An Introduction to JMP® Software, Release 3.1, SAS Institute's Statistical Visualization Software for Both Macintosh and Windows

William Gjertsen, SAS Institute Inc.

Introduction:
JMP is designed as a statistical visualization tool. JMP's goal is to analyze data in as graphical a way as possible, which lets you discover more, interact more, and understand more. The visual presentation of results is the key to doing all these things better.

Discover More: With the data displayed graphically, you look at your data and how data carry the fit. You discover the patterns in the data, and which points don't fit the pattern. There may be some phenomena in your data that you would never discover if you had not looked at a graph.

Interact More: With point-and-click responsiveness, you can interact more with data to explore it. The easier it is to interact, the more you explore, and the more discoveries you make.

Understand More: With results presented graphically, you can gain a better understanding of statistics. You can make informed judgments because you are perceiving your data both statistically and visually. Since JMP is so easy to use, you have less resistance to analyzing data and more confidence to explore.

We called the product JMP to emphasize that it is a big step (or "jump") in a new direction.

JMP is an agile instrument: JMP is built like an instrument, not a language or a dialog script. You point and click rather than program or converse. You are in control. You are encouraged to try things out rather than expecting question-and-answer dialogs to prompt you. Click on things and they will respond. Everything is constantly direct-view. You are always looking at something directly, your data, your analysis. The controls are either kept out of the main field of view, or are integrated with it. You always have a window on your data, and data are viewed in a familiar spreadsheet grid. All the data are loaded into memory.

JMP is capable: JMP is amazingly complete. It has methods for handling all combinations of the three basic modeling types of data: continuous, ordinal, and nominal. Some products are strong in regression, but weak in categorical methods. Some products have it the other way around. JMP handles both kinds of data in a consistent and unified way. For example, JMP treats lack-of-fit in regression in the same manner as goodness-of-fit is treated for categorical models. Tests against base models are done in similar ways between the ANOVA table F-test in regression, and a corresponding likelihood-ratio test for categorical responses. Most regression products don’t do lack-of-fit tests. And many categorical products don’t test against base models. But JMP handles lack-of-fit and tests against base models.

JMP is a pioneer: JMP is a pioneer in statistical graphics. JMP introduced the first implementation of general leverage plots, which show an hypothesis test point by point. JMP pioneered comparison circles, the first graphical means-comparison method that works for unequal-sized samples. JMP is a pioneer in dynamics, with the first mouse-active changing-width histogram bars. It is a pioneer in unification of statistical methods.

JMP is for the introductory user...and for the Expert: By its unified approach to statistical methods; by its integration of statistics with graphics; by its simplicity of operation; JMP rewards the introductory user with analyses that are obtained with ease, and understood readily. Experts appreciate the depth that JMP has. There are five different ways of looking at a linear hypothesis test. One-way analysis goes on and on, even after classical assumptions are left behind. JMP does nonlinear regression and gives confidence intervals that can be trusted. Powerful new features give you the ability to model survival data, request inverse prediction for ED-50 experiments, and produce ternary plots to analyze mixture designs. Version 3.1 does competing cause survival analysis, one and two dimensional density estimation and fast clustering for thousands of points.
Making Discoveries of Outliers:
They are called outliers, because they are different and apart from the rest of the data. They could be miscoded data that might wreck your statistics. Or, they could be real world phenomena waiting to be discovered and put to some wonderful use. Either way, you should look for them.

In one dimension—they are just extreme values. They are easy to spot.
In two dimensions—an outlier can stick out in an unpopular direction. If the variables are correlated, it can stick out in a two dimensional way without sticking out in either individual dimension. You can measure the outlying distance of a point in terms of how many normal density ellipsoids it crosses; this is called the Mahalanobis distance.

In three dimensions—a 3D spinning plot can find the 3D outliers. In more than three dimensions you need more help. If your data are fairly correlated, then the first few principal components describe the popular directions in which the data are spread out. But you are also interested if there are points that stick out in unpopular directions. So get a 3D rotating plot for the last three principal components and see which points stick out.
In N dimensions—Alternatively, you can look at the whole matrix of correlations and score the outlyingness of each point as a Mahalanobis distance. You get the n-dimensional distance from the multivariate mean. But these distances are scored with respect to estimates of the mean, variance, and correlations which involve outlying points themselves. It is better to use jackknifed distances—each point is scored with estimates that do not involve that point.

If you are fitting models, you want to know how each point is affecting the test statistics. The place to look is the leverage plot, which shows the residual for a point and what the residual would be with the effect removed from the model. The differences in the squares of these distances form the test statistic.
Discoveries of Patterns: Good correlations can expose interesting patterns.

Buy patterns based on correlation
Dynamic Linking: If a row in a data table is highlighted in JMP, it highlights everywhere else it is represented.

Linked Rows: When you click on a point in a plot, the point highlights and the label of the point appears in the plot. But that's not all. The row representing that point in the spreadsheet also highlights. The point for that row in other plots highlights, too. Rows are linked so that their attributes are consistent everywhere they appear.

Histogram Highlighting: Bring up a histogram showing the age, sex, and weight of the individuals. Now click on the bar representing the Males in the bar chart for sex. Instantly the bar is highlighted and the rows in the data table for Males also highlight. And that portion of all the other histogram bars and points in all the current plots corresponding to males are highlighted.

Attributes Once you highlight rows, you can assign special attributes to them using the first five commands in the Rows menu. These row states are shown by icons in the spreadsheet rows. Use Hide to hide the selected points, Exclude to exclude them from an analysis, and Label to label the points in plots. Use the Colors and Markers palettes to give points shape and color.

Generate histograms for age and sex. Click on the “M” (males) bar and do the Markers command, selecting a “+” marker. Click on the “F” (females) bar and select the “x” marker. Click on the younger ages and select “Red” color, the middle ages “green” and the older ones “blue”. Now all the points in the height by weight plot are coded so that you can see their sex and age group and notice any patterns.
Brushing  Go to the Tools menu and select the Brush tool. Click in a plot and see a brush rectangle dragging around. As the rectangle encloses points, the points are highlighted in the plot, and they highlight everywhere else they appear too.

Zooming  Get the Magnifier tool from the Tools menu and click on a plot. The plot enlarges and the axes automatically rescale. Click again to look closer at interesting points. Option-click on the plot at any time to restore the original plot.

Correct and Understandable Statistics:
One of the frustrations in statistics is that there are so many methods. Large statistical packages have dozens of procedures documented in thousands of pages. Even with a GUI front end, or even with an expert system front end, it takes a lot of effort to traverse to the method you want. Even then, you may have to assemble a sequence of several methods to get the desired reports. So with most of us unable to program our VCRs, how can we expect to approach a field as large and complex as statistics?
The JMP approach is:
- to choose families of methods that can be unified, that are popular, and that lend themselves to graphical exposition.
- to key the analysis on the modeling type associated with each variable.
- to organize methods into a few platforms where each platform handles a general situation.

Use of the modeling type is the most distinguishing feature of JMP. In the data table you declare how you want JMP to treat a variable's values. There are three choices: Select nominal if the values are unordered categories, ordinal if the values form ordered categories, continuous if the values are to be treated as from a continuous numeric scale. Having the attribute in the data table means that you do not have to make later declarations such as CLASSES statements. It helps direct the kind of analysis.

JMP organizes methods into platforms, one for each general statistical situation. A platform may have several personalities which depend on the modeling type of the variables. For example, when you use Fit Y by X, the platform will do any of four different analyses depending on the combination of models shown on the help screens.

Once the platform produces its initial display, you have a natural context to continue into deeper analyses. For example, you start with the situation of a response measured under various treatments. So you begin by launching the Fit Y by X platform. Since the response Y is continuous and treatment X is nominal, this leads to the one-way personality of the platform, and a vertical scatter plot is shown for each treatment. A pop-up menu offers to fit means and perform a one-way analysis of variance. But you also ask for Fit Quantiles to see the distribution within the groups.
JMP presents the results graphically, and integrates the graphics into the report. This is in contrast to other products, where the statistics and graphics are more disjoint. Textual results are in the same window as the graphs, and they tend to support each other.

### Using JMP to design an experiment:

Experimentation is one of the most important methods in the Quality Movement—the quest for continuous improvement in our products and processes. If you can measure aspects of quality in your product, and if you have factors under your control, then you can perform an experiment to find out which factor settings result in the best quality product. However, experimentation is usually expensive, so you want to get the most information from the least number of runs in the experiment. The statistical area of experimental design has good methods to design experiments. However, the literature in this field gets a bit technical. Good software, such as JMP, can be invaluable in leading you through design choices that result in an efficient design to help you find answers. Good software facilitates good experimental design. Even if you are an expert at experimental design, you still need good software to help you design an experiment. And even if you don’t have the time and resources to spend becoming an expert, you can do a useful and commendable job as a novice with a little guidance and good software. If you have a case that doesn’t fit any of these classical designs, then you can use the D-Optimal design facility in JMP to search for the optimal design for a particular model.

### The first task after you launch Design Experiment is to choose what kind of design you want from the Design Choice dialog. Often, it is a two-level screening design to measure the effect of many factors using only a small number of runs. Occasionally it is a response-surface design to refine optimal factor settings. In some cases you may need a special design that has some factors with three levels, especially if you follow the Taguchi approach. If your experiment involves ingredients in a mixture, use a mixture design and specify mixture portions and constraints.