



JMP Instructional data set 019

Selecting Cities for an A/B Marketing Experiment

JMP platforms and statistical methods

Distribution:	Histogram; box plots; summary statistics
Graph Builder:	Dot plots; bar Chart
Multivariate:	Correlation; scatterplot matrix; parallel coordinate plot
Clustering:	Hierarchical clustering

Problem statement

The marketing analytics team within a national investment brokerage firm has been asked to assist the client services division in designing an experiment to test a new strategy to help the company gain share of the mid-net-worth household (MNW) client segment. This new offering consists of providing this client segment with access to services and tools at no cost that are currently only being offered to the company's high-net-worth client segment (e.g., free quarterly meetings with a financial advisor, online planning tools, etc.).



The plan is to run a pilot of the new program in 7-10 test markets (i.e., large U.S. cities) over the course of a year and evaluate its impact on acquisition and retention of this client segment. The primary question the client service division has is how to determine which cities should serve as the test markets and which non-participating cities should be compared to these test markets to compare to quantify the impact of the program. Clearly the amount of client acquisitions and retention in a city at the end of the year will not only be based upon the impact of the program, but also on many different factors and characteristics of that city. In other words, if the results in a city that receives the new offering are compared to the results of another city that didn't receive the offering, how can those factors and characteristics of each city be controlled for so that it is reasonable to attribute any differences observed in client acquisition and retention to the offering?

There are 20 cities that are candidates to be used in the experiment. Ideally, 10 will be selected to receive the new offering (i.e., test group) and 10 others, paired with each test group, to continue with the current offering (i.e., control group). These 10 pairs are to be selected such that it is reasonable to assume that client acquisition and retention would be similar for each if they did in fact receive the same offering. However, the client services division is amenable to only seven pairs of cities being included in the study to ensure that each city that does receive the new offering in the pilot program has a corresponding city to serve as the control.

Nine different characteristics for each city that could potentially impact the amount of client acquisition and retention that would occur were identified.

Data

IDS_019.jmp

City	Metropolitan Statistical Area (MSA). A city and its surrounding area
Population	Population (in thousands)
PCPI (Adj)	Average Personal Income adjusted for the cost of living of each city
MNW Ret	The % share of the Mid-Net-Worth Retiree segment the company has in that city
MNW Mid	The % share of the Mid-Net-Worth Midlife segment the company has in that city
MNW Young	The % share of the Mid-Net-Worth Young segment the company has in that city
Brand Sent	The sentiment that people in each city has towards the company's brand (1-100 with high number representing a more positive sentiment)
FS Sent	The sentiment that people in each city has towards the financial services industry in general (1-100 with high number representing a more positive sentiment)
Growth Rate	A score representing rate of growth in that city during the past 5 years (a composite score comprised of many variables such as acquisition, retention, asset inflows and balances, customer engagement, among others)
Marketing Eff	A score representing the effectiveness the company's marketing and advertising efforts has been in each city during the past 5 years. (a composite score comprised of many different variables related to the success of various marketing/advertising campaigns. Positive number means the effectiveness of marketing/advertising has been increasing while a negative value indicates it has been declining)

Exercises

1. To get familiar with the data, create histograms, boxplots, and summary statistics for each of the variables.

Instructions:

Analyze > Distribution. Place all the variables in the Y, Columns. Click OK.

2. Are there any cities that stand out as being much different than the others relative to any of the variables? If so, plan to perform the cluster analyses in Exercise 5 with and without those cities to see if these cities have a strong influence on the analysis results.

Instructions:

Graph > Graph Builder. Select all the variables and place them on the X axis.

The graph that gets created are dot plots for each variable all scaled so they can be displayed on the same graph. Note: The minimum and maximum value for each variable is shown.

3. To examine the % market share the company has for the three different client segments across the 20 cities, create these two visualizations: a) Bar Chart showing the % market of the three segments for each city ; b) Scatterplot Matrix and Parallel Coordinate Plot for the three client segment variables. Are there any cities that have some of the highest or smallest % share across all three segments? Are there cities that have very high/low % share in one or more client segments but opposite in the other(s)? Provide a few bullet point description of some of these findings in a way that would be useful for management of client services and marketing to understand.

Instructions:

a) Graph > Graph Builder. Place 'City' on the X axis. Place 'MNW Young' on the Y axis. Place 'MNW Mid' on the Y axis above 'MNW Young'. Place 'MNW Ret' on the Y axis above 'MNW Mid'. Choose the bar chart icon in the graph palette. 

This will result in 3 separate bar charts – one for each segment.

Add data values by selecting Label by Value in the Label drop down menu for all three in the controls panel on the left.

Right-click on the X axis and choose Order By > MNW Young, ascending.

The bars will be reordered accordingly. Now you can see the values for the other two segments comparing to the highest and lowest values for MNW Young. Do the same ordering for the other two segments.

b) Analyze > Multivariate Methods > Multivariate. Place 'MNW Young', 'MNW Mid', and 'MNW Ret' in the Y, Columns and hit OK. Choose Parallel Coordinate Plot under the red triangle. You can now highlight cities with the highest or lowest values for one of the segments in the Parallel coordinate plot, and see how those cities compare to the others across the other two segments.

4. Perform a Hierarchical Cluster Analysis. Identify at least 7 pairs of cities that are “similar” to each other relative to the 9 characteristics. Repeat the analysis removing any cities you identified in Exercise 2 as being very different from the rest. Do the analysis results change much? (i.e., is there much difference in the cities that pair together?).

Instructions:

Analyze > Clustering > Hierarchical Clustering. Place all variables into the Y, Columns role. Click OK. Choose to add a constellation plot by selecting it under the red triangle. Note: The dendrogram and constellation plot are two alternative ways to visualize the grouping of cities.

Select the city or cities you identified as being much different in either the dendrogram, constellation plot, or in the data table. Select Rows > Hide and Exclude. Rerun the clustering analysis while keeping the report for the previous analysis open. Compare the results by examining which cities are pairing together via the dendrogram or constellation plot.

5. For each pair of cities identified in the cluster analysis, how will you decide which city receives the new service offering and which one does not?
6. The plan is to analyze 3 outcome variables from the experiment to decide how effective the new program is: ‘New Client Acquisitions’, ‘Retention Rate’, and ‘Amount of Deposits’. Since the cities have different population sizes, how would you modify the resulting outcome data so that comparing the results across the cities are on the same scale?
7. Are there any other characteristics of these cities that you think would be helpful to include in the cluster analysis to improve upon matching the cities together? Hint: These are factors that would potentially impact the outcome variables.