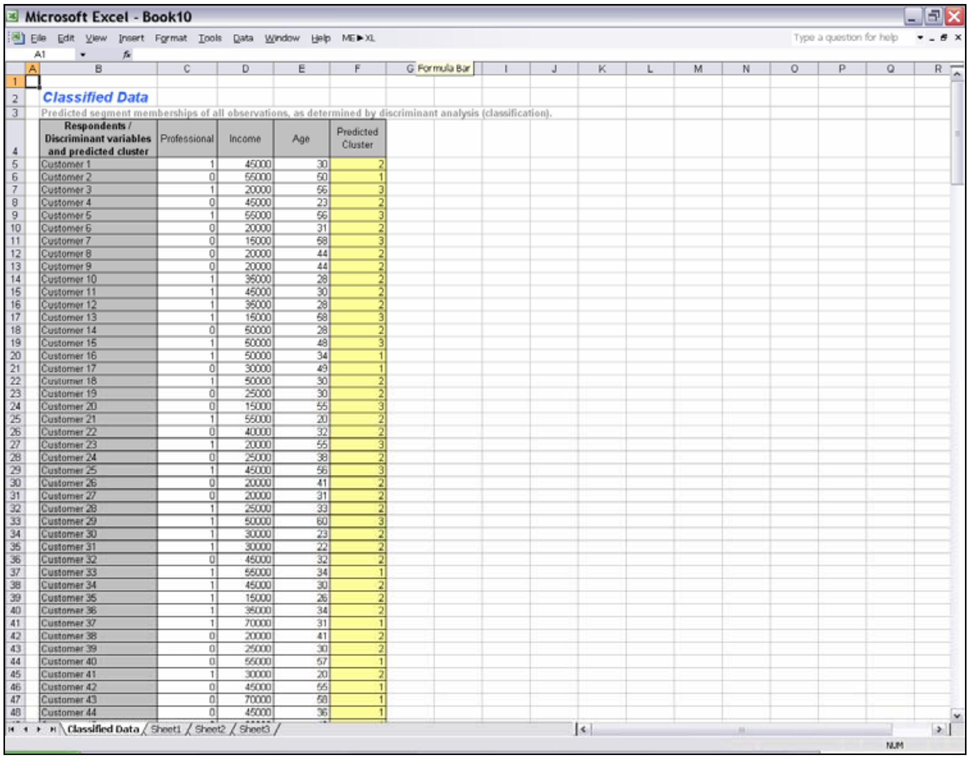
sion).xls]Classification Data'!$B$7:$E$307

Once you are in selecting mode, Excel might not allow you to easily switch between two workbooks. If you require selecting data in different workbooks (as it is usually the case with classification analysis), simply use the Window menu of Excel to select and open another workbook.



Interpreting the results

When you click Ok, a new workbook is generated. This workbook contains the discriminant data used to run classification analysis, and the segment to which each customer is most likely to belong.



Classified Data

Predicted segment memberships of all observations, as determined by discriminant analysis (classification).

Respondents / Discriminant variables and predicted cluster	Professional	Income	Age	Predicted Cluster
Customer 1	1	45000	30	2
Customer 2	0	55000	50	1
Customer 3	1	20000	56	3
Customer 4	0	45000	23	2
Customer 5	1	55000	56	3
Customer 6	0	20000	31	2
Customer 7	0	15000	58	3
Customer 8	0	20000	44	2
Customer 9	0	20000	44	2
Customer 10	1	35000	28	2
Customer 11	1	45000	30	2
Customer 12	1	35000	28	2
Customer 13	1	15000	58	3
Customer 14	0	50000	28	2
Customer 15	1	50000	48	3
Customer 16	1	50000	34	1
Customer 17	0	30000	49	1
Customer 18	1	50000	30	2
Customer 19	0	25000	30	2
Customer 20	0	15000	55	3
Customer 21	1	55000	20	2
Customer 22	0	40000	32	2
Customer 23	1	20000	55	3
Customer 24	0	25000	38	2
Customer 25	1	45000	56	3
Customer 26	0	20000	41	2
Customer 27	0	20000	31	2
Customer 28	1	25000	33	2
Customer 29	1	50000	60	3
Customer 30	1	30000	23	2
Customer 31	1	30000	22	2
Customer 32	0	45000	32	2
Customer 33	1	55000	34	1
Customer 34	1	45000	30	2
Customer 35	1	15000	26	2
Customer 36	1	35000	34	2
Customer 37	1	70000	31	1
Customer 38	0	20000	41	2
Customer 39	0	25000	30	2
Customer 40	0	55000	57	1
Customer 41	1	30000	20	2
Customer 42	0	45000	55	1
Customer 43	0	70000	58	1
Customer 44	0	45000	36	1

Note that this classification of customers across segments is our best guess based on discriminant analysis. It is not perfect, and some customers might be misclassified, that is, they are the closest to segment *A* in terms of needs, but their discriminant variables send us astray and predict they are more likely to belong to segment *B*.