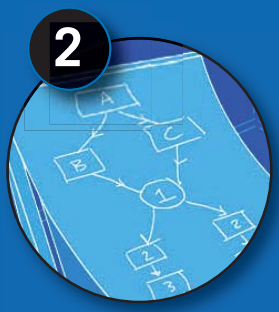
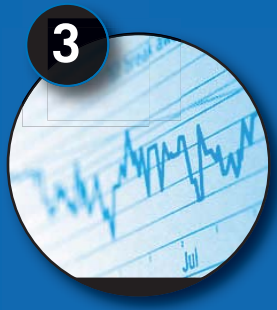
Make Sense of Your Data™

Just Three Simple Steps:

Provide Your Data:
spreadsheet or file



Specify Your Objective:
overview, relationship, classification



Name Your Project:
result is an active report

CASE STUDY: You've been transferred . . . now what?

Your employer has decided to transfer you to headquarters which is located in another part of the country. You love your house, have great neighbors, your children love the school they are in. How do you find a new place in another state that will feel like home?

COMPARING TOWNS

In this case study, EZinfo was used to help find a new town which would feel like home. EZinfo was used to compare publicly available demographic data about the town the family was leaving with data on towns near the new home office.

Their destination was a large metropolitan area with 814 towns within a commutable distance to the office. The demographic data comprised 77 variables such as tax rate, schools, stores, etc. for each town. As you can imagine, having that many towns and variables resulted in a LOT of data to sort through, but with EZinfo, it was simple. Here's how:

OVERVIEW OF DATASET (PCA-X)

The data were available in an Excel spreadsheet which was imported into EZinfo. The dataset (X) had a total of 814 towns, 77 demographic variables and 0 responses.

WHAT IS THE NOISE LEVEL?

Noise In this example, the noise level in the dataset is 40%. Noise level is a measure of how much of the variation in the data is seen in the plot, or map, of the data. .The noise level should be less than 50%. If the noise level is above 50%, the summary plot (score plot) shows less than half of the information (variation) in the data, and the risk of missing important information is substantial. In this case study, the noise level was 40%, i.e., it is acceptable.

THE X-PART SUMMARY

The first report from EZinfo concerns the X-part summary which is comprised of 8 scores. These scores are new indices, similar to the Dow Jones or NASDAQ indices which summarize a table of time points (rows) by stock prices (columns). The scores are weighted averages of the original variables, hence providing a good summary of all the collected demographic variables.

The first scores t[1] and t[2] are the two most important indices in summarizing a dataset. The plot of t[1] vs t[2] gives a picture of the data, an abstract map. The points in the plot are the towns in the data. Towns near each other in the plot have similar profiles; towns far away from each other are dissimilar. The plot shows the possible presence of atypical towns, as well as groups, similarities, trends, and other patterns in the town data. Atypical towns lie outside the ellipse.

How do the demographic variables relate (1) to each other (correlate), and (2) to the abstract town map?

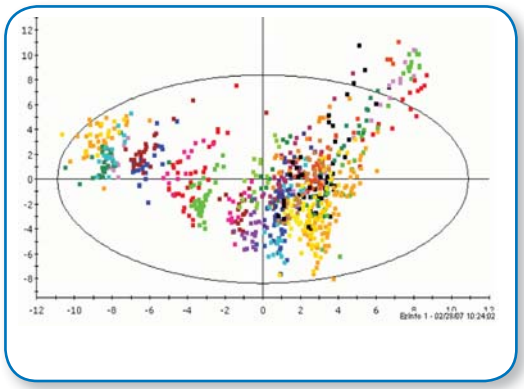


Figure 1 — Scores Plot

To understand the directions in the abstract score plot in Figure 1, we use the weights that combine the variables to form the scores. A two dimensional plot of the first two weight vectors shows directions corresponding to the t[1] – t[2] map : top or the bottom of both plots.

Scrutinizing the loading plot (Figure 2), one finds economic indicators high up and down along the vertical direction, with towns having higher average household income at the bottom of the plot.

The vertical direction contains demographic variables related to school expenditures with high values far out to the right.

Hence, for the relocating family searching for a town with high average household income and high school expenditures, good candidate towns are sitting low and to the right in the score plot. By honing in on these points, and making a second analysis based on just these towns one can take a more detailed look at the selected towns. This is done by marking unwanted points in the score plot, and excluding these points from further analysis. EZinfo automatically remodels the remaining data, updating the report. Save the previous report first!

RESULTS

By analyzing the data on the towns in the new area and comparing them to their home town, the family was able to determine which new towns were most like their current location. By narrowing down the selection process, and focusing on the neighborhood attributes most important to them, the process of selecting a new home town was easier, faster and less complicated.

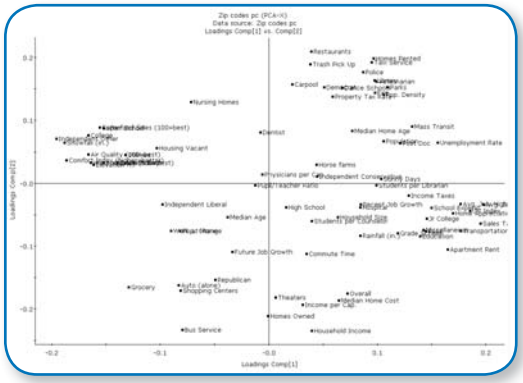


Figure 2 — Loadings Plot

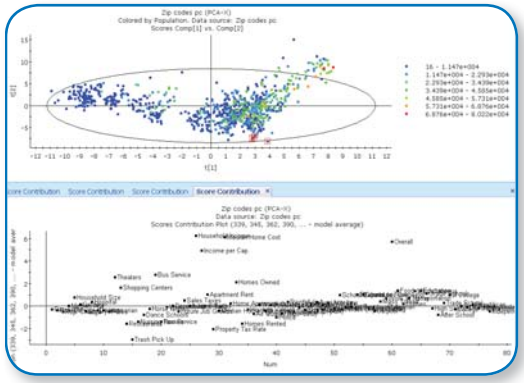


Figure 3 — Comparisons Plot

About Umetrics . . .

Umetrics develops software for design of experiments and multivariate data analysis, for the individual user as well as for on-line continuous and batch processes. We provide training at more than 25 worldwide locations and on-site consulting services. We are committed to supporting our clients in their mission to control data flow by conveying our advanced expertise in multivariate technology.

Umetrics is now owned by MKS Instruments Inc., with the acquisition finalized in January, 2006. Our general manager is Nouna Kettaneh-Wold. Umetrics has offices located in Sweden, United Kingdom and USA, and employs just over 50 people.

UMETRICS www.ezinfo.net www.umetrics.com

Headquarters:

Umetrics Inc.
17 Kiel Ave.
Kinnelon NJ 07405
USA
Phone: +1 973 492 8355
Fax: +1 973 492 8359
Email: info.us@umetrics.com

Offices:

Umetrics AB
Stortorget 21
SE-211 34 Malmö
Sweden
Phone: +46 (0)40 6642580
Fax: +46 (0)40 6642585
Email: info.se@umetrics.com

Umetrics AB
Box 7960
SE-90719 Umeå
Sweden
Phone: +46 (0)90 184800
Fax: +46 (0)90 184899
Email: info.se@umetrics.com

Umetrics UK Ltd.
Woodside House,
Windfield, Windsor
SL4 2DX, UK
Phone: +44 (0)1344 885605
Fax: +44 (0)1344 885410
Email: info.uk@umetrics.com

Umetrics USA
70 Rio Robles
San Jose, CA 95134
USA
Phone: +1 408 750 0300
Fax: +1 408 750 2990
Email: info.us@umetrics.com