# 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.

ME	E►XL		
	Bass F	orecasting Model	►
	Conjoi	nt	→
	Custor	ner Choice (Logit)	→
	Custor	ner Lifetime Value	→
	GE Por	tfolio Matrix	►
	Positio	ning	►
	Resour	ce Allocation	►
	Segme	ntation and Classification	►
<b>S</b>	Getting	Started (PDF)	
0	Market	ing Engineering for Excel Help	
	About	Marketing Engineering for Excel	
	About	DecisionPro	

### **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 Data (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  $\triangleright$  XL  $\rightarrow$  SEGMENTATION AND CLASSIFICATION  $\rightarrow$  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 🗧
Cancel	OK

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.



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.

crosoft Excel -	BOOK3													-
e Edit Yew Inser	t Format Iool	ls Data <u>W</u> in	dow Help M	E►XL								Type a qu	stion for hel	
Import      Data Analys	is •													
В	Ċ	D	E	F	G	н	1	J	K	L	M	N	0	P
Segmentation	Data													
Observations (	(e.q., responde	nt's name) and	segmentation	a variables.										
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 herer								
Enter id 1 here														
Enter id 2 here					1									
Enter id 3 here						i i								
Enter id 4 here					1									
Enter id 5 here						i								
Enter id 6 here					-									
Enter id 7 here			-			- i								
Enter id 8 here			-											
Enter id 9 here	-					i								
Enter id 10 here	-		-											
Enter id 11 here	-		-											
Enter id 12 here	-		-											
Enter id 13 here	-		-											
Enter id 14 here	-				-									
Enter is 15 here	-		-											
Enter of 17 have	-	<u> </u>	-		-									
Enter id 18 here														
Enter id 19 here				-										
Enter id 20 here														
Enter id 21 here						1								
Enter id 22 here														
Enter id 23 here					1	1								
Enter id 24 here														
Enter id 25 here	1				1									
Enter id 26 here														
Enter id 27 here						i i								
Enter id 28 here														
Enter id 29 here					-	i								
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 30 hare	-		-		-	+								
Enter id 39 hare	-		-		-						1			
Enter id 40 here	-					+								
Containe and United				+		+								
	the strategic proton													
and formant attack	Data / Directimin	ation Data / C	hasts /				141							

															1.2
e Edit View Incer	t Format Icol	is Data Win	dow Help ME	►XI.									Type a q	action for	help
• Import • Data Analys	is • 1														
1 - 6	allocated.														
В	Ć	D	E	F	G	н	1	J	K	L	M	N	0	P	Q
Discrimination	n Data														
Enter observations Id	in a responder	nt's name) and	discriminant v	ariables											-
Observations /			[]												-
Discrimination	Enter dis. variable 1 here	Enter dis. variable 2 here	Enter dis. variable 3 heres												
Fetor id 1 horo															
Enter id 2 hare			i												-
Entor id 3 hara															-
Enter id 4 have	-		1 1												-
Enter id 5 hare															
Enter id 6 here															-
Feter id 7 here			1												-
Enter id 8 here															
Enter id 9 here			i i												-
Enter id 10 here															-
Enter id 11 here															
Enter id 12 here															
Enter id 13 here															
Enter ut 14 here			i i												
Enter id 15 here															-
Enter id 16 here			1 1												-
Enter id 17 here	1														
Enter id 18 here															
Enter id 19 here															
Enter id 20 here															
Enter id 21 here	1.00														
Enter id 22 here															
Enter id 23 here															
Enter id 24 here															
Enter id 25 here															
Enter id 26 here															
Enter id 27 here			1												
Enter id 28 here															
Enter id 29 here			i												
Enter id 30 here															-
Enter id 31 here															-
Enter id 32 here			1 1												
Enter id 33 here	-		-												-
Encer id 34 here	-														-
cinter id 35 here			-												
cincer id 36 here	-		- i												+
Enser id 37 here	-														-
Enter id 30 here	-				_			-							-
Enter id sy here															
Enter ig 40 here															
	10 m	1 - 1 12	1 10 1					1.1							

# Step 2 Entering your data



In this tutorial, we use the example file "*OfficeStar Data (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.

	Inclosoft Excel 0	mcestar	Segmenta	tion).xis											
) (	Elle Edit ⊻lew Insert	Format Tool	s <u>D</u> ata <u>W</u> ind	low <u>H</u> elp Mi	i►XL								Type a que	astion for hel	p •
_	A1 • 5														
1	В	C	D	E	F	G	н	1	J	K	L	M	N	0	P
Ļ															
L	Segmentation L	Data													
Γ	Enter observations id (	e.q., responder	it's name) and	segmentation	variables.										
Γ	Observations /	Variaty of		نفتفتقا	Quality of										
	Segmentation	choice	Electronics	Furniture	senice	Low prices	Return policy								
	Variables	0110100			0011100										
	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		9								
	Respondent 6	8	4	3	5	10	6								
	Respondent /	4	2	2											
	Respondent o	7	5	6		4	3								
	Respondent 10	6		0											
	Despendent 11	0		6		6	2								
	Despondent 12			2			2								
	Respondent 13	10	6	6	1		3								
	Respondent 14	6	5	2		3	6								
	Respondent 15	7	3	õ	2	7	6								
	Respondent 16	9	6	7	4	6	2								
	Respondent 17	10	6	7	4	4	3								
	Respondent 10	5	2	1	3	0	7								
	Respondent 19	10	5	4	4	3	3								
	Respondent 20	5	5	2	9	2	6								
	Respondent 21	7	3	) <u>1</u>	9	2	3								
	Respondent 22	8	6	6	2	5	4								
	Respondent 23	8	4	1	4	7	8								
	Respondent 24	4	3	0		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	6	5		-						
	Respondent 31	5	4		5	10	5								
	Respondent 32	10	5	4			2								
	Respondent 33	10	6	5	3		3								
	Prespondent 34	- 10	9			10	5								
	Personal and 25		3			10	6								
	Respondent 37	7	1		2		5								
	Respondent 39	6	4	2		1 2	2								
	Respondent 39	9	6	6	4		3								
	Respondent 40	10	8	5	3	4	5								
	Commentation Da	ta / Discrimina	tion Data / Ch	ants /		1		1.1							

-		<u></u>	-	-	-	-									-	-	
12	A B	С	D	E	F	G	н	1	J	К	L	M	N	0	P	Q	
-	Discrimination						-										-
	Discrimination	dla															
-	Enter observations id	le.g., respond	dent's name] an	d discrimina	nt varia	bles.											-
	Discriptions r	Professional	Income (000%)	4.00													
	I Discrimination	Professional	moonie (000 s)	- Oge													
	Respondent 1	1	40	45													-
	Bespondent 2	i i	20	41													-
	Respondent 3	0	20	31													-
	Respondent 4	1	30	37													-
	Respondent 5	1	45	56													-
	Respondent 6	1	35	28													-
	Respondent 7	1	45	30													
	Respondent 8	0	15	58													
	Respondent 9	0	45	55													
	Respondent 10	0	45	23													
	Respondent 11	1	50	34													
	Respondent 12	0	25	52													
	Respondent 13	1	65	32													
	Respondent 14	1	60	56													-
	Respondent 15	1	30	22													-
	Respondent 16	0	45	36								_					-
	Respondent 17	0	55	48													-
	Respondent 18	0	25	30													
	Hespondent 19	0	40	32													
	Hespondent 20	-	70	00													-
	Respondent 21		00	06													-
	Respondent 22		20	30													-
	Respondent 24	1	50	20			-										-
	Respondent 25	1	70	21													-
	Bespondent 26	1	70	58													-
	Bespondent 27	ó	55	50													-
	Bespondent 28	0	55	57					<u> </u>	1	1		-				1
	Respondent 29	ő	50	60						•							1
	Respondent 30	0	30	32													1
	Respondent 31	0	50	28													1
	Respondent 32	0	30	49													
	Respondent 33	1	55	38													
	Respondent 34	0	65	59													
	Respondent 35	1	25	33													
	Respondent 36	0	20	24													
	Respondent 37	1	40	20													
	Respondent 38	1	20	30													
	Respondent 39	0	45	59													
	Respondent 40	0	70	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 <u>within</u> 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 <u>within</u> 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  $\rightarrow$  XL  $\rightarrow$  SEGMENTATION AND CLASSIFICATION  $\rightarrow$  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 clu	ster analysis.
Analysis Options	Segmentation Data
Number of Clusters: 9 📩	No transformation (raw data)     Standardize data
Hierarchical Clustering     K-Maana	C Reduce data through Factor Analysis
	✓ First column contains respondents' ids
Display Options C No highlights	Discriminant Data
C Lowest vs. highest	Enable discriminant analysis
··· Statistical dinerences	
Segmentation/Targeting Data Select Next, you will be asked to select Ce column containing respondents' ids)	<b>xction</b> Il Ranges for Segmentation Data (with the first and Discriminant Data.
Save choices in current spreadsheet	Cancel <u>N</u> ext >

### 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:

Input	X
Please select Segmentation Data.	
The first column should contain respondent names or IDs.	,
Segmentation Data'!\$B\$4:\$H\$44	
OK Cance	

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 pre-selected.

Input 🛛 🚺	<
Please select Discriminant Data.	
The first column should contain respondent names or IDs.	
Discrimination Data'!\$B\$4:\$E\$44	
OK Cancel	

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.

:	Eile Edit Yiew Insert Fo	rmat <u>T</u> ools	<u>D</u> ata <u>W</u>	indow <u>H</u> elp	ME⊫XL	Adobe PDF							
10	🎽 🔲 🖪 🌧 🎒 🖉	9 101   2	🗈 🙈 • <	<b>y - c</b>	μ	E - Al XI	1 🌆 📣 1	.00% 👻 🌘					
					. 69	21 11			· •				
	A B	<u>^</u>	D	F	F	G	н		1	K		м	N
1				_							_		
2	Cluster Sizes												
3	The following table list	s the size	of the pop	lation and	d of each s	segment, i	n both abs	olute and	relative te	rms.			
4	Size / Cluster	Overall	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9		
5	Number of observations	40	8	3	5	4	7	3	4	2	4		
6	Proportion	1	0.2	0.075	0.125	0.1	0.175	0.075	0.1	0.05	0.1		
7													
8													
9	Segmentation V	ariable/	s										
10	Means of each segmen	tation vari	able for ea	nch segme	nt.								
11	Segmentation variable / Cluster	Overall	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9		
12	Variety of choice	7.53	9.38	7.33	5	7.75	7	5	9.5	10	6.5		
13	Electronics	4.57	5.38	3.67	5	6.75	2.86	3.33	5.75	8	2		
14	Furniture	3.45	5.88	0.667	2.2	5.25	2.14	1	6.75	4.5	0.75		
15	Quality of service	4	1.88	4	8.8	1.75	4	8	3.5	3.5	2.25		
16	Low prices	5.05	2.5	7.67	3	5	9	1.67	4.25	4.5	7.5		
17	Return policy	4.5	3	8	5.2	3.25	5.57	3	2.5	5	6.25		
18													
19													
20	Cluster Membe	rs											
21	The following table list	s the clust	er number	to which (	each obse	rvation be	longs for v	arying clu	ster soluti	ons.			
22	For example, the colum	nn "for 2 c	lusters" gir	/es the clu	ster numb	er of each	observati	on in a 2-c	luster solu	rtion.			
23	The cluster solution you	u have sel	ected is in	bold with	a yellow l	ackgroun	d.						
	Observation / Cluster	With 2	With 3	With 4	With 5	With 6	With 7	With 8	With 9				
24	solution	clusters	clusters	clusters	clusters	clusters	clusters	clusters	clusters				
25	Respondent 1	1	1	1	1	1	1	1	1				
26	Respondent 2	2	2	2	2	2	2	2	2				
27	Respondent 3	2	2	2	5	5	5	5	5				
28	Respondent 4	2	2	2	5	5	5	5	5				
29	Respondent 5	2	3	3	3	3	3	3	3				
30	Respondent 6	2	2	2	5	5	5	5	5				
31	Respondent 7	2	2	2	2	2	2	2	9				
32	Respondent 8	1	1	1	1	1	1	1	1				
33	Respondent 9	1	1	4	4	4	4	4	4				
.34	Respondent 10		7 montation	2 Discrimina	L 2	2 Segmentati	on Data 7	2 Sheet1 / She	et2 / Sheet	3/	• 1		

### 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.

crosoft Excel - Book	(4														
Eile Edit View 1	Insert Fo	rmat <u>T</u> oo	ols <u>D</u> ata	Window	v <u>H</u> elp	ME 🛏 XI	L Adoļ	e PDF							
	ABI	9- 454 IV			0.0		5 -	AIZI	Alia 🚮	70%	- @]				
	∍ ⊡,   √	° ⊪A   ⊅	43 📭	• 🗸 I	-) • (-	· · · 😒	2 *	Z 🕈 A 🕇	<u> </u>	7078	<u> </u>	Ŧ			
H2U -	∱x			_											_
а в	C	0	E	F	Gi	н		J	K	L	M	N	0	P	_
Cluster Circo															-
Chilster Sizes	lists the sit	n of the pr		ndefer	h come	nt in hot	habeels	to and re	Inting tor						
Size / Cluster	Overall	Cluster 1	Cluster 2	Cluster 3	in segme	int, in bot	ii absolo	ice and re	lative teri						-
Number of observations	40	18	14	8											
Proportion	1	0.45	0.35	0.2											
															<u> </u>
D1 1 1 11															-
Discriminant Va	riables														
Means of each disc	riminant va	riable for e	each segm	ent.											
variable / Cluster	Overall	Cluster 1	Cluster 2	Cluster 3											
Age	40.525	44.222	30.525	45											
Income (000's)	42.5	48.333	32.143	47.5											
Professional	0.475	0.333	0.5	0.75											
															-
Discriminant Eu	nction														
Discriminant Fu	hlag with a	ash sisaifi	ant disori	nin ant fu	nation										
fsignificance level of	0.051	ach Sigiliriu	ancuisem	iiiiaiit ru	liction										
Discriminant	Function	Function													-
variable / Function	1	2													
Age	0.91	0.013													
Income (000's)	0.696	0.336													
Variance explained	71.36	28.64													-
Cumulative	71.90	100													
variance explained	11.30	100													
Significance level	0	0.042													
Confusion Matri	v														
Comparison of elus	ter membe	rchin prodi	ations has	ad on die	oriminan	t data									
and actual cluster m	nembership	s. High val	ues in the	diagonal	of the co	nfusion r	natriz fir	n bold)							-
indicates that discri	iminant dat	a is good a	t predictin	g cluster	members	ship.		<b>,</b>							
Actual / Predicted	Cluster 1	Cluster 2	Cluster 3												
cluster	10														
Cluster 1 Cluster 2	0	13	1												
Cluster 3	2	2	4												-
Actual / Predicted	Cluster 1	Cluster 2	Cluster 3												
Cluster Cluster 1	55 60%	16.70%	27.90%												
Cluster 2	00.00%	92.90%	07.10%												-
Cluster 3	25.00%	25.00%	50.00%												
Hit Rate (percent of tota	l cases corre	ctly classified	67.50%												-
															-
Classification 14/	alahta							-							-
	Discrimin	aation /	Segmentati	on / D	iccrimin at	ion Data	/ Sec	mentatio	Data /	Sheet1	/ Sheet?	/ Sheet	21		

Classification V	Veights														
Sum of each segm	ent's projec	tion on eac	h functio	on.											
This matrix was us	ed internally	, and will be	e required	d to run f	urther dis	ciminant	analysis (	i.e., class	ificatio	) on exte	rnal data.				
Clusters / Discriminant Functions	Function 1	Function 2													
Segment 1	2.549721	-0.0937281													-
Segment 2	1.818049	-0.3086759													
Segment 3	2.823745	-0.5134836													
															-
Classification C	oefficient	s													
Coefficient for eac	sh variable ii	n the discri	mination	function											
Coefficient for each	sh variable i	n the discri	mination	function											
Discriminant Variables I	Function 1	Function 2													
Professional	0.2166553	-0.8049017													<u> </u>
Income (000's)	0.0138589	0.0170497													
Age	0.0408766	-0.0146871													
															_
															-
															-
															-
															-
															-
	-														-
D	N						1 -		-	1	1	1	- 1	1.1	

### 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 ightarrow XL ightarrow SEGMENTATION AND CLASSIFICATION ightarrow RUN CLASSIFICATION. The dialog box that appears indicates the next steps required to perform a classification analysis of your data.

Market Classification										
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 Cancel Next >										

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 pre-selected.

Input 🔀									
Please select Classification Weights.									
The first column should contain respondent names or IDs.									
Discrimination'!\$B\$49:\$D\$52									
OK Cancel									

Input 🔀									
Please select Classification Coefficients.									
The first column should contain respondent names or IDs.									
Discrimination'!\$B\$58:\$D\$61									
OK Cancel									

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 Discrimina	ant Data.
The first column should IDs.	contain respondent names or
	OK Cancel

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).

	De for Ten hear of	enar Toole R	iara Tinnow Ga		ALZII	404 eft [109								179	e a question	inormop -	0
		( BA   A 40	12 · 4   • .	(- · ) 2 2 .	* 24 X41	M 49 100		5									
		C	D	E	F	G	н	- 1	1	V		M	N	0	P	0	
. 1	N D	Ų.	U	E	r	Ū.	n		Ĵ.	n.		M	Di	0	P	U U	-
	Discriminant D	ata Hor el	accification	1											-		-
	Discriminant D		assilicatio	<i>•</i>					_							_	-
	This data is used AFTE	R running seg	mentation analy	rsis with discri	iminant d	ata.		Change	Conton Mol	abo and	Class Marriel	an Conffe	lane mar	1			-
	To run classification an	alysis, you wi	Il first need to r	in segmentati	ion analy	as, which	generates a	Class	lication Wei	ghts and	Classificati	on Coeffic	ients matr	ices.			
	Classification analysis	predicts to wh	ich segment a c	ustomer (tor v	whom seg	mentation	data is not	avaita	ble) is most	likely to b	erong						+
	Observations in the	.q., responde	115 Halley and		anapies.						-						-
	Discrimination	Professional	Income (000%)	Ann													
	Variables	Protessional	income (ooo s)	Ada 1													
	Customer 1	1	45	30												-	-
	Customer 2	0	1 55	50												-	
ii.	Customer 3	1	20	56												-	-
1	Customer 4	0	45	23													
2	Customer 5	1	55	56	Input				×								
3	Customer 6	0	20	31	Disco	a cale of Discol	and the second se										
1	Customer 7	0	15	581	Pieas	e select Lischi	ninark Liaca.										
5	Customer 8	0	1 20	44	The f	irst column shi	ould contain res	spondent	names or								
5	Customer 9	0	20	44	IDs.												
	Customer 10	1	35	28													
3	Customer 11	1	45	30	(ion)	ods]Classifical	ion Data'\\$8\$7	7:\$E\$307									
<u>.</u>	Customer 12	1	35	28				_									
٤.	Customer 13	1	15	58			OK	( ) ( )	Cancel								
Ц.	Customer 14	0	50	28	_		_			1							
1	Customer 15	1	50	481													
<u>.</u>	Customer 16	1	50	34													_
1	Customer 17		30	49													
2	Customer 18	1	50	30					_								_
9	Customer 19	0	25	30													
5	Customer 20		1 15	50;					_								+
8	Customer 21		30	201													+
-	Customer 22		20	52													
1	Customer 24		1 26	38										-			-
	Customer 25	1	45	56													+
2	Customer 26	i i	20	41													
	Customer 27	0	20	31												-	T
ε.	Cipetnemar 2R	1 1	26	33												1.	
4	<ul> <li>N Segmentation Data (</li> </ul>	Discrimination Da	Ra)Classificatio	n Data /					•								



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.

Wir	ndow	
	New	Window
	<u>A</u> rra	nge
	Com	pare Side <u>b</u> y Side with
	Hide	
	<u>U</u> nhi	ide
~	<u>1</u> Bo	ok1
	<u>2</u> Off	ficeStar (Segmentation).xls

### 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.

💌 M	icrosoft Excel - Book6																
:	<u>File E</u> dit <u>V</u> iew Ins	sert F <u>o</u> rm	at <u>T</u> ools	<u>D</u> ata <u>V</u>	<u>v</u> indow <u>H</u>	elp ME	►XL Ad	o <u>b</u> e PDF									
	📂 🛃 💪 🔒 🖪	🛕   💞	🕰   🔏 🛛	🐚  🔹 🕯	🍼 i 🌱 👻	6	🧕 Σ 🗸	$\begin{array}{c} \mathbf{A} \\ \mathbf{Z} \\ \mathbf{Z} \\ \mathbf{A} \end{array}$	। 🛄 🦧	75%	• 🕜	Ŧ					
	A1 🔻	f <sub>x</sub>															
	A B	С	D	E	F	G	н	1	J	K	L	M	N	0	P	Q	1
1	1																1
2	Classified Data																
3	Predicted segment m	emberships	of all obse	rvations, a	s determine	ed by disci	riminant a	nalysis fo	lassificat	ion).							
	Respondents /					1											
	Discriminant	Profession	Income	0.00	Predicted												
	variables and	al	(000's)	Age	Cluster												
4	predicted cluster																
5	Customer 1	1	45	30	2												4
6	Customer 2	0	55	50	1												4
7	Customer 3	1	20	56	3												-
8	Customer 4		40	23	2												-
9	Customer 5		50	06	3												+
10	Customer 7	0	20	51	2												+
12	Customer 8	0	20	44	2												+
13	Customer 8	ň	20	44	2												1
14	Customer 10	1	35	28	2												1
15	Customer 11	1	45	30	2												1
16	Customer 12	1	35	28	2												
17	Customer 13	1	15	58	3												
18	Customer 14	0	50	28	2												
19	Customer 15	1	50	48	3												
20	Customer 16	1	50	34	1												
21	Customer 17	0	30	49	1												
22	Customer 18	1	50	30	2												
23	Customer 19	0	25	30	2												4
24	Customer 20	0	15	55	3												4
25	Customer 21	1	55	20	2												-
26	Customer 22		40	32	2												-
27	Customer 24		20	20	2												+
20	Customer 25	1	45	56	3												-
20	Customer 26		40	41	2												-
31	Customer 27	Ŭ Ő	20	31	2												1
32	Customer 28	1	25	33	2												1
33	Customer 29	1	50	60	3												1
34	Customer 30	1	30	23	2												
35	Customer 31	1	30	22	2												
36	Customer 32	0	45	32	2												
37	Customer 33	1	55	34	1												
38	Customer 34	1	45	30	2												
39	Customer 35	1	15	26	2												
40	Customer 36	1	35	34	2												4
41	Customer 37	1	70	31	1												
42	Customer 38	0	20	41	2												4
43	Customer 39		25	30	2											L	4
44	Customer 40		55	57	1												4
40	Customer 41		30	20	2						-	-	-		-		+
14 4	► ► Classified Data	A Sheet1	/ Sheet2 /	Sheet3 /										•			

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.