**Version 13**

# Profilers
## *Second Edition*

*"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."*

Marcel Proust

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

**Technology License Notices**

- Scintilla - Copyright © 1998-2014 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

- Telerik RadControls: Copyright © 2002-2012, Telerik. Usage of the included Telerik RadControls outside of JMP is not permitted.

- ZLIB Compression Library - Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.

- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.

- Packages - Copyright © 2009-2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

– Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar licence but with different Copyright notices) all the files are:

- Regarding the decompression algorithm used for UNIX files:

# Get the Most from JMP®

Whether you are a first-time or a long-time user, there is always something to learn
about JMP.

Visit JMP.com to find the following:

- live and recorded webcasts about how to get started with JMP
- video demos and webcasts of new features and advanced techniques
- details on registering for JMP training
- schedules for seminars being held in your area
- success stories showing how others use JMP
- a blog with tips, tricks, and stories from JMP staff
- a forum to discuss JMP with other users

## http://www.jmp.com/getstarted/

# Contents

Profilers

# Learn about JMP

## Documentation and Additional Resources

This chapter includes the following information:

- book conventions
- JMP documentation
- JMP Help
- additional resources, such as the following:
  - other JMP documentation
  - tutorials
  - indexes
  - Web resources
  - technical support options

# Formatting Conventions

The following conventions help you relate written material to information that you see on your screen:

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in Helvetica font.

- Code appears in Lucida Sans Typewriter font.

- Code output appears in *Lucida Sans Typewriter* italic font and is indented farther than the preceding code.

- **Helvetica bold** formatting indicates items that you select to complete a task:

  – buttons

  – check boxes

  – commands

  – list names that are selectable

  – menus

  – options

  – tab names

  – text boxes

- The following items appear in italics:

  – words or phrases that are important or have definitions specific to JMP

  – book titles

  – variables

  – script output

- Features that are for JMP Pro only are noted with the JMP Pro icon **JMP PRO** . For an overview of JMP Pro features, visit http://www.jmp.com/software/pro/.

**Note:** Special information and limitations appear within a Note.

**Tip:** Helpful information appears within a Tip.

# JMP Documentation

JMP offers documentation in various formats, from print books and Portable Document Format (PDF) to electronic books (e-books).

- Open the PDF versions from the **Help > Books** menu.

- All books are also combined into one PDF file, called *JMP Documentation Library*, for convenient searching. Open the *JMP Documentation Library* PDF file from the **Help > Books** menu.

- You can also purchase printed documentation and e-books on the SAS website:

  http://www.sas.com/store/search.ep?keyWords=JMP

## JMP Documentation Library

The following table describes the purpose and content of each book in the JMP library.

| Document Title | Document Purpose | Document Content |
|---|---|---|
| *Discovering JMP* | If you are not familiar with JMP, start here. | Introduces you to JMP and gets you started creating and analyzing data. |
| *Using JMP* | Learn about JMP data tables and how to perform basic operations. | Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS. |
| *Basic Analysis* | Perform basic analysis using this document. | Describes these Analyze menu platforms: <br><br>• Distribution <br><br>• Fit Y by X <br><br>• Tabulate <br><br>• Text Explorer <br><br>Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze > Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included. |

| Document Title | Document Purpose | Document Content |
| --- | --- | --- |
| *Essential Graphing* | Find the ideal graph for your data. | Describes these Graph menu platforms:<br><br>• Graph Builder<br>• Overlay Plot<br>• Scatterplot 3D<br>• Contour Plot<br>• Bubble Plot<br>• Parallel Plot<br>• Cell Plot<br>• Treemap<br>• Scatterplot Matrix<br>• Ternary Plot<br>• Chart<br><br>The book also covers how to create background and custom maps. |
| *Profilers* | Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface. | Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs. |
| *Design of Experiments Guide* | Learn how to design experiments and determine appropriate sample sizes. | Covers all topics in the DOE menu and the Specialized DOE Models menu item in the Analyze > Specialized Modeling menu. |

| Document Title | Document Purpose | Document Content |
| --- | --- | --- |
| *Fitting Linear Models* | Learn about Fit Model platform and many of its personalities. | Describes these personalities, all available within the Analyze menu Fit Model platform: |

- Standard Least Squares
- Stepwise
- Generalized Regression
- Mixed Model
- MANOVA
- Loglinear Variance
- Nominal Logistic
- Ordinal Logistic
- Generalized Linear Model

| Document Title | Document Purpose | Document Content |
|---|---|---|
| *Predictive and Specialized Modeling* | Learn about additional modeling techniques. | Describes these Analyze > Predictive Modeling menu platforms: |

Describes these Analyze > Predictive Modeling menu platforms:

- Modeling Utilities
- Neural
- Partition
- Bootstrap Forest
- Boosted Tree
- K Nearest Neighbors
- Naive Bayes
- Model Comparison
- Formula Depot

Describes these Analyze > Specialized Modeling menu platforms:

- Fit Curve
- Nonlinear
- Gaussian Process
- Time Series
- Matched Pairs

Describes these Analyze > Screening menu platforms:

- Response Screening
- Process Screening
- Predictor Screening
- Association Analysis

The platforms in the Analyze > Specialized Modeling > Specialized DOE Models menu are described in *Design of Experiments Guide*.

| Document Title | Document Purpose | Document Content |
| --- | --- | --- |
| *Multivariate Methods* | Read about techniques for analyzing several variables simultaneously. | Describes these Analyze > Multivariate Methods menu platforms:<br><br>• Multivariate<br>• Principal Components<br>• Discriminant<br>• Partial Least Squares<br><br>Describes these Analyze > Clustering menu platforms:<br><br>• Hierarchical Cluster<br>• K Means Cluster<br>• Normal Mixtures<br>• Latent Class Analysis<br>• Cluster Variables |
| *Quality and Process Methods* | Read about tools for evaluating and improving processes. | Describes these Analyze > Quality and Process menu platforms:<br><br>• Control Chart Builder and individual control charts<br>• Measurement Systems Analysis<br>• Variability / Attribute Gauge Charts<br>• Process Capability<br>• Pareto Plot<br>• Diagram |

| Document Title | Document Purpose | Document Content |
|---|---|---|
| *Reliability and Survival Methods* | Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products. | Describes these Analyze > Reliability and Survival menu platforms:<br><br>• Life Distribution<br>• Fit Life by X<br>• Cumulative Damage<br>• Recurrence Analysis<br>• Degradation and Destructive Degradation<br>• Reliability Forecast<br>• Reliability Growth<br>• Reliability Block Diagram<br>• Repairable Systems Simulation<br>• Survival<br>• Fit Parametric Survival<br>• Fit Proportional Hazards |
| *Consumer Research* | Learn about methods for studying consumer preferences and using that insight to create better products and services. | Describes these Analyze > Consumer Research menu platforms:<br><br>• Categorical<br>• Multiple Correspondence Analysis<br>• Multidimensional Scaling<br>• Factor Analysis<br>• Choice<br>• MaxDiff<br>• Uplift<br>• Item Analysis |
| *Scripting Guide* | Learn about taking advantage of the powerful JMP Scripting Language (JSL). | Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications. |

| Document Title | Document Purpose | Document Content |
| --- | --- | --- |
| *JSL Syntax Reference* | Read about many JSL functions on functions and their arguments, and messages that you send to objects and display boxes. | Includes syntax, examples, and notes for JSL commands. |

**Note:** The **Books** menu also contains two reference cards that can be printed: The *Menu Card* describes JMP menus, and the *Quick Reference* describes JMP keyboard shortcuts.

## JMP Help

JMP Help is an abbreviated version of the documentation library that provides targeted information. You can open JMP Help in several ways:

- On Windows, press the F1 key to open the Help system window.

- Get help on a specific part of a data table or report window. Select the Help tool ? from the **Tools** menu and then click anywhere in a data table or report window to see the Help for that area.

- Within a JMP window, click the **Help** button.

- Search and view JMP Help on Windows using the **Help > Help Contents**, **Search Help**, and **Help Index** options. On Mac, select **Help > JMP Help**.

- Search the Help at http://jmp.com/support/help/ (English only).

## Additional Resources for Learning JMP

In addition to JMP documentation and JMP Help, you can also learn about JMP using the following resources:

- Tutorials (see "Tutorials" on page 22)
- Sample data (see "Sample Data Tables" on page 22)
- Indexes (see "Learn about Statistical and JSL Terms" on page 22)
- Tip of the Day (see "Learn JMP Tips and Tricks" on page 22)
- Web resources (see "JMP User Community" on page 23)
- JMPer Cable technical publication (see "JMPer Cable" on page 23)
- Books about JMP (see "JMP Books by Users" on page 24)
- JMP Starter (see "The JMP Starter Window" on page 24)

- Teaching Resources (see "Sample Data Tables" on page 22)

## Tutorials

You can access JMP tutorials by selecting **Help > Tutorials**. The first item on the **Tutorials** menu is **Tutorials Directory**. This opens a new window with all the tutorials grouped by category.

If you are not familiar with JMP, then start with the **Beginners Tutorial**. It steps you through the JMP interface and explains the basics of using JMP.

The rest of the tutorials help you with specific aspects of JMP, such as designing an experiment and comparing a sample mean to a constant.

## Sample Data Tables

All of the examples in the JMP documentation suite use sample data. Select **Help > Sample Data Library** to open the sample data directory.

To view an alphabetized list of sample data tables or view sample data within categories, select **Help > Sample Data**.

Sample data tables are installed in the following directory:

On Windows: C:\Program Files\SAS\JMP\13\Samples\Data

On Macintosh: \Library\Application Support\JMP\13\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory. In JMP Shrinkwrap, sample data is installed in the JMPSW directory.

To view examples using sample data, select **Help > Sample Data** and navigate to the Teaching Resources section. To learn more about the teaching resources, visit http://jmp.com/tools.

## Learn about Statistical and JSL Terms

The **Help** menu contains the following indexes:

**Statistics Index**  Provides definitions of statistical terms.

**Scripting Index**  Lets you search for information about JSL functions, objects, and display boxes. You can also edit and run sample scripts from the Scripting Index.

## Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window. This window provides tips for using JMP.

To turn off the Tip of the Day, clear the **Show tips at startup** check box. To view it again, select **Help > Tip of the Day**. Or, you can turn it off using the Preferences window. See the *Using JMP* book for details.

## Tooltips

JMP provides descriptive tooltips when you place your cursor over items, such as the following:

- Menu or toolbar options
- Labels in graphs
- Text results in the report window (move your cursor in a circle to reveal)
- Files or windows in the Home Window
- Code in the Script Editor

**Tip:** On Windows, you can hide tooltips in the JMP Preferences. Select **File > Preferences > General** and then deselect **Show menu tips**. This option is not available on Macintosh.

## JMP User Community

The JMP User Community provides a range of options to help you learn more about JMP and connect with other JMP users. The learning library of one-page guides, tutorials, and demos is a good place to start. And you can continue your education by registering for a variety of JMP training courses.

Other resources include a discussion forum, sample data and script file exchange, webcasts, and social networking groups.

To access JMP resources on the website, select **Help > JMP User Community** or visit https://community.jmp.com/.

## JMPer Cable

The JMPer Cable is a yearly technical publication targeted to users of JMP. The JMPer Cable is available on the JMP website:

http://www.jmp.com/about/newsletters/jmpercable/

## JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website:

http://www.jmp.com/en_us/software/books.html

## The JMP Starter Window

The JMP Starter window is a good place to begin if you are not familiar with JMP or data analysis. Options are categorized and described, and you launch them by clicking a button. The JMP Starter window covers many of the options found in the Analyze, Graph, Tables, and File menus. The window also lists JMP Pro features and platforms.

• To open the JMP Starter window, select **View** (**Window** on the Macintosh) **> JMP Starter**.

• To display the JMP Starter automatically when you open JMP on Windows, select **File > Preferences > General**, and then select **JMP Starter** from the Initial JMP Window list. On Macintosh, select **JMP > Preferences > Initial JMP Starter Window**.

## Technical Support

JMP technical support is provided by statisticians and engineers educated in SAS and JMP, many of whom have graduate degrees in statistics or other technical disciplines.

Many technical support options are provided at http://www.jmp.com/support, including the technical support phone number.

# Introduction to Profilers

## Visualize Response Surfaces and Optimize Processes

Profiling is an approach to visualizing response surfaces by seeing what would happen if you change just one or two factors at a time. Essentially, a profile is a cross-section view. The interactive profilers in JMP promote exploring opportunity spaces. In fitting equations to data, the fitting is only the first step. Interpreting the fit, understanding the fitted response surface, and finding factor values to optimize the responses is desirable.

**Figure 2.1** Examples of Profilers
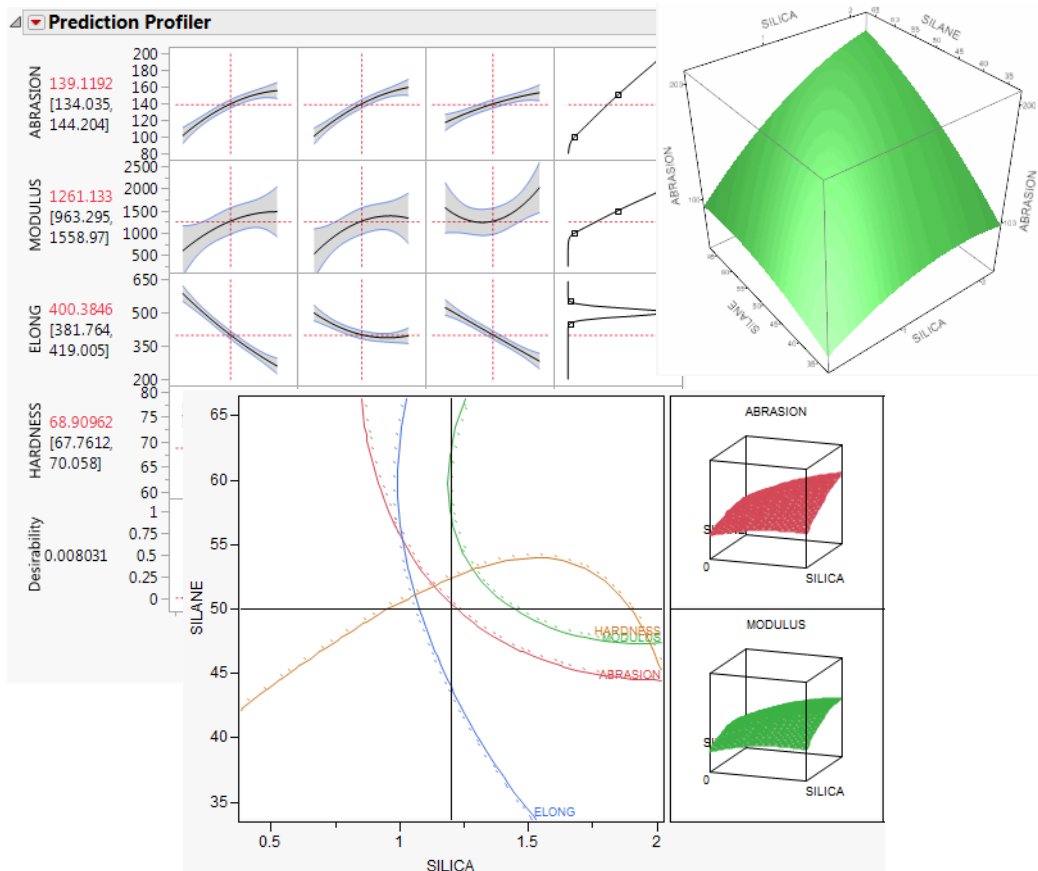
# Profiler Overview

**Note:** For details about the profilers in the Destructive Degradation platform, see the Destructive Degradation chapter in the *Reliability and Survival Methods* book.

The Profiler displays profile traces (see Figure 2.2) for each *X* variable. A *profile trace* is the predicted response as one variable is changed while the others are held constant at the current values. The Profiler recomputes the profiles and predicted responses (in real time) as you vary the value of an *X* variable.

- The vertical dotted line for each *X* variable shows its *current value* or *current setting*. If the variable is nominal, the *x*-axis identifies categories.

  For each *X* variable, the value above the factor name is its current value. Change the current value by clicking in the graph or by dragging the dotted line where you want the new current value to be.

- The horizontal dotted line shows the *current predicted value* of each *Y* variable for the current values of the *X* variables.

- The black lines within the plots show how the predicted value changes when you change the current value of an individual *X* variable. In fitting platforms, the 95% confidence interval for the predicted values is shown by solid blue curves surrounding the prediction trace (for continuous variables) or the height of an error bar (for categorical variables). For continuous variables, the confidence interval region is shaded.

The Profiler is a way of changing one variable at a time and looking at the effect on the predicted response.

**Figure 2.2** Illustration of Traces
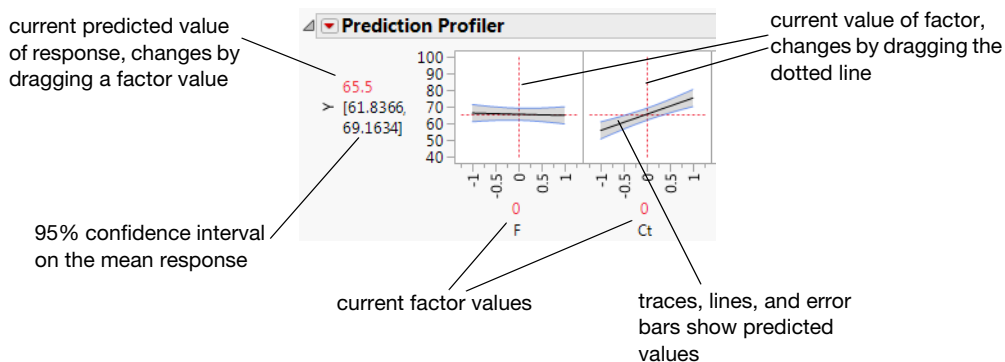


The Profiler in some situations computes confidence intervals for each profiled column. If you have saved both a standard error formula and a prediction formula for the same column, the

Profiler offers to use the standard errors to produce the confidence intervals rather than profiling them as a separate column.

# Introduction to Profiling

It is easy to visualize a response surface with one input factor $X$ and one output factor $Y$. It becomes harder as more factors and responses are added. The profilers in JMP provide a number of highly interactive cross-sectional views of any response surface.

Desirability profiling and optimization features are available to help find good factor settings and produce desirable responses. Most profilers also incorporate multithreading for faster computation. Simulation and defect profiling features are available for when you need to make responses that are robust and high-quality when the factors have variation.

## Profiler Features in JMP

There are several profiler facilities in JMP, accessible from a number of fitting platforms and the main menu under Graph. They are used to profile data column formulas.

**Table 2.1** Profiler Features Summary

|  | Description | Features |
|---|---|---|
| **Profiler (or Prediction Profiler)** | Shows vertical slices across each factor, holding other factors at current values | Desirability, Optimization, Simulator, Propagation of Error |
| **Contour Profiler** | Horizontal slices show contour lines for two factors at a time | Simulator |
| **Surface Profiler** | 3-D plots of responses for 2 factors at a time, or a contour surface plot for 3 factors at a time | Surface Visualization |
| **Mixture Profiler** | A contour profiler for mixture factors | Ternary Plot and Contours |
| **Custom Profiler** | A non-graphical profiler and numerical optimizer | General Optimization, Simulator |
| **Excel Profiler** | Visualize models (or formulas) stored in Excel worksheets. | Profile using Excel Models |

Profiler availability is shown in Table 2.2. The Custom Profiler is available only through the Graph menu. (Model Comparison *does* have Custom Profiler available.)

**Table 2.2** Where to Find JMP Profilers

| Location | Profiler | Contour Profiler | Surface Profiler | Mixture Profiler |
|---|---|---|---|---|
| Graph Menu (as a Platform) | Yes | Yes | Yes | Yes |
| Fit Model: Least Squares | Yes | Yes | Yes | Yes |
| Fit Model: Generalized Regression | Yes | | | |
| Fit Model: Mixed Model | Yes | Yes | Yes | Yes |
| Fit Model: Logistic | Yes | | | |
| Fit Model: Loglinear Variance | Yes | Yes | Yes | |
| Fit Model: Generalized Linear Model | Yes | Yes | Yes | |
| Fit Model: Partial Least Squares | Yes | | | |
| Neural | Yes | Yes | Yes | |
| Model Comparison | Yes | Yes | Yes | Yes |
| Nonlinear: Factors and Response | Yes | Yes | Yes | |
| Nonlinear: Parameters and SSE | Yes | Yes | Yes | |
| Nonlinear: Fit Curve | Yes | | | |
| Gaussian Process | Yes | Yes | Yes | |
| Partial Least Squares | Yes | | | |
| Life Distribution | Yes | | | |
| Fit Life by X | Yes | | Yes | |
| Recurrence Analysis | Yes | | | |
| Choice | Yes | | | |
| Custom Design Prediction Variance | Yes | | Yes | |

**Note:** In this guide, we use the following terms interchangeably:

- factor, input variable, *X* column, independent variable, setting, term
- response, output variable, *Y* column, dependent variable, outcome

The Profiler (with a capital P) is one of several profilers (lowercase p). Sometimes, to distinguish the Profiler from other profilers, we call it the Prediction Profiler.

## Profiler Launch Windows

When a profiler is invoked as a platform from the Graph menu, rather than through a fitting platform, you provide columns with formulas as the Y, Prediction Formula columns. These formulas could have been saved from the fitting platforms.

Figure 2.3 shows an example of most Profiler launch windows.

**Figure 2.3**  Profiler Launch Window
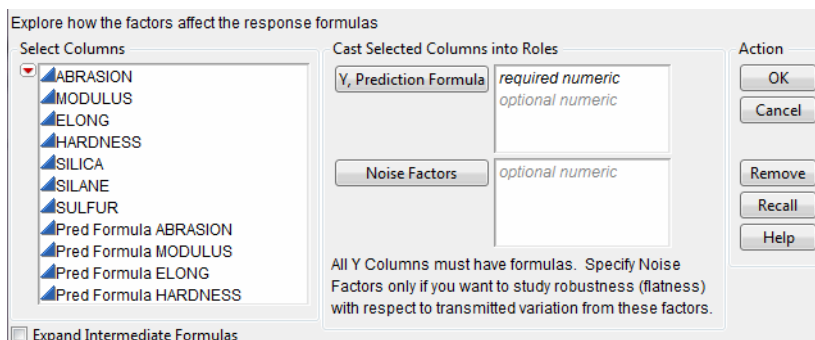


The columns referenced in the formulas become the *X* columns (unless the column is also a *Y*).

**Y, Prediction Formula**   The response columns containing formulas.

**Noise Factors**   Used only in special cases for modeling derivatives. For more information, about noise factors, see "Noise Factors" on page 37.

**Expand Intermediate Formulas**   Tells JMP that if an ingredient column to a formula is a column that itself has a formula, to substitute the inner formula, as long as it refers to other columns. To prevent an ingredient column from expanding, add an Other column property, name it Expand Formula, and assign a value of 0. For more information, see "Expand Intermediate Formulas".

The Surface Plot platform is discussed in the Surface Plot chapter. The Surface Profiler is very similar to the Surface Plot platform, except Surface Plot has more modes of operation. Neither the Surface Plot platform nor the Surface Profiler have some of the capabilities common to other profilers.

### Expand Intermediate Formulas

The Profiler launch window has an Expand Intermediate Formulas check box. When this is checked, when the formula is examined for profiling, if it references another column that has a formula containing references to other columns, then it substitutes that formula and profiles with respect to the end references—not the intermediate column references.

For example, when Fit Model fits a logistic regression for two levels (say A and B), the end formulas (Prob[A] and Prob[B]) are functions of the Lin[x] column, which itself is a function of another column x. If Expand Intermediate Formulas is selected, then when Prob[A] is profiled, it is with reference to x, not Lin[x].

In addition, using the Expand Intermediate Formulas check box enables the Save Expanded Formulas command in the platform red triangle menu. This creates a new column with a formula, which is the formula being profiled as a function of the end columns, not the intermediate columns.

## Fit Group

For the REML and Stepwise personalities of the Fit Model platform, if models are fit to multiple *Y*'s, the results are combined into a Fit Group report. This enables the different *Y*'s to be profiled in the same Profiler. The Fit Group red triangle menu has options for launching the joint Profiler. Profilers for the individual *Y*'s can still be used in the respective Fit Model reports.

Fit Group reports are also created when a By variable is specified for a Stepwise analysis. This allows for the separate models to be profiled in the same Profiler.

The Fit Group scripting command can be used to fit models in different platforms, and have the individual models profiled in the Profiler. For more details, see the JMP Platforms chapter in the *Scripting Guide*.

## Interpret the Profiles

The illustration in Figure 2.4 describes how to use the components of the Profiler. There are several important points to note when interpreting a prediction profile:

- 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 might be curved.

- If you change the value of a factor, the prediction trace for that factor 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.

**Note:** If there are interaction effects or cross-product effects in the model, the prediction traces can shift their slope and curvature as you change current values of other terms. That is what interaction is all about. If there are no interaction effects, the traces only change in height, not slope or shape.

**Figure 2.4**  Changing One Factor from 0 to 0.75
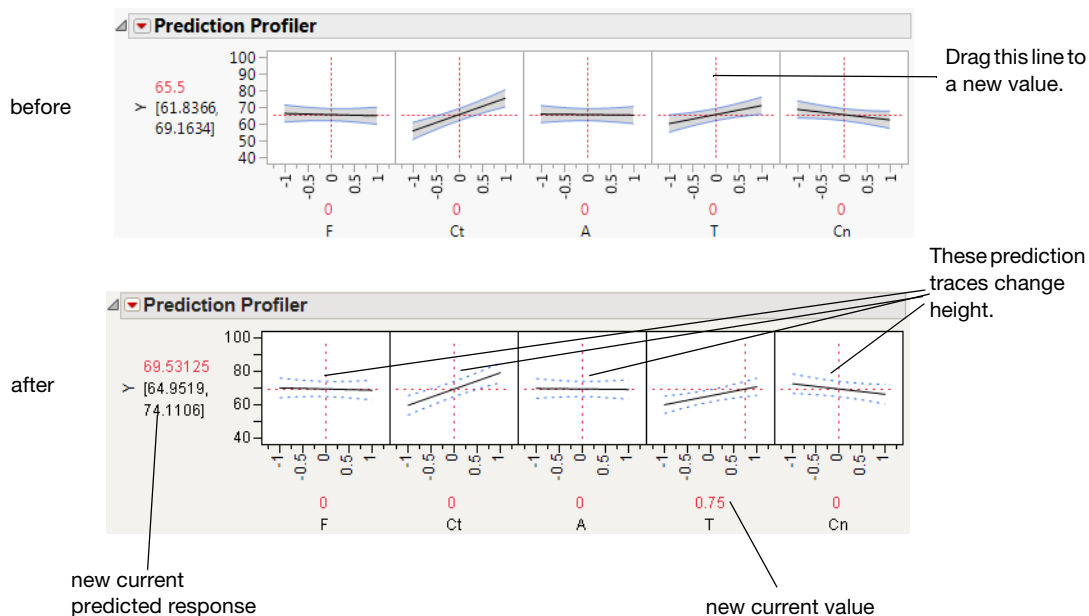


Prediction profiles are especially useful in multiple-response models to help judge which factor values can optimize a complex set of criteria.

Click a graph or drag the current value line right or left to change the factor's current value. The response values change as shown by a horizontal reference line in the body of the graph. Double-click in an axis to bring up a window that changes its settings.

**The Profiler as a Cross-Section**

In the following example using Tiretread.jmp, look at the response surface of the expression for MODULUS as a function of SULFUR and SILANE (holding SILICA constant). Now look at how a grid that cuts across SILANE at the SULFUR value of 2.25. Note how the slice intersects the surface. If you transfer that down below, it becomes the profile for SILANE. Similarly, note the grid across SULFUR at the SILANE value of 50. The intersection when transferred down to the SULFUR graph becomes the profile for SULFUR.

**Figure 2.5**  Profiler as a Cross-Section
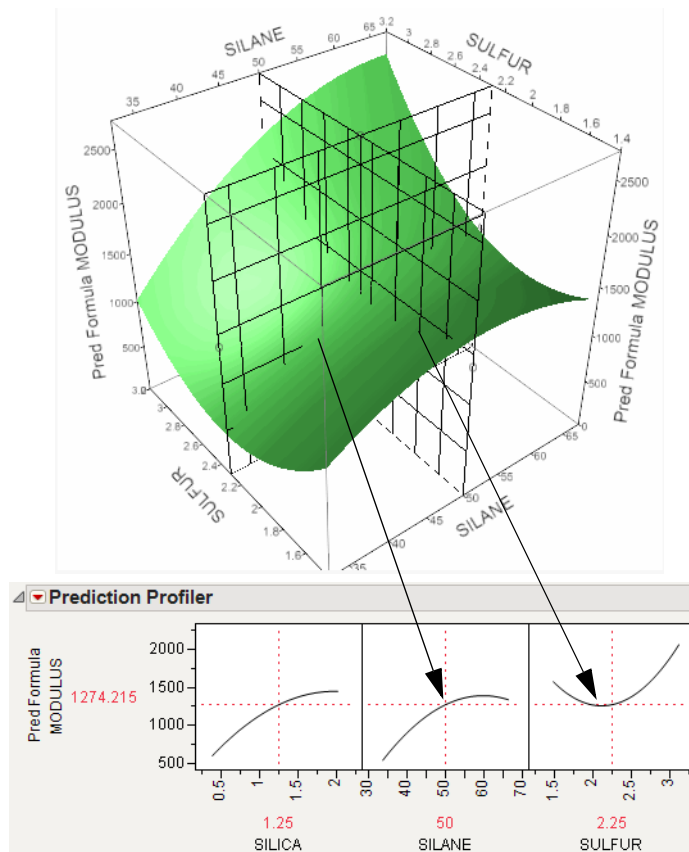


Now consider changing the current value of SULFUR from 2.25 to 1.5.

**Figure 2.6**  Profiler as a Cross-Section
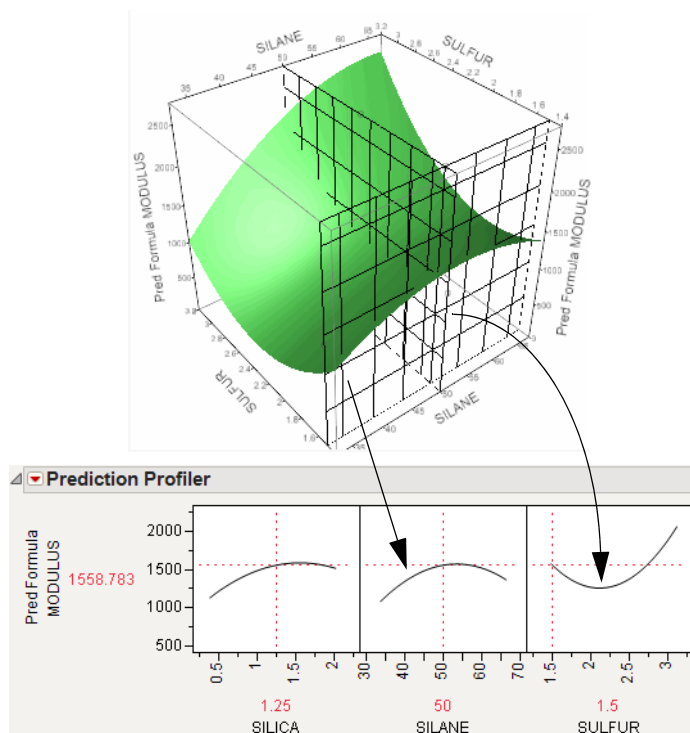


In the Profiler, note the new value just moves along the same curve for SILANE, the SULFUR
curve itself does not change. But the profile for SILANE is now taken at a different cut for
SULFUR. The profile for SILANE is also a little higher and reaches its peak in the different
place, closer to the current SILANE value of 50.

### Set or Lock Factor Values

If you ALT-click (Option-click on the Macintosh) in a graph, a window prompts you to enter
specific settings for the factor.

**Figure 2.7**  Continuous Factor Settings Window

For continuous variables, you can specify:

**Current Value**   The value used to calculate displayed values in the profiler, equivalent to the red vertical line in the graph.

**Minimum Setting**   The minimum value of the factor's axis.

**Maximum Value**   The maximum value of the factor's axis.

**Number of Plotted Points**   Specifies the number of points used in plotting the factor's prediction traces.

**Show**   Show or hide the factor in the profiler.

**Lock Factor Setting**   Locks the value of the factor at its current setting.

## Profiler Platform Options

The red triangle menu on the Profiler title bar has the following options:

**Profiler**   Shows or hides the Profiler.

**Contour Profiler**   Shows or hides the Contour Profiler.

**Custom Profiler**   Shows or hides the Custom Profiler.

**Surface Profiler**   Shows or hides the Surface Profiler.

**Mixture Profiler**   Shows or hides the Mixture Profiler.

**Save for Adobe Flash Platform (.SWF)**   Enables you to save the Profiler (with reduced functionality) as an Adobe Flash file, which can be imported into presentation and web applications. An HTML page can be saved for viewing the Profiler in a browser. The Save for Adobe Flash platform (SWF) command is not available for categorical responses. For more information about this option, go to http://www.jmp.com/support/swfhelp/.

The Profiler accepts any JMP function, but the Flash Profiler only accepts the following functions: Add, Subtract, Multiply, Divide, Minus, Power, Root, Sqrt, Abs, Floor, Ceiling, Min, Max, Equal, Not Equal, Greater, Less, GreaterorEqual, LessorEqual, Or, And, Not, Exp, Log, Log10, Sine, Cosine, Tangent, SinH, CosH, TanH, ArcSine, ArcCosine, ArcTangent, ArcSineH, ArcCosH, ArcTanH, Squish, If, Match, Choose.

**Note:** Some platforms create column formulas that are not supported by the Save As Flash option.

**Show Formulas**   Opens a JSL window showing all formulas being profiled.

**Formulas for OPTMODEL**   Creates code for the OPTMODEL SAS procedure. Hold down CTRL and SHIFT and then select **Formulas for OPTMODEL** from the red triangle menu.

The following options are available in many platforms. See the JMP Reports chapter in the *Using JMP* book for more information.

**Redo**   Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

**Save Script**   Contains options that enable you to save a script that reproduces the report to several destinations.

# Common Profiler Topics

This section contains information on features that are common to more than one of the profiler platforms.

## Linear Constraints

The Prediction Profiler, Custom Profiler, and Mixture Profiler can incorporate linear constraints into their operations. Linear constraints can be entered in two ways, described in the following sections.

### Red Triangle Menu Item

To enter linear constraints via the red triangle menu, select **Alter Linear Constraints** from either the Prediction Profiler or Custom Profiler red triangle menu.

Choose **Add Constraint** from the resulting window, and enter the coefficients into the appropriate boxes. For example, to enter the constraint p1 + 2*p2 ≥ 0.9, enter the coefficients as shown in Figure 2.8. As shown, if you are profiling factors from a mixture design, the mixture constraint is present by default and cannot be modified.

**Figure 2.8**  Enter Coefficients



After you click **OK**, the Profiler updates the profile traces, and the constraint is incorporated into subsequent analyses and optimizations.

If you attempt to add a constraint for which there is no feasible solution, a message is written to the log and the constraint is not added. To delete a constraint, enter zeros for all the coefficients.

Constraints added in one profiler are not accessible by other profilers until saved. For example, if constraints are added under the Prediction Profiler, they are not accessible to the Custom Profiler. To use the constraint, you can either add it under the Custom Profiler red triangle menu, or use the **Save Linear Constraints** command described in the next section.

### Constraint Table Property/Script

If you add constraints in one profiler and want to make them accessible by other profilers, use the **Save Linear Constraints** command, accessible through the platform red triangle menu. For example, if you created constraints in the Prediction Profiler, choose **Save Linear Constraints** under the Prediction Profiler red triangle menu. The Save Linear Constraints command creates or alters a Table Script called Constraint. An example of the Table Property is shown in Figure 2.9.

**Figure 2.9**  Constraint Table Script



The Constraint Table Property is a list of the constraints, and is editable. It is accessible to other profilers, and negates the need to enter the constraints in other profilers. To view or edit Constraint, right-click the red triangle menu and select **Edit**. The content of the constraint from Figure 2.8 is shown below in Figure 2.10.

**Figure 2.10**  Example Constraint



The Constraint Table Script can be created manually by choosing **New Script** from the red triangle menu beside a table name.

**Note:** When creating the Constraint Table Script manually, the spelling must be exactly "Constraint". Also, the constraint variables are case sensitive and must match the column name. For example, in Figure 2.10, the constraint variables are p1 and p2, not P1 and P2.

The Constraint Table Script is also created when specifying linear constraints when designing an experiment.

The Alter Linear Constraints and Save Linear Constraints commands are not available in the Mixture Profiler. To incorporate linear constraints into the operations of the Mixture Profiler, the Constraint Table Script must be created by one of the methods discussed in this section.

## Noise Factors

**Note:** Noise factor optimization is also available in the Prediction Profiler, Contour Profiler, Custom Profiler, and Mixture Profiler.

Robust process engineering enables you to produce acceptable products reliably, despite variation in the process variables. Even when your experiment has controllable factors, there is a certain amount of uncontrollable variation in the factors that affects the response. This is called transmitted variation. Factors with this variation are called noise factors. Some factors you cannot control at all, like environmental noise factors. The mean for some factors can be controlled, but not their standard deviation. This is often the case for intermediate factors that are output from a different process or manufacturing step.

A good approach to making the process robust is to match the target at the flattest place of the noise response surface. Then, the noise has little influence on the process. Mathematically, this is the value where the first derivatives of each response with respect to each noise factor are zero. JMP computes the derivatives for you.

**Figure 2.11**  Noise Factor Example
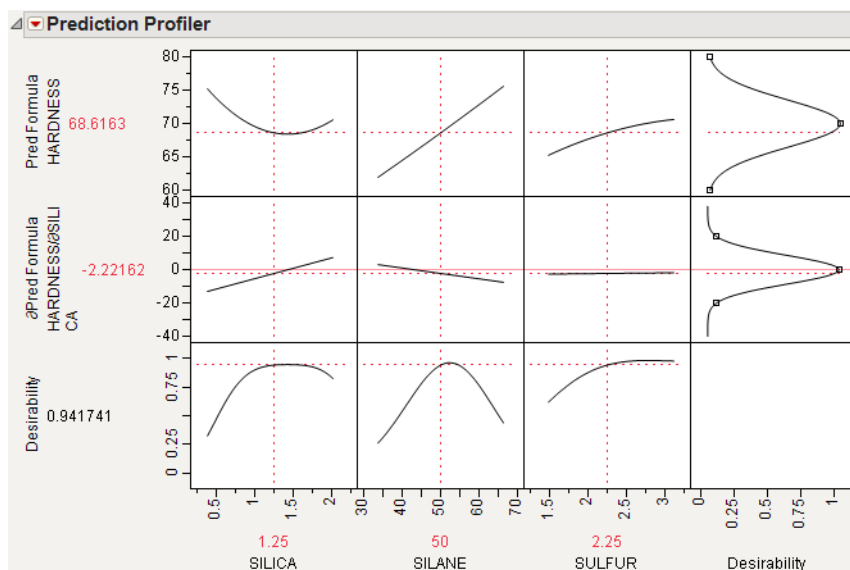


Noise factors (robust process engineering) enables you to produce acceptable products reliably, despite variation in the process variables. Even when your experiment has

controllable factors, there is a certain amount of uncontrollable variation in the factors that affects the response. This is called *transmitted variation*. Factors with this variation are called *noise factors*. You cannot control some factors, like environmental noise factors. The mean for some factors can be controlled, but their standard deviation is uncontrollable. This is often the case for intermediate factors that are output from a different process or manufacturing step.

A good approach to making the process robust is to match the target at the flattest place of the noise response surface. Then, the noise has little influence on the process. Mathematically, this is the value where the first derivatives of each response with respect to each noise factor are zero. JMP computes the derivatives for you.

To analyze a model with noise factors:

1. Fit the appropriate model (for example, using the Fit Model platform).

2. Save the model to the data table with the **Save > Prediction Formula** command.

3. Launch the **Profiler** (from the **Graph** menu).

4. Assign the prediction formula to the **Y, Prediction Formula** role and the noise factors to the **Noise Factor** role.

5. Click **OK**.

The resulting profiler shows response functions and their appropriate derivatives with respect to the noise factors, with the derivatives set to have maximum desirability at zero.

6. Select **Optimization and Desirability > Maximize Desirability** from the **Profiler** menu.

This finds the best settings of the factors, balanced with respect to minimizing transmitted variation from the noise factors.

# Profiler

## Explore Cross Sections of Responses across Each Factor

The Prediction Profiler, or simply, Profiler, gives you a wealth of information about your model. Use the Prediction Profiler to:

- See how your prediction model changes as you change settings of individual factors.
- Set desirability goals for your response or responses, and find optimal settings for your factors.
- Gauge your model's sensitivity to changes in the factors, where sensitivity is based on your predictive model.
- Assess the importance of your factors relative to model predictions, in a way that is independent of the model.
- Simulate your response distribution based on specified distributions for both factors and responses, and control various aspects of the appearance of the profiler.

**Figure 3.1** Profiler for Four Responses with Simulator and Importance Coloring
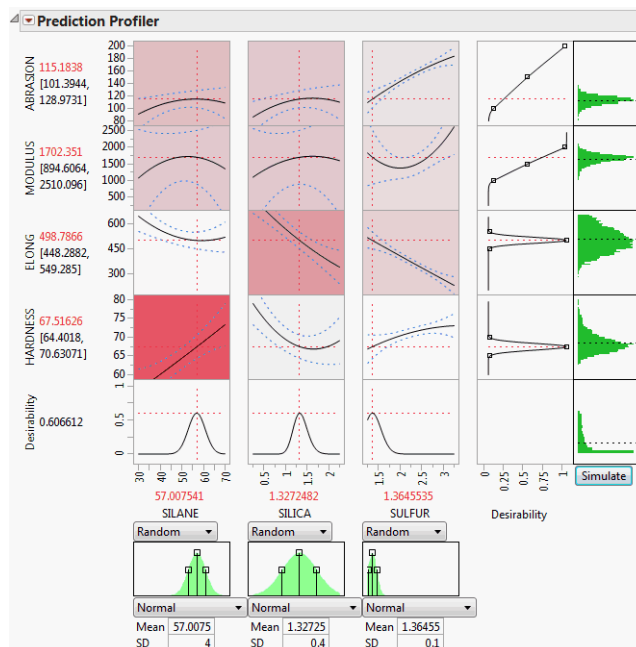
# Example of the Prediction Profiler

This example uses the Tiretread.jmp sample data table. There are three factors (SILICA, SILANE, and SULFUR) and four responses (ABRASION, MODULUS, ELONG, and HARDNESS). The goal is to find the optimal combination of the three factors in the development of a tire tread compound. See Derringer and Suich (1980).

1. Select **Help > Sample Data Library** and open Tiretread.jmp.

2. Select **Graph > Profiler**.

3. Select Pred Formula Abrasion and click **Y, Prediction Formula**.

**Figure 3.2**  Completed Profiler Launch Window
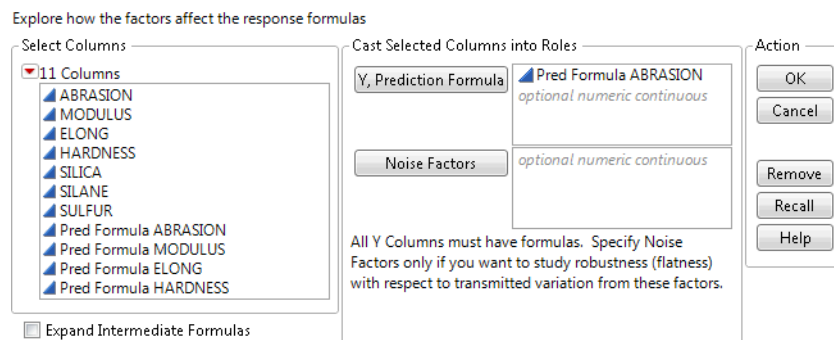


4. Click **OK**.

**Figure 3.3**  Profiler Report Window
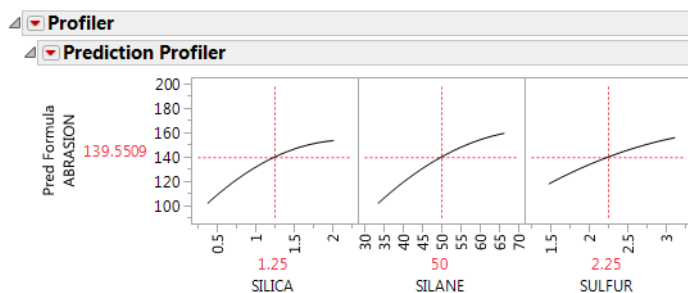


The profiler is interactive. The vertical red lines correspond to the current value of the factors. The current value of each factor is also shown in red below the horizontal axis. The red value on the vertical axis is the predicted response based on the current values of the factors. Click and drag the vertical red lines to change the current values of the factors. Click a factor's red value to set the factor to a specific value. To maximize ABRASION, set

all factors to their highest levels. For details about optimizing the response see "Desirability Profiling and Optimization" on page 47.

## Launch the Prediction Profiler Platform

The Prediction Profiler can be accessed in the following ways:

- The Prediction Profiler can be accessed directly from the Graph menu. When you access the Prediction Profiler in this way, the Prediction Profiler launch window appears. See "Profiler Launch Windows" on page 29 in the "Introduction to Profilers" chapter for details.

- The Prediction Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" on page 28 in the "Introduction to Profilers" chapter for details about the availability of the Prediction Profiler in different platforms.

- The Prediction Profiler can be accessed from the Model Comparisons platform. Select Profiler from the Model Comparisons red triangle menu.

- **JMP PRO** The Prediction Profiler can be accessed from the Formula Depot platform. Select Profiler from the Formula Depot red triangle menu.

## Prediction Profiler Options

The red triangle menu on the Prediction Profiler title bar has the following options:

**Optimization and Desirability**   Submenu that consists of the following options:

**Desirability Functions**   Shows or hides the desirability functions. Desirability is discussed in "Desirability Profiling and Optimization" on page 47.

**Maximize Desirability**   Sets the current factor values to maximize the desirability functions. Takes into account the response importance weights.

**Note:** In many situations, the settings that optimize the desirability function are not unique. The Maximize Desirability option gives one such setting. The Contour Profiler is a good tool for finding alternative factor combinations that optimize desirability. For an example, see "Explore Optimal Settings" on page 83 in the "Contour Profiler" chapter.

**Note:** If a factor has a Design Role column property value of Discrete Numeric, it is treated as continuous in the optimization of the desirability function. To account for the fact that the factor can assume only discrete levels, it is displayed in the profiler as a categorical term and an optimal allowable level is selected.

**Maximize and Remember**   Maximizes the desirability functions and remembers the associated settings.

**Maximization Options**   Opens the Maximization Options window where you can refine the optimization settings.

**Figure 3.4** Maximization Options Window



**Maximize for Each Grid Point**   Used only if one or more factors are locked. The ranges of the locked factors are divided into a grid, and the desirability is maximized at each grid point. This is useful if the model that you are profiling has categorical factors. Then the optimal condition can be found for each combination of the categorical factors.

**Save Desirabilities**   Saves the three desirability function settings for each response, and the associated desirability values, as a Response Limits column property in the data table. These correspond to the coordinates of the handles in the desirability plots.

**Set Desirabilities**   Opens the Response Goal window where you can set specific desirability values.

**Figure 3.5** Response Goal Window



**Save Desirability Formula**   Creates a column in the data table with a formula for Desirability. The formula uses the fitting formula when it can, or the response variables when it cannot access the fitting formula.

**Assess Variable Importance**   Provides different approaches to calculating indices that measure the importance of factors to the model. These indices are independent of the model type and fitting method. Available only for continuous or binary responses. For details, see "Assess Variable Importance" on page 53.

**Bagging**   (Available only when the Prediction Profiler is embedded in certain modeling platforms.) Launches the Bagging window. Bootstrap aggregating (bagging) enables you to create multiple training data sets by sampling with replacement from the original data. For each training set, a model is fit using the analysis platform, and predictions are made. The final prediction is a combination of the results from all of the models. This improves prediction performance by reducing the error from variance. For details, see "Bagging" on page 56.

**Simulator**   Launches the Simulator. The Simulator enables you to create Monte Carlo simulations using random noise added to factors and predictions for the model. A typical use is to set fixed factors at their optimal settings, and uncontrolled factors and model noise to random values. You then find out the rate of responses outside the specification limits. For details, see the "Simulator" chapter on page 127.

**Interaction Profiler**   Shows or hides interaction plots that update as you update the factor values in the Profiler. Use this option to help visualize third degree interactions by seeing how the plot changes as current values for the factors change. The cells that change for a given factor are the cells that do not involve that factor directly.

**Confidence Intervals**   Shows or hides confidence intervals in the Profiler plot. The intervals are drawn by bars for categorical factors, and curves for continuous factors. These are available when the profiler is used inside certain fitting platforms or when a standard error column has been specified in the Profiler launch dialog.

**Prop of Error Bars**   (Appears when a Sigma column property exists in any of the factor and response variables.) This option displays the $3\sigma$ interval that is implied on the response due to the variation in the factor. Propagation of error (POE) is important when attributing the variation of the response in terms of variation in the factor values when the factor values are not very controllable. See "Propagation of Error Bars" on page 75.

**Sensitivity Indicator**   Shows or hides a purple triangle whose height and direction correspond to the value of the partial derivative of the profile function at its current value (see Figure 3.6). This is useful in large profiles to be able to quickly spot the sensitive cells.

**Figure 3.6** Sensitivity Indicators
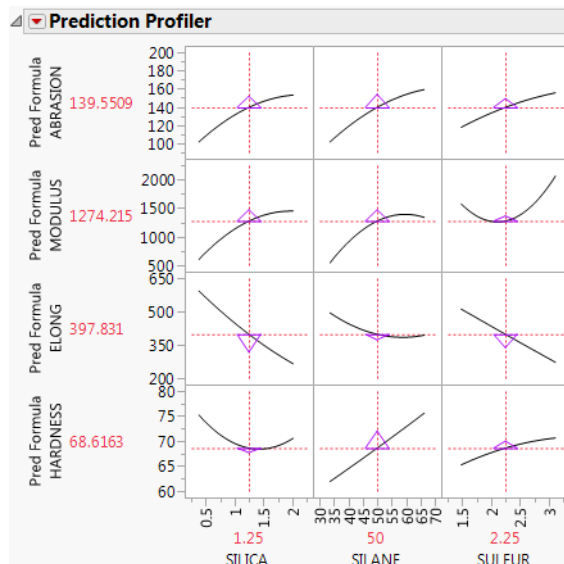


**Profile at Boundary**   When analyzing a mixture design, JMP constrains the ranges of the factors so that settings outside the mixture constraints are not possible. This is why, in some mixture designs, the profile traces turn abruptly.

When there are mixture components that have constraints, other than the usual zero-to-one constraint, a new submenu, called Profile at Boundary, appears on the Prediction Profiler red triangle menu. It has the following two options:

**Turn At Boundaries**   Lets the settings continue along the boundary of the restraint condition.

**Stop At Boundaries**   Truncates the prediction traces to the region where strict proportionality is maintained.

**Reset Factor Grid**   Displays a window for each factor enabling you to enter a specific value for the factor's current setting, to lock that setting, and to control aspects of the grid. See the section "Set or Lock Factor Values" on page 33 for details.

**Figure 3.7** Factor Settings Window
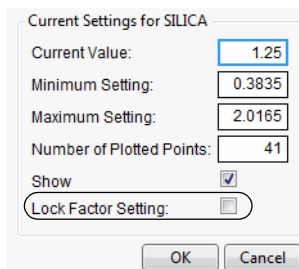


**Factor Settings**   Submenu that consists of the following options:

> **Remember Settings**   Adds an outline node to the report that accumulates the values of the
>   current settings each time the Remember Settings command is invoked. Each
>   remembered setting is preceded by a radio button that is used to reset to those settings.

> **Set To Data in Row**   Assigns the values of a data table row to the *X* variables in the
>   Profiler.

> **Copy Settings Script**   Copies the current Profiler's settings to the clipboard.

> **Paste Settings Script**   Pastes the Profiler settings from the clipboard to a Profiler in
>   another report.

> **Append Settings to Table**   Appends the current profiler's settings to the end of the data
>   table. This is useful if you have a combination of settings in the Profiler that you want
>   to add to an experiment in order to do another run.

> **Link Profilers**   Links all the profilers together. A change in a factor in one profiler causes
>   that factor to change to that value in all other profilers, including Surface Plot. This is a
>   global option, set, or unset for all profilers.

> **Set Script**   Sets a script that is called each time a factor changes. The set script receives a
>   list of arguments of the form:
>
>   ```
>   {factor1 = n1, factor2 = n2, ...}
>   ```
>
>   For example, to write this list to the log, first define a function:
>
>   ```
>   ProfileCallbackLog = Function({arg},show(arg));
>   ```
>
>   Then enter `ProfileCallbackLog` in the Set Script dialog.
>
>   Similar functions convert the factor values to global values:
>
>   ```
>   ProfileCallbackAssign = Function({arg},evalList(arg));
>   ```
>
>   Or access the values one at a time:
>
>   ```
>   ProfileCallbackAccess =
>   Function({arg},f1=arg["factor1"];f2=arg["factor2"]);
>   ```

**Unthreaded**   Enables you to change to an unthreaded analysis if multithreading does not work.

**Default N Levels**   Enables you to set the default number of levels for each continuous factor. This option is useful when the Profiler is especially large. When calculating the traces for the first time, JMP measures how long it takes. If this time is greater than three seconds, you are alerted that decreasing the Default N Levels speeds up the calculations.

**Output Grid Table**   Produces a new data table with columns for the factors that contain grid values, columns for each of the responses with computed values at each grid point, and the desirability computation at each grid point.

If you have a lot of factors, it is impractical to use the Output Grid Table command, because it produces a large table. A memory allocation message might be displayed for large grid tables. In such cases, you should lock some of the factors, which are held at locked, constant values. To get the window to specify locked columns, ALT- or Option-click inside the profiler graph to get a window that has a Lock Factor Setting check box.

**Output Random Table**   Prompts for a number of runs and creates an output table with that many rows, with random factor settings and predicted values over those settings. This is equivalent to (but much simpler than) opening the Simulator, resetting all the factors to a random uniform distribution, then simulating output. This command is similar to Output Grid Table, except it results in a random table rather than a sequenced one.

The prime reason to make uniform random factor tables is to explore the factor space in a multivariate way using graphical queries. This technique is called *Filtered Monte Carlo*.

Suppose you want to see the locus of all factor settings that produce a given range to desirable response settings. By selecting and hiding the points that do not qualify (using graphical brushing or the Data Filter), you see the possibilities of what is left: the opportunity space yielding the result that you want.

Some rows might appear selected and marked with a red dot. These represent the points on the multivariate desirability Pareto Frontier - the points that are not dominated by other points with respect to the desirability of all the factors.

**Alter Linear Constraints**   Enables you to add, change, or delete linear constraints. The constraints are incorporated into the operation of Prediction Profiler. See "Linear Constraints" on page 35.

**Save Linear Constraints**   Enables you to save existing linear constraints to a table script called Constraint. See "Linear Constraints" on page 35.

**Conditional Predictions**   Appears when random effects are included in the model. The random effects predictions are used in formulating the predicted value and profiles.

**Appearance**   Submenu that consists of the following options:

**Arrange in Rows**   Enter the number of plots that appear in a row. This option helps you view plots vertically rather than in one wide row.

**Reorder X Variables**   Opens a window where you can reorder the model main effects by dragging them to the desired order.

**Reorder Y Variables**   Opens a window where you can reorder the responses by dragging them to the desired order.

**Adapt Y Axis**   Re-scales the Y axis if the response is outside the axis range, so that the range of the response is included.

## Desirability Profiling and Optimization

You can define a desirability function for a single response variable or for several response variables. When you are optimizing relative to several responses, there can often be competing criteria. 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 for all responses is defined as the geometric mean of the desirability functions for the individual responses. See Derringer and Suich (1980) for information about combining responses.

To use desirability profiling, select **Optimization and Desirability > Desirability Functions** from the Prediction Profiler red triangle menu.

**Note:** If the response column has a Response Limits property, desirability functions are turned on by default.

This command appends a new row to the bottom of the plot matrix, dedicated to graphing desirability. The row has a plot for each factor showing its *desirability trace*, as illustrated in Figure 3.8. It also adds a column that has an adjustable desirability function for each *Y* variable. The overall desirability measure shows on a scale of zero to one at the left of the row of desirability traces.

**Figure 3.8** The Desirability Profiler
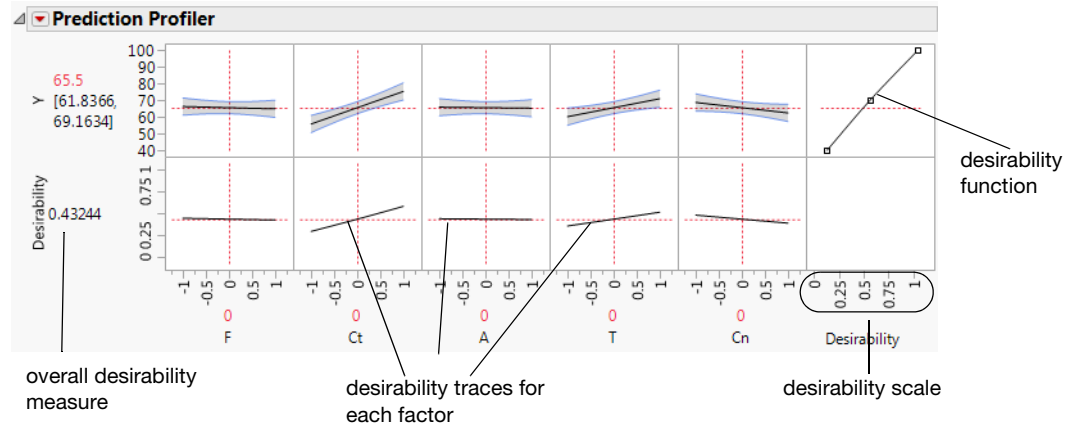


## Construction of Desirability Functions

The individual desirability functions are smooth piecewise functions that pass through three defining points. These points are called control points (Low, Middle, High) and can be used to interactively control the shape of the desirability function.

- The Minimize and Maximize functions are three-part piecewise smooth functions that consist of interpolating cubics between the control points and exponentials in the tails.

- The Target function is a piecewise function that is a scale multiple of a normal density on either side of the Middle value (with different curves on each side), which is also piecewise smooth and fit to the control points. Exponential functions are fit to the tails.

- The None function enables you to specify an arbitrary desirability function. In particular, you can specify desirability to be lower at the Middle value than at the Low and High values. You can also construct custom desirability functions using formulas. See "Customized Desirability Functions" on page 51.

The Low and High control points are not allowed to reach all the way to zero or one. This approach to constructing the desirability functions results in good behavior as the desirability values switch between the maximize, target, and minimize values.

**Note:** JMP does not use the Derringer and Suich (1980) functional forms. Because they are not smooth, they do not always work well with JMP's optimization algorithm.

### Desirability Function for Multiple Optimization

When multiple responses are to be optimized, an overall desirability function is constructed and optimized. The overall desirability for all responses is defined as the geometric mean of the desirability functions for the individual responses.

Denote the individual desirability functions for $k$ responses by $d_1$, $d_2$,...,$d_k$. Then the overall desirability function is the geometric mean of the individual desirability functions:

$$D = d_1^{1/k} d_2^{1/k} ... d_k^{1/k}$$

If Importance values are defined as part of the Response Limits column property or are defined in the Response Goal window, they are integrated into the overall desirability function. The Importance values are scaled so that they sum to 1. Denote the scaled importance values by $w_1$, $w_{2, ...,}$ $w_k$. Then the overall desirability is defined as a weighted geometric mean of the individual desirability functions:

$$D = d_1^{w_1} d_2^{w_2} ... d_k^{w_k}$$

**Optimization Algorithm**

Optimization of the overall desirability function, or of the single desirability function if there is only one response, is conducted as follows.

- For categorical factors, a coordinate exchange algorithm is used.

- For continuous factors, a gradient descent algorithm is used.

- In the presence of constraints or mixture factors, a Wolfe reduced-gradient approach is used.

## How to Use the Desirability Function

To use a variable's desirability function, drag the function handles to represent a response value.

As you drag these handles, the changing response value shows in the area labeled Desirability to the left of the plots. The dotted line is the response for the current factor settings. The overall desirability shows to the left of the row of desirability traces. Alternatively, you can select **Optimization and Desirability > Set Desirabilities** to enter specific values for the points.

Figure 3.9 shows steps to create desirability settings.

**Maximize**   The default desirability function setting is maximize ("higher is better"). The High control point is positioned at the maximum Y value and aligned at the high desirability, close to 1. The Low control point is positioned at the minimum Y value and aligned at a low desirability, close to 0.

**Figure 3.9** Maximizing Desirability
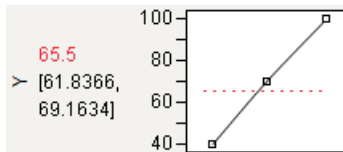


**Target**  You can designate a target value as "best." In this example, the middle control point is
positioned at Y = 70 and aligned with the maximum desirability of 1. Y becomes less
desirable as its value approaches either 40 or 100. The High and Low control points at
Y = 40 and Y = 100 are positioned at the minimum desirability close to 0.

**Figure 3.10** Defining a Target Desirability
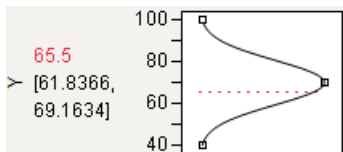


**Minimize**  The minimize ("smaller is better") desirability function associates high response
values with low desirability and low response values with high desirability. The curve is
the maximization curve flipped around a horizontal line at the center of plot.

**Figure 3.11** Minimizing Desirability
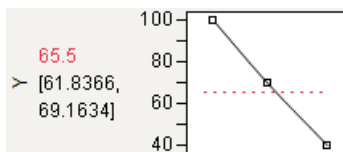


**Note:** Dragging the High or Low control point of a maximize or minimize desirability
function across the *y*-value of the middle point results in the opposite point reflecting. A
Minimize becomes a Maximize, and vice versa.

## The Desirability Profile

The last row of plots shows the desirability trace for each factor. The numerical value beside
the word Desirability on the vertical axis is the geometric mean of the desirability measures.
This row of plots shows both the current desirability and the trace of desirabilities that result
from changing one factor at a time.

For example, Figure 3.12 shows desirability functions for two responses. You want to maximize ABRASION and MODULUS. The desirability plots indicate that you could increase the desirability by increasing any of the factors.

**Figure 3.12** Prediction Profile Plot with Adjusted Desirability and Factor Values
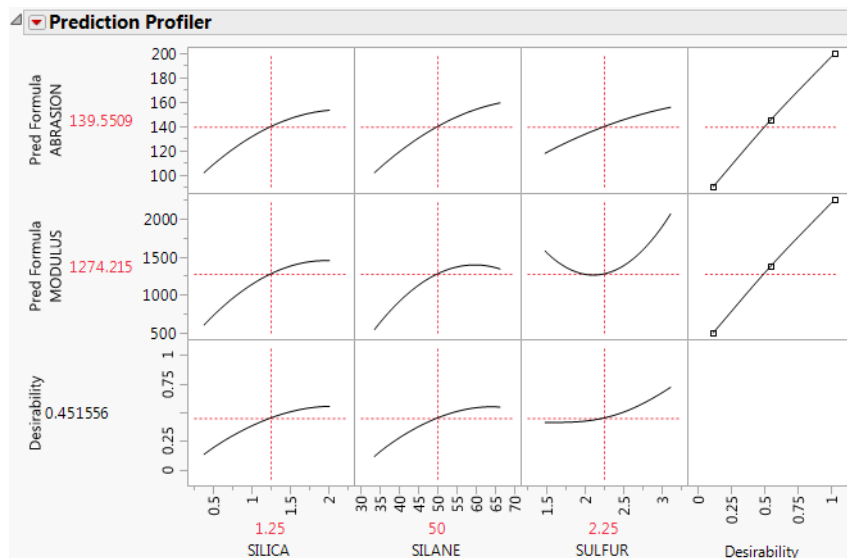


## Customized Desirability Functions

It is possible to use a customized desirability function. For example, suppose you want to maximize using the following function.

**Figure 3.13** Maximizing Desirability Based on a Function

$$\frac{Pred\ Formula\ ABRASION}{96} + \frac{Pred\ Formula\ MODULUS}{700} + \frac{33}{(|Pred\ Formula\ ELONG\text{-}450|+1)} + \frac{2}{(|Pred\ Formula\ HARDNESS\text{-}67|+1)}$$

First, create a column called Custom Desirability that contains the above formula. Then, select **Graph > Profiler** to launch the platform. Select all the Pred Formula columns and the Custom Desirability column and select **Y, Prediction Formula**. Turn on the desirability functions by selecting **Optimization and Desirability > Desirability Functions** from the red triangle menu. All the desirability functions for the individual effects must be turned off. To do this, first

double-click in a desirability plot window, and then select **None** in the window that appears (Figure 3.14). Set the desirability for Custom Desirability to be maximized.

**Figure 3.14**  Selecting No Desirability Goal
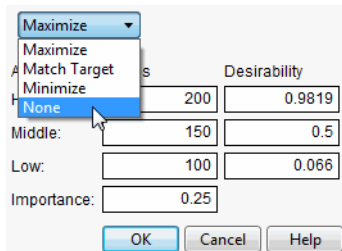


At this point, selecting **Optimization and Desirability > Maximize Desirability** uses only the custom Custom Desirability function.

**Figure 3.15**  Maximized Custom Desirability
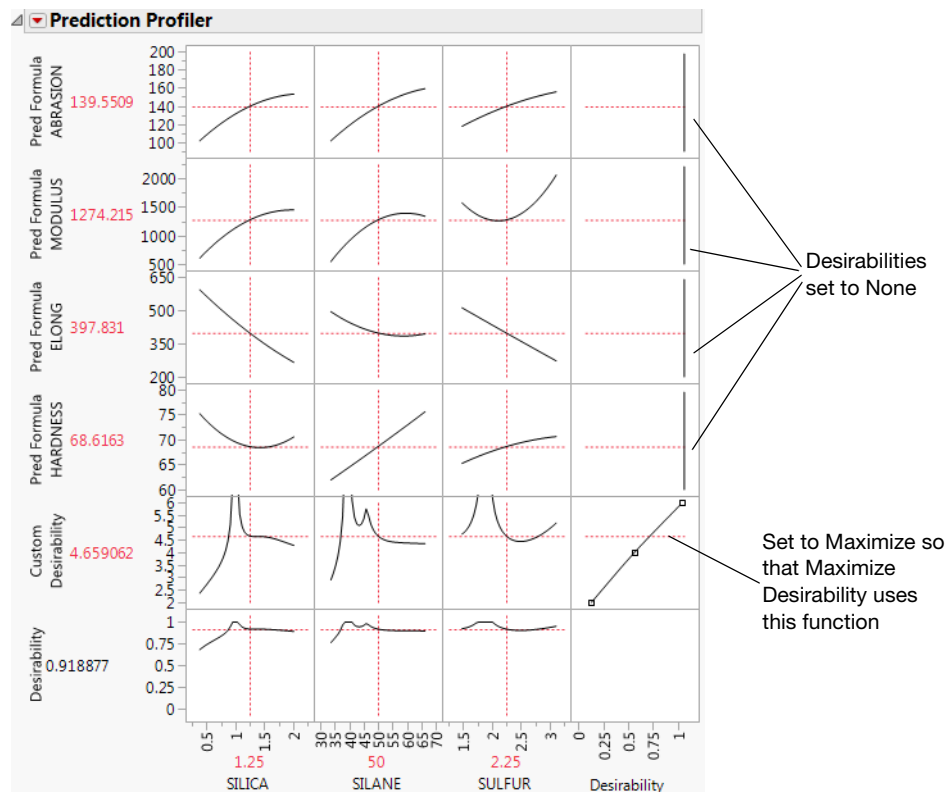
# Assess Variable Importance

For continuous and binary responses, the Variable Importance report calculates indices that measure the importance of factors in a model in a way that is independent of the model type and fitting method. The fitted model is used only in calculating predicted values. The method estimates the variability in the predicted response based on a range of variation for each factor. If variation in the factor causes high variability in the response, then that effect is important relative to the model.

**Note:** Assess Variable Importance is not available for categorical responses with more than two levels.

Assess Variable Importance can also be accessed in the Profiler that is obtained through the Graph menu.

For statistical details, see "Assess Variable Importance" on page 74. See also Saltelli, 2002.

**Note:** Assess Variable Importance requires that all columns reside in the same data table.

### The Assess Variable Importance Report

The Assess Variable Importance menu has the following options that address the methodology used in constructing importance indices:

**Independent Uniform Inputs**   For each factor, Monte Carlo samples are drawn from a uniform distribution defined by the minimum and maximum observed values. Use this option when you believe that your factors are uncorrelated and that their likely values are uniformly spread over the range represented in the study. This is the appropriate option for designed experiments that do not involve constraints or mixture factors.

**Independent Resampled Inputs**   For each factor, Monte Carlo samples are obtained by resampling its set of observed values. Use this option when you believe that your factors are uncorrelated and that their likely values are not represented by a uniform distribution.

**Dependent Resampled Inputs**   Factor values are constructed from observed combinations using a *k*-nearest neighbors approach, in order to account for correlation. This option treats observed variance and covariance as representative of the covariance structure for your factors. Use this option when you believe that your factors are correlated. Note that this option is sensitive to the number of rows in the data table. If used with a small number of rows, the results can be unreliable.

**Note:** The Independent Resampled Inputs and Dependent Resampled Inputs options are intended for observational studies. The Independent option is faster than the Dependent option, but the Dependent option handles multicollinearity better and does not extrapolate into regions far away from the data.

**Linearly Constrained Inputs**   For each factor, Monte Carlo samples are drawn from a uniform distribution over a region defined by linear constraints. The linear constraints can be defined in the Profiler or constructed in connection with a designed experiment. In addition, the samples are restricted to fall within the minimum and maximum observed values. Use this option in the presence of linear constraints, when you believe that these constraints impact the distribution of the inputs.

The speed of these algorithms depends on the model evaluation speed. In general, the fastest option is Independent Uniform Inputs and the slowest is Dependent Resampled Inputs. You have the option to Accept Current Indices when the estimation process is unable to complete instantaneously.

**Note:** In the case of independent and linearly constrained inputs, variable importance indices are constructed using Monte Carlo sampling. For this reason, you can expect some variation in importance index values from one run to another.

## Variable Importance Report

Each Assess Variable Importance option presents a Summary Report and Marginal Model Plots. When the Assess Variable Importance report opens, the factors in the Profiler are reordered according to their Total Effect importance indices. When there are multiple responses, the factors are reordered according to the Total Effect importance indices in the Overall report. When you run several Variable Importance reports, the factors in the Profiler are ordered according to their Total Effect indices in the most recent report.

### Summary Report

For each response, a table displays the following elements:

**Column**   The factor of interest.

**Main Effect**   An importance index that reflects the relative contribution of that factor alone, not in combination with other factors.

**Total Effect**   An importance index that reflects the relative contribution of that factor both alone and in combination with other factors. The Total Effect column is displayed as a bar chart. See "Weights" on page 55.

**Main Effect Std Error**   The Monte Carlo standard error of the Main Effect's importance index. This is a hidden column that you can access by right-clicking in the report and selecting **Columns > Main Effect Std Error**. By default, sampling continues until this error is less than

0.01. Details of the calculation are given in "Variable Importance Standard Errors" on page 75. (Not available for Dependent Resampled Inputs option.)

**Total Effect Std Error**   The Monte Carlo standard error of the Total Effect's importance index. This is a hidden column that you can access by right-clicking in the report and selecting **Columns > Total Effect Std Error**. By default, sampling continues until this error is less than 0.01. Details of the calculation are given in "Variable Importance Standard Errors" on page 75. (Not available for Dependent Resampled Inputs option.)

**Weights**   A plot that shows the Total Effect indices, located to the right of the final column. You can deselect or reselect this plot by right-clicking in the report and selecting **Columns > Weights**.

**Proportion of function evaluations with missing values**   The proportion of Monte Carlo samples for which some combination of inputs results in an inestimable prediction. When the proportion is nonzero, this message appears as a note at the bottom of the table.

---

**Note:** When you have more than one response, the Summary Report presents an Overall table followed by tables for each response. The importance indices in the Overall report are the averages of the importance indices across all responses.

---

**Marginal Model Plots**

The Marginal Model Plots report (see Figure 3.35) shows a matrix of plots, with a row for each response and columns for the factors. The factors are ordered according to the size of their overall Total Effect importance indices.

For a given response and factor, the plot shows the mean response for each factor value, where that mean is taken over all inputs to the calculation of importance indices. These plots differ from profiler plots, which show cross sections of the response. Marginal Model Plots are useful for assessing the main effects of factors.

Note that your choice of input methodology impacts the values plotted on marginal model plots. Also, because the plots are based on the generated input settings, the plotted mean responses might not follow a smooth curve.

The red triangle options enable you to show or hide the following aspects of the plots:

**Estimate**   A smoothed estimate of the mean of the simulated values calculated as a function of the factor values.

---

**Note:** The estimates of the mean are simulated, so the values change when you rerun the analysis.

---

**Confidence Interval**   A 95% confidence band for the simulated means. This band is often narrow and might not be visible unless you expand the scale. Not available for Dependent Resampled Inputs.

> **Note:** The confidence bounds are simulated, so the bands change when you rerun the analysis.

**Data**    The actual (unsimulated) values of the response plotted against the factor values.

**Variable Importance Options**

The Variable Importance report has the following red triangle options:

**Reorder factors by main effect importance**    Reorders the cells in the Profiler in accordance with the importance indices for the main effects (Main Effect).

**Reorder factors by total importance**    Reorders the cells in the Profiler in accordance with the total importance indices for the factors (Total Effect).

**Colorize Profiler**    Colors cells in the profiler by Total Effect importance indices using a red to white intensity scale.

**Note:** You can click rows in the Summary Report to select columns in the data table. This can facilitate further analyses.

# Bagging

Bootstrap aggregating (*bagging*) is a technique to improve predictive performance while also gaining insight into the reliability of predictions. Bagging is especially useful in unstable methods, including neural networks, classification trees, and regression trees.

Bagging creates $M$ training data sets by sampling with replacement from the original data. All training data sets are of the same size as the original. For each training data set, a model is fit using the analysis platform, and predictions are made. Therefore, there are a total of $M$ predictions for each observation in the original data set. The final prediction is the average of the $M$ predictions.

Bagging is available in many analysis platforms. To use bagging, select **Save Bagged Predictions** from the Prediction Profiler red triangle menu. A window appears with the following options for Bagging:

**Number of Bootstrap Samples**    Sets the number of times that you want to resample the data and build a model. A larger number results in more precise predictions. By default, the number of bootstrap samples is 100.

**Random Seed**    Sets a random seed that you can re-enter in subsequent runs of the bagging analysis to duplicate your current results. By default, no seed is set.

**Fractional Weights**    Performs a Bayesian bagging analysis. In each bootstrap iteration, each observation is assigned a nonzero weight. The model that makes the predictions uses the

weighted observations. By default, the Fractional Weights option is not selected, and a simple bagging analysis is conducted.

**Tip:** Use the Fractional Weights option if the number of observations that are used in your analysis is small or if you are concerned about separation in a logistic regression setting.

Suppose that Fractional Weights is selected. For each bootstrap iteration, each observation that is used in the report is assigned a nonzero weight. These weights sum to *n*, the number of observations used in the model. For more information about how the weights are calculated and used, see "Calculation of Fractional Weights" on page 323 in the "Bootstrapping" chapter.

**Save Prediction Formulas**   For each bagged prediction, this option saves the formula used to make that prediction in the column properties. This option is available in only a subset of the analysis platforms that offer bagging.

**Note:** If Save Prediction Formulas is not available, a note appears stating that only the predicted values will be saved.

Bagging automatically creates new columns in the original data table. All *M* sets of bagged predictions are saved as hidden columns. The final prediction is saved in a column named "Pred Formula <colname> Bagged Mean". The standard deviation of the final prediction is saved in a column named "<colname> Bagged Std Dev". The standard error of the bagged mean is saved in a column named "StdError <colname> Bagged Mean." The standard error is the standard deviation divided by $\sqrt{M-1}$. Here, <colname> identifies the column in the report that was bagged.

The standard error gives insight about the precision of the prediction. A very small standard error indicates a precise prediction for that observation. For more information about bagging, see Hastie et al. (2009).

**Figure 3.16**  Bagging Columns
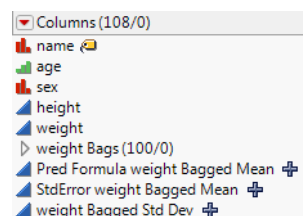


## Additional Examples of the Prediction Profiler

This section contains additional examples illustrating various aspects of the Prediction Profiler platform and the Prediction Profiler embedded in many analysis platforms.

## Example of Desirability Profiling for Multiple Responses

A desirability index becomes especially useful when there are multiple responses. The idea was pioneered by Derringer and Suich (1980), who give the following example. Suppose there are four responses, ABRASION, MODULUS, ELONG, and HARDNESS. Three factors, SILICA, SILANE, and SULFUR, were used in a central composite design.

The data are in the Tiretread.jmp table in the sample data folder. Use the RSM For 4 responses script in the data table, which defines a model for the four responses with a full quadratic response surface. The summary tables and effect information appear for all the responses, followed by the prediction profiler shown in Figure 3.17. The desirability functions are as follows:

1. Maximum ABRASION and maximum MODULUS are most desirable.
2. ELONG target of 500 is most desirable.
3. HARDNESS target of 67.5 is most desirable.

**Figure 3.17**  Profiler for Multiple Responses before Optimization
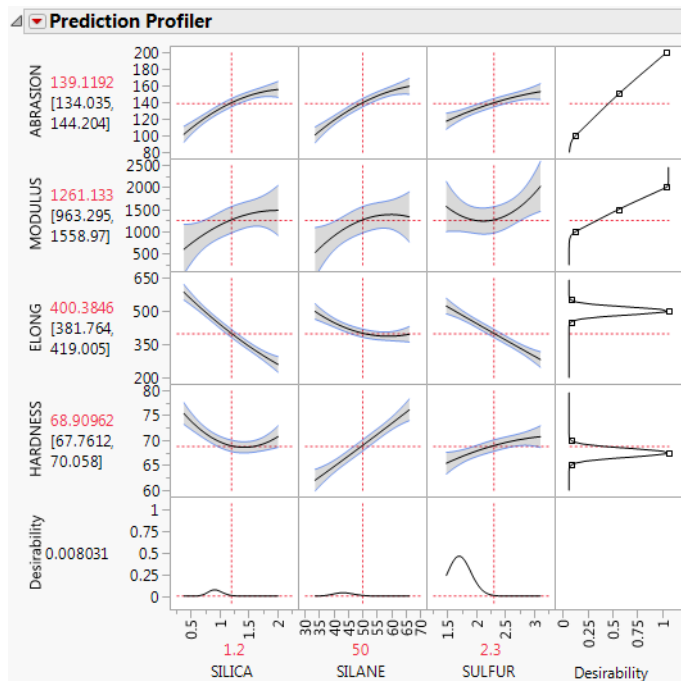


Select **Optimization and Desirability > Maximize Desirability** from the Prediction Profiler red triangle menu to maximize desirability. The results are shown in Figure 3.18. The desirability traces at the bottom decrease everywhere except the current values of the effects, which indicates that any further adjustment could decrease the overall desirability.

**Figure 3.18** Profiler after Optimization
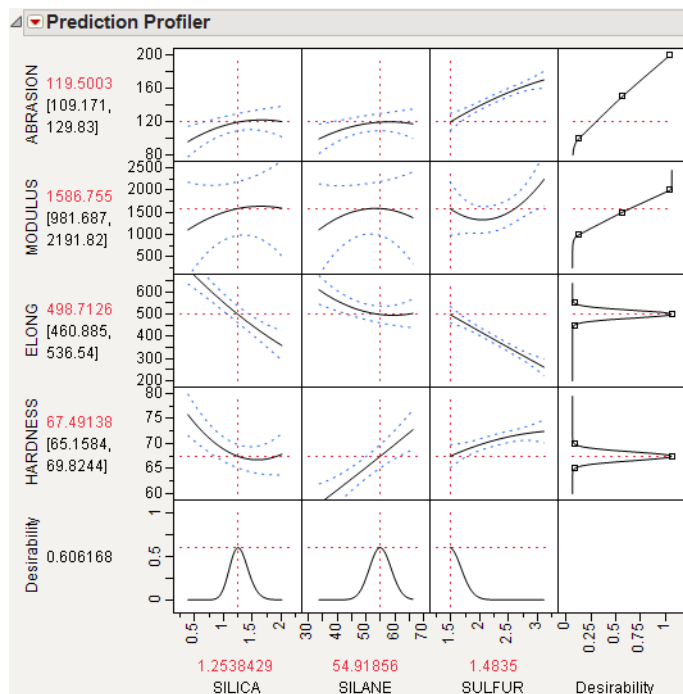


## Example of a Noise Factor in the Prediction Profiler

As an example, use the Tiretread.jmp sample data set. This data set shows the results of a tire manufacturer's experiment whose objective is to match a target value of HARDNESS= 70 based on three factors: SILICA, SILANE, and SULFUR content. Suppose the SILANE and SULFUR content are easily (and precisely) controllable, but SILICA expresses variability that is worth considering.

For comparison, first optimize the factors for hardness without considering variation from the noise factor.

1. Select **Graph > Profiler** to launch the Profiler.

2. Assign Pred Formula HARDNESS to the **Y, Prediction Formula** role.

3. Click **OK**.

4. Select **Optimization and Desirability > Desirability Functions** in the Prediction Profiler menu.

5. Double-click in the Desirability plot to open the Response Goal window. Select **Match Target** from the list.

6. Select **Optimization and Desirability > Maximize Desirability** to find the optimum factor settings for our target value of HARDNESS.

We get the following Profiler display. Notice that the SILICA factor's optimum value is on a sloped part of a profile curve. This means that variations in SILICA are transmitted to become variations in the response, HARDNESS.

**Note:** You might get different results from these because different combinations of factor values can all hit the target.

**Figure 3.19** Maximizing Desirability for HARDNESS
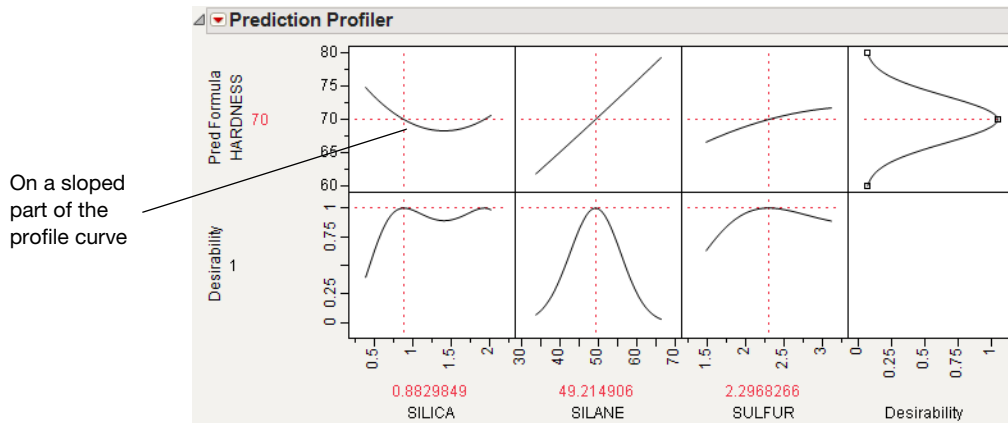


Now, we would like to not just optimize for a specific target value of HARDNESS, but also be on a flat part of the curve with respect to Silica. So, repeat the process and add SILICA as a noise factor.

1. Select **Graph > Profiler**.
2. Select Pred Formula HARDNESS and click **Y, Prediction Formula**.
3. Select SILICA and click **Noise Factors**.
4. Click **OK**.
5. Change the Pred Formula Hardness desirability function as before.

The resulting profiler has the appropriate derivative of the fitted model with respect to the noise factor, set to be maximized at zero, its flattest point.

**Figure 3.20** Derivative of the Prediction Formula with Respect to Silica
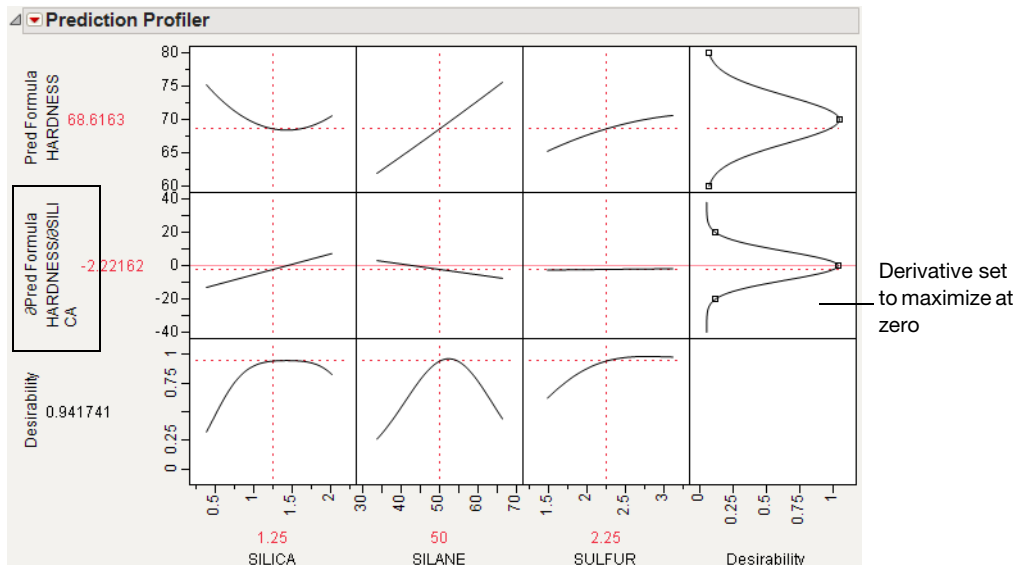


6. Select **Optimization and Desirability > Maximize Desirability** to find the optimum values for the process factor, balancing for the noise factor.

This time, we have also hit the targeted value of HARDNESS, but our value of SILICA is on its flatter region. This means variation in SILICA does not transmit as much variation to HARDNESS.

**Figure 3.21** Maximize Desirability
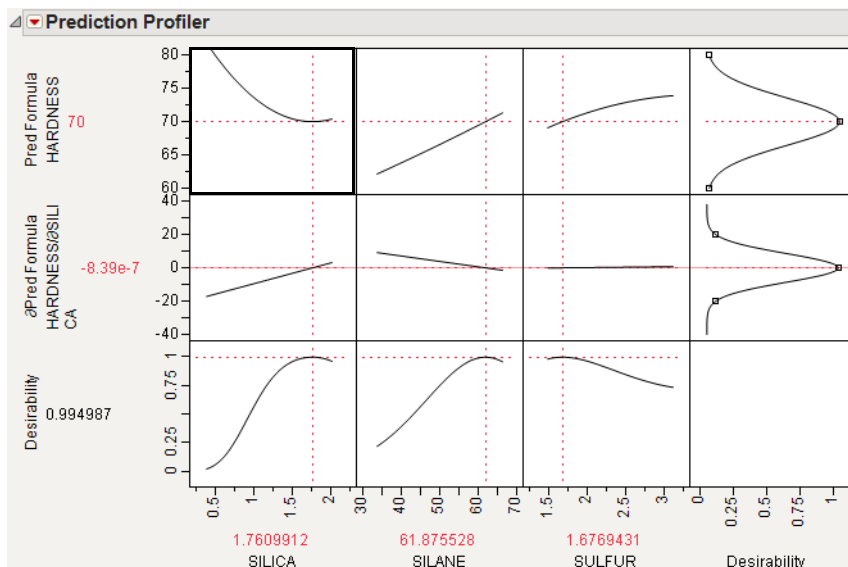
You can see the effect this has on the variance of the predictions by following these steps for each profiler (one without the noise factor, and one with the noise factor):

1.  Select **Simulator** from the platform menu.

2.  Assign SILICA to have a random Normal distribution with a standard deviation of 0.05.

**Figure 3.22** Setting a Random Normal Distribution



Random Normal with SD = 0.05
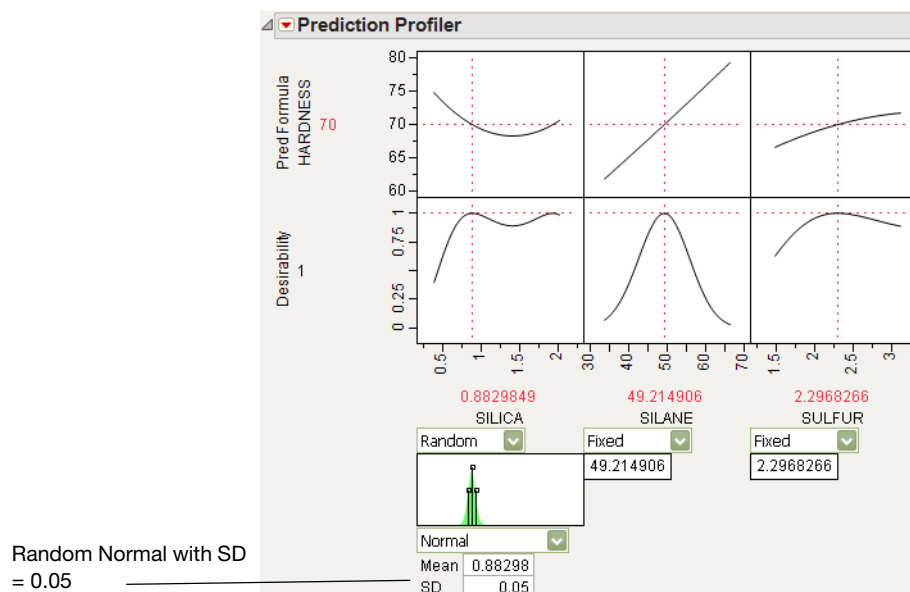
3.  Click **Simulate**.

4.  Click the **Make Table** button under the Simulate to Table node.

    Doing these steps for both the original and noise-factor-optimal simulations results in two similar data tables, each holding a simulation. In order to make two comparable histograms of the predictions, we need the two prediction columns in a single data table.

5.  Copy the Pred Formula HARDNESS column from one of the simulation tables into the other table. They must have different names, like Without Noise Factor and With Noise Factor.

6.  Select **Analyze > Distribution** and assign both prediction columns as **Y**.

7.  When the histograms appear, select **Uniform Scaling** from the Distribution main title bar.

**Figure 3.23** Comparison of Distributions with and without Noise Factors
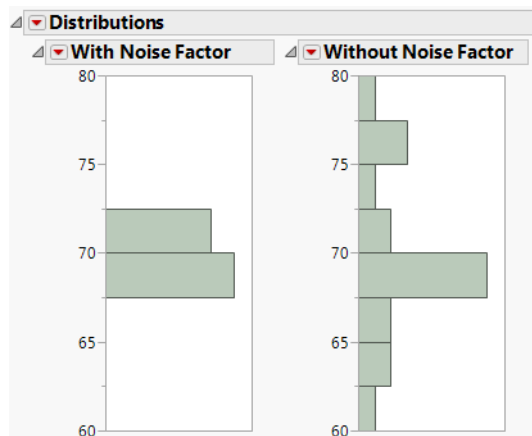


The histograms show that there is much more variation in Hardness when the noise factor was not included in the analysis.

It is also interesting to note the shape of the histogram when the noise factor was included. In the comparison histograms above, note that the With Noise Factor distribution has data trailing off in only one direction. The predictions are skewed because Hardness is at a minimum with respect to SILICA, as shown in Figure 3.24. Therefore, variation in SILICA can make only HARDNESS increase. When the non-robust solution is used, the variation could be transmitted either way.

**Figure 3.24** Profiler Showing the Minima of HARDNESS by SILICA
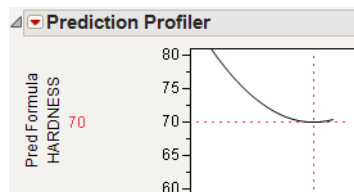


## Example of Variable Importance for One Response

The Boston Housing.jmp sample data table contains data on 13 factors that might relate to median home values. You fit a model using a neural network. Because neural networks do not accommodate formal hypothesis tests, these tests are not available to help assess which variables are important in predicting the response. However, for this purpose, you can use the Assess Variable Importance profiler option.

Note that your results might differ from, but should resemble, those shown here. There are two sources of random variability in this example. When you fit the neural network, *k*-fold

cross validation is used. This partitions the data into training and validation sets at random. Also, Monte Carlo sampling is used to calculate the factor importance indices.

1.   Select **Help > Sample Data Library** and open Boston Housing.jmp.

2.   Select **Analyze > Predictive Modeling > Neural**.

3.   Select mvalue from the Select Columns list and click **Y, Response**.

4.   Select all other columns from the Select Columns list and click **X, Factor**.

5.   Click **OK**.

6.   In the Neural Model Launch panel, select **KFold** from the list under Validation Method.

     When you select KFold, the Number of Folds defaults to 5.

7.   Click **Go**.

8.   From the red triangle menu for the Model NTanH(3) report, select **Profiler**.

     The Prediction Profiler is displayed at the very bottom of the report. Note the order of the factors for later comparison.

     Because the factors are correlated, you take this into account by choosing Dependent Resampled Inputs as the sampling method for assessing variable importance.

9.   From the red triangle menu next to Prediction Profiler, select **Assess Variable Importance > Dependent Resampled Inputs**.

     The Variable Importance: Dependent Resampled Inputs report appears (Figure 3.25). Check that the Prediction Profiler cells have been reordered by the magnitude of the Total Effect indices in the report. In Figure 3.25, check that the Total Effect importance indices identify rooms and lstat as the factors that have most impact on the predicted response.

**Figure 3.25**  Dependent Resampled Inputs Report
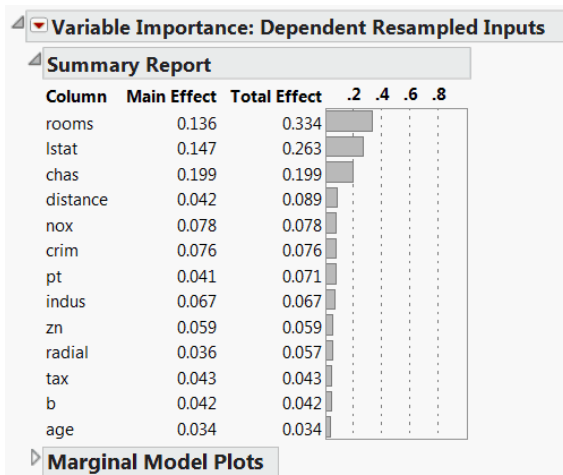
You might be interested in comparing the importance indices obtained assuming that the factors are correlated, with those obtained when the factors are assumed independent.

10. From the red triangle menu next to Prediction Profiler, select **Assess Variable Importance > Independent Resampled Inputs**.

The resampled inputs option makes sense in this example, because the distributions involved are not uniform. The Variable Importance: Independent Resampled Inputs report is shown in Figure 3.26. Check that the two factors identified as having the most impact on the predicted values are lstat and rooms. Note that the ordering of their importance indices is reversed from the ordering using Dependent Resampled Inputs.

**Figure 3.26**  Independent Resampled Inputs Report
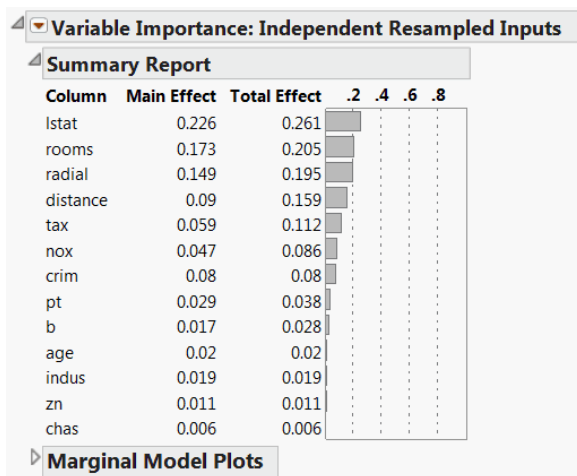


### Example of Variable Importance for Multiple Responses

The data in the Tiretread.jmp sample data table are the result of a designed experiment where the factors are orthogonal. For this reason, you use importance estimates based on independent inputs. Suppose that you believe that, in practice, factor values vary throughout the design space, rather than assume only the settings defined in the experiment. Then you should choose Independent Uniform Inputs as the sampling scheme for your importance indices.

1. Select **Help > Sample Data Library** and open Tiretread.jmp.

2. Run the script RSM for 4 Responses.

   The Prediction Profiler is displayed at the very bottom of the report.

3. From the red triangle menu next to Prediction Profiler, select **Assess Variable Importance > Independent Uniform Inputs**.

The Summary Report is shown in Figure 3.27. Because the importance indices are based on random sampling, your estimates might differ slightly from those shown in the figure.

The report shows tables for each of the four responses. The Overall table averages the factor importance indices across responses. The factors in the Profiler (Figure 3.28) have been reordered to match their ordering on the Overall table's Total Effect importance.

**Figure 3.27** Summary Report for Four Responses
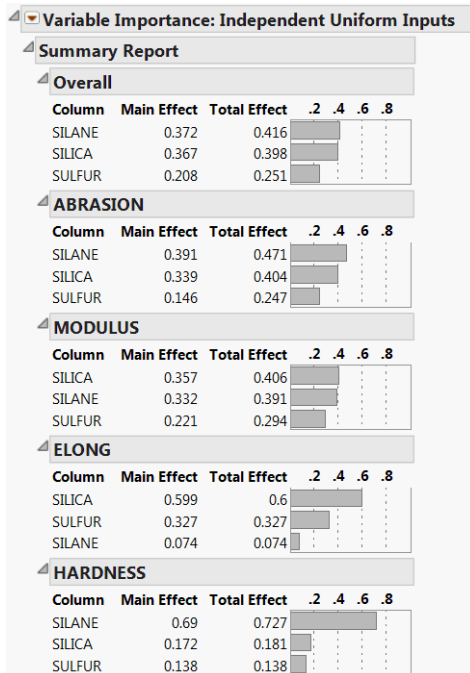


4. From the red triangle menu next to Variable Importance: Independent Uniform Inputs, select **Colorize Profiler**.

Colors from a red to white intensity scale are overlaid on profiler panels to reflect Total Effect importance. For example, you easily see that the most important effect is that of Silane on Hardness.

**Figure 3.28**  Profiler for Four Responses
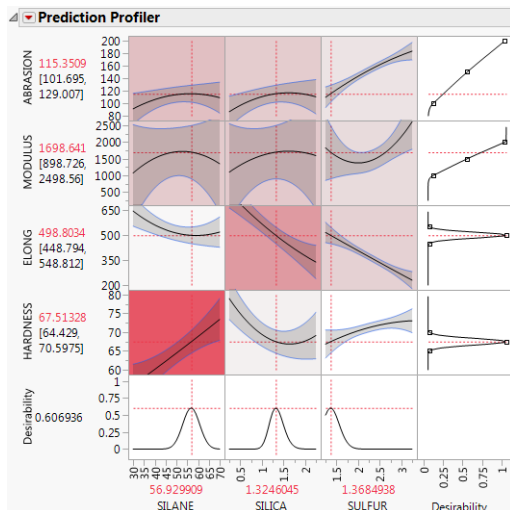


The Marginal Model Plots report (Figure 3.29) shows mean responses for each factor across a uniform distribution of settings for the other two factors.

**Figure 3.29**  Marginal Model Plots for Four Responses
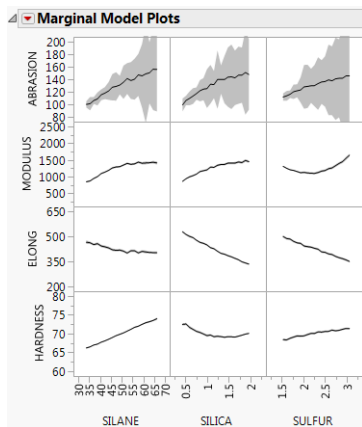


## Example of Bagging to Improve Prediction

Bagging is used in a number of situations, one of which is improving predictive power. Bagging is especially helpful for unstable models. This example uses the Tiretread.jmp sample data table. There are three factors (SILICA, SILANE, and SULFUR) and four responses (ABRASION, MODULUS, ELONG, and HARDNESS). First, you fit a neural network model to simultaneously predict the four response variables as a function of the three factors. Then, you

perform bagging on the neural network model. Last, you compare the predictions to show the improvements obtained through bagging.

**Fit Neural Network Model**

1.  Select **Help > Sample Data Library** and open Tiretread.jmp.

2.  Select **Analyze > Predictive Modeling > Neural**.

3.  Select ABRASION, MODULUS, ELONG, and HARDNESS and click **Y, Response**.

4.  Select SILICA, SILANE, and SULFUR and click **X, Factor**.

5.  Click **OK**.

6.  (Optional) Enter 2121 next to Random Seed.

    **Note:** Results vary due to the random nature of choosing a validation set in the Neural Network model. Entering the seed above enables you to reproduce the results shown in this example.

7.  Click **Go**.

8.  Select **Save Formulas** from the red triangle menu next to Model NTanH(3).

    **Note:** This option saves the predicted values for all response variables from the neural network model to the data table. Later, these values are compared to the predictions that are obtained from bagging.

**Perform Bagging**

Now that the initial model has been constructed, you can perform bagging using that model. Access the Bagging feature through the Profiler.

1.  From the red triangle menu next to Model NTanH(3), select **Profiler**.

    The Prediction Profiler appears at the bottom of the report.

2.  From the red triangle menu next to Prediction Profiler, select **Save Bagged Predictions**.

3.  Enter 100 next to Number of Bootstrap Samples.

4.  (Optional) Enter 2121 next to Random Seed.

    **Note:** Results vary due to the random nature of sampling with replacement. To reproduce the exact results in this example, set the Random Seed.

5.  Click **OK**.

    Return to the data table. For each response variable, there are three new columns denoted as Pred Formula <colname> Bagged Mean, StdError <colname> Bagged Mean, <colname>

Bagged Std Dev. The Pred Formula <colname> Bagged Mean columns are the final predictions.

**Figure 3.30**  Columns Added to Data Table After Bagging
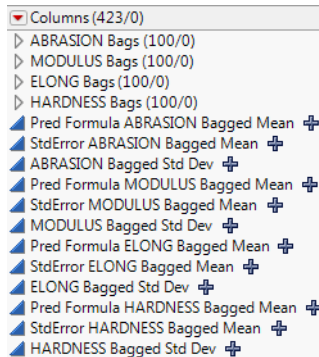


### Compare the Predictions

To see how bagging improves predictive power, compare the predictions from the bagged model to the original model predictions. Use the Model Comparison platform to look at one response variable at a time.

1. Select **Analyze > Predictive Modeling > Model Comparison**.

2. Select Predicted ABRASION and click **Y, Predictors**.

3. Select Pred Formula ABRASION Bagged Mean and click **Y, Predictors.**

4. Click **OK**.

   A window that contains a list of columns appears.

5. Select ABRASION and click **OK**.

6. From the red triangle menu next to Model Comparison, select **Plot Actual by Predicted**.

**Figure 3.31** Comparison of Predictions for ABRASION
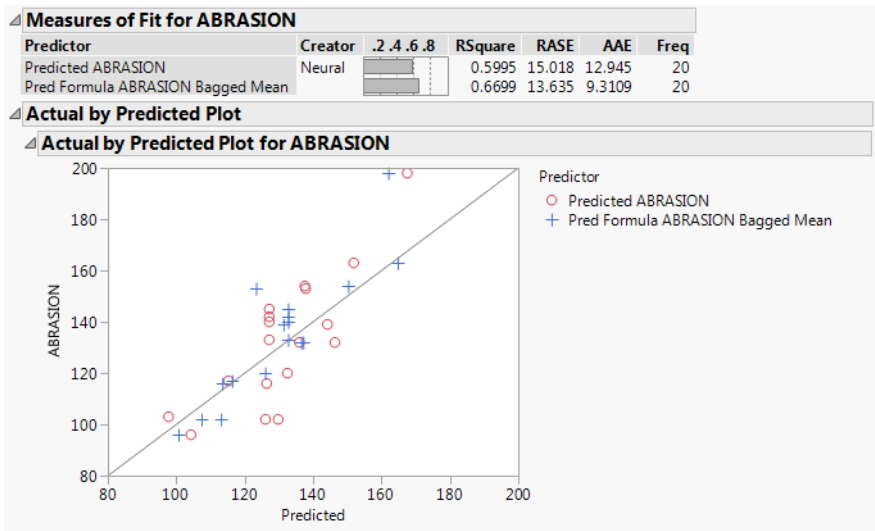


The Measures of Fit report and the Actual by Predicted Plot are shown in Figure 3.31. The predictions that were obtained from bagging are shown in blue. The predictions that were obtained from the original neural network model are shown in red. In general, the bagging predictions are closer to the line than the original model predictions. Because the bagging predictions are closer to the line, the RSquare value of 0.6699 for the bagged predictions is higher than the RSquare value for the original model predictions. You conclude that bagging has improved predictions for ABRASION.

This example compared the predictions for ABRASION. To compare predictions for another response variable, follow step 2 through step 6, replacing ABRASION with the desired response variable. As another example, Figure 3.32 shows the Measures of Fit report for HARDNESS. The report shows similar findings as the Measures of Fit report for ABRASION. The RSquare value for the bagged predictions is slightly higher than the RSquare value for the original model predictions, which indicates a better fit and improved predictions.

**Figure 3.32** Comparison of Predictions for HARDNESS
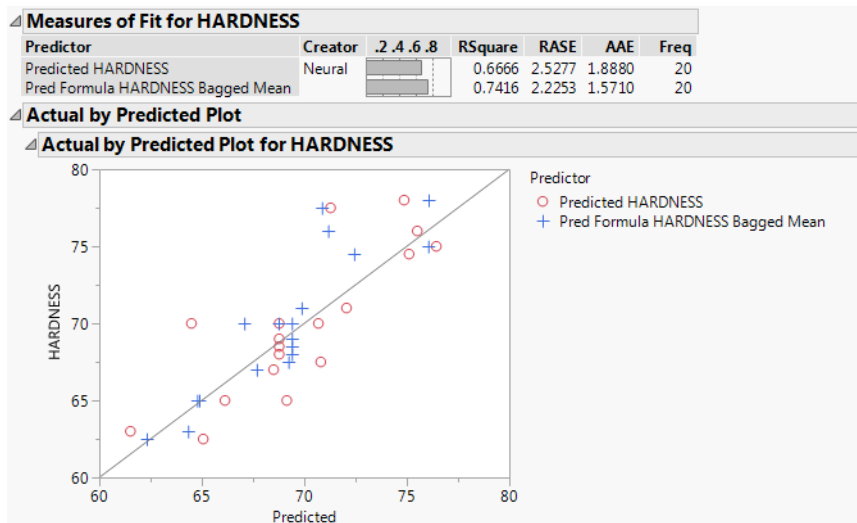


## Example of Bagging to Indicate the Accuracy of Predictions

Bagging is also used to indicate the accuracy of the prediction through standard errors and other distributional measures. In platforms where the Save Predicted Formulas option is available in Bagging, you can make predictions on new observations and determine how accurate they are. The Save Predicted Formulas option is available in the Standard Least Squares, Generalized Regression, and Generalized Linear Models platforms.

In the Tiretread.jmp data table, suppose that you are interested in only predicting ABRASION as a function of the three factor variables. In this example, you fit a generalized regression model to predict ABRASION. Then, you perform bagging on that model. Last, you make a prediction for a new observation and investigate the accuracy of that prediction. This is done by obtaining a confidence interval for the prediction.

### Fit a Generalized Regression Model

1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Select **Analyze > Fit Model**.
3. Select ABRASION and click **Y**.
4. Select **Generalized Regression** from the Personality list.
5. Select SILICA, SILANE, and SULFUR and click **Add**.
6. Click **Run**.
7. Click **Go**.

**Perform Bagging**

1. Select **Profilers > Profiler** from the red triangle menu next to Adaptive Lasso with AICc Validation.

   The Prediction Profiler appears at the bottom of the report.

2. From the red triangle menu next to Prediction Profiler, select **Save Bagged Predictions**.

3. Enter 500 next to Number of Bootstrap Samples.

4. (Optional) Enter 4321 next to Random Seed.

   **Note:** Results vary due to the random nature of sampling with replacement. To reproduce the exact results in this example, set the Random Seed.

5. Confirm that **Save Prediction Formulas** is selected.

6. Click **OK**.

   **Note:** This might take longer to run than the "Example of Bagging to Improve Prediction" on page 67. The larger number of samples gives a better estimate of the prediction distributions.

   Return to the data table. For each response variable, there are three new columns denoted as Pred Formula <colname> Bagged Mean, StdError <colname> Bagged Mean, <colname> Bagged Std Dev. The Pred Formula ABRASION Bagged Mean column is the final prediction.

**Prediction for a New Observation**

You now have predictions for ABRASION for each observation in the data table, as well as the standard errors for those predictions. Suppose that you have an observation with new values of 0.9, 43, and 2 for SILICA, SILANE, and SULFUR, respectively. You can predict the ABRASION response and obtain a confidence interval for that prediction because the Save Prediction Formulas option saves the regression equation for each bagged model. Therefore, $M$ predictions are made with the new factor values to create a distribution of possible predictions. The mean is the final prediction, but analyzing the distribution tells you how accurate the prediction is.

1. In the data table, select **Rows > Add Rows**.

2. Enter 1 in the **How many rows to add** box and click **OK**.

3. Under the SILICA column, type 0.9 in the box for the new row.

4. Under the SILANE column, type 43 in the box for the new row.

5. Under the SULFUR column, type 2 in the box for the new row.

   The prediction columns for the new row are automatically calculated.

**Figure 3.33** Values for New Row

| | | ABRASION | MODULUS | ELONG | HARDNESS | SILICA | SILANE | SULFUR |
|---|---|---|---|---|---|---|---|---|
| • | 1 | 102 | 900 | 470 | 67.5 | 0.7 | 40 | 2.8 |
| • | 2 | 120 | 860 | 410 | 65 | 1.7 | 40 | 1.8 |
| • | 3 | 117 | 800 | 570 | 77.5 | 0.7 | 60 | 1.8 |
| • | 4 | 198 | 2294 | 240 | 74.5 | 1.7 | 60 | 2.8 |
| • | 5 | 103 | 490 | 640 | 62.5 | 0.7 | 40 | 1.8 |
| • | 6 | 132 | 1289 | 270 | 67 | 1.7 | 40 | 2.8 |
| • | 7 | 132 | 1270 | 410 | 78 | 0.7 | 60 | 2.8 |
| • | 8 | 139 | 1090 | 380 | 70 | 1.7 | 60 | 1.8 |
| • | 9 | 102 | 770 | 590 | 76 | 0.3835 | 50 | 2.3 |
| • | 10 | 154 | 1690 | 260 | 70 | 2.0165 | 50 | 2.3 |
| • | 11 | 96 | 700 | 520 | 63 | 1.2 | 33.67 | 2.3 |
| • | 12 | 163 | 1540 | 380 | 75 | 1.2 | 66.33 | 2.3 |
| • | 13 | 116 | 2184 | 520 | 65 | 1.2 | 50 | 1.4835 |
| • | 14 | 153 | 1784 | 290 | 71 | 1.2 | 50 | 3.1165 |
| • | 15 | 133 | 1300 | 380 | 70 | 1.2 | 50 | 2.3 |
| • | 16 | 133 | 1300 | 380 | 68.5 | 1.2 | 50 | 2.3 |
| • | 17 | 140 | 1145 | 430 | 68 | 1.2 | 50 | 2.3 |
| • | 18 | 142 | 1090 | 430 | 68 | 1.2 | 50 | 2.3 |
| • | 19 | 145 | 1260 | 390 | 69 | 1.2 | 50 | 2.3 |
| • | 20 | 142 | 1344 | 390 | 70 | 1.2 | 50 | 2.3 |
| | 21 | • | • | • | • | 0.9 | 43 | 2 |

6.  Select **Tables > Transpose**.

7.  Select ABRASION Bags (500/0) and click **Transpose Columns**.

8.  Click **OK**.

9.  Select **Analyze > Distribution**.

10. Select Row 21 and click **Y, Columns**.

> **Note:** Row 21 corresponds to the predictions from the new observation.

11. Click **OK**.

12. From the red triangle menu next to Row 21, select **Display Options > Horizontal Layout**.

**Figure 3.34** Distribution Report
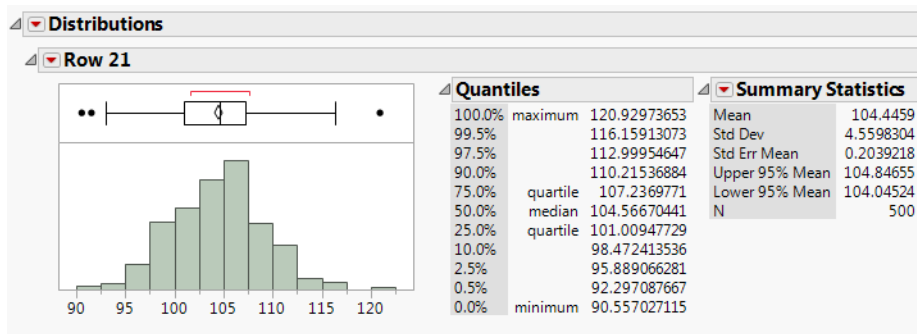
The Distribution Report in Figure 3.34 contains information about the distribution of the predicted values of ABRASION from each bagged model. The final prediction of ABRASION for the new observation is 104.45, which is the mean of all the $M$ bagged predictions. This prediction has a standard error of 4.56. You can also create confidence intervals for the new prediction using the quantiles. For example, a 95% confidence interval for the new prediction is 95.89 to 113.00.

# Statistical Details for the Prediction Profiler

This section contains statistical details for options available in the Prediction Profiler platform.

## Assess Variable Importance

The details that follow relate to the how the variable importance indices are calculated.

### Background

Denote the function that represents the predictive model by $f$, and suppose that $x_1, x_2, ..., x_n$ are the factors, or main effects, in the model. Let $y = f(x_1, x_2 ..., x_n)$.

- The expected value of $y$, $E(y)$, is defined by integrating $y$ with respect to the joint distribution of $x_1, x_2, ..., x_n$.

- The variance of $y$, $Var(y)$, is defined by integrating $(y - E(y))^2$ with respect to the joint distribution of $x_1, x_2, ..., x_n$.

### Main Effect

The impact of the main effect $x_j$ on $y$ can be described by $Var(E(y \mid x_j))$. Here the expectation is taken with respect to the conditional distribution of $x_1, x_2, ..., x_n$ given $x_j$, and the variance is taken over the distribution of $x_j$. In other words, $Var(E(y \mid x_j))$ measures the variation, over the distribution of $x_j$, in the mean of $y$ when $x_j$ is fixed.

It follows that the ratio $Var(E(y \mid x_j))/Var(y)$ gives a measure of the sensitivity of $y$ to the factor $x_j$. The importance index in the Main Effect column in the Summary Report is an estimate of this ratio (see "Adjustment for Sampling Variation" on page 75).

### Total Effect

The Total Effect column represents the total contribution to the variance of $y = f(x_1, x_2 ..., x_n)$ from *all* terms that involve $x_j$. The calculation of Total Effect depends on the concept of functional decomposition. The function $f$ is decomposed into the sum of a constant and functions that represent the effects of single variables, pairs of variables, and so on. These component functions are analogous to main effects, interaction effects, and higher-order effects. (See Saltelli, 2002, and Sobol, 1993.)

Those component functions that include terms containing $x_j$ are identified. For each of these, the variance of the conditional expected value is computed. These variances are summed. The sum represents the total contribution to *Var(y)* due to terms that contain $x_j$. For each $x_j$, this sum is estimated using the selected methodology for generating inputs. The importance indices reported in the Total Effect column are these estimates (see "Adjustment for Sampling Variation" on page 75).

Consider a simple example with two factors, $x_1$ and $x_2$. Then the Total Effect importance index for $x_1$ is an estimate of:

$$\frac{Var(E(y \mid x_1)) + Var(E(y \mid x_1, x_2))}{Var(y)}$$

**Adjustment for Sampling Variation**

Due to the fact that they are obtained using sampling methods, the Main Effect and Total Effect estimates shown in the Summary Table might have been adjusted. Specifically, if the Total Effect estimate is less than the Main Effect estimate, then the Total Effect importance index is set equal to the Main Effect estimate. If the sum of the Main Effect estimates exceeds one, then these estimates are normalized to sum to one.

**Variable Importance Standard Errors**

The standard errors that are provided for independent inputs measure the accuracy of the Monte Carlo replications. Importance indices are computed as follows:

- Latin hypercube sampling is used to generate a set of data values.
- For each set of data values, main and total effect importance estimates are calculated.
- This process is replicated until the estimated standard errors of the Main Effect and Total Effect importance indices for all factors fall below a threshold of 0.01.

The standard errors that are reported are the standard error values in effect when the replications terminate.

# Propagation of Error Bars

Propagation of error (POE) is important when attributing the variation of the response in terms of variation in the factor values when the factor values are not very controllable.

In JMP's implementation, the Profiler first looks at the factor and response variables to see whether there is a Sigma column property (a specification for the standard deviation of the column, accessed through the Cols > Column Info dialog box). If the property exists, then the Prop of Error Bars command becomes accessible in the Prediction Profiler drop-down menu. This displays the $3\sigma$ interval that is implied on the response due to the variation in the factor.

**Figure 3.35** Green Propagation of Errors Bars in the Prediction Profiler
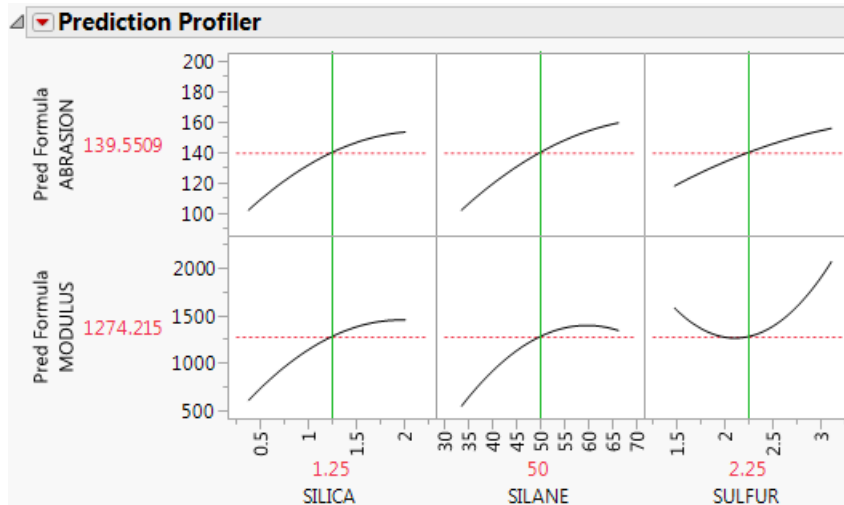


The POE is represented in the graph by a green bracket. The bracket indicates the prediction plus or minus three times the square root of the POE variance. The POE variance can be expressed as:

$$\sum_{i=1}^{N} \left( \sigma_{x_i}^2 \times \left( \frac{\partial f}{\partial x_i} \right)^2 \right) + \sigma_y^2$$

where $\sigma_y$ is the user-specified sigma for the response column, and $\sigma_x$ is the user-specified sigma for the factor column.

Currently, these partial derivatives are calculated by numerical derivatives:

   centered, with δ=xrange/10000

POE limits increase dramatically in response surface models when you are over a more sloped part of the response surface. One of the goals of robust processes is to operate in flat areas of the response surface so that variations in the factors do not amplify in their effect on the response.

# Contour Profiler

## Explore Contours of Responses across Two Factors

The Contour Profiler shows response contours for two factors at a time. The interactive contour profiling facility is useful for optimizing response surfaces graphically.

**Figure 4.1** Contour Profiler Example
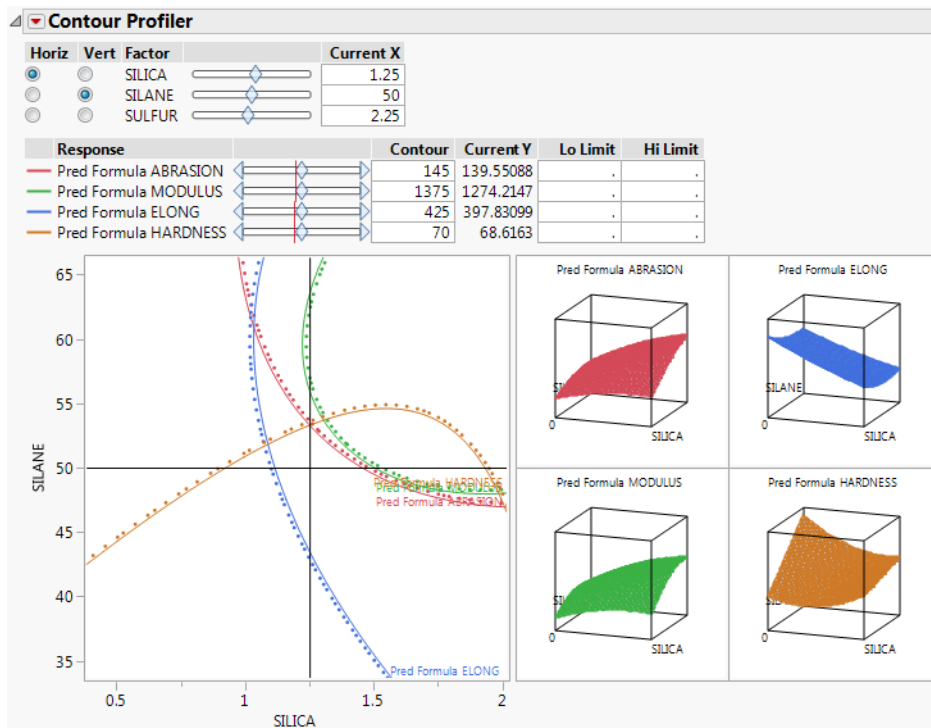
# Contour Profiler Overview

The Contour Profiler shows response contours for two factors at a time. Other factors can be set to specific values to show how this affects the contours. The interactive contour profiling facility is useful for optimizing response surfaces graphically. Separate surface plots for each response are also graphed. Figure 4.2 shows an example of the Contour Profiler for the Tiretread.jmp sample data.

**Figure 4.2**  Contour Profiler
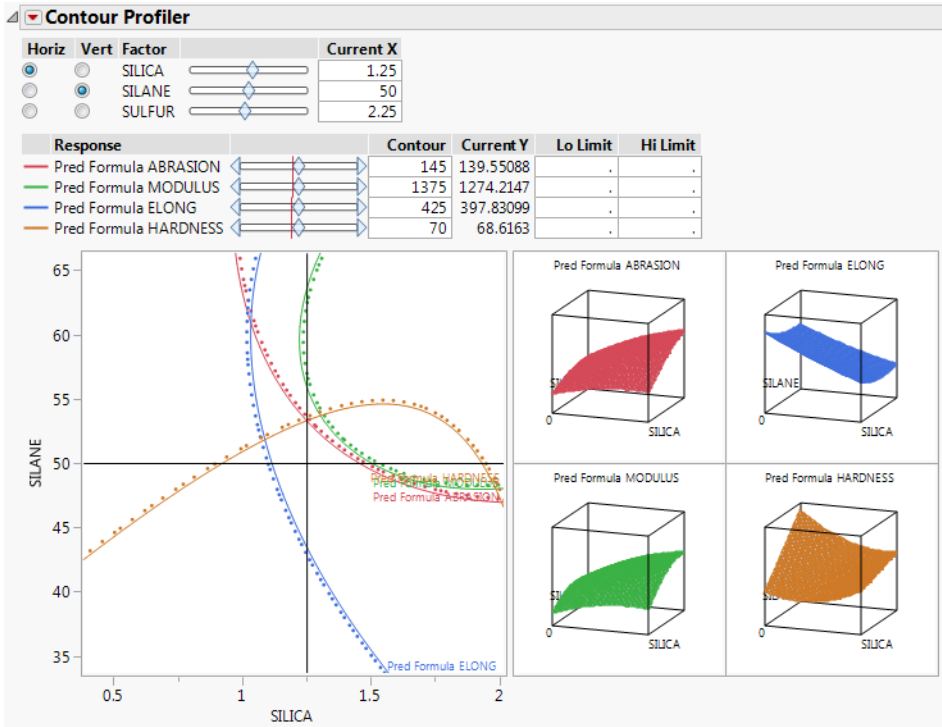


# Example of the Contour Profiler

This example uses the Tiretread.jmp sample data table. There are three factors (SILICA, SILANE, and SULFUR) and four responses (ABRASION, MODULUS, ELONG, and HARDNESS). This data table already contains saved prediction formula columns for the four response variables.

1.  Select **Help > Sample Data Library** and open Tiretread.jmp.

2.  Select **Graph > Contour Profiler**.

3.  Select Pred Formula ABRASION, Pred Formula MODULUS, Pred Formula ELONG, and Pred Formula HARDNESS and click **Y, Prediction Formula**.

4.  Click **OK**.

    The values for SILICA are on the horizontal axis and the values for SILANE are on the vertical axis. The value for SULFUR is fixed at 2.25. See Figure 4.2.

5.  Click the **Current X** box for SULFUR and type 2. Click anywhere outside the box to set the value.

6.  In the **Lo Limit** box for Pred Formula ABRASION, type 100.

7.  In the **Hi Limit** box for Pred Formula ABRASION, type 160.

8.  Click anywhere outside the box to update the plot.

**Figure 4.3**  Contour Profiler for Tiretread.jmp
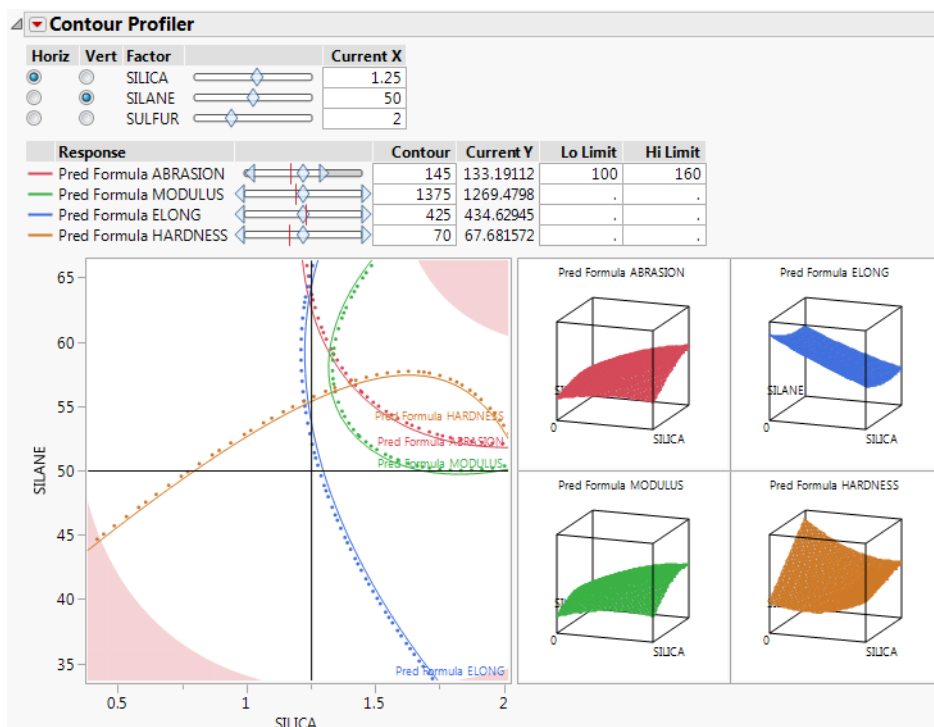


The contour lines and surface plots differ from those in Figure 4.2 because SULFUR is set at a different value. Move the slider bar for SULFUR to see how different values further change the contour lines and surface plots. The shaded area on the contour plot represents the region of the graph that is not feasible, based on the Lo Limit and Hi Limit values for ABRASION,

## Launch the Contour Profiler Platform

The Contour Profiler can be accessed in the following ways:

- The Contour Profiler can be accessed directly from the Graph menu. When you access the Contour Profiler in this way, the Contour Profiler launch window appears. See "Profiler Launch Windows" on page 29 in the "Introduction to Profilers" chapter for details.

- The Contour Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" on page 28 in the "Introduction to Profilers" chapter for details about the availability of the Contour Profiler in different platforms.

- The Contour Profiler can be accessed from the Model Comparisons platform. Select Profiler from the Model Comparisons red triangle menu. Then, select Contour Profiler from the Profiler red triangle menu.

- **JMP PRO**  The Contour Profiler can be accessed from the Formula Depot platform. Select Profiler from the Formula Depot red triangle menu and select the models to be profiled. Then, select Contour Profiler from the Profiler red triangle menu.

## The Contour Profiler Report

The initial Contour Profiler report shows a contour profiler plot, a surface plot for each response, factor settings and controls, and response settings and controls. The colored lines on the graph are the contours for the responses set by the Y slider controls or by entering values in the Contour column. There is a separately colored contour for each response. For each contour, there is a dotted line in the direction of higher response values, so that you get a sense of direction

## Factor Settings and Controls

**Horiz, Vert**   Radio buttons to control factor assignments to the horizontal and vertical axes in the contour profiler and surface plots.

**Factor**   The list of factors.

**Current X**   The current factor settings. Click in a box to change the value of a factor or use the slider controls. Right-click the slider control and select **Rescale Slider** to change the scale of the slider. When one value is changed, the values for the other factors remain the same, but the values in the response settings and controls report change accordingly. The Current X location is shown by the crosshair lines on the graph.

**Lock**   A column that appears for mixture designs with more than three predictors. This column enables you to lock settings for mixture values so that they are not changed when the mixture needs to be adjusted due to other mixture effects being changed. When locked columns exist, the shaded area for a mixture recognizes the newly restricted area.

**Figure 4.4** Lock Column



## Response Settings and Controls

**Response**   The list of one or more responses, with a color legend. To change the color of a response, right-click the color legend for that response.

**Contour**   The current value of the contour on the contour profiler. Click in the box to change the value. The slider control can also be used to change the value of the contour.

**Current Y**   The predicted response based on the current X settings. The value is at the red line on the slider control. This value updates as the factor settings are changed.

**Lo Limit**   Enables you to set a lower limit for your response. Click in a box to set the limit or click the left triangle of the slider control for that response. If a response column's Spec Limits column property has a value for Lower Spec Limit, that value is used as the initial value for Lo Limit.

**High Limit**   Enables you to set an upper limit for your response. Click in a box to set the limit or click the right triangle of the slider control for that response. If a response column's Spec Limits column property has a value for Upper Spec Limit, that value is used as the initial value for Hi Limit.

When response limits are used, the profiler shows regions that are not feasible as shaded in the contour profiler plot.

## Contour Profiler Platform Options

**Grid Density**   Sets the density of the mesh plots in the surface plots.

**Graph Updating**   Contains options to change the frequency of updates for the graph. (The difference might not be noticeable on a fast machine.) The following two options are available:

**Per Mouse Move**   Updates the graph continuously as you drag the mouse. This is the default setting.

**Per Mouse Up**   Updates the graph each time the mouse is released.

**Surface Plot**   Shows or hides the mesh plots.

**Contour Label**   Shows or hides a label for the contour lines. The label colors match the contour colors.

**Contour Grid**   Draws contours on the Contour Profiler plot at intervals that you specify.

**Remove Contour Grid**   Enables you to remove a contour grid once one is drawn.

**Factor Settings**   Provides a submenu of commands that enables you to save and transfer the Contour Profiler's settings to other parts of JMP. See "Factor Settings" on page 45 in the "Profiler" chapter.

**Simulator**   Launches the Simulator. See the "Simulator" chapter on page 127.

**Up Dots**   Shows or hides dotted lines that correspond to each contour. The dotted lines show the direction of increasing response values.

**Set Contours to Current**   Resets the contour lines to be where the current Y values are located. This means that all contour lines cross where the crosshairs are on the contour plot and the controls agree in the Y sliders.

**Arrange X Controls Left**   Rearranges the X and Y controls horizontally with the X controls on the left or vertical with the X controls at the top.

**Hide X Controls**   Shows or hides the X controls (Factor Settings and Controls).

**Hide Y Controls**   Shows or hides the Y controls (Response Settings and Controls).

## Constraint Shading Settings

If you specify limits for the response columns, the areas of the plot that are outside of the feasible region are shaded. The unshaded white area represents the feasible region. See Figure 4.5.

**Figure 4.5** Settings for Contour Shading

Setting appropriate limits on responses

illustrates the feasible region
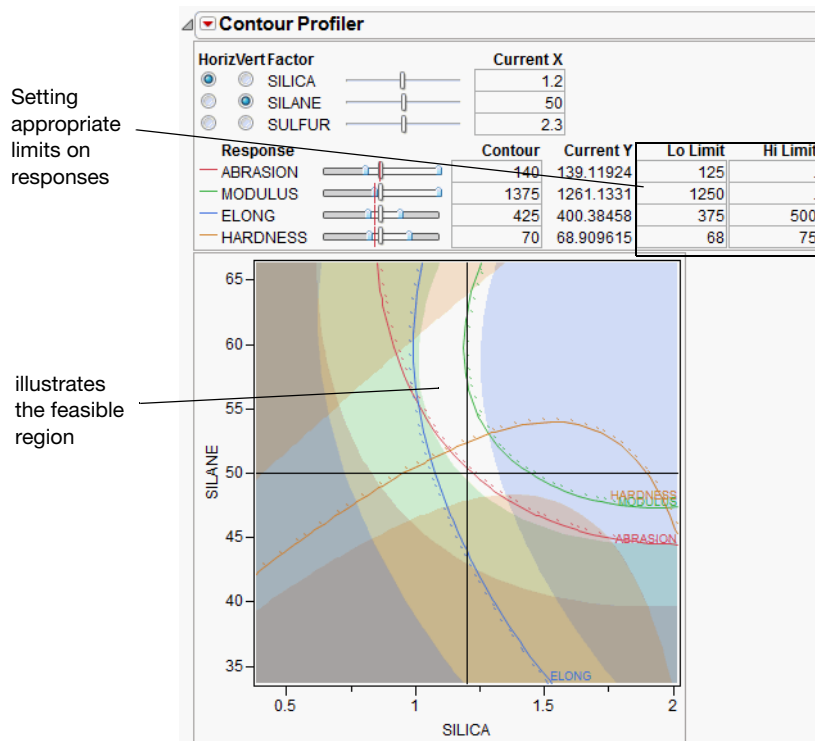


## Additional Examples of the Contour Profiler

### Explore Optimal Settings

Desirability functions are often used to determine the optimal settings for a response surface. The Maximize Desirability feature in the Prediction Profiler provides one combination of factor level settings that results in a predicted response that optimizes the desirability. However, there are often many factor level combinations that can optimize the desirability function. The Contour Profiler is a useful tool for finding alternative settings that optimize desirability.

This example uses the Bounce Data.jmp data table, which contains data for a tire tread experiment. The response, Stretch, is a function of three variables, Silica, Silane, and Sulfur. You will build a model and determine the optimal settings such that the value of Stretch maximizes the desirability function using the Maximize Desirability option in the Prediction Profiler. You will then use the Contour Profiler to explore other optimal settings for the specified value of Stretch.

1.  Select **Help > Sample Data Library** and open Design Experiment/Bounce Data.jmp.

2.  Click the green triangle next to the **Model** script, and then click **Run** in the launch window.

3.  Click the Prediction Profiler red triangle menu and select **Optimization and Desirability > Maximize Desirability**.

**Figure 4.6**  Maximize Desirability Settings
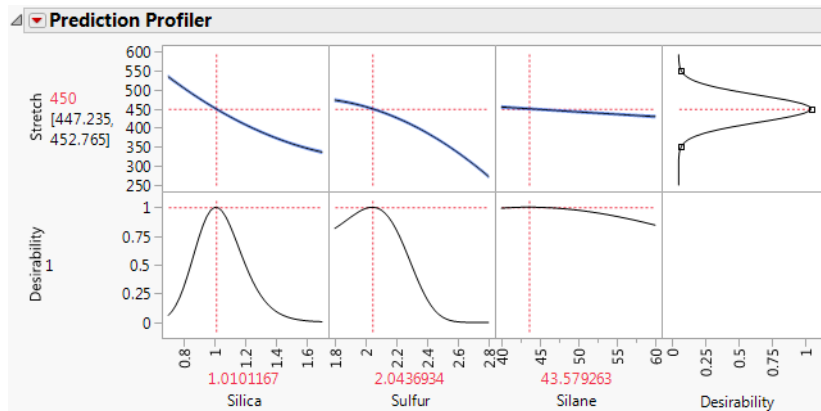


The value of Stretch that maximizes the desirability function is 450. One combination of factor level settings that results in a predicted Stretch of 450 is Silica = 1.0101, Sulfur = 2.0437, and Silane = 43.5792.

4.  Click the red triangle menu next to Response Stretch and select **Factor Profiling > Contour Profiler**.

To ensure that your new setting combinations maximize desirability, you want to make sure that the settings predict Stretch within 2 units of 450. Additionally, you would like a high level of Silane, a fixed value of 60.

**Figure 4.7**  Contour Profiler for Bounce Data.jmp
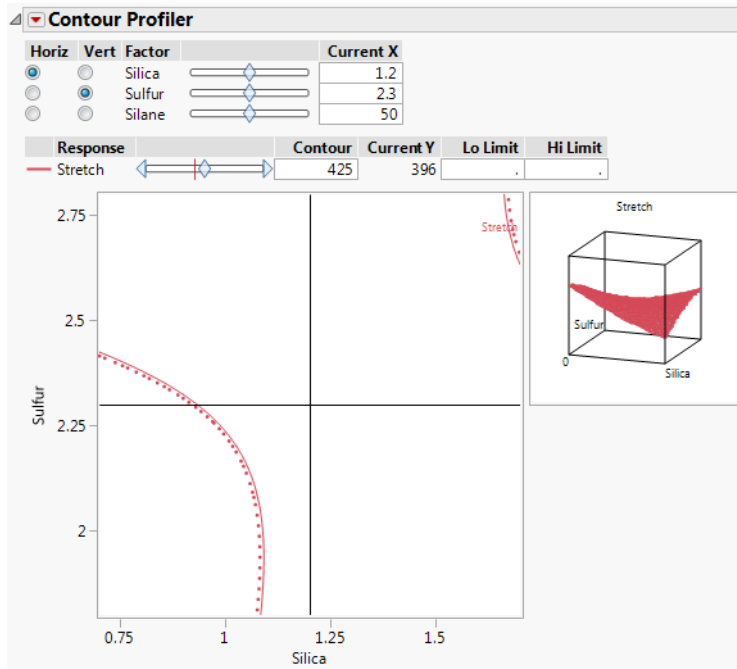


The initial plot shows the contour values of Silica and Sulfur for Stretch at 425 and Silane at 50.

5.  In the response controls, set the **Contour** value for Stretch to 450. Set the **Lo Limit** and **High Limit** for Stretch to 448 and 452, respectively. Press Enter.

6.  Set the **Current X** for Silane to 60.

**Figure 4.8** Contour Profiler for Optimal Settings
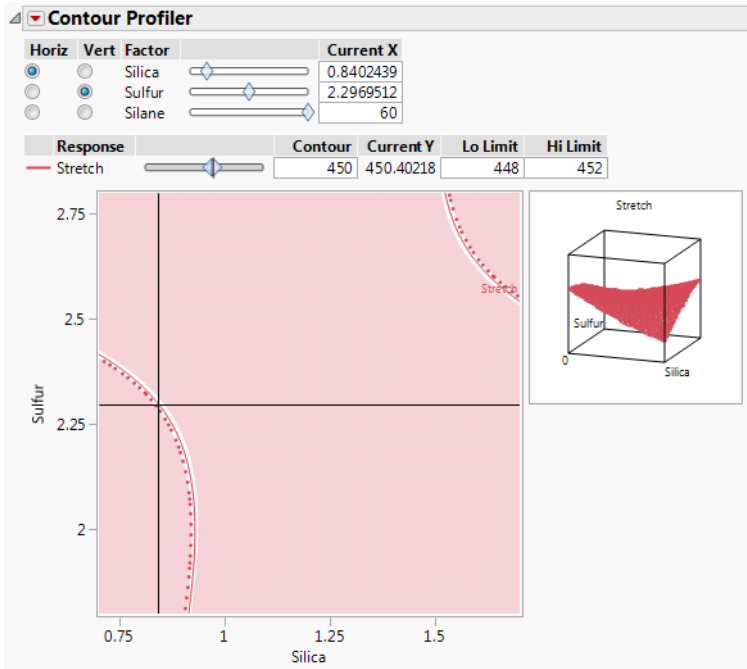


The values on the solid red curve are the Silica and Sulfur values that predict Stretch to be 450. The values within the unshaded band predict values of Stretch between 448 and 452. You can drag the crosshairs that appear on the plot to the unshaded sections to find specific settings for Silica and Sulfur. One such setting is seen in Figure 4.8.

# Surface Plot

## Explore Contours of Responses across Three Factors

A *surface plot* is a three-dimensional plot with one or more dependent variables represented by a smooth surface. You can display up to four dependent variables on the same surface plot.

JMP produces surface plots in the following two situations:

• The Surface Plot platform creates a stand-alone report that contains a surface plot for formulas. The formulas can be formula columns in your data table or mathematical formulas that do not involve any data points.

• The Surface Profiler option in many model fitting platforms produces a surface plot for the fitted model in the existing platform report.

**Figure 5.1** Example of a Surface Plot
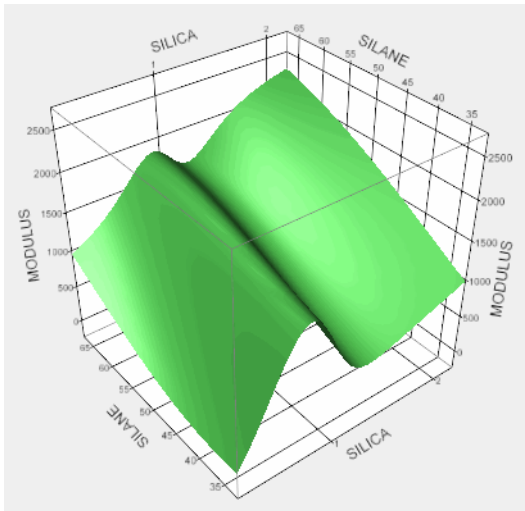
# Overview of the Surface Plot and Surface Profiler Platforms

The Surface Plot and Surface Profiler platforms are used to plot points and surfaces in three dimensions.

Surface plots are available as a separate platform (Graph > Surface Plot) and as options in many reports (known as the Surface Profiler). The surface plot functionality is similar in both situations.

The plots can be of points or surfaces or both. When the points are visible in the surface plot, the points are linked to the data table. You can click on the points or use the Brush tool to select rows in the data table. The colors and markers assigned in the data table are also reflected in the surface plot.

Surfaces can be defined by a mathematical equation, or through a set of points defining a polygonal surface. Each surface can be displayed as a smooth surface or as a mesh, with or without contour lines. The labels, axes, and lighting for the surface plot are fully customizable.

The Surface Plot is built using the 3D scene commands from the JMP Scripting Language (JSL). When you right-click in the surface plot, the pop-up menu contains Open GL Scene commands. For more information about the Open GL Scene commands, see the Three-Dimensional Scenes chapter in the *Scripting Guide*.

# Example of the Surface Plot Platform

This example uses the Tiretread.jmp sample data table. There are three factors (SILICA, SILANE, and SULFUR) and four responses (ABRASION, MODULUS, ELONG, and HARDNESS). This data table already contains saved prediction formula columns for the four response variables. Use these prediction formula columns to examine the surfaces of two of the response variables.

1. Select **Help > Sample Data Library** and open Tiretread.jmp.
2. Select **Graph > Surface Plot**.
3. Select Pred Formula ABRASION and Pred Formula MODULUS, click **Columns**, and then click **OK**.Initial Surface Plot Report
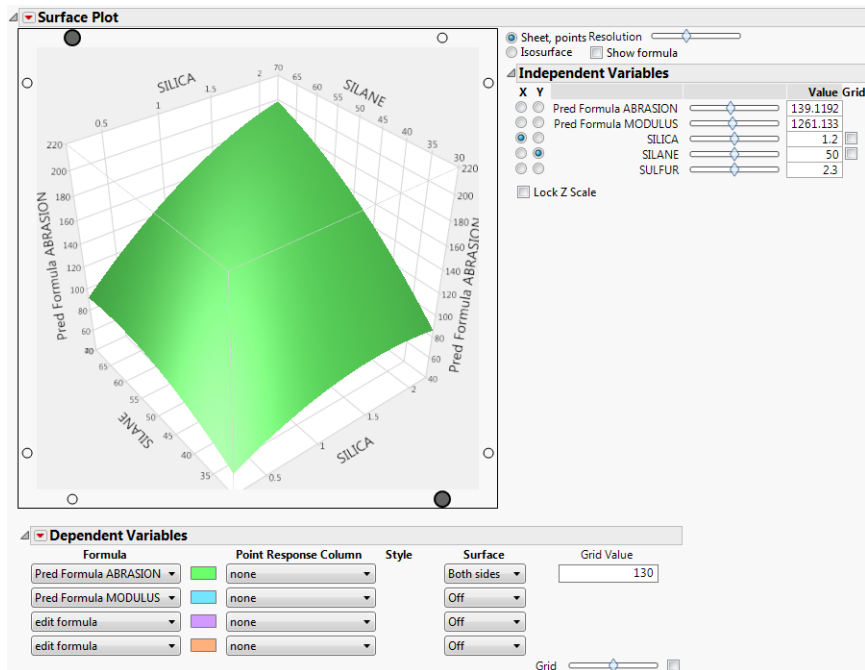
Initially, only the prediction surface for ABRASION appears in the plot. Note that in the Independent Variables report, there is a large difference between the values for ABRASION and MODULUS. The two responses are on different scales.

4. In the Dependent Variables report, select **Both Sides** from the Surface list for Pred Formula MODULUS.

5. Click the Surface Plot red triangle and select **Scale response axes independently**.

**Figure 5.2** Prediction Surface for ABRASION and MODULUS
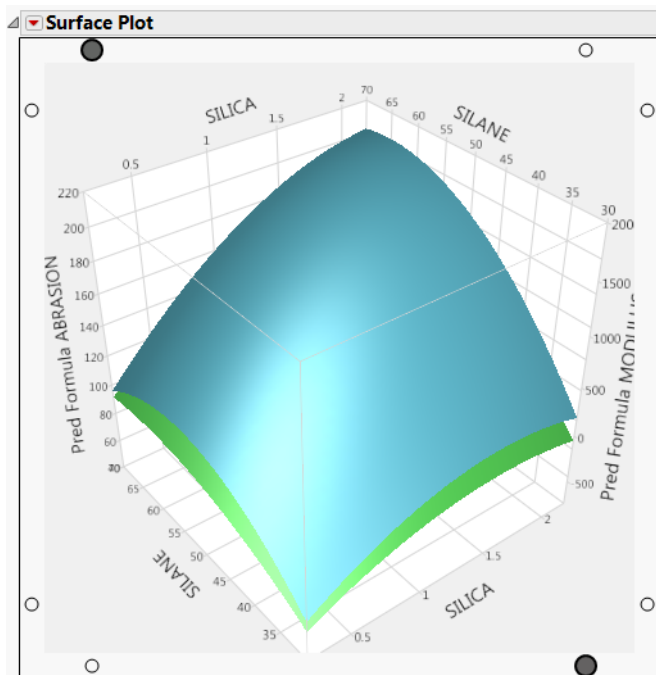


Pred Formula ABRASON is on one $z$-axis and Pred Formula MODULUS is on the other $z$-axis. Each response has its own scale. In this plot, the two factors considered are SILICA and SILANE. You can rotate the surface plot by clicking and dragging inside the graph. This enables you to view the prediction surfaces from different perspectives. For example, in Figure 5.3, you can see that predictions for ABRASION and MODULUS based on the two listed predictors have a similar surface shape, just on a different scale. Use the Independent Variables controls to see how different combinations of predictors affect the surface plot.

**Tip:** To see the original view after rotating, right-click anywhere on the surface plot and select **Reset**.

**Figure 5.3** Rotated Prediction Surfaces
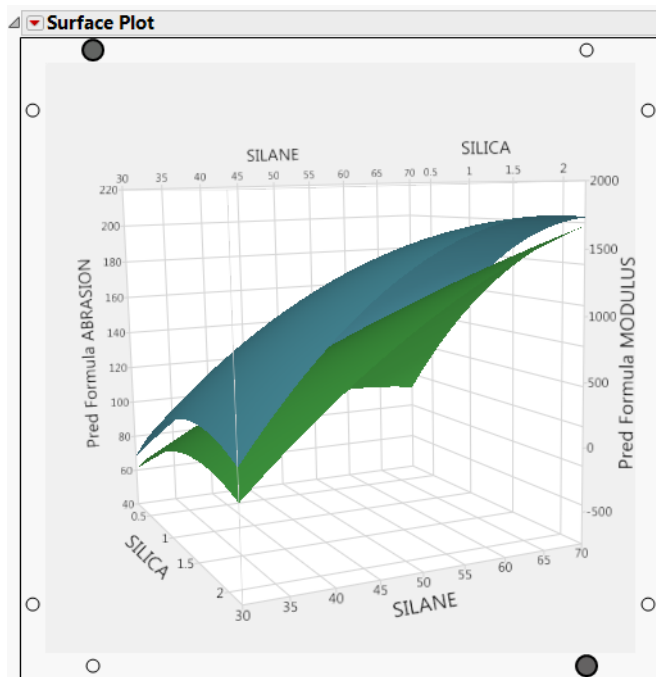


# Launch the Surface Plot Platform

The Surface Plot Profiler can be accessed in the following ways:

- To launch the Surface Plot Profiler platform directly, select Surface Plot from the Graph menu. If there is a data table open, this displays the window in Figure 5.4. If you do not want to use a data table for drawing surface plots, click OK without specifying any columns. If there is no data table open, you are presented with the default surface plot shown in Figure 5.11.

- The Surface Plot Profiler can be accessed as a red triangle menu option in many modeling platforms. In the red triangle menus, it is referred to as Surface Profiler. See "Where to Find JMP Profilers" on page 28 in the "Introduction to Profilers" chapter for details about the availability of the Surface Profiler in different platforms.

- The Surface Plot Profiler can be accessed from the Model Comparisons platform. Select Profiler from the Model Comparisons red triangle menu. Then, select Surface Profiler from the Profiler red triangle menu.

- **JMP PRO** The Surface Plot Profiler can be accessed from the Formula Depot platform. Select Profiler from the Formula Depot red triangle menu and select the models to be profiled. Then, select Surface Profiler from the Profiler red triangle menu.

**Figure 5.4** Surface Plot Launch Window
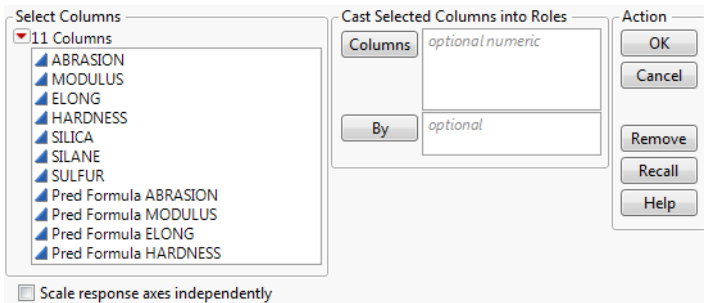


**Columns**   Specifies the columns that you want to plot. Only numeric variables can be
assigned to the Columns role.

**By**   Specifies a variable to create separate surface plots for each level of the variable.

**Scale response axes independently**   Assigns a separate scale to each response on the surface
plot.

Note: When not selected, the axis scale for all responses is the same as the scale for the first
item entered in the Columns role.

## The Surface Plot Report

The initial Surface Plot report shows the surface plot, the appearance controls, the
Independent Variables controls, and the Dependent Variables controls. If more than one
prediction formula column is specified, only the first surface is displayed in the initial report.
However, if more than one observed response column is specified, all points are displayed in
the initial report.

**Figure 5.5** Example of the Surface Plot Report
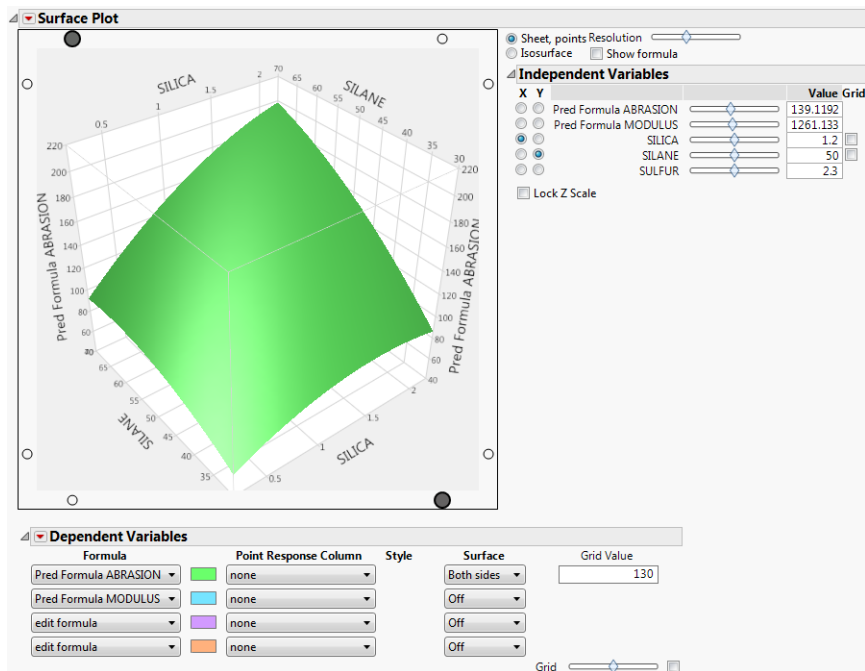


## The Surface Plot

The surface plot has the following controls and settings:

**Rotate**   Click anywhere that the cursor appears with a circular arrow icon ⟳ on the surface plot and drag to rotate the plot in any direction.

> **Note:** The Up, Down, Left, and Right arrow keys can also be used to rotate the plot.

**Axis Settings**   Double-click an axis to reveal the axis settings window. The axis settings window enables you to change the scale, tick mark increment, and axis label settings. Like other JMP graphs, the axes can be adjusted, stretched, and compressed using the grabber tool. Place the cursor over an axis to use the grabber tool.

**Lights**   By default, the plot has lights shining on it. There are eight control knobs on the plot for changing the position and color of the lights. This is useful for highlighting different parts of a plot and creating contrast. Four of the eight knobs are shown in Figure 5.6.

You can perform the following actions to adjust the lights shining on the surface:

– Right-click a knob to turn that light on or off. More lights turned on brighten a plot, and fewer lights darken it.

– Drag a knob to change the position of a light.

– Change the color of a light by right-clicking on the knob. The default color is white.

**Figure 5.6** Control Knobs for Lights
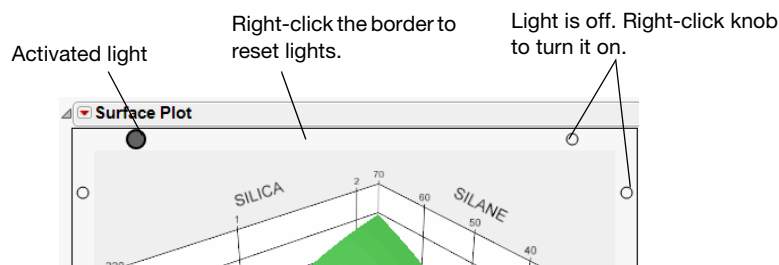


## Appearance Controls

If you select the Surface Profiler option from the red triangle menu of a modeling platform, the Appearance Control options are different than the Appearance Controls in the stand-alone Surface Plot platform. See Figure 5.7 for an example of what the Appearance Controls look like in a modeling platform.

### Surface Plot Appearance Controls

The set of controls in the top right of the report enables you to specify the overall appearance of the surface plot.

**Sheet, points**   The setting for displaying sheets, points, and lines.

**Isosurface**   Changes the display to show isosurfaces. For details, see "Plotting Isosurfaces" on page 103.

**Show formula**   Shows the formula edit box, which enables you to edit the formula of the surface. The formula edit box appears only if you specified a column that contains a formula in the launch window.

**Resolution**   A slider that affects how many points are evaluated for a formula. If the resolution is too coarse, a function with a sharp change might not be represented very well. However, if the resolution is set too high, evaluating and displaying the surface is slower than for lower resolutions.

**Surface Profiler Appearance Controls**

**Figure 5.7** Appearance Controls
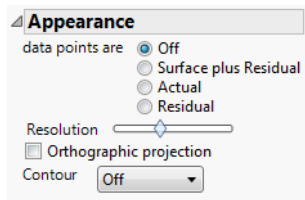


The Appearance Controls in modeling platforms contain the following controls:

**data points are**   Contains the following options that control the appearance of the data points:

**Off**   Turns the data points off.

**Surface plus Residual**   Shows the difference between the predicted value and actual value on the surface.

**Actual**   Shows the actual data points.

**Residual**   Shows the residual values (if they are not off the plot).

**Resolution**   A slider that affects how many points are evaluated for a formula. If the resolution is too coarse, a function with a sharp change might not be represented very well. However, if the resolution is set too high, evaluating and displaying the surface is slower than for lower resolutions.

**Orthogonal Projection**   Changes the view of the graph, such that the axes form right angles.

**Contour**   Determines the placement of the contour lines on the plot in relation to the surface. Options are off, below, above, and on surface.

## Independent Variables

The Independent Variables report contains the following controls:

**X, Y**   Radio buttons that enable you to select which independent variables are displayed on the $x$- and $y$-axes. This feature is useful when you have more than two independent variables.

**Value**   The sliders and text boxes set the current values of each variable. These settings are most important for the variables that are not displayed on the axes. The plot shows the three-dimensional slice of the surface at the value shown in the text box. Move the slider to see different slices.

**Lock Z Scale**   Locks the $z$-axis to its current values. This is useful when moving the sliders that are not on an axis.

**Grid**   Selecting a check box activates a grid that is perpendicular to the selected variables axis. The sliders enable you to adjust the placement of each grid. The resolution of each grid can be controlled by adjusting axis settings. Figure 5.8 shows a surface with the X and Y grids activated.

**Figure 5.8**  Activated X and Y Grids
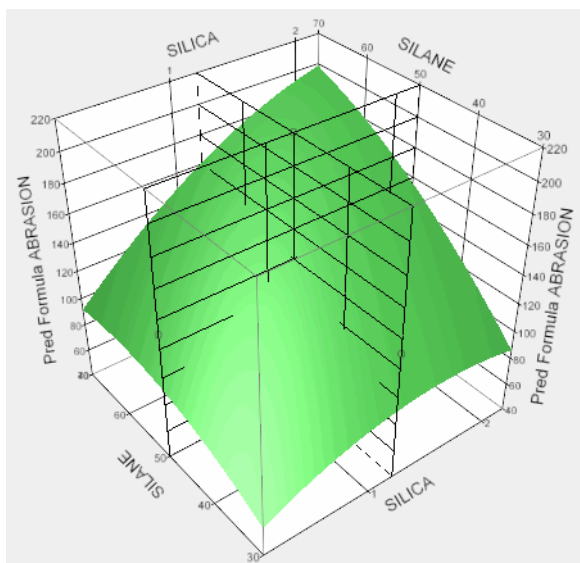


## Dependent Variables

The Dependent Variables controls that are available depend on whether you have selected Sheet, points or Isosurface in the Appearance Controls.

### Controls for Sheet and Points

The Dependent Variables controls that appear by default when Sheet, points is selected are shown in Figure 5.5.

**Formula**   Enables you to select one or more formulas that are displayed as surfaces in the plot.

**Point Response Column**   Enables you to select the column that contains values that are plotted as points.

**Style**   (Menus appear only after you have selected a Point Response Column.) The Style menu enables you to choose how the points are displayed. The menu has the following options:

**Points**   Shows individual points, which change according to the color and marker settings of the row in the data table.

**Needles**   Draws lines from the $x$-$y$ plane to the points, or, if a surface is also plotted, connects the surface to the points.

**Mesh**   Connects the points into a triangular mesh.

**Surface**   Overlays a smooth, reflective surface on the points.

**Off**   Points are not shown at all.

**Surface**   Enables you to show or hide the top or bottom of a surface. If **Above only** or **Below only** is selected, the opposite side of the surface is darkened.

**Grid, Grid Value**   Provides a slider and check box that activate a grid for the dependent variable. You can use the slider to adjust the value at which the grid is drawn. You can also enter the value into the Grid Value box above the slider.

### Controls for Isosurface

Most of the controls for Isosurface are identical to those of Sheet, points. Figure 5.9 shows the default controls, illustrating the slightly different presentation.

**Figure 5.9**  Dependent Variable Controls for Isosurfaces
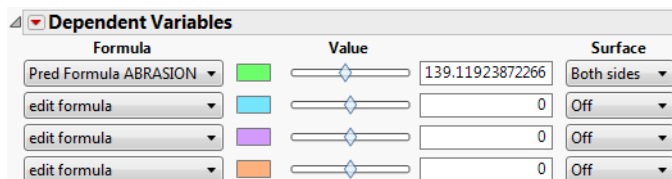


**Value**   Slider and check box that activate an isosurface for the dependent variable. You can use the slider to adjust the value at which the isosurface is drawn. You can also enter the value into the box next to the slider.

## Surface Plot Platform Options

The Surface Plot red triangle menu contains the following options:

**Control Panel**   Shows or hides the Control Panel, which includes the controls for Appearance, Independent Variables, and Dependent Variables.

**Scale response axes independently**   Scales response axes independently. For more information, see "Launch the Surface Plot Platform" on page 91.

**Fit to Window**   Determines whether the plot is resized as you resize the report window.

**Auto (default)**   Bases the scaling on the contents of the plot. For example, a plot with By variables or used as a Surface Profiler in a Fit Model platform does not stretch to fit the resized window; the plot extends beyond the viewing area.

**On**   Always fits the plot inside the window.

**Off**   Prevent the plot from resizing.

See the JMP Reports chapter in the *Using JMP* book for more information about the following options:

**Local Data Filter**   Shows or hides the local data filter that enables you to filter the data used in a specific report.

**Redo**   Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

**Save Script**   Contains options that enable you to save a script that reproduces the report to several destinations.

**Save By-Group Script**   Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

## Pop-Up Menu Options

You can right-click in the plot area to reveal the following pop-up menu options:

**Sheet Properties**   (Option appears only if you right-click on the sheet.) Displays the Sheet Properties window. Available only if Sheet, points is selected in the Appearance Controls. For details, see "Sheet or Surface Properties" on page 99.

**Surface Properties**   (Options appears only if you click on the surface.) Displays the Surface Properties window. Available only if Isosurface is selected in the Appearance Controls. For details, see "Sheet or Surface Properties" on page 99.

**Show Legend**   Shows or hides a legend when the surface is colored using gradients.

**Reset**   Resets the plot to the original viewpoint. Changes in wall and background color are not affected.

**Settings**   Opens a window for changing many plot settings. For more information about the Open GL Scene commands, see the Three-Dimensional Scenes chapter in the *Scripting Guide*.
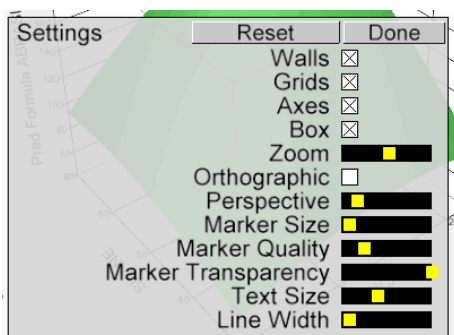
**Hide Lights Border**   Shows or hides lighting controls.

**Wall Color**   Enables you to change the plot wall color.

**Background Color**   Enables you to change the plot background color.

**Rows**   Enables you to change row colors or markers, and also exclude, hide, and label points.

**Use Hardware Acceleration**   Provides for faster rendering of the display. For example, if the plot redraws slowly when rotating, this option can help it redraw faster.

**Show ArcBall**   Provides options for using the ArcBall. The ArcBall is a sphere drawn around the plot to help visualize the directions of rotation.

## Dependent Variables Options

The Dependent Variables red triangle menu provides the following options:

**Formula**   Shows or hides the Formula column in the Dependent Variables controls.

**Surface**   Shows or hides the Surface column in the Dependent Variables controls.

**Points**   Shows or hides the Point Response Column column in the Dependent Variables controls.

**Response Grid**   Shows or hides the Grid controls.

## Sheet or Surface Properties

If you have selected the Sheet, points option in the Appearance Controls, you can right-click on the sheet inside the plot and select **Sheet Properties** to reveal a window that enables you to change the sheet properties. If you are plotting an Isosurface, right-click the surface and select **Surface Properties** to reveal a similar window. The Sheet Properties and Surface Properties windows enable you to modify the appearance of the surface plot, including the surface color, opacity, and contours. You can also show or hide a mesh that appears on the surface plot.

**Figure 5.10** Sheet Properties Window
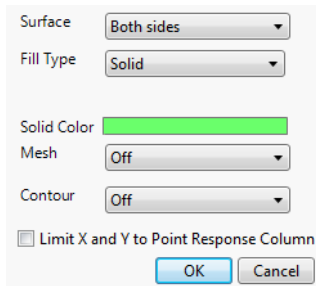


**Surface**   Enables you to show or hide the top or bottom of a surface. If Above only or Below only is selected, the opposite side of the surface is darkened.

**Fill Type**   Enables you to color the surface using a solid color, or continuous or discrete gradients. If a gradient is chosen, the Show Legend option appears when you right-click on the surface.

**Solid Color**   (Available only when Solid is selected for Fill Type.) Enables you to choose a color for the surface.

**Surface Fill**   (Available only when a gradient is selected for Fill Type.) Specifies the dependent variable's surface to which the fill type is applied. When Continuous Gradients is selected for Fill Type, you can also choose the Custom option for Surface Fill.

**Custom**   (Available only when Continuous Gradients is selected for Fill Type and when Custom is selected for Surface Fill.) Enables you to specify an equation that defines a response value. This equation is specified using JSL and can refer to columns in the data table using column scoping.

**Surface Color Theme**   (Available only when a gradient is selected for Fill Type.) Enables you to change the color theme for the surface or define a custom color theme. For details about color themes, see the Enter and Edit Data chapter in the *Using JMP* book.

**Gradient Levels**   (Available only when Discrete Gradients is selected for Fill Type.) Enables you to specify the number of different colors used on the surface plot.

**Surface Color Range**   (Available only when a gradient is selected for Fill Type.) Enables you to choose the endpoints for the color gradient. If you choose Data, the endpoints of the color gradient are determined by the range of the data in the column selected for Surface. If you choose Axis, the endpoints of the color gradient are determined by the range of the axis for the dependent variable.

---

**Caution:** If you choose Axis for Surface Color Range and change the endpoints of the axis in the surface plot, the cutoff values for the color gradient change as well.

---

**Lighting**   Enables you to change the lighting of the surface plot. Choose between None, Low Reflection, and Normal.

**Mesh**   Enables you to turn on or off a surface mesh, for either the *X* or *Y* directions or both.

**Mesh Color**   (Available only when a value other than Off is selected for Mesh.) Enables you to select the color for the surface mesh.

**Contour**   Enables you to turn on or off a contour grid, either above, below, or on the surface. If turned on, the Contour Color option is revealed enabling you to change the color.

**Contour Color**   (Available only when a value other than Off is selected for Contour.) Enables you to select the color for the contour grid.

**Limit X and Y to Point Response Column**   Enables you to limit the range of the plot to the range of the data in the Point Response Column, if one is activated. If checked, this option essentially restricts the plot from extrapolating outside the range of the data in the Point Response Column.

---

**Note:** The equivalent JSL command for this option is `Clip Sheet( Boolean )`. You can send this message to a particular response column by appending the number of the response column. For example, `Clip Sheet2( 1 )` limits the range of the plot to the range of the data of the second response column. See the Scripting Index in the JMP Help menu for an example.

---

## Keyboard Shortcuts

The following keyboard shortcuts can be used to manipulate the surface plot.

**Table 5.1**  Surface Plot Keyboard Shortcuts

| Key | Function |
| --- | --- |
| left, right, up, and down arrows | spin |
| Home, End | diagonally spin |
| Enter (Return) | toggles ArcBall appearance |
| Delete | roll counterclockwise |
| Control | boost spin speed 10X |
| Shift | allows continual spinning |

## Additional Examples of the Surface Plot Platform

This section contains additional examples of the Surface Plot platform.

## Construct a Surface Plot for a Single Mathematical Function

You can use the Surface Plot platform to produce the graph of a mathematical function without any data points.

1. Select **Graph > Surface Plot**.

   **Note:** This can be done with or without a data table open.

2. Click **OK**.

   When you click OK in the Surface Plot launch window without selecting any columns, the default surface plot report appears. If no data tables are open, the launch window does not appear and this step can be skipped.

3. In the Appearance Controls, select the **Show Formula** check box to show the formula space.

**Figure 5.11** Default Surface Plot with Formula
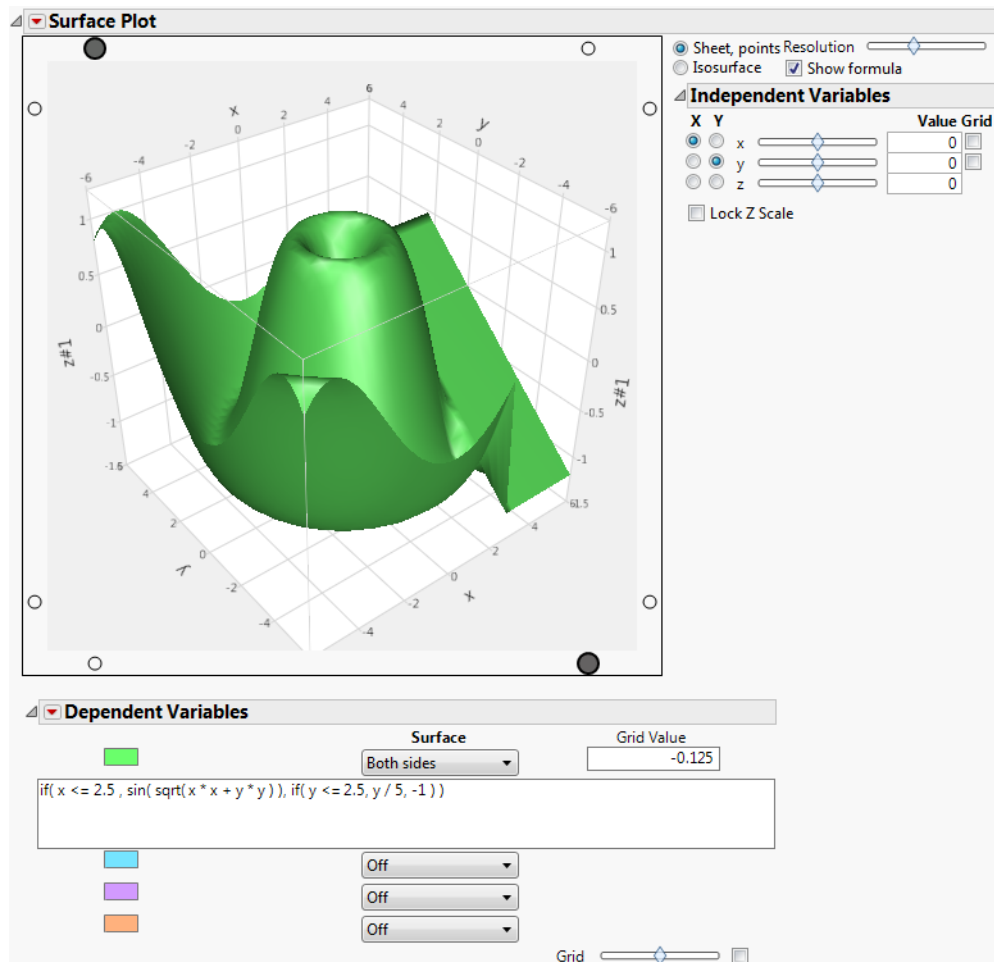


The default function shows in the box. To plot your own function, enter it in this box.

## Plotting Isosurfaces

Isosurfaces are the 3-D equivalent to a 2-D contour plot. An isosurface requires a formula with three independent variables. The Resolution slider determines the $n \times n \times n$ cube of points over which the formula is evaluated. The Value slider in the Dependent Variable section selects the isosurface (that is, the contour level) value.

This example uses the Tiretread.jmp data table. The RSM for 4 Responses script produces a response surface model for dependent variables ABRASION, MODULUS, ELONG, and HARDNESS. The predicted formulas for these four responses are in the data table.

1.  Select **Help > Sample Data Library** and open Tiretread.jmp.

2. Select **Graph > Surface Plot**.

3. Select Pred Formula ABRASION, Pred FormulaMODULUS, and Pred FormulaELONG and click **Columns**.

4. Click **OK**.

5. Select the Isosurface radio button in the appearance controls.

6. In the Dependent Variables report, select **Both Sides** from the Surface menu for all three variables.

**Figure 5.12**  Isosurface of Three Variables
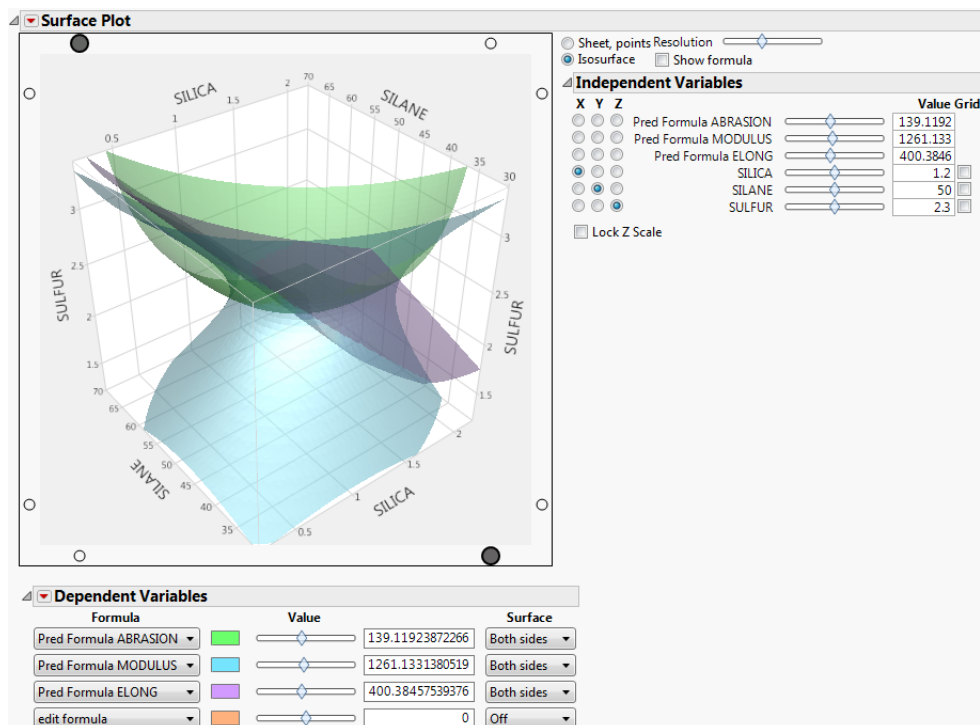


For the tire tread data, one might set the abrasion at a fixed minimum setting and the elongation at a fixed maximum setting. Use the MODULUS slider to see which values of MODULUS are inside the limits set by the other two surfaces.

# Mixture Profiler

## Explore Factor Effects Using Ternary Plots

The Mixture Profiler displays response contours on a ternary plot for mixture models, where three or more factors in the model are components (ingredients) in a mixture. The Mixture Profiler is useful for visualizing and optimizing response surfaces resulting from mixture experiments.

**Figure 6.1** Mixture Profiler Example
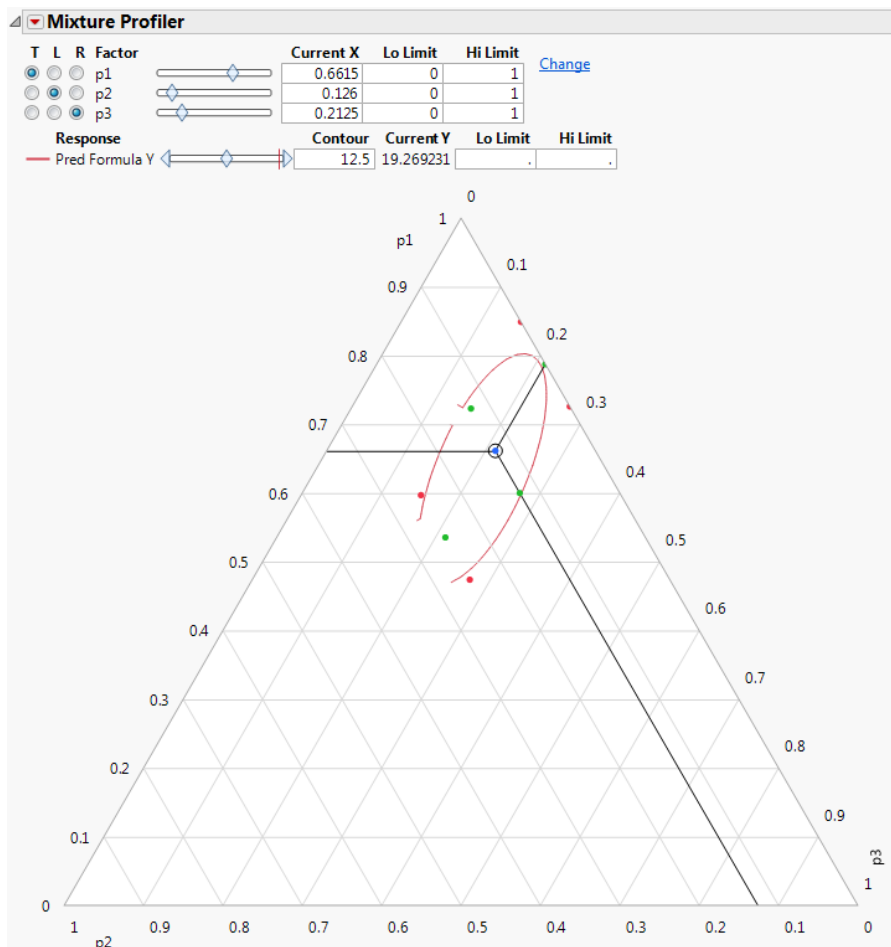
# Mixture Profiler Overview

The Mixture Profiler shows response contours on a ternary plot for mixture models, where three or more factors in the model are components (ingredients) in a mixture. The Mixture Profiler is useful for visualizing and optimizing response surfaces resulting from mixture experiments.

Many of the Mixture Profiler features are similar to those of the Contour Profiler as described in "Contour Profiler Platform Options" on page 81. Unique to the Mixture Profiler is the use of a ternary plot.

## Ternary Plot Overview

A ternary plot is a two-dimensional representation of three mixture components that sum to a constant. The plot is an equilateral triangle with an edge for each component. When unconstrained, each vertex of the triangle corresponds to a pure blend where one component is 1 (100%) and all other components are 0. When components are constrained, the feasible mixtures are represented by a portion of the ternary plot. Shading is used to exclude infeasible portions of the plot.

Figure 6.2 displays three components P1, P2, and P3. The three components are unconstrained with the range of each proportion ranging from 0 to 1 (100%). Three points are labeled with their coordinates (P1, P2, P3). One point (0.2, 0.2, 0.6) includes blue lines along the axis grid lines for each mixture component to guide you in reading the axes.

**Figure 6.2** Ternary Plot Example
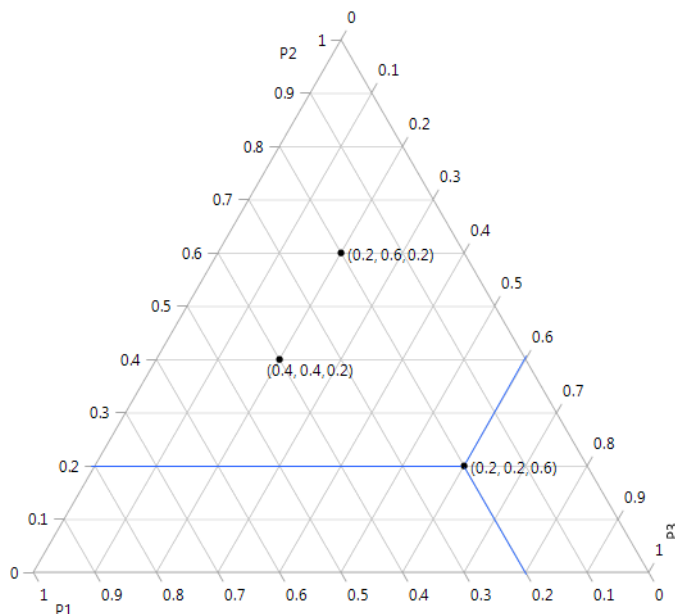


## Ternary Plots with More Than Three Components

The ternary plot can display only three components at a time. For a model with more than three components, the total of the three on-axis (displayed) components is 1 minus the total of the off-axis (non-displayed) components. The plot axes are scaled such that the maximum value a component can attain is 1 minus the total for the off-axis components.

Figure 6.3 shows the Mixture Profiler Ternary Plot for an experiment with 5 factors. The Five Factor Mixture.jmp data table is being used, with the Y1 Predicted column as the formula. The on-axis factors are x1, x2, and x3, and the factors x4 and x5 are off-axis. There are no constraints on any of the mixture components. The value for x4 is 0.1 and the value for x5 is 0.2, for a total of 0.3. The sum of x1, x2, and x3 is equal $1 - 0.3 = 0.7$. I. Note that the maximum value for a plot axis is now 0.7, not 1.

If you change the value for either x4 or x5, then the values for x1, x2, and x3 change, maintaining their relative proportions, to accommodate the constraint that factor values sum to 1.

**Figure 6.3** Ternary Plot with Scaled Axes to Account for Off-Axis Factors
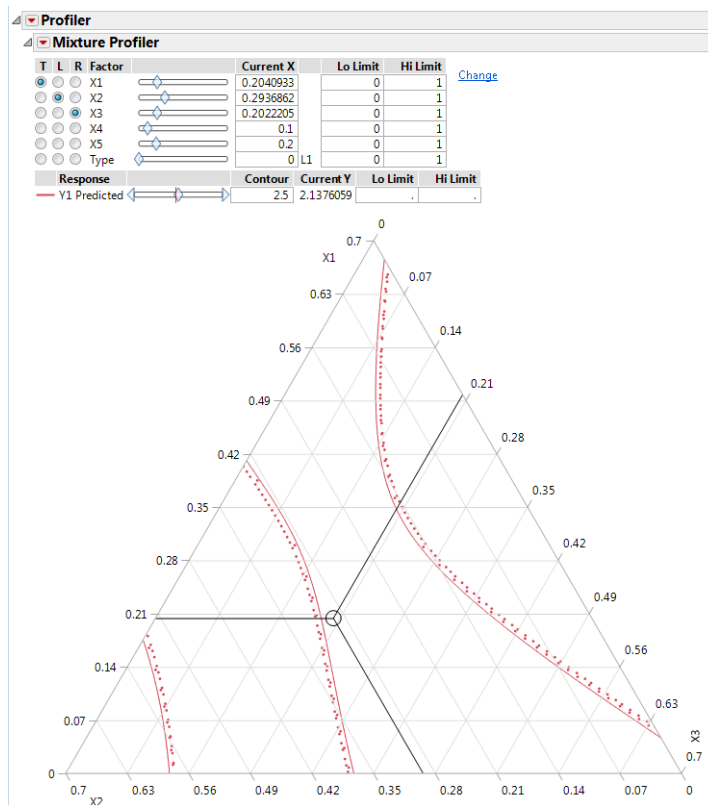


# Example of the Mixture Profiler

This example uses the Plasticizer.jmp sample data table There are three mixture components (p1, p2, and p3) and one response (Y). There are constraints on the component levels. This data table already contains a saved prediction formula column for the response.

1.  Select **Help > Sample Data Library** and open Plasticizer.jmp.

2.  Select **Graph > Mixture Profiler**.

3.  Select Pred Formula Y and click **Y, Prediction Formula**.

4.  Select Y from the Y, Prediction Formula role and click **Remove**.

    **Note:** The Y column role is set to Y in the Columns list. The role automatically assigns columns in launch windows. In this case, the Y column was assigned to the Y role when the mixture profiler was launched. However, the Y column does not contain a formula, and the Mixture Profiler Y column requires a formula column.

5.  Click **OK**.

**Figure 6.4**  Example of Mixture Profiler
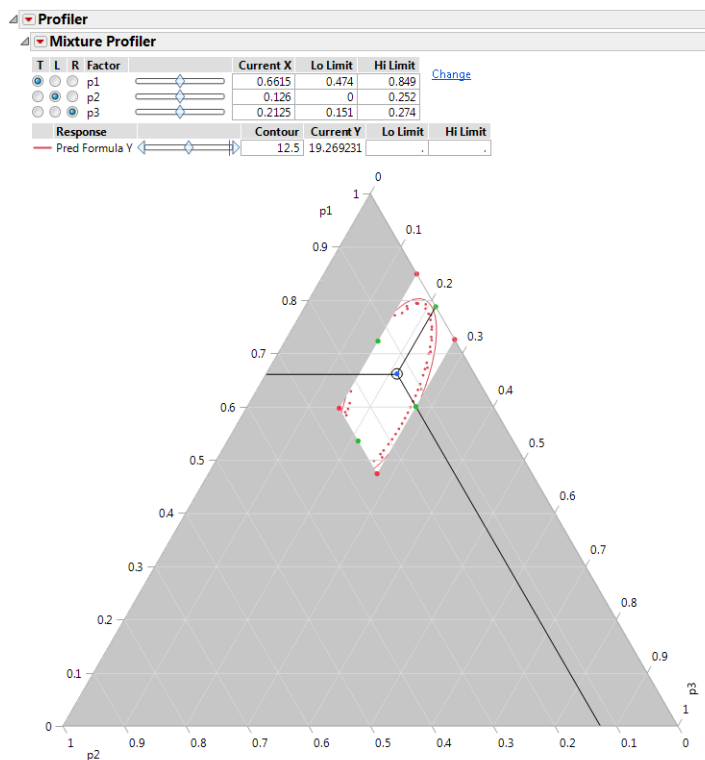


Use the factor sliders to explore how changes impact the response. The unshaded region is the feasible region for this data set.

## Launch the Mixture Profiler Platform

The Mixture Profiler can be accessed in the following ways:

• The Mixture Profiler can be accessed directly from the Graph menu. When you access the Mixture Profiler in this way, the Mixture Profiler launch window appears. See "Profiler Launch Windows" on page 29 in the "Introduction to Profilers" chapter for details.

• The Mixture Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" on page 28 in the "Introduction to Profilers" chapter for details about the availability of the Mixture Profiler in different platforms.

- The Mixture Profiler can be accessed from the Model Comparisons platform. Select Profiler from the Model Comparisons red triangle menu. Then, select Mixture Profiler from the Profiler red triangle menu.

- **JMP PRO** The Mixture Profiler can be accessed from the Formula Depot platform. Select Profiler from the Formula Depot red triangle menu and select the models to be profiled. Then, select Mixture Profiler from the Profiler red triangle menu.

# The Mixture Profiler Report

The initial Mixture Profiler report shows a mixture profiler plot, factor settings and controls, and response settings and controls.

## Factor Settings and Controls

**Figure 6.5** Factor Settings and Controls



| T | L | R | Factor | | Current X | Lo Limit | Hi Limit | |
|---|---|---|--------|---|-----------|----------|----------|---|
| ● | ○ | ○ | p1 | | 0.6615 | 0.474 | 0.849 | Change |
| ○ | ● | ○ | p2 | | 0.126 | 0 | 0.252 | |
| ○ | ○ | ● | p3 | | 0.2125 | 0.151 | 0.274 | |
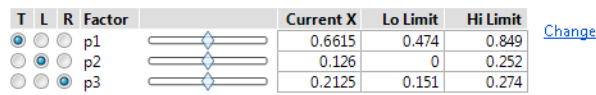
**T, L, R**   Radio buttons to control factor assignments to the axes in the mixture profiler plot. **T** = top, **L** = left, and **R** = right.

**Factor**   The list of factors.

**Current X**   The current factor settings. Click in a box to change the value of a factor. The values for the other factors adjust proportionally to maintain the mixture sum. The slider controls can also be used to change factor settings.

**Lo Limit**   The lower limit on each factor. Click in a box to change the value.

**Hi Limit**   The upper limit on each factor. Click in a box to change the value.

**Note:** When factor limits are used, the profiler shows regions that are feasible as unshaded.

**Change**   Opens the Factor Settings window. See "Mixture Profiler Platform Options" on page 111.

## Response Settings and Controls

**Figure 6.6** Response Settings and Controls



| Response | | Contour | Current Y | Lo Limit | Hi Limit |
|----------|---|---------|-----------|----------|----------|
| — Pred Formula Y | ◁——◇——▷ | 12.5 | 19.269231 | . | . |

**Response**   The list of one or more responses.

**Contour**   The current value of the contour on the mixture profiler. Click in the box to change the value. The slider control can also be used to change the value of the contour.

**Current Y**   The predicted response based on the current X settings. The value is at the center of the cross hairs on the profiler plot. This value updates as the factor settings are changed.

**Lo Limit**   Enables you to set a lower limit for your response.

**High Limit**   Enables you to set an upper limit for your response.

**Note:** When response limits are used, the profiler shows regions that are feasible as unshaded.

## Mixture Profiler Platform Options

**Specify Factor Values**   Opens a dialog for specifying factor values. The entered values should sum to1. If needed, JMP adjusts the specified values to feasible settings. The option is also available from the **Change** button on the report window.

**Show Current Value**   Shows or hides the three-way crosshairs on the plot. The intersection of the crosshairs represents the current factor values. The Current X values above the plot give the exact coordinates of the crosshairs.

**Show Constraints**   Shows or hides the shading resulting from any constraints on the factors. Those constraints can be entered in the Lo Limits and Hi Limits columns above the plot, or in the Mixture Column Property for the factors.

**Up Dots**   Shows or hides the dotted lines corresponding to each contour. The dotted lines show the direction of increasing response values, so that you get a sense of direction.

**Contour Grid**   Draws contours on the plot at intervals that you specify.

**Remove Contour Grid**   Removes the contour grid if one is on the plot.

**Factor Settings**   A submenu of commands that enables you to save and transfer the Mixture Profiler settings to other parts of JMP. Details on this submenu are found in the discussion of the profiler on "Factor Settings" on page 45.

### Customizations for the Mixture Profiler

Customize the Mixture Profiler by right-clicking on the plot, and selecting **Customize**.

**Contour**   Alter line color, fill color, and level of line transparency. If there are multiple contour lines on the plot, each appears individually on the list of options.

**Component Constraints**   Alter component constraint display properties.

**Linear Constraints**   Alter linear constraint display properties.

**Grid Lines**   Use Axis Settings window to modify grid lines.

**Reference Lines**   Use Axis Settings window to modify reference lines.

Double-click an axis on the plot to view the Axis Settings window. Here you can specify axis properties in detail, such as tick marks, grid lines and reference lines.

**Crosshairs**   Alter the display properties of the crosshairs on the plot.

**Marker**   Alter the display properties of the plot marker.

## Linear Constraints

The Mixture Profiler can display linear constraints. To do this, a Constraint Table Script must be part of the data table. For more information about creating the Table Script, see "Linear Constraints" on page 35 in the "Introduction to Profilers" chapter.

When using constraints, unfeasible regions are shaded in the profiler. Figure 6.7 shows an example of a mixture profiler with shaded regions due to four constraints. The unshaded portion is the resulting feasible region. The constraints are below:

- 4*p2 + p3 ≤ 0.8
- p2 + 1.5*p3 ≤ 0.4
- p1 + 2*p2 ≥ 0.8
- p1 + 2*p2 ≤ 0.95

**Figure 6.7**  Shaded Regions Due to Linear Constraints
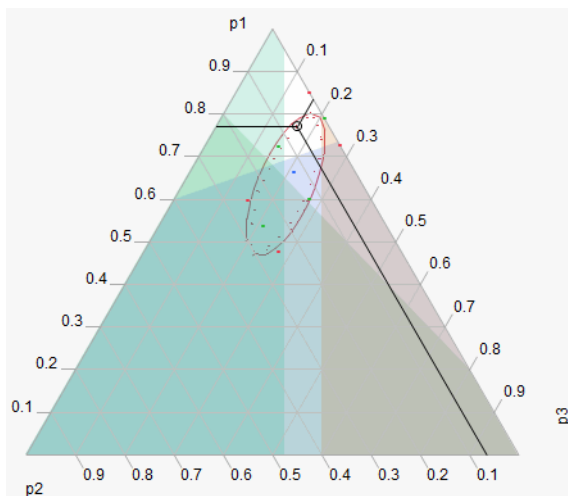
# Additional Examples of the Mixture Profiler Platform

## Mixture Variables with a Process Variable

This example uses the Fish Patty.jmp sample data table. The data, adapted from Cornell (1990), comes from an experiment to optimize the texture of fish patties. The columns Mullet, Sheepshead, and Croaker are mixture components. Each column represents the proportion of a fish type in the fish patty. The Temperature column is a process variable. It is the oven temperature used to bake the patties. The column Rating is the response and is a measure of texture acceptability, where higher is better. A response surface model was fit to the data and the prediction formula was stored in the column Predicted Rating.

1.  Select **Help > Sample Data Library** and open Fish Patty.jmp.
2.  Select **Graph > Mixture Profiler**.
3.  Select  Predicted Rating and click **Y, Prediction Formula**.

**Figure 6.8**  Initial Output for Mixture Profiler
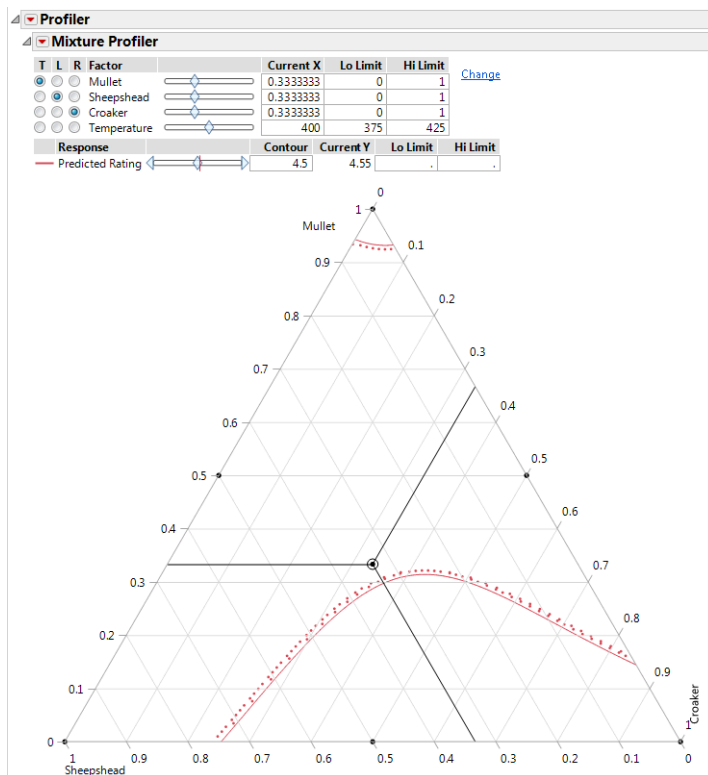
4.  The manufacturer wants the rating to be at least 5. Use the slider control for Predicted
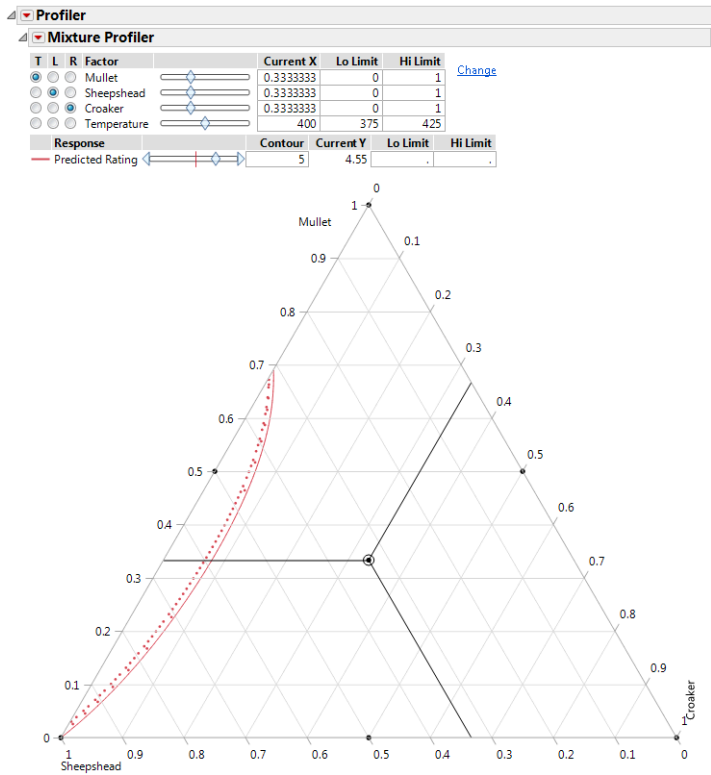    Rating to move the contour close to 5. Alternatively, you can enter 5 in the Contour edit box
    to set the contour to 5.

**Figure 6.9**  Contour Showing a Predicted Rating of 5



The Up Dots shown along the contour indicate the direction of increasing Predicted Rating.

5.  Enter 5 in the Lo Limit edit box for the Predicted Rating. The resulting shaded region
    represents factor combinations that yield a rating less than 5. To produce patties with at
    least a rating of 5, the manufacturer can set the factors values anywhere in the feasible
    (unshaded) region. The region has small proportions of Croaker (<10%), mid to low
    proportions of Mullet (<70%) and mid to high proportions of Sheepshead (>30%). This
    region is for a baking temperature of 400 degrees.

**Figure 6.10** Contour Shading Showing Predicted Rating of 5 or More.
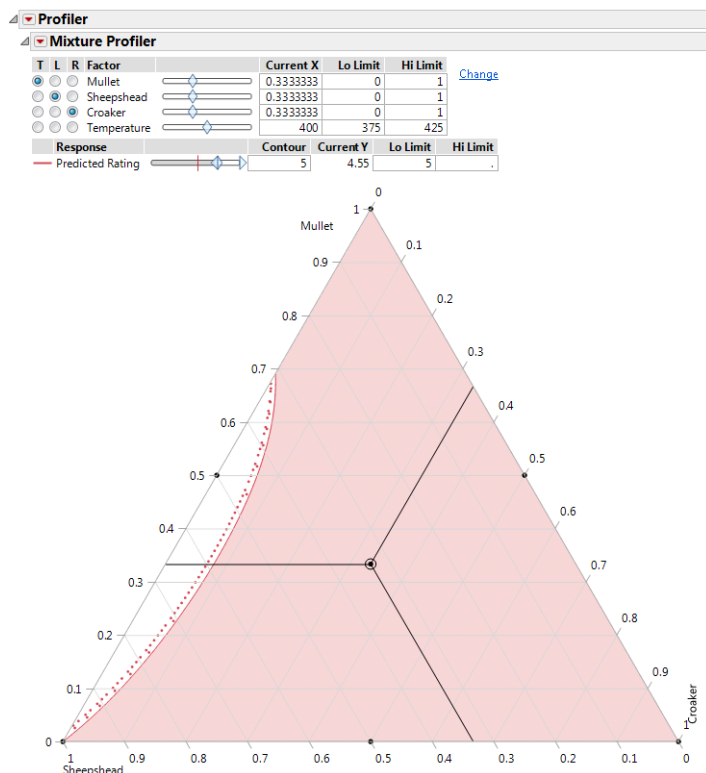


6. Move the slide control for Temperature to observe how the feasible region changes for different temperature settings.

Additional analyses might include:

- Optimize the response across all four factors simultaneously. See the "Custom Profiler" chapter on page 121 or "Desirability Profiling and Optimization" on page 47 in the "Profiler" chapter.

- Simulate the response as a function of the random variation in the factors and model noise. See the "Simulator" chapter on page 127.

## Multiple Responses with Five Mixture Variables

This example uses the Five Factor Mixture.jmp sample data table. There are five components (x1 to x5), one categorical process factor (Type), and three responses, Y1, Y2, and Y3. A response surface model is fit to each response and the prediction equations are saved in Y1 Predicted, Y2 Predicted, and Y3 Predicted.

1. Select **Help > Sample Data Library** and open Five Factor Mixture.jmp.

2.  Select **Graph > Mixture Profiler**.

3.  Select Y1 Predicted, Y2 Predicted, and Y3 Predicted and click **Y, Predicted**.

4.  Click **OK**.

5.  Enter 3 in the Contour edit box for Y3 Predicted so that the contour is visible on the plot.

**Figure 6.11**   Mixture Profiler with Y3 Predicted Contour Set to 3
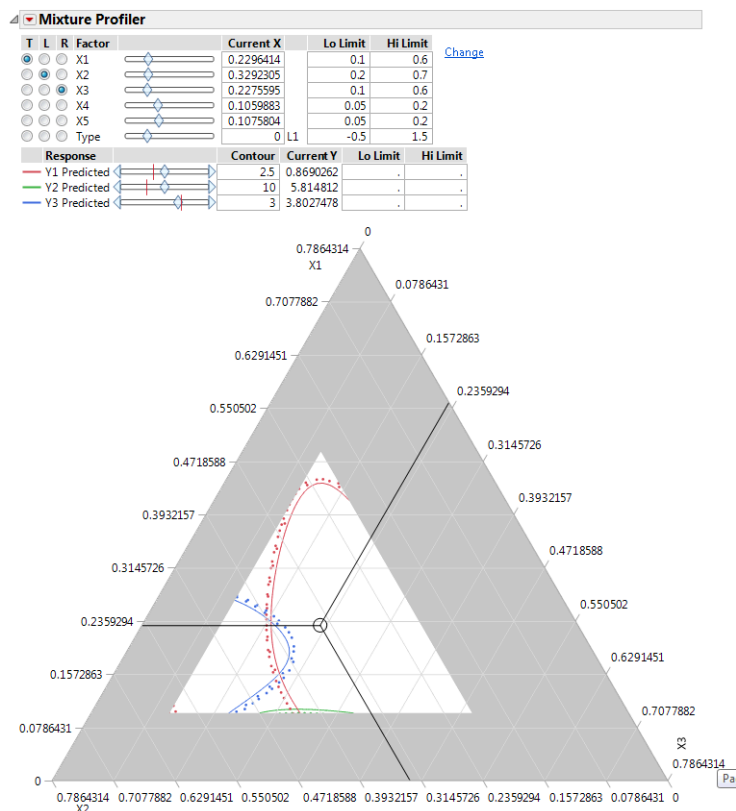


A few items to note about the output in Figure 6.11.

*   The on axis factors are X1, X2, and X3 as indicated by the axes labels and the radio buttons in the factor settings. All of the factors have low and high limits, which were entered previously as Column Properties. See The Column Info Window chapter in the *Using JMP* book for more information about entering column properties. Alternatively, you can enter the low and high limits directly by entering them in the Lo Limit and Hi Limit boxes.

*   The white unshaded region is the feasible region. This is determined by the factor limits.

*   The on-axis factors, x1, x2, and x3, radio buttons are selected.

- The categorical factor, Type, has a radio button, but it cannot be assigned to the plot. The current value for Type is L1, which is listed immediately to the right of the Current X box. The Current X box for Type uses a 0 to represent L1.

- All three prediction equations have contours on the plot and are differentiated by color.

Suppose the manufacturer desires to hold Y1 less than 1, Y2 greater than 8 and Y3 between 4 and 5, with a target of 4.5. The Mixture Profiler can help you investigate the response surface and find optimal factor settings.

6. Enter1 in the Y1 Predicted Hi Limit edit box. Enter 8 in the Y2 Predicted Lo Limit edit box. Enter 4 in theY3 Predicted Lo Limit edit box and 5 in the Y3 Predicted Hi Limit edit box.

7. The feasible region remains white (unshaded). Use the Response slider controls to position the contours in the feasible region. Alternatively, move the cross hair into the feasible region

**Figure 6.12** Feasible Region After Setting Response Limits
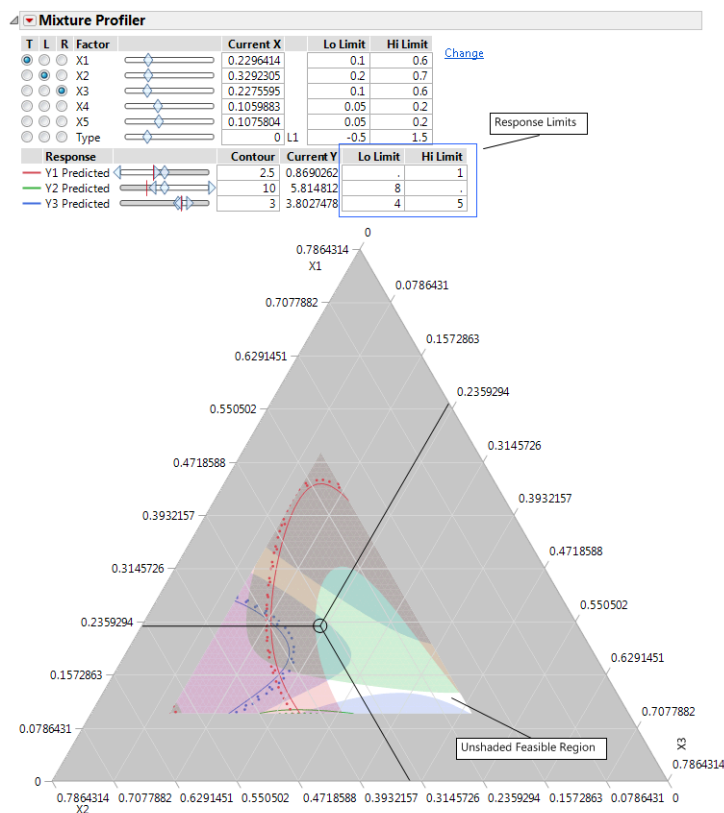


8. Select the magnifier tool 🔍 to zoom in on the feasible region.

**Figure 6.13**  Ternary Plot with Enlarged Feasible Region
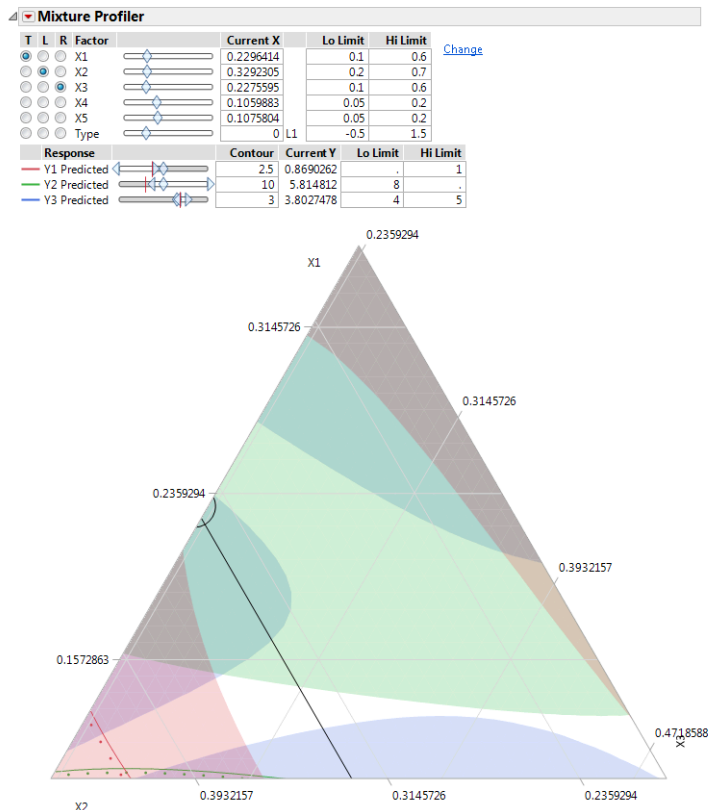


The manufacturer wants to maximize Y1, minimize Y2, and have Y3 at 4.5.

9.  Use the slider controls or Contour edit boxes for Y1 Predicted to maximize the red contour within the feasible region. The Up Dots show direction of increasing predicted response.

10. Use the slider controls or Contour edit boxes for Y2 Predicted to minimize the green contour within the unshaded region.

11. Enter 4.5 in the Contour edit box for Y3 Predicted to set the blue contour to the target value.

The resulting three contours do not all intersect at one spot. Y1 and Y2 cannot be optimized with Y3 on target, so you have to compromise. Position the three-way crosshairs in the middle of the contours to explore the factor levels that produce those response values. Note that these response contours are for the current settings of x4, x5, and Type.

**Figure 6.14** Factor Settings for Optimal Results
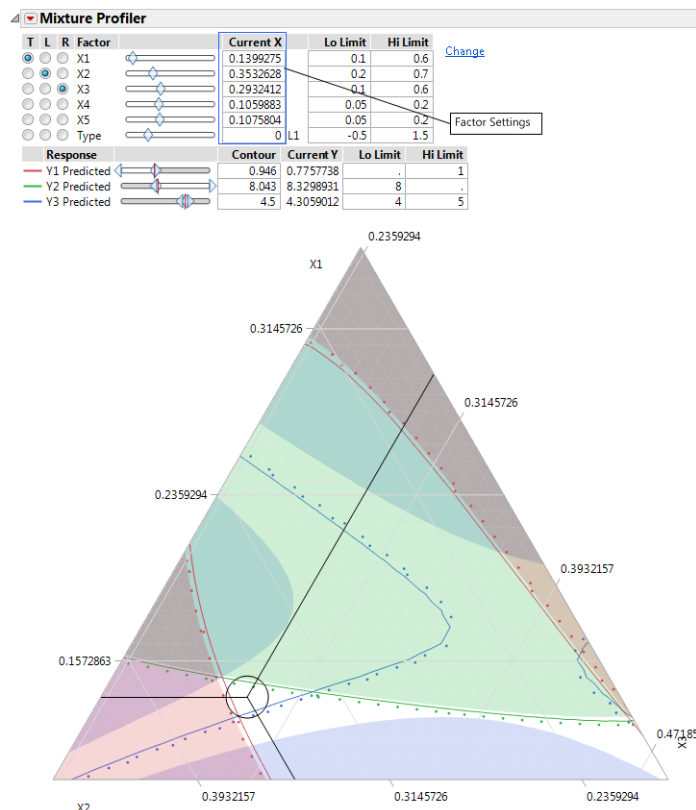


12. Click the Mixture Profiler red triangle and select **Factor Settings > Remember Settings** to save the current settings. The settings are appended to the bottom of the report window.

**Figure 6.15** Remembered Settings

| Setting | X1 | X2 | X3 | X4 | X5 | Type | Y1 Predicted | Y2 Predicted | Y3 Predicted |
|---|---|---|---|---|---|---|---|---|---|
| ○ Setting 1 | 0.1352764 | 0.359975 | 0.29118 | 0.1059883 | 0.1075804 | L1 | 0.9244618 | 8.5514568 | 4.3705523 |

With the current settings saved, you can now change the values of x4, x5 and Type to explore their impact on the feasible region. You can compare the factor settings and response values for each level of Type by referring to the Remembered Settings report.

# Custom Profiler

## Explore Response Surfaces Using a Numerical Calculator

The Custom Profiler enables you to optimize factor settings without graphical output. This is useful for large problems that would have too many graphs to visualize well.

**Figure 7.1** Custom Profiler Example

# Custom Profiler Overview

The Custom Profiler enables you to optimize factor settings without graphical output. The Custom Profiler can be used for problems of any size. It is especially useful for large problems where the standard graphical profiler has too many graphs to visualize well.

The Custom Profiler report has many fields in common with other profilers. The Benchmark field holds a value, or benchmark, of the predicted response. You can compare new results to the benchmark value and update the value based on the current factor settings.

The Optimization report enables you to specify the formula to be optimized and specifications for the optimization.

**Figure 7.2** Custom Profiler
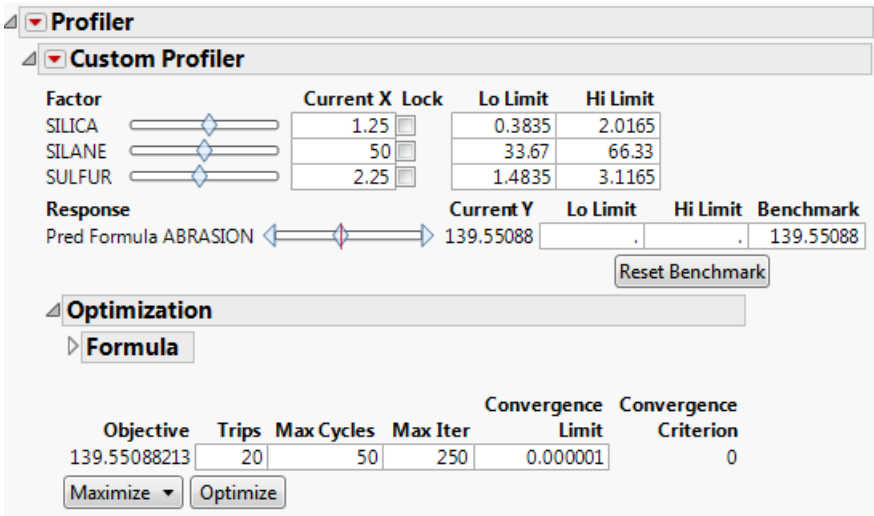


# Example of the Custom Profiler

This example uses data that demonstrates the flow of water through a borehole that is drilled from the ground surface through two aquifers. You want to optimize the predicted value.

1. Select **Help > Sample Data Library > Design Experiment** and open Borehole Latin Hypercube.jmp

2. Select **Graph > Custom Profiler**.

3. Select prediction formula and click **Y, Prediction Formula.**

**Figure 7.3** Completed Custom Profiler Launch Window
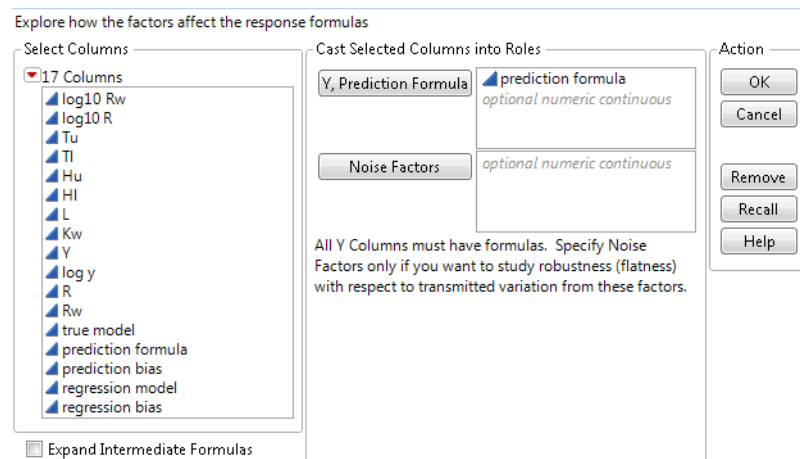


4. Click **OK**.

   Note that the Benchmark value is 70.83. The Current Y value is also 70.83. This value is the predicted response with all factors set to their mean values.

5. In the Custom Profiler Report, click **Optimize**.

**Figure 7.4** Custom Profiler Report after Optimization
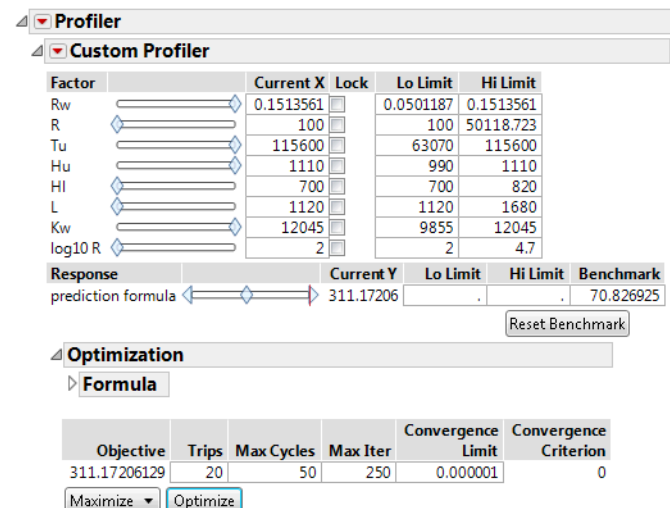


The optimization routine found an optimum predicted response at 311.17. In order to obtain the optimum of 311.17, all factors are set either at their minimum or maximum values. The optimum is greater than the initial Benchmark value of 70.83.

# Launch the Custom Profiler Platform

The Custom Profiler can be accessed in the following ways:

- The Custom Profiler can be accessed directly from the Graph menu. When you access the Custom Profiler in this way, the Custom Profiler launch window appears. See "Profiler Launch Windows" on page 29 in the "Introduction to Profilers" chapter for details.

- The Custom Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" on page 28 in the "Introduction to Profilers" chapter for details about the availability of the Custom Profiler in different platforms.

- The Custom Profiler can be accessed from the Model Comparisons platform. Select Profiler from the Model Comparisons red triangle menu. Then, select Custom Profiler from the Profiler red triangle menu.

- **JMP PRO** The Custom Profiler can be accessed from the Formula Depot platform. Select Profiler from the Formula Depot red triangle menu and select the models to be profiled. Then, select Custom Profiler from the Profiler red triangle menu.

# The Custom Profiler Report

The initial Custom Profiler report shows settings and controls for the factors, responses, and optimization.

## Factor Settings and Controls

**Figure 7.5** Factor Settings and Controls

| Factor | | Current X | Lock | Lo Limit | Hi Limit |
|---|---|---|---|---|---|
| Rw | | 0.1513561 | | 0.0501187 | 0.1513561 |
| R | | 100 | | 100 | 50118.723 |
| Tu | | 115600 | | 63070 | 115600 |
| Hu | | 1110 | | 990 | 1110 |
| Hl | | 700 | | 700 | 820 |
| L | | 1120 | | 1120 | 1680 |
| Kw | | 12045 | | 9855 | 12045 |
| log10 R | | 2 | | 2 | 4.7 |

**Factor** The list of model factors.

**Current X** The current value of each factor. Click in a box to change the value of a factor. The slider controls can also be used to change factor settings.

**Lock** Enables you to lock a factor so that it is fixed when the optimization is performed. You can change a locked factor using the slider or clicking in the box in the Current X column. The lock applies only to the optimization.

**Nominal Column**   Unlabeled column to the right of the Lock column that lists the current value of nominal factors.

---

**Note:** The Current X column for nominal factors displays a coded (numeric) value for the current nominal factor.

---

**Lo Limit**   The lower limit for each factor. Click in a box to change the value.

**High Limit**   The upper limit for each factor. Click in a box to change the value.

## Response Settings and Controls

**Figure 7.6**  Response Settings and Controls



**Response**   The list of one or more responses.

**Current Y**   The predicted response based on the current X settings. This value updates as the factor settings are changed.

**Lo Limit**   Enables you to set a lower limit for your response.

**High Limit**   Enables you to set an upper limit for your response.

**Benchmark**   A saved predicted value of the response. Initially, this value is set to the predicted value when all responses are at their mean value.

**Reset Benchmark**   Updates the benchmark value to the current predicted value.

## Optimization Settings and Controls

**Figure 7.7**  Optimization Settings and Controls



**Formula**   The formula to be optimized. When a single response is used, the expression is the response column name. When multiple responses are used, the expression is a sum of desirability functions. You can edit the objective expression.

**Objective**   The current value of the objective function. When a single response is used, the objective expression is the predicted response. When multiple responses are used, the

objective expression is the desirability function. For more information about desirability functions, see "Desirability Profiling and Optimization" on page 47 in the "Profiler" chapter.

**Trips** The number of random starts in the optimization algorithm. Each trip restarts the algorithm at a different starting point. This guards against finding local solutions.

**Max Cycles** The maximum number of cycles used in the optimization algorithm. Each cycle is single pass through the input parameters and optimizes each one individually.

**Max Iter** The maximum number of optimization iterations per cycle for each input parameter.

**Convergence Limit** The upper limit for the convergence criterion for the optimization algorithm. If the convergence criterion becomes less than this value, the algorithm stops.

**Convergence Criterion** The value of the convergence criterion for the optimization algorithm.

**Maximize** Enables you to choose to maximize or minimize the objective function.

**Optimize** Starts the optimization algorithm.

## Custom Profiler Platform Options

**Factor Settings** Contains options identical to the Factor Settings submenu in the Prediction Profiler. See "Factor Settings" on page 45 in the "Profiler" chapter.

**Log Iterations** Creates a data table that contains iterations of the optimization algorithm. The data table appears after the **Optimize** button is clicked.

**Alter Linear Constraints** Enables you to add, change, or delete linear constraints. The constraints are used in the Custom Profiler. See "Linear Constraints" on page 35 in the "Introduction to Profilers" chapter.

**Save Linear Constraints** Saves existing linear constraints as a data table script that is named Constraint. See "Linear Constraints" on page 35 in the "Introduction to Profilers" chapter.

**Simulator** Launches the Simulator. See Chapter 8, "Simulator".

# Simulator

## Explore the Effects of Variation on Responses

Simulation enables you to discover the distribution of model outputs as a function of the random variation in the factors and model noise. The simulator in the profilers provides a way to set up the random inputs and run the simulations, producing an output table of simulated values. In the Profiler, the Simulator is integrated into the graphical layout. Factor specifications are aligned below each factor's profile. A simulation histogram is shown on the right for each response.

**Figure 8.1** Profiler with Simulator Example
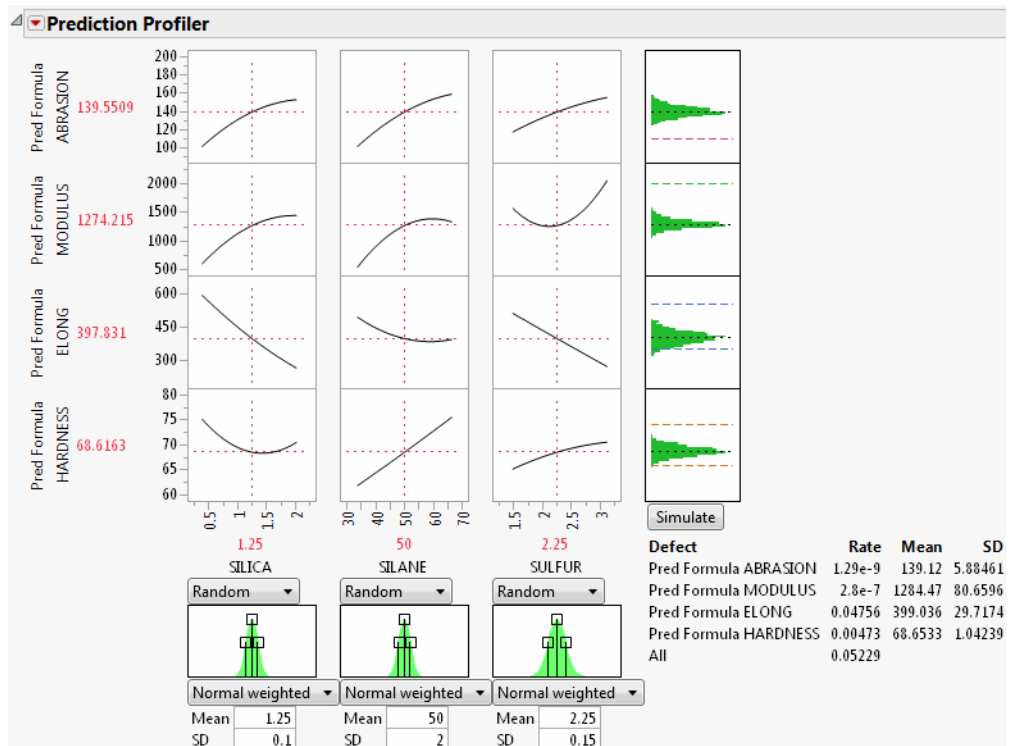
# Simulator Overview

Simulation enables you to discover the distribution of model outputs as a function of the random variation in the factors and model noise. The simulator in the profilers provides a way to set up the random inputs and run the simulations, producing an output table of simulated values.

An example of the simulator would be to model the defect rate of a process, and see how robust the defect rate is with respect to variation in the model factors. If specification limits have been set for the responses, they are carried over into the simulation output. The inclusion of specification limits allows for a prospective capability analysis of the simulated model.

In the Profiler, the Simulator is integrated into the graphical layout. Factor specifications are aligned below each factor's profile. A simulation histogram is shown on the right for each response.

The profiler simulator supports mixture terms. The simulation assigns the random value to each mixture factor and then proportionally adjusts the values to maintain the factor constraints.

**Figure 8.2** Profiler with Simulator
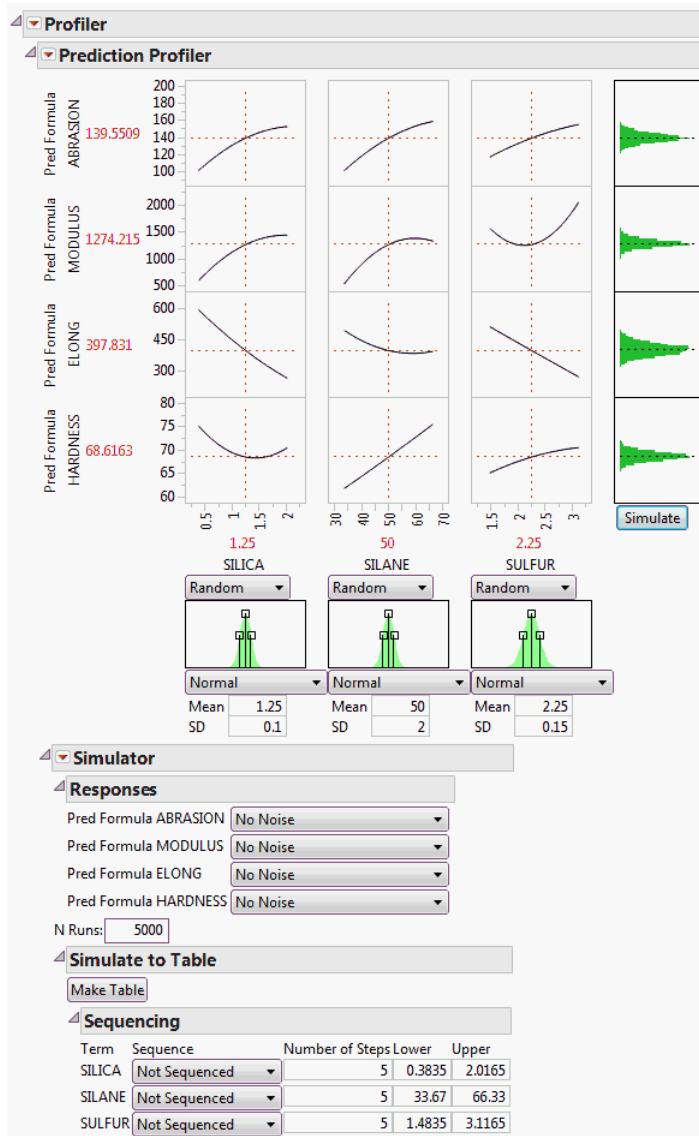


In the contour and custom profilers, the Simulator is less graphical, and kept separate. There are no integrated histograms, and the interface is textual. However, the internals and output tables are the same.

# Example of the Simulator

**Tip:** The Make Table and Sequencing options are most useful when you have random values. Sequencing options are available for the following distributions only: Normal, Uniform, and Triangular.

Specify the number of runs in the simulation by entering it in the N Runs box.

After the factor and response distributions are set, click the **Simulate** button to run the simulation. Or, use the **Make Table** button to create a table with N Runs for the number of rows. Each row is populated with a random draw from the specified distributions, and the corresponding response values are computed. If spec limits are given, the table also contains a column specifying whether a row is in or out of spec.

Use sequencing to examine how the distribution of the response changes when the mean (sequencing location) and variability (sequencing spread) of the inputs change.

1.  Select **Help > Sample Data Library** and open Tiretread.jmp.
2.  Select **Graph > Profiler**.
3.  Select Pred Formula ABRASION and Pred Formula MODULUS and click **Y, Prediction Formula**.
4.  Click **OK**.
5.  From the red triangle menu next to Prediction Profiler, select **Simulator**.
6.  Change each factor to be **Random** instead of **Fixed**.
7.  Change the **N Runs** value to 100.
8.  Open Simulate to Table then Sequencing.

**Figure 8.3**  Simulator Settings



You want to examine how the responses change when the mean values change.

9.  For SILICA, select **Sequence Location**. Keep the number of steps at 5. Because the mean is 1.25, change the values over a range of 1 (lower) to 2 (upper).

10. For SILANE, select **Sequence Location**. Keep the number of steps at 5. Because the mean is 50, change the values over a range of 40 (lower) to 60 (upper).

11. For SULFUR, select **Sequence Location**. Keep the number of steps at 5. Because the mean is 2.25, change the values over a range of 2 (lower) to 3 (upper).

**Figure 8.4**  Sequencing Settings



12. Click **Make Table**.

You can see that the SILICA Mean, SILANE Mean, and SULFUR Mean columns contain five steps for each range of values (Silica Mean is 1, 1.25, 1.5, 1.75, and 2; Silane Mean is 40, 42.5, 45, 47.5, and 60; and so on.) Pred Formula ABRASION and Pred Formula MODULUS values are calculated for each combination of values, so you can see how the responses change as the factor values change.

13. Select **Analyze > Distribution**.

14. Select Pred Formula ABRASION and SILICA Mean and click **Y, Columns**.

15. Click **OK**.

**Figure 8.5**  Distribution of SILICA Mean by Pred Formula ABRASION
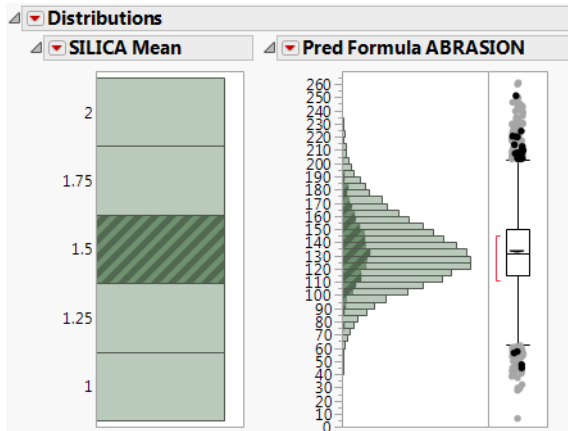


Click a histogram bar that corresponds to a SILICA mean to see how the prediction formula for ABRASION changes given the selected mean.

## Specify Factors

Factors (inputs) and responses (outputs) are already given roles by being in the Profiler. Additional specifications for the simulator are on how to give random values to the factors, and add random noise to the responses.

For each factor, the choices of how to give values are as follows:

**Fixed**   Fixes the factor at the specified value. The initial value is the current value in the profiler, which might be a value obtained through optimization.

**Random**   Gives the factor the specified distribution and distributional parameters.

See the Formula Functions Reference chapter in the *Using JMP* book for descriptions of most of these random functions. If the factor is categorical, then the distribution is characterized by probabilities specified for each category. By default, the categories are given equal probabilities of occurring. You can change the probabilities by entering in different values in the dialog that appears. If the probabilities you enter do not sum to 1, the values are normalized to sum to 1.

**Normal weighted** is normally distributed with the given mean and standard deviation, but a special stratified and weighted sampling system is used to simulate very rare events far

out into the tails of the distribution. This is a good choice when you want to measure very low defect rates accurately. See "Statistical Details for the Simulator" on page 157.

**Normal truncated** is a normal distribution limited by lower and upper limits. Any random realization that exceeds these limits is discarded and the next variate within the limits is chosen. This is used to simulate an inspection system where inputs that do not satisfy specification limits are discarded or sent back.

**Normal censored** is a normal distribution limited by lower and upper limits. Any random realization that exceeds a limit is just set to that limit, putting a density mass on the limits. This is used to simulate a re-work system where inputs that do not satisfy specification limits are reworked until they are at that limit.

**Sampled** means that JMP selects values at random from that column in the data table.

**External** means that JMP selects values at random from a column in another table. You are prompted to choose the table and column.

The **Aligned** check box is used for two or more Sampled or External sources. When checked, the random draws come from the same row of the table. This is useful for maintaining the correlation structure between two columns. If the Aligned option is used to associate two columns in different tables, the columns must have equal number of rows.

In the Profiler, a graphical specification shows the form of the density for the continuous distributions, and provides control points that can be dragged to change the distribution. The drag points for the Normal are the mean and the mean plus or minus one standard deviation. The Normal truncated and censored add points for the lower and upper limits. The Uniform and Triangular have limit control points, and the Triangular adds the mode.

**Figure 8.6** Distributions
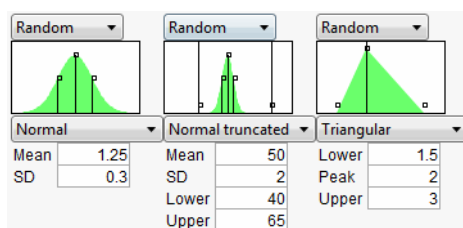


**Expression**   Allows you to write your own expression in JMP Scripting Language (JSL) form into a field. This gives you flexibility to make up a new random distribution. For example, you could create a censored normal distribution that guaranteed nonnegative values with an expression like `Max(0,RandomNormal(5,2))`. In addition, character results are supported, so `If(Random Uniform() < 0.2, "M", "F")` works fine. After entering the expression, click the **Reset** button to submit the expression.

**Multivariate**   Allows you to generate a multivariate normal for when you have correlated
factors. Specify the mean and standard deviation with the factor, and a correlation matrix
separately.

**Figure 8.7**  Using a Correlation Matrix



## Simulator Report Options

The following options are available from the Simulator red triangle menu:

**Automatic Histogram Update**   Toggles histogram update, which sends changes to all
histograms shown in the Profiler, so that histograms update with new simulated values
when you drag distribution handles.

**Defect Profiler**   Shows the defect rate as an isolated function of each factor. This command is
enabled when spec limits are available, as described below.

**Defect Parametric Profile**   Shows the defect rate as an isolated function of the parameters of
each factor's distribution. This command is enabled when the Defect Profiler is launched.
For more information see "Defect Parametric Profiler" on page 143.

**N Strata**   Is a hidden option accessible by holding down the Shift key before clicking the
Simulator red triangle menu. This option enables you to specify the number of strata in
Normal Weighted. For more information also see "Statistical Details for the Simulator" on
page 157.

**Set Random Seed**   Is a hidden option accessible by holding down the Shift key before clicking
the Simulator red triangle menu. This option enables you to specify a seed for the
simulation starting point. This enables the simulation results to be reproducible, unless the
seed is set to zero. The seed is set to zero by default. If the seed is nonzero, then the latest
simulation results are output if the **Make Table** button is clicked.

**Simulation Experiment**   Runs a designed simulation experiment on the locations of the factor
distributions. A window appears, allowing you to specify the number of design points, the
portion of the factor space to be used in the experiment, and which factors to include in the
experiment. For factors not included in the experiment, the current value shown in the
Profiler is the one used in the experiment. For more information, see "Simulation
Experiment" on page 144.

**Spec Limits**   Opens a table to set or edit specification limits.

# Responses Report Options

If the model is only partly a function of the factors, and the rest of the variation of the response is attributed to random noise, then you will want to specify this with the responses. The choices are:

**No Noise**   Evaluates the response from the model, with no additional random noise added.

**Add Random Noise**   Obtains the response by adding a normal random number with the specified standard deviation to the evaluated model.

**Add Random Weighted Noise**   Is distributed like Add Random Noise, but with weighted sampling to enable good extreme tail estimates.

**Add Multivariate Noise**   Yields a response as follows: A multivariate random normal vector is obtained using a specified correlation structure, and it is scaled by the specified standard deviation and added to the value obtained by the model.

# Simulate to Table Report Options

Saves the results of the simulation to a data table. If a response has specification limits, the table includes a Y In Spec column that indicates if the simulated response is within the specification limits. The table also includes an Objective column, labeled Obj. If you have requested Desirability functions prior to simulation, this column is the evaluation of the desirability function that you have created for your responses. If no desirability functions have been requested, the column contains all missing values.

## Sequencing Report Options

The sequencing allows you to run multiple simulations and save them to a data table. Select an option to sequence the location or spread of the term. Select the number of steps and the range for the sequencing. For each combination of the number of steps the simulation will run N Runs. If you have two terms, each with 5 steps, and N Runs = 100 the resulting simulation table will have 2,500 rows.

# Spec Limits Report Options

Allows you to enter or edit specification limits responses. Click the **Save** button to save specification limits to the response column properties in the data table.

# Specification Limits

The profilers support specification limits on the responses, providing a number of features

- In the Profiler, if you do not have the Response Limits property set up in the input data table to provide desirability coordinates, JMP looks for a Spec Limits property and constructs desirability functions appropriate to those Spec Limits.

- If you use the Simulator to output simulation tables, JMP copies Spec Limits to the output data tables, making accounting for defect rates and capability indices easy.

- Adding Spec Limits enables a feature called the Defect Profiler.

In the following example, we assume that the following Spec Limits have been specified.

**Table 8.1**  Spec Limits for Tiretread.jmp Data Table

| Response | LSL | USL |
|----------|-----|-----|
| Abrasion | 110 | |
| Modulus | | 2000 |
| Elong | 350 | 550 |
| Hardness | 66 | 74 |

To set these limits in the data table, highlight a column and select **Cols > Column Info**. Then, click the **Column Properties** button and select the **Spec Limits** property.

If you are already in the Simulator in a profiler, another way to enter them is to use the Spec Limits command in the Simulator red triangle menu.

**Figure 8.8**  Spec Limits
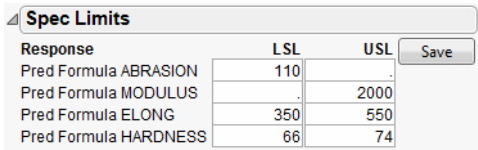


After entering the spec limits, they are incorporated into the profilers. Click the **Save** button if you want the spec limits saved back to the data table as a column property.

With these specification limits, and the distributions shown in Figure 8.2, click the **Simulate** button. Notice the colored spec limit lines in the output histograms. The black dashed lines indicate the mean of the simulations.

**Figure 8.9** Spec Limits in the Prediction Profiler
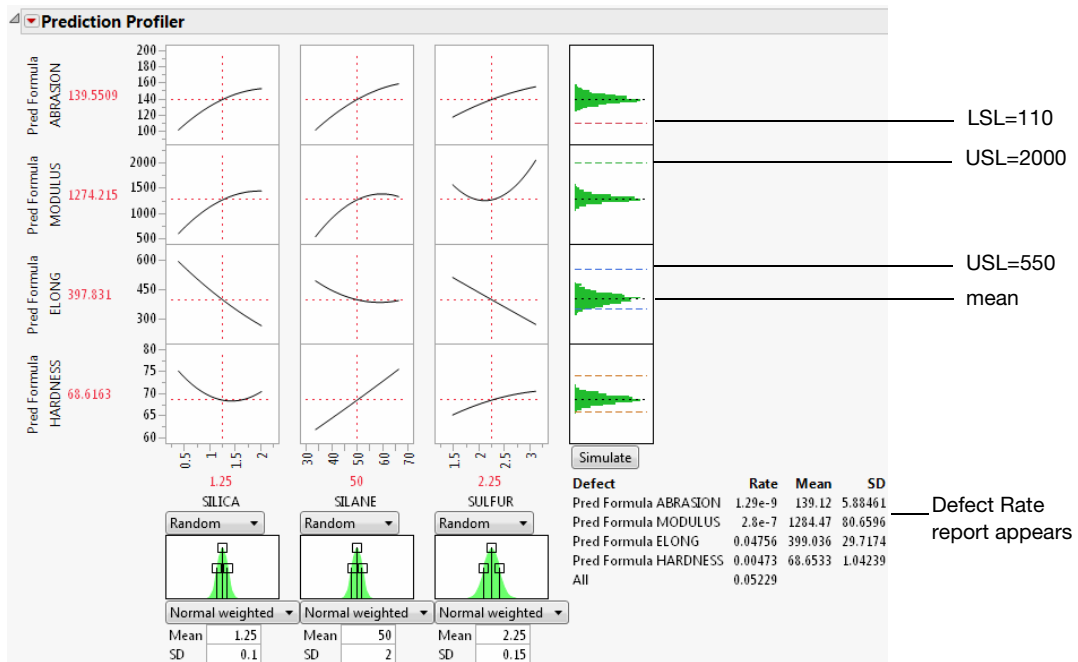


Look at the histogram for Abrasion. The lower spec limit is far below the distribution, yet the Simulator is able to estimate a defect rate for it. This despite only having 5000 runs in the simulation. It can do this rare-event estimation when you use a Normal weighted distribution.

Note that the Overall defect rate (0.05229) is close to the defect rate for ELONG (0.04756), indicating that most of the defects are in the ELONG variable.

To see this weighted simulation in action, click the **Make Table** button and examine the Weight column.

JMP generates extreme values for the later observations, using very small weights to compensate. Because the Distribution platform handles frequencies better than weights, there is also a column of frequencies, which is simply the weights multiplied by $10^{12}$.

The output data set contains a Distribution script appropriate to analyze the simulation data completely with a capability analysis.

## Simulate General Formulas

Though the profiler and simulator are designed to work from formulas stored from a model fit, they work for any formula that can be stored in a column. A typical application of simulation is to exercise financial models under certain probability scenarios to obtain the

distribution of the objectives. This can be done in JMP—the key is to store the formulas into columns, set up ranges, and then conduct the simulation.

**Table 8.2** Factors and Responses for a Financial Simulation

| Inputs (Factors) | Unit Sales | random uniform between 1000 and 2000 |
|---|---|---|
| | Unit Price | fixed |
| | Unit Cost | random normal with mean 2.25 and std dev 0.1 |
| Outputs (Responses) | Revenue | formula: Unit Sales*Unit Price |
| | Total Cost | formula: Unit Sales*Unit Cost + 1200 |
| | Profit | formula: Revenue – Total Cost |

The following JSL script creates the data table below with some initial scaling data and stores formulas into the output variables. It also launches the Profiler.

```
dt = New Table( "Sales Model" );
dt << New Column( "Unit Sales", Values( {1000, 2000} ) );
dt << New Column( "Unit Price", Values( {2, 4} ) );
dt << New Column( "Unit Cost", Values( {2, 2.5} ) );
dt << New Column( "Revenue",
   Formula( :Unit Sales * :Unit Price )
);
dt << New Column( "Total Cost",
   Formula( :Unit Sales * :Unit Cost + 1200 )
);
dt << New Column( "Profit",
   Formula( :Revenue - :Total Cost ),
   Set Property( "Spec Limits", {LSL( 0 )} )
);
Profiler(
   Y( :Revenue, :Total Cost, :Profit ),
   Objective Formula( Profit )
);
```

**Figure 8.10** Data Table Created from Script

Unit Sales*Unit Cost + 1200

Unit Sales*Unit Price

Revenue - Total Cost

| | | Unit Sales | Unit Price | Unit Cost | Revenue | Total Cost | Profit |
|---|---|---|---|---|---|---|---|
| Sales Model | | | | | | | |
| | 1 | 1000 | 2 | 2 | 2000 | 3200 | -1200 |
| | 2 | 2000 | 4 | 2.5 | 8000 | 6200 | 1800 |
| Columns (6/0) | | | | | | | |
| Unit Sales | | | | | | | |
| Unit Price | | | | | | | |
| Unit Cost | | | | | | | |
| Revenue ✚ | | | | | | | |
| Total Cost ✚ | | | | | | | |
| Profit ✚ ✱ | | | | | | | |

Once they are created, select the **Simulator** from the Prediction Profiler. Use the specifications from Figure 8.11 to fill in the Simulator.

**Figure 8.11** Specifications for Profiler

| Unit Sales | Unit Price | Unit Cost |
|---|---|---|
| Random | Fixed | Random |
| | 3 | |
| Uniform | | Normal |
| Lower 1000 | | Mean 2.25 |
| Upper 2000 | | SD 0.1 |

Now, run the simulation, which produces the following histograms in the Profiler.

**Figure 8.12** Simulator



**Note:** Your numbers might differ from those shown in Figure 8.12.

It looks like we are not very likely to be profitable. By putting a lower specification limit of zero on Profit, the defect report would say that the probability of being unprofitable is 62%.

So we raise the Unit Price to $3.25 and rerun the simulation. Now the probability of being unprofitable is down to about 20%.

**Figure 8.13** Results

| Defect | Rate | Mean | SD |
|---|---|---|---|
| Revenue | 0 | 4889.4 | 922.937 |
| Total Cost | 0 | 4587.14 | 658.895 |
| Profit | 0.2006 | 302.252 | 321.969 |
| All | 0.2006 | | |

If unit price cannot be raised anymore, you should now investigate lowering your cost, or increasing sales, if you want to further decrease the probability of being unprofitable.

# Defect Profiler

The defect profiler shows the probability of an out-of-spec output defect as a function of each factor, while the other factors vary randomly. This is used to help visualize which factor's

distributional changes the process is most sensitive to, in the quest to improve quality and decrease cost.

Specification limits define what is a defect, and random factors provide the variation to produce defects in the simulation. Both need to be present for a Defect Profile to be meaningful.

At least one of the Factors must be declared Random for a defect simulation to be meaningful, otherwise the simulation outputs would be constant. These are specified though the simulator Factor specifications.

**Important:** If you need to estimate very small defect rates, use **Normal weighted** instead of just **Normal**. This allows defect rates of just a few parts per million to be estimated well with only a few thousand simulation runs.
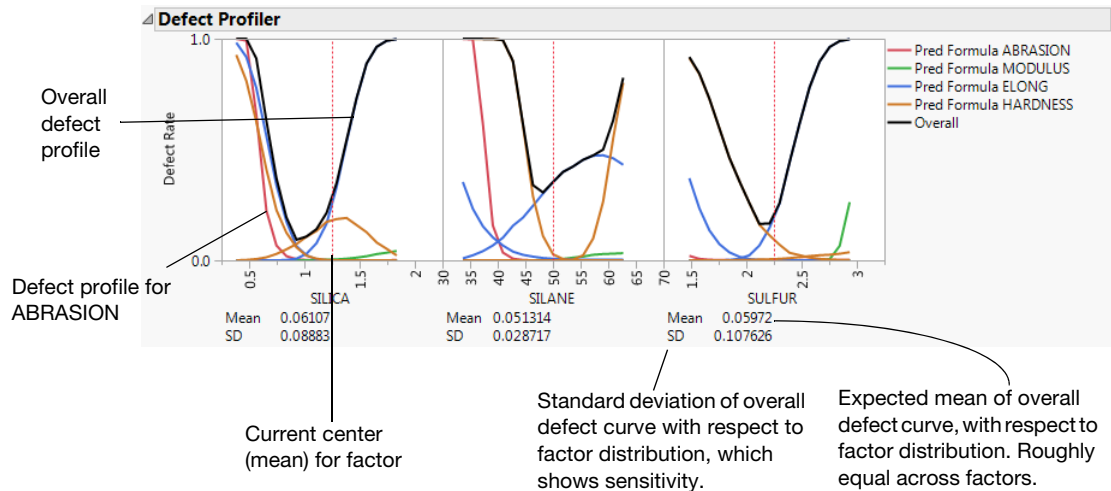
## About Tolerance Design

*Tolerance Design* is the investigation of how defect rates on the outputs can be controlled by controlling variability in the input factors.

The input factors have variation. Specification limits are used to tell the supplier of the input what range of values are acceptable. These input factors then go into a process producing outputs, and the customer of the outputs then judges if these outputs are within an acceptable range.

Sometimes, a Tolerance Design study shows that spec limits on input are unnecessarily tight, and loosening these limits results in cheaper product without a meaningful sacrifice in quality. In these cases, Tolerance Design can save money.

In other cases, a Tolerance Design study might find that either tighter limits or different targets result in higher quality. In all cases, it is valuable to learn which inputs the defect rate in the outputs are most sensitive to.

This graph shows the defect rate as a function of each factor as if it were a constant, but all the other factors varied according to their random specification. If there are multiple outputs with Spec Limits, then there is a defect rate curve color-coded for each output. A black curve shows the overall defect rate—this curve is above all the colored curves.

**Figure 8.14** Defect Profiler



Overall defect profile

Defect profile for ABRASION

Current center (mean) for factor

Standard deviation of overall defect curve with respect to factor distribution, which shows sensitivity.

Expected mean of overall defect curve, with respect to factor distribution. Roughly equal across factors.

## Graph Scale

Defect rates are shown on a cubic root scale, so that small defect rates are shown in some detail even though large defect rates might be visible. A log scale is not used because zero rates are not uncommon and need to be shown.

## Expected Defects

Reported below each defect profile plot is the mean and standard deviation (SD). The mean is the overall defect rate, calculated by integrating the defect profile curve with the specified factor distribution.

In this case, the defect rate that is reported below all the factors is estimating the same quantity, the rate estimated for the overall simulation below the histograms (*that is,* if you clicked the **Simulate** button). Because each estimate of the rate is obtained in a different way, they might be a little different. If they are very different, you might need to use more simulation runs. In addition, check that the range of the factor scale is wide enough so that the integration covers the distribution well.

The standard deviation is a good measure of the sensitivity of the defect rates to the factor. It is quite small if either the factor profile were flat, or the factor distribution has a very small variance. Comparing SD's across factors is a good way to know which factor should get more attention to reducing variation.

The mean and SD are updated when you change the factor distribution. This is one way to explore how to reduce defects as a function of one particular factor at a time. You can click and drag a handle point on the factor distribution, and watch the mean and SD change as you

drag. However, changes are not updated across all factors until you click the **Rerun** button to do another set of simulation runs.

## Simulation Method and Details

Assume we want a defect profile for factor X1, in the presence of random variation in X2 and X3. A series of $n$=N Runs simulation runs is done at each of $k$ points in a grid of equally spaced values of X1. ($k$ is generally set at 17.) At each grid point, suppose that there are $m$ defects due to the specification limits. At that grid point, the defect rate is $m/n$. With normal weighted, these are done in a weighted fashion. These defect rates are connected and plotted as a continuous function of X1.

## Notes

**Recalculation**   The profile curve is not recalculated automatically when distributions change, though the expected value is. It is done this way because the simulations could take a while to run.

**Limited goals**   Profiling does not address the general optimization problem, that of optimizing quality against cost, given functions that represent all aspects of the problem. This more general problem would benefit from a surrogate model and space filling design to explore this space and optimize to it.

**Jagged Defect Profiles**   The defect profiles tend to get uneven when they are low. This is due to exaggerating the differences for low values of the cubic scale. If the overall defect curve (black line) is smooth, and the defect rates are somewhat consistent, then you are probably taking enough runs. If the Black line is jagged and not very low, then increase the number of runs. 20,000 runs is often enough to stabilize the curves.

## Defect Parametric Profiler

The defect parametric profiler shows the impact of process changes on the defect rate. The impacts are based on the simulation parameter settings for each factor. Four scenarios are considered.

**Mean shift**   The impact of shifting the mean is shown by a red curve. The current mean is shown by a red dotted vertical line.

**Std Narrow**   The impact of a reduction in variability is shown by a blue curve. The dotted blue vertical lines are set at the mean plus and minus one standard deviation. The minimum on the curve corresponds to the defect rate when there is no variability.

**LSL Chop**   The impact of inspection to remove all parts below the lower specification limit is shown a green curve.

**USL Chop**   The impact of inspection to remove all parts above the upper specification limit is shown an orange curve.

**Figure 8.15**  Defect Parametric Profile
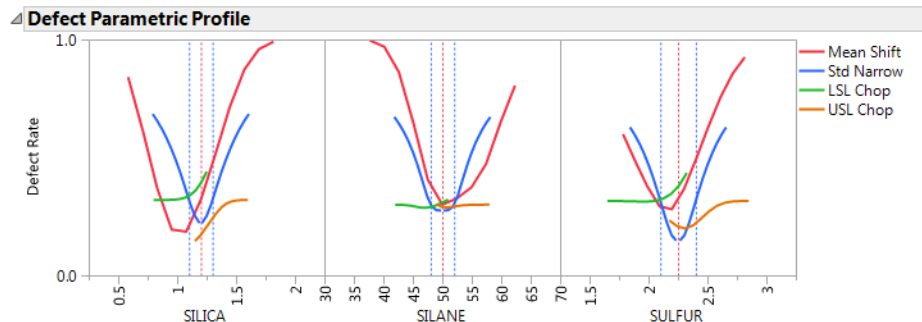


# Simulation Experiment

Used to run a designed simulation experiment on the locations of the factor distributions. A window appears, allowing you to specify the number of design points, the portion of the factor space to be used in the experiment, and which factors to include in the experiment. For factors not included in the experiment, the current value shown in the Profiler is the one used in the experiment.

The experimental design is a Latin Hypercube. The output has one row for each design point. The responses include the defect rate for each response with spec limits, and an overall defect rate. After the experiment, it would be appropriate to fit a Gaussian Process model on the overall defect rate, or a root or a logarithm of it.

A simulation experiment does not sample the factor levels from the specified distributions. As noted above, the design is a Latin Hypercube. At each design point, N Runs random draws are generated with the design point serving as the center of the random draws, and the shape and variability coming from the specified distributions.

# Additional Examples of the Simulator

This section contains additional examples of the Simulator feature in the Prediction Profiler.

## Example of the Defect Profiler

To show a common workflow with the Defect profiler, we use Tiretread.jmp. The experimental data in the Tiretread.jmp sample data table results from an experiment to study the effects of SILICA, SILANE, and SULFUR on four measures of tire tread performance.

1.  Select **Help > Sample Data Library** and open Tiretread.jmp.

2.  Select **Graph > Profiler**.

3.  Select Pred Formula ABRASION, Pred Formula MODULUS, Pred Formula ELONG, and Pred Formula HARDNESS and click **Y, Prediction Formula**.

4.  Click **OK**.

5.  From the Prediction Profiler's red triangle menu, select **Simulator**.

6.  Select **Spec Limits** from the Simulator red triangle menu.

7.  For each response, enter the spec limits as shown in Figure 8.8 on page 136. Click **Save**.

8.  For each factor, enter the random specifications as shown in Figure 8.16.

**Figure 8.16**  Profiler Random Specifications
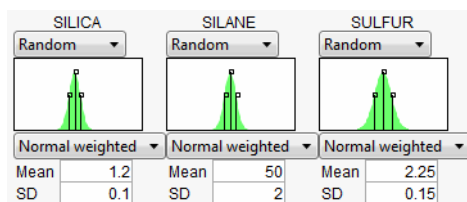


9.  Select **Defect Profiler** from the Simulator red triangle menu to see the defect profiles. The curves, Means, and SDs will change from simulation to simulation, but will be relatively consistent.

**Figure 8.17**  Defect Profiler



---
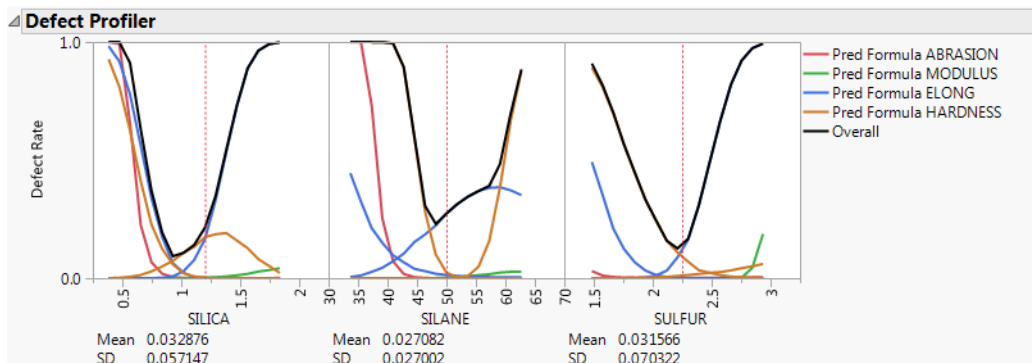
**Note:** Your numbers might differ from those shown in Figure 8.17.

---

The black curve on each factor shows the defect rate if you could fix that factor at the *x*-axis value, but leave the other features random.

Look at the curve for SILICA. As its values vary, its defect rate goes from the lowest 0.001 at SILICA=0.95, quickly up to a defect rate of 1 at SILICA=0.4 or 1.8. However, SILICA is itself random. If you imagine integrating the density curve of SILICA with its defect profile curve, you could estimate the average defect rate 0.033, also shown as the Mean for SILICA. This is estimating the overall defect rate shown under the simulation histograms, but by numerically integrating, rather than by the overall simulation. The Means for the other factors are similar. The numbers are not exactly the same. However, we now also get an estimate of the standard deviation of the defect rate with respect to the variation in SILICA. This value (labeled SD) is 0.057. The standard deviation is intimately related to the sensitivity of the defect rate with respect to the distribution of that factor.

Looking at the SDs across the three factors, we see that the SD for SULFUR is higher than the SD for SILICA, which is in turn much higher than the SD for SILANE. This means that to improve the defect rate, improving the distribution in SULFUR should have the greatest effect. A distribution can be improved in three ways: changing its mean, changing its standard deviation, or by chopping off the distribution by rejecting parts that do not meet certain specification limits.

10. Select **Defect Parametric Profile** from the Simulator red triangle menu. This command shows how single changes in the factor distribution parameters affect the defect rate.

**Figure 8.18**  Defect Parametric Profile
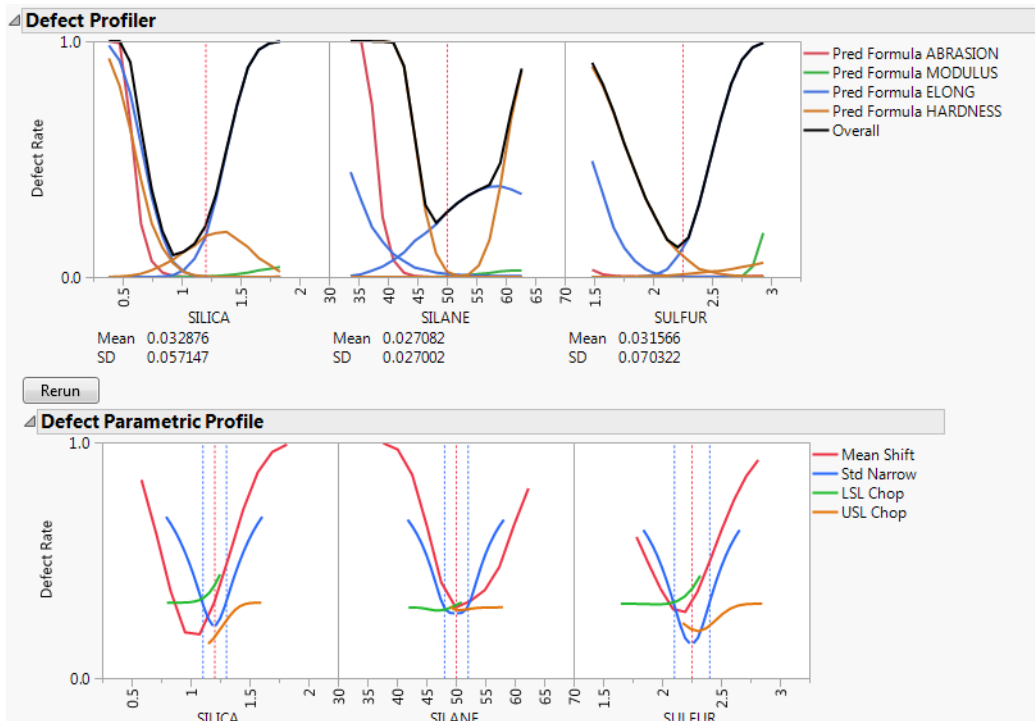
Let's look closely at the SULFUR situation. You might need to enlarge the graph to see more detail.

First, note that the current defect rate (0.03) is represented in four ways corresponding to each of the four curves.

For the red curve, Mean Shift, the current rate is where the red solid line intersects the vertical red dotted line. The Mean Shift curve represents the change in overall defect rate by changing the mean of SULFUR. One opportunity to reduce the defect rate is to shift the mean slightly to the left. If you use the crosshair tool on this plot, you see that a mean shift reduces the defect rate to about 0.02.

For the blue curve, Std Narrow, the current rate represents where the solid blue line intersects the two dotted blue lines. The Std Narrow curves represent the change in defect rate by changing the standard deviation of the factor. The dotted blue lines represent one standard deviation below and above the current mean. The solid blue lines are drawn symmetrically around the center. At the center, the blue line typically reaches a minimum, representing the defect rate for a standard deviation of zero. That is, if we totally eliminate variation in SULFUR, the defect rate is about 0.003. This is much better than 0.03. If you look at the other Defect parametric profile curves, you can see that this is better than reducing variation in the other factors, something that we suspected by the SD value for SULFUR.

For the green curve, LSL Chop, there are no interesting opportunities in this example, because the green curve is above current defect rates for the whole curve. This means that reducing the variation by rejecting parts with too-small values for SULFUR will not help.

For the orange curve, USL Chop, there are good opportunities. Reading the curve from the right, the curve starts out at the current defect rate (0.03), then as you start rejecting more parts by decreasing the USL for SULFUR, the defect rate improves. However, moving a spec limit to the center is equivalent to throwing away half the parts, which might not be a practical solution.

Looking at all the opportunities over all the factors, it now looks like there are two good options for a first move: change the mean of SILICA to about 1, or reduce the variation in SULFUR. Because it is generally easier in practice to change a process mean than process variation, the best first move is to change the mean of SILICA to 1.
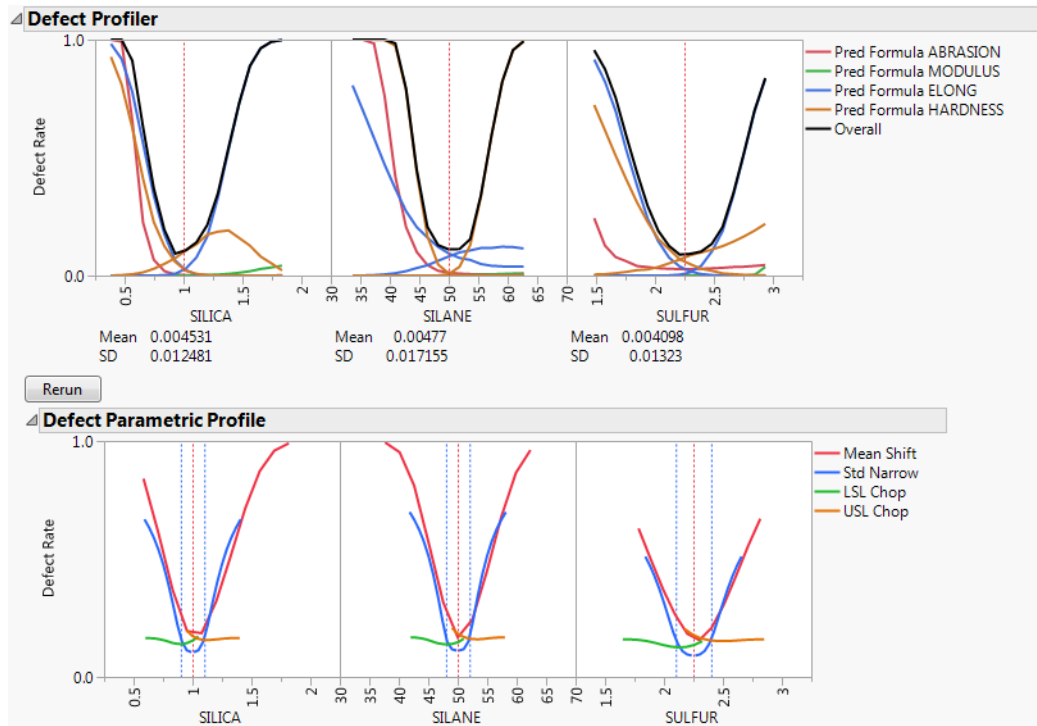
11. Adjust the Mean of SILICA from 1.2 to 1.0. Click **Rerun**.

**Figure 8.19**  Adjusting the Mean of SILICA
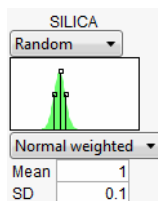
After clicking **Rerun**, we get a new perspective on defect rates.

**Figure 8.20**  Adjusted Defect Rates



Now, the defect rate is down to about 0.004, much improved. Further reduction in the defect rate can occur by continued investigation of the parametric profiles, making changes to the distributions, and rerunning the simulations.

As the defect rate is decreased further, the mean defect rates across the factors become relatively less reliable. The accuracy could be improved by reducing the ranges of the factors in the Profiler a little so that it integrates the distributions better.

This level of fine-tuning is probably not practical, because the experiment that estimated the response surface is probably not at this high level of accuracy. Once the ranges have been refined, you might need to conduct another experiment focusing on the area that you know is closer to the optimum.

## Example of Stochastic Optimization

This example is adapted from Box and Draper (1987) and uses Stochastic Optimization.jmp. A chemical reaction converts chemical "A" into chemical "B". The resulting amount of chemical "B" is a function of reaction time and reaction temperature. A longer time and hotter

temperature result in a greater amount of "B". But, a longer time and hotter temperature also result in some of chemical "B" getting converted to a third chemical "C". What reaction time and reaction temperature will maximize the resulting amount of "B" and minimize the amount of "A" and "C"? Should the reaction be fast and hot, or slow and cool?

**Figure 8.21** Chemical Reaction
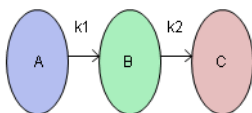


The goal is to maximize the resulting amount of chemical "B". One approach is to conduct an experiment and fit a response surface model for reaction yield (amount of chemical "B") as a function of time and temperature. But, due to well known chemical reaction models, based on the Arrhenius laws, the reaction yield can be directly computed. The column Yield contains the formula for yield. The formula is a function of Reaction Time (hours) and reaction rates k1 and k2. The reaction rates are a function of Reaction Temperature (degrees Kelvin) and known physical constants $\theta_1$, $\theta_2$, $\theta_3$, $\theta_4$. Therefore, Yield is a function of Reaction Time and Reaction Temperature.

1. Select **Help > Sample Data Library** and open Stochastic Optimization.jmp
2. Select **Graph > Profiler**.
3. Assign Yield to **Y, Prediction Formula**.
4. Click **Expand Intermediate Formulas**. Click **OK**.

    The Prediction Profiler with Desirability Functions enabled appears. For more information on Desirability Functions, see the "Desirability Profiling and Optimization" on page 47 in the "Profiler" chapter.

5. From the Prediction Profiler red triangle menu, select **Optimization and Desirability > Maximize Desirability**.

    The Profiler maximizes Yield and sets the graphs to the optimum value of Reaction Time and Reaction Temperature.

**Figure 8.22** Yield Maximum
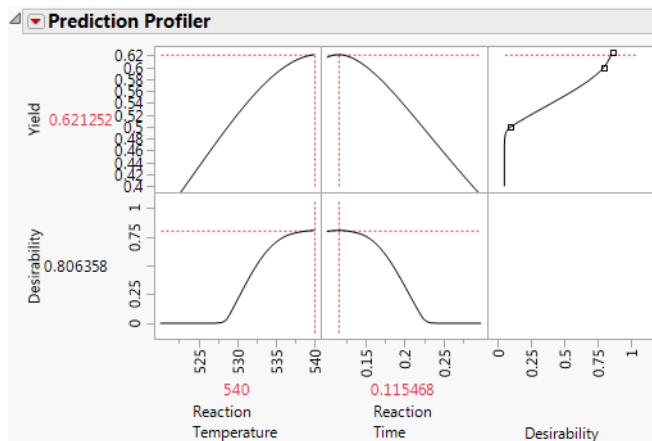


The maximum Yield is approximately 0.62 at a Reaction Time of 0.115 hours and Reaction Temperature of 540 degrees Kelvin, or hot and fast. (Your results might differ slightly due to random starting values in the optimization process.)

In a production environment, process inputs cannot always be controlled exactly. What happens to Yield if the inputs (Reaction Time and Reaction Temperature) have random variation? Furthermore, if Yield has a spec limit, what percent of batches will be out of spec and need to be discarded? The Simulator can help us investigate the variation and defect rate for Yield, given variation in Reaction Time and Reaction Temperature.

6. Deselect **Desirability Functions** from the Prediction Profiler red triangle menu.

7. Select **Simulator** from the Prediction Profiler red triangle menu.

8. As shown in Figure 8.23, fill in the factor parameters so that Temperature is **Normal weighted** with standard deviation of 1, and Time is **Normal weighted** with standard deviation of 0.03. The Mean parameters default to the current factor values.

9. Change the number of runs to 15,000.

**Figure 8.23** Initial Simulator Settings
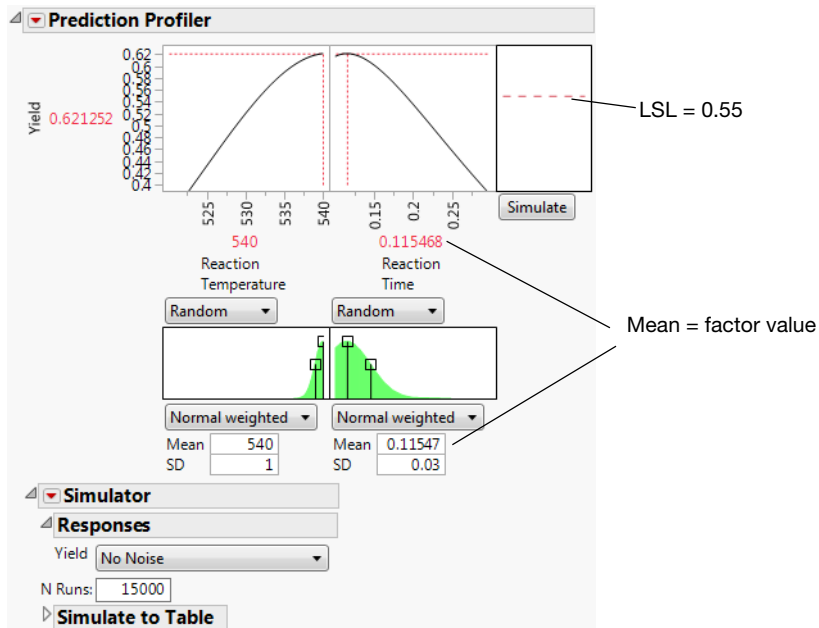


Yield has a lower spec limit of 0.55, set as a column property, and shows in Figure 8.23 as a red line. With the random variation set for the input factors, you are ready to run a simulation to study the resulting variation and defect rate for Yield.

10. Click the **Simulate** button.

**Figure 8.24** Simulation Results
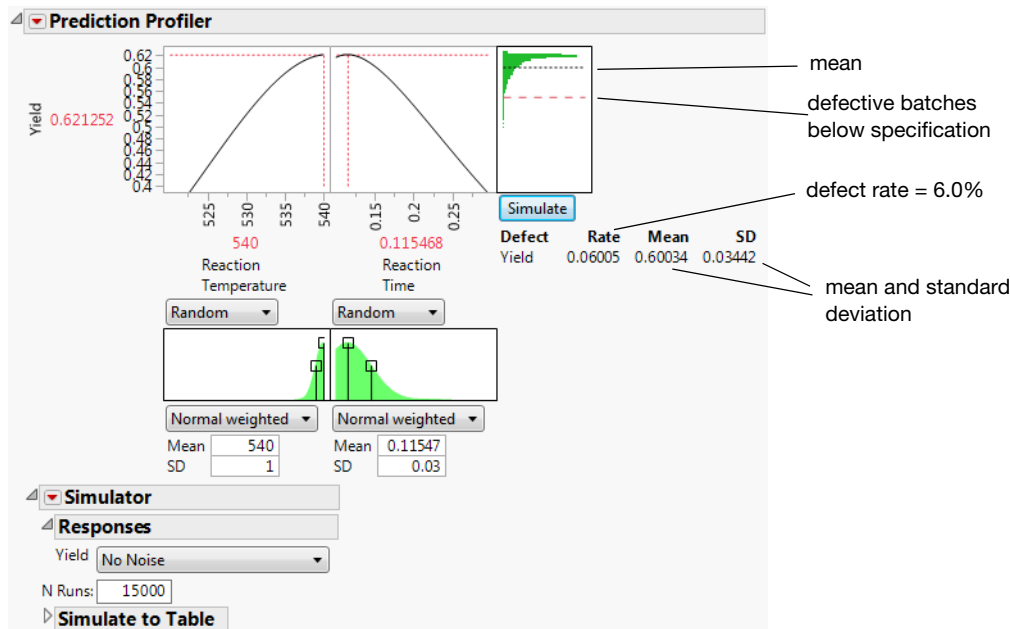


As shown in Figure 8.24, the predicted Yield is 0.62, but if the factors have the given variation, the average Yield is 0.60 with a standard deviation of 0.03.

The defect rate is about 6.0%, meaning that about 6.0% of batches are discarded. A defect rate this high is usually not acceptable.

What is the defect rate for other settings of Reaction Temperature and Reaction Time? Suppose you change the Reaction Temperature to 535, then set Reaction Time to the value that maximizes Yield?

Before changing settings, we should save these factor settings that give us the maximum Yield.

11. From the Prediction Profiler red triangle menu, select **Factor Settings > Remember Settings**.

12. Type "Max Yield" and click **OK**.

The settings are appended to the report window.

**Figure 8.25** Remembered Settings for Maximum Yield



13. Now, change the Mean value for Reaction Temperature to 535.

14. Move the dashed red line in the Reaction Time plot to the value that maximizes Yield (around 0.16). Change the Mean value of Reaction Time to 0.16.

15. Click **Simulate**.

**Figure 8.26**  Defect Rate for Temperature of 535
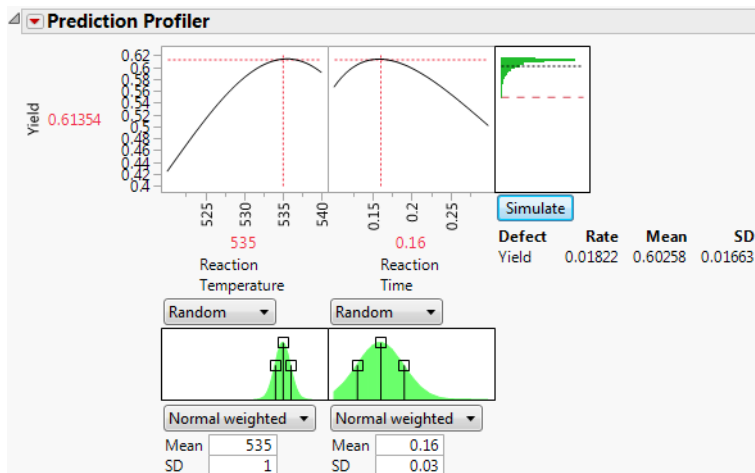


As illustrated in Figure 8.26, the defect rate decreases to about 1.8%, which is much better than 6.0%. So, what you see is that the fixed (no variability) settings that maximize Yield are not the same settings that minimize the defect rate in the presence of factor variation.

By running a Simulation Experiment you can find the settings of Reaction Temperature and Reaction Time that minimize the defect rate. To do this you simulate the defect rate at each point of a Reaction Temperature and Reaction Time design, then fit a predictive model for the defect rate and minimize it.

16. Select **Simulation Experiment** from the Simulator red triangle menu.

17. Enter 80 runs, and 1 to use the whole factor space in the experiment. Click **OK**.

A Latin Hypercube design with 80 design points is chosen within the specified factor space, and N Runs random draws are taken at each of the design points. The design points are the center of the random draws, and the shape and variance of the random draws coming from the factor distributions.

A table is created with the results of the experiment. The Overall Defect Rate is given at each design point. You can now fit a model that predicts the defect rate as a function of Reaction Temperature and Reaction Time.

Do not close the Stochastic Optimization Profiler window.

18. From the new table, run the attached Gaussian Process script.

The results are shown in Figure 8.27. Your results will be slightly different due to the random draws in the simulation. The Gaussian Process platform automatically opens the Prediction Profiler.

**Figure 8.27**  Results of Gaussian Process Model Fit



Model Report

| Column | Theta | Total Sensitivity | Main Effect | Reaction Temperature Interaction | Reaction Time Interaction |
|---|---|---|---|---|---|
| Reaction Temperature | 0.0985232 | 0.9025854 | 0.2229226 | . | 0.6796628 |
| Reaction Time | 9.4040305 | 0.7770774 | 0.0974146 | 0.6796628 | . |

| μ | σ² | Nugget |
|---|---|---|
| -0.689765 | 0.4038852 | 0 |

-2*LogLikelihood
-71.19388

Fit using the Cubic correlation function.
Nugget estimated via maximum likelihood.

19. To find the settings of Temperature and Time that minimizes the defect rate, select **Optimization and Desirability > Maximize Desirability** from the Prediction Profiler red triangle menu.

The desirability function is already set up to minimize the defect rate.

**Figure 8.28**  Settings for Minimum Defect Rate
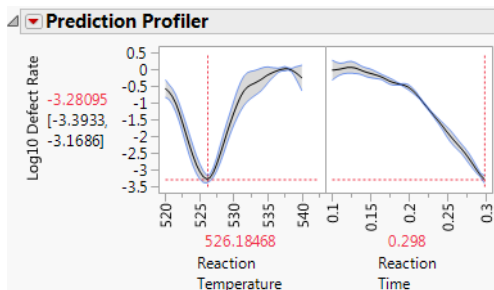


The settings that minimize the defect rate are approximately Reaction Temperature = 526 and Reaction Time = 0.3.

20. Click the **Transfer Factor Settings Back** button.

   This sets Temperature and Time to those settings that minimize the defect rate in the original Profiler report window.

21. Return to the original Profiler report window.

22. From the Prediction Profiler red triangle menu, select **Factor Settings > Remember Settings**.

23. Type "Min Defect" and click **OK**.

**Figure 8.29**  Minimum Defect Settings
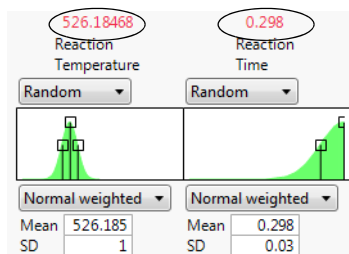


24. With the new settings in place, click the **Simulate** button to estimate the defect rate at the new settings.

**Figure 8.30**  Lower Defect Rate
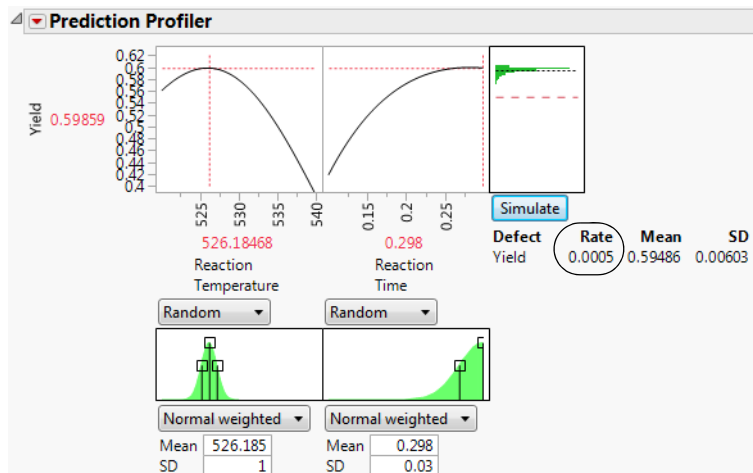


At the new settings the defect rate is 0.05%, much better than the 6.0% for the settings that maximize Yield. That is a reduction of about 120x. Recall the average Yield from the first settings is 0.60 and the new average is 0.59. The decrease in average Yield of 0.01 is very acceptable when the defect rate decreases by 120x.

Because we saved the settings using Remember Settings, we can easily compare the old and new settings. The Differences report summarizes the difference.

25. Click the **Remembered Settings** radio buttons to view the profiler for each setting.

**Figure 8.31** Settings Comparison
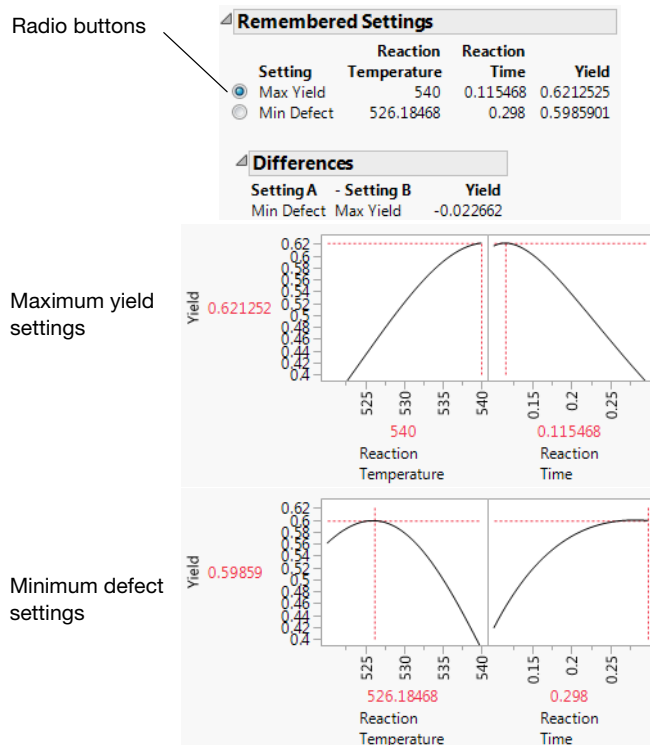


The chemist now knows what settings to use for a quality process. If the factors have no variation, the settings for maximum Yield are hot and fast. But, if the process inputs have variation similar to what we have simulated, the settings for maximum Yield produce a high defect rate. Therefore, to minimize the defect rate in the presence of factor variation, the settings should be cool and slow.

## Statistical Details for the Simulator

This section contains statistical details for the Simulator profiler.

### Normal Weighted Distribution

JMP uses the multivariate radial strata method for each factor that uses the **Normal Weighted** distribution. This seems to work better than a number of Importance Sampling methods, as a multivariate Normal Integrator accurate in the extreme tails.

First, define strata and calculate corresponding probabilities and weights. For $d$ random factors, the strata are radial intervals as follows.

**Table 8.3** Strata Intervals

| Strata Number | Inside Distance | Outside Distance |
|:---:|:---:|:---:|
| 0 | 0 | $\sqrt{d}$ |
| 1 | $\sqrt{d}$ | $\sqrt{d + \sqrt{2d}}$ |
| 2 | $\sqrt{d + \sqrt{2d}}$ | $\sqrt{d + 2\sqrt{2d}}$ |
| $i$ | $\sqrt{d + (i-1)\sqrt{2d}}$ | $\sqrt{d + i\sqrt{2d}}$ |
| $N_{Strata} - 1$ | previous | $\infty$ |

The default number of strata is 12. To change the number of strata, a hidden command N Strata is available if you hold the Shift key down while clicking on the red triangle next to Simulator. Increase the sample size as needed to maintain an even number of strata.

For each simulation run:

1. Select a strata as $\mathrm{mod}(i - 1, N_{Strata})$ for run $i$.

2. Determine a random $n$-dimensional direction by scaling multivariate Normal (0,1) deviates to unit norm.

3. Determine a random distance using a chi-square quantile appropriate for the strata of a random uniform argument.

4. Scale the variates so that the norm is the random distance.

5. Scale and re-center the variates individually to be as specified for each factor.

The resulting factor distributions are multivariate normal with the appropriate means and standard deviations when estimated with the right weights. Note that you cannot use the Distribution standard deviation with weights, because it does not estimate the desired value. However, multiplying the weight by a large value, like $10^{12}$, and using that as a Freq value results in the correct standard deviation.

# Excel Profiler
## Visualize Models Saved in Microsoft Excel

The JMP Add-In for Excel uses the JMP Profiler to visualize models (or formulas) stored in Excel worksheets. You can install the Excel add-in when you install JMP.

**Figure 9.1** Example of a Profiler Using an Excel Model
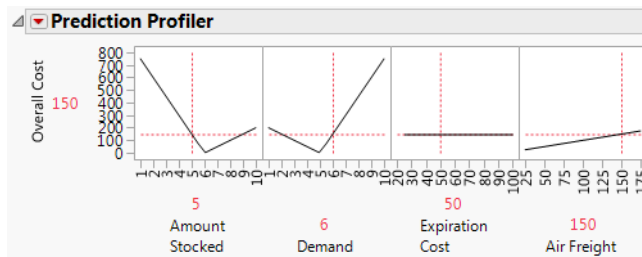
# Excel Profiler Overview

The JMP Add-In for Excel uses the JMP Profiler to visualize models (or formulas) stored in Excel worksheets. The Excel add-in is automatically installed when you install JMP. Profiling in the Excel Add-In is a two-step process:

1. Click the **Create/Edit Model** button (Excel 2007 through 2016) to enter information about the model that JMP needs. This needs to be done only once per model. For more information, click **Help** in the Create/Edit Model window.

2. Click the **Run Model** button (Excel 2007 through 2016) to launch the JMP Profiler and run the Excel model. For more information, see "Run the JMP Profiler" on page 162.

**Notes:**

- The Preferences, Data Table, Graph Builder, and Distribution buttons are not needed to profile an Excel model. For more information about these features, see the Import Your Data chapter in the *Using JMP* book.

- A JMP ribbon is added to Microsoft Excel when the add-in is installed. If there is no JMP ribbon, install the add-in by double-clicking the JMP setup.exe file, selecting **Modify**, selecting **Excel Add-In**, and clicking **Next**.

# Example of an Excel Model

An Excel model consists of one or more Excel formulas. Each formula must be a function of one or more other cells. This example uses the Demand.xls sample import data located in the Samples/Import Data folder.

**Figure 9.2** Demand Model in Excel

**About the Demand.xls Sample Import Data**

The formula in cell B8 is a calculation of the Overall Cost associated with having different amounts of product in stock. The formula, which is shown in the Formula Bar, is a function of four cells:

- Amount Stocked is the amount of product in stock.
- Demand is the customer demand for the product.
- Air Freight is the cost per unit to ship additional product by air when the demand exceeds the amount in stock.
- Expiration Cost is the cost per unit of disposing of unused product when the demand is less than the amount in stock.

The calculations of the formula are as follows:

- If Amount Stocked is less than Demand, then the company has to ship additional units, at a cost of (Demand-Amount Stocked) × Air Freight. For example, if the demand is 8, but the company has only 6 in stock, then it has to ship 8 - 6 = 2 units at a cost of 2 × 150 = 300.
- If Amount Stocked is greater than Demand, then the company has to dispose of unused product, at a cost of (Amount Stocked - Demand) × Expiration Cost. For example, if the demand is 5, but the company has 8 in stock, then it has to dispose of 8 - 5 = 3 units at a cost of 3 × 50 = 150.
- If Amount Stocked is equal to Demand, then there is no shipping cost or disposal cost.
- There is never both a shipping cost and a disposal cost at the same time.

**Create the Model in Excel**

1. Select **Help > Sample Data Library** and navigate up one level to the Sample Import Data folder.
2. Double-click Demand.xls to open the file in Microsoft Excel.
3. In Microsoft Excel, click the JMP ribbon.
4. Click the **Create/Edit Model** button .

   The name of the worksheet is displayed in the Model and Model Name fields.

   The Inputs and Outputs fields are populated with data from the worksheet.
5. Enter Customer Demand in the Model Name field and click **Apply**.

   The Model field is updated.
6. Select **Air Freight** in the Inputs box and then click the down arrow button .

   Air Freight is moved to the bottom of the list because you want it displayed last in the Profiler.
7. Click **OK**.

The Excel model is saved to the worksheet.

**Notes:**

- If the fields in the Create/Edit Model window are not populated when you set up your model, click **Choose** and select the cell that contains the input name. Specify the values and click **Apply**.

- You can create more than one model for a worksheet. In the Create/Edit Model window, click the plus button ⊞ next to the Model name. In the Model Name field, change the name if necessary and click the **Apply** button. You can then change the inputs and outputs as necessary and click **OK**.

- You must define the entire model on one worksheet. A model cannot reference cells on another worksheet.

# Run the JMP Profiler

Once you create the model using the Excel Add-In, you can run it in the JMP Profiler. From the Excel Add-In, perform the following actions:

1. In Microsoft Excel, click the JMP ribbon.
2. Click the **Run Model** button.
3. Select the model that you want to run.
4. Click **Profile in JMP**.
5. Use the JMP Profiler to simultaneously see the effect of all inputs on the output. You can also simulate a range of input combinations to see the resulting range of output values.

**Figure 9.3** Example of the Profiler Using Excel Models
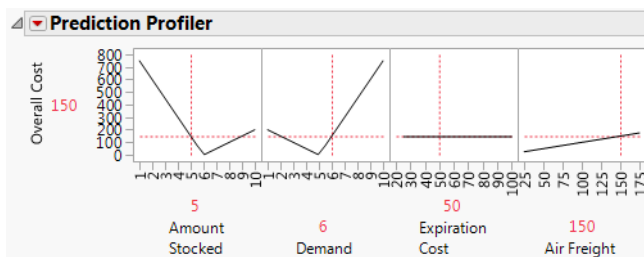


**Note:** To ensure that your original Excel worksheet is not altered, JMP runs a hidden copy of Excel in the background that controls all of the Profiler calculations.

## Use Linear Constraints

Within the JMP Profiler, you can alter the linear constraints in order to restrict the model input values. You are prompted to save the constraints to the Excel workbook. After constraints are saved to the Excel workbook, whenever the model is profiled from the Excel Add-In, the constraints are incorporated.

1. From the red triangle menu next to Prediction Profiler, select **Alter Linear Constraints**.

2. Click **Add Constraint**.

3. Type in the constraining values.

4. Click **OK**.

5. From the red triangle menu next to Prediction Profiler, select **Save Linear Constraints**.

   You are prompted to save the constraints to the Excel workbook.

6. Click **Yes**.

**Note:** When you save the .xls file, you might see a compatibility error. If so, click **Continue** to save the file.

The workbook opens in Excel. When you run the model, the constraints are reflected in the JMP Profiler. For more information about linear constraints, see "Linear Constraints" on page 35 in the "Introduction to Profilers" chapter.

**Tip:** To delete a linear constraint, set all constraint values to zero.

## Resolution of Profile Lines

The Default N Levels option on the red triangle menu next to Prediction Profiler affects the resolution of the profile lines. Note the following information:

• This option defaults to 17 when the Profiler runs a model stored in Excel.

• This option defaults to 41 when the model is stored directly in JMP.

If the same model is stored in both Excel and JMP, then the profile lines can appear differently when the models are profiled. Increasing this value causes the Excel Profiler to run slower.

# Use the Excel Profiler from JMP

After you have defined model input and outputs in an Excel file, you can profile the model from within JMP.

1. Select **Graph** > **Excel Profiler**.

2. Locate the Excel file containing the model and then click **Open**.

3. If the Excel file contains multiple models, you are prompted to select the model that you want to profile.

Note that the Excel Profiler is also scriptable, as follows:

```
Excel Profiler( "path to workbook", <"model name"> ) ;
```

If more than one model exists, and no model is specified, a window with the list of available models appears. For more information about scripting the Excel Profiler, see the Extending JMP chapter in the *Scripting Guide*.

# Appendix **A**

# **References**

Box, G.E.P. and Draper, N.R. (1987), *Empirical Model–Building and Response Surfaces*, New York: John Wiley and Sons.

Cornell, J.A. (1990), *Experiments with Mixtures*, Second Edition, New York: John Wiley and Sons.

Derringer, D. and Suich, R. (1980), "Simultaneous Optimization of Several Response Variables," *Journal of Quality Technology*, 12:4, 214-219.

Hastie, T., Tibshirani, R., and Friedman, J.H.(2009), *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, New York: Springer Science and Business Media.

Saltelli, A. (2002), "Making best use of model evaluations to compute sensitivity indices," *Computer Physics Communications*, 145, 280-297.

Sobol, I.M. (1993), "Sensitivity Estimates for Nonlinear Mathematical Models," *MMCE*, 1:4, 407-414.