POINTS OF INTEREST

See Structure
through graphical visualization

Uncover Relationships
through statistical discovery

Explore Opportunities
through designed experiments

Introduction

JMP (pronounced "jump") is a statistical visualization and discovery tool that graphically displays and analyzes data and results. By using graphics, JMP lets you discover more, understand more, and interact more.

Discover More

JMP helps you discover more with graphically-displayed data and analyses, which help you visualize data and see how it carries a fit. With graphs, you can discover patterns in the data and see 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.

Understand More

With graphically-presented results, the statistics themselves become easier to understand. This encourages analyzing data and provides confidence to explore. By perceiving data both statistically and visually, you can make informed judgments.

Interact More

While analyzing and exploring data, you can interact with it using JMP's point-and-click responsiveness. Because JMP makes it easy to interact, you explore more. The more you explore, the more discoveries you make.

Built with Your Needs in Mind

JMP is built like an instrument, not a language or programming batch. You point and click rather than program or converse. You are in control. You are encouraged to try things rather than react to question-and-answer dialogs.

Some statistical software packages are strong in some techniques, such as regression, but weak in others, such as categorical methods. JMP handles all kinds of data in a consistent and unified way. It handles all combinations of continuous, ordinal, and nominal data.

For example, JMP treats lack-of-fit in regression the same as it does goodness-of-fit in categorical models. Most regression products don't do lack-of-fit tests, and many categorical products don't test against base models. JMP handles both lack-of-fit and tests against base models.

Investigation without Frustration

JMP keeps you moving smoothly throughout your analysis. Start by describing the basic situation, obtain a graph and the initial results, and then progress to a detailed analysis as you lead the investigation naturally. Contrast this with the interrupted style in older packages, where at every step you fight your way through dialogs to tell the software what to do.

JMP is a complete stage for investigating data in an effortless, uninterrupted process.

JMP is for the introductory user and for the expert. JMP's unified approach to statistical methods and integration of statistics with graphics gives its operation ease enough for the beginner. It rewards the beginner with analyses that are obtained with ease and are readily understood. On the other hand, JMP gives experts the depth they want, such as

- one-way analyses that continue even after classical assumptions are left behind
- nonlinear regression and confidence intervals that can be trusted
- tools for reliability, Six Sigma, and many other industrial applications, including capability analysis and a large array of control charts that can receive data live from a measuring instrument
- powerful Design of Experiments (DOE) capabilities, including classical, optimal, and supersaturated designs
- *k*-means, hierarchical, and normal mixtures clustering, including self-organizing maps
- exploratory modeling (sometimes known as "data mining") through classification and regression trees, and recursive partitioning
- a scripting language that allows for automation of repetitive tasks, extensions of JMP's built-in analyses, and creation of instructional simulations

Interactivity Among Graphs, Data Tables, and Histograms

Most statistical packages are not responsive—the data sit dormant, there is a brief flurry of action in response to commands, then the reports are deposited in a non-reactive state.

JMP, however, is designed as a dynamic data analysis system. Windows on your desktop are ready for action. Data stay in front of you. Analyses are ready to respond to your interests. Everything is alive until closed. When you click in a window, JMP responds instantly.

Using Linked Data

Highlight a point, bar, or row, and JMP highlights its corresponding data everywhere else in the session it is represented. Data are linked so you can easily notice patterns.

For example, click a point in a plot. The point is highlighted, and when the arrow is over the point, its label appears. But that's not all. The row representing that point in the data table is highlighted. The point for that row in other plots is highlighted, too.

Click a histogram bar and that bar is highlighted along with the corresponding rows in the data table. And also, all other open histograms and plots that contain the corresponding part show the same data highlighted.

Linked data is useful when you want to look at more than one point. In JMP, you can click and drag the cursor to select a group of points. As the rectangle encloses the points, they are highlighted in the plot and everywhere else they appear.

You can also tell JMP to hide or exclude points from an analysis, label them, or give them specific colors or shapes. When you do this in a data table, all the points in your previously-made plots and histograms assume these characteristics too, so you can easily notice any patterns.

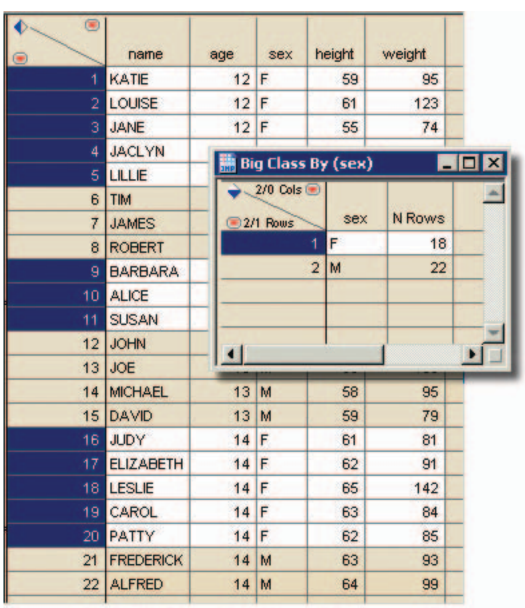
Summary Tables

Summary tables are linked to their parent table, so clicking in the summary table highlights associated rows in the parent table.

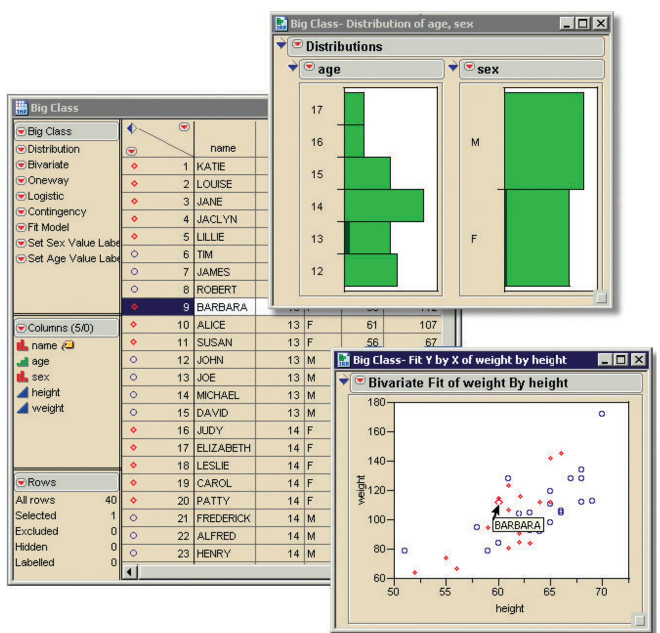
Manipulating Histograms

In analysis reports and histograms, JMP lets you

- click and drag in the corner of a plot to resize it
- double-click a plot axis to rescale it
- access appropriate statistical commands
- click and drag with the hand tool in a histogram to change intervals and shift the histogram bar positions



Click a row in a summary table, and all corresponding rows in the parent table are highlighted.



Data in JMP are linked: click one area in a histogram, and all coordinating points are highlighted in open data tables and graphs.

Data and Data Tables

JMP keeps your data table visible in a familiar spreadsheet format. You do not have to use inquiry functions to see it.

Editing and Highlighting

Data tables in JMP are responsive. You can

- click a cell to edit its contents
- click a column name and start typing to change it
- click rows or column headings to select them
- click and drag over a set of cells to select them
- click a group of cells to drag them elsewhere
- copy and paste
- click and drag on a column border to resize it
- drag and drop columns to rearrange them

Linked Data

JMP allows you several views of your data. You can

- highlight a summary row and all the observations in the original data table are also highlighted. Then you can create a new data table containing only that highlighted data.
- bring up several windows on the same data table and view them independently
- edit a value in one window and it changes automatically in the other windows

	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128

A JMP data table.

Assigning Attributes to Rows

To assign special attributes to rows, highlight them and select an option from the Rows menu.

After assigning attributes to a row, an icon appears in the data table to indicate its attribute(s), which include

Hide—prevents rows from showing in plots

Exclude—prevents rows from being included in an analysis

Label—labels rows (points) in plots

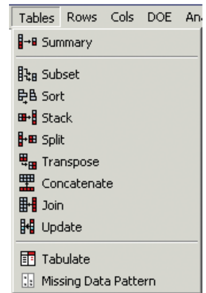
Color—adds one of any of 65 colors to points

Marker—replaces dots in graphs with one of 16 shaped markers

	name	age	s
1	JOE	13	M
2	MICHAEL	13	M
3	DAVID	13	M
4	JUDY	14	F
5	ELIZABETH	14	F

Rows with assigned attributes.

Manipulating and Forming New Tables



The Tables menu.

The Tables menu allows you to

summarize columns from the active tables

subset selected rows and columns

sort by any number of variables in either ascending or descending order

stack columns to create a long, narrow table

split columns to create a shorter, wider table

transpose to interchange rows and columns

concatenate to append tables end to end

join tables side by side, or match values from one table in the other

update one table with values from another table

drag and drop variables to create a table of counts and statistics

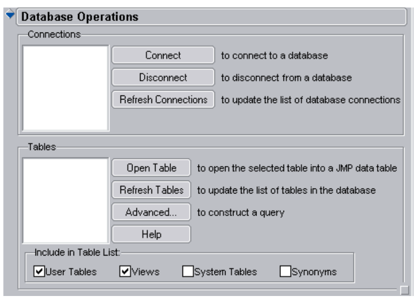
shows pattern of missing values in data table

Data Access Capabilities

JMP communicates seamlessly with other applications and data formats. For example, it can access data that was saved in a variety of formats.

Using ODBC

On Windows and Linux, JMP implements ODBC (Open DataBase Connectivity). It can open any file type that has a corresponding ODBC driver installed on your machine. It can access either a file or external databases.

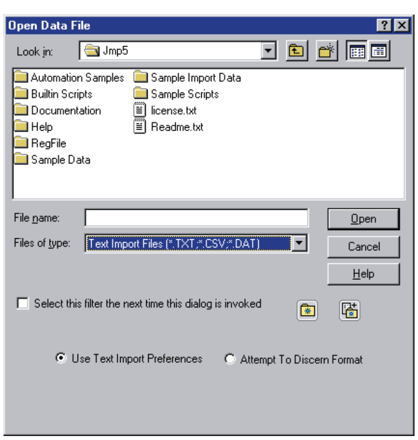


JMP implements ODBC.

In addition to accessing relational databases, JMP allows you to issue SQL commands to the database to extract only the data you want. Once connected, reading and writing data is seamless.

Importing Data from Text Files

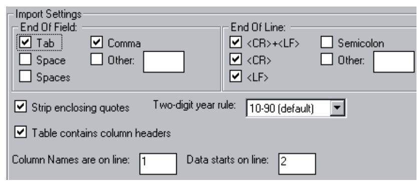
If your data is stored as plain text, JMP's flexible text import facility transforms it into a JMP table.



JMP can open plain text files as a data table.

Whether you are importing from a consistently-formatted source, or from varied text layouts, JMP's flexible import method can read it.

- If you know the format of the data ahead of time, you can set JMP's preferences so that opening any file of this format is a single command.

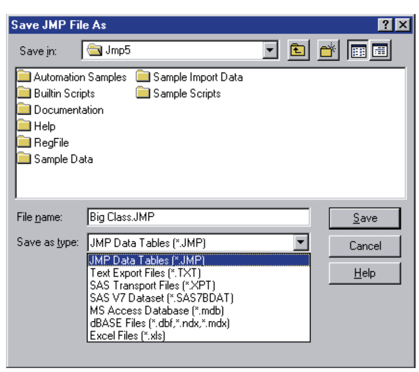


Under JMP's Preferences, you can change the default import settings.

- If you are retrieving data from sources with different layouts, JMP will attempt to discern the format of the data. It provides a preview of the data, allowing you to tailor the settings on a table-by-table basis.

Reading and Writing Excel Files

JMP can open Excel files. On Windows and Macintosh, JMP can save data tables in Excel format. By giving you this flexibility, JMP provides a way for you to transfer your data to a wide variety of users. JMP for Linux imports and exports OpenOffice spreadsheets.



JMP can access and save as several different formats.

Reading and Writing SAS Files

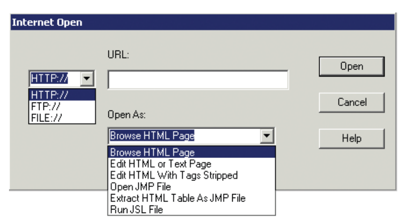
On Windows, JMP can open SAS datasets that are version 5 or later. On the Macintosh and Linux, JMP only opens version 6 SAS datasets. All versions of JMP can open SAS transport files that were saved using the SAS XPORT engine.

Using the Save As command, JMP can save files as SAS Transport files. JMP on Windows can also directly save SAS datasets.

JMP can also interact directly with the SAS system. Using JSL, you can log into a SAS session, construct and submit SAS code, and retrieve output and log files. There are also JSL commands to manipulate SAS variable names and librefs.

Accessing Remote Machines and Web Servers

The Windows edition of JMP can open HTML, JMP, or text files stored on a web server or FTP site, which gives you the ability to import web pages and files as JMP data tables.



Access files stored on a web server.

Live Data Collection

JMP can communicate with external devices through the serial port on your computer. Through this port, JMP can gather data and add it to data tables through scripts.

Report Surface Accessibility

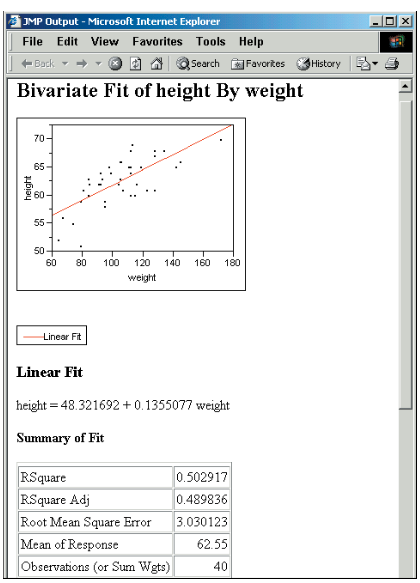
Once you have acquired and analyzed data, it is likely that you need to communicate your results to others. JMP has several capabilities that make this task easy.

Saving Reports

Using the journal and layout features, JMP can save a report in many formats, including HTML, PNG, JPEG, and RTF. (On Linux and Macintosh, supported formats are JRN, RTF, and HTML.)

Selecting an area of an analysis window lets you save only the selected area.

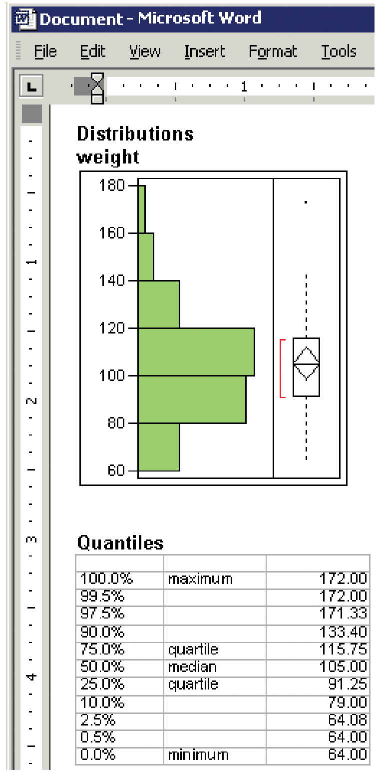
JMP reports can be saved as HTML, where graphics are automatically collected in a separate folder.



RSquare	0.502917
RSquare Adj	0.489836
Root Mean Square Error	3.030123
Mean of Response	62.55
Observations (or Sum Wgts)	40

JMP output as viewed in a web browser.

JMP preserves the graphs as pictures and the text output as tables when you save a report as an RTF file or drag and drop a selection into a Microsoft Word or PowerPoint window.



100.0%	maximum	172.00
99.5%		172.00
97.5%		171.33
90.0%		133.40
75.0%	quartile	115.75
50.0%	median	105.00
25.0%	quartile	91.25
10.0%		79.00
2.5%		64.08
0.5%		64.00
0.0%	minimum	64.00

Graphs and tables are preserved as pictures and tables when you drag and drop into Word.

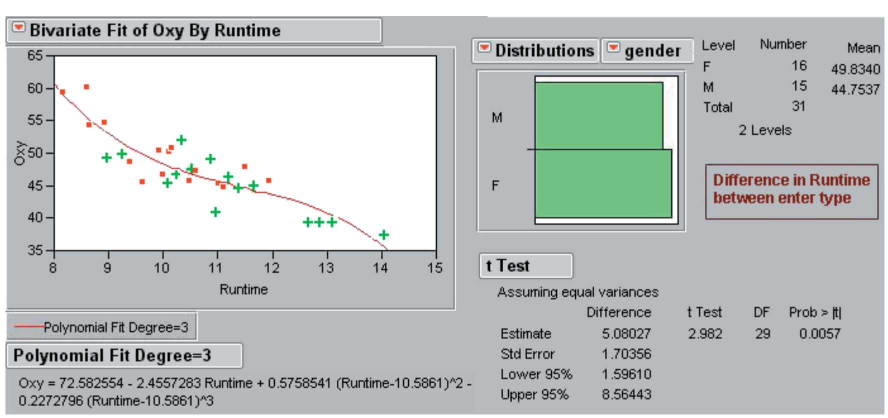
Manipulating Reports Before Saving

With its Layout window and menu, JMP gives you the freedom to open an output report in a separate window and manipulate its appearance before you save it. You can move parts of the reports around, add other sections to it, and delete sections.

Using the layout feature, you can

- click an object to select it and move it anywhere in the layout window. The layout window has as many pages as you want, outlined with gray boundary lines
- double-click a title bar or report table column header and edit it
- rerun the analysis in a new window or edit the script

The example below shows results extracted, combined, and organized in a compact form from a bivariate analysis (scatterplot and polynomial fit), a distribution analysis (histogram and frequency count), and a one-way analysis (means and *t*-test). The result gives a clean and concise summary of oxygen uptake as a function of run time in an exercise experiment. This layout was arranged by forming a layout window for one analysis, then dragging the other desired analysis elements into the layout window, arranging the desired pieces, and deleting unwanted elements.



Results from multiple analyses that were edited and combined into the layout window.

Specialty Tools

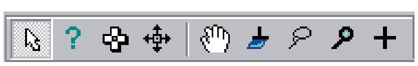
As you move the cursor around the window, it changes shapes to indicate the available action.

The different shapes mean that you can

- click inside a cell to edit its contents
- click and drag inside cells to highlight a set of cells
- click and drag the border between columns to resize column width
- click the red triangle icon and select from the pull-down menus
- click rows and columns to assign them attributes

Tool Palette

You can also change the function of the cursor manually by clicking JMP's tool palette, which contains specialty tools. You can use these tools to brush over points on a plot to highlight them both in the plot and in all other open windows. You can also use the magnifier tool to take a closer look at points by rescaling the axes and enlarging the plot.



JMP's tool palette.

The following is a list of cursors and their function in JMP.

The arrow pushes buttons and selects when you click with it. It is the default tool.

The question mark displays help text on the area or object clicked.

The fat plus selects rows or regions of rows when you click and drag. When you click a section of a plot or report, that section highlights and can be copied or dragged.

The scroller provides an easy way to scroll to all parts of large reports. Click and drag with it to move up and down on the report.

The hand moves objects with each click and drag. It directly manipulates plots, charts, axes, and formula components. It re-bins histograms, repositions plot scales, and spins plots.

The brush highlights points in rectangular areas of a plot. When you click with it, a rectangle appears. Move the rectangle over points to highlight them and shift-click to extend the selection.

The lasso highlights irregular areas in a plot when you click and drag to enclose them.

The magnifier zooms on points of interest. The point or area selected becomes the center of a new view of the data.

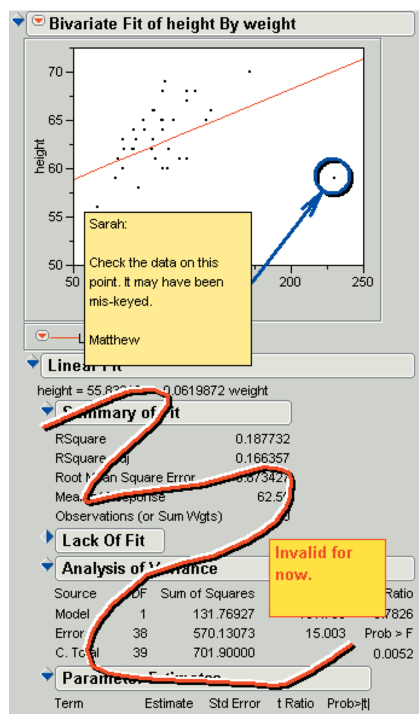
The crosshairs display a moveable set of axes measuring plot points and distances.

The annotate tool places a text box wherever you click in a report, journal, or layout window. You can type notes into that text box. The box and its text can be any color or can be transparent.

The line draws thin, thick, or dashed lines, which can have arrows at either or both ends.

The polygon draws any shaped polygon and can be spline smoothed.

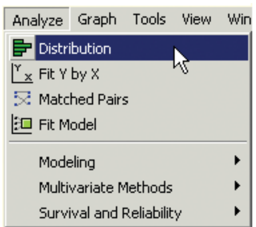
The simple shape draws either ovals or rectangles. Polygons, rectangles, and ovals (simple shapes) can be filled and raised to give a three-dimensional appearance.



The annotate, line, and simple shape tools were used on this JMP output report.

Distribution of Values

Often, your first statistical data inquiry consists of checking the distribution of values. To help you accomplish this, JMP provides you with the distribution command.



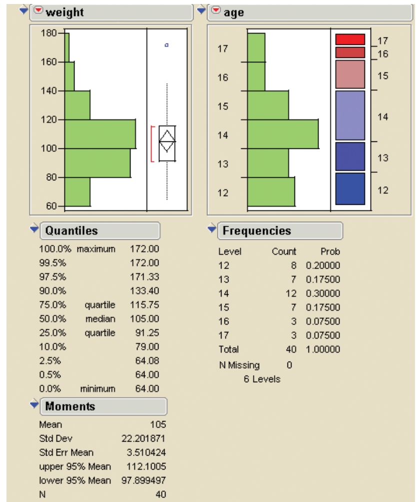
The Distribution command.

After selecting the distribution command, you identify the variables that interest you.

Then, JMP decides how to treat the variables based on their modeling types.

- For continuous variables, JMP calculates the mean, standard deviations, and quantiles. It also graphically presents results in the form of histograms and box plots.
- For nominal and ordinal variables, JMP counts frequencies and calculates percentages. It shows results with a frequency bar chart and, upon request, gives a divided bar chart (also called a mosaic plot).

Results are presented as both graphs and text, and they are conveniently placed next to each other in the same window. Furthermore, the report responds to your actions. For example, clicking a histogram bar highlights the row(s) in the report and in the corresponding data table, which



Continuous and categorical distributions.

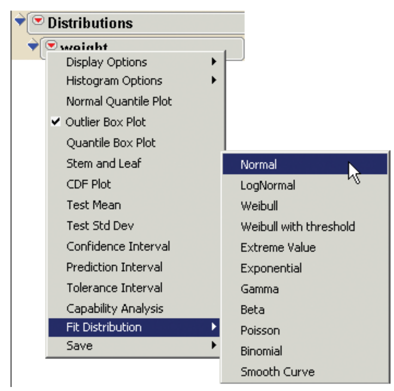
helps you see patterns among variables. Additionally, you can click and drag the histogram bins to modify their widths and locations.

Step-by-Step Analyses

JMP analyses unfold. Additional statistics and plots are available from the report surface in the form of drop-down menus, allowing you to access analyses one at the time as discoveries are made. Unlike other statistical products, you don't have to re-run an analysis to add additional options to an existing report—they are always available via the red triangle icon on the title bar.

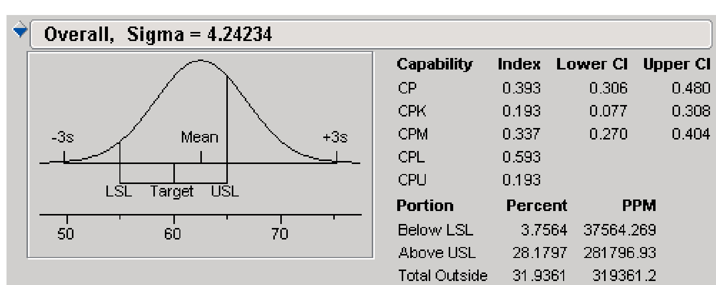
Similarly, JMP contains commands in those drop-down menus that describe the shape of the distribution.

- The **Fit Distribution** option superimposes a fitted distribution curve on the histogram. Normal, LogNormal, Weibull, and many other distributions—along with the appropriate goodness-of-fit statistics—are only a click away.
- The **Normal Quantile Plot** option shows the observed values against the expected normal order statistics. If the data are normal, the points on this plot tend to follow a straight line. This is a quick and accurate visual way of confirming normality.
- The **Smooth Curve** option lets you apply a kernel density smoother to the data, which can then be interactively adjusted to get the desired level of fit.



Commands that help describe the distribution.

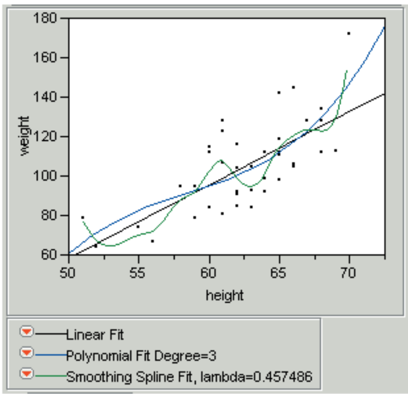
A complete capability analysis, based on specification limits that you set, is also available. As with other analyses, it is presented both graphically and textually.



Example of a capability analysis.

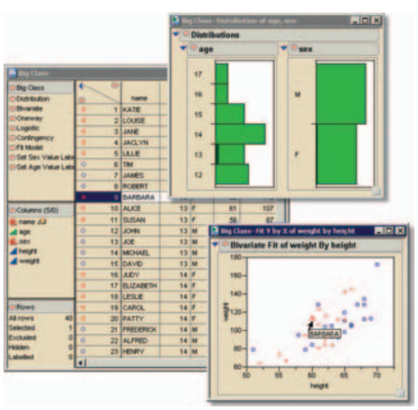
Bivariate Relationships

When you initiate the Bivariate platform in JMP, it starts with a scatterplot of two variables.



Bivariate scatterplot with three fits.

- Several fits can be overlaid on a single graph for easy comparison of their fit.
- The scatterplot can be resized by clicking and dragging the plot's sides or corners.
- Hovering over a point in the scatterplot reveals its label.
- You can change the point's color, marker, and label.
- As with other platforms in JMP, clicking on a point highlights its corresponding point in every open graph, as well as the row in the data table.



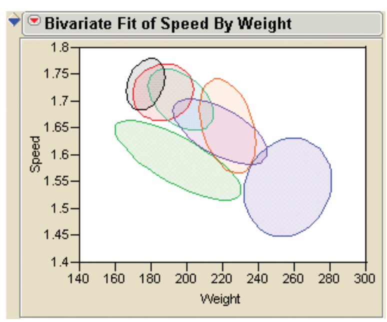
Click a point in a plot and the same data is highlighted in every open graph and table.

Performing Fits

After displaying the scatterplot, JMP gives you fitting options to perform on the data.

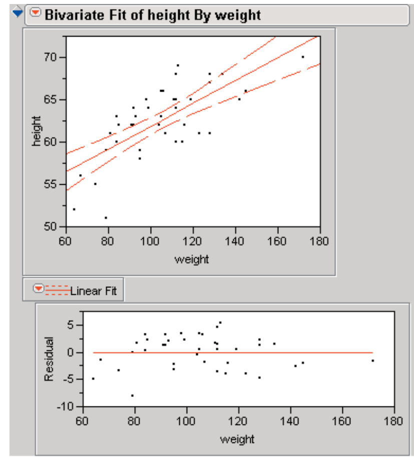
Each available fitting command adds a fit along with statistical reports detailing the fit. In particular, the **Group By** option allows you to perform several fits based on the value of a grouping variable.

For example, a few years ago, some Utah researchers were looking at the bivariate distribution of football players' weight and speed by position. With JMP, this task involves two mouse clicks: clicking the drop-down menu and choosing the **Group By** option, then clicking it again and choosing the **Density Ellipses** option. Furthermore, the graph can be easily copied and pasted into a presentation program.



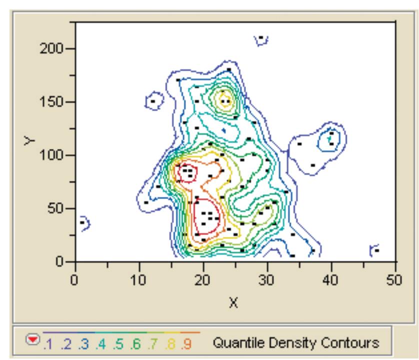
Bivariate ellipses by position.

Thus, as you make discoveries and see patterns, exploration options are available to you. JMP makes it simple to look at several aspects of a fit and determine the effectiveness of your model. In the next example, a residual plot was requested after a bivariate fit to check model assumptions. Then, confidence curves were overlaid on the plot. These options were available from the drop-down menus on the plot surface—the analysis did not have to be re-run.



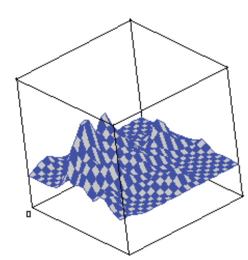
Confidence curves and a residual plot.

JMP even lets you graphically explore points' bivariate density. You can also overlay nonparametric densities with a single command, helping you determine if your data is piling up in a small area.



Nonparametric density.

You can also see the data spring to life in three dimensions.



A three-dimensional graph in JMP.

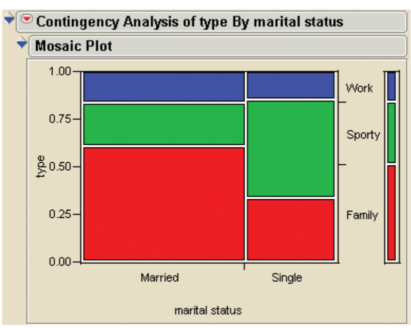
Contingency Tables and Correspondence Analyses

Nominal responses are modeled according to the probability that each response level will occur. One question you might ask when looking at nominal responses is if the probability for the various response levels is a function of the x level or if it is constant across the x levels.

Some statistical packages treat the analysis of two nominal variables with pages after pages of tables. JMP, however, makes these analyses come to life with integrated graphics.

Mosaic Plots

JMP presents mosaic plots to help you determine if the probability for the various response levels is a function of the *x* level.



A JMP mosaic plot.

The mosaic plot is divided on the *x*-axis by the population sizes of the *x* levels. The *y*-axis is subdivided by the response rates for each response category. If the response rates are similar across the *x* groups, then the partitions will align horizontally.

The same information is displayed non-graphically in a crosstab report for each cell.

Cheese- Contingency

Contingency Analysis of Response By Cheese

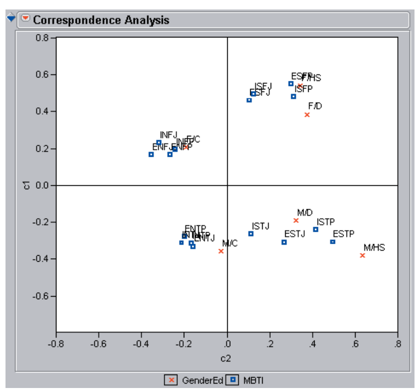
Contingency Table

Cheese	Count / Total % / Col % / Row %	1	2	3	4	5	6	7	8	9	
A	Count	0	0	1	7	8	8	19	8	1	52
	Total %	0.00	0.00	0.48	3.37	3.85	3.85	9.13	3.85	0.48	25.00
	Col %	0.00	0.00	5.26	25.93	19.51	28.57	48.72	32.00	8.33	
	Row %	0.00	0.00	1.92	13.46	15.38	15.38	36.54	15.38	1.92	
B	Count	6	9	12	11	7	6	1	0	0	52
	Total %	2.88	4.33	5.77	5.29	3.37	2.88	0.48	0.00	0.00	25.00
	Col %	85.71	90.00	63.16	40.74	17.07	21.43	2.56	0.00	0.00	
	Row %	11.54	17.31	23.08	21.15	13.46	11.54	1.92	0.00	0.00	
C	Count	1	1	6	8	23	7	5	1	0	52
	Total %	0.48	0.48	2.88	3.85	11.06	3.37	2.40	0.48	0.00	25.00
	Col %	14.29	10.00	31.58	29.63	56.10	25.00	12.82	4.00	0.00	
	Row %	1.92	1.92	11.54	15.38	44.23	13.46	9.62	1.92	0.00	
D	Count	0	0	0	1	3	7	14	16	11	52
	Total %	0.00	0.00	0.00	0.48	1.44	3.37	6.73	7.69	5.29	25.00
	Col %	0.00	0.00	0.00	3.70	7.32	25.00	35.90	64.00	91.67	
	Row %	0.00	0.00	0.00	1.92	5.77	13.46	26.92	30.77	21.15	
		7	10	19	27	41	28	39	25	12	208
		3.37	4.81	9.13	12.98	19.71	13.46	18.75	12.02	5.77	

A JMP contingency table.

Correspondence Analysis Graphs

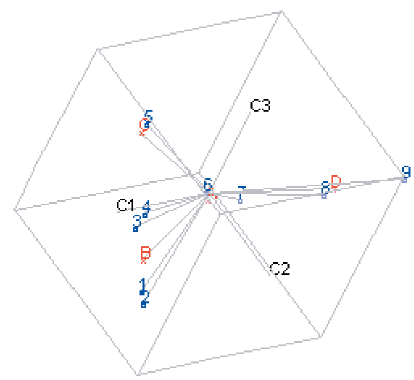
In situations where there are many different levels of *x* and you need to see what the differences are, JMP lets you request a correspondence analysis graph, a technique championed by the French for showing relationships of rows and columns in a contingency table.



A JMP correspondence analysis graph.

Each point represents a row and column in the data table. Row points that are close together represent rows that have similar profiles. The technique is symmetric, enabling you to see column profile similarities. This correspondence analysis plot shows the relationship in a way that is easier to read than a mosaic plot.

Additionally, you can save the correspondence coordinates and display a three-dimensional spinning plot to further explore the correspondences.



Three-dimensional spinning plot.

Test Statistics

JMP also provides test statistics that support the graphical presentations. For example, Chi-square tests of marginal homogeneity are reported in text reports in JMP. This test is equivalent to the test of independence if both variables are considered responses.

Freq: Count

Tests

Source	DF	-LogLike	RSquare (U)
Model	24	84.38105	0.1963
Error	176	345.51986	
C. Total	200	429.90090	
N	208		

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	168.762	<.0001
Pearson	162.482	<.0001

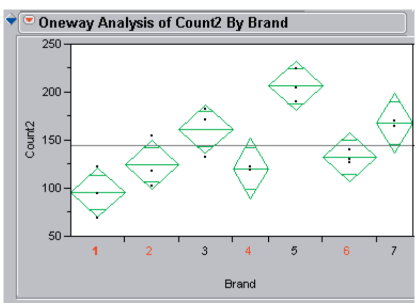
Warning: 20% of cells have expected count less than 5, ChiSquare suspect

JMP contingency analysis test statistics.

The Contingency Analysis platform also offers the Cochran-Mantel-Haenszel test for testing if there is a relationship between two categorical variables after blocking across a third classification.

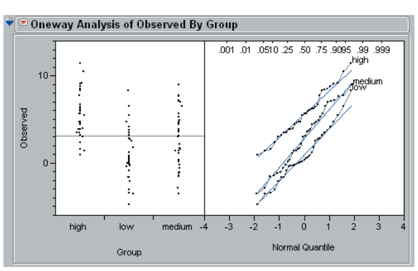
Oneway Layouts

The richness of JMP lies in its ability to deeply address a set of questions and the vast amount of insight you can gain in the results. Running a oneway analysis demonstrates this richness. If you have a continuous response from several groups, you can use JMP's to graphically show the estimation of means and quantiles. It shows the group means using diamonds whose centers are at the group means. The diamonds' top and bottom ends show the confidence intervals, and their widths show the sample size. Quantile and outlier box plots show quantiles and outliers, respectively.



The diamonds' height show confidence intervals and the widths reflect sample size.

When you have large amounts of data, normal quantile plots often reveal more differences among groups. The following figure shows a oneway analysis of 1,000 points with points "jittered" in the dot plot on the left.



A normal quantile plot.

Comparison Procedures

As a researcher, you might also want to know if all the means are the same. You would use the analysis of variance report with the *F*-test, then further clarify which means are different by looking at comparison circles. JMP pioneered comparison circles, a graphical technique that shows the significant relationships on the scale of the response for unequal sample sizes. Comparison circles help you determine which means are different.

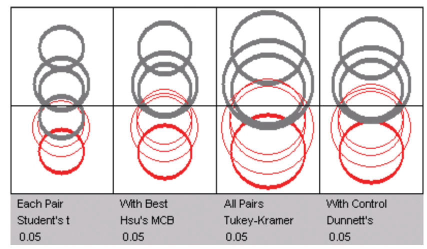
Since determining which means are different could involve many comparisons, there are many ways to control for the level of the tests using various multiple comparisons procedures. JMP allows you to choose among several procedures.

- To look at all the comparisons, JMP offers the Tukey-Kramer HSD approach. The circles are larger with multiple comparisons because the test guards against falsely declaring significant differences among more comparisons.
- To determine if a mean is not the maximum (or minimum) in the group, JMP offers the Hsu MCB test.
- To compare many treatments against one control group, JMP offers Dunnett's test. This is done using the actual sample sizes rather than a harmonic mean approximation.

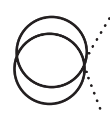
If you are unwilling to assume normality in researching if all the means are the same, JMP offers three different nonparametric tests: Wilcoxon (Kruskal-Wallace, Mann-Whitney), Median, and van der Waerden rank tests. Similarly, if you are unwilling to assume equal variances, there are four tests for homogeneity of variance.

And, if those tests are significant, you can examine a Welch ANOVA.

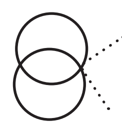
JMP also provides options to calculate the Least Significant Number (LSN), Least Significant Value (LSV), and power analysis.



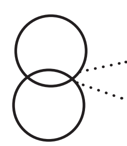
If angle > 90°, not significantly different



If angle = 90°, boardline significantly different



If angle < 90°, significantly different

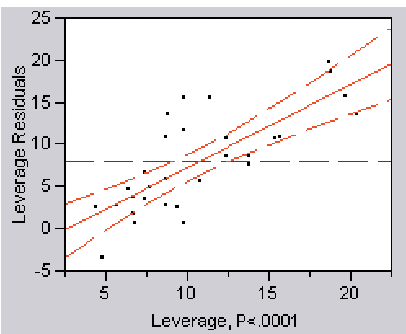


Effect Exploration: A Way to Discover More

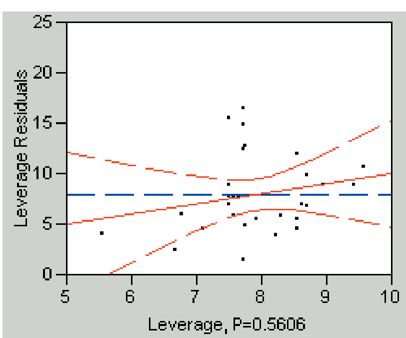
If you are fitting a linear model, such as a regression or analysis of variance model, you might want to investigate many details about each effect in the model.

What makes JMP complete is that it can answer questions other statistical software packages do not—and it does so in a deep, approachable way using leverage plots.

The following two graphs show a leverage plot for a significant effect and a non-significant effect.



Leverage plot showing significant effect.



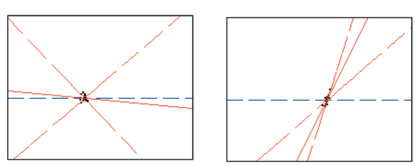
Leverage plot showing non-significant effect.

About JMP's Leverage Plots

Leverage plots show how each point contributes to the hypothesis test statistic. You can see which points support the test statistic. You can glance at the plot and understand if your effect is significant. For each observation, you can see the residual and what the residual would be if you removed that effect from the model.

The distance from each point to the sloped straight line is the residual. The distance from each point to the horizontal line is what the residual would be if that effect were removed from the model. The difference in sums of squares of these two sets of residuals is the hypothesis sums of squares for the *F*-test. The dashed lines show significance at the 0.05 level; they cross the horizontal line if the effect is significant and straddle the horizontal line if the effect is not significant.

In a multiple regression setting, leverage plots help diagnose collinearity by how much the points shrink horizontally to the center of the plot. The following two leverage plots from the famous Longley data set show how collinearity manifests itself in a leverage plot.

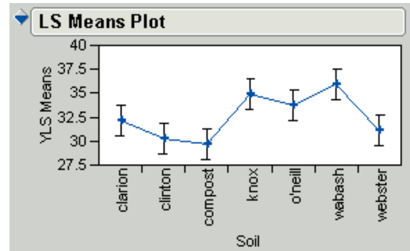


Leverage plots showing collinearity.

Categorical Effects

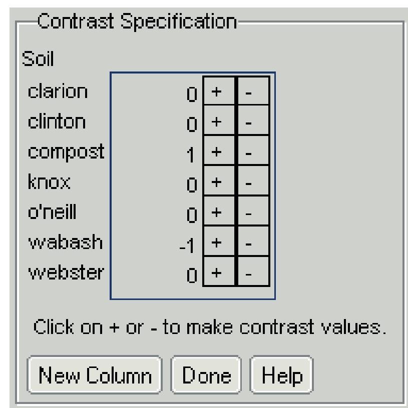
With categorical effects, the main interest is usually in comparing the levels. JMP automatically computes least-squares means for all the levels.

JMP shows these means in a profile plot.



Least-squares means plot.

You can also easily view specific contrasts. JMP computes the contrasts and shows the results.



Linear contrasts.

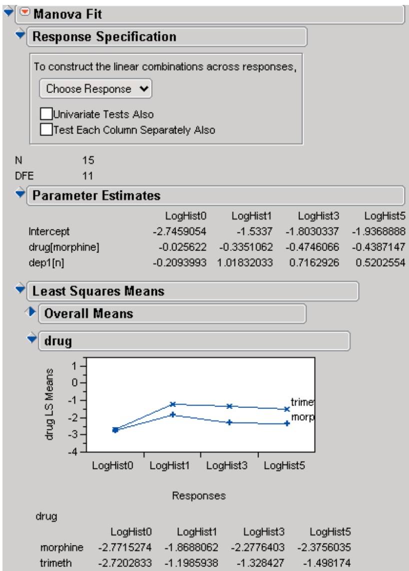
Furthermore, a single mouse click produces a multiple comparisons report, showing details of which levels actually differ.

Level					Least Sq Mean
wabash	A				35.966667
knox	A				34.900000
o'neill	A	B			33.800000
clarion		B	C		32.166667
webster			C	D	31.100000
clinton			C	D	30.300000
compost				D	29.666667

Multiple comparison report.

MANOVA

When many responses are fit to the same linear model, the analysis can be approached as a MANOVA (Multivariate Analysis of Variance), which considers relationships across responses.



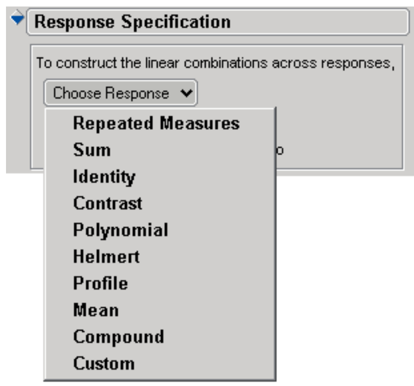
The initial fit in a JMP MANOVA analysis.

One common use of multivariate fitting is analyzing repeated measures data. A subject is measured repeatedly across time and the data are arranged so each time measurement forms a variable.

Take, for example, 16 dogs that are given different drugs. The dogs are assigned to two groups defined by the variable "drug" and the dependent variable is the logarithm of the blood histamine concentration at zero, one, three, and five minutes after injection of the drug. To analyze the data using JMP, you run a fit model analysis. JMP fits the responses to the effects using least squares. The initial fit shows parameter estimates and least squares means estimates.

It also graphically displays the least squares means. You can quickly see that the histamine level of dogs given trimeth is greater than that of the dogs given morphine.

The next step in this example is to choose the response specification to use in the MANOVA. In JMP, you can select the form of the linear combinations across responses.

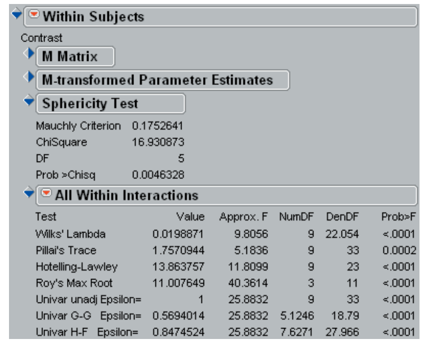


Response specifications.

There are several common pre-set combinations, or you can customize your own.

If you choose the repeated measures linear combination, JMP models two response functions: the between-subjects analysis using the sum across the *y*'s and the within-subjects analysis using the contrast across the *y*'s. You can also construct contrasts or save predicted values, residuals, and canonical scores (with the discriminant formula) as columns in the original data table.

JMP also provides access to a Sphericity Test table and adjusted univariate *F*-tests in the multivariate report tables.

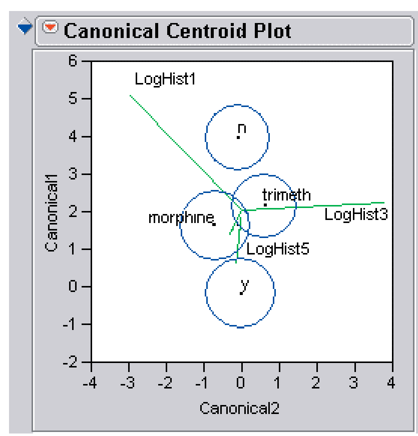


A Sphericity test table and adjusted univariate *F*-tests in the multivariate report tables.

Centroid Plots

To show the multivariate least-squares means, JMP provides a centroid plot. The centroid plot shows the centroids (multivariate least-squares means) on the first two canonical variables formed from the test space.

In the example below, the least squares means are shown for each nominal variable with biplot rays showing the directions of the original variables.



Centroid plot.

95% confidence ellipses surround each least squares mean. Since they are plotted in canonical space, they appear as circles.

Stepwise Regression Models

Sometimes you need a way to wade through a large number of effects (some with many levels) and determine which ones are most important. JMP's stepwise regression methods do just that. You give JMP all the effects you have, and it proceeds with a step-by-step examination of regression models composed of different subsets of variables and advises you on the most important effects.

Types of Stepwise Regression

JMP implements three stepwise techniques: forward, backwards, and mixed.

At any time, you can click a check box to bring a variable into the model or remove it from the model. You set a direction, then take one step at a time by clicking the **Step** button. A Current Estimates report shows the state of the model.

You can force a term to be entered or removed from the model, or lock it so that it is no longer considered. Click the **Go** button to proceed through the steps automatically.

When you are finished, click the **Make Model** button to get a complete analysis.

Categorical Effects

Categorical variables with more than two levels are a special case. Most stepwise regression software handles only continuous variables, but JMP forms indicator variables for levels of categorical variables. Using special rules, levels (or combinations of levels) can form new columns and be included in the model.

Compound Effects

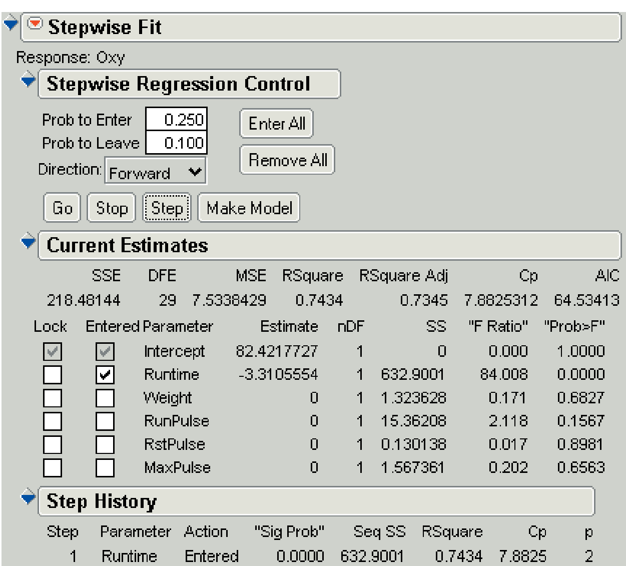
If you have crossed or polynomial terms, you can force the model's steps to proceed according to hierarchy rules.

Stepwise Logistic Regression

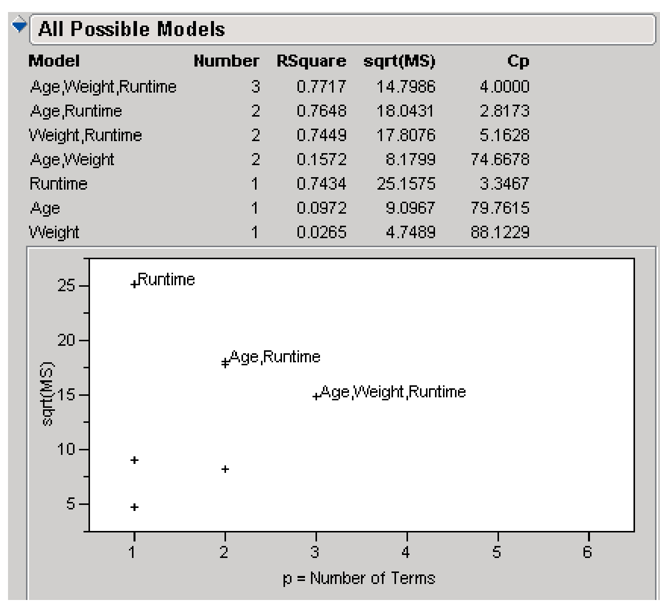
You can also run a stepwise regression if you have a binary or ordinal response. This regression works similarly to the least-squares version.

All Possible Models

JMP takes advantage of modern computer speed, so now it is feasible to examine all possible combinations of candidate variables that may be useful in prediction. As with other analyses in JMP, the results in a stepwise regression are portrayed graphically and in text reports.



The report after one click of the Step button.

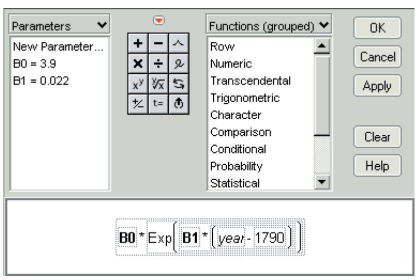


A stepwise fit run on a binary or ordinal response.

Nonlinear Modeling

Models that are nonlinear in their parameters are more difficult to fit than linear models. Unfortunately, there is no foolproof method for estimating nonlinear regression parameters that guarantees a globally-optimal solution.

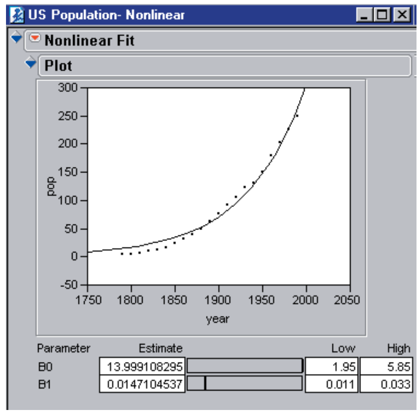
Entering the model to be estimated is easy in JMP—you enter a formula into a column using the formula editor. Derivatives are calculated automatically. Once you enter the formula and the initial values for the parameters, JMP does the rest.



The formula editor in JMP.

The great thing about JMP's nonlinear platform is its interactivity. At first, you might try some starting values and it does not converge. To try other values, you enter different values, click **Reset**, then click **Go**. You can lock some parameter and estimate the rest. The grid and slider features interact in real time. Because nonlinear may not work the first time, it is important that exploring be interactive and easy.

In one single step, you can stop the iterations at any point, examine the values, change the parameters, or restart the iterations. Because the iterations run in the background, you can continue with other analyses while the nonlinear regression runs.



The slider at the bottom adjusts parameters.

JMP's Modeling Features

JMP also handles the features listed below differently than other software packages.

Derivatives

A troublesome part of using most nonlinear regression software is often in programming the derivatives. But in JMP, after you enter the prediction formula, the first and, if desired, second derivatives are symbolically computed whenever possible by the software. If JMP is unable to symbolically compute the exact expressions for the derivatives, it computes finite-difference derivatives instead. JMP supports the Newton-Raphson/Gauss-Newton/ Marquardt methods.

Confidence Intervals

JMP offers confidence intervals based on searching the likelihood surface for each parameter, whereas other nonlinear regression programs give you a confidence interval or standard error for each parameter that is based on approximating the nonlinear surface with a linear surface.

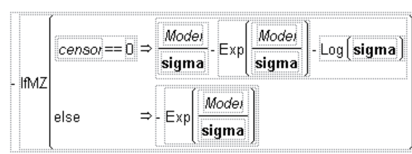
Plots of the Fitted Function

To get a plot of the fitted function in most packages, you have to use a separate plotting procedure designed to create scatterplots of individual points instead of smooth functions. In JMP, if your model is a function of one variable, the software automatically produces a plot of the data and fitted function. The function is plotted smoothly, not just at the observed values. The points are interactive, and the plot can be customized using the slider scale at the bottom of the report.

Maximum Likelihood

In addition to least squares fitting, the nonlinear platform can also minimize with respect to a user-specified loss function, which can have additional parameters to estimate.

You can maximize likelihood fitting by specifying the loss function so that its sum is the negative log likelihood. For example, to use nonlinear for a censored Weibull survival model, you could specify a loss function such as:



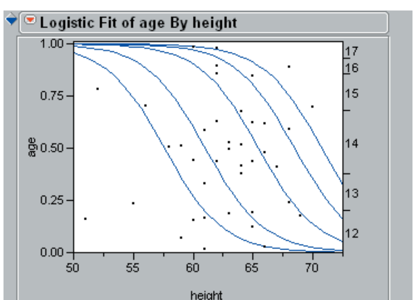
Model Library

With the click of a button, you can conveniently set up nonlinear regression model columns using JMP's model library. The library helps you quickly choose, construct, and save the appropriate model for a set of predefined models.

Ordinal Regression

Logistic regression models the probabilities of a nominal response as a continuous function of a continuous x *variable. The method 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 one. If* x *does not affect the distribution of* y*, then the slope parameters tend to be non-significant, and the lines of fit tend to be horizontal.*

The ordinal cumulative logistic regression technology in JMP is able to handle a large number of response levels. This technology models ordinal responses similarly to how it models nominal responses, but the ordinal response 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.



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	11.160573	1	22.32115	<.0001
Full	56.105777			
Reduced	67.266350			

RSquare (U)	0.1659
Observations (or Sum Wgts)	40

Converged by Gradient

Parameter Estimates

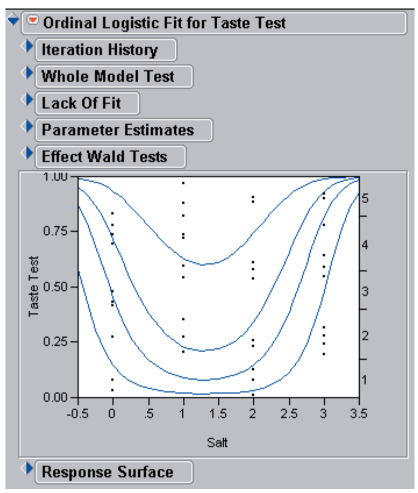
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	23.5438501	6.0933968	14.93	0.0001
Intercept	24.948157	6.2145054	16.12	<.0001
Intercept	26.7625977	6.3879331	17.55	<.0001
Intercept	27.9892409	6.4905911	18.60	<.0001
Intercept	28.9204074	6.5543938	19.47	<.0001
height	-0.4089295	0.0996045	16.86	<.0001

An ordinal logistic fit of a person's age by his height.

You can graphically see the probability that a given height is in an age group by looking at the distance between curves.

An Example

Suppose a popcorn manufacturer wants to maximize the probability of a favorable taste response as a function of how much salt is added. Imagine an experiment with the salt controlled at zero, one, two, and three teaspoons. Respondents could rate taste on a scale of one to five, where five is a positive taste. In this case, the manufacturer would use JMP to build a model that fits the ordinal taste test to the surface effect of salt.



A graph showing how the quadratic model would fit the response probabilities.

Instead of being shifted logistic curves, the lines in the figure above are a folded pile of curves, where each curve achieves its optimum at the same point. That point, the critical value, is 1.29 teaspoons of salt. The probability for the highest response level is the distance from the top curve to the top of the plot rectangle. This distance reaches a maximum when salt is 1.29 teaspoons, where the curve bottoms out. All curves share the same critical point. This combination of ordinal logistic regression technology with response surface technology is unique to JMP.

Binary Outcomes

When there are only two outcomes, JMP has some extra capabilities for exploring the fit.

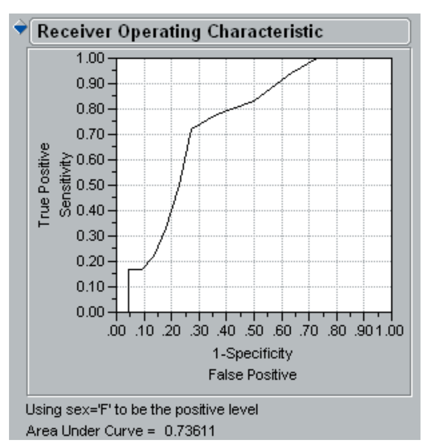
The **Inverse Prediction** option enables you to request an exact inverse prediction. If there is a strong relationship between the variables, the inverse prediction table displays upper and lower confidence limits for the inverse prediction.

Inverse Prediction

Probability	Predicted height	Lower Limit	Upper Limit	1-Alpha
0.50000000	61.5254338	22.34139	67.0376482	0.9500
0.90000000	50.5057971	-331.90144	56.9724139	

An inverse prediction table showing an analysis of a person's gender as predicted by their height.

You can also request an ROC (Receiver Operating Characteristic) curve for your analysis. JMP allows you to specify which level is "positive," then produces the curve.



A requested ROC curve.

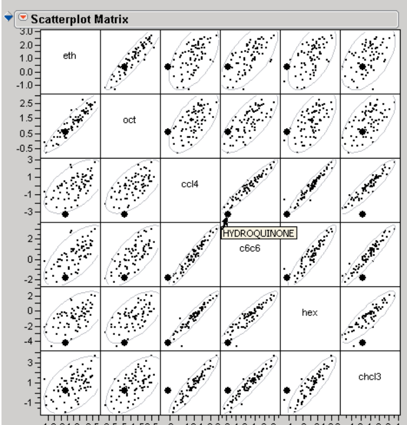
Multivariate Analyses

Are you missing patterns among your variables? JMP presents correlations in graphical ways so that you can easily see patterns that may be missed by looking at text reports.

JMP shows multivariate correlations using a scatterplot matrix to support a table of Pearson correlation coefficients. The scatterplots show the bivariate relationships between each pair of variables.

Correlations

	eth	oct	ccl4	c6c6	hex	chcl3
eth	1.0000	0.9293	0.5452	0.6495	0.5553	0.5114
oct	0.9293	1.0000	0.6153	0.7172	0.5992	0.5960
ccl4	0.5452	0.6153	1.0000	0.9578	0.9507	0.9224
c6c6	0.6495	0.7172	0.9578	1.0000	0.9092	0.9411
hex	0.5553	0.5992	0.9507	0.9092	1.0000	0.8719
chcl3	0.5114	0.5960	0.9224	0.9411	0.8719	1.0000



Multivariate analyses produce a table and graph.

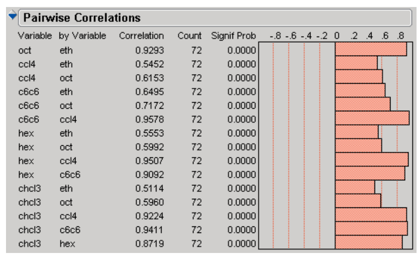
The normal density ellipses show the correlation visually.

- High correlations correspond to ellipses that are narrow across the diagonal.
- Small correlations lead to more circular ellipses.
- Negative correlations show up as ellipses with a negative slope.

As with all the other plots in JMP, the scatterplot matrix is connected to other plots. Highlight a point in a scatterplot matrix and it becomes highlighted in every other active plot (including those generated by other platforms). Place the cursor over the point, and its identity is revealed.

Correlation Methods

Using the red triangle's drop-down menu options, you can request tables of inverse and partial correlations, pairwise correlations listed with their probabilities and a bar chart to compare them, and three kinds of nonparametric correlation: Spearman's ρ, Kendall's τ, and Hoeffding's D.



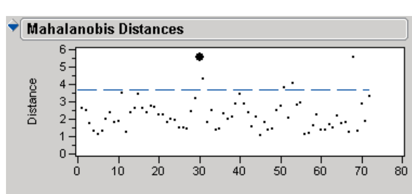
Pairwise Correlations

Variable	by Variable	Correlation	Count	Signif Prob
oct	eth	0.9293	72	0.0000
ccl4	eth	0.5452	72	0.0000
ccl4	oct	0.6153	72	0.0000
c6c6	eth	0.6495	72	0.0000
c6c6	oct	0.7172	72	0.0000
c6c6	ccl4	0.9578	72	0.0000
hex	eth	0.5553	72	0.0000
hex	oct	0.5992	72	0.0000
hex	ccl4	0.9507	72	0.0000
hex	c6c6	0.9092	72	0.0000
chcl3	eth	0.5114	72	0.0000
chcl3	oct	0.5960	72	0.0000
chcl3	ccl4	0.9224	72	0.0000
chcl3	c6c6	0.9411	72	0.0000
chcl3	hex	0.8719	72	0.0000

Table of pairwise correlations.

Outlier Analysis

The correlation structure can be used to help discover points that are outliers in a high number of dimensions. Visualize a series of nested ellipses corresponding to many normal density contours. The number of ellipses you cross when moving from the multivariate mean to a data point is a distance measure called the Mahalanobis distance. This distance can be calculated for any number of dimensions. The figure below is JMP's plot of the Mahalanobis distances, used in multivariate outlier analysis. Points above the dotted line exceed the threshold for outliers.

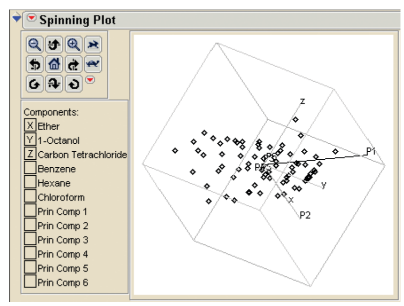


The Mahalanobis distance is a measure to determine outliers.

Principal Components

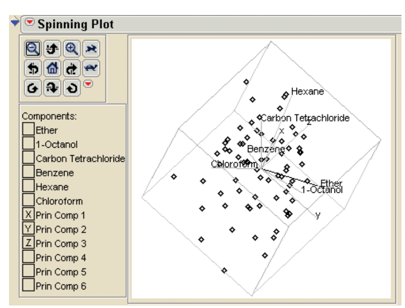
Another way to examine relationships when there are more than two variables is with principal components. Principal components are a set of variables computed as linear combinations of the original variables. They organize the total variability so it can be represented with fewer variables.

JMP can calculate principal components on correlations, covariances, or unscaled (raw) data. It reports eigenvalues so you can determine how many components are necessary to account for a large amount of variability.



A spinning plot showing the principal components plotted with three of the original variables.

For three components at a time, JMP can even produce a spinning biplot showing the orientation of the original variables in principal component space. Just drag the axis markers on the left of the plot to reassign their roles.



This biplot shows the original variables' orientation.

Cluster Analysis

Clustering groups the points (rows) into clusters whose values are close to each other relative to those of other clusters.

JMP provides five different clustering methods:

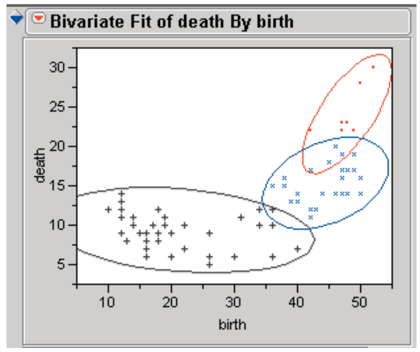
- hierarchical clustering
- *k*-Means clustering
- normal mixtures clustering
- Self-Organizing Maps (SOMs)
- modal clustering (bivariate only)

Hierarchical Clustering

To show a visual representation of the cluster's process, JMP provides dendrograms, which are tree diagrams that display the results of a cluster analysis. The dendrogram is linked to its original data: if you click in the dendrogram to highlight points in a cluster, the corresponding points highlight in all other open analyses and data tables. See the figure on the right for an example of hierarchical clustering using dendrograms.

You also can identify a specific number of clusters by clicking and dragging a small diamond dotted icon at the top and bottom of the dendrogram. Optionally, JMP can automatically assign colors and markers based on the number of clusters assigned in the dendrogram. JMP will also let you save the cluster number assigned to each data point as a new column in the data table.

With sufficient memory in your computer, you can cluster thousands of rows. The color map, showing the values of every cell of data, identify which clusters are associated with low or high values of which columns. You can also cluster columns to obtain a two-way clustering.



This plot's density ellipses show clustering.

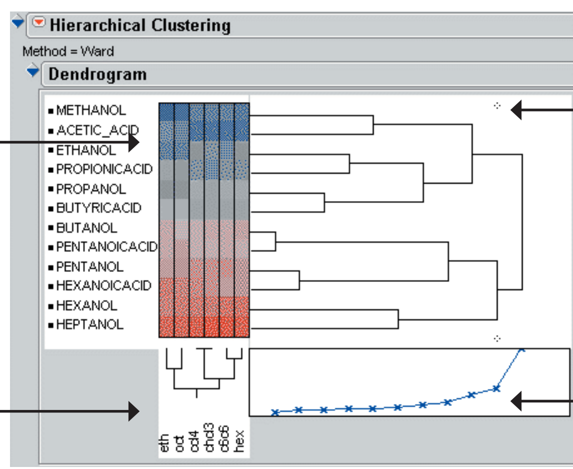
K-Means Clustering

The *k*-means approach to clustering is through an iterative alternating fitting process (a special case of the *EM* algorithm). The *k*-means approach is well-suited to larger data tables.

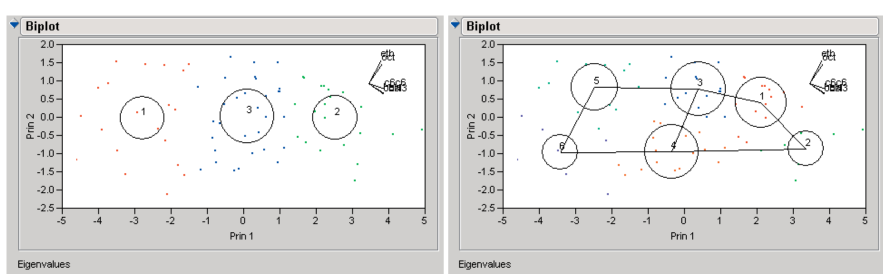
In addition to standard *k*-means, JMP offers Self-Organizing Maps (SOMs). They form clusters on a cluster grid so that the points near each other in the grid are also near each other in mulivariate space. The grid structure may be invaluable in interpreting the clusters in a two dimensional sense.

Normal Mixtures

Normal mixtures is similar to *k*-means, but it estimates probabilities of group membership instead of each point being a strict member. It is especially useful in cases where the clusters overlap. In these cases, normal mixtures clustering provides a far more accurate estimate of the total population in each group.



Hierarchical clustering using dendrograms.



Biplot of regular *k*-means clustering (left). Biplot of Self-Organizing Maps (SOMs) (right).

Recursive Partitioning

JMP recursively partitions data according to a relationship between the x *and* y *values, creating a tree of partitions. The technique (also called decision trees) is often taught as a data mining technique because*

- it is good for exploring relationships without having a good prior model
- it handles large problems easily
- the results are easily interpretable

JMP provides a unique and useful decision tree graph that shows the partition clearly.

A classic application of partitioning is when you want to turn a data table of symptoms and diagnoses of a certain illness into a hierarchy of questions to ask new patients in order to make a quick initial diagnosis.

The factor columns *(x's)* can be either continuous or categorical (e.g., nominal or ordinal). If an *x* is continuous, then the splits (partitions) are created by a cutting value. The sample is divided into values below and above this cutting value. If the *x* is categorical, then the sample is divided into two groups of levels.

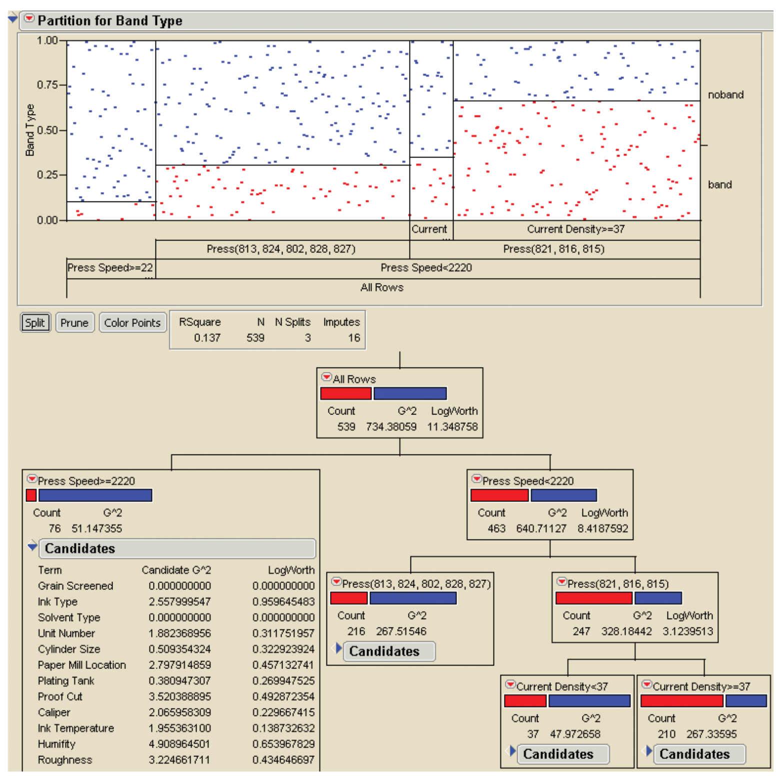
The response column *(y)* can also be either continuous or categorical (and nominal or ordinal). If *y* is continuous, then the platform fits means and creates splits which most significantly separate the means by examining the sums of squares due to the mean differences. If *y* is categorical, then the response rates (the estimated probability for each response level) become the fitted value, and the most significant split is determined by the largest likelihood-ratio Chi-square statistic. In either case, the split is chosen to maximize the difference in the responses between the two branches of the split.

An Example

The following example and figure show the Partition platform using data from Evans and Fisher[1]. In this example, high-speed printing press operators were interested in causes of "banding"—strips on the final product where the plate did not produce an image. To explore, they followed these steps:

1. They gathered data on over forty variables, including ink type, press speed, press type, ink color, and many more.
2. The first time the Split button is pressed, JMP splits the effects by press speed. This shows that press speeds less than 2,220 rpm have a higher chance of producing banding. It tells the analysts that press speed affects banding more than any of the other variables.
3. A second click of the **Split** button splits by press type. It seems that some brands of press are more likely to produce banding than others.
4. The next split is on the variable that represents density of the paper.

Note that the entire process in JMP is interactive. You control when and where the variables are split. You evaluate when the analysis is complete.



An example of the Partition platform in action.

1 Evans, B. & Fisher, D. (1994). "Process delay analysis using decision tree induction." *IEEE Expert*, 9, 1, 60-66.

Three-Dimensional Plots

JMP has three ways of visualizing three variables at a time: through spinning plots (3-D scatterplots), surface plots, and contour plots.

Spinning Provides Visualization

You can spin JMP's three-dimensional scatterplots in real time so data can be viewed from any desired aspect with one mouse movement.

In spinning plots, you can also

- show and hide the axes
- change the size of the points, background color, and depth cueing
- overlay a box that makes seeing axis direction easy
- add rays and variables
- save principal and rotated components
- choose a different set of three variables by clicking and dragging the *x*-, *y*-, and *z*-axes

If you highlight a point in a spinning plot, JMP highlights its corresponding data everywhere else it is represented. You can also select the brush tool and click and drag it to select a group of points. As the rectangle encloses the points, they highlight in the plot and everywhere else they appear.

Principal Components and Biplots

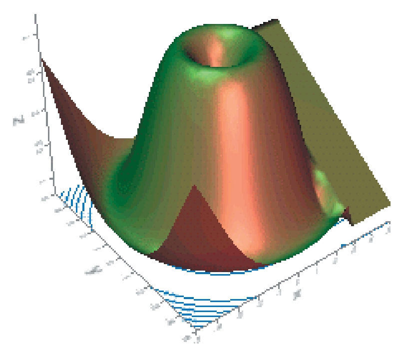
You can't have a six-dimensional plot, but if the variables are fairly correlated, the principal components can capture most of the six-dimensional variation in a smaller number of components.

JMP allows you to display principal components in a biplot; JMP computes the principal components and displays them in a three-dimensional plot. You can also manipulate the plot and see it from the components' point of view. This gives visual feedback on what the components represent. The plot shows rays representing and approximating the variables in the principal component space.

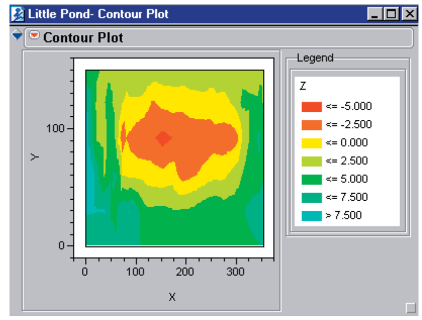
Surface Plots

You can plot points or surfaces using surface plots. JMP displays surfaces either smoothly or as a mesh, with or without contour lines. You are given the opportunity to customize the labels, axes, and lighting.

Additionally, the Surface Plot platform functions both as a separate platform and as an option in model fitting platforms. And, if you have columns that contain formulas, you can plot the formulas on a surface plot.



Example of a surface plot in JMP.



A contour plot of water depth in a pond.

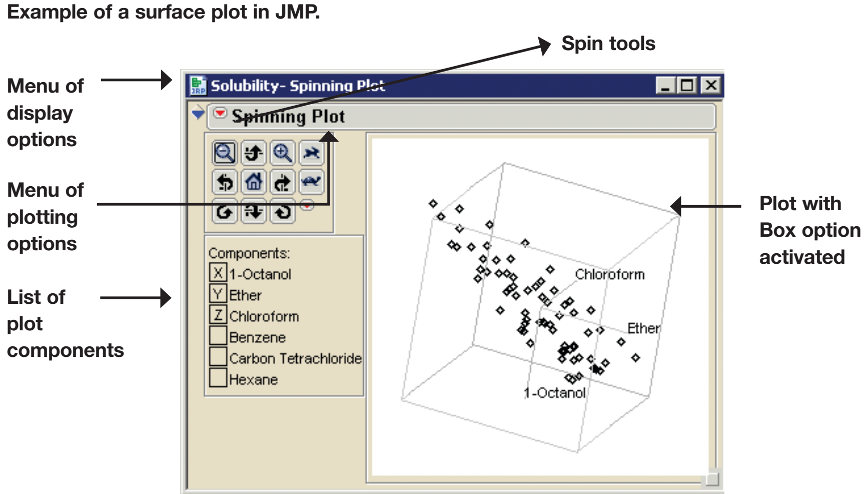
Contour Plots

You can also visualize three-dimensional data using JMP's contour plot. Two variables are plotted on the horizontal and vertical axes, and colors represent the third variable.

JMP can use any set of *x* and *y* points—they don't have to be a precise grid. The plot below shows a contour of the water depth in a pond. The *x*- and *y*-values are coordinates at which the water depth was measured.

You can customize the contour plot to

- show or hide data points, contours, and boundaries
- fill areas
- reverse colors
- label, change, and save contours



JMP's three-dimensional spinning plot.

Linear Model Creation

JMP uses a general computational approach that works for any linear model—balanced or unbalanced, complete or with missing cell values, with main effects, with interactions, or with nested terms. The following are unique ways JMP gives insight on linear models.

Automatic Lack-of-Fit Test

JMP looks for multiple observations with the same factor values, and if it finds them and the model is not saturated, it calculates the pure error and automatically produces a lack-of-fit test.

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	28	5482.4324	195.801	15.3069
Pure Error	4	51.1667	12.792	Prob > F
Total Error	32	5533.5991		0.0083
				Max RSq
				0.9973

Example of a lack-of-fit test in JMP.

Effective Hypothesis Test

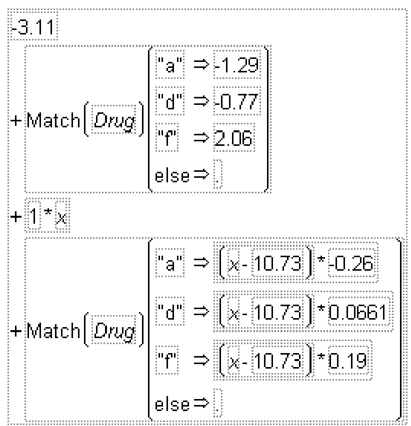
If a model has no missing cells, JMP produces a hypothesis test, such as SAS PROC GLM's Types III and IV, as well as Type I sequential tests. But when there are missing cells, instead of giving up or hunting for subsets of cells to use for testing main effects, JMP uses R. R. Hocking's "effective hypothesis test." This brings the tests into a close relationship with the least squares means: testing an effective hypothesis for main effects is equivalent to testing that the least squares means are equal. Testing an effective hypothesis for main effects is unique and unaffected by reordering the levels (unlike Type IV sums of squares) or reordering the effects (unlike Type I sums of squares).

Polynomial Centering

When you have polynomial terms in a model, such as squares and cross products of continuous regressors, it is sometimes difficult to produce good test statistics for them; the main effect tests are not meaningful in the presence of effects that contain them. But with JMP, polynomial terms are automatically centered by the mean so that tests on main effects are meaningful. You can also turn off the centering.

Formula Output

Predicted values can be saved into a JMP data table as a formula, not just as data. This allows you to compute predicted values as new rows are added.



Predicted values can be saved into a data table as a formula, not just as data.

Inverse Prediction

For nominal or ordinal models with two response levels, JMP computes an *x*-value that results from a specified response (*y*) probability. It also fiducially gives confidence limits for the inverse prediction.

Ordinal Coding Option

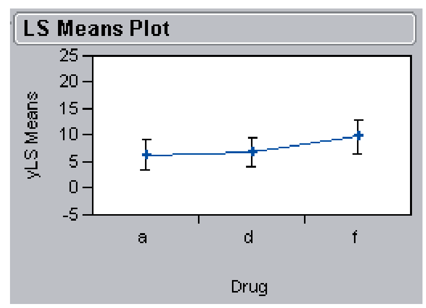
If you have dose treatment effects, you can use JMP's ordinal parameterization for better interpretability of the parameters. If there are interaction effects, the ordinal coding option gives further advantages because the main effects better tolerate missing cells that are located away from the baseline level.

Parameterization

JMP uses a parameterization that is more compact and more interpretable than full design matrices. Each JMP parameter measures its level against the average of all levels. This parameterization can be much more efficient than those of other software packages. For example, a 2^5 factorial in JMP takes less than one-seventh the number of parameters than parameterization with one parameter for each level in each effect.

Least Squares Means Plots

When you want to see main effect or interaction graphs, JMP produces a profile plot of the least-squares means.



Interaction plots show main effects.

Random Effects and Tests

JMP supports the Kackar-Harville and Kenward-Roger improvements in REML estimates, which produce the best test in a wide variety of situations. The result is an automated general method for all cases that gets the best possible answer with a minimum of user expertise. You declare which effects are random, and everything else is automatic.

This automated method gives correct answers without the requirements of balanced design. The results are correct even if whole plot fixed effects are purely nested in whole plot random effects.

Scripting

Advanced JMP users often want to automate repetitive analyses. Some also want to create custom analyses and animate displays for use in instructional situations.

To address these needs, JMP contains a complete scripting language—JMP Scripting Language (JSL). It is both easy to learn and powerful. In fact, JMP can write some of these scripts itself.

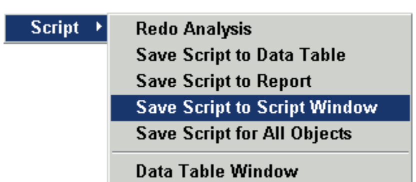
Scripting Features

Some highlights of JSL include

- all of JMP's analysis platforms are scriptable
- the display surface of JMP is scriptable, so custom reports have the look and feel of native JMP reports
- data tables can be read, written, and manipulated through scripting
- dialog boxes and platform dialogs can be integrated into a script to obtain input from users

Automatic Creation of Scripts

Every report window contains a script menu, as shown below.



JMP script menu.

After running an analysis, JMP can write a script that duplicates the steps involved in arriving at the report.

For example, suppose you need to combine the capabilities of two JMP platforms into your own custom platform. JSL makes this easy:

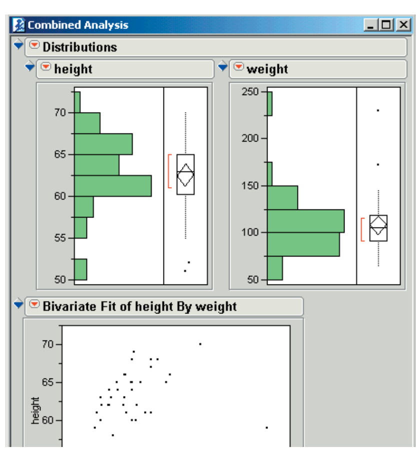
The first step in scripting is to produce the analyses.

1. Produce the desired output. For example, the following figure shows output produced by choosing **Analyze > Distribution,** then **Analyze > Fit Y by X.**

2. Click the triangle icon and select Save Script to Script Window in both reports. This creates a script that duplicates the steps involved in arriving at the report.

3. Type a New Window command around the generated scripts so they appear in the same window:

```
New Window(“Analysis”,
Distribution(Continuous
Distribution(Column( :
height), Quantiles(0),
Moments (0)),
Continuous
Distribution(Column( :
weight), Quantiles(0),
Moments (0))); Bivariate(Y(
:weight), X( :height));
```



The New Window scripting command places analyses in the same window.

On Windows, you can then copy and paste this script to a menu or toolbar icon (using **Edit > Customize > Menus/Toolbars**) for others to use.

Creating Custom Analyses

JMP's scripting language includes an extensive list of operators on matrices. In addition, some of JMP's platforms expose their internal matrices for use by scripts. Combining these capabilities allows you to add complex analyses that are not included in JMP's standard package.

Furthermore, you can translate data tables into matrices so manipulations are transparent to the script user.

Loops, conditionals, logical operations, random number generators, and probability functions are among the tools available to custom analyses.

Instructional Simulations

The JMP CD includes several example scripts that illustrate statistical concepts. For example, the demokernal.jsl script graphically shows how kernel density estimation works. Each data point is represented by a small normal distribution. These normal distributions are summed to create the larger, red curve, which represents the kernel density function. You can adjust the spread of the individual distributions— and therefore the smoothness of the spline fit—by clicking and dragging the handle. Furthermore, you can script the drawing of graphics and capture clicks and drags, allowing for animated output.

Partial Least Squares (PLS)

Partial Least Squares (PLS) balances the two objectives of explaining response variation and explaining predictor variation. It is especially useful since it is appropriate in analyses that have more x-variables than observations.

The number of factors to extract in PLS depends on the data. Basing the model on more extracted factors (successive linear combinations of the predictors, also called components or latent vectors) optimally addresses one or both of two goals

- explaining response variation
- explaining predictor variation

In particular, PLS balances the two goals, seeking for factors that explain both response and predictor variation.

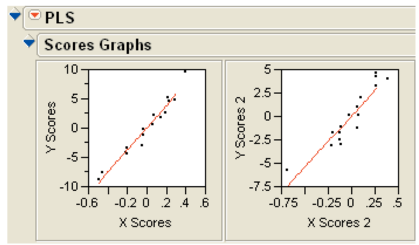
Furthermore, PLS works in cases where OLS does not. If the data have more *x*-variables than observations, OLS cannot produce estimates, whereas PLS can.

For example, researchers studying pollution in the Baltic Sea would like to use a spectra of samples of sea water to determine the amounts of three compounds present in samples from the Baltic Sea: lignin sulfonate (pulp industry pollution), humic acids (natural forest products), and detergent (optical whitener).

Spectrometric calibration is a type of problem in which partial least squares can be very effective. The predictors are the spectra emission intensities at different frequencies in sample spectrum (v1–v27), and the responses are the amounts of various chemicals in the sample. In this case, there are only 16 observations.

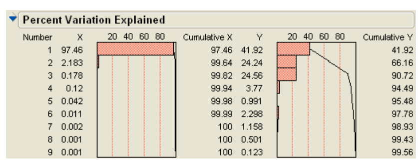
Diagnostic Graphs

JMP displays graphs that plot each *x*-variable against its corresponding *y*-variable. These graphs automatically update as your analysis continues.



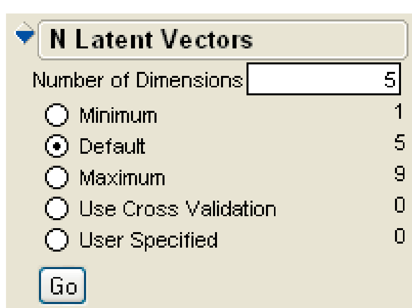
JMP provides scores plots.

In addition, graphical displays show the amount of variance explained by different numbers of components. This display helps decide on how many latent vectors are needed to characterize the data.



Percent variance is explained in the output report.

In the example above, JMP shows that most of the variation in *x* is explained by two factors, and most of the variation in *y* is explained by four. Based on this information, you can select the number of components to be analyzed.



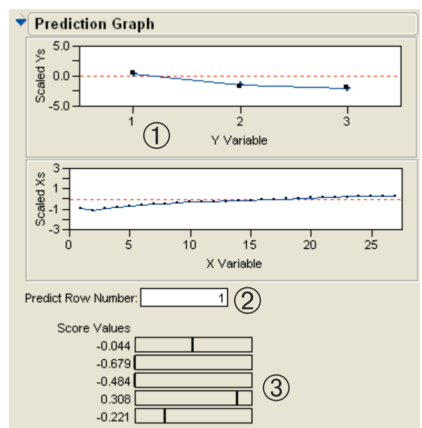
You can select the number of components to be analyzed.

Cross-Validation

You can choose the number of factors by using cross validation, in which the data set is divided into two or more groups. You fit the model to all groups except one, then you check the capability of the model to predict responses for the group omitted. Repeating this for each group, you then can measure the overall capability of a given form of the model. JMP displays the results using the prediction root mean square error for each number of factors.

Prediction Profiling

JMP also lets you interactively view your predictions.



① **Graphically view predictions for x and y based on current settings**

② **Predict specific row.**

③ **Use sliders to interactively adjust components and see their effect on predictions**

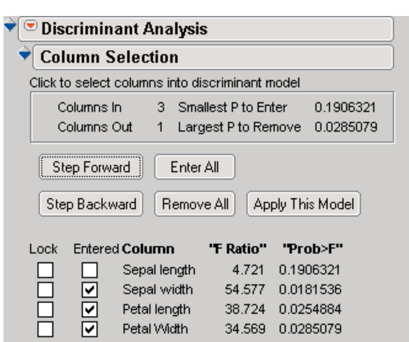
You can make predictions for a specific row in the data table. In addition, sliders appear for each of the components, which can be interactively adjusted and viewed in the prediction graph. Using this technique, you can gain insight into the underlying traits that each score represents.

Discriminant Analysis

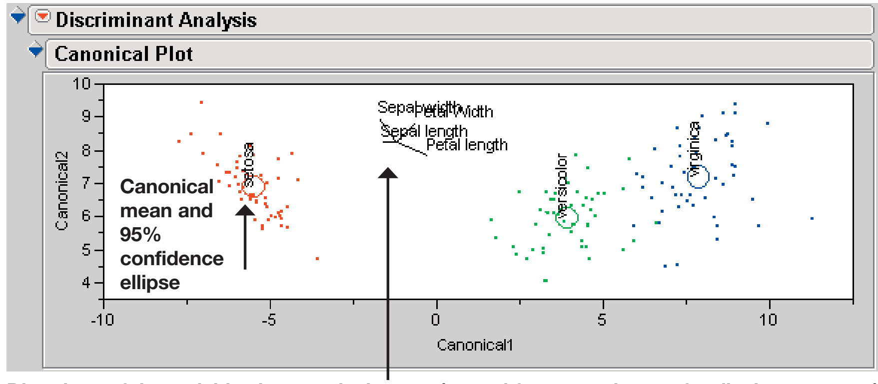
Discriminant analysis is an alternative to logistic regression. In logistic regression, the classification variable is random and predicted by the continuous variables, whereas in discriminant analysis the classifications are fixed, and the y's are realizations of random variables. However, in both cases, the categorical value is predicted by the continuous.

There are several varieties of discriminant analysis. JMP implements the simplest case, which is linear discriminant analysis. In linear discriminant analysis, it is assumed that the *y*'s are normally distributed with the same variances and covariances, but that there are different means for each group defined by *x*. The method measures the distance from each point in the data set to each group's multivariate mean (often called a centroid) and classifies the point to the closest group. The distance measure used is the Mahalanobis distance, which takes into account the variances and covariances between the variables.

Fisher's iris data set is the classic example to use for a discriminant analysis. Four measurements are taken from a sample consisting of three different species. The goal is to identify



JMP's stepwise capability after three iterations.



Directions of the variables in canonical space (moved from grand mean for display purposes)

the species accurately using the values of the four measurements. In this example, we use stepwise selection in JMP to determine the best classifiers.

When satisfied with the model, you click **Apply This Model** to see a centroid plot and reports on the discriminant scores.

Note the following features of the canonical plot shown at the bottom of the page.

- There is a 95% confidence ellipse around each multivariate mean. Since the plots are in centroid space, the ellipses appear as a circle.
- The directions of the variables in the canonical space are shown by labeled rays emanating from the grand mean.

JMP's discriminant options allow you to view the results of the analysis. For example, you can show only "interesting" rows—those that have fitted probabilities of 0.05 or those that are misclassified. You can also save the discriminant formula in your original data table and use it for prediction of data gathered in the future.

Consider New Levels

JMP can also consider new levels of the variables. Rather than classifying points into existing groups, it considers the possibility that points are from a new, unscored group.

For example, suppose you gather a new data point representing an unknown iris. Using the Discriminant platform, you can either determine the probability it is virginica, setosa, or versicolor, or determine the probability that it is from a new species of iris, different from the three that have been considered.

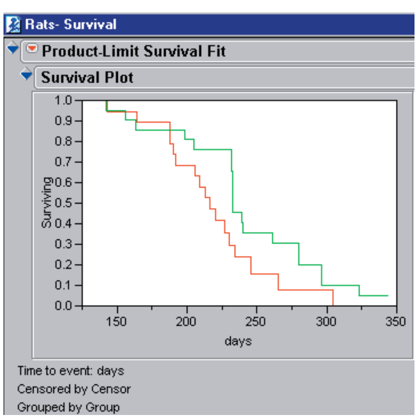
Survival Data

Survival data contains time measurements that show the occurrence of a specific event, such as the failure of an engine or a limit on what the event time would be—a censored value. JMP offers special modeling techniques to analyze left-censored, right-censored, and interval-censored observations correctly.

Survival/Reliability

You can compute life-table estimates of survival functions using the product-limit (Kaplan-Meier) method for one or more groups of right-censored data. Right-censored observations occur due either to units withdrawing from a study or termination of the study before all units have failed. All that is certain about this kind of observation is that the lifetime exceeds the final time value recorded for the observation. Although the exact lifetime of the observation is not known, the observation should not be ignored.

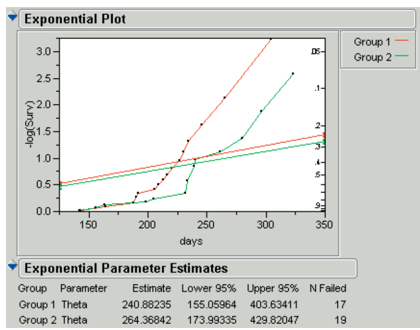
JMP's product-limit method gives a life table and an overlay plot of the estimated survival function for each group as well as the whole sample. Optionally, you can include confidence intervals for each group or a plot of the combined function.



A JMP survival plot.

JMP supports interval censoring and many graphics and estimation features by

- fitting exponential, Weibull, or LogNormal distributions
- providing plots for each fit
- providing survival, density, and hazard functions
- testing homogeneity between groups via the log rank and Wilcoxon tests
- analyzing competing causes



A JMP exponential plot.

Fit Parametric Survival

Parametric survival fits a survival distribution scaled to a linear model. You can choose from exponential, Weibull, and Log-normal. The estimation uses an iterative maximum likelihood method.

Nonlinear

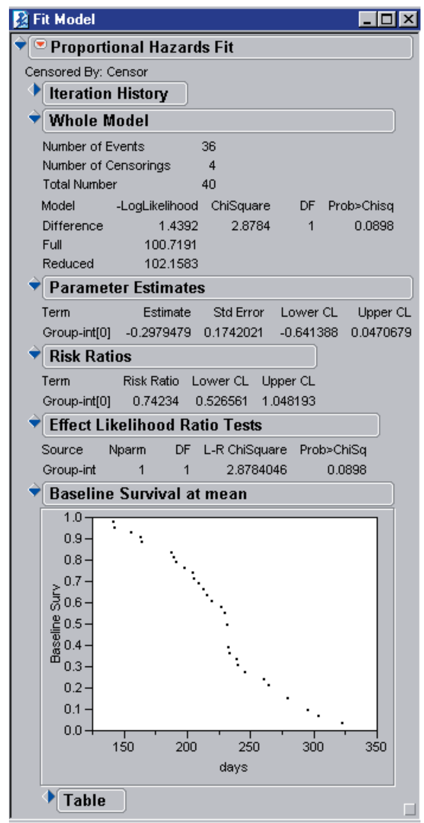
Alternatively, you can fit a nonlinear model. JMP allows you to create an equation that represents a loss function adjusted for censored observations. You can then use this function to estimate the parameters using maximum likelihood.

JMP also includes a sample library of ready-made tables with loss functions already in them, such as exponential, Weibull, and log-logistic functions.

You paste data into the table, run the Nonlinear analysis, and give the parameters initial values. JMP takes care of the rest.

Fit Proportional Hazards (Regression Models)

JMP lets you define proportional hazard models for a failure time response with right- or interval- censoring and covariates. Proportional hazard models are popular regression models for survival data with covariates.



JMP's proportional hazards reports.

The regression parameters associated with the explanatory variables and their standard errors are estimated using the maximum likelihood method. JMP also computes the conditional risk ratio (or hazard ratio) and its confidence limits from the parameter estimates.

Reliability Analysis

Reliability data needs to be analyzed using specialized methods for two reasons:

- The survival times have specialized non-normal distributions, like the exponential, Weibull, and lognormal.
- Some of the data could be censored—that is, you don't know the exact survival time, but you know that it is greater than a specified time (right-censored) or within an interval of time (interval-censored). Censoring happens when the study ends without all the units failing, or when a unit is removed from the study before the study finishes. If the censored observations are ignored, the analysis is biased.

An Example

An example from Meeker and Escobar[1] describes a fleet of 41 diesel engines. The data are engine age (in days) at the time that a valve seat was replaced.

JMP uses a data column to mark which observations are censored.

1 Meeker, W.Q. and Esobar, L.A. (1998), *Statistical Methods for Reliability Data*, pp. 60-62, New York: John Wiley and Sons.

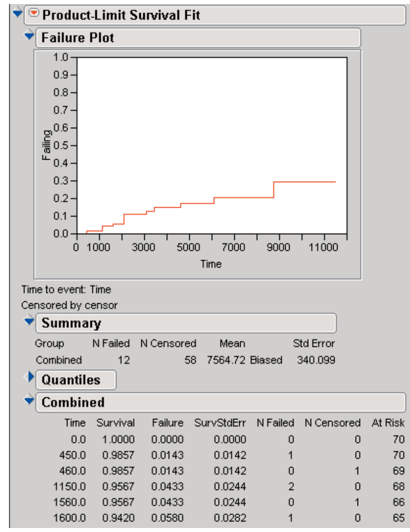
Obtaining the reliability analysis in JMP is simple. Specify **Time** as the time to event (the event being failure) and the **censor** column as the one holding the censoring information.

JMP then draws a failure plot, showing summary information, a display of failure probabilities, and their standard errors.

The next step is to explore distributional fits. You do this by clicking the red triangle icon and selecting one of JMP's fitting commands.

If you choose to fit a Weibull distribution, JMP will display the report shown below. You can use this plot to judge if the distribution fits well.

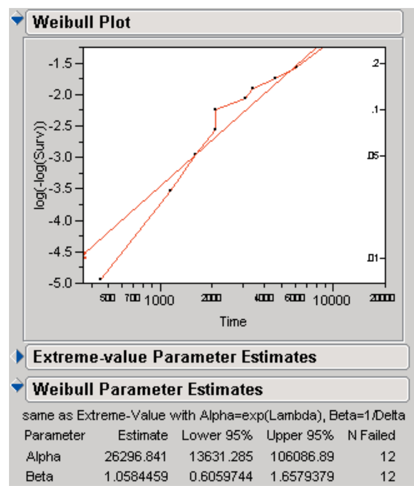
Since the fit is reasonable and the *beta* estimate is near one, you can conclude that this looks like an exponential distribution, which would have a constant hazard rate. In JMP, you can get three views of the distributional fit: survival, density, and hazard curves.



Group	N Failed	N Censored	Mean		Std Error
Combined	12	58	7564.72	Biased	340.099

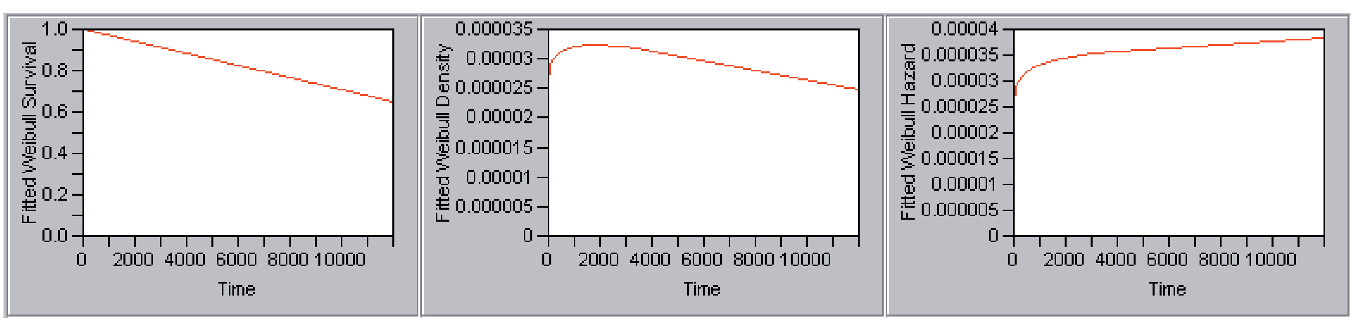
Time	Survival	Failure	SurvStdErr	N Failed	N Censored	At Risk
0.0	1.0000	0.0000	0.0000	0	0	70
450.0	0.9857	0.0143	0.0142	1	0	70
460.0	0.9857	0.0143	0.0142	0	1	69
1150.0	0.9567	0.0433	0.0244	2	0	68
1560.0	0.9567	0.0433	0.0244	0	1	66
1600.0	0.9420	0.0580	0.0282	1	0	65

A JMP failure plot.



Parameter	Estimate	Lower 95%	Upper 95%	N Failed
Alpha	26296.841	13631.285	106086.89	12
Beta	1.0584459	0.6059744	1.6579379	12

Weibull fit and plot.



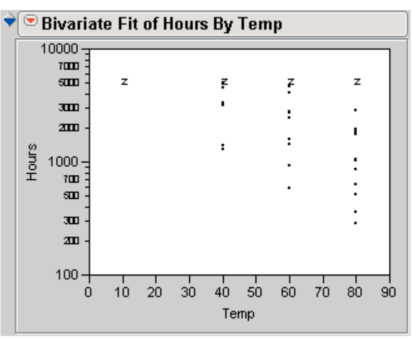
Three views of the distributional fit can be displayed using the Fitted Distribution Plots command.

Reliability: Accelerated Testing

Accelerated tests have become increasingly important as products' time-to-market decreases. One important type of test is the accelerated life test, where units are stressed to make them fail more rapidly.

The data shown in the figures below is temperature-accelerated lifetime data for an unnamed device. It uses the Arrhenius model and is taken from Meeker and Escobar[1].

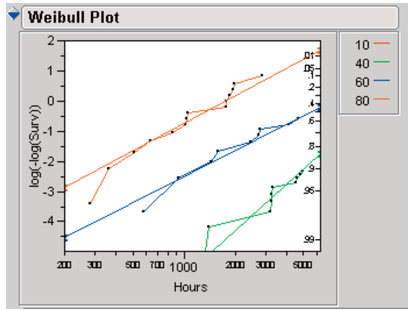
When the data is entered into JMP, the Bivariate platform gives an initial view of the data, plotting log(hours) by temperature (in degrees Celsius). Censored points—those that did not fail during the test—are designated by a "z" marker.



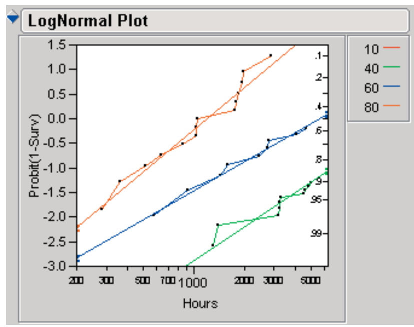
Device-A data with censored observations.

The above plot reveals that units fail sooner at higher temperature levels. Since there is heavy censoring at some temperatures, the appropriate model is not immediately apparent. As recommended by Meeker and Escobar, distributions to the data were fit at separate levels of the temperature variable.

JMP gives a Weibull probability plot and a lognormal probability plot. Both are reasonable fits, but the lognormal gives a better fit for the individual probability groups.



Weibull probability plot.

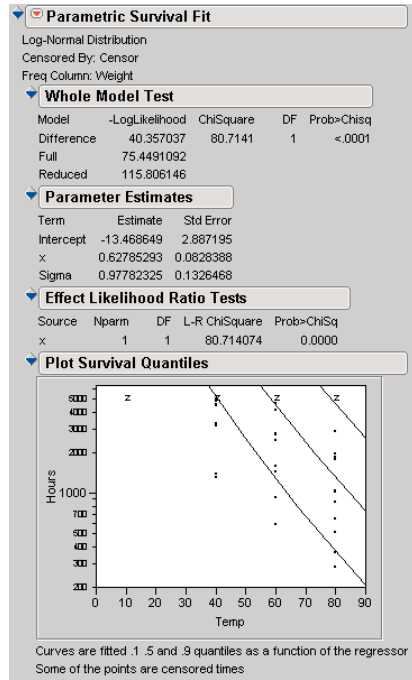


Lognormal probability plot.

Continuing this strategy involves fitting a single model using a regressor for temperature. The regressor, labeled *x*, is the Arrhenius transformation of temperature calculated by a formula stored in a data table column:

$$\frac{11605}{\text{Temperature} + 273.15}$$

Selecting JMP's **Fit Parametric Survival** option provides output shown in the following figure.



Parametric survival output.

The output shows tests of the adequacy of the model and a plot of the 0.1, 0.5, and 0.9 survival quantiles as a function of temperature.

With the fitted model, we can now obtain estimates of the survival probabilities for times beyond our test. JMP prompts for the times (for instance, 10,000 and 30,000 hours) and produces survival and failure probabilities, along with 95% confidence intervals for an *x* of 40.98, which corresponds to a temperature of 10 degrees Celsius.

Estimates of Survival

x	Time	Prob Failure	Lower 95%	Upper 95%	Prob Survival
40.9853	10000	0.00090	0.00006	0.01347	0.99910
	30000	0.02278	0.00320	0.14473	0.97722

Probabilities for the estimate of survival.

1 Meeker, W.Q. and Esobar, L.A. (1998), *Statistical Methods for Reliability Data*, pp. 60-62, New York: John Wiley and Sons.

Reliability: Competing Causes

In any structured quality program (such as Six Sigma), it is important to look at special causes of failure. Knowing how a product fails is important, and a careful statistical analysis of these causes can help in eliminating them.

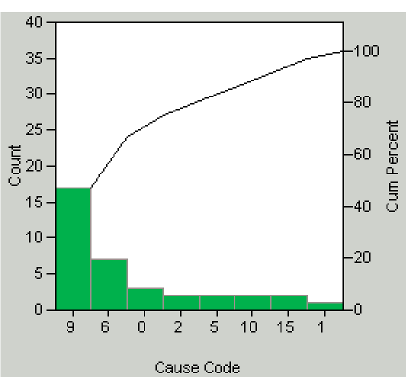
Sometimes there are multiple causes of failure in a system. If the different causes are independent, the failure times can be modeled by an estimation of the survival distribution for each cause. These multiple causes are frequently termed competing failure modes or competing causes. JMP's reliability analysis allows you to explore the effect on the system when causes are omitted.

An Example

Nelson[1] gives an example of a small electrical appliance that has several causes of failure. The objectives of the analysis are to

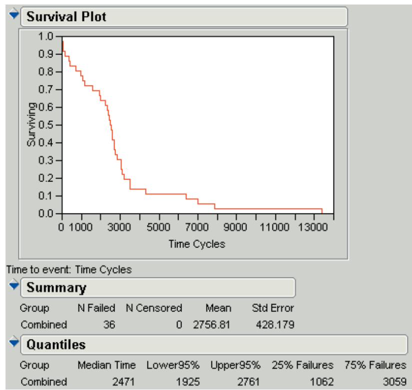
- identify which causes of failure are most prevalent
- explore the survival distribution if these failure causes were corrected

One way of identifying the leading cause of failure is through a Pareto plot. JMP's Pareto plot of the appliance's failures clearly shows cause nine as an area on which to focus.



A Pareto plot.

The initial reliability analysis shows a survival plot that is steep, with many failures between 2,000 and 3,000 hours.

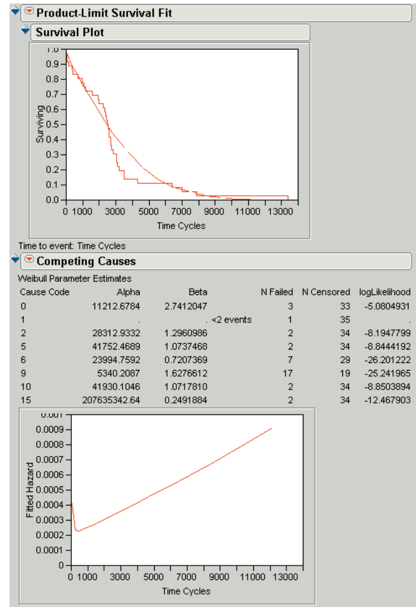


Initial analysis.

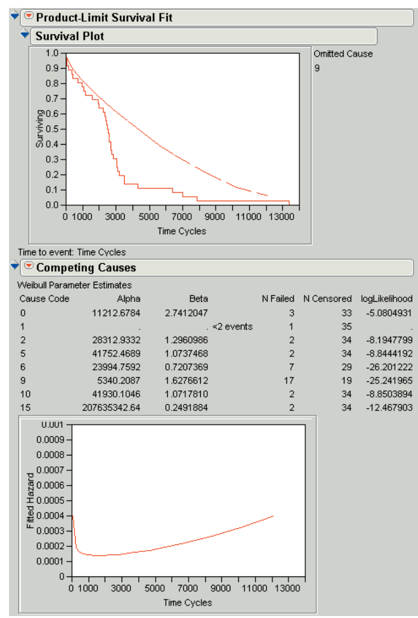
With competing causes, each cause of failure can have a separate failure distribution. Each failure becomes a censored time for all the causes that are not the cause of failure for that particular event. You could analyze this by creating separate censoring columns for each cause. However, JMP has a command that will do all this automatically, estimating the Weibull distribution for each cause.

Alpha is the estimate of the 63.2% quantile of failure time. Cause nine in the example below stands out with 17 observed failures and a low alpha estimate.

What would change if cause nine were eliminated? By selecting a single JMP menu option (**Omit Causes**) you can omit one or more causes. After omitting cause nine, the estimated Weibull curve and the hazard plot both change to reflect the omission.



Competing causes report.



Competing causes report with cause nine omitted.

The value of fixing cause nine is now apparent. The operational gain per unit is roughly the area between the original survival curve and the new curve omitting cause nine.

1 Nelson, W. (1982), *Applied Life Data Analysis*, New York: John Wiley and Sons.

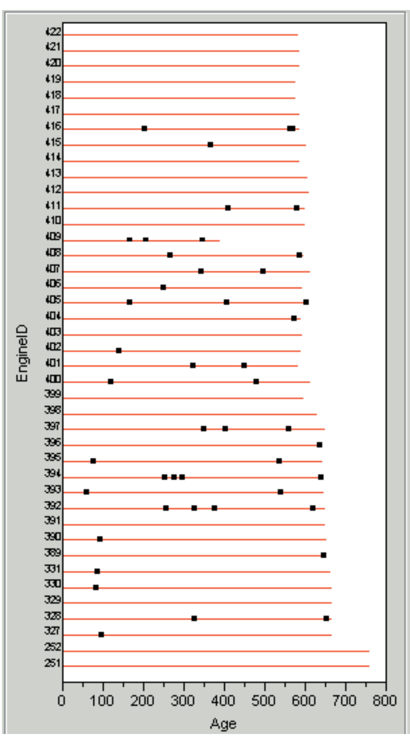
Recurrence Analysis for Repairable Systems

Recurrence analysis analyzes event times like survival or reliability analysis, but recurrence analyses contain events that can recur several times for each unit. Typically, these events are when a unit breaks down and needs to be repaired then placed back into service. The units are followed until they are ultimately taken out of service.

The goal of recurrence analysis is to obtain the MCF (Mean Cumulative Function), which shows the total cost per unit as a function of time, where cost can be just the number of repairs, or it can be the actual cost of repair.

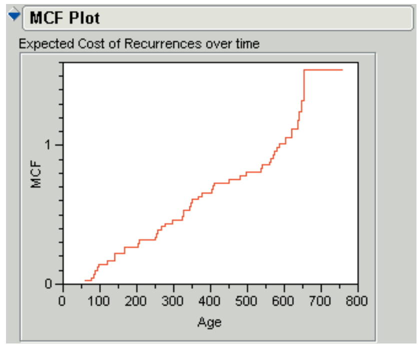
An Example

An example from Meeker and Escobar[1] describes a fleet of 41 diesel engines. The data are engine age (in days) at the time that a valve seat was replaced. When the data are placed in JMP, an event plot shows the times for repair on each of the systems.



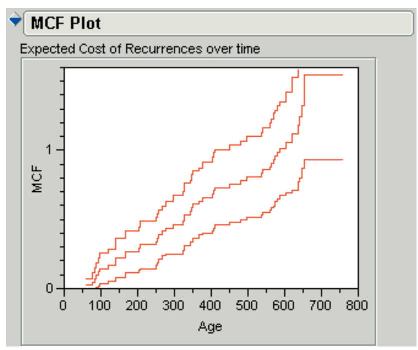
Event plot.

The MCF plot appears automatically with the recurrence analysis.



MCF plot.

After you have received the plot above, you can produce confidence intervals, as shown on the fit below.



MCF plot with confidence intervals.

As shown in the above plot, the confidence intervals are quite wide for the later times. A look back at the data shows that there were only two units under observation at these times, making the estimate more uncertain.

The MCF estimates, their standard errors, and confidence intervals are available in text form as well.

MCF Table

Age	MCF	MCF Std	MCF LCL	MCF UCL
61	0.0243902	0.024091	-0.022827	0.0716077
76	0.0487805	0.0336412	-0.017155	0.114716
84	0.0731707	0.0406702	-0.006541	0.1528829
87	0.097561	0.0463399	0.0067364	0.1883855
92	0.1219512	0.0511047	0.0217879	0.2221145
98	0.1463415	0.0551993	0.0381528	0.2545302
120	0.1707317	0.0587642	0.055556	0.2859074
139	0.2195122	0.0732698	0.075906	0.3631184
165	0.2439024	0.0754167	0.0960885	0.3917164
166	0.2682927	0.0773166	0.1167549	0.4198305
202	0.2926829	0.0789875	0.1378702	0.4474956
206	0.3170732	0.0875268	0.1455237	0.4886226
249	0.3414634	0.0886796	0.1676545	0.5152723
254	0.3658537	0.0896559	0.1901312	0.5415761
258	0.3902439	0.0904615	0.2129426	0.5675452
265	0.4146341	0.0911008	0.2360798	0.5931884
276	0.4390244	0.097858	0.2472263	0.6308225
298	0.4634146	0.1096073	0.2485883	0.6782409
323	0.4878049	0.1097396	0.2727193	0.7028905
326	0.5121951	0.1097396	0.2971095	0.7272807
328	0.5365854	0.1149066	0.3113726	0.7617981
344	0.5609756	0.1146538	0.3362584	0.7856929
348	0.5853659	0.1242496	0.3418411	0.8288906
349	0.6097561	0.1237816	0.3671486	0.8523636

Text table provided alongside MCF plots.

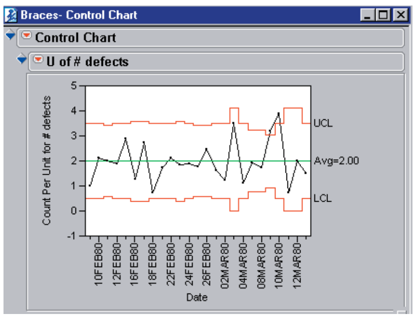
Furthermore, if you have a grouping variable, you can create a plot of the difference of MCF's, including a 95% confidence interval for that difference. The MCF's are significantly different where the confidence interval lines do not cross the zero line.

1 Meeker, W.Q. and Esobar, L.A. (1998), *Statistical Methods for Reliability Data*, pp. 60-62, New York: John Wiley and Sons.

Quality Control

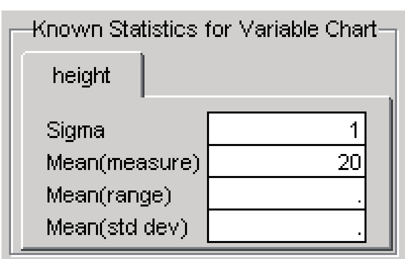
JMP offers several methods designed to measure and monitor quality control. These methods are control charts, including

- $\bar{X}$, R-, X-, MR-, and *s*-charts
- UWMA and EWMA charts
- *p*-, *np*-, *u*-, and *c*-charts for attributes
- one- and two-sided CUSUM charts
- Leyey-Jennings plots



Example of a JMP quality control chart.

All control charts in JMP allow you to specify known statistics ahead of time if you have historical data about the process.

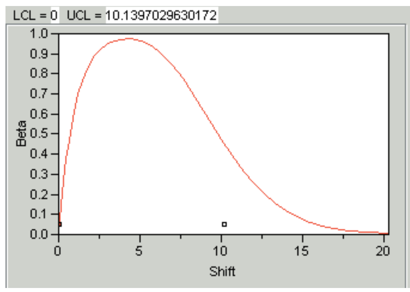


JMP can handle historical data.

In addition, JMP can analyze unequal batch sizes.

OC Curves

JMP has built-in scripts for plotting operating characteristic curves. These curves show graphically the probability of not detecting a shift to a given number of non-conformities. The scripts are interactive, so you can change the values of the LCL and UCL and instantly see the effect of your change.



A chart showing operating characteristics.

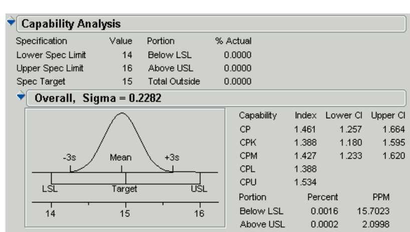
Capability Analysis

You can obtain a capability analysis of your process. There are four ways of specifying sigma for the analysis

- long-term
- specified, or imported from control chart
- short-term, grouped by fixed

subgroup size

- short-term, grouped by variable



A capability analysis.

Instrument Friendliness and Real-Time Information

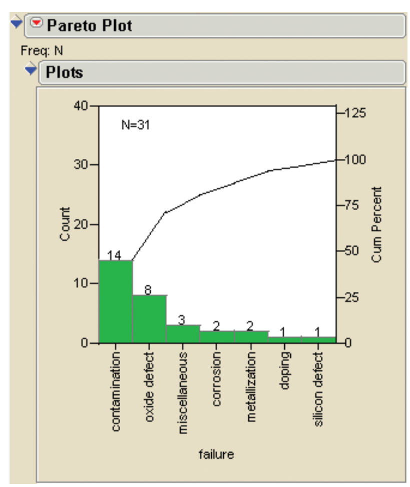
You can connect a measuring instrument to a serial port and have a script direct the data to extend a data table. Each time the instrument makes a measurement, it creates the script. This happens in real time as measurements are made. Each time a batch of measurements is added (usually 5-20), JMP adds a point to the Shewhart chart. If the chart needs more space, JMP contracts the scale and reformats the plot.

Out-of-Control Alarm

If a process goes out of control as defined by any combination of the available criteria (the "Western Electric" rules or the Westgard rules), then JMP can call a script to play a message, show a message in a window, or email a message to a person or pager for immediate attention.

Pareto Charts

JMP provides a complete comparative Pareto charts implementation, including two-way comparative charts.



A Pareto chart.

Measurement Systems

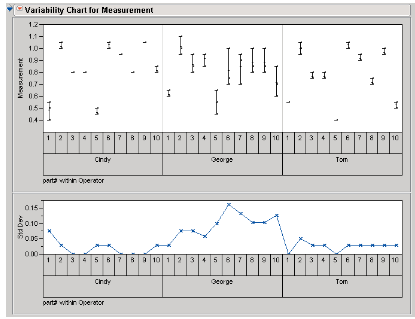
Measurement systems analysis is an important step in any quality control application. Before studying the process itself, it is important to make sure that the variation due to measurement errors is small relative to the variation in the process. If most of the variation you see comes from the measuring process itself, then you aren't learning reliably about the process.

The instruments that take the measurements are called gauges, and the analysis of their variation is a Gage R&R (Repeatability and Reproducibility) study, which examines the following quantities:

- The process variation, from one part to another. This is the ultimate variation that you want to be studying if your measurements are reliable.
- The variability inherent in making multiple measurements—that is, repeatability.
- The variability due to having different operators measure parts—that is, reproducibility.

Although there are some approaches to Gage R&R that use ranges to estimate variability, JMP implements Gage R&R studies through its variability charts. This method uses the variance component method, which allows for estimation of operator by part interactions.

In the same way that a Shewhart control chart can identify processes that are going out of control over time, a variability chart can help identify operators, instruments, or part sources that are systematically different in mean or variance. By default, JMP shows a range chart of the measurements and a chart showing the standard deviations of each measurement.



Measurement and standard deviation charts.

From these charts, you can request a variance components analysis. Since operator and part are regarded as random effects, variance components can be estimated for them and their interaction.

Analysis of Variance

Source	DF	SS	Mean Square	F Ratio	Prob > F
Operator	2	0.054889	0.02744	1.3150	0.2931
part#	9	2.633583	0.29262	14.0209	<.0001
Operator*part#	18	0.375667	0.02087	5.0425	<.0001
Within	60	0.248333	0.00414		
Total	89	3.312472	0.03722		

Variance Components

Component	Var Component	% of Total	Sqrt(Var Comp)
Operator	0.00021914	0.5	0.01480
part#	0.03019444	75.2	0.17377
Operator*part#	0.00557716	13.9	0.07468
Within	0.00413889	10.3	0.06433
Total	0.04012963	100.0	0.20032

Analysis of variance and variance components.

In addition, you can generate a Gage R&R report. You can specify any σ multiplier (frequently equal to 5.15) and, optionally, values for the tolerance interval (USL–LSL).

Gage R&R

Measurement	Variation		which is k*sqrt of
Repeatability (EV)	0.3313211	Equipment Variation	V(Within)
Reproducibility (AV)	0.0762367	Appraiser Variation	V(Operator)
Operator * Part (IV)	0.3846040	Interaction Variation	V(Oper*Part)
Gage R&R (RR)	0.5133283	Measurement Variation	V(Within)+V(Oper)+V(Oper*Part)
Part Variation (PV)	0.8948923	Part Variation	V(Part)
Total Variation (TV)	1.0316676	Total Variation	V(Within)+V(Oper)+V(Oper*Part)+V(Part)

5.15 k
49.7571 % Gage R&R = 100*(RR/TV)
0.57362 Precision to Process Width Ratio = RR/PV
Using column 'Operator' for Operator, and column 'part#' for Part.

Gage R&R report.

Note that in this report, values (repeatability, reproducibility, and so on) are calculated from the variance components.

JMP provides additional information: the discrimination ratio. The discrimination ratio characterizes the relative usefulness of a given measurement for a specific product. It compares the total variance of the measurement with the variance of the measurement error.

Discrim Ratio

Source	Discrim Ratio
Operator	1.1727709
part#	4.0606574

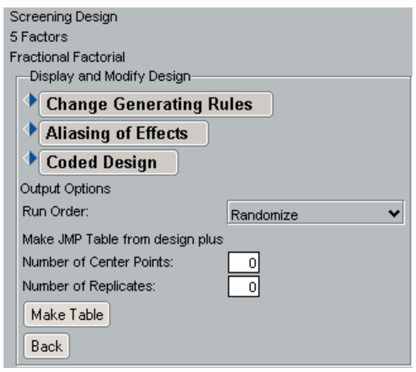
Discrimination ratio report.

DOE: Classic Designs

JMP offers several types of classical designs that you can use in your experiment.

Screening Designs

Screening experiments separate the group of factors having a significant influence on the response from the rest of the factors. JMP supplies a list of screening designs for an unlimited number of factors and designs that group the experimental runs into blocks of equal sizes where the size is a power of two.



Screening design output options.

Response Surface Designs

JMP offers the Response Surface Methodology (RSM) technique to help you find the optimal response within the specified ranges of the factors. These designs are capable of fitting a second order prediction equation for the response.

Space Filling Designs

JMP contains this powerful tool for modeling complex, nonlinear systems. Space filling designs are useful for developing simpler approximations of a complex computer simulation models. These designs are also useful for modeling deterministic systems, where the relationship between input and output variables are unknown but thought to be highly nonlinear.

Full Factorial Designs

These designs contain all possible combinations of a set of factors. JMP supports both continuous and categorical factors with up to nine levels.

Taguchi Designs

These designs help you find control factor settings that generate acceptable responses despite natural environmental and process variability. JMP supports the standard features of the Taguchi method, including inner and outer array designs, control and noise factors, and signal-to-noise ratios.

Mixture Designs

JMP offers a way for you to design a set of factors that are ingredients in a mixture. You choose among several mixture design approaches, such as simplex centered, simplex lattice, and extreme vertices.

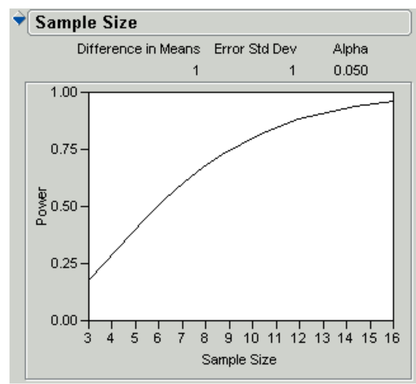
Furthermore, JMP has specialized tools for building designs using a set of factors that are ingredients in a mixture.

Augment Designs

JMP can add new runs to existing designs with the options of adding center points, replicating the design, creating a foldover design, adding axial points, or using a model that has more terms than the original model. Adding runs to a design is powerful because you can achieve the objectives of response surface methodology by changing a linear model to a full quadratic model and adding the necessary number of runs. For example, suppose you start with a two-factor, two-level, four-run design. If you add quadratic terms and five new points to the model, JMP generates the three-by-three full factorial as the optimal augmented design.

Sample Size and Power

JMP computes power, sample size, or the effect size you want to detect given a significance level and error standard deviation. You supply two of these values and the JMP computes the third. If you can only supply one of these values, JMP gives a plot of the other two.



A JMP graph showing sample size.

Six Sigma Calculator

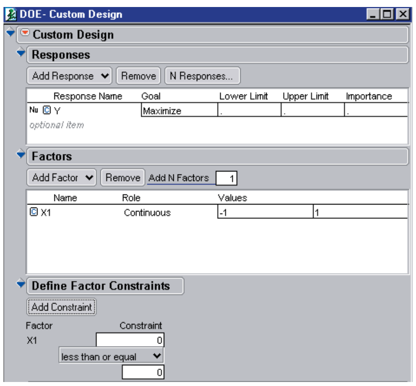
JMP also provides a Six Sigma Quality Level calculator. This provides you with a convenient way to get the sigma-level of your process.

Factors and Responses

Once a design has been specified, the factors, responses, and constraints can be saved for later use. Furthermore, responses can be simulated. This is useful in testing designs or in teaching situations.

DOE: Custom Designer

Rarely is a pre-defined design an exact match for your unique industrial process. You usually have to modify the process description to suit the design or make ad hoc modifications to the design so it models your process better. To give you the flexibility to tailor a design to match your specific circumstances, JMP offers the custom designer. You describe the process variables and constraints, then JMP tailors a design that fits.



The Custom Designer tailors a design to fit your needs.

This approach is general and requires less experience and expertise in statistical design of experiments. When creating designs with the custom designer, you can

- mix factor roles as required by your engineering situation, such as having non-mixture factors in a mixture experiment
- add factors with any role in any experiment
- have categorical factors with as many levels as you need
- specify any number of runs per block
- have any design with continuous or categorical covariate factors (factors whose values are fixed in advance of the experiment)
- disallow certain regions of the factor space by defining linear inequality constraints

Custom designs generate an appropriate design for the requirements you specify. In cases when a classic design (such as factorial) are optimal, JMP finds them. Therefore, custom designs serve any number or combinations of factors.

After generating a design, JMP gives you a Prediction Variance Profiler to use as a diagnostic tool to assess the quality of the design. You can use this tool to compare many candidate designs and choose the one that best meets your needs.

How Does the Custom Designer Work?

You can use the custom designer for routine factor screening, response optimization, and mixture problems. It uses the following approach to help you customize your design.

The custom designer starts with a random design with each point inside the range of each factor. The computational method is an iterative algorithm called coordinate exchange. Each iteration of the algorithm tests every value of every factor in the design to determine if replacing that value increases the optimality criterion. If so, the new value replaces the old. Iteration continues until no replacement occurs in an entire iterate.

To help ensure the solution is a global (rather than local) optimum, the whole process is repeated several times using a different random start. The designer displays the best of these designs.

If a design problem has several equivalent solutions (designs with equal precision for estimating the model coefficients as a group), the design algorithm can generate different, but equivalent, designs.

I-Optimal Designs

JMP allows you to employ an *I*-Optimal design, which improves predictions in response-surface settings. *I*-Optimal are the default for RSA designs.

Supersaturated Designs

In some cases, there are more factors that need to be explained than observations. Traditionally, this was an impossible situation.

However, JMP has the ability to generate supersaturated designs, which can, for example, examine dozens of factors using less than half as many runs using Bayesian *D*-Optimality.

Mixing Factor Roles

The ability to mix factor roles is what makes custom designs so flexible. Preformulated designs rely on the assumption that the experimenter controls all the factors. JMP's custom designer is flexible enough to allow you to include all design factors that affect experimental responses.

Split Plot Design and Analysis

Sometimes split plot experiments are not recognized nor analyzed correctly. The common trait in split-plot designs is that a factor is constant within a block but not across blocks, which means there is a restriction on the randomization of runs. In JMP, you specify a whole plot identifier effect in the model and declare it as a random effect. Everything else is automatic. JMP can then design an **optimal** split-plot experiment and analyze any split plot data with the best statistical methods and with a minimum of expertise required from the user.

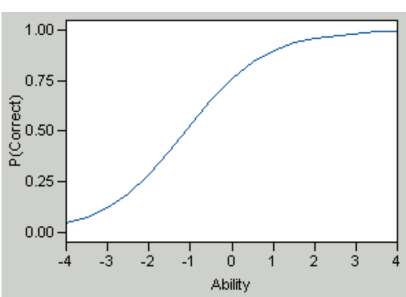
Item Response Theory (IRT)

Item Response Theory (IRT) is a method of scoring tests of ability, such as educational achievement tests. Although classical test theory methods have been widely used for a century, IRT provides a better, more scientifically-based scoring procedure. Advantages include

- scoring tests at the item level, giving insight into the contributions of each item on the total test score
- producing scores of both the test takers and the test items on the same scale
- identifying which questions or items yield the most information about ability
- accounting for not only the total number of correct answers, but the quality of the correctly-answered questions

Item Characteristic Curves

IRT models the probability of getting a question correct based on the trait or ability being measured. This probability is expressed graphically in JMP with Item Characteristic Curves (ICCs).



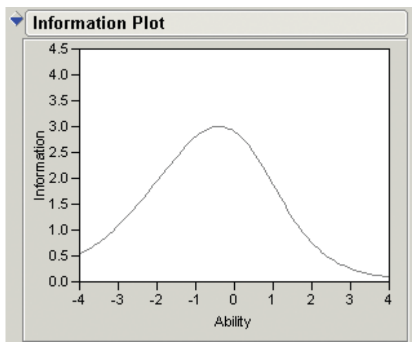
Item characteristic curve in JMP.

The ICCs appear in an array for easy comparison.

Furthermore, JMP implements 1PL, 2PL, and 3PL models.

Information Curves

In addition to ICCs, JMP plots information curves (light gray in the array of plots below). These curves give you an idea of which ability levels the item measures the best. JMP also displays the sum of all the information curves, giving you an idea of the ability levels for which the entire test is appropriate.



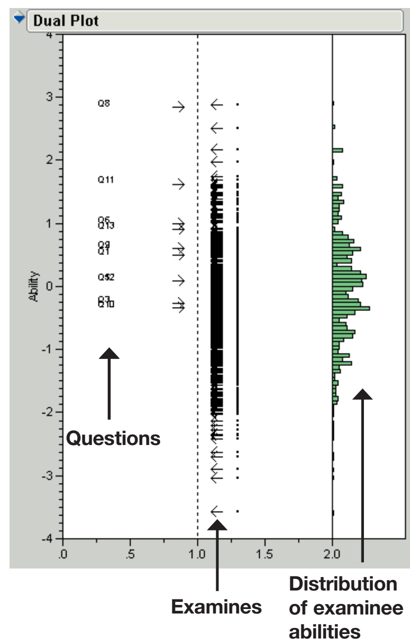
Total information curve in JMP.

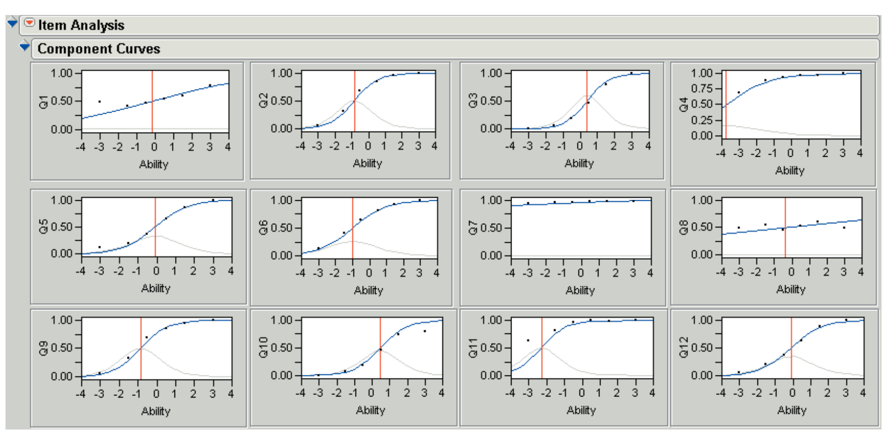
The total information curve can be used in conjunction with the individual item curves to tailor-make a test for certain ability levels. The example above gives the most information for examinees that have slightly below average ability. If more information is needed at other ability levels, questions can be added to increase the information at those levels.

Dual Plots

The information gained from item difficulty parameters in IRT models can be used to construct an increasing scale of questions, from easiest to hardest, on the same scale as the examinees. This structure gives information on which items are associated with low levels of the trait, and which are associated with high levels of the trait.

JMP shows this correspondence with a dual plot.



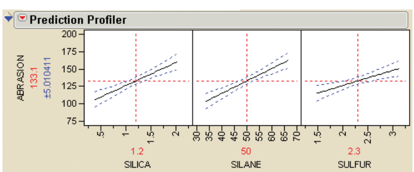


An array of ICC (blue) and information curves (gray).

The Profiler

JMP's philosophy recognizes that it is easier to interpret graphical representations of statistics rather than tables. In addition, point-and-click interfaces are more intuitive than complicated programming languages.

An example of this philosophy is JMP's profiler. For models that involve a complicated surface (such as a quadratic response surface design or a neural net), the profiler allows you to view prediction traces (slices through the response surface), adjust the values of these traces, and calculate optimal values of the *x*-values. It allows you to change one variable at a time and see its effect on the other variables.



The profiler shows the current value of the response, which changes as the *x*-values change.

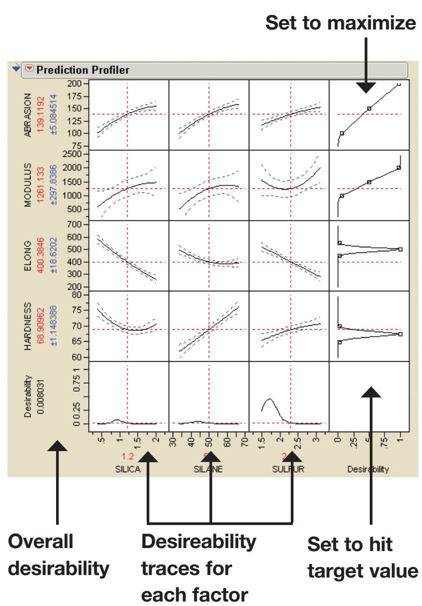
Since the traces are displayed graphically, you can quickly evaluate the factors' effect on the model.

- The importance of a factor can be assessed to some extent by the steepness of the prediction trace. If the model has curvature terms (such as squared terms), then the traces can be curved.
- If you change a factor's value, then its prediction trace is not affected, but the prediction traces of all the other factors can change. The Y response line must cross the intersection points of the prediction traces with their current value lines.

Prediction profiles are especially useful in multiple-response models to help judge which factor values can optimize a complex set of criteria.

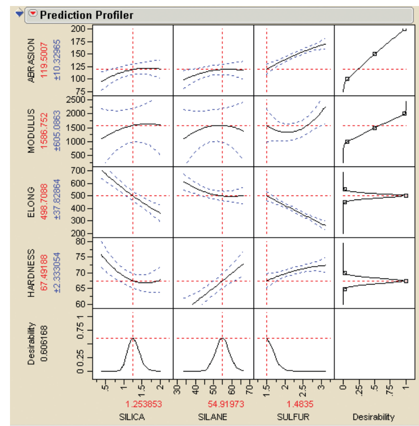
Desirability Profiling

Often there are multiple responses measured for each set of experimental conditions, and the desirability of the outcome involves several or all of these responses. For example, you might want to maximize one response, minimize another, and keep a third response close to some target value. In desirability profiling, you specify a desirability function for each response. The overall desirability can be defined as the geometric mean of the desirability for each response.



Once the desirability functions are set, JMP can automatically find the optimal settings for each factor.

JMP makes optimizing values simple. The desirability profiles are set easily by dragging curves to their desired values, and the **Maximize Desirability** command quickly finds their optimum settings.



This example's curves were dragged to their desired values and the optimum setting found.

JMP Headquarters
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
USA
Tel: +1 919.677.8000
Fax: +1 919.677.4444
jmpsales@jmp.com
www.jmp.com

JMP Europe
SAS Institute
Henley Road
Medmenham
Marlow
SL7 2EB
United Kingdom
Tel: +44 (0)1628 486 933
Fax: +44 (0)1628 483 203
jmpsaleseur@jmp.com
www.jmp.com

JMP Japan
SAS Japan Head Office
Inui Bldg. Kachidoki
1-13-1 Kachidoki
Chuo-ku Tokyo 104-0054
Japan
Tel: +81 3 3533 3887
Fax: +81 3 3533 1600
jmpjapan@jmp.com
www.jmp.com/japan

JMP China
SAS China
25/F POS Plaza
1600 Century Avenue
Pudong New District
Shanghai 200122
PRC
Tel: + 86 21 6876 5353
Fax: + 86 21 6876 9010
jmpsalesprc@jmp.com
www.jmp.com