Version 17

# Profilers

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

Marcel Proust

JMP Statistical Discovery LLC
SAS Campus Drive
Cary, North Carolina 27513-2414

17.1

The correct bibliographic citation for this manual is as follows: JMP Statistical Discovery LLC 2022–2023. JMP® 17 *Profilers*. Cary, NC: JMP Statistical Discovery LLC

**JMP® 17 Profilers**

Copyright © 2022–2023, JMP Statistical Discovery LLC, Cary, NC, USA

All rights reserved. Produced in the United States of America.

JMP Statistical Discovery LLC, SAS Campus Drive, Cary, North Carolina 27513-2414.

October 2022

March 2023

JMP® and all other JMP Statistical Discovery LLC product or service names are registered trademarks or trademarks of JMP Statistical Discovery LLC in the USA and other countries. ® indicates USA registration.

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

JMP 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 JMP software, refer to http://support.sas.com/thirdpartylicenses.

# 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
- the JMP user community, resources for users including examples of add-ins and scripts, a forum, blogs, conference information, and so on

**jmp.com/getstarted**

# Contents

## Profilers


1 Learn about JMP ..... 11

Documentation and Additional Resources

Formatting Conventions in JMP Documentation ..... 13
JMP Help ..... 14
JMP Documentation Library ..... 14
Additional Resources for Learning JMP ..... 21
Search JMP ..... 21
JMP Tutorials ..... 21
Sample Data Tables ..... 22
Learn about Statistical and JSL Terms ..... 22
Learn JMP Tips and Tricks ..... 22
JMP Tooltips ..... 22
JMP User Community ..... 23
Free Online Statistical Thinking Course ..... 23
JMP New User Welcome Kit ..... 23
Statistics Knowledge Portal ..... 23
JMP Training ..... 24
JMP Books by Users ..... 24
The JMP Starter Window ..... 24
JMP Technical Support ..... 24

2 Introduction to Profilers ..... 25

Visualize Response Surfaces and Optimize Processes

Introduction to Profiling ..... 27
Profiler Features in JMP ..... 27
Profiler Launch Windows ..... 29
Profilers in Fit Group Reports ..... 31
Interpret Prediction Profiler ..... 31
The Profiler as a Cross-Section ..... 32
Set or Lock Factor Values ..... 34
Profiler Platform Options ..... 35
Common Profiler Topics ..... 36
Linear Constraints ..... 36

Noise Factors . . . 38

3 Profiler . . . 41
Explore Cross Sections of Responses across Each Factor

Overview of the Prediction Profiler . . . 43
Example of the Prediction Profiler . . . 44
Launch the Prediction Profiler Platform . . . 45
Prediction Profiler Options . . . 46
Maximization Options Window . . . 55
Desirability Profiling and Optimization . . . 56
Construction of Desirability Functions . . . 57
How to Use the Desirability Function . . . 58
The Desirability Profile . . . 60
Customized Desirability Functions . . . 60
Assess Variable Importance . . . 62
Bagging . . . 66
Design Space Profiler . . . 67
Design Space Profiler Report . . . 68
Design Space Profiler Options . . . 68
Additional Examples of the Prediction Profiler . . . 70
Example of a Desirability Profiler for Multiple Responses . . . 70
Example of a Noise Factor in the Prediction Profiler . . . 72
Example of Variable Importance for One Response . . . 78
Example of Variable Importance for Multiple Responses . . . 80
Example of Bagging to Improve Prediction . . . 82
Example of Bagging to Indicate the Accuracy of Predictions . . . 86
Example of Confidence Intervals Using Bagging . . . 89
Example of Confidence Intervals Using StdErr Pred Formula Column . . . 91
Example of Design Space Profiler . . . 92
Statistical Details for the Prediction Profiler . . . 94
Statistical Details for Assess Variable Importance . . . 94
Statistical Details for Propagation of Error Bars . . . 96
Statistical Details for Extrapolation Control Metrics . . . 97

4 Contour Profiler . . . 101
Explore Contours of Responses across Two Factors

Overview of the Contour Profiler . . . 103
Example of the Contour Profiler . . . 103
Launch the Contour Profiler Platform . . . 105
The Contour Profiler Report . . . 105
Factor Settings and Controls . . . 106

Response Settings and Controls . . . 106
Contour Profiler Platform Options . . . 107
Constraint Shading Settings . . . 109
Additional Example of the Contour Profiler . . . 109

5 Surface Plot . . . 113
Explore Contours of Responses across Three Factors
Overview of the Surface Plot Profiler . . . 115
Example of the Surface Plot . . . 115
Launch the Surface Plot . . . 118
The Surface Plot Report . . . 120
The Surface Plot . . . 120
Appearance Controls . . . 121
Independent Variables . . . 122
Dependent Variables . . . 123
Surface Plot Platform Options . . . 125
Right-Click Menu Options . . . 125
Dependent Variables Options . . . 126
Sheet or Surface Properties . . . 127
Keyboard Shortcuts . . . 129
Additional Examples of the Surface Plot . . . 129
Example of a Surface Plot for a Single Mathematical Function . . . 129
Example of Isosurfaces . . . 130

6 Mixture Profiler . . . 133
Explore Factor Effects Using Ternary Plots
Overview of the Mixture Profiler . . . 135
Ternary Plot Overview . . . 135
Ternary Plots with More Than Three Components . . . 136
Example of the Mixture Profiler . . . 137
Launch the Mixture Profiler Platform . . . 138
The Mixture Profiler Report . . . 139
Factor Settings and Controls . . . 139
Response Settings and Controls . . . 140
Mixture Profiler Platform Options . . . 140
Customizations for the Mixture Profiler . . . 141
Linear Constraints . . . 141
Additional Examples of the Mixture Profiler . . . 142
Example of Mixture Variables with a Process Variable . . . 142
Example of Multiple Responses with Five Mixture Variables . . . 145

7 Custom Profiler ..... 151
Explore Response Surfaces Using a Numerical Calculator

Overview of the Custom Profiler ..... 153
Example of the Custom Profiler ..... 153
Launch the Custom Profiler Platform ..... 155
The Custom Profiler Report ..... 156
Factor Settings and Controls ..... 156
Response Settings and Controls ..... 157
Optimization Settings and Controls ..... 157
Custom Profiler Platform Options ..... 158

8 Simulator ..... 159
Explore the Effects of Variation on Responses

Overview of the Simulator ..... 161
Example of the Simulator ..... 161
Launch the Simulator ..... 164
Simulation Settings for Factors ..... 164
Simulation Settings for Responses ..... 167
The Simulator Report ..... 168
The Simulator Report Options ..... 169
Defect Profiler ..... 170
Defect Parametric Profile ..... 172
Simulation Experiment ..... 173
Specification Limits ..... 174
Additional Examples of the Simulator ..... 174
Example of the Defect Profiler ..... 174
Example of Stochastic Optimization ..... 178
Example of Simulating General Formulas ..... 187
Statistical Details for the Simulator ..... 190

9 Excel Profiler ..... 193
Visualize Models Saved in Microsoft Excel

Overview of the Excel Profiler ..... 195
Example of an Excel Model ..... 195
Run the Excel Profiler ..... 197
Use Linear Constraints ..... 198
Resolution of Profile Lines ..... 198
Use the Excel Profiler from JMP ..... 199

**A** **References** ..... 201
**B** **Technology License Notices** ..... 203

Chapter 1

# Learn about JMP

## Documentation and Additional Resources

Learn about JMP documentation, such as book conventions, descriptions of each JMP document, the Help system, and where to find additional support.

## Contents

Formatting Conventions in JMP Documentation . . . 13
JMP Help . . . 14
JMP Documentation Library . . . 14
Additional Resources for Learning JMP . . . 21
Search JMP . . . 21
JMP Tutorials . . . 21
Sample Data Tables . . . 22
Learn about Statistical and JSL Terms . . . 22
Learn JMP Tips and Tricks . . . 22
JMP Tooltips . . . 22
JMP User Community . . . 23
Free Online Statistical Thinking Course . . . 23
JMP New User Welcome Kit . . . 23
Statistics Knowledge Portal . . . 23
JMP Training . . . 24
JMP Books by Users . . . 24
The JMP Starter Window . . . 24
JMP Technical Support . . . 24

# Formatting Conventions in JMP Documentation

These 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 (or sans-serif online) font.
- Code appears in `Lucida Sans Typewriter` (or monospace online) font.
- Code output appears in *`Lucida Sans Typewriter`* italic (or monospace italic online) font and is indented farther than the preceding code.
- **Helvetica bold** formatting (or bold sans-serif online) 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
- Features that are for JMP Pro only are noted with the JMP Pro icon JMP PRO. For an overview of JMP Pro features, visit jmp.com/software/pro.

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

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

# JMP Help

JMP Help in the Help menu enables you to search for information about JMP features, statistical methods, and the JMP Scripting Language (or *JSL*). You can open JMP Help in several ways:

- Search and view JMP Help on Windows by selecting **Help > JMP Help**.
- On Windows, press the F1 key to open the Help system in the default browser.
- 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.

**Note:** The JMP Help is available for users with Internet connections. Users without an Internet connection can search all books in a PDF file by selecting **Help > JMP Documentation Library**. See "JMP Documentation Library" for more information.

# JMP Documentation Library

The Help system content is also available in one PDF file called *JMP Documentation Library*. Select **Help > JMP Documentation Library** to open the file. You can also download the Documentation PDF Files add-in if you prefer searching individual PDF files of each document in the JMP library. Download the available add-ins from community.jmp.com.

The following table describes the purpose and content of each document 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. Also learn how to share your results. |
| *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. |

| Document Title | Document Purpose | Document Content |
|---|---|---|
| *Basic Analysis* | Perform basic analysis using this document. | Describes these Analyze menu platforms:<br>• Distribution<br>• Fit Y by X<br>• Tabulate<br>• Text Explorer<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. |
| *Essential Graphing* | Find the ideal graph for your data. | Describes these Graph menu platforms:<br>• Graph Builder<br>• Scatterplot 3D<br>• Contour Plot<br>• Bubble Plot<br>• Parallel Plot<br>• Cell Plot<br>• Scatterplot Matrix<br>• Ternary Plot<br>• Treemap<br>• Chart<br>• Overlay Plot<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. |

| Document Title | Document Purpose | Document Content |
|---|---|---|
| *Design of Experiments Guide* | Learn how to design experiments and determine appropriate sample sizes. | Covers all topics in the DOE menu. |
| *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:<br>• Standard Least Squares<br>• Stepwise<br>• Generalized Regression<br>• Mixed Model<br>• Generalized Linear Mixed Model<br>• MANOVA<br>• Loglinear Variance<br>• Nominal Logistic<br>• Ordinal Logistic<br>• 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:<br>• Neural<br>• Partition<br>• Bootstrap Forest<br>• Boosted Tree<br>• K Nearest Neighbors<br>• Naive Bayes<br>• Support Vector Machines<br>• Model Comparison<br>• Model Screening<br>• Make Validation Column<br>• Formula Depot<br>Describes these Analyze > Specialized Modeling menu platforms:<br>• Fit Curve<br>• Nonlinear<br>• Functional Data Explorer<br>• Gaussian Process<br>• Time Series<br>• Matched Pairs<br>Describes these Analyze > Screening menu platforms:<br>• Explore Outliers<br>• Explore Missing Values<br>• Explore Patterns<br>• Response Screening<br>• Predictor Screening<br>• Association Analysis<br>• Process History Explorer |

| Document Title | Document Purpose | Document Content |
|---|---|---|
| *Multivariate Methods* | Read about techniques for analyzing several variables simultaneously. | Describes these Analyze > Multivariate Methods menu platforms:<br>• Multivariate<br>• Principal Components<br>• Discriminant<br>• Partial Least Squares<br>• Multiple Correspondence Analysis<br>• Structural Equation Models<br>• Factor Analysis<br>• Multidimensional Scaling<br>• Multivariate Embedding<br>• Item Analysis<br>Describes these Analyze > Clustering menu platforms:<br>• Hierarchical Cluster<br>• K Means Cluster<br>• Normal Mixtures<br>• Latent Class Analysis<br>• Cluster Variables |

| Document Title | Document Purpose | Document Content |
|---|---|---|
| *Quality and Process Methods* | Read about tools for evaluating and improving processes. | Describes these Analyze > Quality and Process menu platforms:<br>• Control Chart Builder and individual control charts<br>• Measurement Systems Analysis (EMP and Type 1 Gauge)<br>• Variability / Attribute Gauge Charts<br>• Process Screening<br>• Process Capability<br>• Model Driven Multivariate Control Chart<br>• Legacy Control Charts<br>• Pareto Plot<br>• Diagram<br>• Manage Limits<br>• OC Curves |
| *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>• Life Distribution<br>• Fit Life by X<br>• Cumulative Damage<br>• Recurrence Analysis<br>• Degradation<br>• Repeated Measures Degradation<br>• 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 |

| Document Title | Document Purpose | Document Content |
|---|---|---|
| *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>• Categorical<br>• Choice<br>• MaxDiff<br>• Uplift<br>• Multiple Factor Analysis |
| *Genetics* | Learn about methods that are available in JMP to help you analyze your genetic data and use that data to simulate a breeding program to predict the optimum genetic crosses to make. | Describes these Analyze > Genetics menu platforms:<br>• Marker Statistics<br>• Marker Simulation |
| *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. |
| *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. |

# Additional Resources for Learning JMP

In addition to reading JMP help, you can also learn about JMP using the following resources:

- "Search JMP"
- "JMP Tutorials"
- "Sample Data Tables"
- "Learn about Statistical and JSL Terms"
- "Learn JMP Tips and Tricks"
- "JMP Tooltips"
- "JMP User Community"
- "Free Online Statistical Thinking Course"
- "JMP New User Welcome Kit"
- "Statistics Knowledge Portal"
- "JMP Training"
- "JMP Books by Users"
- "The JMP Starter Window"

## Search JMP

If you are not sure where to find a statistical procedure, do a search across JMP. Results are tailored to the window that you launch the search from, such as a data table or report.

1. Click **Help** > **Search JMP**. Or, press Ctrl+comma.
2. Enter your search text.
3. Click the result that contains the procedure that you want.

   On the right, you can see a description and the location of the procedure.
4. Click the corresponding button to open or go to a result.

## JMP Tutorials

You can access JMP tutorials by selecting **Help > Tutorials**.

If you are not familiar with JMP, 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 Folder** to open the sample data directory.

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

Sample data tables are installed in the following directory:

On Windows: C:\Program Files\SAS\JMP\17\Samples\Data

On macOS: \Library\Application Support\JMP\17\Samples\Data

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

To view examples using sample data, select **Help > Sample Index** and navigate to the Teaching Resources section.

## Learn about Statistical and JSL Terms

For help with statistical terms, select Help > Statistics Index. For help with JSL scripting and examples, select **Help > Scripting Index**.

**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 and get help on the commands.

## 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.

## JMP Tooltips

JMP provides descriptive tooltips (or *hover labels*) when you hover 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 macOS.

## 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 on the Web > JMP User Community** or visit https://community.jmp.com.

## Free Online Statistical Thinking Course

Learn practical statistical skills in this free online course on topics such as exploratory data analysis, quality methods, and correlation and regression. The course consists of short videos, demonstrations, exercises, and more. Visit jmp.com/statisticalthinking.

## JMP New User Welcome Kit

The JMP New User Welcome Kit is designed to help you quickly get comfortable with the basics of JMP. You'll complete its thirty short demo videos and activities, build your confidence in using the software, and connect with the largest online community of JMP users in the world. Visit jmp.com/welcome.

## Statistics Knowledge Portal

The Statistics Knowledge Portal combines concise statistical explanations with illuminating examples and graphics to help visitors establish a firm foundation upon which to build statistical skills. Visit jmp.com/skp.

## JMP Training

SAS offers training on a variety of topics led by a seasoned team of JMP experts. Public courses, live web courses, and on-site courses are available. You might also choose the online e-learning subscription to learn at your convenience. Visit jmp.com/training.

## JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website. Visit jmp.com/books.

## 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 macOS) **> 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 macOS, select **JMP > Preferences > General > Initial JMP Starter Window**.

# JMP 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 jmp.com/support, including the technical support phone number.

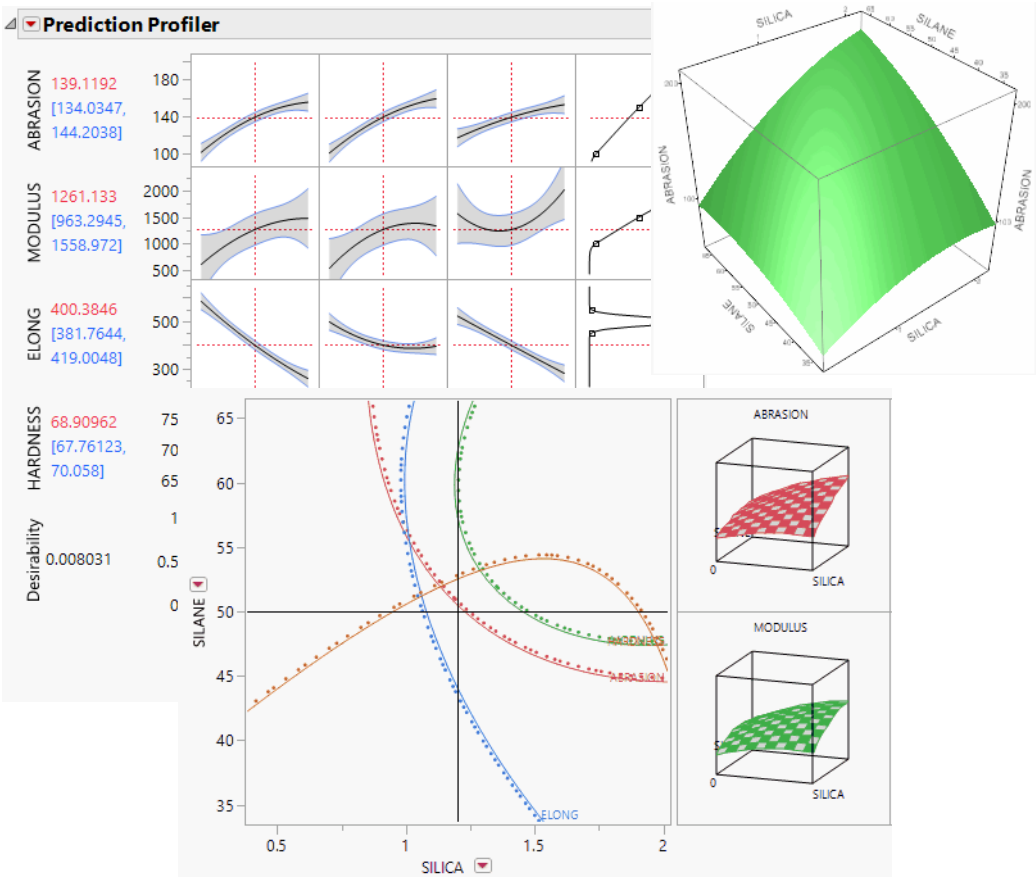
Chapter **2**

# 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. Use the interactive profilers in JMP to exploring how changes in inputs impact your response(s). 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



## Contents

Introduction to Profiling . . . . . 27
Profiler Features in JMP . . . . . 27
Profiler Launch Windows . . . . . 29
Profilers in Fit Group Reports . . . . . 31
Interpret Prediction Profiler . . . . . 31
The Profiler as a Cross-Section . . . . . 32
Set or Lock Factor Values . . . . . 34
Profiler Platform Options . . . . . 35
Common Profiler Topics . . . . . 36
Linear Constraints . . . . . 36
Noise Factors . . . . . 38

# 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. 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

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"
- "Profiler Launch Windows"
- "Profilers in Fit Group Reports"

## Profiler Features in JMP

There are several profilers in JMP and each have a different set of features. The profilers are accessible from most fitting platforms and the main menu under Graph. When you launch a profiler from the Graph menu, you must specify formula columns.

**Table 2.1** Profiler Features Summary

| | Description | Features |
|---|---|---|
| **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 |

**Table 2.1** Profiler Features Summary *(Continued)*

| | Description | Features |
|---|---|---|
| **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 | |
| Partition | Yes | | | |
| Bootstrap Forest | Yes | | | |
| Boosted Tree | Yes | | | |
| Support Vector Machines | Yes | Yes | Yes | |
| Model Screening | Yes | | | |
| Model Comparison | Yes | Yes | Yes | Yes |
| Nonlinear: Factors and Response | Yes | Yes | Yes | |
| Nonlinear: Parameters and SSE | Yes | Yes | Yes | |

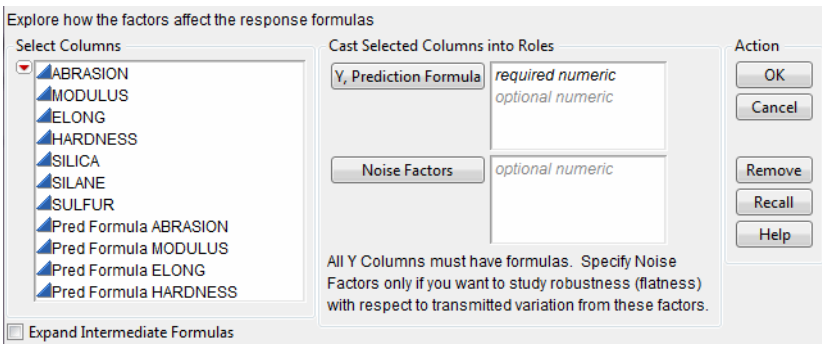
**Table 2.2** Where to Find JMP Profilers *(Continued)*

| Location | Profiler | Contour Profiler | Surface Profiler | Mixture Profiler |
|---|---|---|---|---|
| Nonlinear: Fit Curve | Yes | | | |
| Functional Data Explorer | Yes | | | |
| Gaussian Process | Yes | Yes | Yes | |
| Discriminant | Yes | | | |
| Partial Least Squares | Yes | | | |
| Structural Equation Models | Yes | | | |
| Life Distribution | Yes | | | |
| Fit Life by X | Yes | | Yes | |
| Recurrence Analysis | Yes | | | |
| Choice | Yes | | | |
| Custom Design Prediction Variance | Yes | | Yes | |

## Profiler Launch Windows

When a profiler is launched directly from the Graph menu, you provide columns with formulas as the Y, Prediction Formula columns. Often, these formulas have been saved from a fitting platform.

**Figure 2.2** Profiler Launch Window




For more information about the options in the Select Columns red triangle menu, see *Using JMP*.

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.

> **Note:** You can also add columns that contain the formula and the values for the standard errors of the corresponding response columns. These columns are used to create confidence intervals. The formula columns must be named PredSE <colname>. After you click **OK**, specify whether you want to use PredSE <colname> to construct confidence intervals for Pred Formula <colname>. Otherwise, JMP creates a separate profiler plot for PredSE <colname>.

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

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

The Surface Plot platform is discussed in "Surface Plot". 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 box is checked, the formula that is being profiled is handled differently. If the profiled formula contains another formula column that references other columns, then the original columns are substituted into the inner formula. Therefore, the formula being profiled is profiled with respect to the original columns instead of the intermediate column references. Expand Intermediate Formula also expands formula columns that have the Vector modeling type.

For example, when Fit Model fits a logistic regression for two levels (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.

## Profilers in Fit Group Reports

Fit Group reports contain results of multiple models that can be explored jointly in a profiler. To do this, launch a profiler from the Fit Group red triangle menu. This enables different *Y*'s to be profiled in the same profiler. Profilers for the individual *Y*'s can still be used in the respective Fit Model reports.

Fit Group reports are created in two ways. In some modeling platforms, when models are fit to multiple *Y*'s, the results are combined into a Fit Group report. 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 profile the individual models in the Profiler. See the *Scripting Guide*.

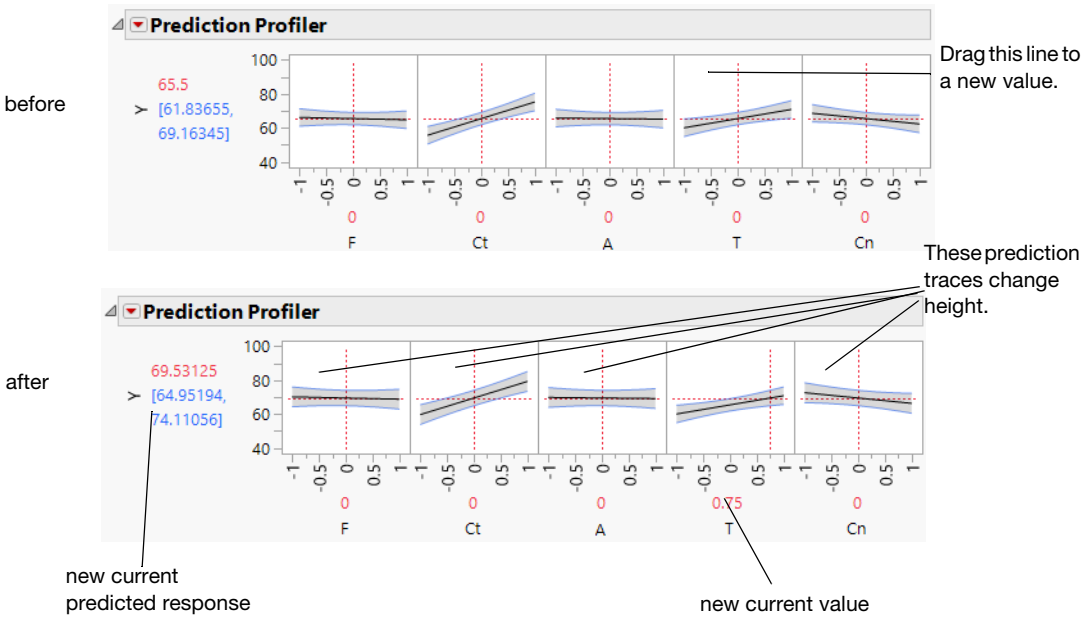
# Interpret Prediction Profiler

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 change only in height, not slope or shape.

**Figure 2.3** Changing One Factor from 0 to 0.75



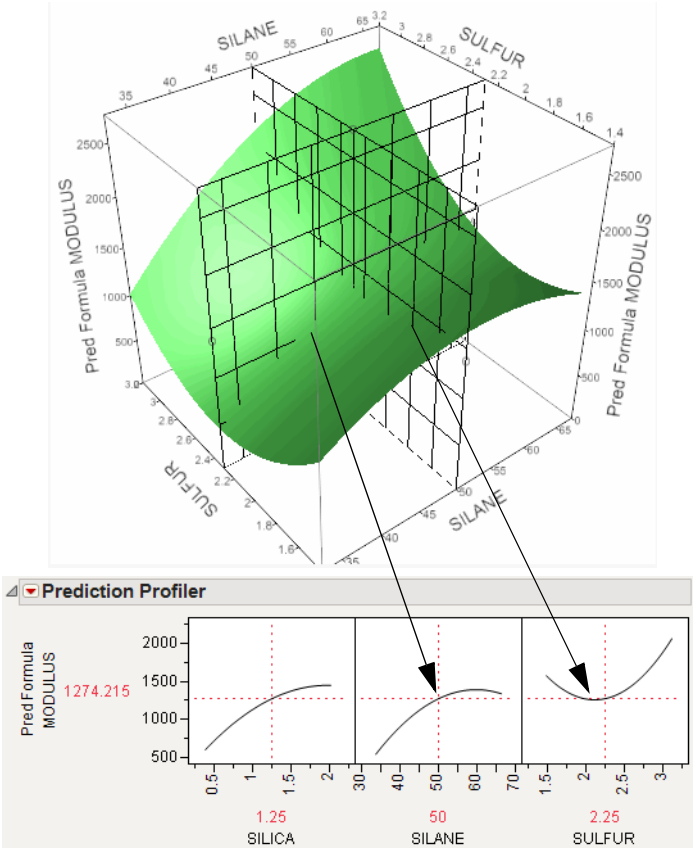

Prediction profiles are especially useful in multiple-response models to help judge which factor values can optimize a complex set of criteria.

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

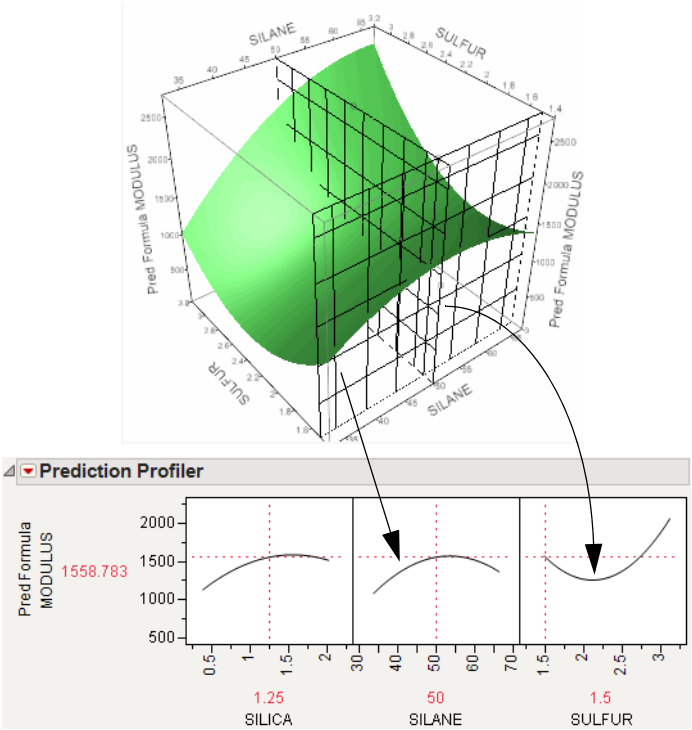
The profilers can display traces of a response surface in two dimensions. The traces are cross-sections of higher dimension surfaces. Consider a three dimensional response surface. Look at the response surface of the expression for MODULUS as a function of SULFUR and SILANE (holding SILICA constant). A superimposed grid helps to visualize the translation of the higher dimensions to the two dimensional traces.

**Figure 2.4** Profiler as a Cross-Section




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

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

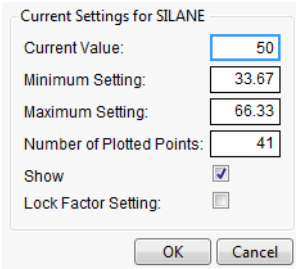



In the Prediction Profiler, note the new setting for SULFUR while the SULFUR curve itself does not change. However, the profile trace for SILANE is now where SULFUR is set at 1.5. The profile for SILANE is now shifted higher and reaches its peak close to the current SILANE value of 50.

## Set or Lock Factor Values

If you press Alt (Option on macOS) and click in a profiler, a window prompts you to enter specific settings for the factor.

**Figure 2.6** Continuous Factor Settings Window



For continuous variables, you can specify the following:

**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 Profiler red triangle menu contains the following options:

**Profiler** Shows or hides the Prediction 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.

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

**Formulas for OPTMODEL** Creates code for the OPTMODEL SAS procedure. Press Ctrl+Shift and then select **Formulas for OPTMODEL** from the red triangle menu.

The following options are available in many platforms. See *Using JMP*.

**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

The following topics are common across multiple types of profilers.

- "Linear Constraints"
- "Noise Factors"

## Linear Constraints

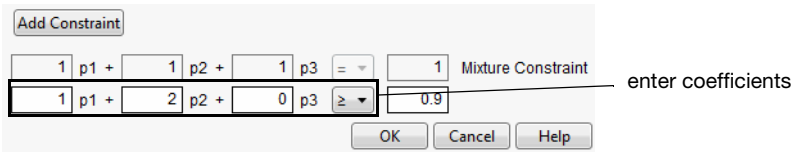
The Prediction Profiler, Custom Profiler, and Mixture Profiler can incorporate linear constraints into their operations. Linear constraints can be entered in two ways, described in the following sections.

### 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 2.7. As shown, if you are profiling factors from a mixture design, the mixture constraint is present by default and cannot be modified.

**Figure 2.7** Enter Coefficients




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

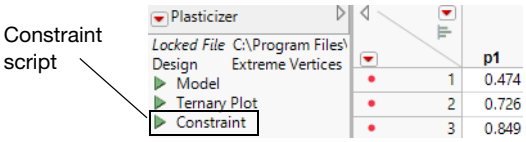
If you attempt to add a constraint for which there is no feasible solution, a message is written to the log and the constraint is not added. To delete a constraint, enter zeros for all the coefficients.

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

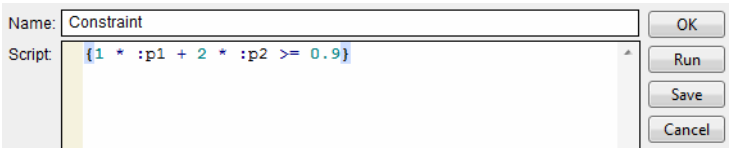
If you add constraints in one profiler and want to make them accessible by other profilers, use the **Save Linear Constraints** command, accessible through the platform red triangle menu. For example, if you created constraints in the Prediction Profiler, choose **Save Linear Constraints** under the 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 2.8.

**Figure 2.8** 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 green triangle and select **Edit**. The content of the constraint from Figure 2.7 is shown below in Figure 2.9.

**Figure 2.9** 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 2.9, 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.

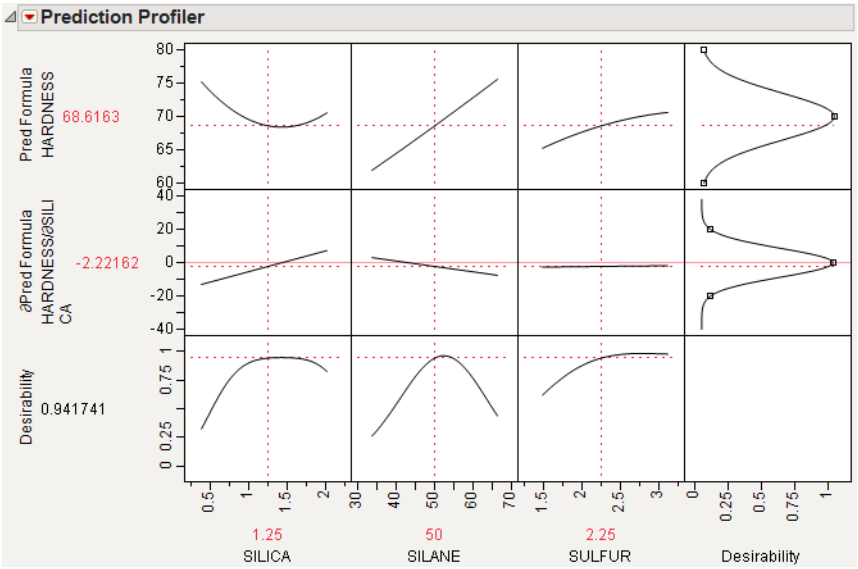
The Alter Linear Constraints and Save Linear Constraints commands are not available in the Mixture Profiler. To incorporate linear constraints into the operations of the Mixture Profiler, the Constraint Table Script must be created by one of the methods discussed in this section.

# Noise Factors

Noise factor optimization is available in the Prediction Profiler, Contour Profiler, Custom Profiler, and Mixture Profiler. Noise factors (robust process engineering) enable 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 2.10** Noise Factor Example




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. The derivatives are 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.

# Chapter 3

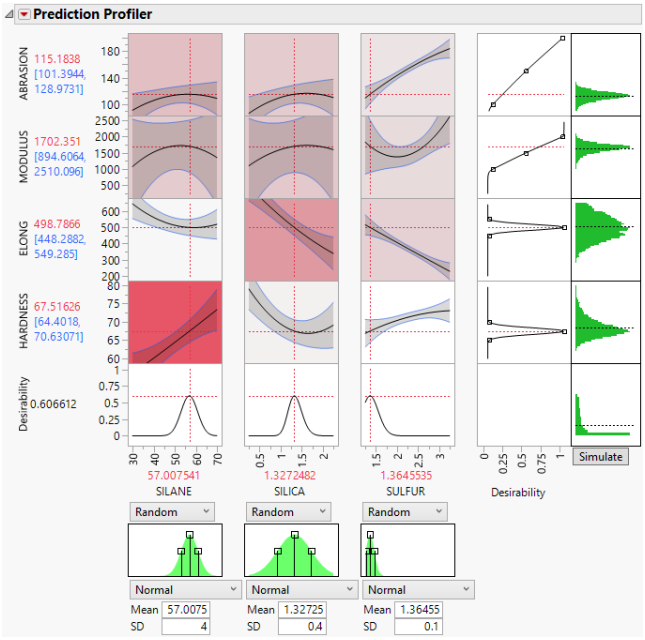
# Profiler

## Explore Cross Sections of Responses across Each Factor

The Prediction Profiler gives you a wealth of information about your model. Use the Prediction Profiler to do the following:

- 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** Prediction Profiler for Four Responses with Simulator and Importance Coloring



## Contents

Overview of the Prediction Profiler . . . 43
Example of the Prediction Profiler . . . 44
Launch the Prediction Profiler Platform . . . 45
Prediction Profiler Options . . . 46
Maximization Options Window . . . 55
Desirability Profiling and Optimization . . . 56
Construction of Desirability Functions . . . 57
How to Use the Desirability Function . . . 58
The Desirability Profile. . . . 60
Customized Desirability Functions . . . 60
Assess Variable Importance. . . . 62
Bagging . . . 66
Design Space Profiler . . . 67
Design Space Profiler Report. . . . 68
Design Space Profiler Options. . . . 68
Additional Examples of the Prediction Profiler . . . 70
Example of a Desirability Profiler for Multiple Responses. . . . 70
Example of a Noise Factor in the Prediction Profiler. . . . 72
Example of Variable Importance for One Response . . . 78
Example of Variable Importance for Multiple Responses. . . . 80
Example of Bagging to Improve Prediction . . . 82
Example of Bagging to Indicate the Accuracy of Predictions. . . . 86
Example of Confidence Intervals Using Bagging. . . . 89
Example of Confidence Intervals Using StdErr Pred Formula Column . . . 91
Example of Design Space Profiler. . . . 92
Statistical Details for the Prediction Profiler . . . 94
Statistical Details for Assess Variable Importance . . . 94
Statistical Details for Propagation of Error Bars. . . . 96
Statistical Details for Extrapolation Control Metrics . . . 97

# Overview of the Prediction Profiler

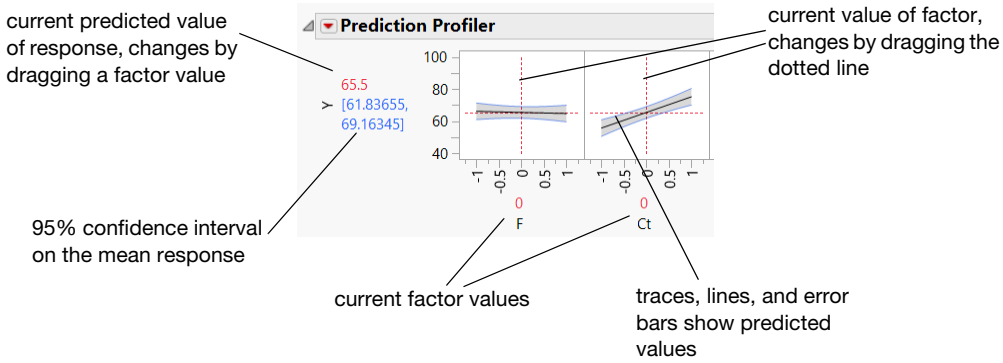
The Prediction Profiler displays profile traces (Figure 3.2) for each *X* variable. A *profile trace* is the predicted response as one variable is changed while the others are held constant at the current values. The Prediction Profiler recomputes the profiles and predicted responses (in real time) as 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 Prediction Profiler is a way of changing one variable at a time and looking at the effect on the predicted response.

**Figure 3.2** Illustration of Traces



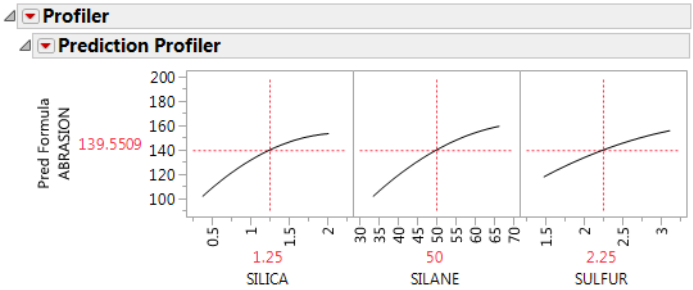
The Prediction Profiler in some situations computes confidence intervals for each profiled column. If you have saved both a standard error formula and a prediction formula for the same column, the Prediction Profiler offers to use the standard errors to produce the confidence intervals rather than profiling them as a separate column.

# Example of the Prediction Profiler

Use the Prediction Profiler to find the optimal combination of three factors in the development of a tire tread compound.

1. Select **Help > Sample Data Folder** and open Tiretread.jmp.
2. Select **Graph > Profiler**.
3. Select Pred Formula Abrasion and click **Y, Prediction Formula**.
4. Click **OK**.

**Figure 3.3** Prediction Profiler Report Window

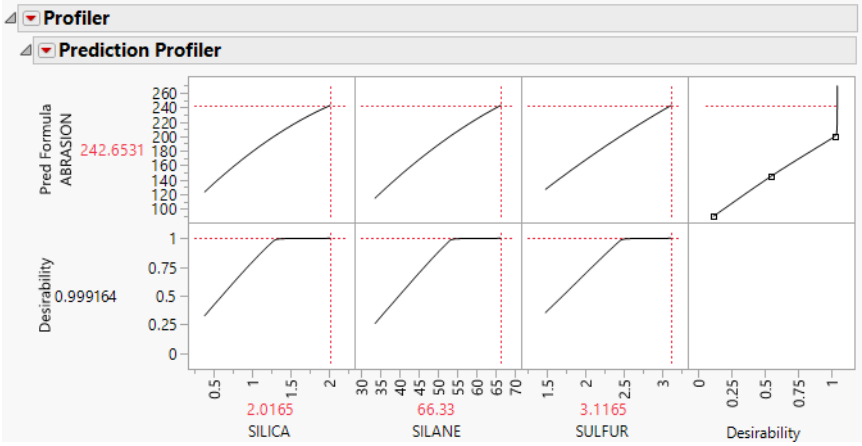



The profiler is interactive. The vertical red lines correspond to the current value of the factors. The current value of each factor is also shown in red below the horizontal axis. The red value on the vertical axis is the predicted response based on the current values of the factors. Click and drag 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. Alternatively, you can animate the profiler to cycle through a variety of factor values.

5. Click the Prediction Profiler red triangle and select **Factor Settings > Animation**.
6. Click the play button to sequentially cycle through the factor values, one factor at a time.
7. Click the Prediction Profiler red triangle and select **Optimize and Desirability > Desirability Functions**.
8. Click the Prediction Profiler red triangle and select **Optimize and Desirability > Maximize Desirability.**

**Tip:** You might need to adjust the vertical axis of the prediction profiler to view the maximum value of ABRASION in the plot.

**Figure 3.4** Maximize Desirability Report




By maximizing the desirability, you see that to maximize ABRASION, all factors are set to their highest levels.

# Launch the Prediction Profiler Platform

The Prediction Profiler can be accessed from the Graph menu or the red triangle menus of certain platforms.

- 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".
- The Prediction Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" for more information 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 Prediction Profiler red triangle menu contains 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".

**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 "Additional Example of the Contour Profiler".

---

**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.

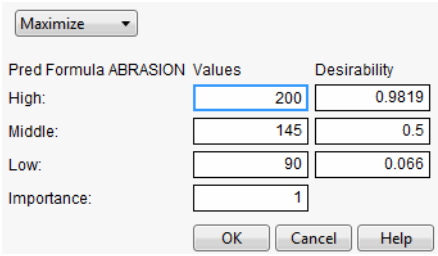
**Maximization Options** Opens the Maximization Options window where you can refine the optimization settings. See "Maximization Options Window".

**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. See "Assess Variable Importance".

**Save Bagged Predictions** (Available only when the Prediction Profiler is embedded in select 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. See "Bagging".

**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. See "Simulator".

**Design Space Profiler** (Available only if there is at least one continuous factor.) Launches the Design Space Profiler. Use the Design Space Profiler to determine operation limits for the factors that honor the specification limits on the response variables. See "Design Space Profiler".

**Interaction Profiler** Shows or hides interaction plots that update as you update the factor values in the Prediction Profiler. Use this option to 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. If there is more than one response, there is a separate tab in the Interactions Profilers report for each response.

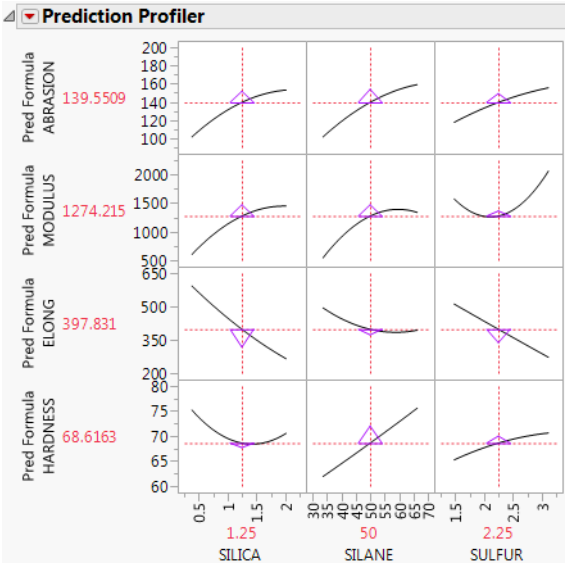
**Confidence Intervals** Shows or hides confidence intervals in the Prediction Profiler plot. The intervals are drawn by bars for categorical factors, and curves for continuous factors. The interval values are also displayed on the vertical axis in blue. These are available when the

profiler is used inside certain fitting platforms or when a standard error column of the form PredSE<colname> has been specified in the Y, Prediction Formula role of the Prediction Profiler launch dialog.

**Prop of Error Bars** (Appears when a Sigma column property exists in any of the factor and response variables.) Shows or hides the 3σ interval that is implied on the response due to the variation in the factor. The interval values are also displayed on the vertical axis in green. 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 "Statistical Details for Propagation of Error Bars".

**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. 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.

JMP PRO **Extrapolation Control** Shows a submenu of options for extrapolation control. This feature helps identify possible extrapolated predictions. A prediction is considered an extrapolation when it is made using a combination of factor points that are not within the factor space of the original data. In the extrapolation control feature, the metric used to determine if a point is an extrapolation depends on the type of model fit. For models that are fit in the Standard Least Squares personality of the Fit Model platform, the leverage at the factor settings is used as the extrapolation metric. For all other models, the regularized Hotelling's $T^2$ value is used as the extrapolation metric. See "Statistical Details for Extrapolation Control Metrics".

Extrapolation Control is available in profilers embedded in the following platforms: Fit Least Squares, Neural, Naive Bayes, Partial Least Squares, Support Vector Machines, Structural Equation Models, and Generalized Regression. It is also available in profilers launched from the Graph menu. If Extrapolation Control is not available in a particular platform, you can save the prediction formula to the data table and launch the Profiler from the Graph menu. The data used for the extrapolation control metrics depends on the type of profiler.

- When a model is built with validation, the embedded profiler and extrapolation control metrics are based on the training data.
- If you launch a profiler from the Graph menu the extrapolation control metrics are based on all data, unless you specifically exclude certain rows.
- When a model is built in a platform that ignores missing values during model fitting, those rows are excluded from the embedded profiler and extrapolation control metrics.
- When a model is built with Informative Missing, the embedded profiler and extrapolation control metrics reflect the informative missing.
- To include informative missing in the extrapolation control metrics when launching the profiler from the Graph menu use the Informative Missing column property.
- If you call Extrapolation Control from a profiler launched from the Graph menu, the regularized Hotelling's $T^2$ value is always used as the extrapolation metric, regardless of the type of model fit. Therefore, the extrapolation control results from a profiler embedded in the Standard Least Squares platform will not match those from the Graph menu profiler.

The extrapolation control red triangle menu includes options to either warn of possible extrapolation or to restrict the factor settings so that extrapolated predictions are not shown.

**Off** Turns off all extrapolation controls and warnings.

**On** Turns on extrapolation control. When this option is selected, it is indicated at the top of the profiler and the profile traces are restricted to factor combinations that do not lead to extrapolations.

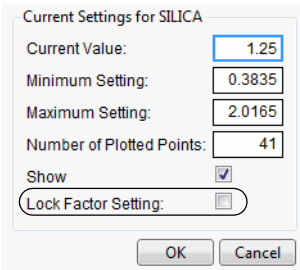
**Warning On** Turns on extrapolation warnings. When this option is selected, it is indicated at the top of the profiler. If a factor combination is selected that produces an extrapolation, an alert appears that reads --Possible Extrapolation--.

**Extrapolation Details** Shows or hides the extrapolation control details above the prediction profiler. The extrapolation control details include the value of the extrapolation metric at the current point, the value of the extrapolation threshold, the type of extrapolation metric, and the definition of the extrapolation threshold.

**Set Threshold Criterion** Opens a window that enables you to adjust the extrapolation threshold. When the extrapolation metric is the leverage at the factor settings, you can specify how the leverage is computed and the value of the corresponding multiplier. When the extrapolation metric is the regularized Hotelling's $T^2$ value, you can specify the multiplier. See "Statistical Details for Extrapolation Control Metrics".

**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".

**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. There are options to remove selected settings or all settings in the Remembered Settings red triangle menu. The names of the remembered settings are also customizable.

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

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

**Paste Settings Script** Pastes the Prediction Profiler settings from the clipboard to a Prediction 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 Prediction Profiler that you want to add to an experiment in order to do another run.

**Broadcast Factor Settings** Sends the current profiler's factor settings to all other profilers, but does not link the profilers. A change in a factor in one profiler does not cause changes in any other profilers unless Broadcast Factor Settings is selected again.

**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.

**Animation** Shows or hides animation controls that enable you to easily cycle through a variety of factor settings. See "Animation Controls".

**Default N Levels** Enables you to set the default number of levels for each continuous factor. This option is useful when the Prediction 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 any of the factors or responses have specification limits, there are columns that indicate if the row is within the specification

limits. The new data table contains scripts that can be used to visualize the in-spec regions of the factors or responses.

If you have a large number of factors, a memory allocation message might be displayed for the grid table. In such cases, you could lock some of the factors, which are held at the locked, constant values in the grid table. 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** Creates a data table of random factor settings and predicted values over those settings. If any of the factors or responses have specification limits, there are columns that indicate if the row is within the specification limits. The new data table contains scripts that can be used to visualize the in-spec regions of the factors or responses.

When you select the Output Random Table option, you are prompted to specify the number of runs. There is also an option to add random noise to one or more of the responses using the specified Std Dev values. This options adds a normal random value with mean zero and specified standard deviation to the predicted response. If the response column contains a Predicting column property that includes a Std Dev value, the Std Dev value is automatically populated with the value from the column property.

This option is a simpler equivalent to opening the Simulator, resetting all the factors to a random uniform distribution, then simulating responses (with or without added random noise).

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. The selected rows correspond to rows that have a value of 1 in the Dominant column.

JMP PRO **Shapley Values** A submenu of options to calculate Shapley values. Shapley values explain individual predictions of a model. For each independent variable, $x_j$, a vector of Shapley values, $\phi_j$, is calculated so that there is a value for each individual prediction. These values give the contribution of the independent variable to a prediction compared to the average prediction of the model fit on the background data set. Shapley values are additive and each prediction can be written as a sum of the Shapley values plus the

average prediction. The average prediction is referred to as the Shapley Intercept. For more information on Shapley values, see Shapley (1953) and Lundberg and Lee (2017).

The Profiler uses the Permutation SHAP method to calculate the Shapley values. See Lundberg (2018).

**Save Shapley Values** Adds a new column to the original data table for each independent variable in the predictive model. Each new column contains the Shapley values for that factor, calculated using the current estimation option settings. There is also a hidden column for the Shapley Intercept. By default, Shapley values are not calculated for rows that are excluded in the data table.

**Set Shapley Values Options** Opens a window that enables you to specify options for the calculation of the Shapley values.

**Background Data Choice** Specifies how much background data is used in the calculation of the Shapley values. You can specify a percentage of the training data or a specific number of rows in the training data. The Shapley values calculation uses all of the training data by default.

**Shapley Estimation Method Options** Provides options to specify the number of permutations used in Permutation SHAP and to set a random seed for reproducibility. By default, the number of permutations is 10.

There is an option to calculate the Shapley values for all rows, including excluded rows. There is also an option to add a script to the data table that produces graphs of the Shapley values. A script is added for each response variable.

Click **OK** to save the option settings. Click **OK and Run** to save the option settings and calculate the Shapley values.

**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".

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

**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.

---

**Note:** To set a default number of plots to appear in a row, go to File > Preferences > Platforms > Profiler and edit the Arrange in Rows preference.

---

**Graph Spacing** Opens a window that enables you to set the amount of horizontal space between graph panels.

**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.

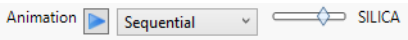
**Hide Y Variables** (Available only for continuous responses.) Specifies the response variables that you would like to show or hide in the profiler.

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

**Show Creator** Shows or hides the name of the platform that created the formula in the response column. The platform name appears on the vertical axis. (Available only if the response column contains a "Creator" named argument in the "Predicting" column property.)

## Animation Controls

**Figure 3.8** Animation Controls




**Play/Pause** Press play to animate the profiler. Moves through a cycle of factor settings and loops back to the beginning when the cycle is complete. Press pause to stop the animation.

**Cycle Type** Lists the types of cycles for factor settings.

**Sequential** Cycles through values for each factor, one factor after another. The name of the factor the animation is currently cycling through is displayed next to the speed slider bar.

**Single Factor** Cycles through values for the selected factor while all other factors are held constant. The name of the selected factor is displayed next to the speed slider bar.

**Random** Randomly cycles through different combinations of factor settings/values.

**Data Sequential** Sets the factor values to a row in the data table, one row at a time, starting with row 1. The row number of the current factor setting is displayed next to the speed slider bar.

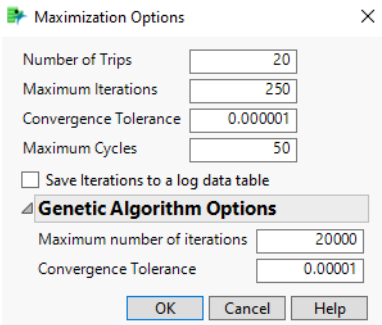
**Data Random** Sets the factor values to a random row in the data table, one row at a time. The row number of the current factoring setting is displayed next to the speed slider bar.

**Speed Slider Bar** Use the slider bar to adjust the speed of the animation.

## Maximization Options Window

Refine the optimization settings using the Maximization Options in the Prediction Profiler red triangle menu.

**Figure 3.9** Maximization Options Window




**Number of Trips** The number of random starts in the optimization algorithm. Each trip starts with a randomly chosen value for each of the factors in the model.

**Maximum Iterations** The maximum number of steps taken within each trip. If the algorithm does not converge before the maximum iterations is reached, a new trip is started.

**Convergence Tolerance** The upper limit for the convergence criterion for the optimization algorithm. If the convergence criterion is less than this value for two consecutive iterations, the algorithm stops.

**Maximum Cycles** (Not applicable if there are linear constraints.) The maximum number of cycles in the algorithm.

**Save Iterations to a log data table** Saves the results from each iteration of the optimization to a new data table.

**Genetic Algorithm Options** (Available only if there is a disallowed combination constraint or if one of the extrapolation control features is turned on.) Provides additional options for the genetic algorithm. The genetic algorithm is used only if the solution from the regular optimization algorithm is not feasible.

**Maximum number of iterations** The maximum number of iterations for the genetic algorithm.

**Convergence Tolerance** The upper limit for the convergence criterion for the genetic algorithm.

**Caution:** The solution found by the genetic algorithm is not guaranteed to be the global optimum. This is true for all genetic algorithms. A genetic algorithm is needed for optimization problems that have more general constraints. For example, JMP uses the genetic algorithm when a disallowed combinations constraint is present because these types of constraints can involve categorical factors, be nonlinear, or be any general constraint entered via JSL.

# Desirability Profiling and Optimization

In the Prediction Profiler, 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.

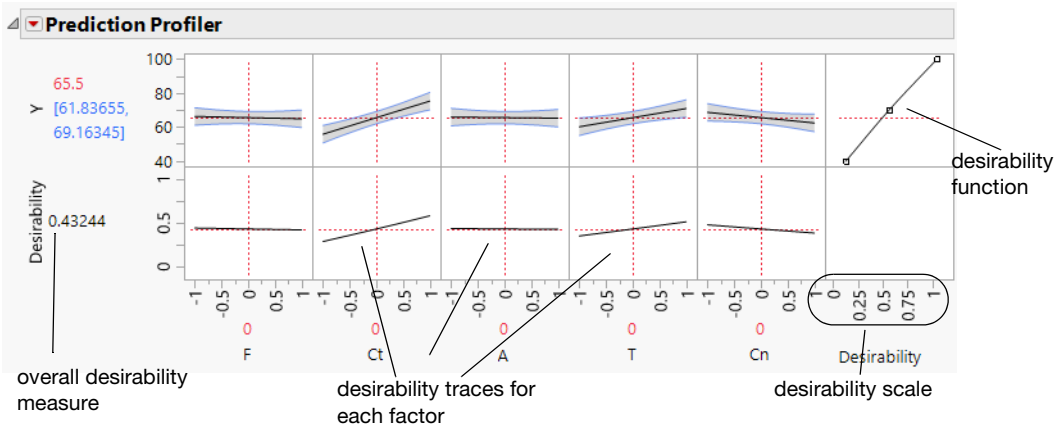
In desirability profiling, you specify a desirability function for each response. The overall desirability for all responses is defined as the geometric mean of the desirability functions for the individual responses. See Derringer and Suich (1980) for information about combining responses.

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 3.10. 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 3.10** The Desirability Prediction Profiler




## Construction of Desirability Functions

The individual desirability functions that are added to the Prediction Profiler 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".

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:

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

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:

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

### Optimization Algorithm

The method used for optimization of the overall desirability function, or of the single desirability function if there is only one response, depends on the factor types.

- 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.
- To reduce the risk of finding local optima, JMP uses multiple random starts.

## How to Use the Desirability Function

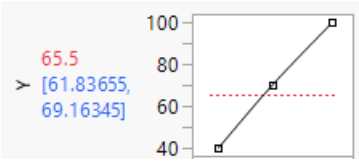
To use a variable's desirability function in the Prediction Profiler, drag the function handles to represent a response value.

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 3.11 shows steps to create desirability settings.

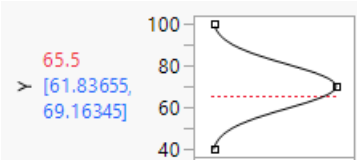
**Maximize** The default desirability function setting is maximize ("higher is better"). The High control point is positioned at the maximum Y value and aligned at the high desirability, close to 1. The Low control point is positioned at the minimum Y value and aligned at a low desirability, close to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

**Figure 3.11** Maximizing Desirability



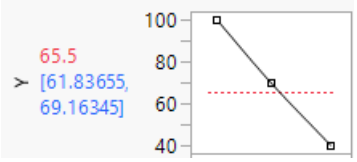

**Target** You can designate a target value as "best." In this example, the middle control point is positioned at Y = 70 and aligned with the maximum desirability of 1. Y becomes less desirable as its value approaches either 40 or 100. The High and Low control points at Y = 40 and Y = 100 are positioned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

**Figure 3.12** 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 3.13** Minimizing Desirability



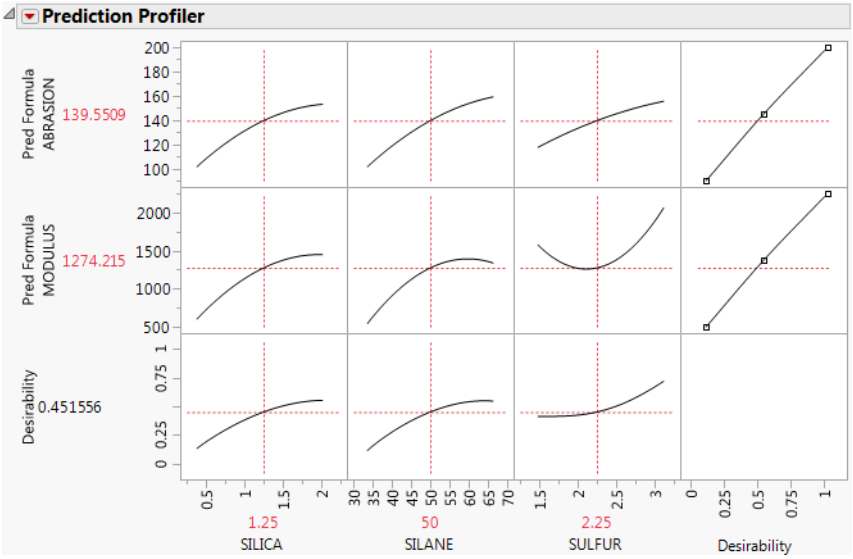

**Note:** Dragging the High or Low control point of a maximize or minimize desirability function across the *y*-value of the middle point results in the opposite point reflecting. A Minimize becomes a Maximize, and vice versa.

# The Desirability Profile

The last row of plots in the Prediction Profiler 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 3.14 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 3.14** Prediction Profile Plot with Adjusted Desirability and Factor Values




# Customized Desirability Functions

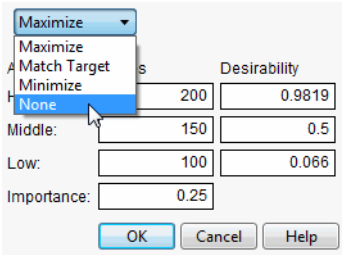
It is possible to use a customized desirability function in the Prediction Profiler. For example, suppose you want to maximize desirability using the following function.

**Figure 3.15** Maximizing Desirability Based on a Function

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

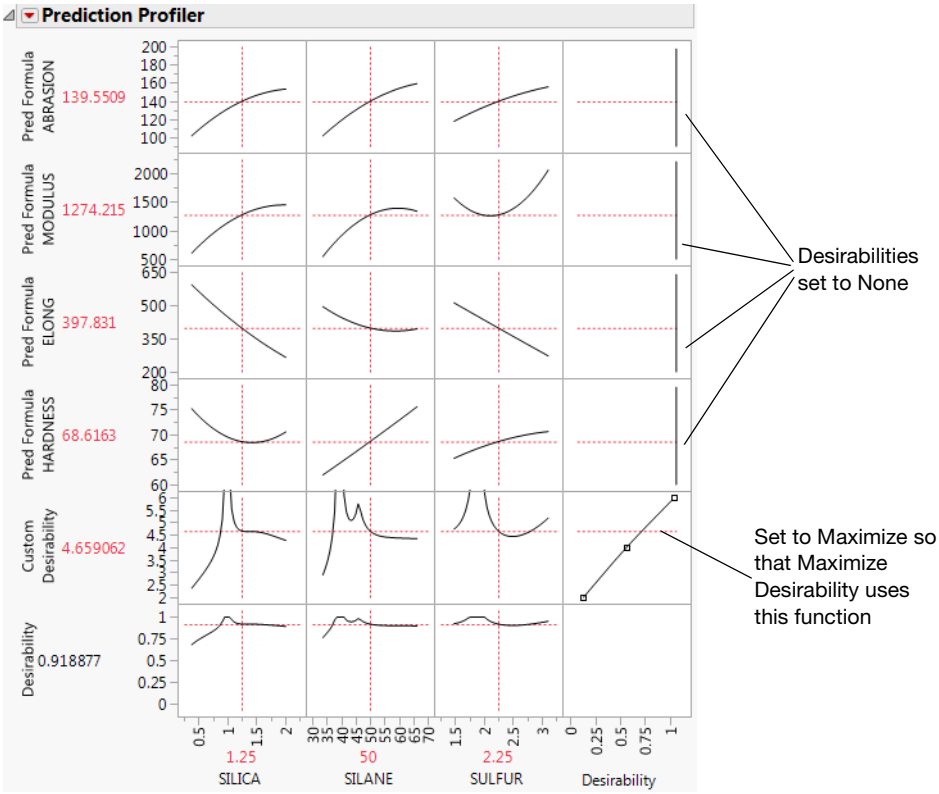
First, create a column called Custom Desirability that contains the above formula. Then, select **Graph > Profiler** to launch the platform. Select all the Pred Formula columns and the Custom Desirability column and select **Y, Prediction Formula**. Check the **Expand Intermediate Formula** option. Turn 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. Set the desirability for Custom Desirability to be maximized.

**Figure 3.16** Selecting No Desirability Goal




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

**Figure 3.17** Maximized Custom Desirability




# Assess Variable Importance

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:** In some platforms, Assess Variable Importance is not available for categorical responses with more than two levels.

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

For statistical details, see "Statistical Details for Assess Variable Importance". 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 Prediction 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:** 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 Prediction 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 Prediction 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".

**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". (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". (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 (Figure 3.43) 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 menu 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** Shows or hides 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 red triangle menu contains the following options:

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

**Reorder factors by total importance** Reorders the cells in the Prediction 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.

# JMP PRO Bagging

Available through the Prediction Profiler red triangle menu, 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 *Basic Analysis*.

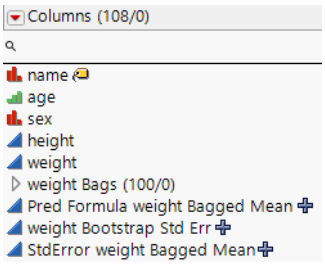
**Save Prediction Formulas** For each bagged prediction, this option saves the formula used to make that prediction in the column properties. This option is available in only a subset of the analysis platforms that offer bagging.

**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 "StdError <colname> Bagged Mean". The standard error of the bagged mean is saved in a column named "<colname> Bootstrap Std Err." 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 3.18** Bagging Columns




# Design Space Profiler

Use the Design Space Profiler to explore the impact of factor limits on the expected rate of future runs passing specifications (in-spec rate) of the responses. To determine the in-spec rate of the responses, thousands of points that are uniformly distributed through the factor space are simulated. Then, the number of simulated points that produce a response, or set of responses, that are within the response specification limits are counted to calculate the in-spec rate.

**Notes:**

- The Design Space Profiler requires specification limits on the responses. If the response columns do not have specification limits defined in a column property, you are prompted to enter specification limits for each response when you select Design Space Profiler option.
- If there are spec limits in a column property of the X variables, these are set as the initial factor operating limits in the Design Space Profiler.
- The Design Space Profiler is not designed to handle mixture factors.

## Design Space Profiler Report

The Design Space Profiler report contains a profiler graph for the expected proportion of responses that would be in-spec given the limits on the factors. For each factor, there are two solid curves that model the in-spec rate and two dashed vertical lines that represent the current operating limits. The blue curve represents the response in-spec rate as the lower operating limit is changed for the corresponding factor. The red curve represents the response in-spec rate as the upper operating limit is changed for the corresponding factor.

**Note:** The Design Space Profiler respects locked factors if the factor is locked in the Prediction Profiler before the Design Space Profiler is launched for the first time. If a factor is locked, the profiler for that factor contains only a solid vertical line at the locked value. See "Set or Lock Factor Values".

The Design Space Profiler report also contains several tables. The first table contains the in-spec rate for the current limits. If there is more than one response, this value is the combined in-spec rate for all of the responses. There is a factor table that shows the current lower and upper operating limits for each of the factors. These limits can be edited, but are automatically updated if the limits are changed on the graph or by using the Move Inward or Move Outward buttons. There is a response table that shows the current lower and upper specification limits for the responses, as well as the error standard deviation. You can edit the response specification limits if desired and the Design Space Profiler report is updated automatically.

You can change the factor operating limits in a number of ways.

- In the profilers, you can drag the vertical lines to change the operating limits.
- You can manually update the operating limits by entering new values in the factor table.
- Use the Move Inward and Move Outward buttons for optimal limit adjustment. When you use the Move Inward button, the platform chooses which factor and operating limit to adjust based on maximizing the improvement of the in-spec rate. The goal is to narrow the ranges of the operating limits enough to achieve an acceptable in-spec rate while maintaining a wide enough range that allows for expected variation in the factors.

The Design Space Profiler also supports categorical factors. Instead of setting operating limits, you can restrict the levels of the categorical factors to explore how the in-spec rate is affected.

## Design Space Profiler Options

The Design Space Profiler red triangle menu contains the following options:

**Show Portion for Each Response** Shows or hides a column in the table of responses that contains the in-specification portion for each response at the current lower and upper limits for the factors. If there is more than one response, this column is shown by default.

**Send Midpoints to Profiler** Sends the midpoints to the Prediction Profiler and sets the factors to the midpoint values. These values are the midpoints between the current Design Space Profiler lower and upper limits for the factors.

**Send Limits to Simulator** Sends the current limits for the continuous factors and a specified distribution to the Simulator. You can select a normal distribution with the limits at 2σ or 3σ or a uniform distribution. The Simulator uses the Design Space Profiler factor limits and the specified distribution to define the parameters for the random factor distribution. If you select a normal distribution, the factor limits are used to determine the mean and the standard deviation, σ. If you select a uniform distribution, the factor limits are used to determine the lower and upper values of the distribution. This option also sends the Error StdDev value for each response to the Simulator. The values are used as the standard deviation of the added random noise for the corresponding response.

If there are categorical factors, this option sends an equal probability multinomial random distribution to the simulator. Any factor levels that have been excluded in the Design Space Profiler have probabilities set to zero in the Simulator.

**Note:** This option launches the Simulator if it has not been launched yet.

**Make and Connect Random Table** Creates a new data table that contains uniformly distributed factor settings and their corresponding simulated responses. The new data table is connected to the design space profiler and the rows that contain factors that are within the current lower and upper limits are selected.

**Connect to Table** Connects the Design Space Profiler report to the specified data table. The rows that contain factors that are within the current lower and upper limits are selected in the connected table. The row selections are automatically updated as changes are made in the Design Space Profiler report.

**Connect Hide Mode** (Available only after the Connect to Table option is selected.) Hides the rows in the connected data table that contain factors that are outside of the current lower and upper limits.

**Save X Spec Limits** Saves the current operating limits for the factors as spec limits column properties in the original data table.

# Additional Examples of the Prediction Profiler

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


- "Example of a Desirability Profiler for Multiple Responses"
- "Example of a Noise Factor in the Prediction Profiler"
- "Example of Variable Importance for One Response"
- "Example of Variable Importance for Multiple Responses"
- "Example of Bagging to Improve Prediction"
- "Example of Bagging to Indicate the Accuracy of Predictions"
- "Example of Confidence Intervals Using Bagging"
- "Example of Confidence Intervals Using StdErr Pred Formula Column"
- "Example of Design Space Profiler"

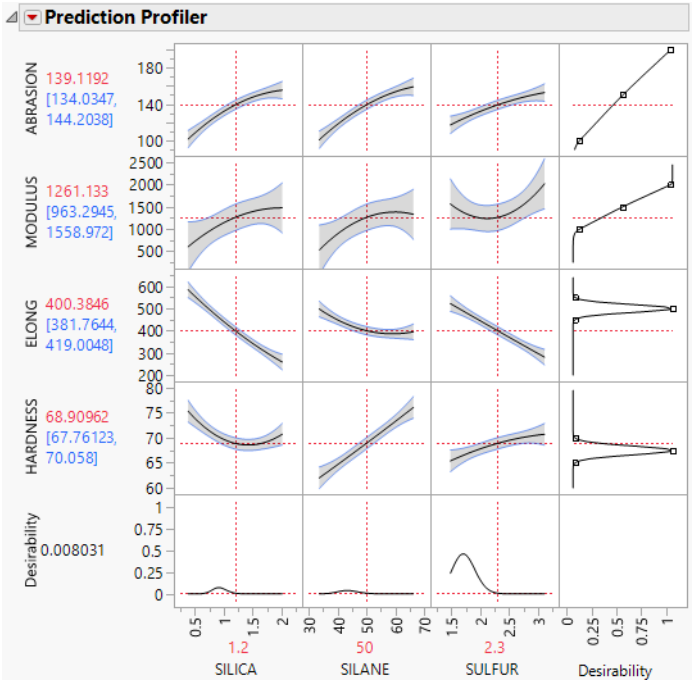

## Example of a Desirability Profiler for Multiple Responses

This example uses a desirability function to optimize multiple responses. Maximize Desirability is an option in the Prediction Profiler red triangle menu.

1. Select **Help > Sample Data Folder** and open Tiretread.jmp.
2. Click the green triangle next to RSM for 4 Responses to run the script.

   This script 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 3.19.

**Figure 3.19** Prediction Profiler for Multiple Responses before Optimization

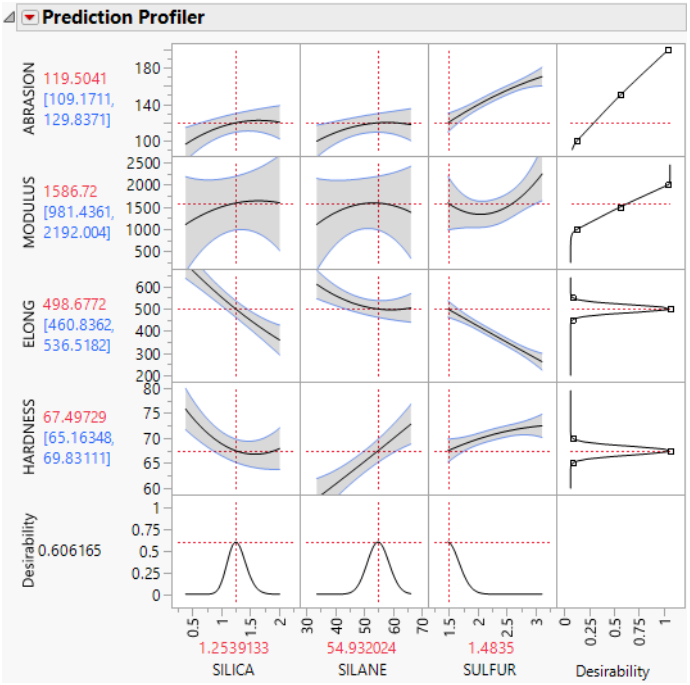



Note the following about the desirability functions:

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

3. Select **Optimization and Desirability > Maximize Desirability** from the Prediction Profiler red triangle menu to maximize desirability.

**Figure 3.20** Prediction Profiler after Optimization




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.

## Example of a Noise Factor in the Prediction Profiler

This example compares the optimization of factors for a response with and without considering the variation from a noise factor. The 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 is not easily controlled and is therefore considered a noise factor.

1. Select **Help > Sample Data Folder** and open Tiretread.jmp.
2. Select **Graph > Profiler** to launch the Prediction Profiler.
3. Assign Pred Formula HARDNESS to the **Y, Prediction Formula** role.
4. Click **OK**.
5. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Desirability Functions**.

6. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Set Desirabilities** to open the Response Goal window**.**

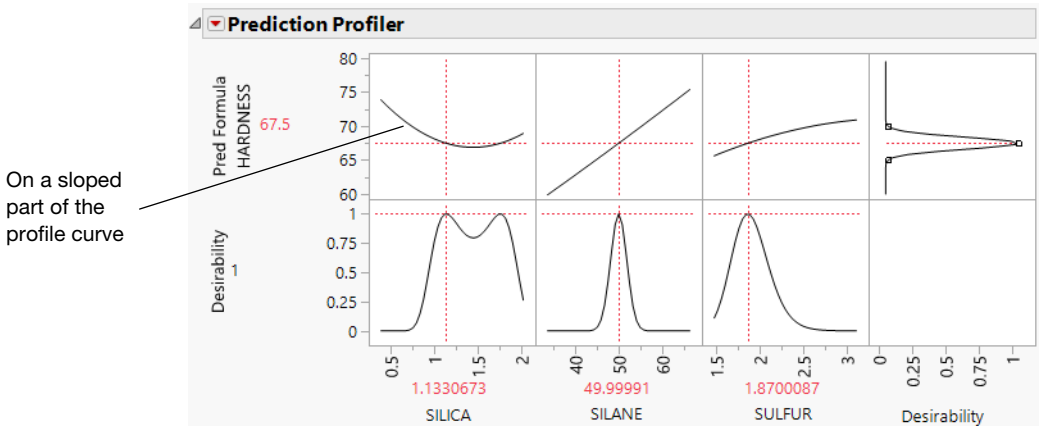
   **Tip:** Alternatively, you can double-click in the Desirability plot to open the Response Goal window.

7. Select **Match Target** from the list and click **OK**.
8. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Maximize Desirability** to find the optimum factor settings for the target value of HARDNESS.

   We get the following Prediction 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 settings impact the response, HARDNESS.

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

**Figure 3.21** Maximizing Desirability for HARDNESS



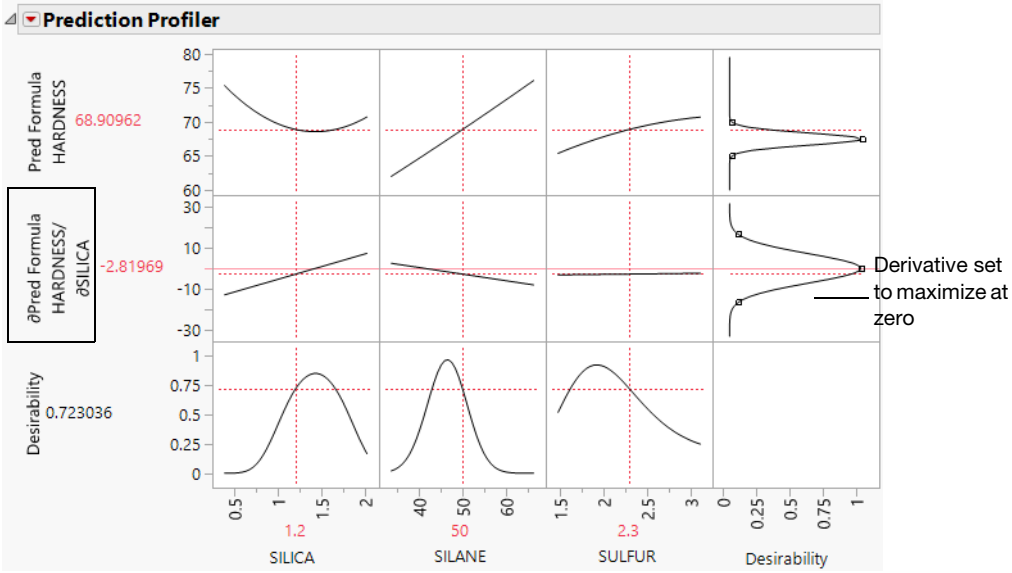

Now, we would like to not just optimize for a specific target value of HARDNESS, but also find a setting that is insensitive to the variability in the setting of Silica. Repeat the optimization process with SILICA as a noise factor.

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. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Desirability Functions**.
6. Double-click in the Desirability plot for the response Pred Formula HARDNESS to open the Response Goal window.
7. Select **Match Target** from the list and click **OK**.

   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.

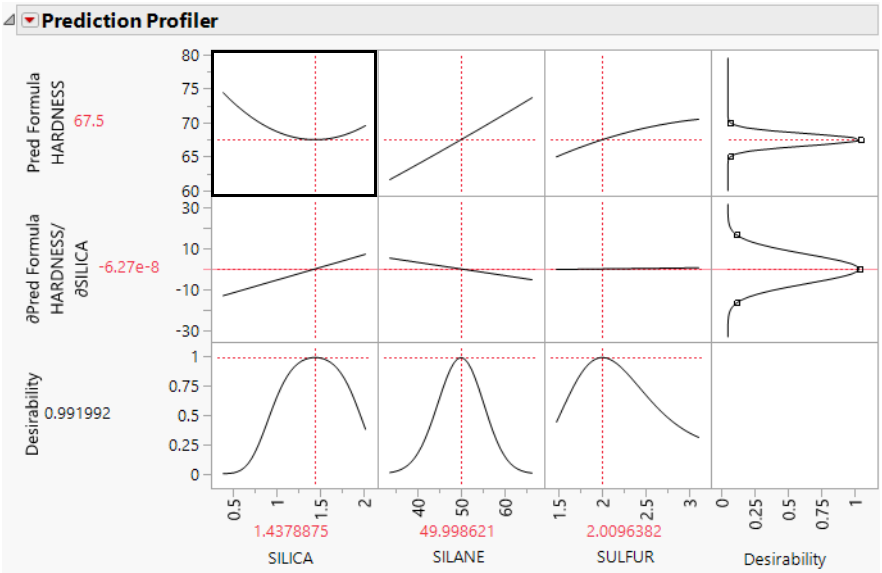
**Figure 3.22** Derivative of the Prediction Formula with Respect to Silica




8. Click the Prediction Profiler red triangle and 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 as when it was set at a value on the steep part of the SILICA curve.

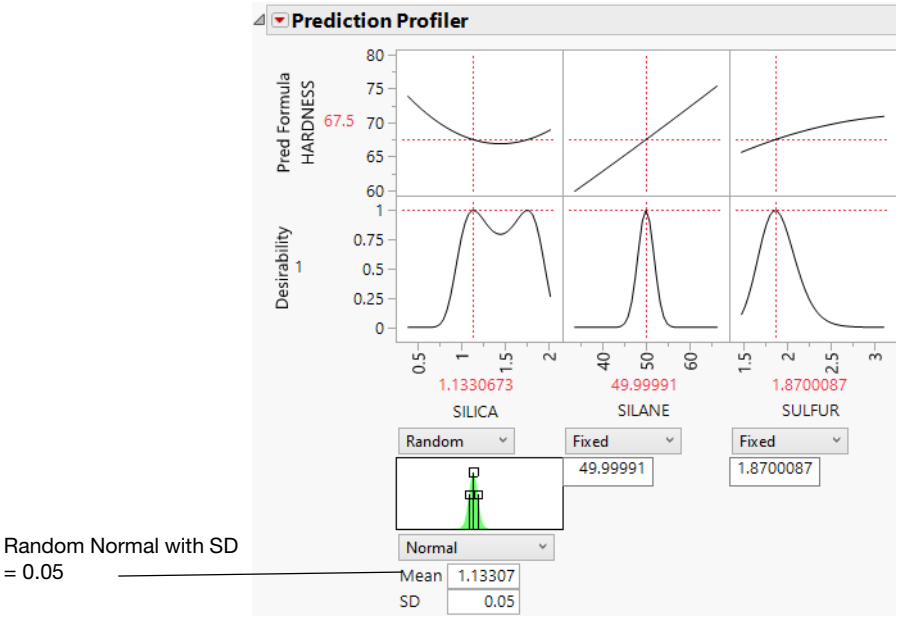
**Figure 3.23** Maximize Desirability




You can see the effect of accounting for the SILICA variation 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. In the Profiler without the noise factor, click the Prediction Profiler red triangle and select **Simulator**.
2. Assign SILICA to have a random Normal distribution with a standard deviation of 0.05. Use the default setting for the mean.

**Figure 3.24** Setting a Random Normal Distribution



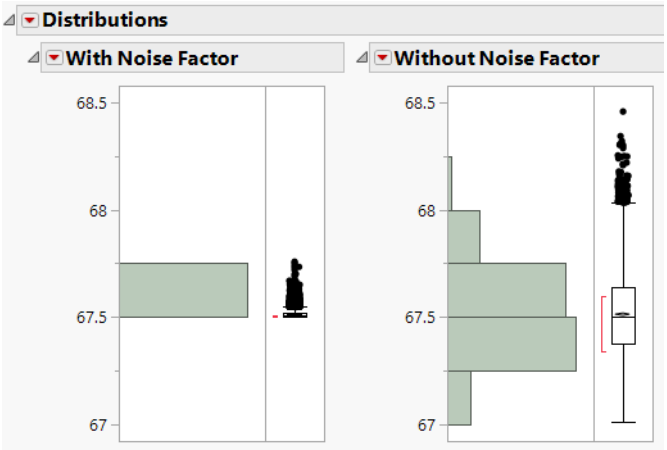

3. Click **Simulate**.
4. Click the Simulate to Table gray triangle and then click the **Make Table** button.
5. Rename the Pred Formula HARDNESS column to Without Noise Factor.
6. In the Profiler with the noise factor, click the Prediction Profiler red triangle and select **Simulator**.
7. Assign SILICA to have a random Normal distribution with a standard deviation of 0.05. Use the default setting for the mean.
8. Click **Simulate**.
9. Click the Simulate to Table gray triangle and then click the **Make Table** button.

Completing these steps for both the original and noise-factor-optimal simulations results in two similar data tables, one for each simulation. We want to compare the distributions of the predicted hardness from the two simulations. To facilitate the comparison we put the two prediction columns in a single data table.

1. Select the Pred Formula HARDNESS column in the current simulation table and select **Edit > Copy With Full Precision**.
2. In the other current simulation table, select **Cols > New Columns**.
3. Type With Noise Factor next to Column Name and click **OK**.
4. Select the With Noise Factor column and select **Edit > Paste**.
5. Select **Analyze > Distribution**.

6. Select With Noise Factor and Without Noise Factor and click **Y, Columns**.
7. Click **OK**.
8. Click the Distribution red triangle and select **Uniform Scaling**.

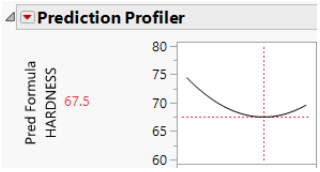
**Figure 3.25** Comparison of Distributions with and without Noise Factors




The histograms show that there is much more variation in predicted 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 skewness in the With Noise Factor is clear if you deselect Uniform Scaling from the Distributions red triangle menu. The predictions are skewed because Hardness is at a minimum with respect to SILICA, as shown in Figure 3.26. Therefore, variation in SILICA can make only HARDNESS increase. When the non-robust solution is used, the variation could be transmitted either way.

**Figure 3.26** Prediction Profiler Showing the Minima of HARDNESS by SILICA



## Example of Variable Importance for One Response

This example uses the Assess Variable Importance option in the Prediction Profiler to assess which variables are important in predicting a response modeled by a neural network. This option is helpful for models, like neural networks, that do not accommodate formal hypothesis tests.

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 Folder** 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. (Optional) Enter 123 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.
8. Click **Go**.
9. Click the red triangle next to Model NTanH(3) and 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.
10. Click the Prediction Profiler red triangle and select **Assess Variable Importance > Dependent Resampled Inputs**.

    The Variable Importance: Dependent Resampled Inputs report appears. Check that the Prediction Profiler cells have been reordered by the magnitude of the Total Effect indices in the report. In Figure 3.27, check that the Total Effect importance indices identify rooms and lstat as the factors that have most impact on the predicted response.

**Figure 3.27** Dependent Resampled Inputs Report

**Variable Importance: Dependent Resampled Inputs**

**Summary Report**

| Column | Main Effect | Total Effect |
|---|---|---|
| lstat | 0.147 | 0.279 |
| rooms | 0.127 | 0.217 |
| chas | 0.2 | 0.2 |
| distance | 0.043 | 0.108 |
| pt | 0.043 | 0.091 |
| nox | 0.077 | 0.077 |
| crim | 0.077 | 0.077 |
| indus | 0.068 | 0.068 |
| zn | 0.061 | 0.061 |
| tax | 0.045 | 0.045 |
| radial | 0.037 | 0.044 |
| b | 0.042 | 0.042 |
| age | 0.032 | 0.032 |

**Marginal Model Plots**

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.

11. Click the Prediction Profiler red triangle and 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 3.28. Check that the two factors identified as having the most impact on the predicted values are lstat and radial.

**Figure 3.28** Independent Resampled Inputs Report

**Variable Importance: Independent Resampled Inputs**

**Summary Report**

| Column | Main Effect | Total Effect |
|---|---|---|
| lstat | 0.168 | 0.339 |
| radial | 0.151 | 0.33 |
| distance | 0.104 | 0.213 |
| nox | 0.042 | 0.15 |
| rooms | 0.059 | 0.113 |
| pt | 0.041 | 0.073 |
| age | 0.019 | 0.071 |
| tax | 0.051 | 0.071 |
| b | 0.012 | 0.023 |
| chas | 0.008 | 0.022 |
| crim | 0.01 | 0.02 |
| zn | 0.007 | 0.013 |
| indus | 0.002 | 0.004 |

**Marginal Model Plots**

## Example of Variable Importance for Multiple Responses

This example uses importance estimates based on independent inputs on data from a designed experiment where the factors are orthogonal. This is done by choosing Independent Uniform Inputs as the sampling scheme for your importance indices in the Prediction Profiler. Generally, this specification is useful if you believe that, in practice, factor values vary throughout the design space, rather than assume only the settings defined in the experiment.

1. Select **Help > Sample Data Folder** 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. Click the Prediction Profiler red triangle and select **Assess Variable Importance > Independent Uniform Inputs**.

   The Summary Report is shown in Figure 3.29. 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 Prediction Profiler (Figure 3.30) have been reordered to match their ordering on the Overall table's Total Effect importance.

**Figure 3.29** Summary Report for Four Responses

**Variable Importance: Independent Uniform Inputs**

**Summary Report**

**Overall**

| Column | Main Effect | Total Effect |
|---|---|---|
| SILANE | 0.366 | 0.436 |
| SILICA | 0.341 | 0.41 |
| SULFUR | 0.18 | 0.245 |

**ABRASION**

| Column | Main Effect | Total Effect |
|---|---|---|
| SILANE | 0.389 | 0.49 |
| SILICA | 0.313 | 0.41 |
| SULFUR | 0.127 | 0.259 |

**MODULUS**

| Column | Main Effect | Total Effect |
|---|---|---|
| SILANE | 0.338 | 0.411 |
| SILICA | 0.338 | 0.408 |
| SULFUR | 0.2 | 0.291 |

**ELONG**

| Column | Main Effect | Total Effect |
|---|---|---|
| SILICA | 0.583 | 0.601 |
| SULFUR | 0.308 | 0.324 |
| SILANE | 0.063 | 0.081 |

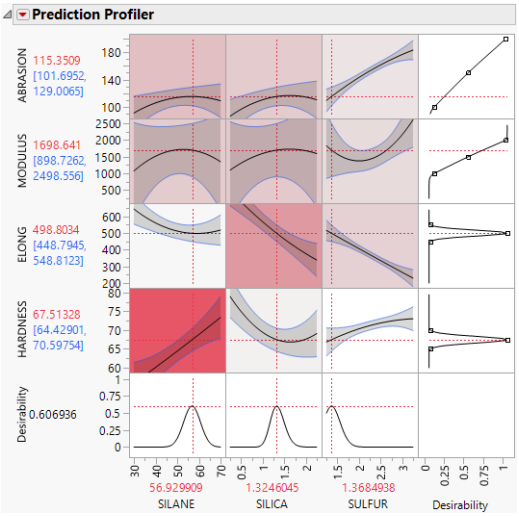
**HARDNESS**

| Column | Main Effect | Total Effect |
|---|---|---|
| SILANE | 0.672 | 0.763 |
| SILICA | 0.129 | 0.219 |
| SULFUR | 0.084 | 0.108 |

4. Click the red triangle next to Variable Importance: Independent Uniform Inputs and select **Colorize Profiler**.

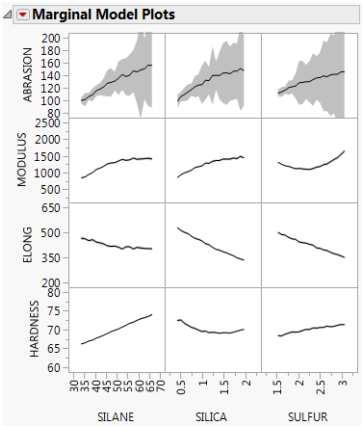
   Colors from a red to white intensity scale are overlaid on profiler panels to reflect Total Effect importance. For example, you easily see that the most important effect is that of Silane on Hardness.

**Figure 3.30** Prediction Profiler for Four Responses




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

**Figure 3.31** Marginal Model Plots for Four Responses




## Example of Bagging to Improve Prediction

This example uses bagging through the Prediction Profiler. The goal is to improve predictions from a neural network model that simultaneously models four response variables as a function of three factors. Improving predictive power is one of a number of situations where bagging is used. Bagging is also especially helpful for unstable models.

## Fit Neural Network Model

1. Select **Help > Sample Data Folder** 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. Click the red triangle next to Model NTanH(3) and select **Save Formulas**.

   **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 Prediction Profiler.

1. Click the red triangle next to Model NTanH(3) and select **Profiler**.

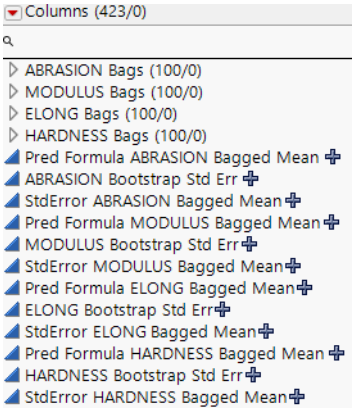
   The Prediction Profiler appears at the bottom of the report.
2. Click the Prediction Profiler red triangle and select **Save Bagged Predictions**.
3. Enter 100 next to Number of Bootstrap Samples.
4. (Optional) Enter 2121 next to Random Seed.

   **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 3.32** Columns Added to Data Table After Bagging



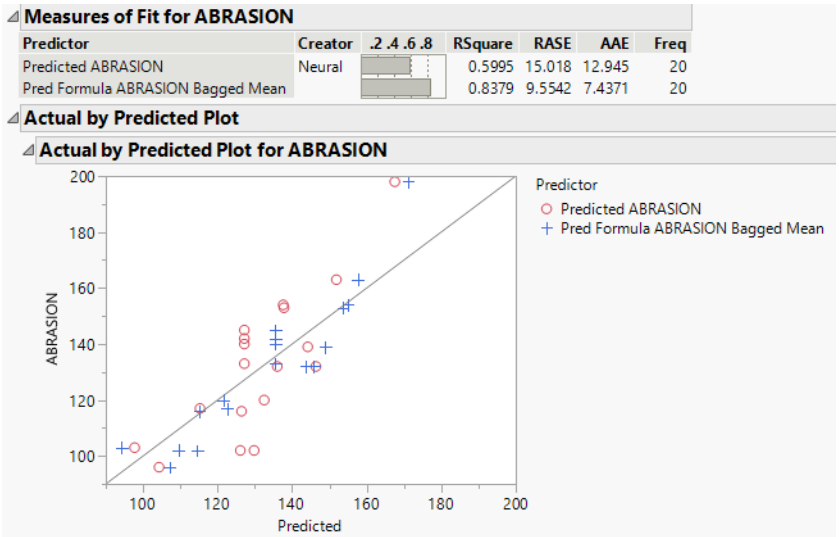

## Compare the Predictions

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. Click the Model Comparison red triangle and select **Plot Actual by Predicted**.

**Figure 3.33** Comparison of Predictions for ABRASION



**Measures of Fit for ABRASION**

| Predictor | Creator | .2 .4 .6 .8 | RSquare | RASE | AAE | Freq |
|---|---|---|---|---|---|---|
| Predicted ABRASION | Neural | | 0.5995 | 15.018 | 12.945 | 20 |
| Pred Formula ABRASION Bagged Mean | | | 0.8379 | 9.5542 | 7.4371 | 20 |

**Actual by Predicted Plot**

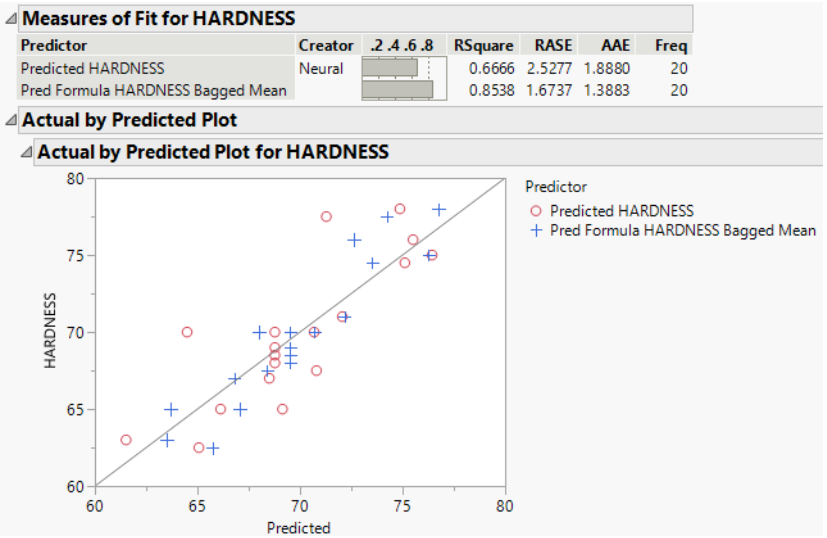
**Actual by Predicted Plot for ABRASION**



The Measures of Fit report and the Actual by Predicted Plot are shown in Figure 3.33. 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 0.8379 for the bagged predictions is higher than the RSquare value for the original model predictions. You conclude that bagging improves 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 3.34 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 3.34** Comparison of Predictions for HARDNESS



Measures of Fit for HARDNESS

| Predictor | Creator | .2 .4 .6 .8 | RSquare | RASE | AAE | Freq |
|---|---|---|---|---|---|---|
| Predicted HARDNESS | Neural | | 0.6666 | 2.5277 | 1.8880 | 20 |
| Pred Formula HARDNESS Bagged Mean | | | 0.8538 | 1.6737 | 1.3883 | 20 |

Actual by Predicted Plot

Actual by Predicted Plot for HARDNESS



## JMP PRO 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 Folder** 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 **Macros > Full Factorial**.

   This adds all terms, including interactions, to the model.
6. Click **Run**.

7. Click **Go**.

**Figure 3.35** Parameter Estimates From Generalized Regression Report

**Parameter Estimates for Original Predictors**

| Term | Estimate | Std Error | Wald ChiSquare | Prob > ChiSquare | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|
| Intercept | -46.05865 | 7.526913 | 37.44455 | <.0001* | -60.81113 | -31.30617 |
| SILICA | 32.987289 | 2.2710897 | 210.97194 | <.0001* | 28.536036 | 37.438543 |
| SILANE | 1.7880765 | 0.1194432 | 224.10381 | <.0001* | 1.5539721 | 2.0221809 |
| (SILICA-1.2)*(SILANE-50) | 1.025 | 0.2316574 | 19.577413 | <.0001* | 0.5709598 | 1.4790402 |
| SULFUR | 21.813077 | 1.4913761 | 213.92398 | <.0001* | 18.890033 | 24.73612 |
| (SILICA-1.2)*(SULFUR-2.3) | 28.5 | 4.6331484 | 37.838796 | <.0001* | 19.419196 | 37.580804 |
| (SILANE-50)*(SULFUR-2.3) | 1.575 | 0.2316574 | 46.224125 | <.0001* | 1.1209598 | 2.0290402 |
| (SILICA-1.2)*(SILANE-50)*(SULFUR-2.3) | 1.55 | 0.4633148 | 11.192085 | 0.0008* | 0.6419196 | 2.4580804 |

| Normal Distribution Parameters | Estimate | Std Error | Wald ChiSquare | Prob > ChiSquare | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|
| Scale | 5.2681724 | 0.7759427 | 46.095727 | <.0001* | 3.7473527 | 6.7889921 |

## Perform Bagging

1. Click the red triangle next to Normal Lasso with AICc Validation and select **Profilers > Profiler**.

   The Prediction Profiler appears at the bottom of the report.
2. Click the Prediction Profiler red triangle and 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". 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, <colname> Bootstrap Std Err, and StdError <colname> Bagged Mean. 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.

   All of the prediction columns for the new row are automatically calculated.

**Figure 3.36** Values for New Row

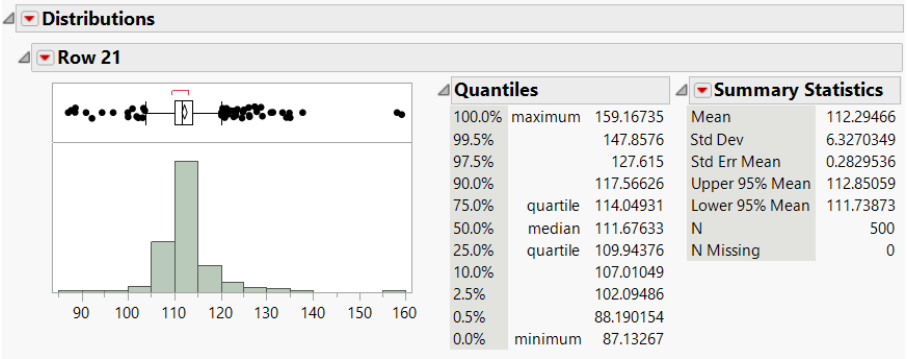
| | ABRASION | MODULUS | ELONG | HARDNESS | SILICA | SILANE | SULFUR | Pred Formula ABRASION | Pred Formula MODULUS |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 102 | 900 | 470 | 67.5 | 0.7 | 40 | 2.8 | 96.7475395 | 747.638719 |
| 2 | 120 | 860 | 410 | 65 | 1.7 | 40 | 1.8 | 113.421752 | 1074.97186 |
| 3 | 117 | 800 | 570 | 77.5 | 0.7 | 60 | 1.8 | 114.695993 | 1011.176 |
| 4 | 198 | 2294 | 240 | 74.5 | 1.7 | 60 | 2.8 | 195.496359 | 2173.94727 |
| 5 | 103 | 490 | 640 | 62.5 | 0.7 | 40 | 1.8 | 104.934463 | 865.669653 |
| 6 | 132 | 1289 | 270 | 67 | 1.7 | 40 | 2.8 | 133.734829 | 1333.44092 |
| 7 | 132 | 1270 | 410 | 78 | 0.7 | 60 | 2.8 | 138.00907 | 1310.64507 |
| 8 | 139 | 1090 | 380 | 70 | 1.7 | 60 | 1.8 | 143.683282 | 1497.9782 |
| 9 | 102 | 770 | 590 | 76 | 0.3835 | 50 | 2.3 | 101.492758 | 600.398164 |
| 10 | 154 | 1690 | 260 | 70 | 2.0165 | 50 | 2.3 | 155.361002 | 1476.17966 |
| 11 | 96 | 700 | 520 | 63 | 1.2 | 33.67 | 2.3 | 100.727591 | 525.749229 |
| 12 | 163 | 1540 | 380 | 75 | 1.2 | 66.33 | 2.3 | 159.12617 | 1330.8286 |
| 13 | 116 | 2184 | 520 | 65 | 1.2 | 50 | 1.4835 | 117.116503 | 1564.51067 |
| 14 | 153 | 1784 | 290 | 71 | 1.2 | 50 | 3.1165 | 152.737257 | 2020.06715 |
| 15 | 133 | 1300 | 380 | 70 | 1.2 | 50 | 2.3 | 139.119239 | 1261.13314 |
| 16 | 133 | 1300 | 380 | 68.5 | 1.2 | 50 | 2.3 | 139.119239 | 1261.13314 |
| 17 | 140 | 1145 | 430 | 68 | 1.2 | 50 | 2.3 | 139.119239 | 1261.13314 |
| 18 | 142 | 1090 | 430 | 68 | 1.2 | 50 | 2.3 | 139.119239 | 1261.13314 |
| 19 | 145 | 1260 | 390 | 69 | 1.2 | 50 | 2.3 | 139.119239 | 1261.13314 |
| 20 | 142 | 1344 | 390 | 70 | 1.2 | 50 | 2.3 | 139.119239 | 1261.13314 |
| 21 | • | • | • | • | 0.9 | 43 | 2 | 114.489091 | 931.321632 |

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. Click the red triangle next to Row 21 and select **Display Options > Horizontal Layout**.

**Figure 3.37** Distribution Report




| Quantiles | | |
|---|---|---|
| 100.0% | maximum | 159.16735 |
| 99.5% | | 147.8576 |
| 97.5% | | 127.615 |
| 90.0% | | 117.56626 |
| 75.0% | quartile | 114.04931 |
| 50.0% | median | 111.67633 |
| 25.0% | quartile | 109.94376 |
| 10.0% | | 107.01049 |
| 2.5% | | 102.09486 |
| 0.5% | | 88.190154 |
| 0.0% | minimum | 87.13267 |

| Summary Statistics | |
|---|---|
| Mean | 112.29466 |
| Std Dev | 6.3270349 |
| Std Err Mean | 0.2829536 |
| Upper 95% Mean | 112.85059 |
| Lower 95% Mean | 111.73873 |
| N | 500 |
| N Missing | 0 |

The Distribution Report in Figure 3.37 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 112.3, which is the mean of all the *M* bagged predictions. This prediction has a standard error of 6.33. You can also create confidence intervals for the new prediction using the quantiles. For example, a 95% confidence interval for the new prediction is 102.09 to 127.62.

## JMP PRO Example of Confidence Intervals Using Bagging

In this example, use the bagged standard errors to obtain confidence intervals around predictions for a Neural Network model.

### Fit Neural Network

1. Select **Help > Sample Data Folder** and open Tiretread.jmp.
2. Select **Analyze > Predictive Modeling > Neural**.
3. Select ABRASION and click **Y, Response**.
4. Select SILICA, SILANE, and SULFUR and click **X, Factor**.
5. Click **OK**.
6. (Optional) Enter 123456 next to Random Seed.

**Note:** Results vary due to the random nature of choosing a validation set in the Neural Network model. To reproduce the results shown in this example, set the Random Seed.

7. Click **Go**.

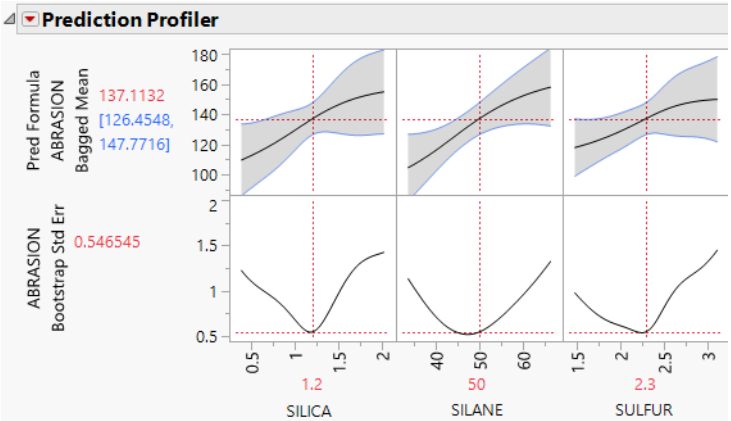
## Perform Bagging and Launch the Profiler

1. Click the red triangle next to Model NTanH(3) and select **Profiler**.
2. Click the Prediction Profiler red triangle and select **Save Bagged Predictions**.
3. Enter 100 next to Number of Bootstrap Samples.
4. (Optional) Enter 123456 next to Random Seed.

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

5. Click **OK**.
6. In the data table, click the green triangle next to the Bagging Profiler script.

**Figure 3.38** Profiler of Bagged Neural Predictions with Confidence Intervals




A profiler is created with confidence intervals using the values from the StdError ABRASION Bagged Mean column.

# Example of Confidence Intervals Using StdErr Pred Formula Column

In this example, use the saved Prediction Formula and StdErr Prediction Formula columns from a standard least squares model to create a prediction profiler with confidence intervals.

## Fit the Model

Fit a standard least squares regression model using data on diabetes patients.

1. Select **Help > Sample Data Folder** and open Diabetes.jmp.
2. Select **Analyze > Fit Model**.
3. Select Y and click **Y**.

   By default, the Personality is set to Standard Least Squares.
4. Select Age through Total Cholesterol and click **Add**.
5. Select Validation and click **Validation**.
6. Click **Run**.
7. Click the Response Y red triangle and select **Save Columns > Prediction Formula**.
8. Click the Response Y red triangle and select **Save Columns > StdErr Pred Formula**.

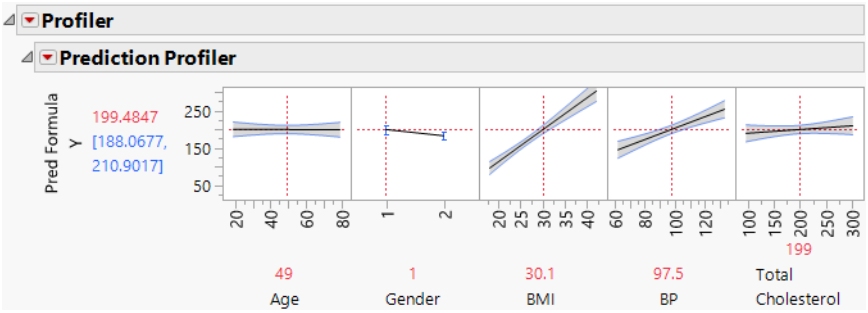
## Launch the Profiler

The data table now contains two additional columns named Pred Formula Y and PredSE Y. Use these formula columns in the Profiler.

1. Select **Graph > Profiler**.
2. Select Pred Formula Y and PredSE Y and click **Y, Prediction Formula**.
3. Click **OK**.

   A JMP Alert window appears asking if you want to use the PredSE column to make intervals for the Pred Formula Y column.
4. Click **Yes** in the JMP Alert window.

**Figure 3.39** Profiler with Confidence Intervals



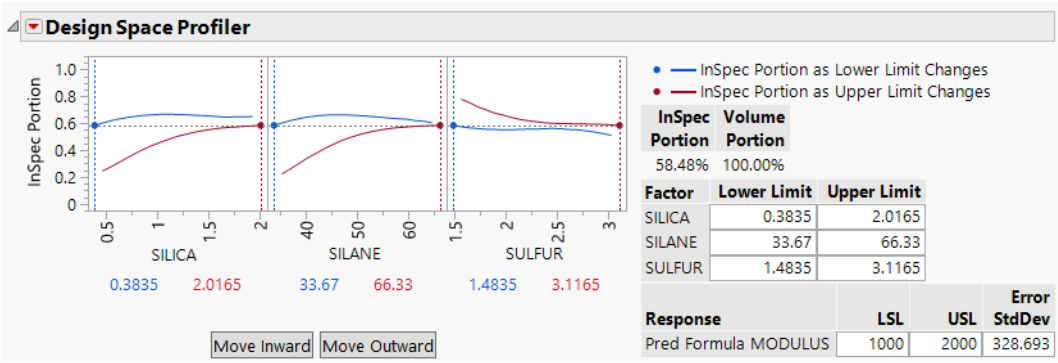

A prediction profiler with confidence intervals constructed using the standard error of the predicted values is shown.

## Example of Design Space Profiler

Use the Design Space Profiler option in the Prediction Profiler to find operational ranges for three factors that predict an in-spec response at least 75% of the time.

1. Select **Help > Sample Data Folder** and open Tiretread.jmp.
2. Select **Graph > Profiler**.
3. Select Pred Formula MODULUS and click **Y, Prediction Formula**.
4. Click **OK**.
5. Click the Prediction Profiler red triangle and select **Design Space Profiler**.
6. In the Enter Spec Limits window, specify 1000 as the lower specification limit (LSL) and 2000 as the upper specification limit (USL). These are the specification limits for Modulus.
7. Click **OK**.

**Figure 3.40** Initial Design Space Profiler Report



The initial Design Space Profiler report shows a profiler graph for each factor and several tables. With the factors unconstrained over their full ranges, the InSpec Portion is 58.48%. This means that without defining an operating range, the factors predict responses that fall between 1000 and 2000 about 58% of the time. Now explore setting operating limits on the factors.

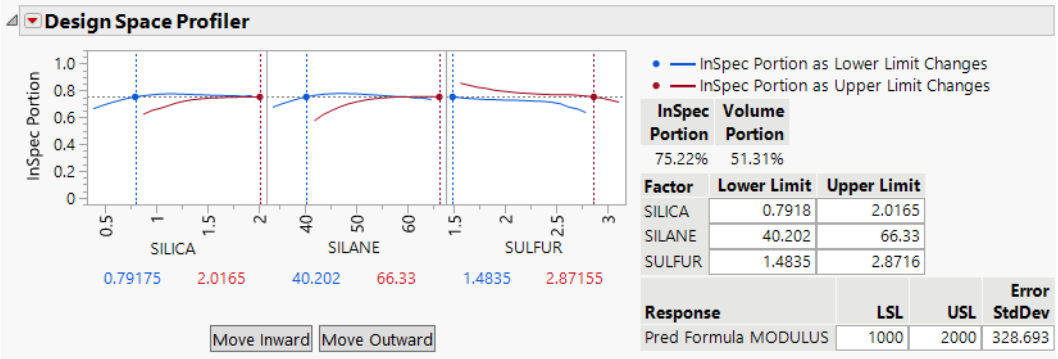
8. Click **Move Inward**.

   The first specification limit to move is the LSL of SILANE.

9. Click **Move Inward** until the InSpec Portion is above 0.75.

**Note:** This should take a total of about 11 clicks.

**Figure 3.41** Updated Design Space Profiler



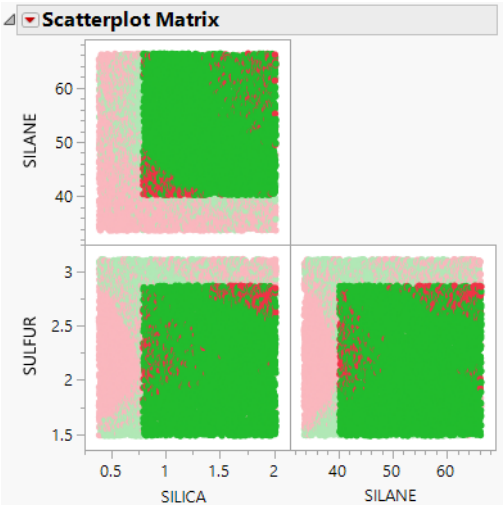

The updated Design Space Profiler report shows that setting operating limits improves the InSpec Portion to about 75%. The factor SULFUR keeps a relatively wide range, with only it's USL moving slightly inward. The other factors, SILICA and SILANE, keep their USLs at their respective maximums while their LSLs moved inward.

10. Click the Design Space Profiler red triangle and select **Make and Connect Random Table**.
11. Click **OK** in the Output Random Table window.

    The data table contains 10,000 rows of uniformly distributed factor settings with their corresponding simulated responses.

12. In the new data table, click the green triangle next to the Scatterplot Matrix X script to view the setting combinations that lead to in-spec material.

**Figure 3.42** Scatterplot Matrix Highlighting Desirable Operating Space




The scatterplots provide a visual representation of the operating limits and their impact on the in spec results. Predicted responses that are within the specification limits are colored green and predicted responses that are outside of the specification limits are colored red. The selected points contain factors at values that are within the operating limits.

# Statistical Details for the Prediction Profiler

This section contains statistical details for the Prediction Profiler:

- "Statistical Details for Assess Variable Importance"
- "Statistical Details for Propagation of Error Bars"
- "Statistical Details for Extrapolation Control Metrics"

## Statistical Details for Assess Variable Importance

This section contains details that relate to the how the variable importance indices in the Prediction Profiler 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").

### 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); 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").

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 the following:

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

### 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.

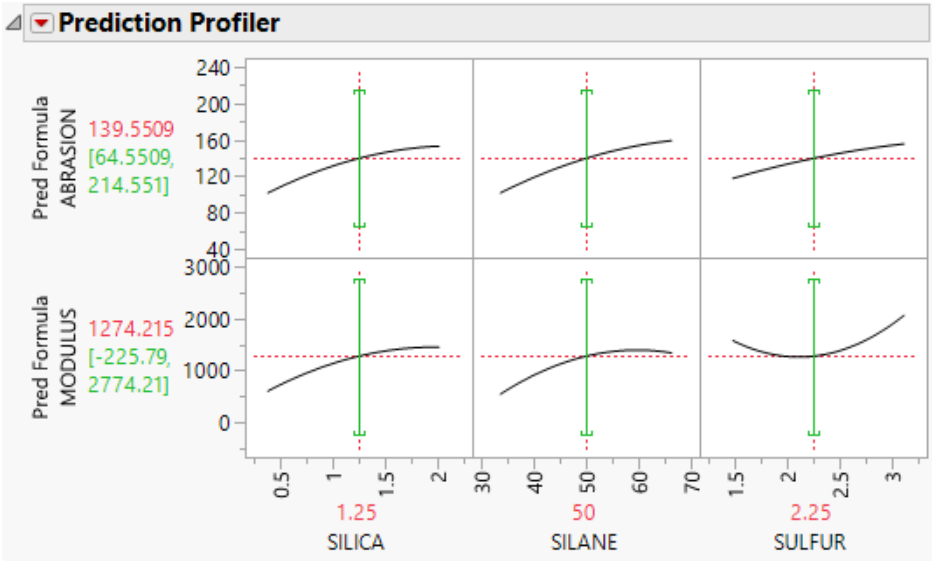
The standard errors that are reported are the standard error values in effect when the replications terminate.

## Statistical Details for 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 Prediction 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 3.43** 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 follows:

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

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.

## JMP PRO Statistical Details for Extrapolation Control Metrics

The Extrapolation Control option in the Prediction Profiler has two metrics that are used to determine if a point is an extrapolation. The type of metric used depends on the type of model fit.

## Leverage

In models that are fit in the Standard Least Squares personality of the Fit Model platform, the leverage at the factor settings is used as the extrapolation metric.

The leverage of the $i^{th}$ observation, $h_{ii}$, is the $i^{th}$ diagonal entry of the matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, sometimes called the hat matrix. The leverage for a new prediction point is calculated as $h_{pred} = \mathbf{x}'_{pred}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{pred}$. The following two criteria can be used to determine if a prediction with leverage $h_{pred}$ is an extrapolation:

- $h_{pred} > K \times \max(h_{ii})$, where $K$ is a customizable multiplier
- $h_{pred} > L \times p/n$, where $L$ is a customizable multiplier, $p$ is the number of variables, $n$ is the number of observations, and $p/n$ is the average leverage

You can use the Set Threshold Criterion option to specify which criterion is used and the value of the multiplier. The default values of the multipliers are $K = 1$ and $L = 3$.

**Note:** Extrapolation control on profilers run from the graph menu using a saved least squares model do not implement the leverage methodology. Instead, the Regularized Hotelling's $T^2$ methodology is used.

## Regularized Hotelling's $T^2$

In models other than least squares models, the Regularized Hotelling's $T^2$ value is used as the extrapolation metric. The $T^2$ value for the training data and $T^2$ values for the prediction points are calculated as follows:

$$T^2 = (x - \bar{x})'\hat{\Sigma}^{-1}(x - \bar{x})$$

$$T^2_{pred} = (x_{pred} - \bar{x})'\hat{\Sigma}^{-1}(x_{pred} - \bar{x})$$

where $\hat{\Sigma}$ is the Schafer and Strimmer regularized covariance matrix estimator estimated on the training data. The target matrix used for the Schafer Strimmer estimator is a diagonal covariance matrix. See Schafer and Strimmer (2005). In platforms that train models using observations with missing values, the covariance matrix is estimated with pairwise deletion.

**Note:** Categorical variables are converted to indicator variables for these calculations.

The calculation of the threshold depends on the number of nonmissing $T^2$ values computed on the training data.

- If there are ten or more nonmissing $T^2$ values, the threshold is set as follows:

$$\overline{T^2} + K\hat{\sigma}_{T^2}$$

where

$K$ is a customizable multiplier and is set to 3 by default

$\hat{\sigma}_{T^2}$ is the standard deviation of the $T^2$ values.

- If there are less than ten nonmissing $T^2$ values, the threshold is set using an $F$ distribution quantile equivalent to a $K\sigma$ limit.

$$Q_{T^2}(q)$$

where

$q = \Phi(K)$

$\Phi(\cdot)$ is the standard normal distribution

$K$ is a customizable multiplier and is set to 3 by default

$$T^2 \sim \frac{(n+1)(n-1)p}{n(n-p)} F(p, n-p)$$

$p$ is the number of parameters

$n$ is the number of nonmissing $T^2$ values

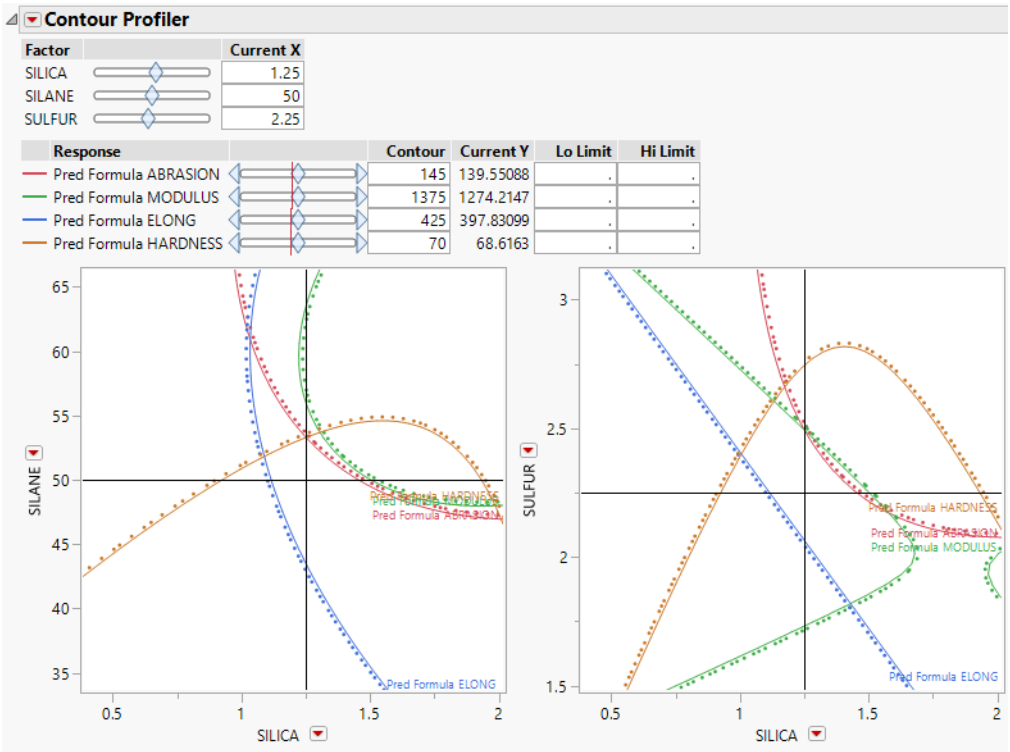
Chapter 4

# Contour Profiler

## Explore Contours of Responses across Two Factors

The Contour Profiler shows plots of response contours for two factors at a time. Multiple contour plots for different combinations of factors can be viewed at the same time. The interactive contour profiling facility is useful for optimizing response surfaces graphically.

**Figure 4.1** Contour Profiler Example



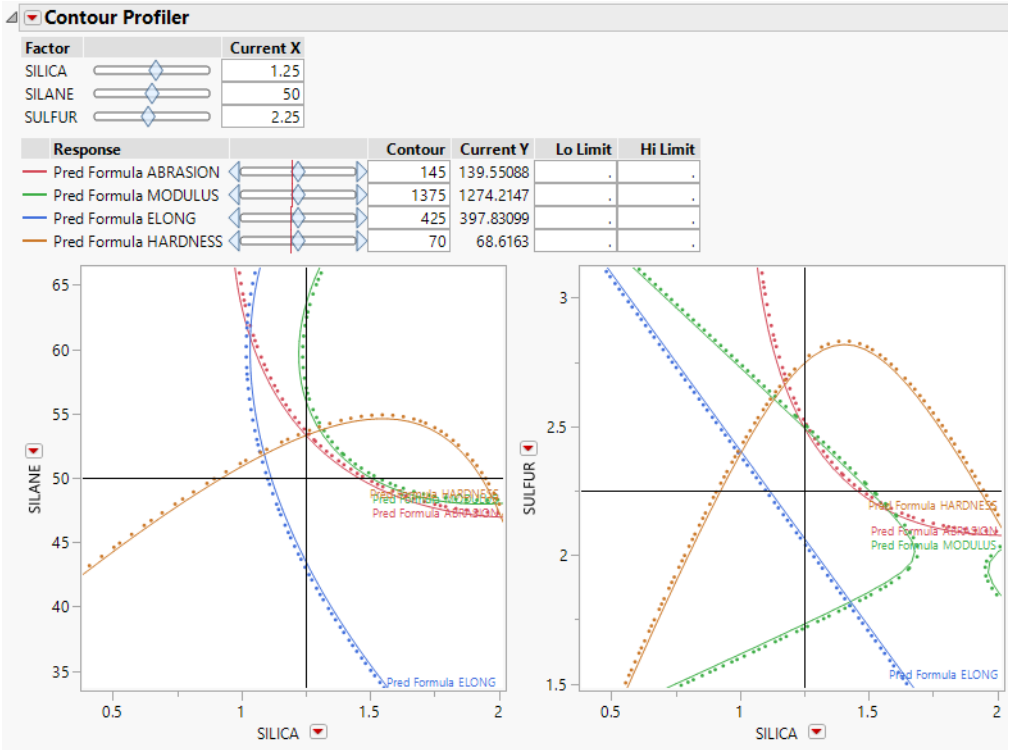
## Contents

Overview of the Contour Profiler . . . . . 103
Example of the Contour Profiler . . . . . 103
Launch the Contour Profiler Platform . . . . . 105
The Contour Profiler Report . . . . . 105
Factor Settings and Controls . . . . . 106
Response Settings and Controls . . . . . 106
Contour Profiler Platform Options . . . . . 107
Constraint Shading Settings . . . . . 109
Additional Example of the Contour Profiler . . . . . 109

# Overview of the Contour Profiler

The Contour Profiler shows plots of response contours for two factors at a time. Multiple contour plots for different combinations of factors can be viewed at the same time. 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 available. 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

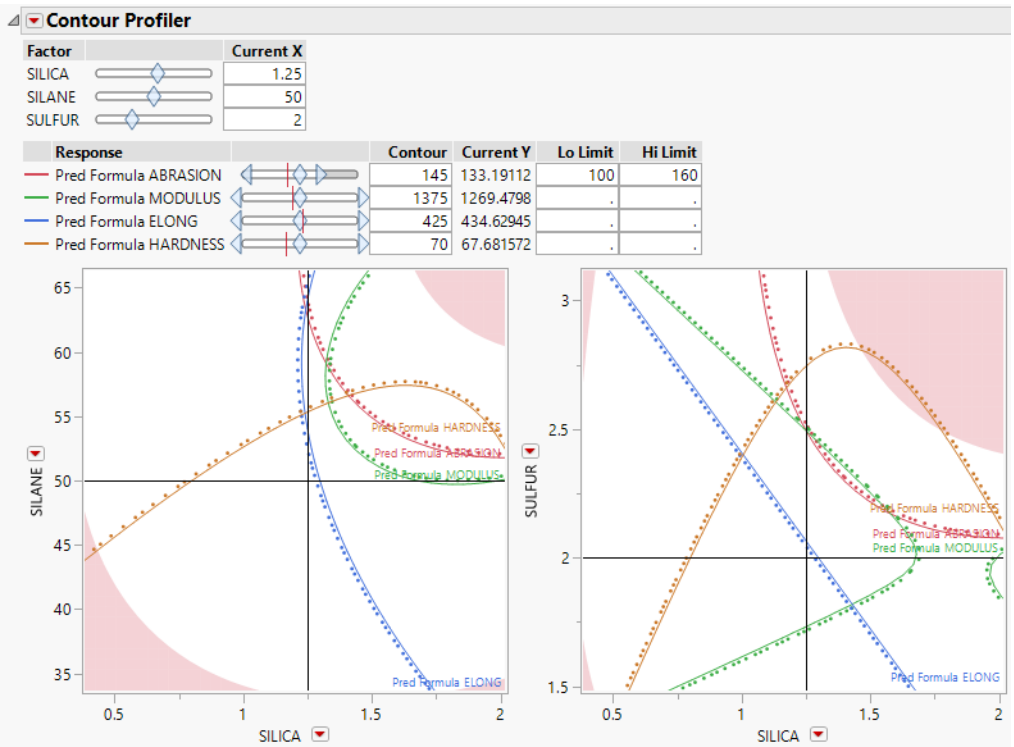
This example explores the relationship between three factors (SILICA, SILANE, and SULFUR) and four responses (ABRASION, MODULUS, ELONG, and HARDNESS). The data table contains saved prediction formula columns for the four response variables.

1. Select **Help > Sample Data Folder** 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**.
5. Click the Contour Profiler red triangle and select **Multiple Contour Frames**.
6. Under Horizontal Factor, select SILICA and under Vertical Factor, select SULFUR.
7. Click **OK**.

   On both contour plots, the values for SILICA are on the horizontal axis. The values for SILANE and SULFUR are on the vertical axes. In the first contour plot, the value for SULFUR is fixed at 2.25. In the second contour plot, the value for SILANE is fixed at 50 (Figure 4.2).
8. Click the **Current X** box for SULFUR and type 2. Click anywhere outside the box to set the value.
9. In the **Lo Limit** box for Pred Formula ABRASION, type 100.
10. In the **Hi Limit** box for Pred Formula ABRASION, type 160.
11. Click anywhere outside the box to update the plot.

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



The contour lines in the first contour plot differ from those in Figure 4.2 because SULFUR is set at a different value. This can be seen in the placement of the horizontal line at 2 in the second contour plot. Move the slider bar for SULFUR to see how different values further change the contour lines in the first contour plot. Move the slider bar for SILANE to see how different values change the contour lines in the second contour plot. The shaded area on the contour plots represent the regions of the graphs that are not feasible, based on the Lo Limit and Hi Limit values for ABRASION.

# Launch the Contour Profiler Platform

The Contour Profiler can be accessed from the Graph menu or the red triangle menus of certain platforms.

- 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".
- The Contour Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" for more information 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, 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. To change the factor assignments to the horizontal and vertical axes of a contour plot, click the red triangle on the axis and select a factor. If you select the factor that is on the opposite axis, the factor assignments are interchanged.

- "Factor Settings and Controls"
- "Response Settings and Controls"

## Factor Settings and Controls

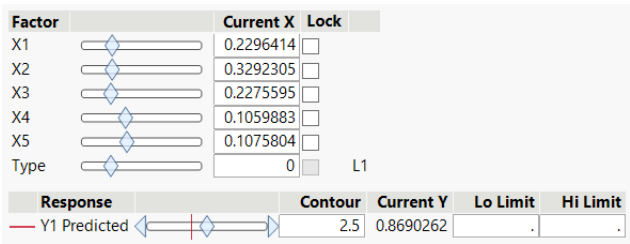
You can change the settings for the factors directly from the Contour Profiler report.

**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




**Note:** If one or more of the factors have specification limits, a dotted rectangle is drawn on the contour plot to show the joint spec limit range.

## Response Settings and Controls

You can change the settings for the responses directly from the Contour Profiler report.

**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.

**Note:** If your response has a Response Probability column property, the High Limit is automatically set to 0.5. In this case, the shaded region is the region where that response level is predicted.

# Contour Profiler Platform Options

The Contour Profiler red triangle menu contains the following 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".

**Simulator** Launches the Simulator. See "Simulator".

**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).

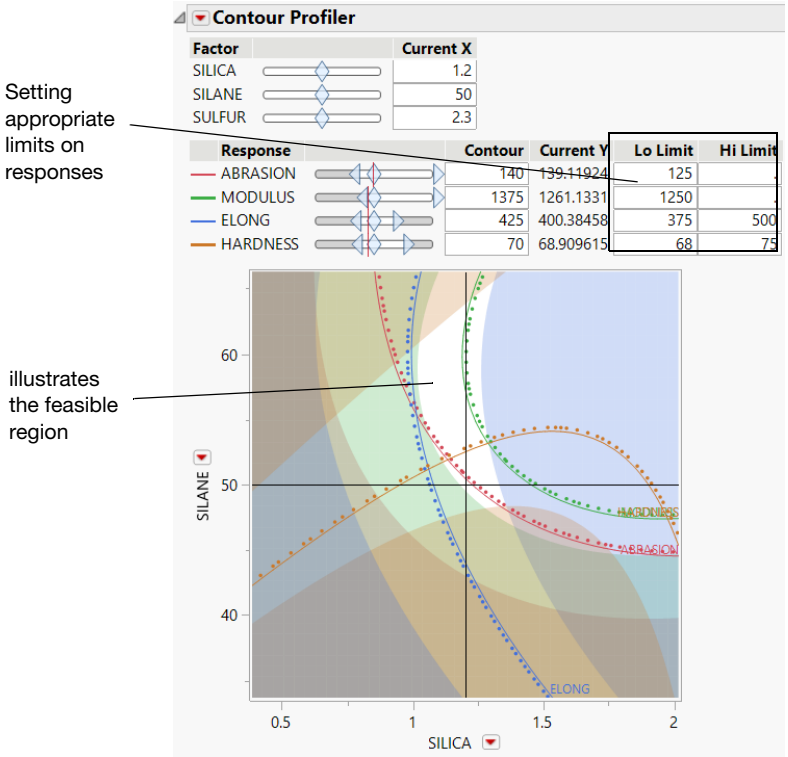
**Hide Y Controls** Shows or hides the Y controls (Response Settings and Controls).

**Multiple Contour Frames** Enables you to add one or more contour plots to the report that represent different combinations of factor settings. When you select this option, the Multiple Contour Frames dialog appears. Use the dialog to specify additional individual contour plots or a matrix of contour plots. You can specify a lower triangular, upper triangular, or full square matrix of contour frames. By default, the new contour plots will be added to the current contour plots in the report. Click the **Remove previous frames** check box to remove the current contour plots and replace them with your new selections.

# Constraint Shading Settings

In the Contour Profiler, 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.

**Figure 4.5** Settings for Contour Shading

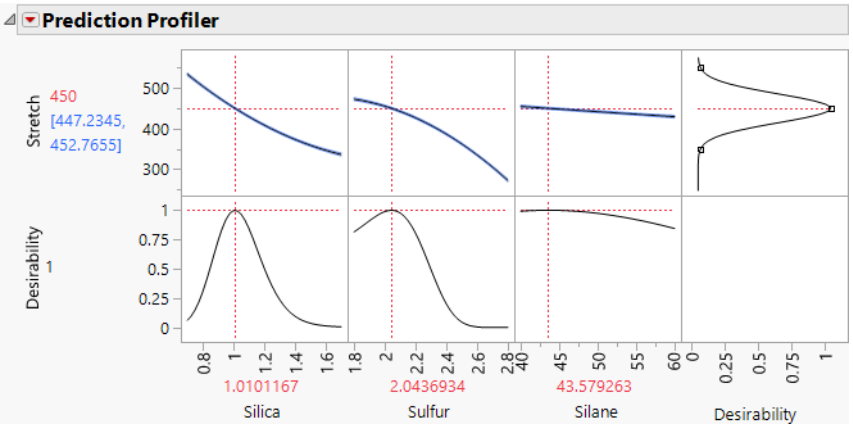



# Additional Example of the Contour Profiler

This example explores optimal factor settings by determining the value of the response that maximizes the desirability functions. 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.

1. Select **Help > Sample Data Folder** 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 and select **Optimization and Desirability > Maximize Desirability**.

**Figure 4.6** Maximize Desirability Settings

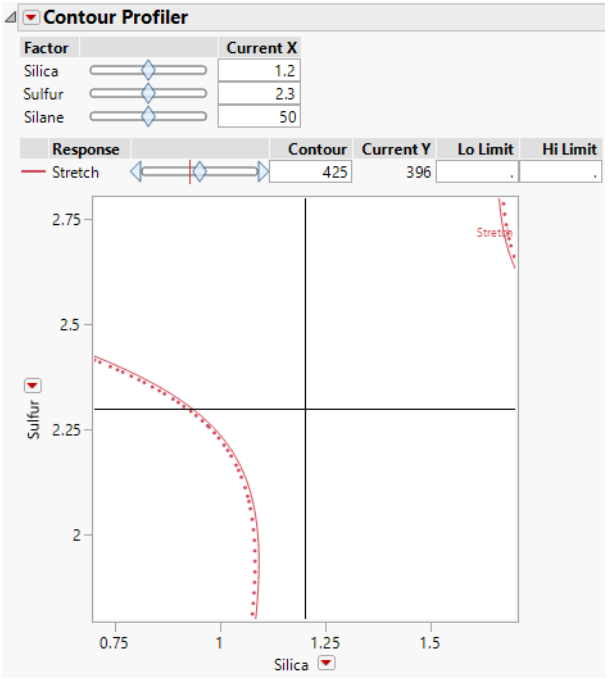



The value of Stretch that maximizes the desirability function is 450. One combination of factor level settings that results in a predicted Stretch of 450 is Silica = 1.0101, Sulfur = 2.0437, and Silane = 43.5792.

4. Click the Response Stretch red triangle 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. In addition, you would like a high level of Silane, a fixed value of 60.

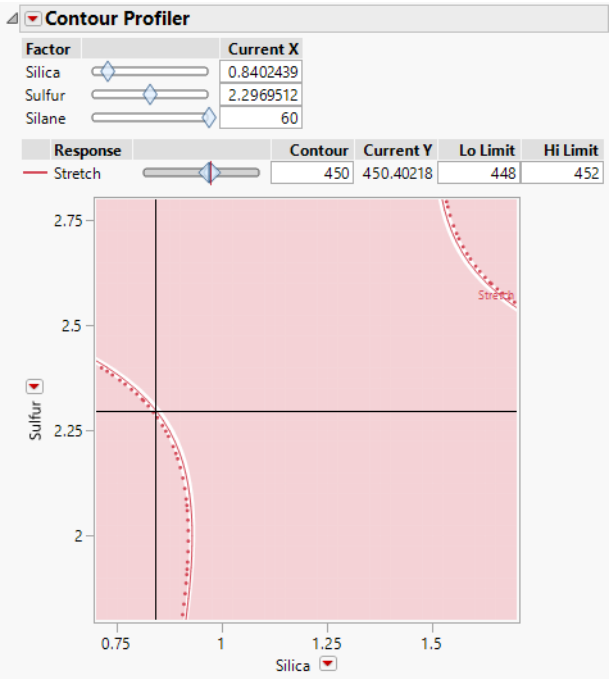
**Figure 4.7** Contour Profiler for Bounce Data.jmp




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.

Chapter **5**

# Surface Plot

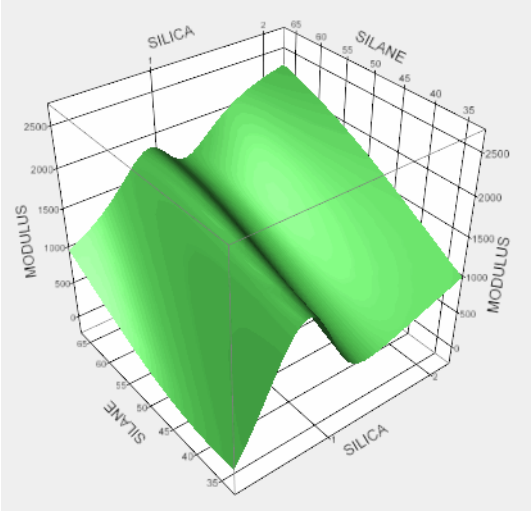
## Explore Contours of Responses across Three Factors

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



## Contents

Overview of the Surface Plot Profiler . . . . . 115
Example of the Surface Plot . . . . . 115
Launch the Surface Plot . . . . . 118
The Surface Plot Report . . . . . 120
The Surface Plot . . . . . 120
Appearance Controls . . . . . 121
Independent Variables . . . . . 122
Dependent Variables . . . . . 123
Surface Plot Platform Options . . . . . 125
Right-Click Menu Options . . . . . 125
Dependent Variables Options . . . . . 126
Sheet or Surface Properties . . . . . 127
Keyboard Shortcuts . . . . . 129
Additional Examples of the Surface Plot . . . . . 129
Example of a Surface Plot for a Single Mathematical Function . . . . . 129
Example of Isosurfaces . . . . . 130

# Overview of the Surface Plot Profiler

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

Surface plots are available directly (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 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.

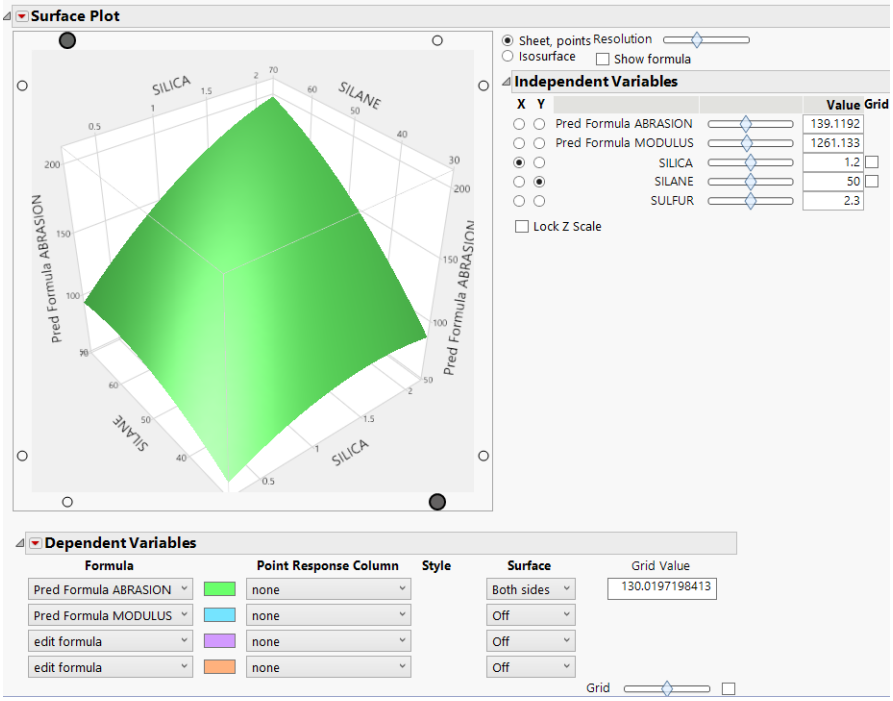
The Surface Plot is built using the 3D scene commands from the JMP Scripting Language (JSL). When you right-click in the surface plot, the pop-up menu contains Open GL Scene commands. For more information about the Open GL Scene commands, see the *Scripting Guide*.

# Example of the Surface Plot

In this example, you use prediction formula columns to examine the surfaces of two response variables.

1. Select **Help > Sample Data Folder** and open Tiretread.jmp.
2. Select **Graph > Surface Plot**.
3. Select Pred Formula ABRASION and Pred Formula MODULUS, click **Columns**, and then click **OK**.

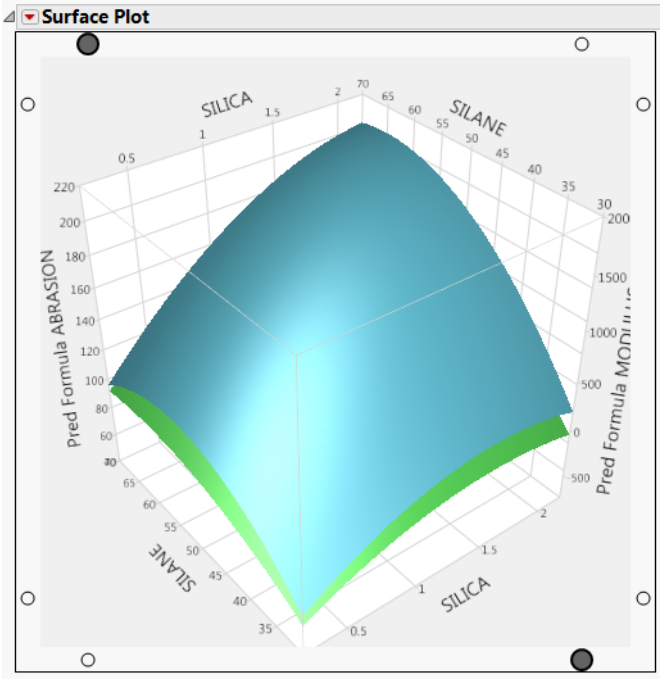
**Figure 5.2** Initial Surface Plot Report




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**.

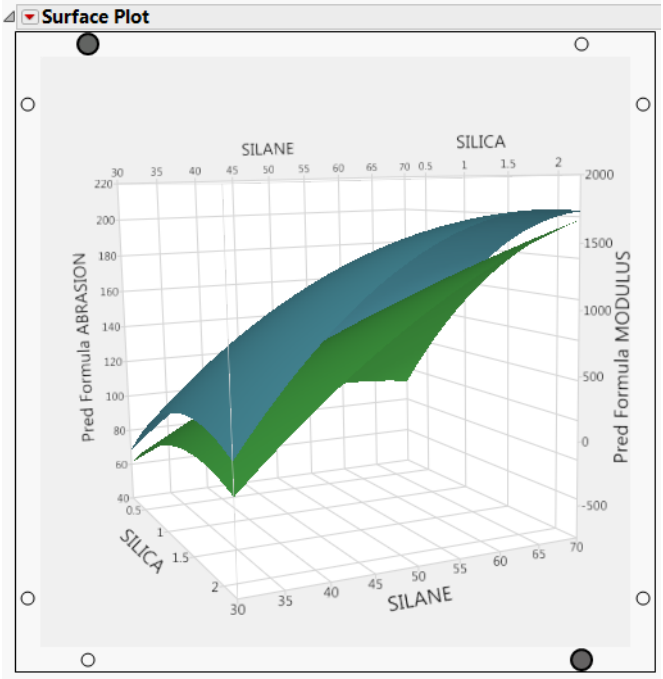
**Figure 5.3** Prediction Surface for ABRASION and MODULUS




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 5.4, 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 5.4** Rotated Prediction Surfaces

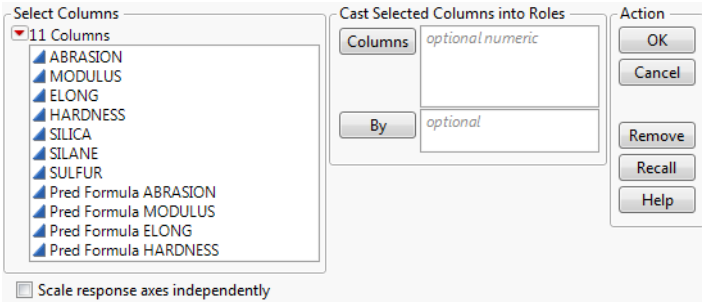



# Launch the Surface Plot

The Surface Plot Profiler can be accessed from the Graph menu or the red triangle menus of certain platforms.

- To launch the Surface Plot Profiler directly, select Surface Plot from the Graph menu. If there is a data table open, this displays the window in Figure 5.5. 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 5.13.
- 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" for more information 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 5.5** 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.

---

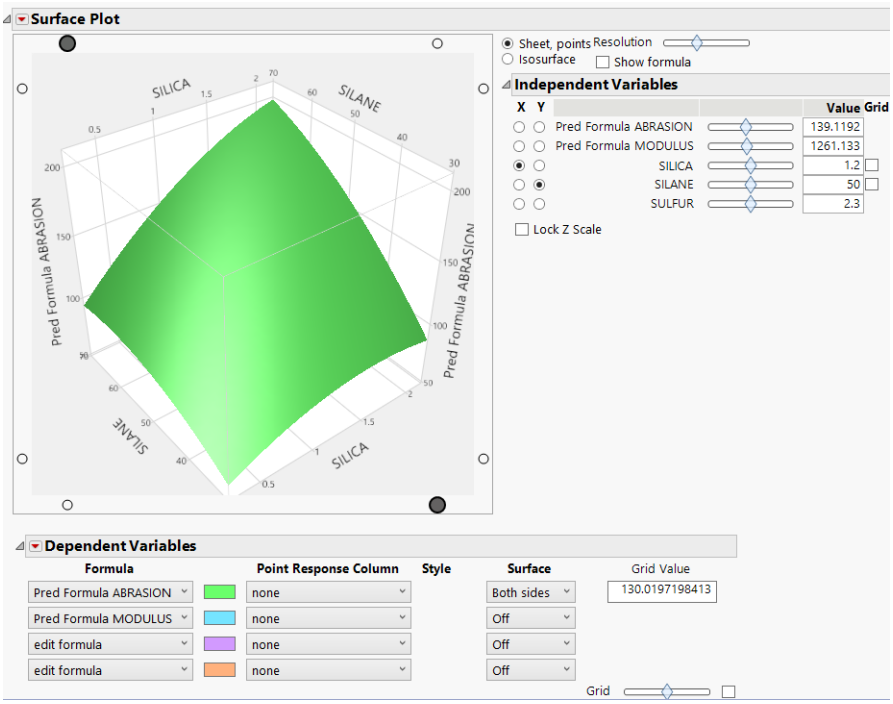
**Note:** When not selected, the axis scale for all responses is the same as the scale for the first item entered in the Columns role.

---

# 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 5.6** 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

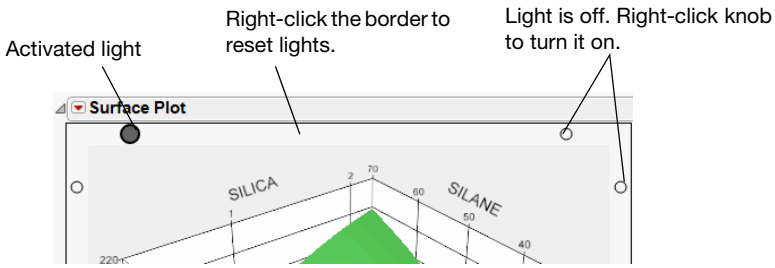
other JMP graphs, the axes can be adjusted, stretched, and compressed using the grabber tool. Hover over an axis to use the grabber tool.

**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 5.7.

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 5.7** 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 5.8 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.

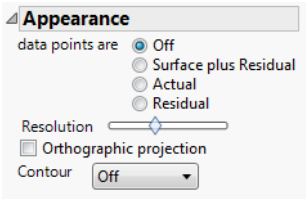
**Isosurface** Changes the display to show isosurfaces. See "Example of Isosurfaces".

**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 5.8** 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 section in the Surface Plot report contains the following controls:

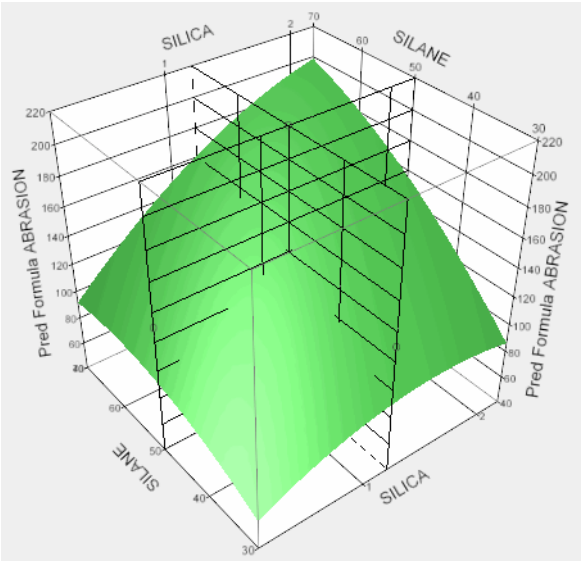
**X, Y** Radio buttons that enable you to select which independent variables are displayed on the $x$- and $y$-axes. This feature is useful when you have more than two independent variables.

**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 5.9 shows a surface with the X and Y grids activated.

**Figure 5.9** Activated X and Y Grids




# Dependent Variables

The Dependent Variables controls that are available in the Surface Plot report 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 5.6.

**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.

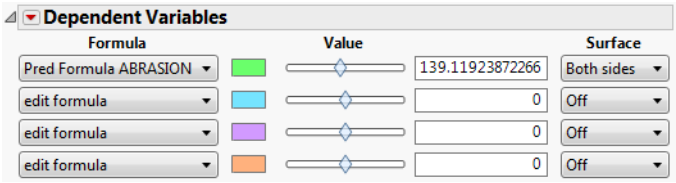
**Surface** Enables you to show or hide the top or bottom of a surface. If **Above only** or **Below only** is selected, the opposite side of the surface is darkened.

**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 text box above the slider.

## Controls for Isosurface

Most of the controls for Isosurface are identical to those of Sheet, points. Figure 5.10 shows the default controls, illustrating the slightly different presentation.

**Figure 5.10** Dependent Variable Controls for Isosurfaces




**Value** Slider and text 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 text 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. See "Launch the Surface Plot".

**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 *Using JMP* 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.

**Platform Preferences** Contains options that enable you to view the current platform preferences or update the platform preferences to match the settings in the current JMP report.

**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.

## Right-Click Menu Options

You can right-click in the plot area of the Surface Plot to reveal the following menu options:

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

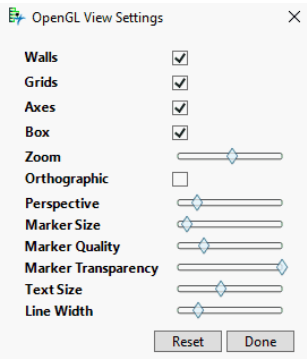
**Surface Properties** (Options appears only if you click the surface.) Displays the Surface Properties window. Available only if Isosurface is selected in the Appearance Controls. See "Sheet or Surface Properties".

**Show Legend** Shows or hides a legend.

**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 OpenGL Scene commands, see the *Scripting Guide*.

**Figure 5.11** OpenGL View Settings




**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

In the Surface Plot report, the Dependent Variables section contains the following red triangle menu 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.

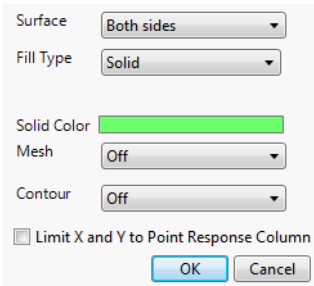
**Response Grid** Shows or hides the Grid controls.

# Sheet or Surface Properties

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.

- If you have selected the Sheet, points option in the Appearance Controls section of the Surface Plot report, you can right-click 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.

**Figure 5.12** 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.

**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 more information about color themes, see *Using JMP*.

**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

This section contains examples of surface plots.

- "Example of a Surface Plot for a Single Mathematical Function"
- "Example of Isosurfaces"

## Example of a Surface Plot for a Single Mathematical Function

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

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

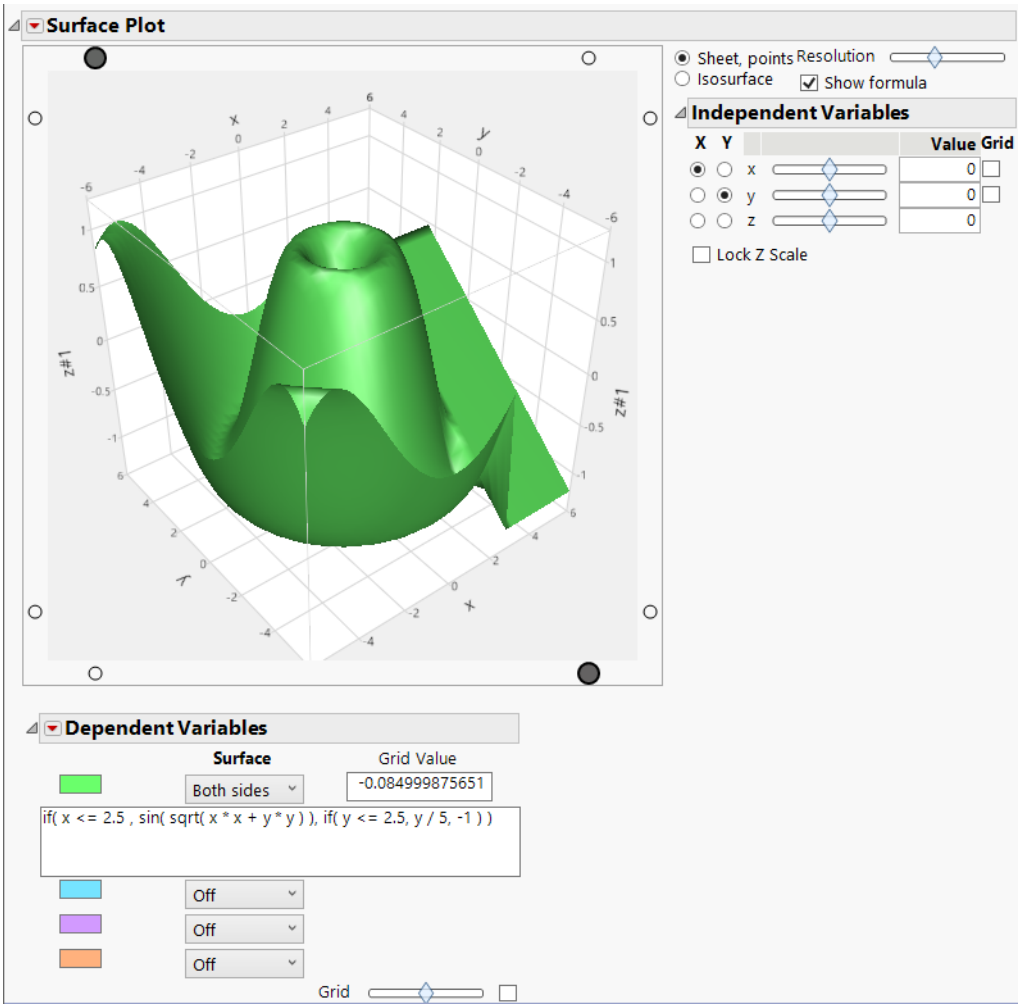
   **Note:** This can be done with or without a data table open.

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 5.13** Default Surface Plot with Formula




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

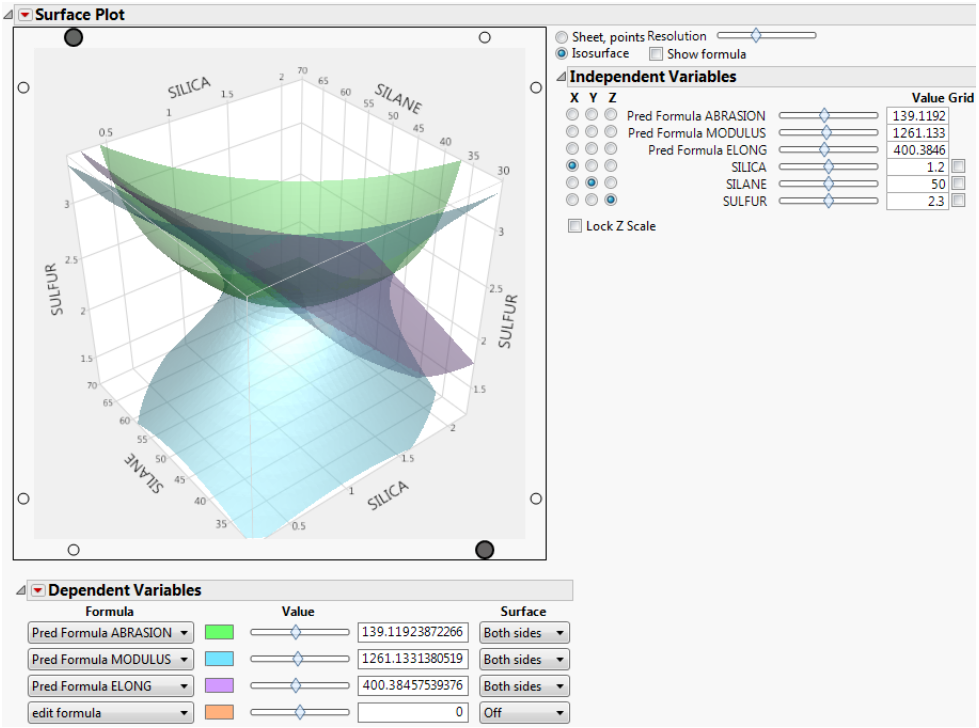
## Example of Isosurfaces

In this example, the goal is to determine which values of a particular variable are inside the limits set by two surfaces. 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.

1. Select **Help > Sample Data Folder** 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 5.14** 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.

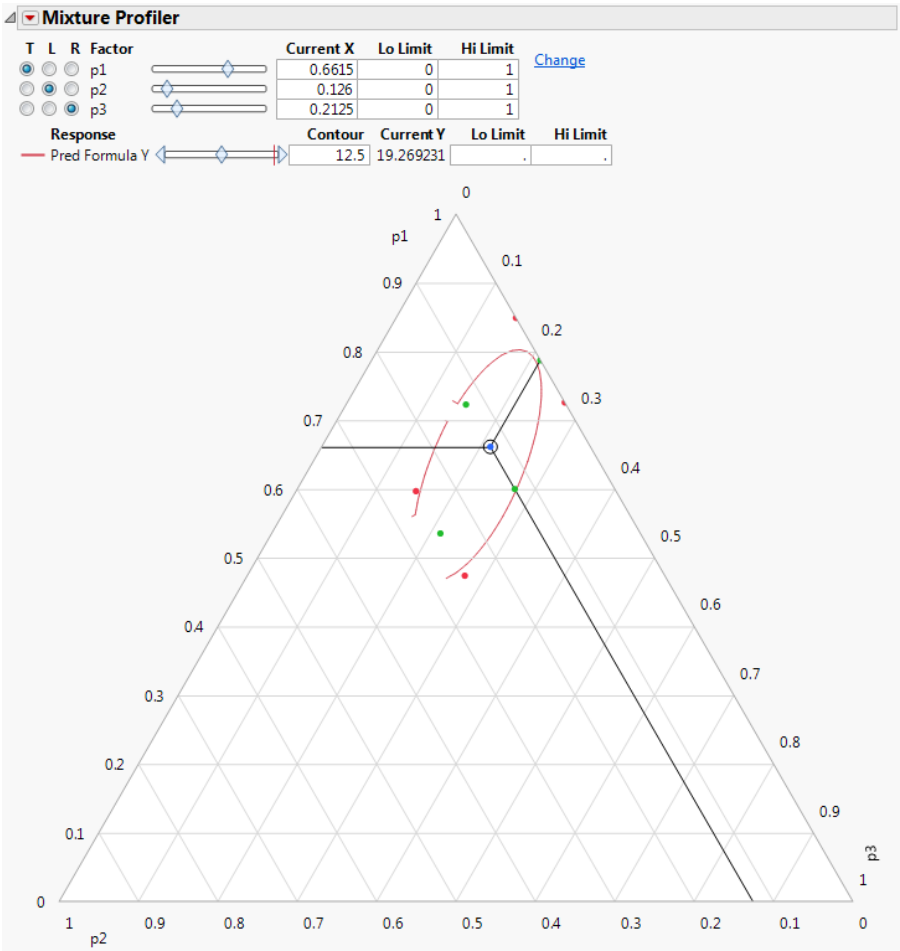
Chapter 6

# 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



## Contents

Overview of the Mixture Profiler . . . . . 135
Ternary Plot Overview . . . . . 135
Ternary Plots with More Than Three Components . . . . . 136
Example of the Mixture Profiler . . . . . 137
Launch the Mixture Profiler Platform . . . . . 138
The Mixture Profiler Report . . . . . 139
Factor Settings and Controls . . . . . 139
Response Settings and Controls . . . . . 140
Mixture Profiler Platform Options . . . . . 140
Customizations for the Mixture Profiler . . . . . 141
Linear Constraints . . . . . 141
Additional Examples of the Mixture Profiler . . . . . 142
Example of Mixture Variables with a Process Variable . . . . . 142
Example of Multiple Responses with Five Mixture Variables . . . . . 145

# Overview of the Mixture Profiler

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.

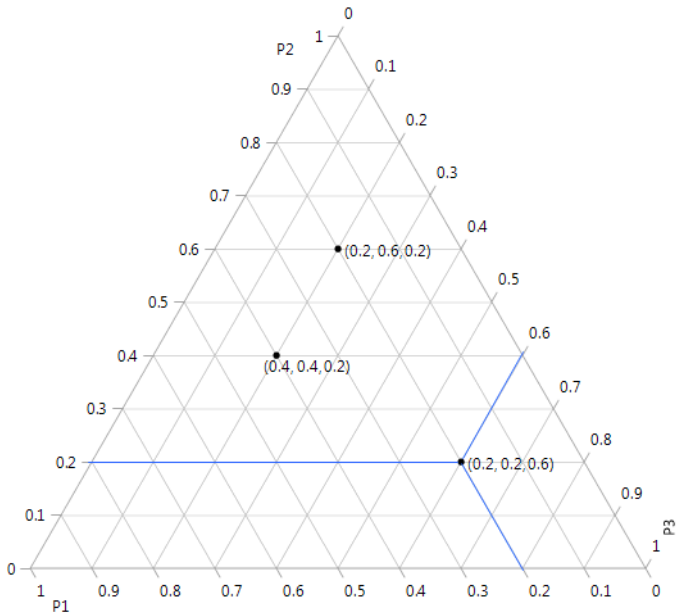
Many of the Mixture Profiler features are similar to those of the Contour Profiler as described in "Contour Profiler Platform Options". Unique to the Mixture Profiler is the use of a ternary plot.

## 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



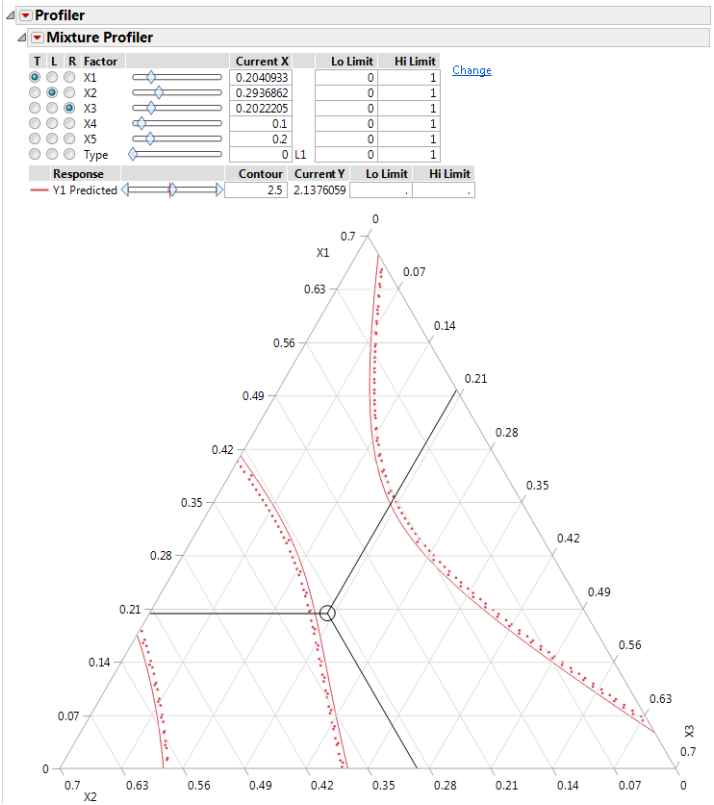

## Ternary Plots with More Than Three Components

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. Note that all Lo Limit values are set to 0 and all Hi Limit values set to 1. 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. 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

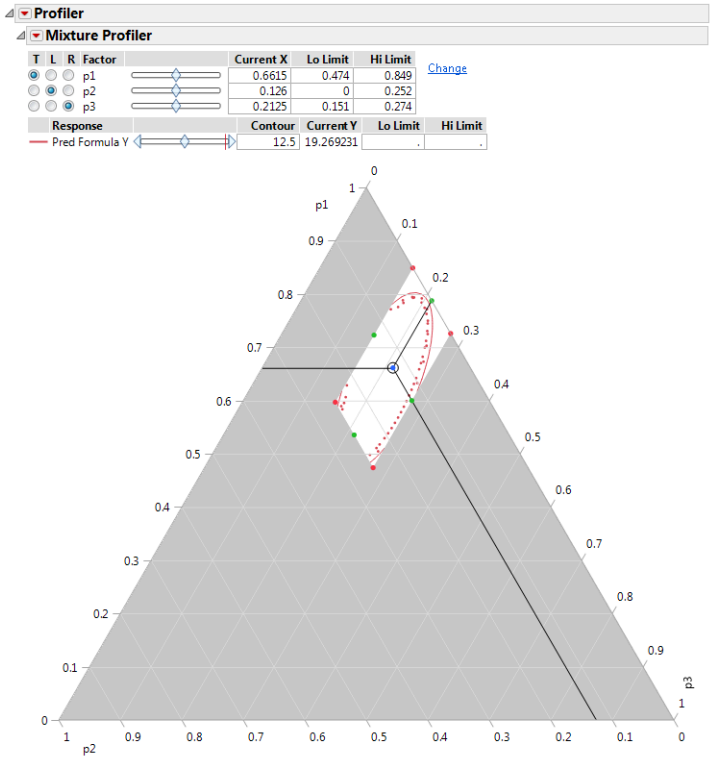
This example explores how the feasible region of a response is impacted by three mixture components. There are constraints on the levels of the mixture components. Use the saved prediction formula column for the response that is in the data table.

1. Select **Help > Sample Data Folder** 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 from the Graph menu or the red triangle menus of certain platforms.

- 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".
- The Mixture Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" for more information 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

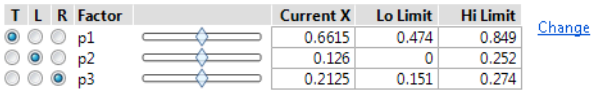
The initial Mixture Profiler report shows a mixture profiler with settings and controls for both factors and responses.

- "Factor Settings and Controls"
- "Response Settings and Controls"

## Factor Settings and Controls

You can change the settings for the factors directly from the Mixture Profiler report.

**Figure 6.5** Factor Settings and Controls



| T | L | R | Factor | Current X | Lo Limit | Hi Limit |
|---|---|---|---|---|---|---|
| ● | ○ | ○ | p1 | 0.6615 | 0.474 | 0.849 |
| ○ | ● | ○ | p2 | 0.126 | 0 | 0.252 |
| ○ | ○ | ● | p3 | 0.2125 | 0.151 | 0.274 |

Change

**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.

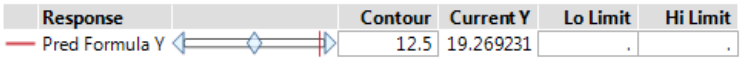
**Note:** When factor limits are used, the profiler shows regions that are feasible as unshaded.

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

## Response Settings and Controls

You can change the settings for the responses directly from the Mixture Profiler report.

**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

The Mixture Profiler red triangle menu contains the following 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 Points** Shows or hides the individual points on the plot.

**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. For more information on this submenu, see the discussion of the profiler in "Factor Settings".

## 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.

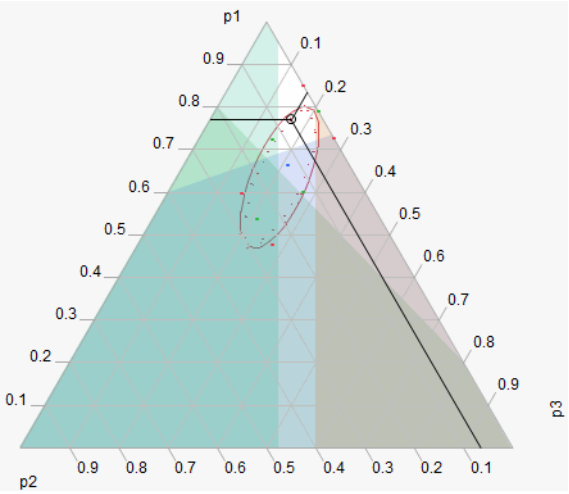
# Linear Constraints

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".

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:

- $4*p2 + p3 \leq 0.8$
- $p2 + 1.5*p3 \leq 0.4$
- $p1 + 2*p2 \geq 0.8$
- $p1 + 2*p2 \leq 0.95$

**Figure 6.7** Shaded Regions Due to Linear Constraints




# Additional Examples of the Mixture Profiler

This section contains examples using a process variable and multiple variables.

- "Example of Mixture Variables with a Process Variable"
- "Example of Multiple Responses with Five Mixture Variables"

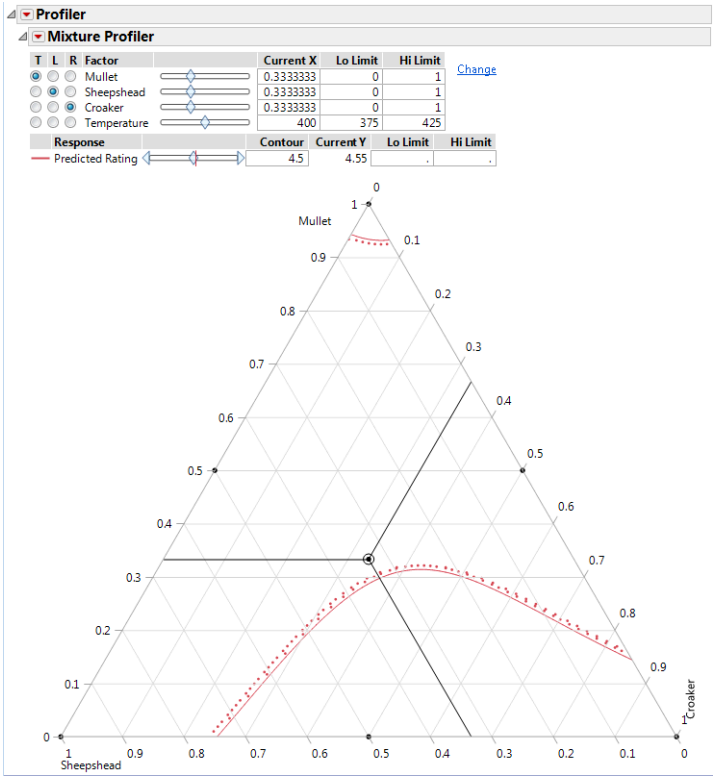
## Example of Mixture Variables with a Process Variable

This example uses the prediction formula from a response surface model to create a Mixture Profiler. You set a feasible region that achieves a certain response and explore the different combinations of factor values that are within the feasible region.

The data comes from an experiment to optimize the texture of fish patties. There are three mixture components that each represent the proportion of a fish type in the fish patty. There is also a process variable that is the oven temperature used to bake the patties. The response is a measure of texture acceptability, where higher is better.

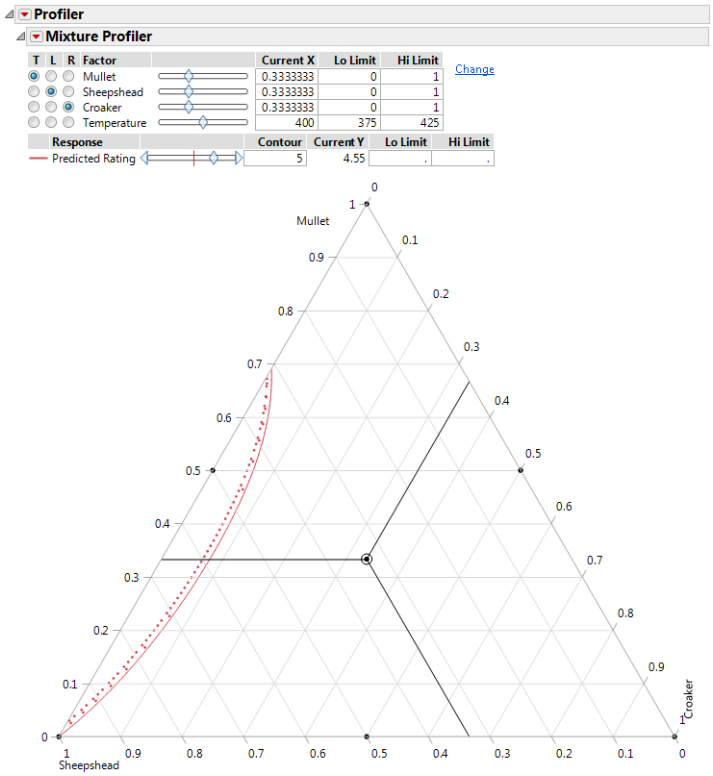
1. Select **Help > Sample Data Folder** and open Fish Patty.jmp.
2. Select **Graph > Mixture Profiler**.
3. Select Predicted Rating and click **Y, Prediction Formula**.
4. Click **OK**.

**Figure 6.8** Initial Output for Mixture Profiler




5. 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.

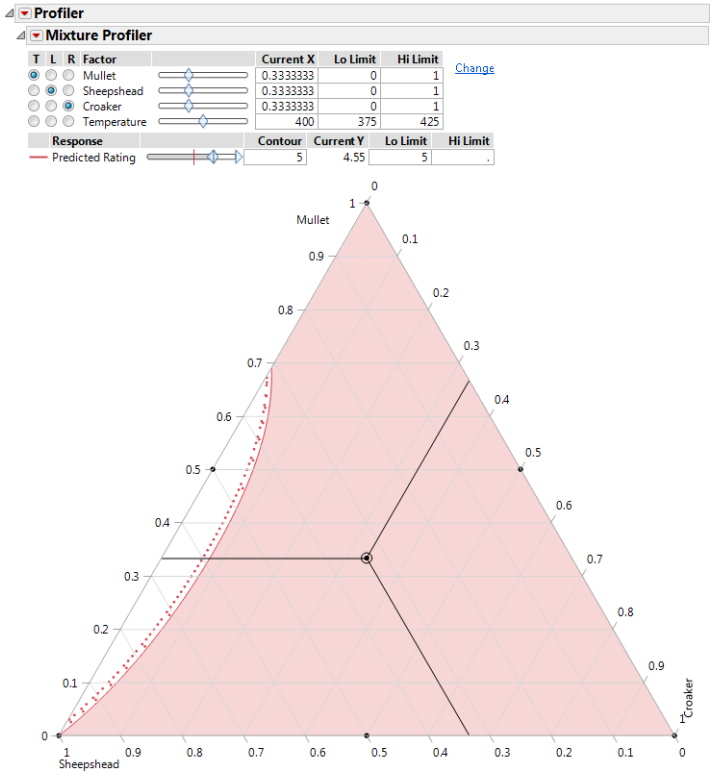
**Figure 6.9** Contour Showing a Predicted Rating of 5




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

6. 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.




7. 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 "Custom Profiler" or "Desirability Profiling and Optimization".
- Simulate the response as a function of the random variation in the factors and model noise. See "Simulator".

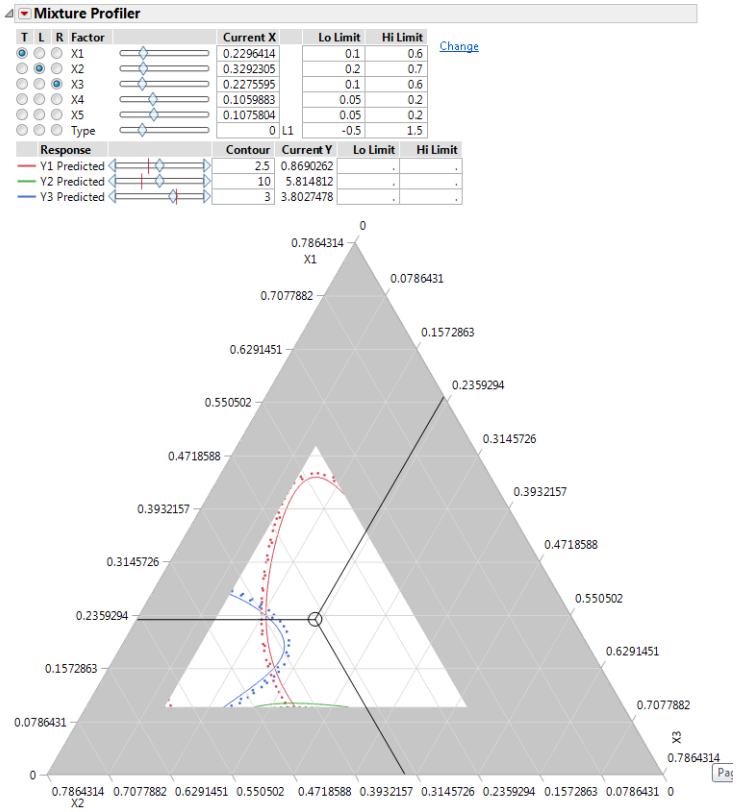
## Example of Multiple Responses with Five Mixture Variables

This example uses the Mixture Profiler to visualize the feasible region of five mixture components after setting limits for the three responses. Prediction formulas from a response surface model are used.

1. Select **Help > Sample Data Folder** 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.

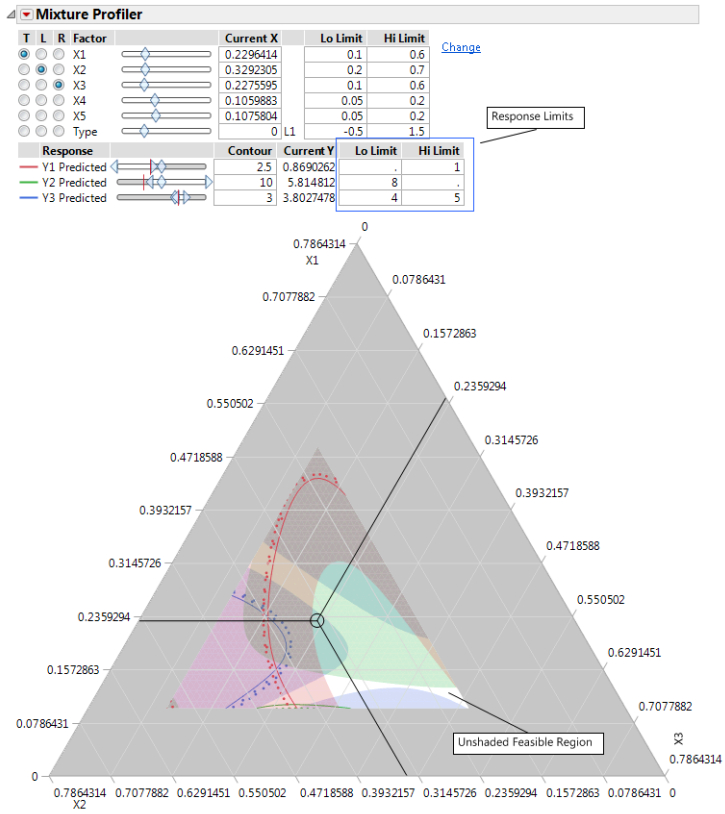
- The on axis factors are X1, X2, and X3 as indicated by the axes labels and the radio buttons in the factor settings. All of the factors have low and high limits, which were entered previously as Column Properties. See *Using JMP* for more information about entering column properties. Alternatively, 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.

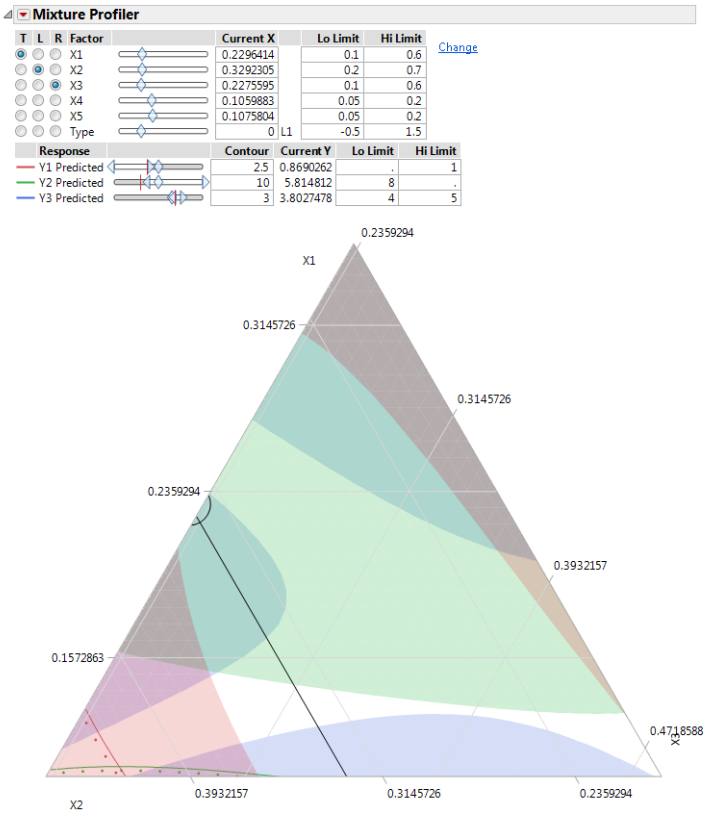
6. Enter 1 in the Y1 Predicted Hi Limit edit box. Enter 8 in the Y2 Predicted Lo Limit edit box. Enter 4 in theY3 Predicted Lo Limit edit box and 5 in the Y3 Predicted Hi Limit edit box.
7. The feasible region remains white (unshaded). Use the Response slider controls to position the contours in the 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

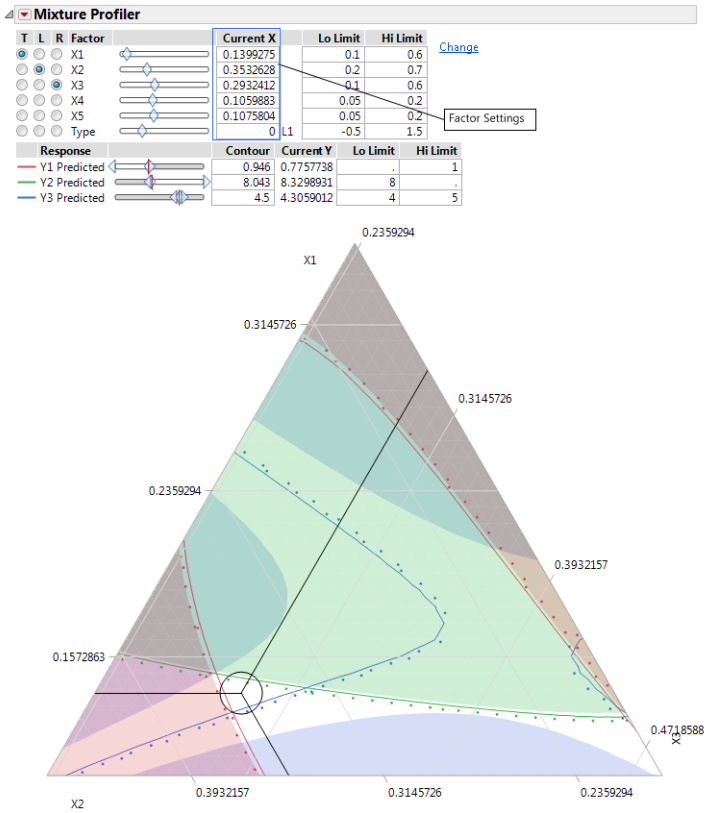



The manufacturer wants to maximize Y1, minimize Y2, and have Y3 at 4.5.

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.

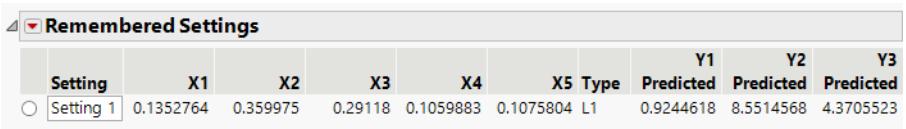
The resulting three contours do not all intersect at one spot. Y1 and Y2 cannot be optimized with Y3 on target, so you have to compromise. Position the three-way crosshairs in the middle of the contours to explore the factor levels that produce those response values. Note that these response contours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



**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.

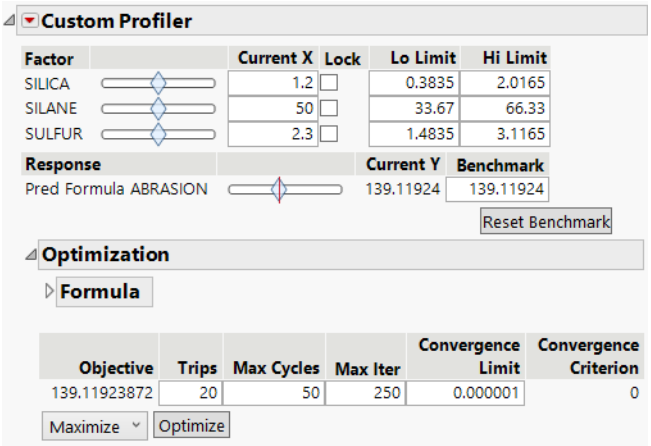
Chapter 7

# 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



## Contents

Overview of the Custom Profiler . . . . . 153
Example of the Custom Profiler . . . . . 153
Launch the Custom Profiler Platform . . . . . 155
The Custom Profiler Report. . . . . 156
Factor Settings and Controls . . . . . 156
Response Settings and Controls . . . . . 157
Optimization Settings and Controls. . . . . 157
Custom Profiler Platform Options . . . . . 158

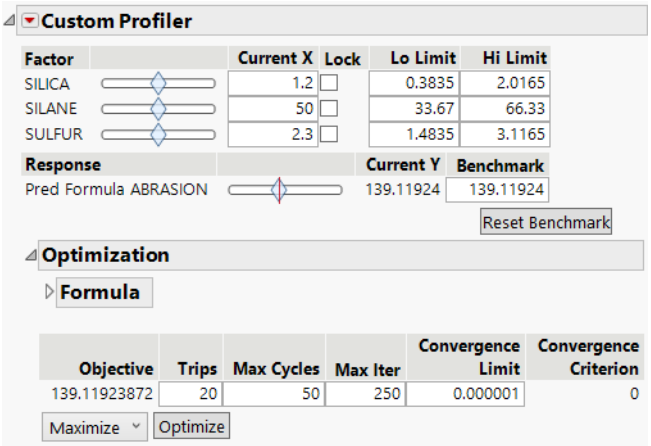
# Overview of the Custom Profiler

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

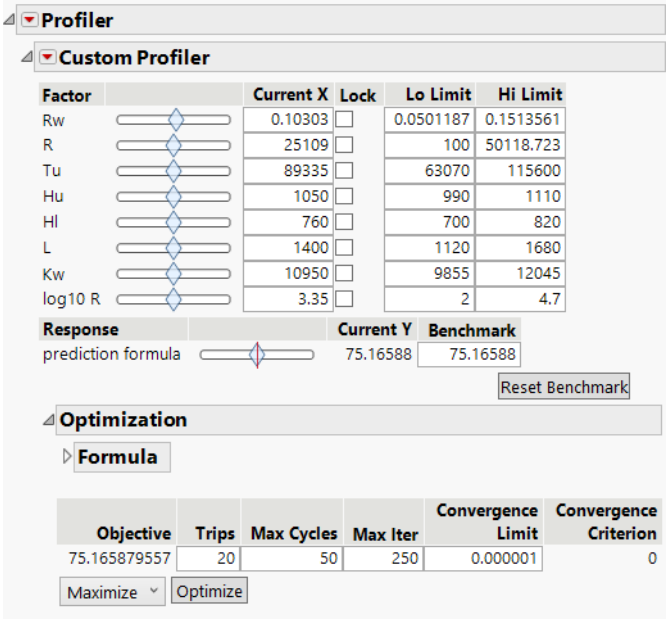



# Example of the Custom Profiler

In this example, you want to optimize the predicted value of the response. The data demonstrates the flow of water through a borehole that is drilled from the ground surface through two aquifers.

1. Select **Help > Sample Data Folder** and open Design Experiment/Borehole Latin Hypercube.jmp
2. Select **Graph > Custom Profiler**.
3. Select prediction formula and click **Y, Prediction Formula.**
4. Click **OK**.

**Figure 7.3** Custom Profiler Report




Note that the Benchmark value is 75.17. The Current Y value is also 75.17. This value is the predicted response with all factors set to their mean values.

You would like to find the X values of the factors that optimize the predicted value. However, of the factors in the model, only those related to the borehole can be manipulated in practice. The other factors are out of human control. You can lock the uncontrollable factors at their mean values and optimize the factors that can be controlled.

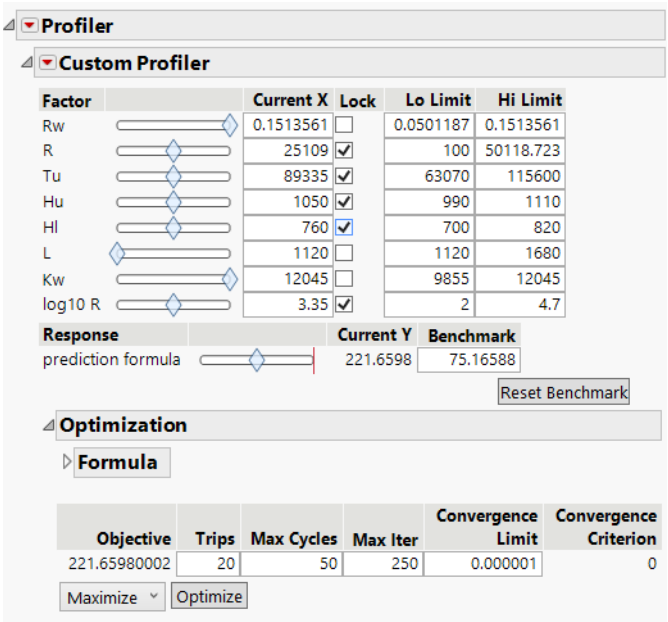
---

**Note:** To view the description of a factor, right-click the corresponding column in the data table and select **Column Info**.

---

5. In the Custom Profiler Report, select the Lock check box next to R, Tu, Hu, Hl, and log10 R.
6. In the Custom Profiler Report, click **Optimize**.

**Figure 7.4** Optimized Custom Profiler Report




The optimization routine found an optimum predicted response at 221.66. In order to obtain the optimum of 221.66, Rw and Kw are set to their maximum values and L is set to its minimum value. The optimum is greater than the initial Benchmark value of 75.17.

# Launch the Custom Profiler Platform

The Custom Profiler can be accessed from the Graph menu or the red triangle menus of certain platforms.

- 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".
- The Custom Profiler can be accessed as a red triangle menu option in many modeling platforms. See "Where to Find JMP Profilers" for more information 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"
- "Response Settings and Controls"
- "Optimization Settings and Controls"

## Factor Settings and Controls

You can change the settings for the factors directly from the Custom Profiler report.

**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

You can change the settings for the responses directly from the Custom Profiler report.

**Figure 7.6** Response Settings and Controls

| Response | | Current Y | Benchmark |
|---|---|---|---|
| prediction formula | | 311.17206 | 75.16588 |

Reset Benchmark

**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.

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

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

## Optimization Settings and Controls

The Custom Profiler report includes an Optimization section that enables you to specify different criterion to optimize the prediction formula.

**Figure 7.7** Optimization Settings and Controls

**Optimization**

**Formula**

| Objective | Trips | Max Cycles | Max Iter | Convergence Limit | Convergence Criterion |
|---|---|---|---|---|---|
| 311.17206129 | 20 | 50 | 250 | 0.000001 | 0 |

Maximize Optimize

**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".

**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

The Custom Profiler red triangle menu contains the following options:

**Factor Settings** Contains options identical to the Factor Settings submenu in the Prediction Profiler. See "Factor Settings".

**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".

**Save Linear Constraints** Saves existing linear constraints as a data table script that is named Constraint. See "Linear Constraints".

**Simulator** Launches the Simulator. See "Simulator".

# Chapter 8

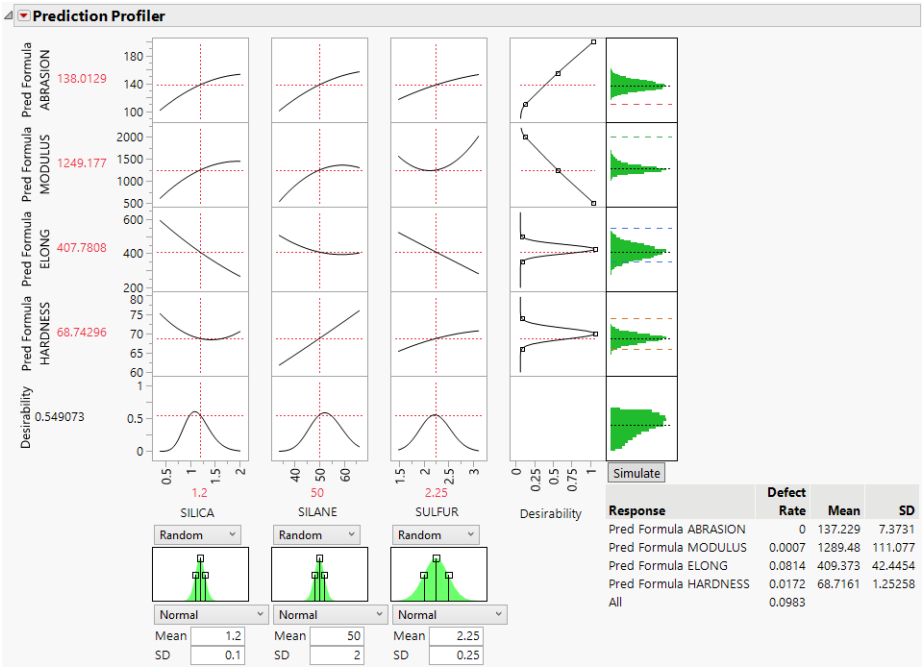
# Simulator

## 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 model inputs and noise. The Simulator in the profilers enables you to define random inputs, run simulations, and produce output tables of simulated values.

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

**Figure 8.1** Prediction Profiler with Simulator Example



## Contents

Overview of the Simulator . . . . . 161
Example of the Simulator . . . . . 161
Launch the Simulator . . . . . 164
Simulation Settings for Factors . . . . . 164
Simulation Settings for Responses . . . . . 167
The Simulator Report . . . . . 168
The Simulator Report Options . . . . . 169
Defect Profiler . . . . . 170
Defect Parametric Profile . . . . . 172
Simulation Experiment . . . . . 173
Specification Limits . . . . . 174
Additional Examples of the Simulator . . . . . 174
Example of the Defect Profiler . . . . . 174
Example of Stochastic Optimization . . . . . 178
Example of Simulating General Formulas . . . . . 187
Statistical Details for the Simulator . . . . . 190

# Overview of the Simulator

Simulation enables you to discover the distribution of model outputs as a function of the random variation in the model inputs and noise. The Simulator in the profilers enables you to define random inputs, run simulations, and produce output tables of simulated values.

For example, you can use the simulator to model the defect rate of a process and to explore the robustness of the defect rate with respect to variation in the model factors. If specification limits have been set for the responses, they are displayed in the simulation output. The inclusion of specification limits enables you to perform a prospective capability analysis of the simulated model. For information about adding specification limits, see *Using JMP*.

You can access the Simulator through the Prediction, Contour, and Custom Profilers. The appearance of the Simulator depends on the profiler platform in which it appears.

### Prediction Profiler

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

### Contour Profiler and Custom Profiler

In the Contour and Custom profilers, the Simulator is not integrated into the graphical layout. There are no integrated histograms, and the interface is textual. However, the process is the same and the resulting output tables are the same as in other profilers.

**Note:** The 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.

# Example of the Simulator

Launch the Simulator through the Prediction Profiler. Then, use sequencing to examine how the distribution of the response changes when the mean (sequencing location) and variability (sequencing spread) of the inputs change.

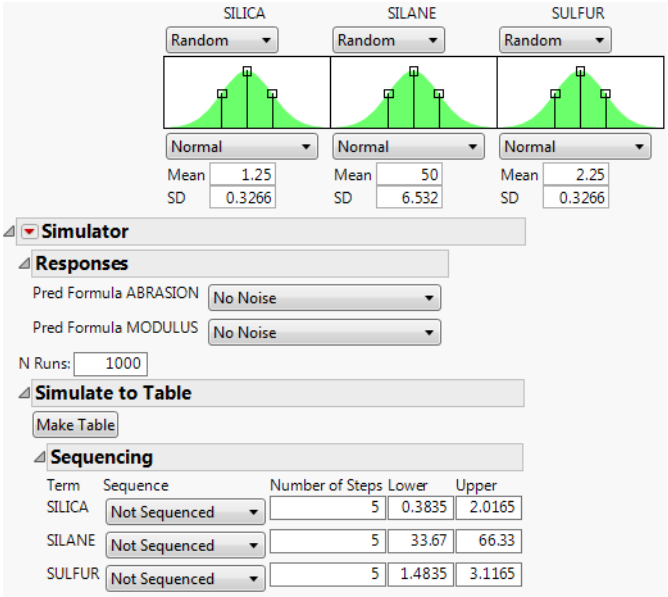
1. Select **Help > Sample Data Folder** and open Tiretread.jmp.
2. Select **Graph > Profiler**.
3. Select Pred Formula ABRASION and Pred Formula MODULUS and click **Y, Prediction Formula**.

4. Click **OK**.
5. Click the Prediction Profiler red triangle and select **Simulator**.
6. Beneath each factor select **Random**.

**Note:** The default random distribution is a normal distribution.

7. Change the **N Runs** value to 1000.
8. Open Simulate to Table outline and then open the Sequencing outline.

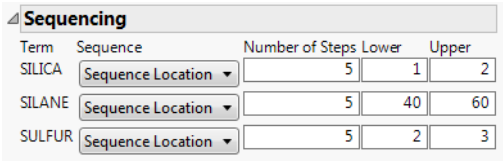
**Figure 8.2** Simulator Settings




You want to examine how the responses change as the input factors change locations. To explore changes in factor mean, use sequence location. To instead explore changes in the variability of the factor, use sequence spread.

9. For SILICA, select **Sequence Location**. Leave the number of steps set to 5. Set the **Lower** limit to 1 and the **Upper** limit to 2.
10. For SILANE, select **Sequence Location**. Leave the number of steps set to 5. Set the **Lower** limit to 40 and the **Upper** limit to 60.
11. For SULFUR, select **Sequence Location**. Leave the number of steps set to 5. Set the values **Lower** limit to 2 and the **Upper** limit to 3.

**Figure 8.3** Sequencing Settings

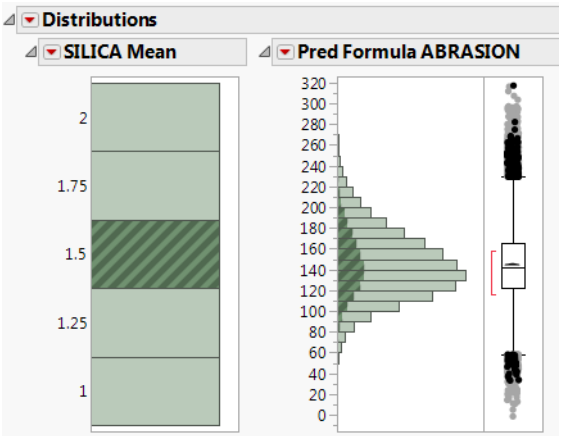



12. Click **Make Table**.

    The SILICA Mean, SILANE Mean, and SULFUR Mean columns contain five steps across their respective range of values. For example, the Silica Mean values are 1, 1.25, 1.5, 1.75, and 2. The SILICA, SILANE, and SULFUR columns are the simulated values from a normal distribution with a mean defined by the corresponding mean columns and a fixed standard deviation. Pred Formula ABRASION and Pred Formula MODULUS values are calculated for set of simulated values, so that you can explore how the responses change as the factor values change.

13. Select **Analyze > Distribution**.
14. Select SILICA Mean and Pred Formula ABRASION and click **Y, Columns**.
15. Click **OK**.

**Figure 8.4** Distribution of SILICA Mean by Pred Formula ABRASION




Click a histogram bar that corresponds to a SILICA Mean to see how the predicted value for abrasion varies across the simulated silane and sulfur ranges given the selected value of silica.

# Launch the Simulator

The Simulator can be accessed as a red triangle menu option in the Prediction Profiler, the Contour Profiler, and the Custom Profiler. When you launch the Simulator from the Prediction Profiler, factor specifications are displayed below each factor's profile. When you launch the Simulator from the Contour or Custom profilers, factor specifications appear in a Factors report, located below the respective profiler report. See "Simulation Settings for Factors".

Once the Simulator is accessed, you can define simulation parameters for the responses in the Responses report. See "Simulation Settings for Responses". The Responses report also enables you to set the number of simulations.

A histogram is also added to the right of each response when the Simulator is accessed from the Prediction Profiler. Below the histograms is a **Simulate** button. Once you specify settings for the factors and responses, click **Simulate** to simulate response values. The histograms of the simulated responses are updated each time the simulate button is clicked. To simulate responses in the Contour and Custom profilers, open the Simulate to Table report and click **Make Table**. See "The Simulator Report".

## Simulation Settings for Factors

By default, factors (inputs) in the Simulator are defined by the functions that you select when you launch a profiler. There are options for changing the location, error, and noise that enable you to modify how your factors are simulated.

**Fixed** Fixes the factor at the current value in the profiler for all simulation runs.

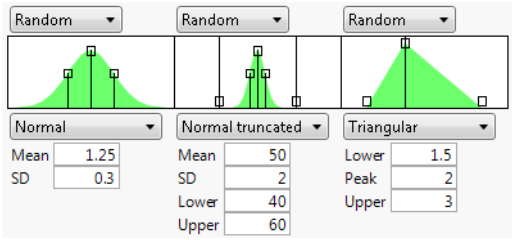
**Random** Draws a random value of the factor for each simulation run from the specified distribution and distributional parameters. When you select Random, distribution options appear. For more information about the options, see "Continuous Factors" and "Categorical Factors".

---

**Note:** The default random distribution is a normal distribution with a mean at the current factor setting and a standard deviation estimated by the range of the factor divided by 5. When mixture factors are involved, the standard deviation is estimated by the range of the mixture limits divided by 5. The mean of the normal distribution is automatically updated if you change the factor setting in the profiler. There are other random distributions that also update parameters automatically when the factor settings are changed. The Triangular distribution updates all three parameters, the Cauchy distribution updates the mode, and the Johnson Su distribution updates θ.

---

**Figure 8.5** Example of Three Random Distributions



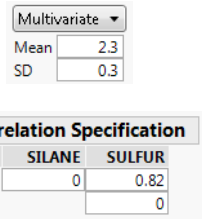

**Expression** Generates values for the factor based on a JSL expression. This gives you flexibility to use a random distribution of your choice. For example, you could create a censored normal distribution that guaranteed nonnegative values with an expression such as the following:

```
Max(0,RandomNormal(5,2))
```

After entering the expression, click the **Reset** button to submit the expression.

**Multivariate** Generates values for the factor based on a multivariate normal distribution. Specify the mean and standard deviation for each factor. The correlation matrix is defined separately in the X Correlation Specification report. You can use this option to accommodate correlated factors.

**Figure 8.6** Defining a Correlation Matrix for the Multivariate Option




**Tip:** To specify the same setting for all of the factors, press Ctrl and select the desired setting.

## Continuous Factors

For continuous factors the random distributions available use standard random functions that are described in *Using JMP*. Descriptions of the specialized distributions available follow:

**Normal weighted** Generates values from a weighted normal distribution with the given mean and standard deviation. The weighting is a specific sampling scheme used to simulate rare events from the tails of the distribution. This is a good choice when you want to simulate very low defect rates. See "Statistical Details for the Simulator".

**Normal truncated** Generates values from a normal distribution limited by lower and upper limits. Any randomly generated value that exceeds a limit is discarded and another value is generated. This is a good choice when you want to simulate an inspection system where inputs that do not satisfy specification limits are discarded.

**Normal censored** Generates values from a normal distribution limited by lower and upper limits. Any randomly generated value that exceeds a limit is set to that limit, putting a density mass at the limits. This is a good choice when you want to simulate systems where inputs that do not satisfy specification limits are reworked until they are at the specification limit.

**Sampled** Generates factor values by selecting a value at random from that factor's column in the data table.

**External** Generates factor values by selecting a value at random from a column in another table. You are prompted to choose the table and column.

**Tip:** To specify the same distribution for all of the factors, press Ctrl and select the desired distribution.

**Note:** The Aligned check box appears when you select Sampled or External. It is used when two or more factors are set to Sampled or External. When checked, the random draws come from the same row of the table. This maintains the correlation structure between two columns. If the Aligned option is used to associate two columns in different tables, the columns must have the same number of rows.

## Categorical Factors

For a categorical factor the random distribution is characterized by probabilities specified for each category. By default, the probabilities are set to the observed probabilities in the data table. You can change the probabilities using the handles in the plot or by changing the numerical values in the Prob column. The probabilities must sum to 1.

**Note:** If the probabilities that you select do not sum to 1, they are automatically regularized to sum to 1 once you click the Simulate button.

When the simulation for a factor is set to Random, a graphical representation of the random distribution density is shown. The graph 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.

### Mixture Factors

For mixture factors there are two random distributions available. Mixture factors must be generated from the same random distribution. However, there can be a combination of fixed and random mixture factors.

**Mixture Uniform** Generates random uniform points over a convex polytope, given the specified lower and upper bounds for each factor. This ensures that each generated row upholds the mixture constraints. This option uses the Polytope Uniform Random JSL function. See the *JSL Syntax Reference*.

**Mixture Dirichlet** Generates random points from a Dirichlet distribution with the specified α parameter for each factor. The probability density function of the Dirichlet distribution is calculated as follows:

$$\frac{1}{B(\alpha)} \prod_{i=1}^{K} x_i^{(\alpha_i - 1)}, x_i > 0, \sum_{i=1}^{K} x_i = 1$$

where B(α) is the multivariate beta function, which acts as a normalizing constant. The Mixture Dirichlet option is available only if the bounds on the factors create a mixture region that is a simplex, or generalized triangle.

## Simulation Settings for Responses

You can use the Simulator to simulate random noise in responses. Often your factors explain part of the variation in the response with the rest of the variation of the response attributed to random noise. You can specify the random noise using the following options:

**No Noise** Generates the response from the model based on the factor settings. No random noise added to the response.

**Add Random Noise** Adds a normal random value with mean zero and specified standard deviation to the generated response. If the response column contains a Predicting column property that includes a Std Dev value, the field used to specify the standard deviation in the Simulator is automatically populated with the value from the column property.

**Add Random Weighted Noise** Adds an error term to the generated response based on a weighted sampling scheme used to simulate rare events from the tails of the distribution. If the response column contains a Predicting column property that includes a Std Dev value, the field used to specify the standard deviation in the Simulator is automatically populated with the value from the column property.

**Add Multivariate Noise** Adds an error term to the generated response based on a multivariate normal distribution. Specify the standard deviation for the response. Specify

the correlation structure between responses in the Y Correlation Specification report. If the response column contains a Predicting column property that includes a Std Dev value, the field used to specify the standard deviation in the Simulator is automatically populated with the value from the column property.

**Random by Model** (Available only within the Standard Least Squares and Generalized Regression personalities of the Fit Model platform.) Adds an error term to the generated response based on the model distribution specified in the Fit Model launch window.

**Tip:** To specify the same noise setting for all of the responses, press Ctrl and select the desired noise setting.

# The Simulator Report

Once you launch the Simulator from one of the profilers, the Simulator report contains the following items:

**N Runs** Enables you to specify the number of runs for the simulation.

**Simulate to Table** Contains a Make Table button that 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 whether the simulated response is within the specification limits. The table also includes an Objective column, labeled Obj. If you added Desirability functions to your profiler prior to simulation, this column is the evaluation of the desirability function. If the Desirability functions were not present prior to simulation, the column contains all missing values.

**Sequencing** (Available only when the distribution is Normal, Uniform, or Triangular random.) Sequencing enables you to simulate across a range of locations or spreads for each factor. The simulations are saved to a data table. Select an option to sequence the location or spread of each factor. Specify the number of steps and the range for the sequencing. For each combination of the number of steps, the simulation runs N Runs. If you have two terms, each with 5 steps, and N Runs = 100, the resulting simulation table has 2,500 rows.

# The Simulator Report Options

Once you launch the Simulator from one of the profilers, the Simulator red triangle menu contains the following options:

**Automatic Histogram Update** Toggles the histogram update. When this option is selected, the histograms update with new simulated values when you drag factor setting handles. By default, this option is not selected.

**Defect Profiler** (Available only when spec limits are specified.) Shows the defect rate as an isolated function of each factor. See "Defect Profiler".

**Defect Parametric Profile** (Available when the Defect Profiler is launched.) Shows the defect rate as an isolated function of the parameters of each factor's distribution. See "Defect Parametric Profile".

**N Strata** (Appears only when you press Shift before clicking the Simulator red triangle menu.) Enables you to specify the number of strata for the Normal Weighted distribution. See "Statistical Details for the Simulator".

**Set Random Seed** (Appears only when you press Shift before clicking the Simulator red triangle menu.) 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 when the Make Table button is clicked.

**Simulation Experiment** (Available for Normal, Uniform, and Triangular random distributions.) Runs a designed simulation experiment on the locations of the factor distributions. A window appears that enables you to specify the number of design points, the portion of the factor space to be used in the experiment, and which factors to vary in the experiment. Factors that are not selected for the experiment are set to their current Profiler value. See "Simulation Experiment".

**Spec Limits** Opens a table to set or edit specification limits for the responses. Click the **Save** button to save specification limits to the response column properties in the data table. See "Specification Limits".

# Defect Profiler

From the Simulator red triangle menu, you can launch the Defect Profiler. This 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 the changes in a factor's distribution that the process is most sensitive to, often as part of tolerance design.

Specification limits are used to define a defect, and the random distributions assigned to each factor are used to simulate responses. To use the Defect Profiler, specification limits must be defined for at least one response.

Define at least one of the factors as Random in the simulation settings to obtain useful results. Otherwise, the simulation outputs are constant and the defect rate is zero if the outcome falls within spec or one if the outcome falls out of spec.

**Tip:** If you need to estimate very small defect rates, consider the Normal weighted random option. This random distribution estimates stable defect rates of just a few parts per million with only a few thousand simulation runs.

## Introduction to Tolerance Design

*Tolerance design* is the investigation of how defect rates can be controlled by controlling variability in the input factors.

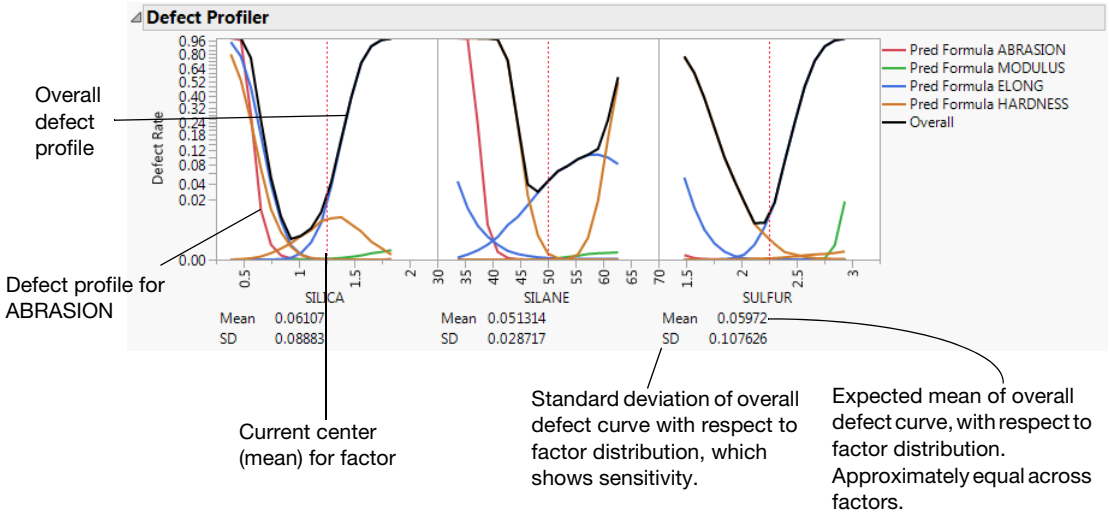
Input factors have variation. Specification limits are used to define acceptable ranges for each input. The variability in the input factors impacts the outputs. Specification limits are also used to define the acceptable range for output variables.

Sometimes, a tolerance design study shows that spec limits on an input are unnecessarily tight. Loosening such limits can result in reduced costs without sacrificing product 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 understand how the inputs and their variability impact that the defect rates.

The Defect Profiler shows the defect rate as a function of each factor fixed at the specified mean value while all other factors are varied according to their random specification. If there are multiple outputs with spec limits, each has a different colored defect rate curve. A black curve shows the overall defect rate.

**Figure 8.7** Defect Profiler




## Graph Scale

Defect rates are shown on a cubic root scale, so that both large and small defect rates are shown in some detail.

## Defect Rates

The mean and standard deviation (SD) of the simulated overall defect rate are reported below each defect profile plot. This mean is calculated by integrating the defect profile curve with the specified factor distribution. Because of numerical error in the estimations, the overall mean defect rate reported under each factor can differ slightly.

The defect rates that are reported in the Defect Profiler are estimates of the simulated overall defect rate. This rate is also reported in the Rate column of the defect table that is in the Prediction Profiler outline.

**Note:** The defect table is added to the Prediction Profiler outline after you run a simulation and one or more of your responses has defined spec limits.

Because the rate estimates are obtained differently, by integration and simulation, they might differ slightly. If they are very different, you might consider increasing the number of simulation runs. In addition, verify that the range of the factor scale is wide enough so that the integration covers the distribution well.

The standard deviation is a measure of the sensitivity of the defect rate to the factor. It is small when either the factor profile is flat or the factor distribution has a small variance. The larger the standard deviation, the larger the impact that changes in that factor have on the defect rate variability. Comparing SDs across factors enables you to select factors for improvement to reduce defect rates.

The mean and SD values are updated when you change the factor distribution settings. 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 generate updated simulation runs.

### Simulation Method and Details

Suppose you 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. (By default, $k$ is 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 $m/n$. These defect rates are connected and plotted as a continuous function of X1.

### Notes

- The profile curve is not recalculated automatically when distributions change. Click **Rerun** to update the curve.
- Defect Profiling does not address the general optimization problem 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.
- The defect profiles tend to be uneven when they are low. This is due, in part, to the use of the cubic scale where differences in low values are exaggerated. Jagged defect profiles could be due to limited simulation runs. If the overall defect curve (black line) is smooth, and the individual defect rates are fairly consistent, then your simulation probably has enough runs to provide a stable solution. If the overall defect rate curve is jagged, then consider increasing the number of runs. 20,000 runs is generally enough to stabilize the curves.

## Defect Parametric Profile

From the Simulator red triangle menu, you can launch the Defect Parametric Profile. This 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.

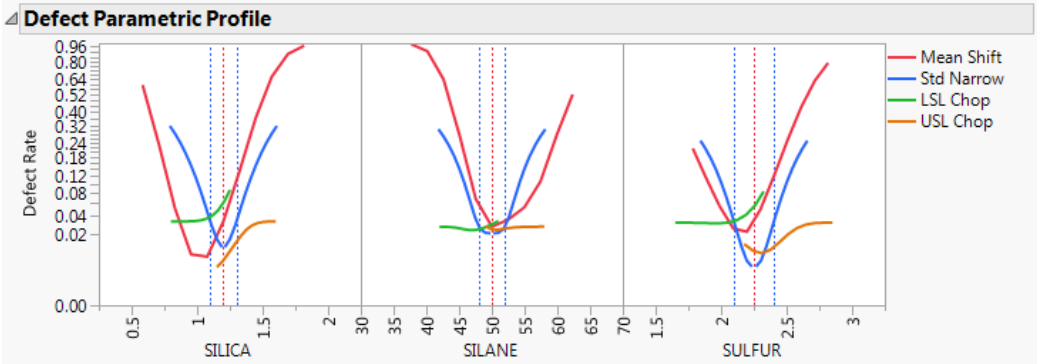
**Mean Shift** The impact of shifting the mean is shown by a red curve. The current mean is shown by a red dotted vertical line.

**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 value 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 by a green curve.

**USL Chop** The impact of inspection to remove all parts above the upper specification limit is shown by an orange curve.

**Figure 8.8** Defect Parametric Profile




## Simulation Experiment

Use the Simulation Experiment option in the Simulator red triangle menu to run a designed simulation experiment on the locations of the factor distributions. When you select Simulation Experiment, 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 that are not included in the experiment, the current value set in the Profiler is used in the experiment.

The experimental 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. The shape and variability come from the specified distributions. 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. You can use a Gaussian Process model to model the overall defect rate from the simulation experiment. For more information about Gaussian Process models, see *Predictive and Specialized Modeling*.

## Specification Limits

You can define Specification Limits using the Spec Limits option in the Simulator red triangle menu. Alternatively, right-click the desired column and select **Column Properties > Spec Limits.** In the Column Properties window, there are boxes to enter values for the Lower Spec Limit, Target, and Upper Spec Limit.

The profilers support specification limits on the responses and provide a number of features:

- In the Profiler, if the Response Limits are not set up in the input data table to provide desirability coordinates, JMP looks for a Spec Limits column property and constructs desirability functions appropriate to those spec limits. For more information about Response Limits see *Using JMP*.
- If you use the Simulator to output simulation tables, JMP includes the Spec Limits column properties in the output data tables. This makes accounting for defect rates and capability indices easy.
- Adding spec limits enables the Defect Profiler.

# Additional Examples of the Simulator

This section contains examples of the Simulator.

- "Example of the Defect Profiler"
- "Example of Stochastic Optimization"
- "Example of Simulating General Formulas"

## Example of the Defect Profiler

This example demonstrates a possible workflow of launching the Defect Profiler from the Simulator red triangle menu. The data comes from an experiment to study the effects of SILICA, SILANE, and SULFUR on four measures of tire tread performance.

### Add Specification Limits in the Simulator

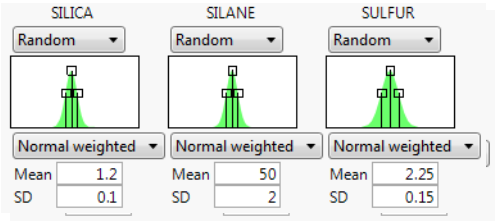
1. Select **Help > Sample Data Folder** and open Tiretread.jmp.
2. Select **Graph > Profiler**.
3. Select Pred Formula ABRASION, Pred Formula MODULUS, Pred Formula ELONG, and Pred Formula HARDNESS and click **Y, Prediction Formula**.
4. Click **OK**.
5. Click the Prediction Profiler red triangle and select **Simulator**.

6. Click the Simulator red triangle and select **Spec Limits**.

**Note:** If your columns have spec limits saved as a column property, those spec limits are shown in the Simulator Spec Limits table.

7. Set the spec limits:
   - For ABRASION set the LSL to 110.
   - For MODULUS set the USL to 2000.
   - For ELONG set the LSL to 350 and the USL to 550.
   - For HARDNESS set the LSL to 66 and the USL to 74.
8. Click **Save** to save the Spec Limits to the data table.
9. Beneath each factor, select **Random**.
10. For each factor, enter the random specifications as shown in Figure 8.9.

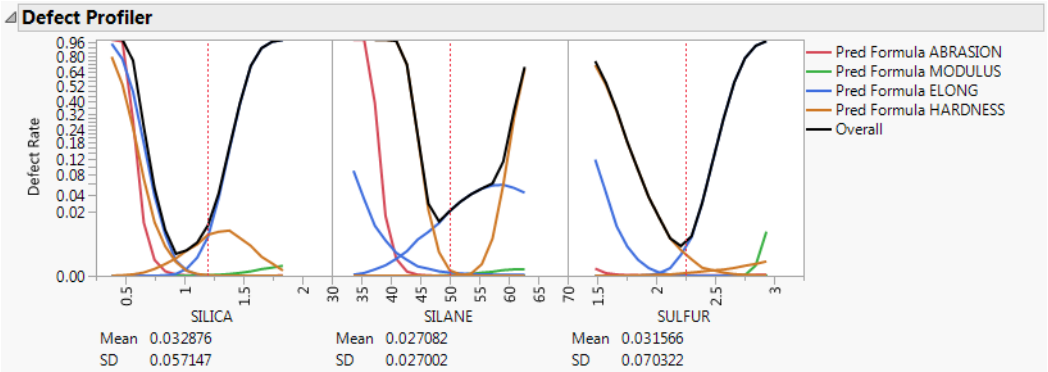
**Figure 8.9** Profiler Random Specifications




## Open the Defect Profiler and Defect Parametric Profile

1. Click the Simulator red triangle and select **Defect Profiler.**

**Figure 8.10** Defect Profiler



**Note:** The defect profiler is based on simulation with random inputs. The values that you obtain might differ from those shown in Figure 8.10.

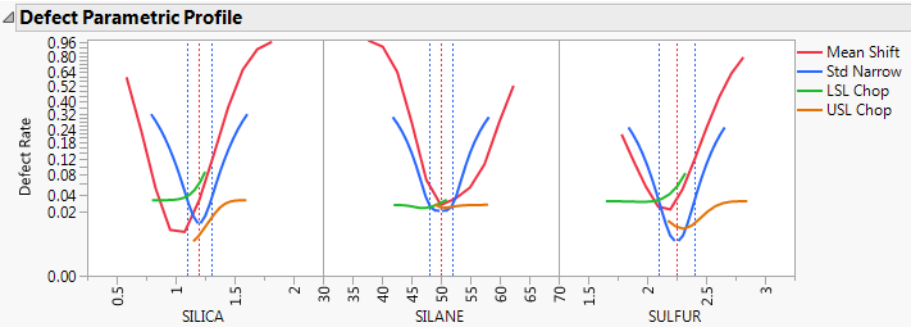
The black curves show the overall defect rate for each factor fixed at the X axis value while all other factors vary.

Consider the overall curve for SILICA. As silica varies, the defect rate goes from the lowest rate of 0.001 when silica is about 1 and increases quickly up to a defect rate of nearly 1 as silica increases or decreases from 1. However, SILICA is itself random. If you integrate the density curve of SILICA, you would estimate the average defect rate to be about 0.03, which is shown as the Mean for SILICA. This estimate of the overall defect rate estimated by the simulation is shown in the defect table under the simulation histograms. The Mean value for the overall defect rate for all factors are similar.

The Defect Profiler also includes an estimate of the standard deviation of the defect rate with respect to the variation in each factor. This value (labeled SD) is 0.057 for SILICA. The standard deviation is related to the sensitivity of the defect rate with respect to the distribution of that factor. Comparing the SD values across the three factors, the SD for SULFUR is higher than the SD values for SILICA and SILANE. This indicates that to improve the defect rate, shifting the distribution in SULFUR should have the greatest effect. A distribution can be shifted by changing its mean, changing its standard deviation, or by truncating the distribution by rejecting inputs that do not meet certain specification limits.

2. Click the Simulator red triangle and select **Defect Parametric Profile**.

**Figure 8.11** Defect Parametric Profile




Consider SULFUR and note that the current defect rate (0.03) is represented in four ways corresponding to each of the four curves in the Parametric Profiler.

For the red curve, Mean Shift, the current rate is where the red curve intersects the vertical red dotted line. The Mean Shift curve represents the change in overall defect rate as the mean of SULFUR changes. 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 shift down in the mean reduces the defect rate to about 0.02.

For the blue curve, Std Narrow, the current defect rate is where the blue curve intersects the two dotted blue lines. The Std Narrow curves represent the change in defect rate as the standard deviation changes. The dotted blue lines represent one standard deviation below and above the current mean. The blue curve is drawn symmetrically around the center. At the center, the blue curve 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 lower than the current rate of 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 for improvement in the defect rate. The green curve is above current defect rates for the entire range of the curve. This indicates that reducing the variation by rejecting parts with too-small values for SULFUR does not help reduce the defect rate.

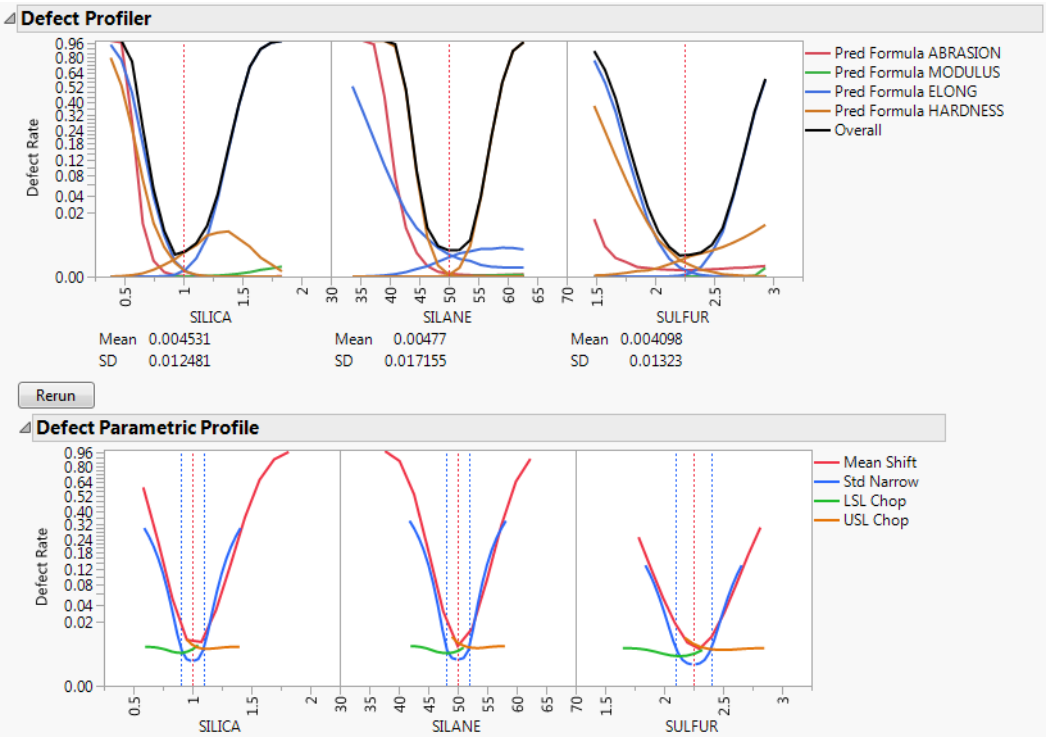
The orange curve, USL Chop, suggests a way to improve the defect rate. Reading the curve from the right, the curve starts out at the current defect rate (0.03). Then as you start rejecting more parts by decreasing the USL for SULFUR, the defect rate improves. However, moving a spec limit to the center of the distribution is equivalent to throwing away half the parts, which might not be a practical solution.

Looking at all the opportunities over all factors, it now looks like there are two good options for further investigation. You could shift 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 a process variation, the best first adjustment might be to shift the mean of SILICA to 1.

3. In the Prediction Profiler, adjust the Mean of SILICA from 1.2 to 1.0.
4. Below the Defect Profiler, click **Rerun**.

   The Defect Profiler updates based on the adjusted SILICA value.

**Figure 8.12** Adjusted Defect Rates

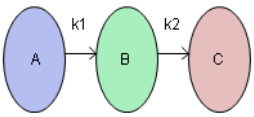



By shifting the mean of SILICA from 1.2 to 1.0, the defect rate has decreased from 0.03 to about 0.004, which is a good improvement. Investigating other defect reduction scenarios is easy to do by making changes to the distributions and rerunning the simulations.

## Example of Stochastic Optimization

Use the profiler and the Simulator to explore factor settings to optimize the yield of an intermediary chemical in a chemical reaction. A chemical reaction converts chemical A into chemical B and B into C. The resulting amount of chemical B is a function of reaction time and reaction temperature.

**Figure 8.13** Chemical Reaction



For this reaction the yield of B can be computed using the Arrhenius laws. 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.

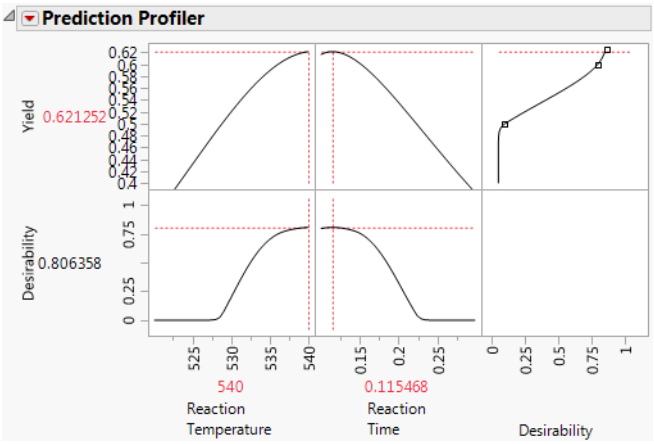
## Maximize Yield Using Desirability Function

1. Select **Help > Sample Data Folder** and open Stochastic Optimization.jmp.
2. Select **Graph > Profiler**.
3. Select Yield and click **Y, Prediction Formula**.
4. Click **Expand Intermediate Formulas** and then click **OK**.

   The Prediction Profiler with Desirability Functions enabled appears. For more information about Desirability Functions, see the "Desirability Profiling and Optimization".
5. Click the Prediction Profiler red triangle and select **Optimization and Desirability > Maximize Desirability**.

   The Prediction Profiler maximizes Yield and sets the graphs to the optimum value of Reaction Time and Reaction Temperature.

**Figure 8.14** Yield Maximum



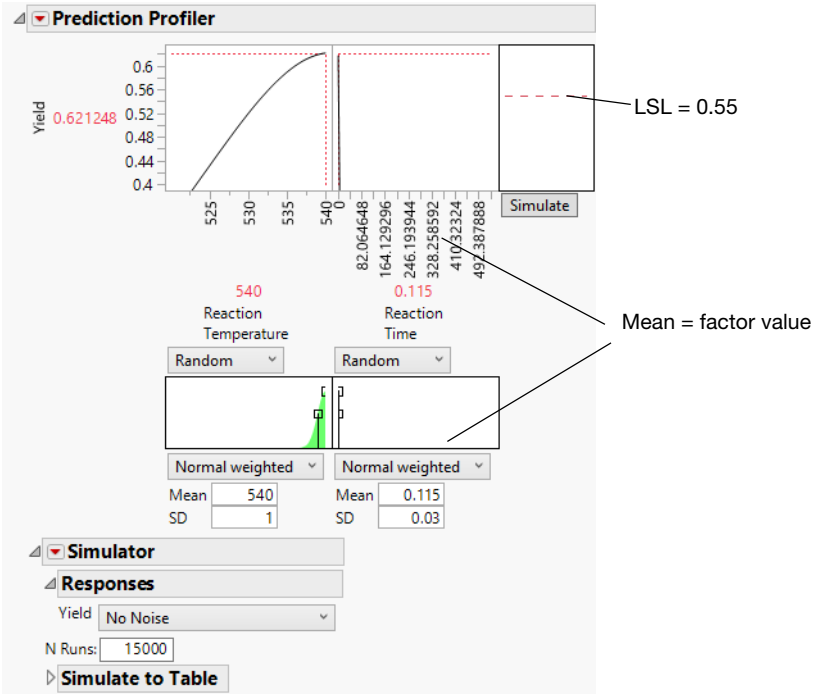

The maximum Yield is approximately 0.62 at a Reaction Time of 0.115 hours and Reaction Temperature of 540 degrees Kelvin, or hot and fast. (Your results might differ slightly due to random starting values in the optimization process.)

## Simulate Defect Rate for Maximum Yield

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 are out of spec? The Simulator can help you investigate the variation and defect rate for Yield, given variation in Reaction Time and Reaction Temperature.

1. Click the Prediction Profiler red triangle and deselect **Optimization and Desirability > Desirability Functions**.
2. Click the Prediction Profiler red triangle and select **Simulator**.
3. Set the simulation parameters for Reaction Temperature to **Random > Normal Weighted** with Mean = 540 and SD = 1.
4. Set the simulation parameters for Reaction Time to **Random > Normal Weighted** with Mean = 0.115 and SD = 0.03.
5. Set the **N Runs** value to 15,000.

**Figure 8.15** Initial Simulator Settings

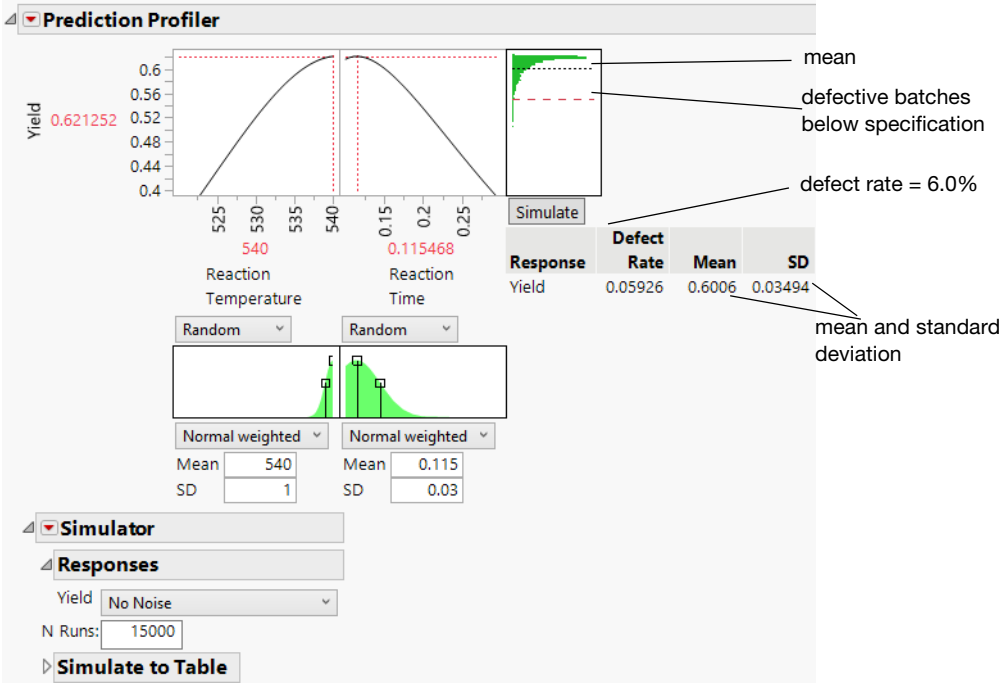



Yield has a lower spec limit of 0.55, set as a column property, and shows in Figure 8.15 as a dashed red line.

6. Click the **Simulate** button.

**Note:** Your numbers might differ from those shown in Figure 8.16 due to the random draws in the simulation.

**Figure 8.16** Simulation Results



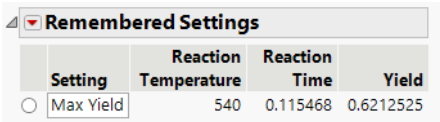

The predicted Yield is 0.62 for Reaction Temperature of 540 and Reaction Time of 0.115. The simulation estimates an average defect rate of about 6% with a standard deviation of 0.03 given the assumed variation in the temperature and time. The defect rate of 6.0%, indicates that about 6.0% of batches would be out of specification.

Use the simulator to explore other settings of Reaction Temperature and Reaction Time that might maintain a high Yield but with a lower defect rate. Before changing settings, save these factor settings that give us the maximum Yield for later use.

7. Click the Prediction Profiler red triangle and select **Factor Settings > Remember Settings**.
8. Type “Max Yield” and click **OK**.

   The settings are appended to the report window.

**Figure 8.17** Remembered Settings for Maximum Yield



Remembered Settings

| Setting | Reaction Temperature | Reaction Time | Yield |
|---|---|---|---|
| Max Yield | 540 | 0.115468 | 0.6212525 |

9. Set the Mean value for Reaction Temperature to 535.

   Using the dashed red line in the Reaction Time plot, you can explore and determine the approximate value that maximized Yield. This value should be around 0.16.

10. Set the Mean value of Reaction Time to 0.16.

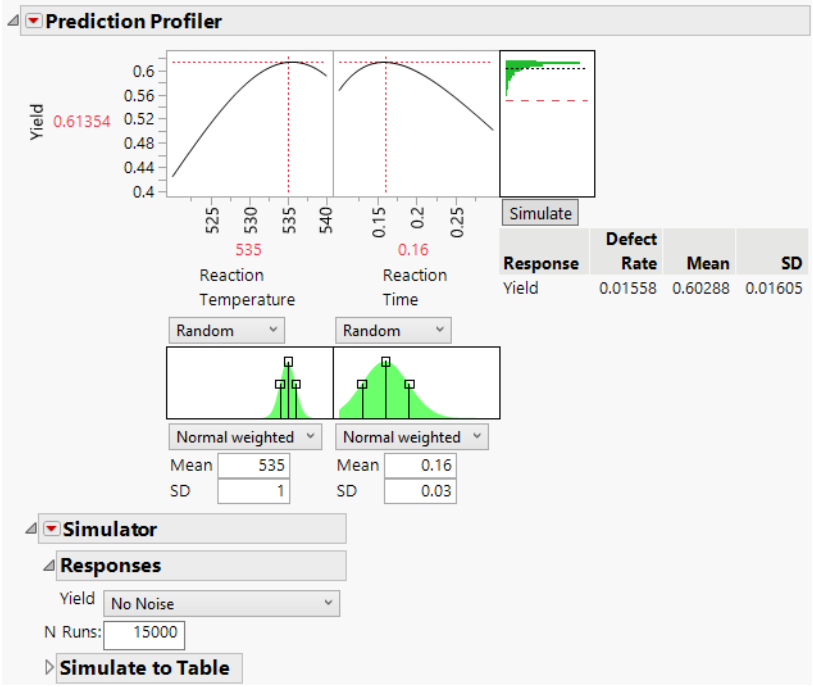
11. Click **Simulate**.

---

**Note:** Your numbers might differ from those shown in Figure 8.18 due to the random draws in the simulation.

---

**Figure 8.18** Defect Rate for Temperature of 535




By making slight changes to the input factor settings, the defect rate decreases to about 1.6% with a decrease of less than 0.01 in the predicted yield. The fixed (no variability) settings that maximize Yield are not the same settings that minimize the defect rate in the presence of factor variation.

### Simulation Experiment

You can run a simulation experiment to find the settings of Reaction Temperature and Reaction Time that minimize the defect rate. To do this, simulate the defect rate at each point in an experimental design for Reaction Temperature and Reaction Time. Then fit a predictive model for the defect rate and find factor settings to minimize the defect rate.

1. Click the Simulator red triangle and select **Simulation Experiment**.
2. Set the number of experimental runs to 80 and the portion of the design space to 1 to use the whole factor space in the experiment.
3. Click **OK**.

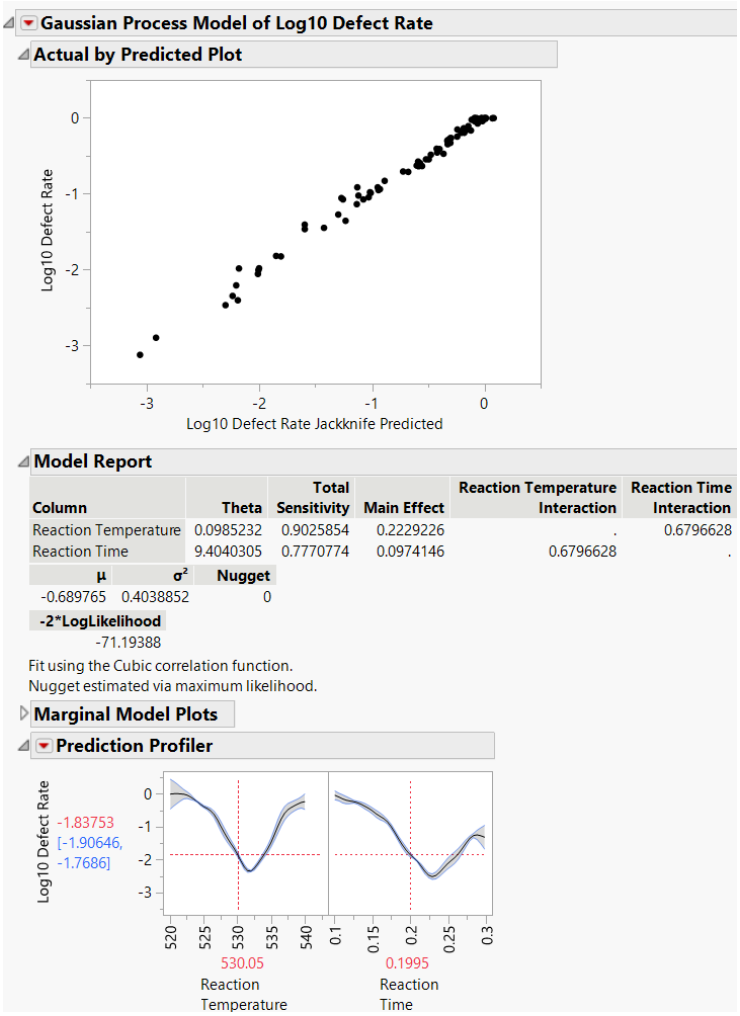
   A Latin Hypercube design with 80 design points is generated within the specified factor space, and N Runs random draws are taken at each of the design points. The design points are the center of the random draws, and the shape and variance of the random draws come from the factor distribution settings.

   A data 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.

   **Note:** Do not close the Stochastic Optimization Profiler window. You come back to it later.

4. From the new data table, click the green triangle next to the Gaussian Process script.

**Figure 8.19** Results of Gaussian Process Model Fit

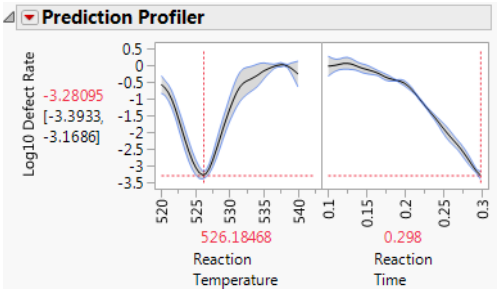



**Note:** Your results might be slightly different due to the random draws in the simulation.

5. To find the settings of Reaction Temperature and Reaction Time that minimizes the defect rate, click the Prediction Profiler red triangle and select **Optimization and Desirability > Maximize Desirability**.

   The desirability function is already set up to minimize the defect rate.

**Figure 8.20** Settings for Minimum Defect Rate



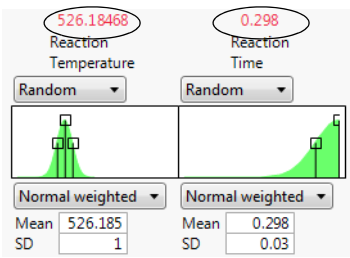

The settings that minimize the defect rate are approximately Reaction Temperature = 526 and Reaction Time = 0.3.

6. Click the **Transfer Factor Settings Back** button.

   This updates the original Profiler report to use the setting for Reaction Temperature and Reaction Time that minimize the defect rate.

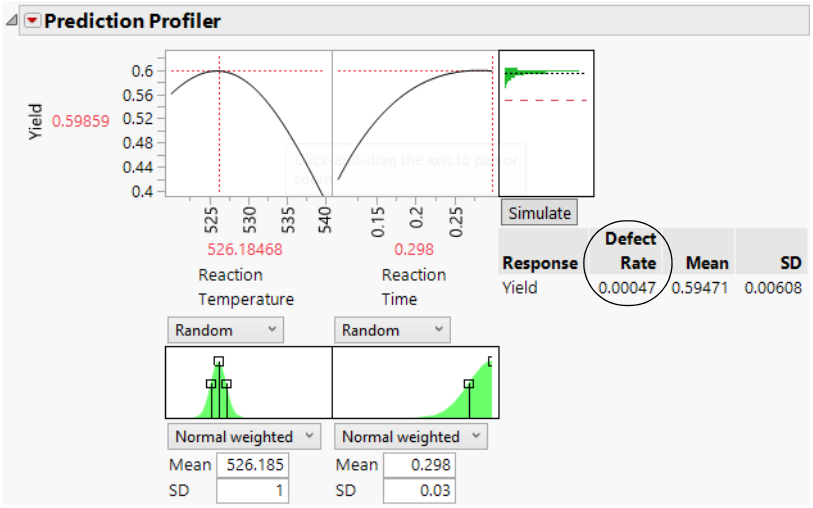
7. Return to the original Profiler report window.
8. Click the Prediction Profiler red triangle and select **Factor Settings > Remember Settings**.
9. Type “Min Defect” and click **OK**.

**Figure 8.21** Minimum Defect Settings




10. With the new settings in place, click the **Simulate** button to estimate the defect rate at the new settings.

**Figure 8.22** Lower Defect Rate

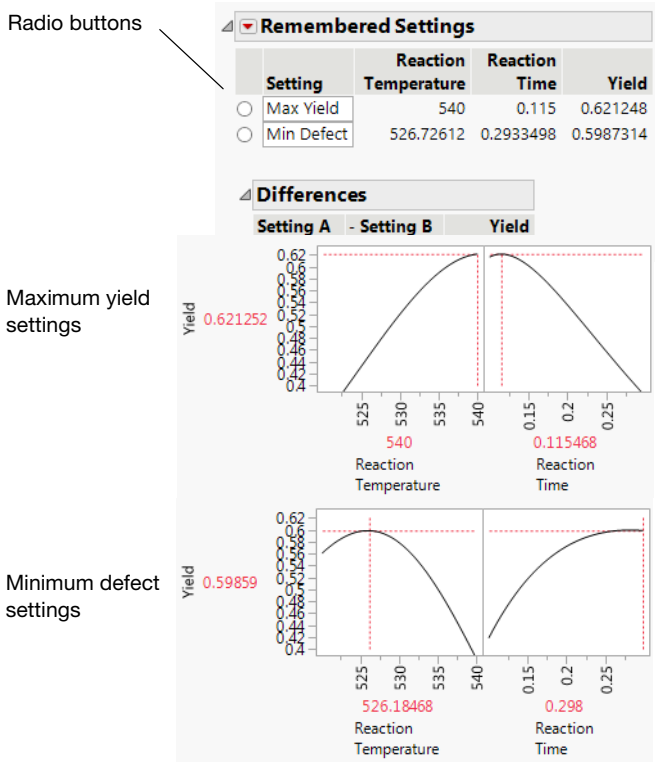



At the new settings the defect rate is 0.05%. This is much lower than the 6.0% for the settings that maximize Yield. That is a reduction of about 120x. Recall that the average Yield from the first settings is 0.62, and the new average is 0.59. The decrease in average Yield of 0.03 is the trade off for lowering the defect rate by 120x.

Because you saved the settings using Remember Settings, you can easily compare the old and new settings. The Differences report summarizes the difference.

11. Click the **Remembered Settings** radio buttons to view the profiler for each setting.

**Figure 8.23** 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.

## Example of Simulating General Formulas

Obtain the distribution of certain objectives by using the Simulator to run a financial model under different probability scenarios. In this example, you store the formulas into columns, set up ranges, and then use the Simulator to conduct a simulation.

1. In the JMP Home Window, select **File > New > Script**. This opens a new script window.
2. Copy and paste the following JSL script into the new script window.

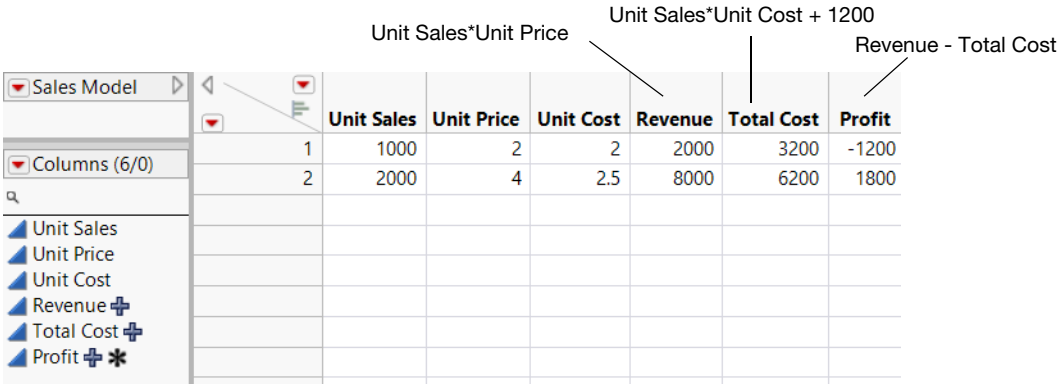
```
dt = New Table( "Sales Model" );
dt << New Column( "Unit Sales", Values( {1000, 2000} ) );
dt << New Column( "Unit Price", Values( {2, 4} ) );
dt << New Column( "Unit Cost", Values( {2, 2.5} ) );
```

```
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 )
);
```

3. Click the Run Script icon to run the script. Alternatively, you can select Ctrl-R.

   The script creates the data table in Figure 8.24 with some initial scaling data and stores formulas into the output variables. It also launches the Prediction Profiler.

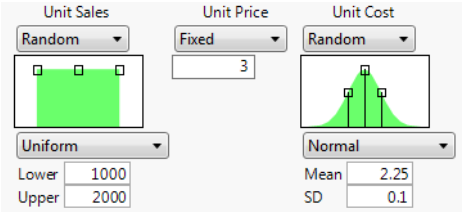
**Figure 8.24** Data Table Created from Script




| | Unit Sales | Unit Price | Unit Cost | Revenue | Total Cost | Profit |
|---|---|---|---|---|---|---|
| 1 | 1000 | 2 | 2 | 2000 | 3200 | -1200 |
| 2 | 2000 | 4 | 2.5 | 8000 | 6200 | 1800 |

4. Click the Prediction Profiler red triangle and select the **Simulator**.
5. In the menus beneath Unit Sales and Unit Cost, select **Random**.
6. Fill in the factor parameters:
   - Unit Sales is **Uniform** with Lower limit 1000 and Upper limit 2000.
   - Unit Price is **Fixed** at 3.
   - Unit Cost is **Normal** with mean of 2.25 and standard deviation of 0.1.

**Figure 8.25** Specifications for Prediction Profiler




7. Click the **Simulate** button.

**Note:** Your numbers might differ from those shown in Figure 8.26 due to the random draws in the simulation.

**Figure 8.26** Simulator

It appears that the models are unlikely to be profitable. By putting a lower specification limit of zero on Profit, the defect report tells you that the probability of being unprofitable is 62%.

8. Change the fixed value of Unit Price to 3.25.
9. Click the **Simulate** button.

**Figure 8.27** Results

| Response | Defect Rate | Mean | SD |
|---|---|---|---|
| Revenue | 0 | 4891.91 | 942.932 |
| Total Cost | 0 | 4587.98 | 671.168 |
| Profit | 0.2114 | 303.933 | 328.727 |
| All | 0.2114 | | |

Now the probability of being unprofitable is down to about 21%.

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.

# Statistical Details for the Simulator

The Simulator 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 and is accurate at estimating in the extreme tails.

First, define the number of strata. The strata are a net of hyperspheres that are centered around 0. For $d$ random factors, the strata are defined by their radial intervals. See Table 8.1.

**Table 8.1** Strata Intervals

| Strata Number | Inside Distance | Outside Distance |
|---|---|---|
| 0 | 0 | $\sqrt{d}$ |
| 1 | $\sqrt{d}$ | $\sqrt{d+\sqrt{2d}}$ |
| 2 | $\sqrt{d+\sqrt{2d}}$ | $\sqrt{d+2\sqrt{2d}}$ |
| $i$ | $\sqrt{d+(i-1)\sqrt{2d}}$ | $\sqrt{d+i\sqrt{2d}}$ |
| $N_{\text{Strata}}-1$ | $\sqrt{d+\{(N_{\text{Strata}}-1)-1\}\sqrt{2d}}$ | $\sqrt{d+(N_{\text{Strata}}-1)\sqrt{2d}}$ |

The default number of strata is 16. 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, the following is done:

1. Select a strata as mod($i$ – 1, $N_{\text{Strata}}$) 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.

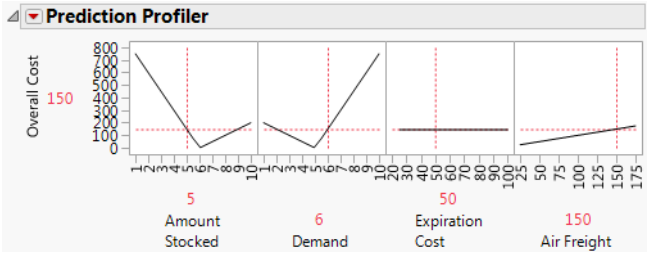
Chapter 9

# Excel Profiler

## Visualize Models Saved in Microsoft Excel

The Excel Profiler is used 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 Prediction Profiler Using an Excel Model



## Contents

Overview of the Excel Profiler . . . . . 195
Example of an Excel Model . . . . . 195
Run the Excel Profiler . . . . . 197
Use Linear Constraints . . . . . 198
Resolution of Profile Lines . . . . . 198
Use the Excel Profiler from JMP . . . . . 199

# Overview of the Excel Profiler

Use the Excel Profiler to visualize models (or formulas) stored in Excel worksheets. You can choose to install the Excel Add-In when you install JMP. Profiling in the Excel Add-In is a two-step process:

1. Click the **Create/Edit Model** button (Excel 2010 through 2021) 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 2010 through 2021) to launch the JMP Profiler and run the Excel model. See "Run the Excel Profiler".

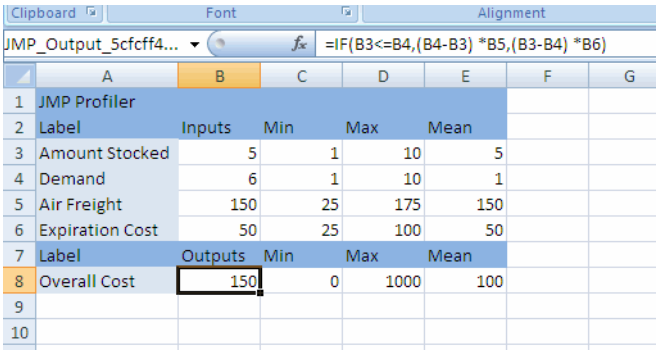
**Notes:**

- The Preferences, Data Table, Graph Builder, Distribution, Fit Y by X, Fit Model, Time Series, and Control Chart buttons are not needed to profile an Excel model. For more information about these features, see *Using JMP*.
- 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 installer, 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




| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | JMP Profiler | | | | | | |
| 2 | Label | Inputs | Min | Max | Mean | | |
| 3 | Amount Stocked | 5 | 1 | 10 | 5 | | |
| 4 | Demand | 6 | 1 | 10 | 1 | | |
| 5 | Air Freight | 150 | 25 | 175 | 150 | | |
| 6 | Expiration Cost | 50 | 25 | 100 | 50 | | |
| 7 | Label | Outputs | Min | Max | Mean | | |
| 8 | Overall Cost | 150 | 0 | 1000 | 100 | | |
| 9 | | | | | | | |
| 10 | | | | | | | |

## 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 depend on the relationship between Amount Stocked and Demand.

- 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 Folder** and navigate up one level to the Samples/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 workbook 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:**

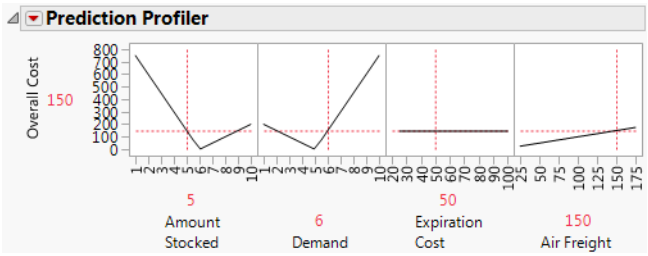
- If the fields in the Create/Edit Model window are not populated when you set up your model, click **Choose** and select the cell that contains the input or output name.
  - For inputs, specify the values and click **Apply**. Inputs must be values, not formulas.
  - For outputs, the specified cell must be a formula containing only information from the Input cells.
- 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 Excel Profiler

Once you create a model using the Excel Add-In, you can run the model in the 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 Prediction 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. Since the Excel worksheet and the Profiler are not dynamically linked, after you update a formula in the worksheet, you must relaunch the JMP Profiler to reflect the changes.

# 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 the constraints are saved to the Excel workbook, whenever the model is profiled from the Excel Add-In, the constraints are incorporated.

1. Click the Prediction Profiler red triangle and select **Alter Linear Constraints**.
2. Click **Add Constraint**.
3. Type in the constraining values.
4. Click **OK**.
5. Click the Prediction Profiler red triangle and 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".

**Tip:** To delete a linear constraint, set all constraint values to zero.

# Resolution of Profile Lines

The Default N Levels option on the Prediction Profiler red triangle menu 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 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.

Use the following code to script the Excel Profiler:

```
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 *Scripting Guide*.

# Appendix A

## References

The following sources are referenced in *Profilers*.


Box, G. E. P., and Draper, N. R. (1987). *Empirical Model–Building and Response Surfaces*. New York: John Wiley & Sons.

Box, G. E. P., and Draper, N. R. (2007). *Response Surfaces, Mixtures, and Ridge Analyses*. 2nd ed. New York: John Wiley & Sons.

Cornell, J. A. (1990). *Experiments with Mixtures*. 2nd ed. New York: John Wiley & Sons.

Derringer, G., 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). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. New York: Springer-Verlag.

Lundberg, S. M. (2018). Permutation Explainer [GitHub]. Retrieved from https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/explainers/Permutation.html

Lundberg, S. M., and Lee, S. I. (2017). "A unified approach to interpreting model predictions." *Advances in Neural Information Processing Systems* 30.

Saltelli, A. (2002). "Making best use of model evaluations to compute sensitivity indices." *Computer Physics Communications* 145:280–297.

Schafer, J., and Strimmer, K. (2005). "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics." *Statistical Applications in Genetics and Molecular Biology* 4:1–30.

Shapley, L. S. (2016). *Contributions to the Theory of Games (AM-28), Volume II*. Chapter 17 "A Value for n-Person Games". Princeton University Press.

Sobol, I. M. (1993). "Sensitivity Estimates for Nonlinear Mathematical Models." *MMCE* 1.4:407–414.

# Appendix B

## Technology License Notices

- Scintilla is Copyright © 1998–2017 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.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is 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.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software is Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). 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 OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

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

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

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

  3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. 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 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.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF

USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- 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 license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

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

2. 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.

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

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

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

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

  2. 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.

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

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS "AS IS" WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. 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 Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

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

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

  2. 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.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

  You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

  - You must give any other recipients of the Work or Derivative Works a copy of this License; and
  - You must cause any modified files to carry prominent notices stating that You changed the files; and
  - You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
  - If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
  - You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF

ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2021, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

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

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

- On the Windows operating system, JMP utilizes the OpenBLAS library. OpenBLAS is licensed under the 3-clause BSD license. Full license text follows: Copyright © 2011-2015, The OpenBLAS Project

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

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

  2. 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.

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

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE OPENBLAS PROJECT OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.