

Data Mining and CRM Applications (MKTG 5963)



Dr. Goutam Chakraborty
Professor (Marketing)
Applications of Hierarchical Clustering (Demo)

Acknowledgement: Many of the slides used in this presentation are adapted from a business knowledge series course that I teach for SAS – details: <https://support.sas.com/edu/schedules.html?id=1718&ctry=US>.

58

Objectives

- State the variables and the business issues in the hydraulics company survey data set.
- Look at Raw base variables distribution
- Understand rationale for specialized transformations such as double-standardization
- Use the average linkage method in SAS Enterprise Guide to conduct a cluster analysis on this data set using double-standardized variables.

59

Business Problem and Data Description

- XYZ is a supplier of hydraulic and pneumatic products serving 50,000+ customers in the USA.
- XYZ wants to segment their customers based on what the customers perceive to be important in choosing a supplier for hydraulic and pneumatic products.
 - Business goal: to create customized sales communication for each segment
- Anonymous mail surveys were sent to 2,500 customers chosen at random from XYZ's database.
 - 800+ surveys were returned.
 - Not all surveys are usable due to missing data

60

Variables (Selected based on business objectives)

Bases (perceived by customers as important to be used for creating segments):

- **Av_pay, Av_br, Av_spec, Credit, Price, Reliab, Return, Talk_dir, Time, and Warranty**

Descriptors (to be used for profiling segments):

- **Sales, Num_emp, and Industry**

Other managerially important variables (can also be used for profiling segments):

- **Satisf and Rate**

61



Exploring Base Variables

This demonstration illustrates using SAS Enterprise Guide to look at distributions of base variables.

62

A Typical Survey Scale and Response Style

Likert Scale Example

Strongly Disagree

Neither Agree or Disagree

Strongly Agree



63

...

A Typical Survey Scale and Response Style

Likert Scale Example

Strongly Disagree

Neither Agree or Disagree

Strongly Agree



Optimal Response Style



64

...

A Typical Survey Scale and Response Style

Likert Scale Example

Strongly Disagree

Neither Agree or Disagree

Strongly Agree



Optimal Response Style



Extreme Response Styles



65

...

A Typical Survey Scale and Response Style

Likert Scale Example

Strongly Disagree

Neither Agree or Disagree

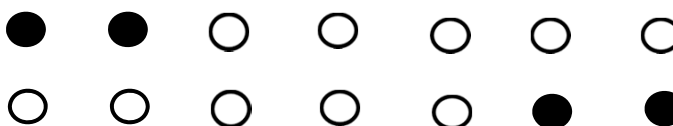
Strongly Agree



Optimal Response Style



Extreme Response Styles



66

...

A Typical Survey Scale and Response Style

Likert Scale Example

Strongly Disagree Neither Agree or Disagree Strongly Agree



Optimal Response Style



Extreme Response Styles



Midrange Response Styles



67

Solution to Response Styles for Segmentation

Centering or standardization (there are many possibilities):

- Row-centering
- Row-standardization
- Double-centering
- Double-standardization

What works best might depend on the nature of response styles, the severity of response styles, and so on. For a comparison of various methods, see *“Eliminating Response Style Segments in Survey Data via Double Standardization Before Clustering,”* by Pagolu and Chakraborty, SAS Global Forum paper, 165-2011.

68



Hierarchical Clustering (Average Method)

This demonstration illustrates using SAS Enterprise Guide to conduct hierarchical clustering on double-standardized base variables.

69

Summary of Results with Transformed Data

- Overall, there is some improvement in the distributions of variables.
- All clusters (*not single observations*) are joining at the end of the cluster history table.
- Potential number of clusters as identified by:
 - Peak in Pseudo F : 5, 7
 - Peak in Pseudo t^2 : $2+1=3$, $4+1=5$, $6+1=7$
 - % change in RMS distance: does not help here
- Five cluster solutions are carried forward to check number of observations.
 - Frequency of observations are skewed with one cluster having very few observations (unacceptable)

70