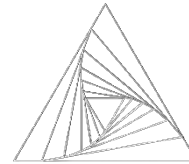# SegmentSolve

DECISION SUPPORT SCIENCES

User Manual

Using the Decision Support Sciences Segmentation Decision Support Tool to Develop Operational Segments That Yield Real-Word Outcomes (Profit, Sales, Growth and Retention)

Decision Support Sciences.   Better Science.   Better Solutions.

SegmentSolve

DECISION SUPPORT SCIENCES

Table of Contents

SegmentSolve

DECISION SUPPORT SCIENCES

I. **Getting Started**

A. Theory

    1. What is Market Segmentation?

    2. Definitions

SegmentSolve

DECISION SUPPORT SCIENCES

What is Market Segmentation?

Market Segmentation is the process of classifying a population into segments based on similarities in the data. Any data can be used to segment a population, but historically most data used has proven to be inactionable or disconnected to Real-World Outcomes (RWO's) like revenue, sales, retention and profit. For more on what data is best (including the seven critical psychometric dimensions see Decision Support Sciences website under Consulting-Segmentation.

Technically, SegmentSolve™ uses 13 variants of multivariate clustering to objectively uncover which clustering of the data maximizes separation between clusters and minimizes spread within a cluster. When the best clusters are determined, they can be analyzed in ProductSolve™ to evaluate product concepts that appeal to those clusters.
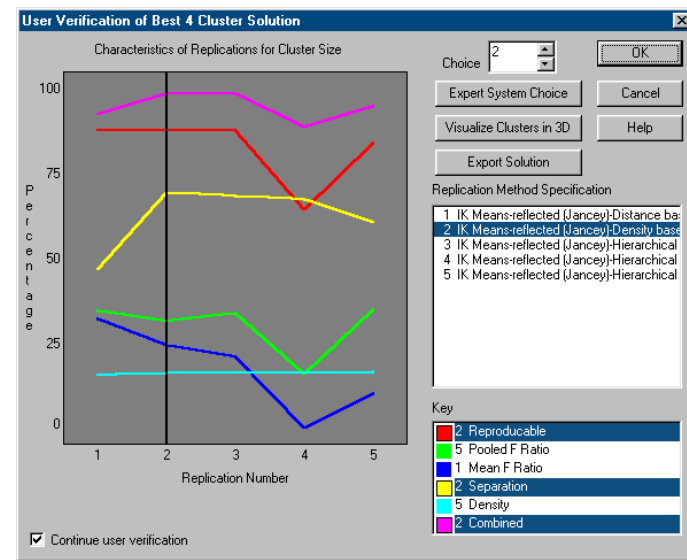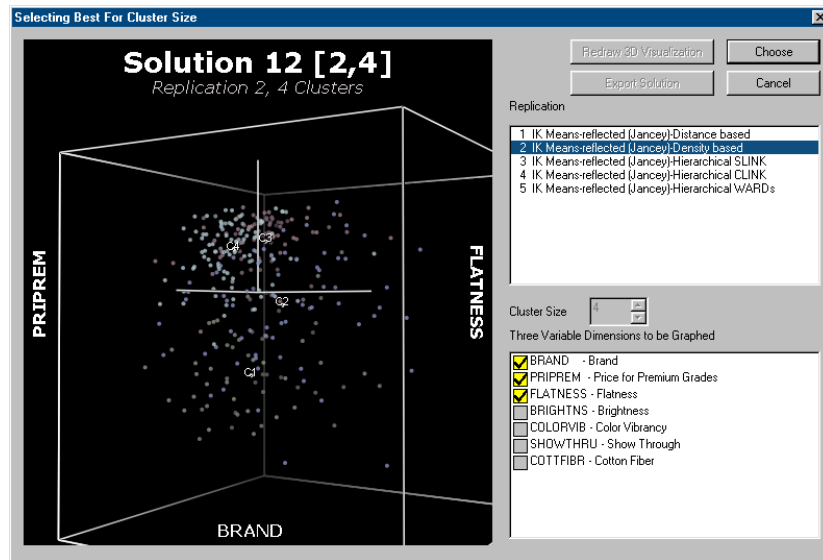
Definitions:

CLINK: Complete Linkage; A hierarchical agglomerative 'sorting strategy' that compares the case to be clustered with every point in a cluster and assigns the case to the cluster with the minimum total variance. This method tends to produce compact clusters composed of highly similar cases.

City Block: A type of distance measure, defined as $d_{ij}= \sum_{k=1} | x_{ik} - x_{jk} |$

Cluster: (as a method) The process of using a statistical method to categorize a collection of respondents into segments, or clusters.

Density: A property of a cluster that defines it as a relatively thick swarm of data points in a space when compared to other areas of the cluster space that may have comparatively few or no points. (Aldenderfer, 34)

Distance: A dissimilarity measure used with the iterative cluster methods.

Euclidean: A method of determining distance, defined as $d_{ij}= (\sum_{k=1}^{p}(x_{ik} - x_{jk})^2)^{1/2}$,which considers two cases to be more similar the smaller d_{ij} is.

Farthest Neighbor: A sorting strategy that compares the case to be classified with the farthest point from it in each cluster, and assigns it to the cluster with the lowest distance. This tends to produce tighter clusters than the Nearest Neighbor method.

Forgy (Simple): An Iterative K-means method that recalculates the centroid using the average of the points in the cluster.

Hierarchical Agglomerative: A clustering method that makes one pass through the data and agglomerates cases into non-overlapping clusters. A general form that describes linkage rules is:
$d(h,k) = A(i) \cdot d(h,i) + A(j) \cdot d(h,j) + B \cdot d(i,j) + C \cdot ABS(d(h,i) - d(h,j))$

Iterative K means (IK means): An iterative partitioning method which assigns cases to the cluster with the nearest centroid. This method handles large data sets better than the hierarchical agglomerative method because it does not require the calculation and storage of a matrix of similarity. There are two methods for using IK-Means: Forgy and Jancy.

SegmentSolve

DECISION SUPPORT SCIENCES

Definitions, con't:

Jancey (Reflected): Recalculates the centroid by doubling the difference between the old centroid and what would be the new Forgy centroid.

Nearest Neighbor: A sorting strategy that compares the case to be classified with the nearest point to it in each cluster, and assigns it to the cluster with the lowest distance.

Replication: A combination of a clustering algorithm and a starting solution, such as Hierarchical Agglomerative, Nearest Neighbor SLINK.

SLINK: Single Linkage; A hierarchical agglomerative method of clustering whose rule is that any case is joined to an existing cluster on the basis of the highest level of similarity of any one member of the existing group (only a single link is required). This method is the only one that is invariant to monotonic transformations and is unaffected by ties in the data, but has a tendency to form elongated clusters. (Jardine and Sibson, 1971)

Squared Euclidean: Denoted d_{ij}^2, this measure avoids the use of the square root by using the square of the Euclidean measure.

Standardize: the variables being analyzed are transformed to have a standard mean (0) and variance (1.0) to reduce the effect of the relative size of the variables.

Weighting: The manipulation of the value of a variable such that it plays a greater or lesser role in the measurement of similarity between two cases. (Williams, 1971)

Ward's method: A hierarchical agglomerative method that optimizes the minimum variance within clusters by joining cases that minimize the increase in the error sum of squares (ESS). This method tends to create clusters of relatively equal size. $ESS = x_i^2 - 1/n(\Sigma x_i)^2$ In conditions of cluster overlap, Ward's is the most effective method for minimizing the overlap.

SegmentSolve

DECISION SUPPORT SCIENCES

B. **Using SegmentSolve™**

       1. Installing SegmentSolve™
       2. Required Data Files

SegmentSolve
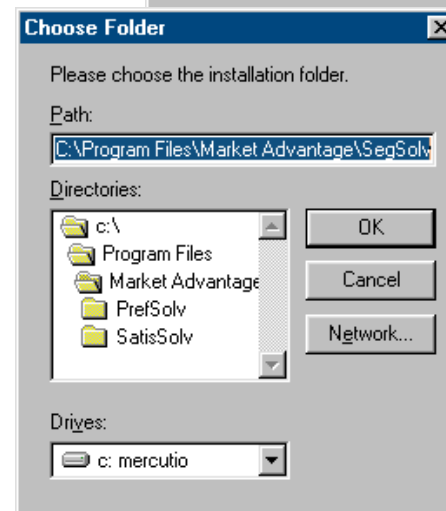
DECISION SUPPORT SCIENCES

Installing SegmentSolve^TM

System Requirements:

- SegmentSolve^TM runs on any modern computer that runs Window 7 or above.

- It is recommended that you have at least 10 GB of hard drive space, and 4 GB of RAM.

Installation Steps

- Whether you are installing from a CD or files already on your hard drive, imply run the file setup.exe.

- During installation, you may choose to which directory SegmentSolve^TM is installed and where you desire the shortcut to appear.

Choose Destination Location

Setup will install SegSolv in the following folder.

To install to this folder, click Next.

To install to a different folder, click Browse and select another folder.

You can choose not to install SegSolv by clicking Cancel to exit Setup.

Destination Folder

C:\...\Market Advantage\SegSolv   [Browse...]

[< Back] [Next >] [Cancel]

Choose Folder

Please choose the installation folder.

Path:

C:\Program Files\Market Advantage\SegSolv

Directories:

- c:\
- Program Files
- Market Advantage
- PrefSolv
- SatisSolv

[OK] [Cancel] [Network...]

Drives:

c: mercutio

SegmentSolve

DECISION SUPPORT SCIENCES

Required Data Files

There are 3 options you may use to input files into SegmentSolve™ for cluster analysis:

- SPSS file—A single SPSS file can be used for the clustering .

- SatisfactionSolve™ NUT file– A NUT file can be created in SatisfactionSolve™ that contains the importance or performance utilities for selected attributes. The NUT file stores the data, and a PRF file or TXT file that contains the variable labels and definitions must be used with the NUT file.

- TEXT file. An ASCII text file that contains a respondent number for each case and a value for each of the clustering variables can be used along with a text file that contains the variable names and labels. See pages 14-16 for instructions about the data file format and using the text import wizard.

SegmentSolve

DECISION SUPPORT SCIENCES

II.

Using the SegmentSolve™ Wizard
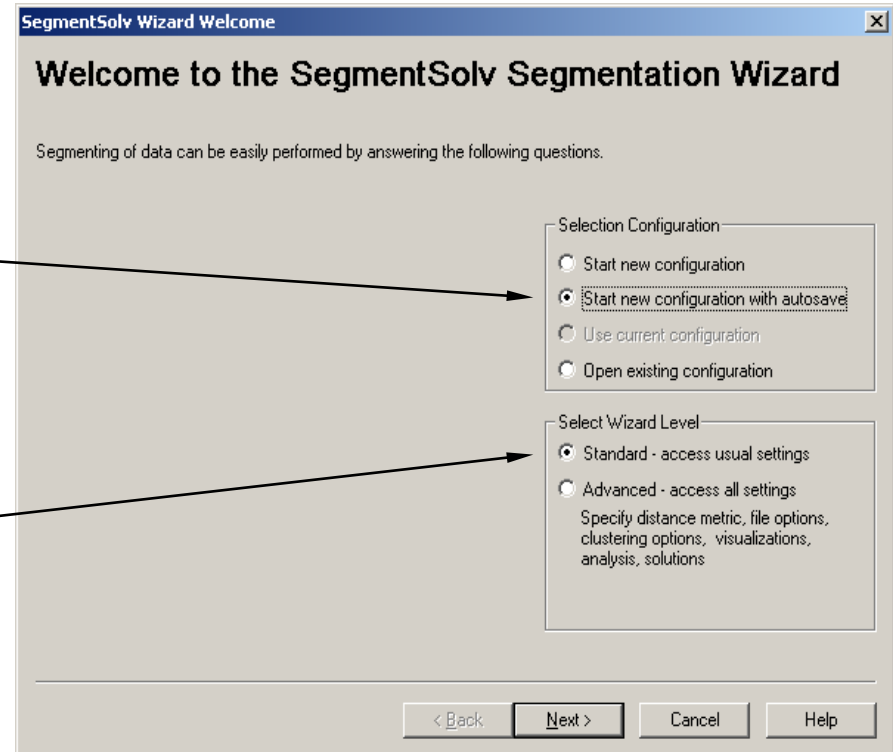
SegmentSolve

DECISION SUPPORT SCIENCES

II-A.

The Standard SegmentSolve^{TM} Wizard

SegmentSolve™ Wizard Step 1
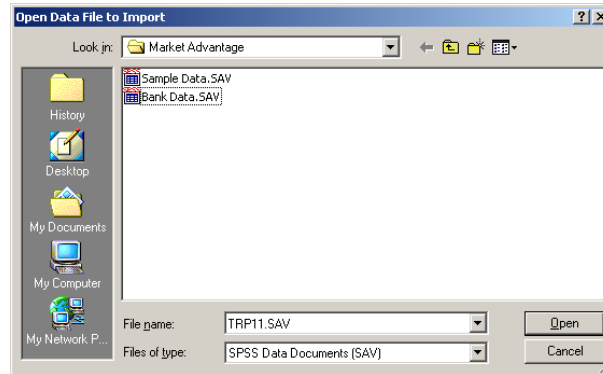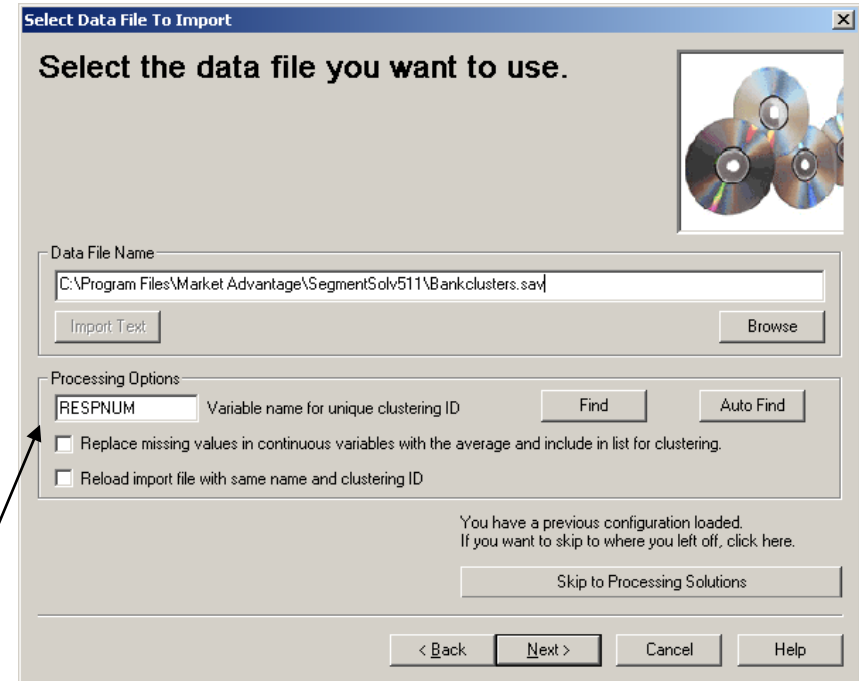
■ **Introduction**

- As you go through the wizard, all of the settings you choose can be saved in a SegmentSolve™ configuration file.

- If you have previously created a configuration file, it can be loaded at this step. Select the option to open an existing configuration, then select the Next button, and an Open File dialog will appear to let you browse for the file.

- The wizard can be used two ways:

  - The standard settings are used if you don't want to modify the default statistical settings for the cluster analysis. The standard level still allows the customization to select variables and filters for the analysis.

  - The advanced level of the wizard can be used if you want to change the data settings such as standardizing, centering, or changing which algorithms are used to determine the clusters.

SegmentSolv Wizard Welcome

Welcome to the SegmentSolv Segmentation Wizard

Segmenting of data can be easily performed by answering the following questions.

Selection Configuration
- ○ Start new configuration
- ⊙ Start new configuration with autosave
- ○ Use current configuration
- ○ Open existing configuration

Select Wizard Level
- ⊙ Standard - access usual settings
- ○ Advanced - access all settings
  - Specify distance metric, file options, clustering options, visualizations, analysis, solutions

< Back   Next >   Cancel   Help

SegmentSolve

DECISION SUPPORT SCIENCES

Standard SegmentSolve™ Wizard Step 2

■ **Select the Data File**

- From this dialog, the data file can be selected. Use the Browse button to find the directory and file to use for the analysis. The file type is selected in the Open File dialog.

- If the clustering data is in a text file, select 'comma delimited text file' or 'delimited text' as the file type. Select 'Open' from the Open file dialog, then select 'Import Text' under the filename. The text import wizard will automatically follow to convert the file to from a text file to an SPSS data file. See the next page for instructions on the text import wizard.

- There must be at least one variable that has a different value for each record in the data. In the Processing Options box, select the Find button to display all the variables select the variable to use for the clustering ID.

- This same variable will be used to export the clustering data back to the original data file.

Select Data File To Import

Select the data file you want to use.

Data File Name
C:\Program Files\Market Advantage\SegmentSolv511\Bankclusters.sav

Import Text   Browse

Processing Options
RESPNUM   Variable name for unique clustering ID   Find   Auto Find
☐ Replace missing values in continuous variables with the average and include in list for clustering.
☐ Reload import file with same name and clustering ID

You have a previous configuration loaded.
If you want to skip to where you left off, click here.

Skip to Processing Solutions

< Back   Next >   Cancel   Help

Open Data File to Import

Look in: Market Advantage

Sample Data.SAV
Bank Data.SAV

History
Desktop
My Documents
My Computer
My Network P...

File name: TRP11.SAV   Open
Files of type: SPSS Data Documents (SAV)   Cancel

SegmentSolve

DECISION SUPPORT SCIENCES

Standard SegmentSolve^{TM} Wizard Step 2: Import Text Wizard

■ Text File Format

- In order to be read by SegmentSolve^{TM}, a text data file must have a specific format. First, the file must be delimited in some way (e.g. with commas, tabs, spaces) that separates each data field. The sample data file on the right is delimited by commas.

- The data file also must have all of the variable names, labels, and formats at the top of the file. There should be one line at the top of the file for each variable. For each variable there must be a variable name, variable label, and either an N or S to designate a numeric or string variable. Those three fields must be specified for all variables.

- The variable level values and labels can also be specified after the format. There can be a text qualifier, such as quotation marks, but that is not necessary. The same delimiter used for the data should be used for the variable information at the top of the file. (See the example above for age and income).

- There must be one blank line after the variable information and before the data to indicate the number of variables.


```
sample cluster XLS.csv - WordPad

File  Edit  View  Insert  Format  Help

RESPNUM,Respondent Number,N
AGE,Age,N,1,Less than 35 Years,2,35-44 Years,3,45-54 Years,4,55-64 Years,5,
INCOME,Income,N,1,Less than $49,999,2,$50,000 to $99,999,3,$100,000 to $199
MFUND,Have Mutual Fund Account,N
PRIMPROV,Primary Provider,N
INVVX,Value of Investment Accounts,N
MFUNDVX,Value of Mutual Funds,N
VX401K,Value of Retirement Accounts,N
ASSETVX,Total Investable Assets,N
TENUREDB,Tenure at Provider,N

1000,3,6,1,40,1250000.00,1250000.00,750000.00,3250000.00,1.95
1001,2,3,1,13,175000.00,375000.00,175000.00,725000.00,1.51
1003,4,3,1,23,375000.00,175000.00,30000.00,580000.00,1.80
1005,2,2,1,37,343320.00,501129.44,229183.10,1073632.54,1.81
1007,4,2,1,37,30000.00,750000.00,75000.00,855000.00,.33
1008,3,3,1,40,343320.00,175000.00,229183.10,747503.10,1.73
1009,5,1,1,40,375000.00,375000.00,229183.10,979183.10,.99
1010,3,4,1,37,75000.00,175000.00,30000.00,280000.00,1.59
1012,2,7,1,37,30000.00,30000.00,175000.00,235000.00,1.58
1013,2,3,1,37,5000.00,30000.00,30000.00,65000.00,999.00
1014,5,3,1,13,175000.00,375000.00,75000.00,625000.00,1.68
1015,3,1,1,37,343320.00,501129.44,229183.10,1073632.54,1.12
1017,3,4,1,9,175000.00,175000.00,750000.00,1100000.00,1.68

For Help, press F1
```

Standard SegmentSolve™ Wizard Step 2: Import Text Wizard

- **Import Text Wizard**
  - After a text filename has been specified in the second step of the SegmentSolve™ wizard, SegmentSolve™ reads in the text file, when the 'Import Text' option is selected.

  - SegmentSolve™ attempts to determine the delimiter in the text file, and applies variable names to each of the data fields based on the information in the text file. If there is not a standard delimiter for the data, select the 'Other' delimiter option and specify which character separates the data in the box to the right of the Other category.

  - The first twenty cases of the data file can be previewed in the step using the vertical and horizontal scroll bars. Check the data to make sure that the variable names match the data fields and that the data has been imported correctly.

  - Select Next to import all the records in the data file into SegmentSolve™. If any data is found that does not match the variable definitions, SegmentSolve™ will describe the problems and present options for solutions.

Text Import Wizard - Define Delimiter

Indicate the delimiter and text qualifier used in your data. The preview pane shows how the variable names and the first few lines of data will be imported.

Delimiter: ◉ Comma ○ Semicolon ○ Space ○ Tab ○ Other: [ ]

Text Qualifier: [▼]

Preview

RESPNUM	AGE	INCOME	MFUND	PRIMPROV	INVVX	MFL
1000	3	6	1	40	1250000.00	125l
1001	2	3	1	13	175000.00	375l
1003	4	3	1	23	375000.00	175l
1005	2	2	1	37	343320.00	501l
1007	4	2	1	37	30000.00	750l
1008	3	3	1	40	343320.00	175l
1009	5	1	1	40	375000.00	375l
1010	3	4	1	37	75000.00	175l
1012	2	7	1	37	30000.00	300l
1013	2	3	1	37	5000.00	300l
1014	5	3	1	13	175000.00	375l
1015	3	1	1	37	343320.00	501l

Invalid Data - Numeric Variables with String Data

The following variables were marked as being numeric, but contained data in one or more cases that was non-numeric (string).
- Select "Continue" to automatically replace invalid data with the system missing value.
- Select "Retry" to adjust your file delimiter and text qualifier selections.
- Select "Exit Wizard" to exit the text import wizard and fix the source file before continuing.

Below is a list of the variables that had invalid data, along with the first case number and line number in which it appeared for that variable.
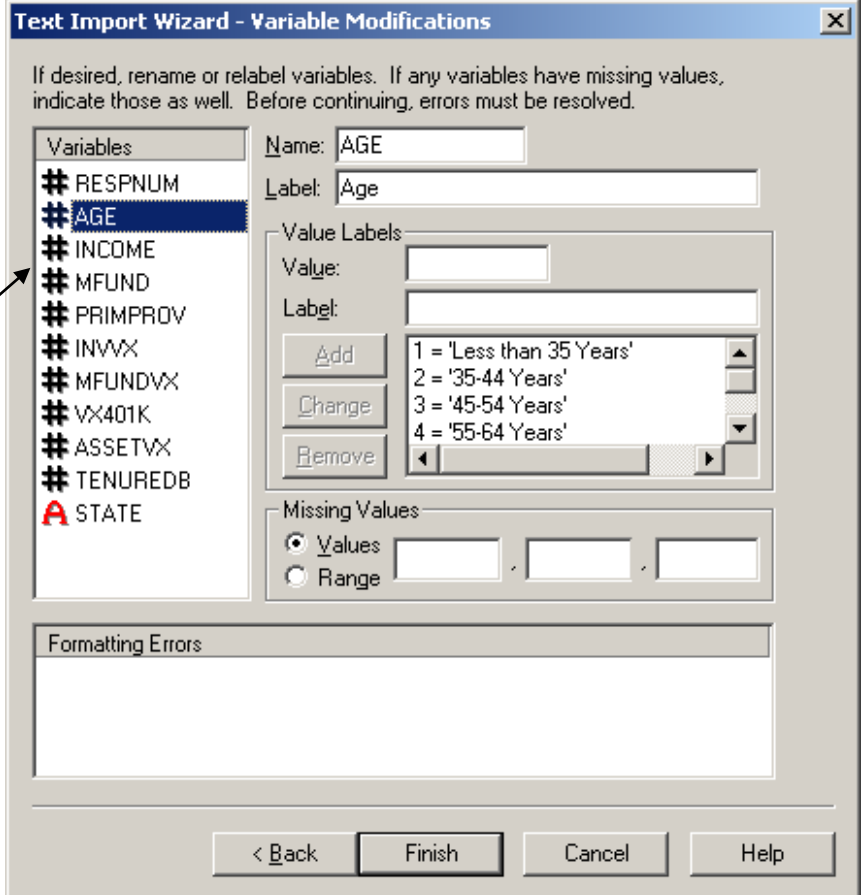
Field # (Variable Name)	Case #	Line #
3 (MFUND)	17	28

[Continue]   [Retry]   [Exit Wizard]

SegmentSolve

DECISION SUPPORT SCIENCES

Standard SegmentSolve™ Wizard Step 2: Import Text Wizard

■ **Variable Modification**

- The second step of the text import wizard allows the user to modify the settings for any of the variables. Highlight a variable in the left Variables box to view or modify the values or labels or that variable.

- The # or A symbol to the left of the variable name indicates whether the variable is a numeric or string variable. Numeric variables cannot have any letters as values.

- Missing values can be specified in the missing values range. Up to three specific values can be specified using the 'values' option (i.e. 0,999,9999). A range and distinct value can be specified using the 'range' option (i.e. 0-10,999).

- Select the Finish button to import the data and return to the second step of the SegmentSolve™ wizard

Text Import Wizard - Variable Modifications

If desired, rename or relabel variables. If any variables have missing values, indicate those as well. Before continuing, errors must be resolved.

Variables:
- # RESPNUM
- # AGE
- # INCOME
- # MFUND
- # PRIMPROV
- # INVVX
- # MFUNDVX
- # VX401K
- # ASSETVX
- # TENUREDB
- A STATE

Name: AGE
Label: Age

Value Labels
Value:
Label:

[Add] [Change] [Remove]

1 = 'Less than 35 Years'
2 = '35-44 Years'
3 = '45-54 Years'
4 = '55-64 Years'

Missing Values
○ Values
○ Range

Formatting Errors

[< Back] [Finish] [Cancel] [Help]

SegmentSolve

DECISION SUPPORT SCIENCES

Standard SegmentSolve™ Wizard Step 2

■ **Select the Clustering ID**
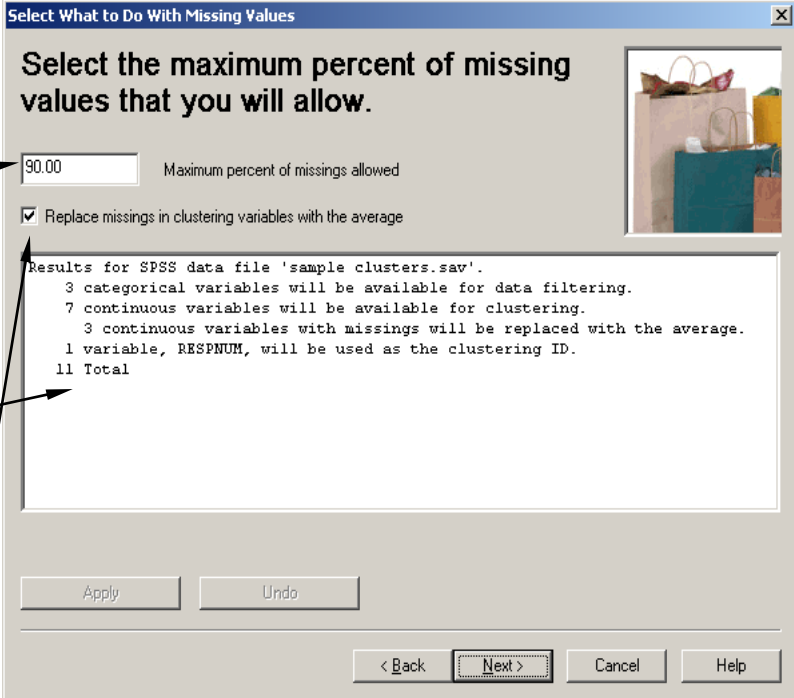
- One variable must be selected as the clustering ID that has a unique value for each record in the data. This dialog displays all the variables in the selected data file, and recommends which variables would be appropriate for a clustering ID.

- Use the mouse to select a variable in the left box, and a message at the bottom states whether that variable has missing values for any of the cases.

- To verify that the selected variable has a unique value for each record, use the 'Verify Unique' button. When it determines whether the variable is unique, the result will be described in the message at the bottom of the dialog.

- In the right text box, the value of the variable for each record is shown, in file order. Use the scroll bar on the right to view all the cases. This preview of the data is only shown if the box "Previewing 448 values" is checked.

- When the variable for clustering ID has been determined, select the "Use" button at the top.

Find Variable for Clustering ID

Select cluster ID from list of 1321 variables.   Use   Cancel   Help

☑ Previewing 448 values

RESPNUM - Respondent Number	1000
RESPNUM1 - RESPNUM1	1001
RESPNUM2 - RESPNUM2	1003
FRSTHALF - FRSTHALF	1005
SCNDHALF - SCNDHALF	1007
STATUS$ - STATUS$	1008
TIME$ - Survey Time	1009
TIMEANS$ - TIMEANS$	1010
SECUR1 - Employment Security	1012
DECMKR - Primary Decision Maker	1013
PROV01 - Inv Prov: AIM	1014
PROV02 - Inv Prov: Alliance Capital	1015
PROV03 - Inv Prov: American Century	1017
PROV04 - Inv Prov: American Express	1019
PROV05 - Inv Prov: American Funds	1020
PROV06 - Inv Prov: Bank of America	1021
PROV07 - Inv Prov: Brown & Company	1022
PROV08 - Inv Prov: Charles Schwab	1024
PROV09 - Inv Prov: Dreyfus	
PROV10 - Inv Prov: E-Trade	
PROV11 - Inv Prov: Evergreen (First Union	

RESPNUM1 has no missings, so it might make a good clustering ID.
You should verify that all its values are unique.

Verify Unique

SegmentSolve

DECISION SUPPORT SCIENCES

SegmentSolve™ Wizard Step 3

■ SPSS Options

- The "Maximum percent of missing allowed" refers to the maximum percentage of cases that can be missing in order for any variables to be included in the analysis. To change this value, type in a new number in the text box, and click on the Apply button.

- The number of categorical and continuous variables included in or excluded from the analysis with the current maximum percentage of missing's will be displayed in the large text box. The categorical variables are those that are used for filtering or selecting portions of the data. The continues variables are those that will be used for clustering.

- The option 'Replace missing's in clustering variables with the average' must be checked to include variables with missing values in the analysis. Any case that has a missing value will be replaced with the mean. The default setting is for this option to be checked. If it is not checked, then variables are only included if none of the cases have missing values. When this option is checked on or off, the apply button must selected for the change to take affect.

Select What to Do With Missing Values

Select the maximum percent of missing values that you will allow.

`90.00` Maximum percent of missings allowed

☑ Replace missings in clustering variables with the average

```
Results for SPSS data file 'sample clusters.sav'.
    3 categorical variables will be available for data filtering.
    7 continuous variables will be available for clustering.
    3 continuous variables with missings will be replaced with the average.
    1 variable, RESPNUM, will be used as the clustering ID.
   11 Total
```

Apply   Undo

< Back   Next >   Cancel   Help

SegmentSolve

DECISION SUPPORT SCIENCES

SegmentSolve™ Wizard Step 4

■ Clustering Variables

- The continuous variables available for the analysis are displayed on this dialog. Variables with a checkmark to the left will be included. Click on the checkmark if you want to 'turn off' a variable.

- Use the Preview Variables option to view the data values for all cases in the data file. The cases are shown in file order, with the clustering ID shown first and the value of the highlighted variable in the second column.

- Select Advanced Data Import Options to bring up a dialog where standardizing, centering, and weighting options can be set. The default setting is for the data NOT to be standardized or centered; it is recommended that the user does not change those settings unless he/she understands how those processes change the data.

- Weighting the data refers only to the filtering variables, and not to the cluster development process.

SegmentSolve

DECISION SUPPORT SCIENCES

SegmentSolve™ Wizard Step 5

■ Select Subgroups

- All categorical variables from the data file will be shown in the left box of this dialog. To include or exclude certain cases, simply select or deselect variable levels. When a variable does not have all of its levels selected, there will be a blue checkmark by the variable name (for example, AGE).

- If the option to compute level frequencies is checked, the frequencies are shown in brackets by the variable levels. Once that option is checked, frequencies will be shown for all variables.

- Levels from multiple variables may be selected or de-selected, but cases that are in ANY deselected group will be excluded. In other words, only the intersection of the selected levels from all variables will be used in the cluster analysis.

- The advanced data import options available here are the same as the previous wizard step.

SegmentSolve

DECISION SUPPORT SCIENCES
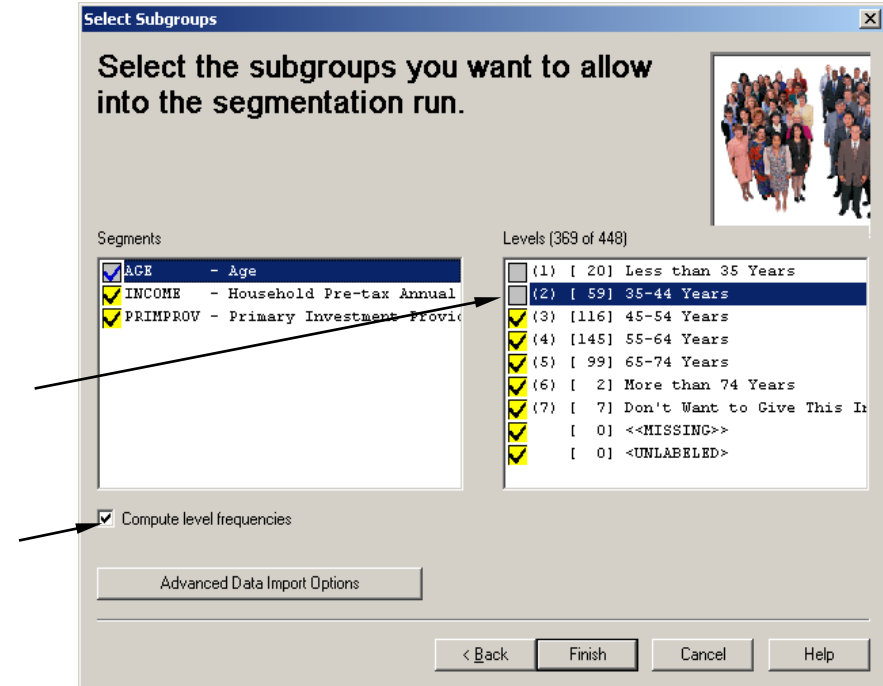
II-B.
The Advanced SegmentSolve™ Wizard

SegmentSolve™ Wizard Step 1

■ Introduction

- As you go through the wizard, all of the settings you choose can be saved in a SegmentSolve™ configuration file.

- If you have previously created a configuration file, it can be loaded here. Select the option to open an existing configuration, then select the Next button, and an Open File dialog will appear.

- The wizard can be used in two modes.

  - The standard settings are used if you don't want to modify the default statistical settings. The standard mode still allows the customization to select variables and filters for the analysis.

  - The advanced mode provides additional options such as data standardizing and centering. It also allows you to change which algorithms are used to determine the clusters. If the Advanced mode is selected, an option appears to customize which steps the wizard will show.

SegmentSolv Wizard Welcome

Welcome to the SegmentSolv Segmentation Wizard

Segmenting of data can be easily performed by answering the following questions.

Selection Configuration
- ○ Start new configuration
- ● Start new configuration with autosave
- ○ Use current configuration
- ○ Open existing configuration

Select Wizard Level
- ○ Standard - access usual settings
- ● Advanced - access all settings
  Specify distance metric, file options, clustering options, visualizations, analysis, solutions
  [Customize]

[< Back] [Next >] [Cancel] [Help]

SegmentSolve

DECISION SUPPORT SCIENCES

SegmentSolve™ Step 1 Wizard

■ Introduction

- When the Customize option is selected on the first screen of the wizard, this dialog appears to allow some screens of the wizard to be omitted.

- Use the mouse to select or deselect wizard pages. The default is for all wizard pages to be shown.

- The box on the right of the dialog indicates the function of the selected wizard pages and describes the circumstances where the page should be included. Click on a page to have the description appear in the box.

- Pages that must always be included in the analysis cannot be deselected. When you have made your final choices, select the OK button.

Select Pages To Be Shown

Select pages to be included in the wizard.
Defaults will be used for all unchecked pages.

OK   Cancel   Help

Description: Select options for handling missing data

- Select initial wizard options
- Select distance metric
- Select data file to import
- Select options for handling missing data
- Select clustering variables
- Select data subgroups
- Select clustering methods
- Select clustering visualization
- Select analysis options
- View and modify expert system choices
- Select solutions to report and export

Include this page to select the variables to be included when computing the clustering solutions.

The default is to use all variables that qualify.

Since usually you want to limit the number of variables, you probably should include this page.

SegmentSolve

DECISION SUPPORT SCIENCES

SegmentSolve™ Wizard Step 2

■ **Clustering Metric**

- Choose the distance measure for the cluster analysis.

- Squared Euclidean is the default setting.

- See the definition pages at the beginning of the manual for definitions of each distance metric.

SegmentSolve

DECISION SUPPORT SCIENCES
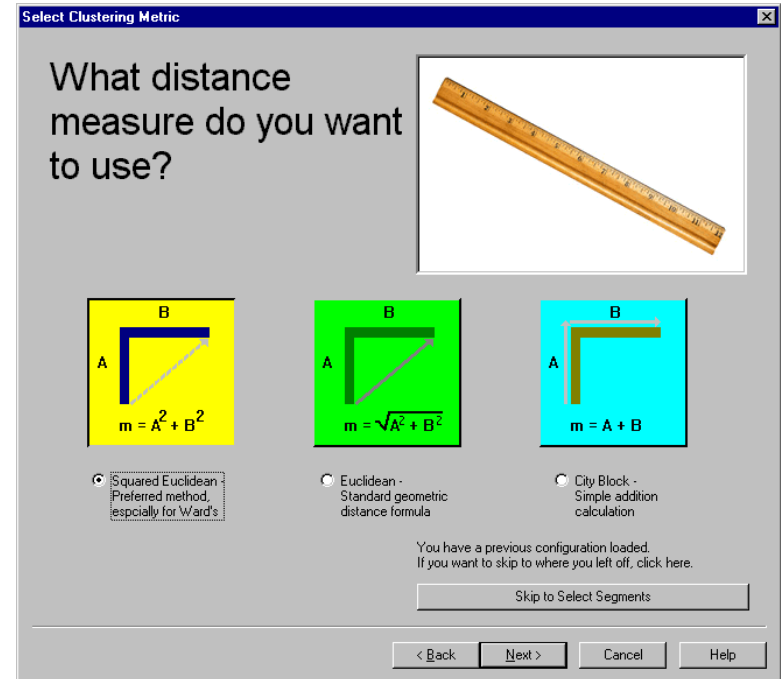
Advanced SegmentSolve™ Wizard Step 3

■ Select the Data File

- Select the data file from this dialog. Use the Browse button to find the directory and file to use for the analysis.

- If the clustering data is in a text file, select 'comma delimited text file' or 'delimited text' as the file type. Select 'Open' from the Open file dialog, then select 'Import Text' under the filename. See pages 14 to 16 for instructions on the text import wizard.

- There must be at least one variable that has a different value for each record in the data. In the Processing Options box, select the Find button to display all the variables. Select the variable to use for the clustering ID.

- This same variable will be used to export the clustering data back to the original data file.

- If there are any missing values in the data, SegmentSolve™ can replace them with the mean for that variable. Do so by checking the option to replace missing values.

Select Data File To Import

Select the data file you want to use.

Data File Name
C:\Market Advantage\sample cluster.csv

Import Text   Browse

Processing Options
Variable name for unique clustering ID   Find   Auto Find
☑ Replace missing values in continuous variables with the average and include in list for clustering.

Open Data File to Import

Look in: Market Advantage

Sample Data.SAV
Bank Data.SAV

File name: TRP11.SAV   Open
Files of type: SPSS Data Documents (SAV)   Cancel

SegmentSolve

DECISION SUPPORT SCIENCES

Advanced SegmentSolve™ Wizard Step 3

■ Select the Clustering ID

- One variable must be selected as the clustering ID. It unique value for each record in the data. This dialog displays all the variables in the selected data file, and recommends which variables would be appropriate for a clustering ID.

- Use the mouse to select a variable in the left box, and a message at the bottom states whether that variable has missing values for any of the cases.

- To verify that the selected variable has a unique value for each record, use the 'Verify Unique' button. When it determines whether the variable is unique, the result will be described in the message at the bottom of the dialog.

- In the right text box, the value of the variable for each record is shown, in file order. Use the scroll bar on the right to view all the cases. This preview of the data is only shown if the box "Previewing 448 values" is checked.

- When the variable for clustering ID has been determined, select the "Use" button at the top.

Find Variable for Clustering ID

Select cluster ID from list of 1321 variables.   [Use] [Cancel] [Help]

```
RESPNUM   - Respondent Number
RESPNUM1  - RESPNUM1
RESPNUM2  - RESPNUM2
FRSTHALF  - FRSTHALF
SCNDHALF  - SCNDHALF
STATUS$   - STATUS$
TIME$     - Survey Time
TIMEANS$  - TIMEANS$
SECUR1    - Employment Security
DECMKR    - Primary Decision Maker
PROV01    - Inv Prov: AIM
PROV02    - Inv Prov: Alliance Capital
PROV03    - Inv Prov: American Century
PROV04    - Inv Prov: American Express
PROV05    - Inv Prov: American Funds
PROV06    - Inv Prov: Bank of America
PROV07    - Inv Prov: Brown & Company
PROV08    - Inv Prov: Charles Schwab
PROV09    - Inv Prov: Dreyfus
PROV10    - Inv Prov: E-Trade
PROV11    - Inv Prov: Evergreen (First Union
```

☑ Previewing 448 values

```
1000
1001
1003
1005
1007
1008
1009
1010
1012
1013
1014
1015
1017
1019
1020
1021
1022
1024
```

RESPNUM1 has no missings, so it might make a good clustering ID. You should verify that all its values are unique.   [Verify Unique]

SegmentSolve

DECISION SUPPORT SCIENCES

Advanced SegmentSolve™ Wizard Step 4

■ SPSS Options

- The "Maximum percent of missing allowed" refers to the maximum percentage of cases that can be missing in order for any variables to be included in the analysis. To change this value, type in a new number in the text box, and click on the Apply button.

- The number of categorical and continuous variables **included** in or **excluded** from the analysis with the current maximum percentage of missing's will be displayed in the large text box. (Categorical variables are those are used to select portions of the data..

- The option 'Replace missing's in clustering variables with the average' must be checked to include any variables with missing values in the analysis. Cases with missing variable values will be replaced with the mean. If it is not checked, then variables are only included if none of the cases have missing values.

Select What to Do With Missing Values

Select the maximum percent of missing values that you will allow.

`90.00`   Maximum percent of missings allowed

☑ Replace missings in clustering variables with the average

```
Results for SPSS data file 'sample clusters.sav'.
    3 categorical variables will be available for data filtering.
    7 continuous variables will be available for clustering.
    3 continuous variables with missings will be replaced with the average.
    1 variable, RESPNUM, will be used as the clustering ID.
   11 Total
```

Apply   Undo

< Back   Next >   Cancel   Help

SegmentSolve

DECISION SUPPORT SCIENCES

SegmentSolve™ Wizard Step 5

■ Clustering Variables

- Continuous variables are displayed on this dialog. Variables with a checkmark to the left will be included. Use the mouse to click on the checkmark if you want to 'turn off' a variable.

- Use the Preview Variables option to view the data values for all cases in the data file. Use the mouse to check the option on or off. The cases are shown in file order, with the clustering ID shown first and the value of the highlighted variable in the second column.

- Select the "Advanced Data Import Option" to standardize, center, and weight options can be set. The default setting is for the data NOT to be standardized or centered. Do not change those settings unless you fully understand these data normalization options.

Select Variables

Select the variables you want to use to determine the segments.

Utilizing 7 of 7 Variables Available for Analysis

- MFUND   - MFUND
- INVVX   - Investments-Non MF Value - Discrete
- MFUNDVX - Mutual Fund Value - Discrete
- VX401K   - Pension or 401K Value - Discrete
- ASSETVX - Total Investable Assets - Value
- BALVPV37 - Balance Value: T. Rowe Price
- TENUREDB - TENUREDB

☑ Preview variables

RESPNUM	MFUND
1000	1
1001	1
1003	1
1005	1
1007	1
1008	1
1009	1
1010	1

Advanced Data Import Options

Advanced Data Options

Standardizing
- ◉ None
- ○ N-1 Standard Deviatic
- ○ N Standard Deviation
- ○ Arithmetic   Set Min/Max

Scaling
- ◉ None
- ○ Center
- ○ Remove All Effects

Weighting
- ◉ None
- ○ Weighted   Set Weighting

Set to Defaults

OK   Cancel   Help

SegmentSolve

DECISION SUPPORT SCIENCES

SegmentSolve™ Wizard Step 6

■ Select Subgroups

- All categorical variables from the data file appear in this wizard step. The variable name is on the left box and its levels on the right. By default, all levels are selected. To exclude one level from the analysis, click on the yellow box to deselect it.

- When a variable does not have all of its levels selected, there will be a blue checkmark by the variable name (for example, AGE).

- If the option to compute level frequencies is checked, the frequencies are shown in brackets by the variable levels. Once that option is checked, frequencies will be shown for all variables.

- Levels from multiple variables may be selected or de-selected. When this is done, only the intersection of the selected levels from all variables will be used in the cluster analysis.

- The advanced data import options available from this dialog are the same as those available from the previous wizard step.
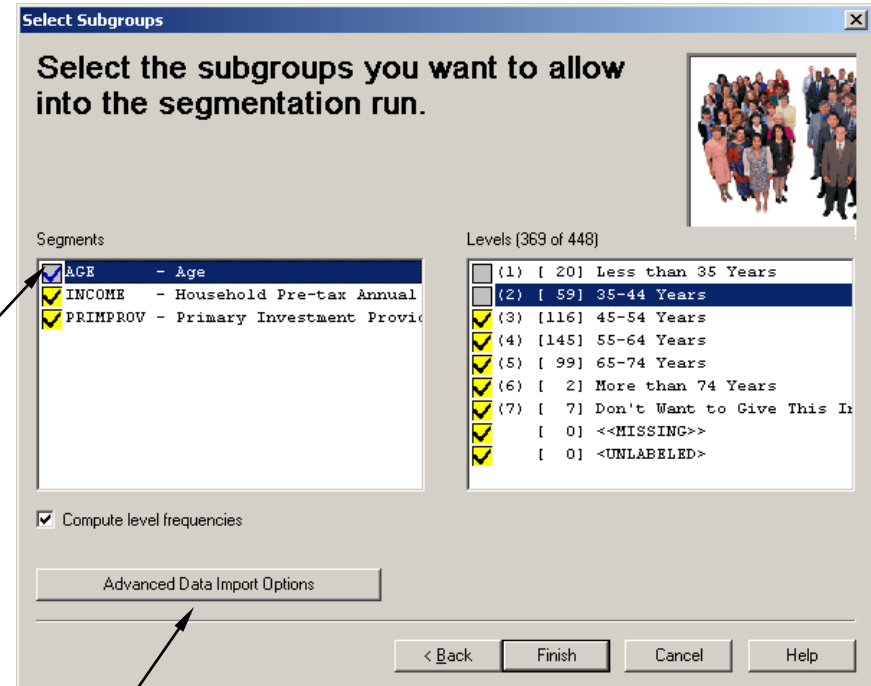
SegmentSolve™ Wizard Step 7

■ **Clustering Options**

- All of the clustering methods used for the analysis are selected at this step. There are 13 possible methods. One or more methods can be included multiple times. Some methods use a random seed, and thus may yield different cluster solutions if run multiple times.

- To add a method, highlight a method and specification, and click the Add button. To add all of the possible methods, use the Add All button. See the definitions at the beginning of the manual for information about each method. For the best clusters, it is recommended that all the combinations of methods and specifications are used to develop the clusters.

- The Advanced Clustering Options button leads to the lower dialog. Here, the minimum and maximum cluster size can be changed, along with the sample size and iterations.

- A set of clustering methods can be saved for future use by clicking the Save button in the advanced options dialog. This file can be loaded in subsequent runs by selecting Importing Segmentations .

SegmentSolve

DECISION SUPPORT SCIENCES

SegmentSolve™ Wizard Step 8

■ **Clustering Visualization**

- The clustering process can be viewed using a detailed real-time status bar or a standard status bar. Use the radio buttons to select the option you want.

- The real-time display shows at the top of the dialog the current method and replication that is being run, as well as the current iteration and cluster size. The time elapsed and the time remaining are shown on the right side of the status display.

- The circular pictures of people show relative sizes of the clusters currently being generated.

- The analysis will run slightly faster using the standard status bar.
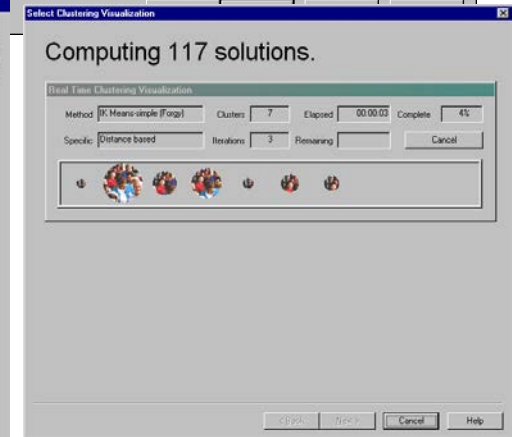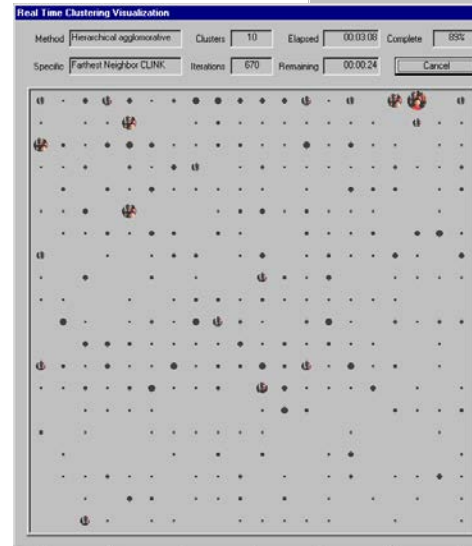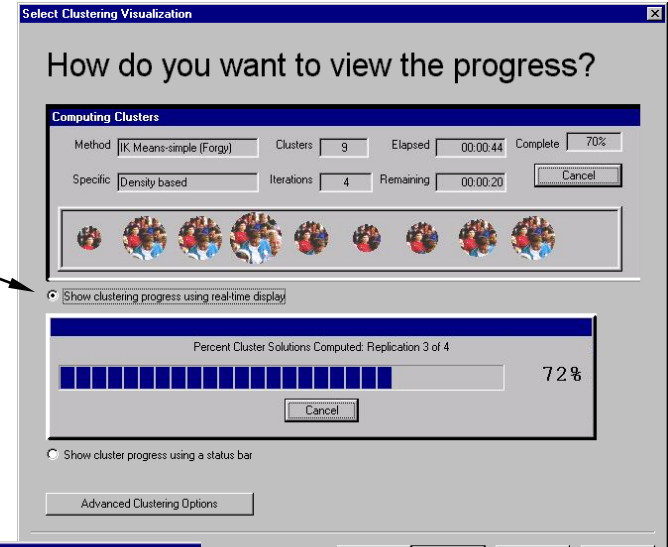
SegmentSolve™ Wizard Step 9

■ Analysis Options

- Of all the solutions generated, SegmentSolve™ chooses one from each cluster size as the "expert system choice" for that cluster size.

- Choosing the first option will cause SegmentSolve™ to skip to the processing solutions step where reports are created.

- Selecting the second option will allows the user to evaluate only the expert system choices based on several criteria before any reports are generated. The manual review of all cluster solutions is skipped.

- The last option allows the user to manually review all the cluster solutions produced, and change the choices that are marked as the 'expert system choices'.

- The minimum refers to the minimum number of cases that can be in a cluster in order for SegmentSolve™ to consider it a valid cluster in the analysis. If percentage is selected, then each cluster must contain at least the specified minimum percentage of respondents.

Select Expert System Options

Select whether you want to review how the Expert System chose the best solutions.

SegmentSolv is now ready to choose a best solution for cluster sizes 2 through 10 with a minimum cluster size of 1, and then a best overall solution from the 13 replications.

○ Use expert system choices for best solutions without reviewing them
○ Review expert system choice for best over all solution for all cluster sizes
◉ Review all expert system choices for best solutions

Minimum Cluster Size
```
1.00
```
◉ Size is expressed as a raw score
○ Size is expressed as a percentage

Currently no cluster will be ignored due to size when computing a solution's statistics.

	A
1	
2	Clusters
3	2
4	
5	List of
6	Number

< Back   Next >   Cancel   Help

SegmentSolve

DECISION SUPPORT SCIENCES

SegmentSolve™ Wizard Step 10

- **Verifying Best Replication for Each Cluster Size**

  - This optional step allows the user to see all of the cluster solutions evaluated on critical several criteria. There is one dialog for each cluster size — the example on the right shows results for all of the 5 cluster solutions.

  - Each line on the chart represents one criteria. (e.g. the blue line shows the pooled significance for each replication.) In the Key box, the replication with the highest score for each criteria is shown.

  - To see the number of respondents in each cluster, select the Show Allocations option, and a dialog will appear that shows the allocations for each replication.

  - The clusters can be named and edited at this step. To do so, open that option, select Edit Solution and Cluster Labels.

  - To save a specific replication for reporting or exporting, highlight it then select the Record as Significant button.

SegmentSolve

DECISION SUPPORT SCIENCES

SegmentSolve™ Wizard Step 11

■ Report and Export

- Reports can be created to show cluster differences and display the means for each cluster. By default, these reports can be viewed in SegmentSolve™ and saved as an Excel file.

- Solutions are also exported to an ASCII data file using the clustering ID specified in the second step of the SegmentSolve™ wizard. By default, an SPSS syntax file is created to import the cluster solutions into the original data file.

- First, select the report, export, and recommendation options on the right of the dialog.

- The left box designates which solutions are included in the report or exported data file. Any cluster solutions that were marked as significant in the user verification of the clusters is included in the 'Best solutions for each of the x cluster groups' options.

- To specify which options are included in the report select the 'Advanced Solution Processing' option.
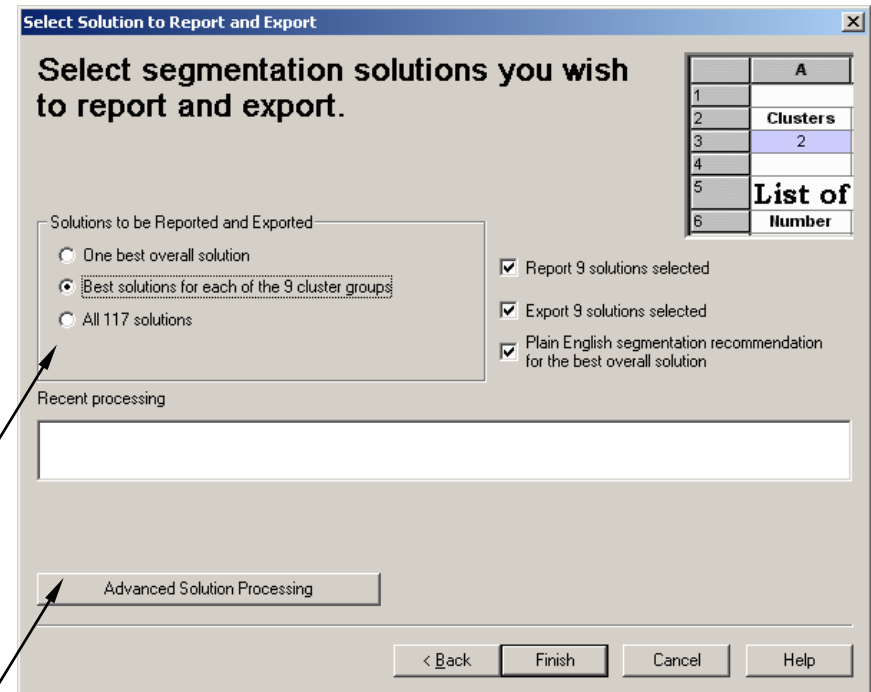
SegmentSolve™ Wizard Step 11

■ Advanced Solution Processing

- This dialog appears when the Advanced Solution Processing option is selected. Use this option to customize the solutions and formatting in the reports.

- The left box displays all of the cluster solutions. Click on a solution to deselect it from the current report, or highlight the solution and use the Remove button to remove it from the list of possible solutions to be included in the reports.

- To add solutions to those to be included in the reports and export files, select a cluster size and replication from the boxes at the right. Multiple replications can be selected for one cluster size. Once the sizes and replications have been selected use the 'Add Marked Solutions' button to add them. They will then appear in the included solutions at the left.

- Use the bottom 'Processing' box to change the file name or location of the report, export or recommendation files. The default location is the directory where SegmentSolve™ was installed.

SegmentSolve

DECISION SUPPORT SCIENCES
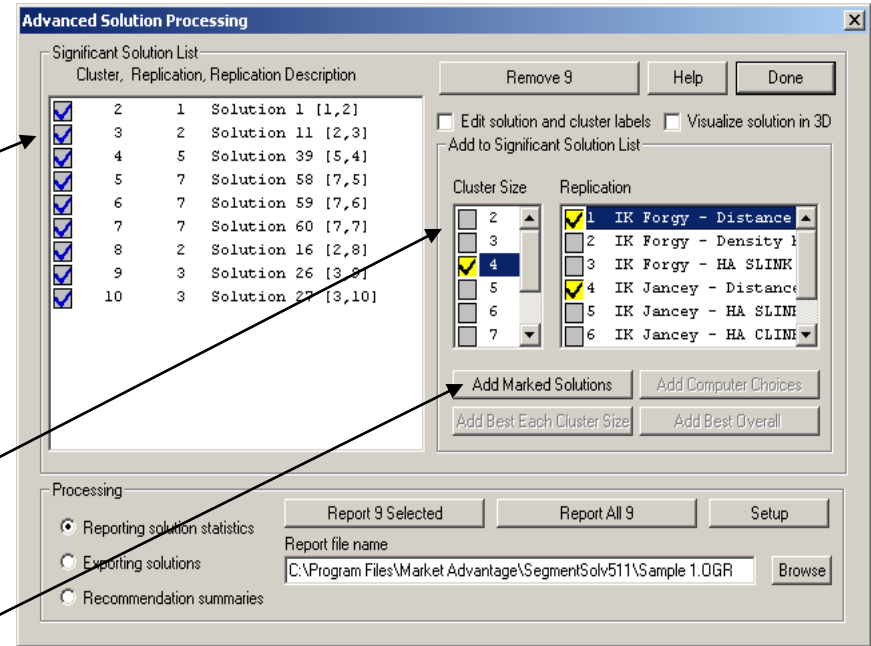
B. Practice

1. Identifying the Best Overall Market Segmentation
2. Viewing Results with Tables and Reports

SegmentSolve

DECISION SUPPORT SCIENCES

Example 1: Identifying the Best Overall Market Segmentation

Example Question:

- What is the "natural segmentation" of the marketplace? Which customers value similar things about a product or service, and how many of those groups are there?

Example Input Data:

- Importance or performance data collected using conjoint analysis

Steps:

- Define Variables
- Define Styles
- Select Method
- Select Replications
- Execute Cluster Analysis
- Create Reports

Results:

- Cluster identification that can be seen in report or graph form
- Identification of best solution within a given cluster set
- Identification of the overall best cluster solution

SegmentSolve

DECISION SUPPORT SCIENCES

Example 1: Identifying The Best Overall Market Segmentation

- **Open Decision Support Sciences Model**

  - Click on Start - Programs – SegmentSolve™

- **Setup Clustering**

  - Use the wizard to select the clustering data file, attributes, and filters.

  - If you want the data to be standardized, centered, or weighted, select those options from the Advanced Data Import Options dialog.

SegmentSolve

DECISION SUPPORT SCIENCES

Example 1: Identifying The Best Overall Market Segmentation

■ Analyze Reports

- One replication for each cluster solution is selected as the best solution, and an overall best solution is selected. (SegmentSolve™ selects a best solution based on the high scores in the criteria listed in the Key combined with a weighting scheme listed under Edit Preferences.)

- In the User Verification dialog box, the user has the opportunity to modify the solution that is designated as the best solution. The value displayed in the choice box is the best solution as defined above. When OK is selected to go to the next cluster solution, the current value in the Choice box is stored as the best solution.

- The best solution can be determined and selected in a number of ways:

  - If the level of the statistical results are important to you, click the mouse on the graph in the User Verification dialog box above the replication number. Use the Key to evaluate which replications have performed better by certain methods. When you click on the graph, the black vertical line will jump to the replication number closest to the mouse click. In the Key box, the statistical methods which that replication scored highest in will be highlighted.

  - If you are interested in a specific statistical measure, use the mouse to click on that measure in the Key box. The replication that scored highest in that criteria will be highlighted, the vertical bar on the graph will move to that number, and the number will appear in the Choice box.
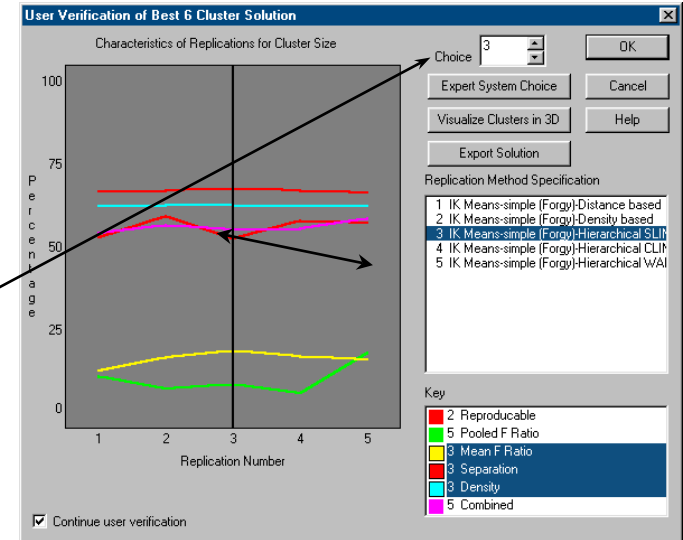
  - A third way to select the best solution is by selecting from the Replication Method box. The value in the choice box will change according to the Replication method that is highlighted.

  - The final way to determine the best solution is from the Visualize Clusters option, which is explained on the following page.

SegmentSolve

DECISION SUPPORT SCIENCES

Example 1: Identifying The Best Overall Market Segmentation

■ Visualize Cluster Solutions

- From the User Verification dialog box, select the Visualize Clusters in 3D button. When this is done, a three dimensional visualization od the data appears in the Graph dialog box. See the appendix for rotation and zoom control instructions. To change the variables, check or uncheck the variables, and select the Redraw 3D Visualization button.

- All of the replications being analyzed are listed in the Replications box, and the one currently graphed is highlighted. To view clusters using a different replication method, select a replication and click on the Redraw 3D Visualization button.

- To designate one replication as the best solution, select the Choose button. SegmentSolve[TM] will return to the User Verification dialog box, and the solution selected in the Visualization dialog will be selected. The best solution can still be modified here by highlighting a different replication. Press OK when the desired replication appears in the Choose box. Then you may move on to visualize the next cluster solution.

- See the graphing section in the appendix for other graph control options (e.g. changing the text properties.

SegmentSolve

DECISION SUPPORT SCIENCES
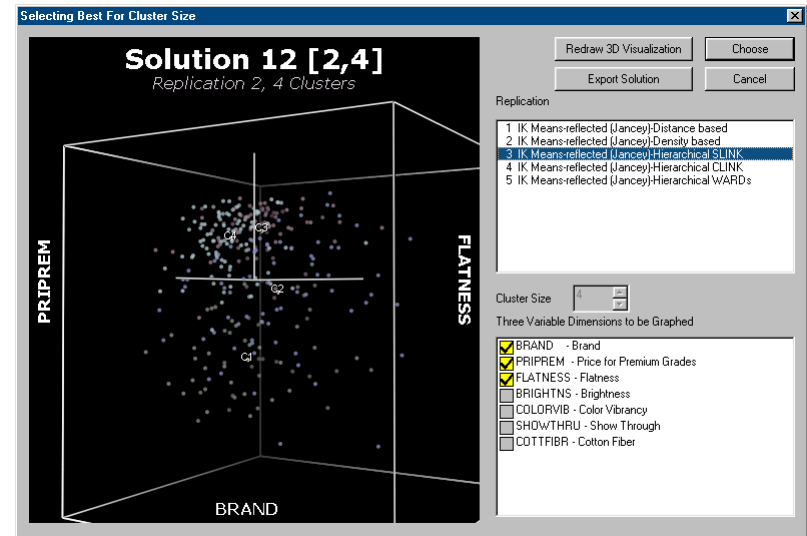
Example 1: Identifying The Best Overall Market Segmentation

■ **Best Overall Cluster Solution**

- When all the cluster solutions have been computed, SegmentSolve™ presents the replications that were verified by the user to be the best for each cluster size.

- In the box that lists the replications, the first number on each line refers to the cluster size.

- Select the Expert System Choice button to highlight the overall best replication.
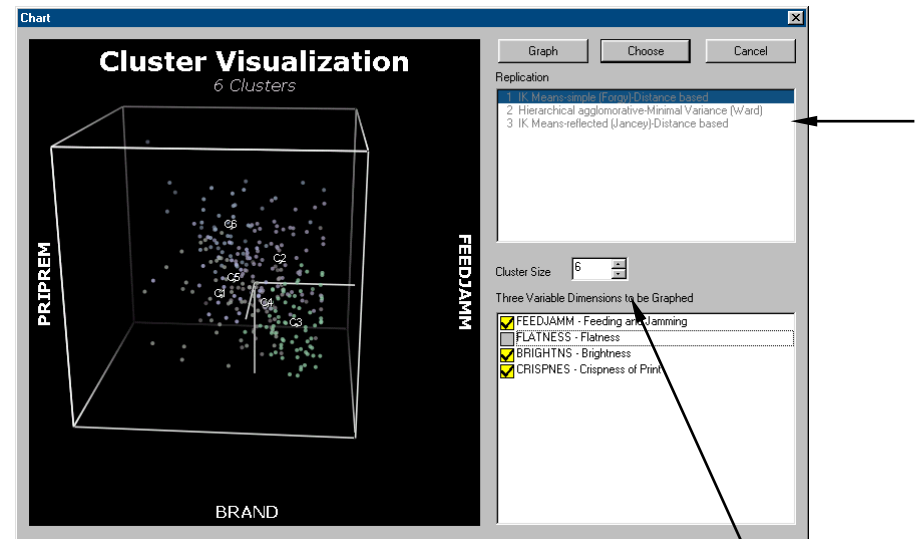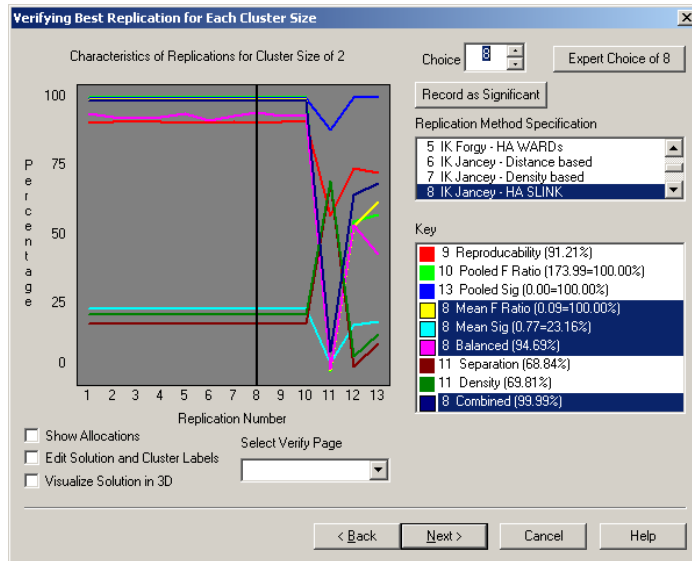
■ **Visualize Cluster Solutions**

- Select the Visualize Clusters in 3D button from the User Verification dialog box,

- When viewing the overall best clusters, the Cluster Size box is active, and any cluster solution can be reached by scrolling through the cluster solutions.

- Only the replication which produced the best solution is viewable for each cluster solution. The selected replication will be highlighted.

- To change the selected variables, check or uncheck the variables, and select the Graph button. Exactly 3 variables must be selected.

SegmentSolve

DECISION SUPPORT SCIENCES

Example 2: Viewing Results with Reports and Tables

■ Create Reports

• In the last step of the wizard, select which solutions you want included in the report or export file. Reports will automatically be generated when you select the 'finish' button.

• The default choice is to report and export the 'Best solutions for each of the cluster groups'. If the user verification process was used for evaluating the clusters, then all of the clusters marked as significant by the user will be included in the report in addition to the best solution for each cluster size determined by SegmentSolve™'s expert verification system. If the user verification process was not used, only the best solution from each cluster size will be included in the report or export files.

• The report can be saved as an Excel file and opened later in either Excel or SegmentSolve™.
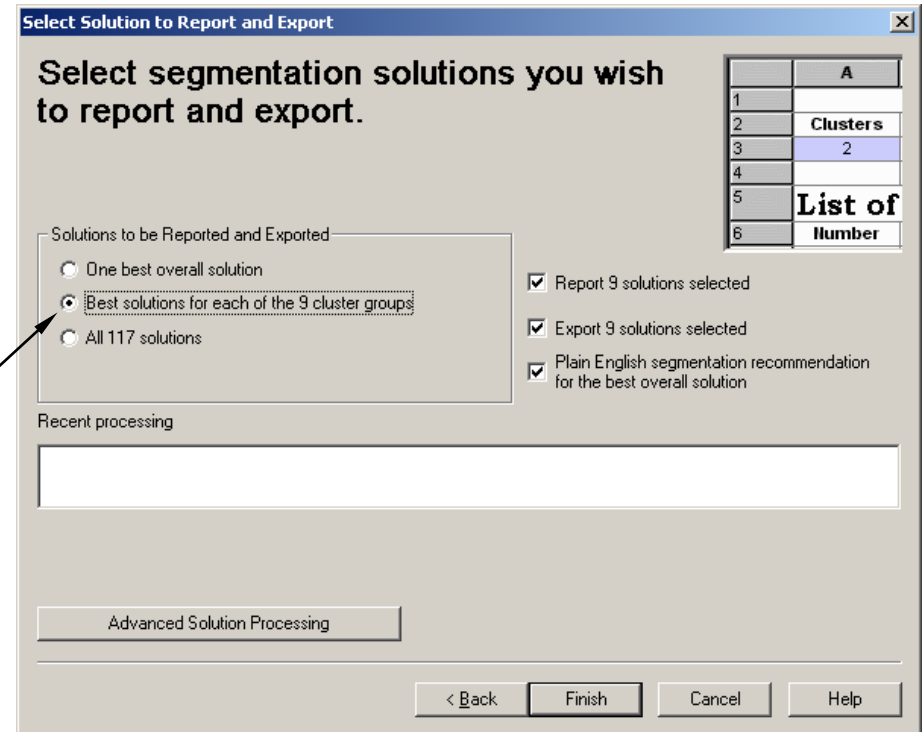
Example 2: Viewing Results with Reports and Tables

■ Create Tables

- At the beginning of the report, general information about that run are listed (e.g. number of variables special treatment of the data, such as centering or weighting, etc.)

- A list of all the replications and specifications is listed at the top of the report.

- The analysis for each cluster solution and replication combination is shown in a separate section.

- The table is broken into major categories by cluster solution, and then each cluster section is divided by the selected replications. At the beginning of each cluster size section, there is a comparison of the replications by various criteria for that cluster size.

- This example shows the cluster means results for the first replication of the 3 cluster solution.

- Specific means and deviations for each variable within each cluster is also given in the 'Variables Sorted as Deviations of Cluster's Means' section.

SegmentSolve

**DECISION SUPPORT SCIENCES**

IV.   Appendix - Reference Manual

(dialog boxes that are not in the Clustering Wizard are described here)

SegmentSolve

DECISION SUPPORT SCIENCES

1. File: Save Configuration

- A configuration file stores the 'state' of the SegmentSolve™ settings. To Create a configuration file, select File, New, Configuration before you start the analysis.
- To save a configuration file, enter a name for the file in the 'File name' box, then select
- The default name for configuration files is untitled.ssc, but this default name can be modified in the Edit Preferences menu.
- Configuration files can be opened using the File Open Configuration file option, or by selecting one of the recently used configuration files from the bottom of the File menu.
- Browse through available directories using the up button and selecting desired directories in the same way that a file would be saved in most standard Windows applications.

- If the Cancel button is selected, SegmentSolve™ will exit this dialog box and cancel any changes that were made.

SegmentSolve

DECISION SUPPORT SCIENCES
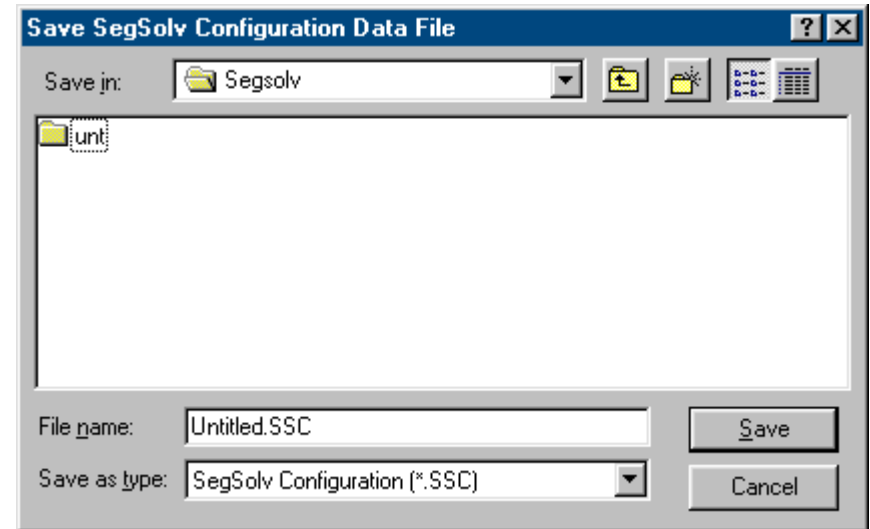
2. **Edit: Preferences**

- Select Preferences from the Edit menu brings up this dialog box.
- On the **Graph** tab, the colors to be used can be customized by highlighting a color with the mouse, then specifying the hue mixture and selecting Apply.
- Select the **Reports** tab to change the colors that are used in the titles and highlighting of the reports. Customize the colors using the same process as that for graphs. (See above)
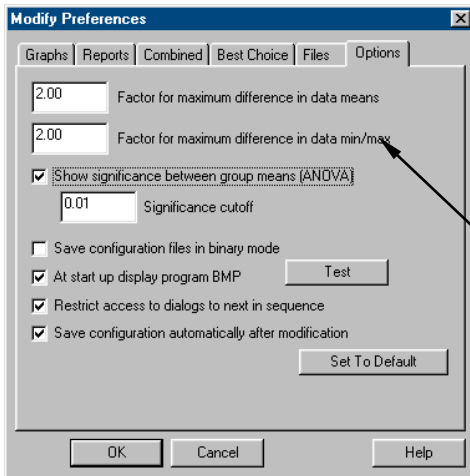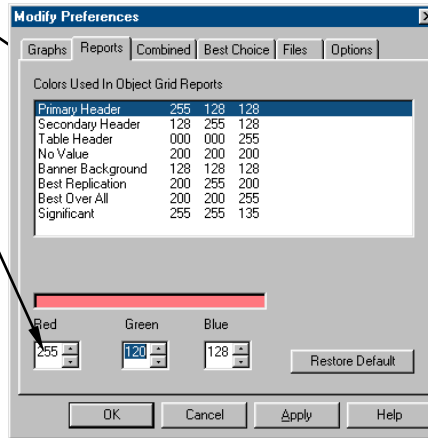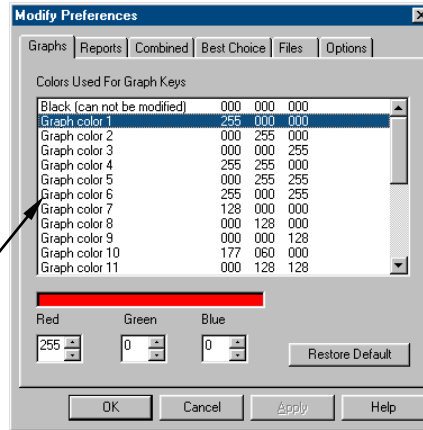
- On the **Options** tab, some of the statistical options can be viewed and modified. For numerical values, use the mouse to click in the text box to change the value. For other options, use the mouse to check or uncheck the options.

- Under the **Combined** tab, the current weightings are listed. When the cluster solutions are being evaluated in the User Verification dialog box and in the reports, the solutions (a specific replication and number of clusters) are given a score based on a combination of these analysis methods. To change any of the weights, use the mouse to select the desired method, then change the value in the Weighting box, and select Apply. The Apply button is only accessible when a new value has been entered in the Weighting box. The sum of all the weights should be 100.

SegmentSolve

DECISION SUPPORT SCIENCES
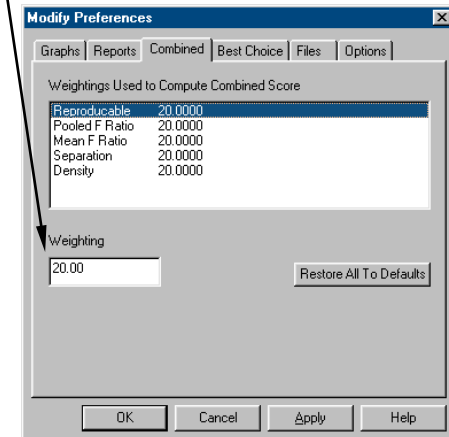
3. Segment: Cluster Options

- When Clusters are selected as the analysis method, Cluster options will be available under the Select menu. The replications for the analysis are activated here, and cluster size options can be set.
- To activate a new replication, select a cluster method by highlighting it with the mouse, then highlight a starting solution from the Specifications box, and select Add.
- The active replications appear in the large text box in the lower part of the dialog box.
- To delete a replication, highlight the title in the lower box, and select Delete.
- **Size Options:** To change the current options, use the mouse to click in the box corresponding to the option to change the value.
- The size options do not apply to all replications. For example, the Maximum Iterations option applies only to the IK Means methods.
- The minimum and maximum number of clusters must be less than the sample size of respondents.
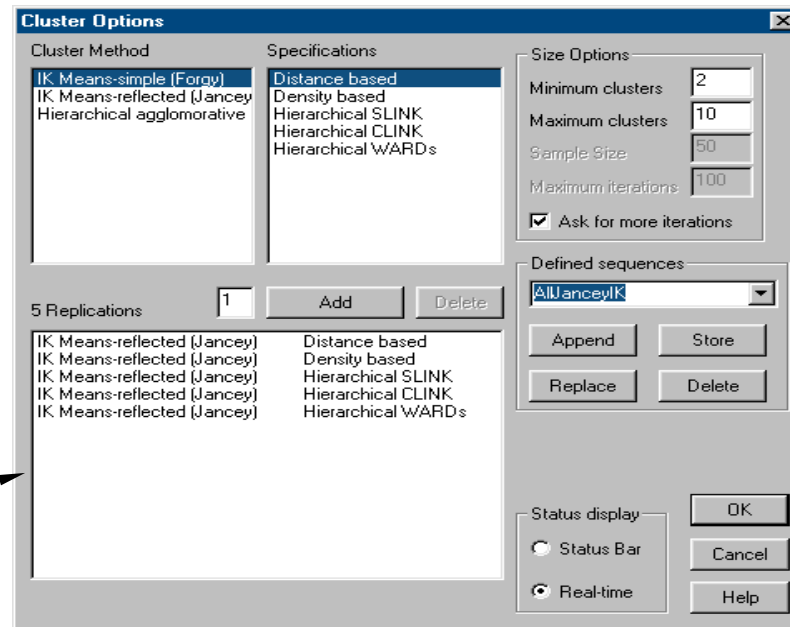
- The OK button saves the groups that has been formed and sets the active replications as those to be used in analysis.
- If the cancel button is selected, all changes that have been made will be discarded.

- **Defined Sequences**: This feature has several replications stored under one sequence name, which is useful if the same set of replications will be used multiple times.
- To use a defined sequence, click on the down arrow, and select the sequence name. Then select the Append button to add the replications in the sequence to any replications already selected.
- To define a new sequence, make sure that all of the active replications are to be included in the new sequence. Type in a sequence name in the text box under Defined sequences, then select the Store button.
- Defined sequences are stored with the configuration file, so in order use a non-default sequence repeatedly, save the configuration file before exiting SegmentSolve[TM].

SegmentSolve

DECISION SUPPORT SCIENCES

4. Report: Edit Cluster Names

- This dialog box is an option under the Report menu.
- To change a solution name, use the mouse to click in the Solution Name box and edit the text.
- To change the Cluster Name, highlight the cluster so that the cursor appears on the cluster name, then type in the new text.
- Use the scroll arrows to modify the Solution or Cluster names for a different combination of Replication and Cluster Size.
- Usually, the default settings are used here, but changes can be made for efficiency or preference.
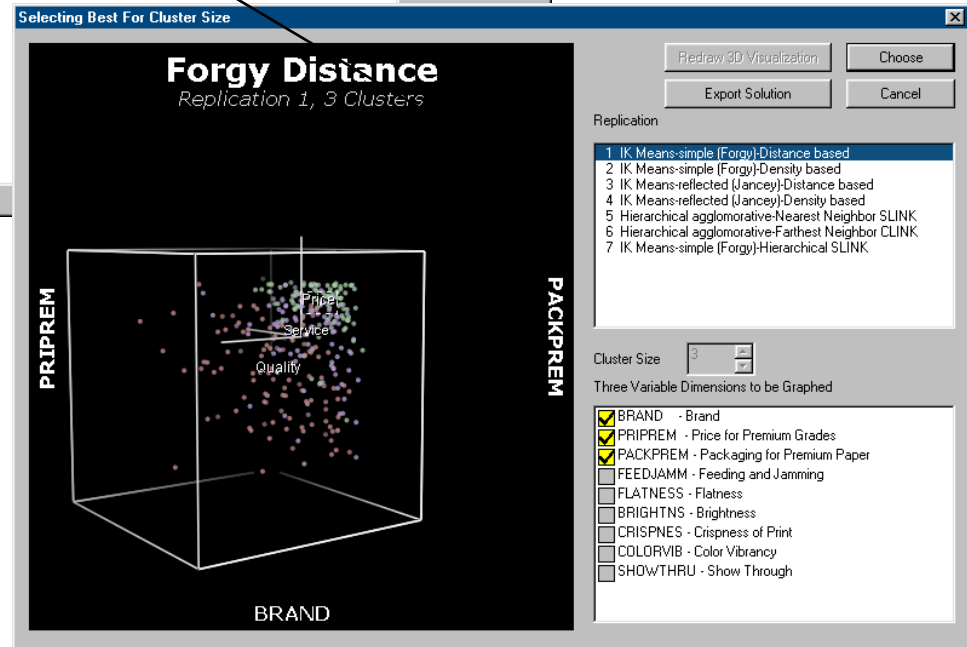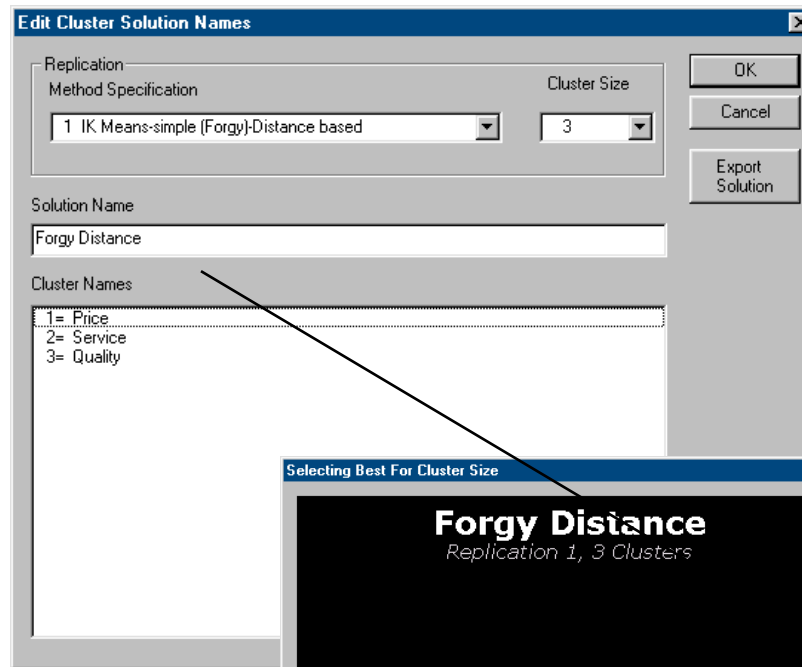
- The OK button saves the settings that has been chosen and exits the dialog box.
- If the Cancel button is selected, SegmentSolve™ will exit this dialog box and cancel any selections that were made.

Edit Cluster Solution Names

Replication
Method Specification
1 IK Means-simple (Forgy)-Distance based

Cluster Size
3

OK
Cancel
Export Solution

Solution Name
Forgy Distance

Cluster Names
1= Price
2= Service
3= Quality

Selecting Best For Cluster Size

Forgy Distance
Replication 1, 3 Clusters

PRIPREM

PACKPREM

Price
Service
Quality

BRAND

Redraw 3D Visualization
Choose
Export Solution
Cancel

Replication
1 IK Means-simple (Forgy)-Distance based
2 IK Means-simple (Forgy)-Density based
3 IK Means-reflected (Jancey)-Distance based
4 IK Means-reflected (Jancey)-Density based
5 Hierarchical agglomorative-Nearest Neighbor SLINK
6 Hierarchical agglomorative-Farthest Neighbor CLINK
7 IK Means-simple (Forgy)-Hierarchical SLINK

Cluster Size 3
Three Variable Dimensions to be Graphed

☑ BRAND - Brand
☑ PRIPREM - Price for Premium Grades
☑ PACKPREM - Packaging for Premium Paper
☐ FEEDJAMM - Feeding and Jamming
☐ FLATNESS - Flatness
☐ BRIGHTNS - Brightness
☐ CRISPNES - Crispness of Print
☐ COLORVIB - Color Vibrancy
☐ SHOWTHRU - Show Through

SegmentSolve

DECISION SUPPORT SCIENCES

Using Charts: Movement

- There is powerful movement control over all 3-D visualization created in SegmentSolve[TM].
- Moving the mouse vertically or using the up and down arrows will rotate the chart around a horizontal axis. Likewise, moving the mouse horizontally or using the left and right arrows will rotate the chart around a vertical axis.
- To zoom into the clusters, hold down the Control key and move the mouse vertically. To change the centering of the clusters, hold down the Shift key and move the mouse to the desired position.

- Clicking on the right mouse button anywhere in the background of the chart will bring up a menu. Selecting 'Properties' will bring up the Chart Properties dialog box. The Geometry tab can be used to set precise specifications on position, and the extent of the rotation.

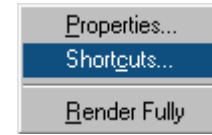# SegmentSolve

DECISION SUPPORT SCIENCES
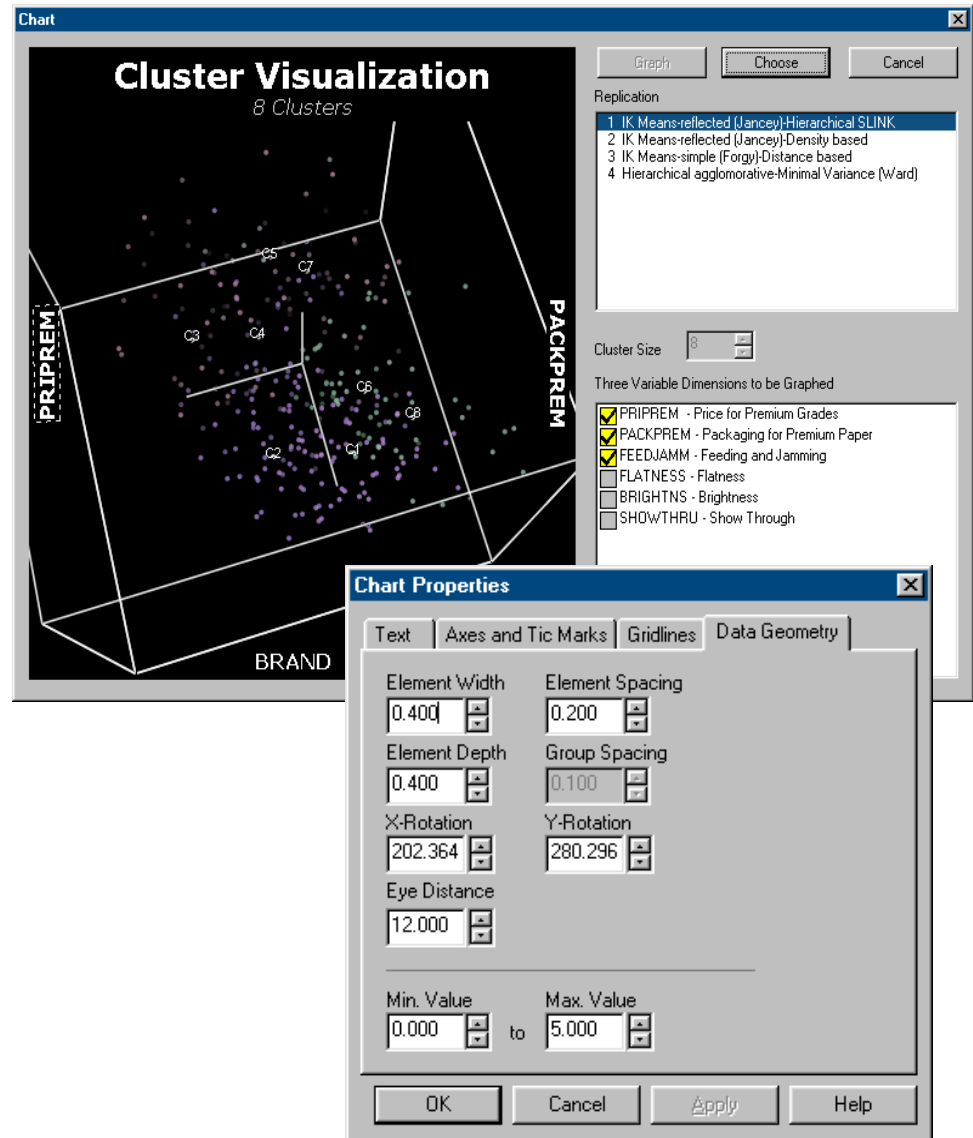
Using Charts: Text Properties

- Using the right mouse button anywhere on the chart screen will bring up the this menu.
- Selecting Properties will lead to a dialog box with the capability to change most of the chart properties that were determined before the chart was displayed, and some additional.
- All of the text on the chart can be changed. Use the text tab here to change the words that are displayed. The check-boxes on each line are used to elect whether that title is displayed.
- The font, color, size, and position of all titles on the chart can be modified. Select the text with the left mouse button. Doing so will highlight the text with a dashed box.
- Use the right mouse button on any selected text to bring up the text menu. The font and size can be modified by the Change Font feature. The Change Angle feature is not yet supported.
- When the text is selected, the position can be changed by dragging the title to the desired location. The titles can not be moved to a different side of the chart from their original location.
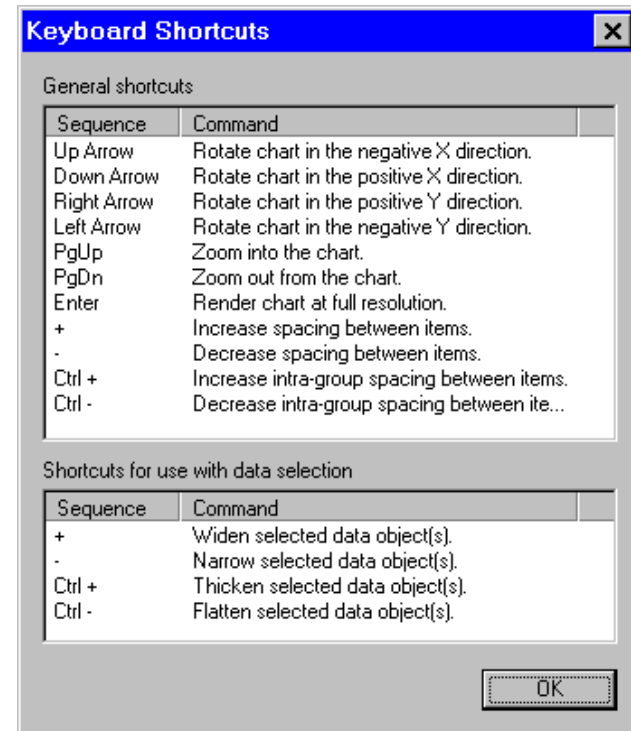

SegmentSolve

DECISION SUPPORT SCIENCES

51
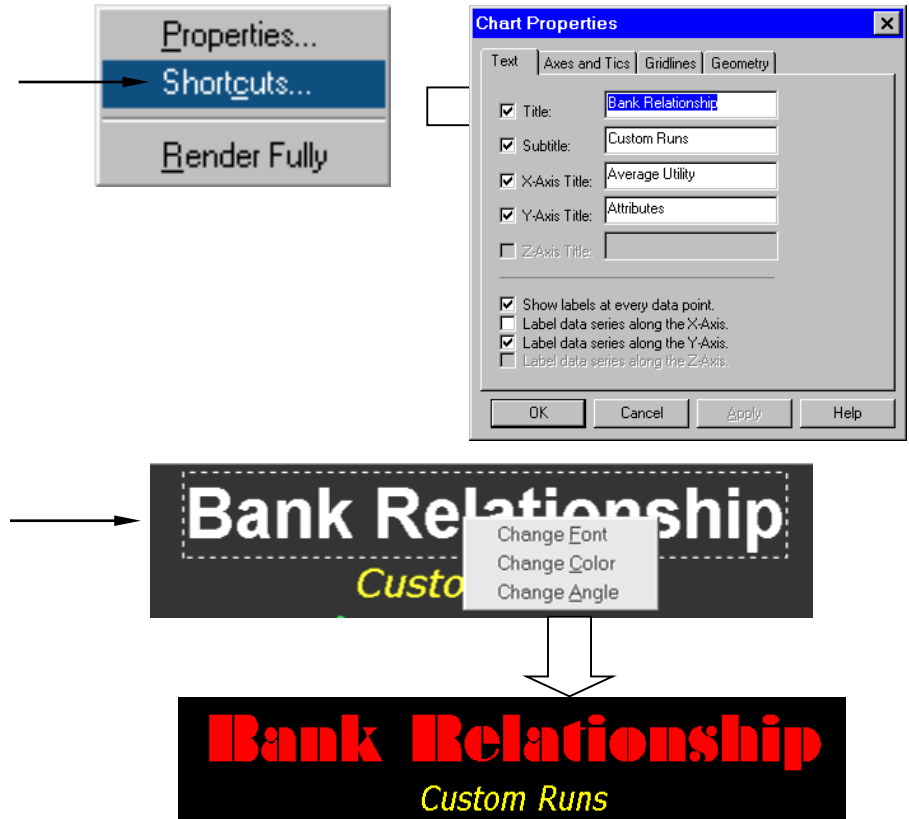
Using Charts: Shortcuts

- Using the right mouse button anywhere on the chart screen will bring up the this menu.

Properties...
Shortcuts...
Render Fully

- The Render Fully function from the right-click menu is no longer necessary. The program now redraws each time a change to the visual properties of the chart are changed, so the 'render fully' function is automatic.

- There are keystrokes associated with the zooming and geometry functions on the toolbars in SegmentSolve^{TM}. A list of the keystrokes is accessed by selecting 'Shortcuts…" from the right-mouse click menu.

- Once you are familiar with using charts, using keystrokes is the fastest way to make changes to the appearance of the chart.

- The rotation functions can be performed by holding down the left mouse button and moving the mouse in a specific direction. See the page on chart movement for more instructions.

Keyboard Shortcuts

General shortcuts

Sequence	Command
Up Arrow	Rotate chart in the negative X direction.
Down Arrow	Rotate chart in the positive X direction.
Right Arrow	Rotate chart in the positive Y direction.
Left Arrow	Rotate chart in the negative Y direction.
PgUp	Zoom into the chart.
PgDn	Zoom out from the chart.
Enter	Render chart at full resolution.
+	Increase spacing between items.
-	Decrease spacing between items.
Ctrl +	Increase intra-group spacing between items.
Ctrl -	Decrease intra-group spacing between ite...

Shortcuts for use with data selection

Sequence	Command
+	Widen selected data object(s).
-	Narrow selected data object(s).
Ctrl +	Thicken selected data object(s).
Ctrl -	Flatten selected data object(s).

OK

SegmentSolve

DECISION SUPPORT SCIENCES

Bibliography (For Definitions):

Aldenderfer, Mark S. and Roger K. Blashfield. Cluster Analysis.   Newbury Park, CA: Sage
        Publications, Inc., 1984.

Jardine, N. and R. Sibson. Mathematical Taxonomy. New York: John Wiley, 1971.

Williams, W. "Principles of Clustering." Annual Review of Ecology and Systematics 2:303-326, 1971.

SegmentSolve

DECISION SUPPORT SCIENCES