What is JMP?
JMP is designed as a statistical visualization tool. JMP’s goal is to analyze data in as graphical a way as possible so that you can:

- discover more
- interact more
- understand more

The visual presentation of the results is the key to doing all these things better.

- With the data displayed graphically, you look at your data and how it carries the fit. You discover the patterns in the data, and which points don’t fit the pattern.
- With point-and-click responsiveness, you can interact more with the data to explore its phenomena.
- With the results presented graphically, you can understand your statistics more. Since the tool is so easy to use, you have less resistance to analyzing data, more confidence to explore it.

We called the product JMP to emphasize that it is a big step in a new direction.

Why the Macintosh?
JMP was implemented on the Apple Macintosh because the Mac provided the most supportive environment for the product. The Macintosh pioneered the graphical user interface. The Macintosh is a personal machine anyone can get at a local store and learn to use.

Product Attributes
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 it is viewed in a familiar spreadsheet grid. All the data is loaded into memory.

JMP is a small product. JMP fits on one diskette with its sample data. A second diskette holds the help cards. JMP does basic distribution and fitting techniques, but no specialized methods like clustering or discriminant analysis or time series methods.

Though small, JMP is amazingly complete in a very important way. It has methods for handling all combinations of the three basic measurement levels of data: interval, ordinal, and nominal. Some products are strong in regression, but are weak in categorical methods. Some products have it the other way around. JMP handles both kinds of data in a way that is consistent and unified. For example JMP treats Lack-of-fit in regression in the same manner that Goodness-of-fit is done for categorical. Tests against base models is done in similar ways between the ANOVA-table F test in regression, and a corresponding likelihood-ratio test in categorical. Most regression products don’t even do lack-of-fit tests. And many categorical products don’t test against background models.

Try it out.
JMP must be experienced first-hand to appreciate fully. You can try it in the exhibit hall here at SUGI. There is a video available, and we are working on a demo disk.

The Data Table
JMP presents the data table in a familiar spreadsheet grid, so you are always seeing your data. That is important. It is not like a data base where you have to think of "retrieving" your data through inquiry procedures. Your data is the source of all the analysis, all the discoveries, so it is something that should be right in front of you. Too long, we have had the data base tradition where you see the programming, and the data is hidden. Welcome to the spreadsheet era where you see the data and the programming is behinds the scenes.

The data table window has a very responsive control surface. Point and click on a cell to edit it. Point and click on a column name to change it. Pull down a menu to transform it. Point and click on row or column headings to select them. Click and drag on a set of cells to select them. And then you can copy-and-paste. Click and drag on a column border to resize it. Double-click on a column to get and change information.

Imagine you want to calculate a new variable. On the Columns menu, issue New Column.

A New Column dialog appears to let you define the name and attributes of the column. Check the box Formula, then OK.

Table Name: STUDENTS
Col Home: [ ] Numeric [ ] Character [ ] New Data
Field Width: [ ] Format: [ ] Best [ ] Fixed Dec
Measurement Type: [ ] Interval [ ] Ordinal [ ] Nominal
Notes: compute new column as weight/height

Use Formula

OK

Cancel
Now a calculator window appears. To form the ratio of height to weight you click on height in the columns panel, click on the divide button, and on weight. You now have the formula beautifully displayed in typeset form. Click on the close box and the formula automatically performs the calculation as the window is closed.

Change a value of weight, and the ratio column is automatically updated! The formulas are live, like in a spreadsheet.

First, because the measurement levels were already committed, we had a much shorter path to the analysis. The path related to the situation directly, rather than through a catalog of statistical methods. This gave it almost an expert-system aspect, but with out the complexity and the question-and-answer. It also reduced the control surface area, since we do not need things like the CLASSES statement of GLM, or the DISCRETE option of PROC CHART.

Second, we tried to anticipate and produce almost every report that would be needed, rather than burdening the user with asking for it. When you fit a model with nominal effects, you automatically get least squares means; you don't have to ask for them. If there are exact replicates in the observations, and the model is unsaturated, you automatically get a lack-of-fit report; you don't have to ask for it. When options are available, they are available interactively; you don't have to rerun the analysis.

Third, we tried to present the results graphically, and integrate the graphics into the report. This is in contrast to other products, where the statistics and graphics are more disjoint. Instead of having a plotting platform that happens to fit regressions, we have a fitting platform whose centerpiece is a plot. Instead of the same plotting platform able to have categorical axes, we have the fitting platform present a graphic tailored for a nominal x axis. There are advantages either way, of course, but we do feel that the focus on analysis has advantages over a more mechanical focus. The textual results are in the same window as the graphs, and they tend to support each other.

The top of the Help tree for Analysis is this picture:

The Analysis

One of the frustrations in statistics is that there are so many methods. Large statistical packages like SAS have dozens of procedures documented in thousands of pages. Even with a menued front end, or even with an expert-system front end, it takes a lot of effort to traverse the method you want. Even then, you may have to assemble a sequence of several methods to get the desired reports.

So the important question is how to choose and organize statistical methods so that one can have a small but useful consumer product? Our 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 measurement level associated with each variable.
- to organize methods into a few platforms where each platform handles a situation. One situation is to treat each variable by itself. Another to look at relationships in pairs.

In choosing methods, we gave up approaches like Nonparametric methods. The nonparametric literature seemed to present too loose a collection of methods; the small sample probabilities were difficult to calculate; and most of the methods do not lend themselves to a graphical exposition. We expect to make add-on products later to recover lost areas like this.

Using measurement levels is the most distinguishing feature of JMP. You declare whether a variable is nominal, ordinal, or interval in the data table so you don't have to redefine it every time you analyze it. Also, this obligated us to implement methods for every measurement level. In particular, JMP has addressed ordinal data in a much more complete way than other packages.

The platform approach meant several things to us.

The first statistical inquiry of a set of data is often just checking out the distribution of values. Choose "Distribution of Y's" in the Analyze menu.

JMP will prompt you for the variables, and then present the results in a window. Or you can select the variables ahead of time by setting their role. Most packages support either preselected or prompted variable selections, but JMP supports both.

The resulting report is displayed in a new window. Notice that the nominal and interval variables are treated differently. Each has both graphics and text reports showing the distribution.
The surface of the report bristles with interactivity. Click in the corner "stretch box" of a graph and drag the mouse to change its size. Click on buttons titling reports to alternately reveal and conceal the reports. Get various tools from the Tools menu to interact in specialized ways. Click on a question mark to get help. (There are three other ways to get help too.) Click and drag with a scissors to select areas for cut-and-paste.

Get theGrabber tool (the hand), and click and drag to instantly change the granularity and position of the histogram bars. Drag to the right for narrower intervals. Drag to the left for wider bars. Drag up and down to set the positioning.

Click on a bar with the standard arrow cursor, and the rows corresponding to that bar are selected in the spreadsheet, and in all the other graphs, including the other histograms.

Click the mouse for action. This scatterplot can resized by clicking and dragging in the corner stretch box. You can click on a point and it will

The "Fit Y by X" platform is for studying the distribution of a Y variable as it is affected by values of an X variable. This platform is really four platforms, because it does different methods depending on the measurement level specified for the X and Y variable.

The Help Picture for "Fit Y by X".

Fit Interval Y by Interval X

Let's study the distribution of height by weight, both declared as intervally measured. The platform starts with a scatterplot of the two variables.
highlight the row in the spreadsheet and that point in every graph. As long as you hold down the button, it will display the ID of the row. You can highlight areas of rows with a brush tool. The points inside the brush rectangle highlight as you move it around. You can change attributes about the point like color, marker, and several others. (By holding down the option and command keys when you mouse, you can even get the brush to glide and bounce around the graph like the ball in a Pong game, highlighting points as it passes over them.)

The submenu icon is clicked to reveal the following menu of fits that can be performed on the data.

- Show Points
- Fit Mean
- Fit Line
- Fit Polynomial
- Fit Spline
- Density Ellipses

Each time, a fit is added to the graph, along with statistical reports detailing the fit. You can remove a fit, or change its color. You can add or remove confidence curves.

**Fit Interval Y by Nominal X**

If the X variable is given the nominal measurement level, then the "Fit Y by X" platform presents a very different graph. The central question is: Are the means of the response different in the different groups? The Y axis is the same. The nominal X axis is divided to show the relative frequency for each level.

Below the graphs are the reports detailing the statistics. The reports are for the one-way analysis-of-variance model for the means.

**Summary of Fit**

- R^2 = 0.4457445
- Root Mean Square Error = 3.382558
- Mean of Response = 62.55
- Observations (or Sum Wgts) = 40
The Fit Y by X platform can also handle the situation of modeling the probabilities of a nominal response as a continuous function of an interval X variable. The method for doing this, logistic regression, fits a series of S-shaped curves. The probability for each response level is the distance between each curve at a particular X value. The probabilities sum to 1.

If X does not affect the distribution of Y, then the slope parameters will tend to be nonsignificant, and the lines of fit tend to be more horizontal.

The ordinal cumulative logistic regression technology in JMP is especially capable, able to handle a large number of response levels.

Fit Nominal Y by Interval X

An ordinal response is modeled in a manner similar to a nominal response, but the model is simplified by having only one common slope parameter. The curves that separate the probability areas are the same curve just shifted horizontally by an amount that scores the spacing of the response levels.

The measurement levels in JMP are practical and convenient method indexes, not inherent and immutable data descriptors. We expect users to try out several measurement levels for a given variable; in fact we make it easy and obvious through a popup menu on each column of the
data table. If you have an ordinal response, you might consider trying it both as ordinal and nominal, perhaps even interval too. Then evaluate.

Platform for "Fit Y by X's" and "Specify Model"

There is a general fitting platform that is accessed through two menu items: "Fit Y by X's" is for a quick specification of a straight-forward model. "Specify Model" is more complex, and more powerful.

You can model the response as nominal, or ordinal, but the results are not very graphical. The interval response model is supported very richly. Here is where you can to multiple regression and Anova models. Here is the rough equivalent to GLM in SAS. You can specify multiple regressions, main effects, cross effects nested effects, polynomial effects, even response surface effects.

The results are shown with the traditional tables of statistics. But the showcase feature is an innovative plot that shows point-by-point the composition of each hypothesis test.

We present the leverage plot, named as a generalization of the partial regression residual leverage plot (of Belsley, Kuh and Welch, Regression Diagnostics). The plot is always drawn with two reference lines, one horizontal, the other sloped. The points on the plot are positioned so that the distance from each point to the sloped line is the residual, and the distance from each point to the horizontal line is the residual that would result from constraining the model by the hypothesis. The difference of the sums of squares of these two distances is, of course, the hypothesis sums of squares.

The leverage plot that does a joint test for all effects in the model (except the intercept) is called the "Whole Model" leverage plot, and happens to be simply a plot of actual vs. predicted values. The leverage plot for a single simple regressor effect is the same as a partial regression residual leverage plot. We show a leverage plot for almost every hypothesis tested. We also include confidence curves which show whether the hypothesis is significant at the .05 level by whether they cross the horizontal line or not.

If you have nominal effects, the platform will automatically produce a least squares means table.

You can also ask for specific contrasts through a contrast dialog.

with the results:
Spinning Plots and Biplots

Graphics can reach into 3 dimensions by exploiting dynamics. With fast machines and graphics, we can spin 3D scatterplots in real time, e.g. 10 to 15 frames per second. The Spin platform in JMP does this. It is actually pretty fun to do this. You can spin it using the buttons. You can also get the grabber tool and just click and drag, and the plot will spin to follow your mouse motion in real time. You can shift-click and drag-release and the plot will spin by itself. You can click on points or brush in, as with the other in JMP. But this is limited to three variables at a time, of course. You can change variables by dragging the X/Y labels to other variables.

To get a visual view from the components, drag the XYZ to the principal components, and the plot will show not only the points according to the components, but also rays representing and approximating the variables in that subspace. This simultaneous display of principal component points and variables is called a Gabriel biplot, and it is a particularly powerful way to explore multivariate data.

Y's by Y's

The Y's by Y's platform shows the relationships among many pairs of variables. The first report is just the correlations among the variables.

But correlations are more interesting if you can see them visually. JMP produces a scatterplot matrix showing scatterplots for each pair of variables. You can brush in one plot, and the corresponding points highlight in the other plots. The normal-density ellipsoid shows the correlation visually, with high correlation indicated by diagonal flatness.
These correlations can be used to help measure each point's outlyingness with respect to the correlation structure in a high number of dimensions. The result is a plot of the Mahalanobis distances, showing in this case the four-dimensional outlyingness of each point.

Convenience Features

A product needs convenience features to help you find your way, to help you manage data and results. Data must be manipulated and saved. Reports must be assembled and printed.

One of the most compelling features of Macintosh software in general is that you can cut-and-paste freely. JMP allows you to cut and paste data table values to and from its spreadsheets, and to-and-from other spreadsheets such as Excel. JMP allows you to copy pictures from windows or selected parts of windows, and paste them into other applications. You could paste them into a draw program, for example, to annotate and embellish them, and then paste them into a word processing or publishing application to make a report.

JMP also has a Journal facility for saving results to a MacWrite file. With this file, the results are in text form, so they can be easily edited with any word processing program. The text is saved with tab settings so that the tables display well from the beginning.

**JMP IN**

We believe that JMP is ideally suited for student users, since it is so easy to use, and promotes a natural visual understanding of statistics. So we offer a version of JMP called JMP IN that has all the features, but is limited in the size of the data table that can be saved or pasted (500 cells).

**Future**

We are working on more business graphics, multiple comparisons, and quality-control charts. Also we are studying mixed models, and certain other high-end statistical methods that could become add-on products. Further in the future, we will be working on making JMP more programmable, customizable, and extensible.