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 and 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
• schedules for seminars being held in your area
• success stories showing how others use JMP
• a blog with tips, tricks, and stories from JMP staff
• a forum to discuss JMP with other users

http://www.jmp.com/getstarted/
Release 9

Quality and Reliability Methods

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

Marcel Proust

JMP, A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513
# Contents

Quality and Reliability Methods

## 1 Lifetime Distribution

**The Life Distribution Platform** ................................................................. 1

- Introduction ......................................................................................... 3
- Launching the Life Distribution Platform ......................................... 3
  - Launch Window Buttons ................................................................. 4
- Working with the Life Distribution Report Window ......................... 4
  - Event Plot ......................................................................................... 5
  - Compare Distributions ..................................................................... 6
  - Statistics ......................................................................................... 11
- Understanding Platform Options ...................................................... 13
- Navigating Report Details ................................................................. 14
  - Profilers ......................................................................................... 14
  - Parametric Statistics Report .......................................................... 15
  - The Tabbed Report Option ............................................................. 17
- Mixed Censoring Example ................................................................. 18
  - Illustration of Censoring Differences ............................................ 20
- Estimating Competing Causes ............................................................ 22
  - Additional Life Distribution Launch Options ................................. 23
  - Navigating the Competing Cause Report Window .......................... 24
- Statistical Details .............................................................................. 29
  - Nonparametric Fit .......................................................................... 30
  - Parametric Distributions ................................................................. 30

## 2 Lifetime Distribution II

**The Fit Life by X Platform** ................................................................. 41

- Introduction to Accelerated Test Models ............................................ 43
- Launching the Fit Life by X Platform Window .................................... 43
- Platform Options .............................................................................. 45
Navigating the Fit Life by X Report Window ........................................... 46
  Scatterplot ................................................................. 46
  Nonparametric Overlay .................................................... 49
  Wilcoxon Group Homogeneity Test ........................................... 50
  Comparisons ............................................................... 50
  Results ............................................................................ 54
Viewing Profilers and Surface Plots .................................................... 61
Using the Custom Relationship for Fit Life by X ................................. 63

3 Recurrence Analysis
  The Recurrence Platform ....................................................... 65
  Recurrence Data Analysis ...................................................... 67
    Launching the Platform ...................................................... 67
  Examples ............................................................................ 68
    Valve Seat Repairs Example .............................................. 68
    Bladder Cancer Recurrences Example ............................... 71
  Fit Model ........................................................................... 73

4 Degradation
  Using the Degradation Platform ............................................... 77
  Overview of the Degradation Platform ...................................... 79
  Launching the Degradation Platform ........................................ 79
  The Degradation Report ...................................................... 80
  Model Specification ........................................................... 82
    Simple Linear Path .......................................................... 82
    Nonlinear Path ............................................................... 84
  Inverse Prediction ............................................................. 91
  Prediction Graph ............................................................. 92
  Platform Options .............................................................. 93
  Model Reports ................................................................. 96
    Model Lists ................................................................. 96
    Reports ........................................................................... 97
  Destructive Degradation ....................................................... 98
  Stability Analysis .............................................................. 101
5 Reliability and Survival Analysis

Univariate Survival .......................................................... 103

Introduction to Survival Analysis ........................................... 105

Univariate Survival Analysis .................................................. 105

Selecting Variables for Univariate Survival Analysis ................. 106

Example: Fan Reliability ..................................................... 106

Example: Rats Data ........................................................... 109

Overview of the Univariate Survival Platform ......................... 109

Statistical Reports for the Univariate Analysis ....................... 110

Platform Options ............................................................ 112

Fitting Distributions ......................................................... 116

Interval Censoring ........................................................... 119

WeiBayes Analysis ......................................................... 121

Estimation of Competing Causes ........................................... 121

Omitting Causes .............................................................. 123

Saving Competing Causes Information ................................... 123

Simulating Time and Cause Data ......................................... 123

6 Reliability and Survival Analysis II

Regression Fitting .............................................................. 125

Parametric Regression Survival Fitting .................................. 127

Example: Computer Program Execution Time ......................... 127

Launching the Fit Parametric Survival Platform ....................... 130

Options and Reports .......................................................... 132

Example: Arrhenius Accelerated Failure Log-Normal Model ........ 133

Example: Interval-Censored Accelerated Failure Time Model ....... 138

Example: Right-Censored Data; Tobit Model ......................... 140

Proportional Hazards Model ................................................ 141

Example: One Nominal Effect with Two Levels ....................... 141

Example: Multiple Effects and Multiple Levels ....................... 144

Nonlinear Parametric Survival Models ................................. 148

Loss Formulas for Survival Distributions ............................... 148

Tobit Model Example ....................................................... 152

Fitting Simple Survival Distributions .................................... 154
7 **Variability Charts**

Variability Chart and Gauge R&R Analysis ............................................. 159

Variability Charts .............................................................................. 161
  Launch the Variability Platform ....................................................... 161
  Variability Chart .......................................................................... 162
  Variability Platform Options .......................................................... 163
  Heterogeneity of Variance Tests ....................................................... 165

Variance Components ........................................................................ 168
  Variance Component Method .......................................................... 169

R&R Measurement Systems ............................................................... 171
  Gauge R&R Variability Report ......................................................... 172
  Misclassification Probabilities ........................................................... 174
  Bias ............................................................................................... 174
  Linearity Study ............................................................................. 176

Discrimination Ratio Report ............................................................... 177

Attribute Gauge Charts ..................................................................... 178
  Data Organization ......................................................................... 178
  Launching the Platform .................................................................. 178
  Attribute Gauge Plots .................................................................... 179
  Agreement ..................................................................................... 181
  Effectiveness Report ..................................................................... 184

8 **Capability Analyses**

The Capability Platform ................................................................. 187

Launch the Platform .......................................................................... 189
  Entering Limits ............................................................................. 189

Capability Plots and Options ............................................................. 191
  Goal Plot ...................................................................................... 191
  Capability Box Plots ..................................................................... 192
  Normalized Box Plots .................................................................... 194
  Individual Detail Reports ............................................................... 194
  Make Summary Table .................................................................... 195
  Capability Indices Report ............................................................... 196
12 Shewhart Control Charts

Variables and Attribute Control Charts .................................................. 255

Shewhart Control Charts for Variables .................................................. 257
  XBar-, R-, and S-Charts ........................................................................ 257
  Run Charts .............................................................................................. 262
  Individual Measurement Charts .............................................................. 262
  Presummarize Charts ............................................................................ 265

Moving Average Charts ............................................................................ 267
  Uniformly Weighted Moving Average (UWMA) Charts .......................... 267
  Exponentially Weighted Moving Average (EWMA) Charts .................... 269

Shewhart Control Charts for Attributes .................................................. 271
  P- and NP-Charts .................................................................................. 272
  U-Charts ................................................................................................. 274
  C-Charts ................................................................................................. 275
  Levey-Jennings Charts ........................................................................... 277

Phases ....................................................................................................... 277
  Example .................................................................................................. 278
  JSL Phase Level Limits .......................................................................... 280

13 Cumulative Sum Control Charts

CUSUM Charts ......................................................................................... 281

Cumulative Sum (Cusum) Charts ............................................................... 283
  Launch Options for Cusum Charts ........................................................ 284
  Cusum Chart Options ............................................................................ 286
Formulas for CUSUM Charts .................................................. 290
Notation ................................................................. 290
One-Sided CUSUM Charts ................................................. 290
Two-Sided Cusum Schemes .................................................. 291

14 Multivariate Control Charts
Quality Control for Multivariate Data ....................................... 293
Launch the Platform .......................................................... 295
Control Chart Usage ......................................................... 295
  Phase 1—Obtaining Targets ............................................... 295
  Phase 2—Monitoring the Process ........................................ 296
  Monitoring a Grouped Process .......................................... 297
Change-Point Detection ..................................................... 300
  Method ................................................................. 300
  Example .............................................................. 300
Platform Options ............................................................. 302
Statistical Details ............................................................. 302
  Ungrouped Data ....................................................... 302
  Grouped Data ........................................................ 303
  Additivity .............................................................. 304
  Change-Point Detection ................................................ 304

Index
Quality and Reliability Methods ........................................... 319
Credits and Acknowledgments

Origin

JMP was developed by SAS Institute Inc., Cary, NC. JMP is not a part of the SAS System, though portions of JMP were adapted from routines in the SAS System, particularly for linear algebra and probability calculations. Version 1 of JMP went into production in October 1989.

Credits

JMP was conceived and started by John Sall. Design and development were done by John Sall, Chung-Wei Ng, Michael Hecht, Richard Potter, Brian Corcoran, Annie Dudley Zangi, Bradley Jones, Craige Hales, Chris Gotwalt, Paul Nelson, Xan Gregg, Jianfeng Ding, Eric Hill, John Schroedl, Laura Lancaster, Scott McQuiggan, Melinda Thielbar, Clay Barker, Peng Liu, Dave Barbour, Jeff Polzin, John Ponte, and Steve Amerige.

In the SAS Institute Technical Support division, Duane Hayes, Wendy Murphrey, Rosemary Lucas, Win LeDinh, Bobby Riggs, Glen Grimme, Sue Walsh, Mike Stockstill, Kathleen Kiernan, and Liz Edwards provide technical support.

Nicole Jones, Kyoko Keener, Hui Di, Joseph Morgan, Wenjun Bao, Fang Chen, Susan Shao, Yusuke Ono, Michael Crotty, Jong-Seok Lee, Tonya Mauldin, Audrey Ventura, Ani Eloyan, Bo Meng, and Sequola McNeill provide ongoing quality assurance. Additional testing and technical support are provided by Noriki Inoue, Kyoko Takenaka, and Masakazu Okada from SAS Japan.

Bob Hickey and Jim Borek are the release engineers.

The JMP books were written by Ann Lehman, Lee Creighton, John Sall, Bradley Jones, Erin Vang, Melanie Drake, Meredith Blackwelder, Diane Perhac, Jonathan Gatlin, Susan Conaghan, and Sheila Loring, with contributions from Annie Dudley Zangi and Brian Corcoran. Creative services and production was done by SAS Publications. Melanie Drake implemented the Help system.

Jon Weisz and Jeff Perkinson provided project management. Also thanks to Lou Valente, Ian Cox, Mark Bailey, and Malcolm Moore for technical advice.

Thanks also to Georges Guirguis, Warren Sarle, Gordon Johnston, Duane Hayes, Russell Wolfinger, Randall Tobias, Robert N. Rodriguez, Ying So, Warren Kuhfeld, George MacKensie, Bob Lucas, Warren Kuhfeld, Mike Leonard, and Padraic Neville for statistical R&D support. Thanks are also due to Doug Melzer, Bryan Wolfe, Vincent DelGobbo, Biff Beers, Russell Gonsalves, Mitchel Soltys, Dave Mackie, and Stephanie Smith, who helped us get started with SAS Foundation Services from JMP.

Acknowledgments

We owe special gratitude to the people that encouraged us to start JMP, to the alpha and beta testers of JMP, and to the reviewers of the documentation. In particular we thank Michael Benson, Howard
Yetter (d), Andy Mauromoustakos, Al Best, Stan Young, Robert Muenchen, Lenore Herzenberg, Ramon Leon, Tom Lange, Homer Hegedus, Skip Weed, Michael Emptage, Pat Spagan, Paul Wenz, Mike Bowen, Lori Gates, Georgia Morgan, David Tanaka, Zoe Jewell, Sky Alibhai, David Coleman, Linda Blazeck, Michael Friendly, Joe Hockman, Frank Shen, J.H. Goodman, David Iklé, Barry Hembree, Dan Obermiller, Jeff Sweeney, Lynn Vanatta, and Kris Ghosh.

Also, we thank Dick DeVeaux, Gray McQuarrie, Robert Stine, George Fraction, Avigdor Cahaner, José Ramirez, Gudmundur Axelsson, Al Fulmer, Cary Tuckfield, Ron Thisted, Nancy McDermott, Verónica Czitrom, Tom Johnson, Cy Wegman, Paul Dwyer, DaRon Huffaker, Kevin Norwood, Mike Thompson, Jack Reese, Francois Mainville, and John Wåss.

We also thank the following individuals for expert advice in their statistical specialties: R. Hocking and P. Spector for advice on effective hypotheses; Robert Mee for screening design generators; Roselinde Kessels for advice on choice experiments; Greg Piepel, Peter Goos, J. Stuart Hunter, Dennis Lin, Doug Montgomery, and Chris Nachtsheim for advice on design of experiments; Jason Hsu for advice on multiple comparisons methods (not all of which we were able to incorporate in JMP); Ralph O’Brien for advice on homogeneity of variance tests; Ralph O’Brien and S. Paul Wright for advice on statistical power; Keith Muller for advice in multivariate methods, Harry Martz, Wayne Nelson, Ramon Leon, Dave Trindade, Paul Tobias, and William Q. Meeker for advice on reliability plots; Lijian Yang and J.S. Marton for bivariate smoothing design; George Milliken and Yuri Bulavski for development of mixed models; Will Potts and Cathy Maahs-Fladung for data mining; Clay Thompson for advice on contour plotting algorithms; and Tom Little, Damon Stoddard, Blanton Godfrey, Tim Clapp, and Joe Ficalora for advice in the area of Six Sigma; and Josef Schmee and Alan Bowman for advice on simulation and tolerance design.

For sample data, thanks to Patrice Strahle for Pareto examples, the Texas air control board for the pollution data, and David Coleman for the pollen (eureka) data.

Translations
Trish O’Grady coordinates localization. Special thanks to Noriki Inoue, Kyoko Takenaka, Masakazu Okada, Naohiro Masukawa and Yusuke Ono (SAS Japan); and Professor Toshiro Haga (retired, Tokyo University of Science) and Professor Hirohiko Asano (Tokyo Metropolitan University) for reviewing our Japanese translation; Professors Fengshan Bai, Xuan Lu, and Jianguo Li at Tsinghua University in Beijing, and their assistants Rui Guo, Shan Jiang, Zhicheng Wan, and Qiang Zhao; and William Zhou (SAS China) and Zhonggou Zheng, professor at Peking University, for reviewing the Simplified Chinese translation; Jacques Goupy (consultant, ReConFor) and Olivier Nuñez (professor, Universidad Carlos III de Madrid) for reviewing the French translation; Dr. Byung Chun Kim (professor, Korea Advanced Institute of Science and Technology) and Duk-Hyun Ko (SAS Korea) for reviewing the Korean translation; Bertram Schäfer and David Meintrup (consultants, StatCon) for reviewing the German translation; Patrizia Omodei, Maria Scaccabarozzi, and Letizia Bazzani (SAS Italy) for reviewing the Italian translation. Finally, thanks to all the members of our outstanding translation teams.

Past Support
Many people were important in the evolution of JMP. Special thanks to David DeLong, Mary Cole, Kristin Nauta, Aaron Walker, Ike Walker, Eric Gjertsen, Dave Tilley, Ruth Lee, Annette Sanders, Tim Christensen, Eric Wasserman, Charles Soper, Wenjie Bao, and Junji Kishimoto. Thanks to SAS Institute quality assurance by Jeanne Martin, Fouad Younan, and Frank Lassiter. Additional testing for Versions 3 and 4 was done by Li Yang, Brenda Sun, Katrina Hauser, and Andrea Ritter.
Also thanks to Jenny Kendall, John Hansen, Eddie Routten, David Schlotzhauer, and James Mulherin. Thanks to Steve Shack, Greg Weier, and Maura Stokes for testing JMP Version 1.

Thanks for support from Charles Shipp, Harold Gugel (d), Jim Winters, Matthew Lay, Tim Rey, Rubin Gabriel, Brian Ruff, William Lisowski, Tom Esposito, Susan West, Chris Fehily, Dan Chilko, Jim Shook, Ken Bodner, Rick Blahunka, Dana C. Aultman, and William Fehlner.

Technology License Notices

Scintilla is Copyright 1998-2003 by Neil Hodgson <neilh@scintilla.org>. 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.

XRender is Copyright © 2002 Keith Packard. KEITH PACKARD DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL KEITH PACKARD 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.

ImageMagick software is Copyright © 1999-2010 ImageMagick Studio LLC, a non-profit organization dedicated to making software imaging solutions freely available.

bzlib software is Copyright © 1991-2009, Thomas G. Lane, Guido Vollbeding. All Rights Reserved.

FreeType software is Copyright © 1996-2002 The FreeType Project (www.freetype.org). All rights reserved.
Using the Life Distribution platform, you can discover distributional properties of time-to-event data. You can access the Life Distribution platform by selecting it from the Reliability and Survival menu.

With this platform, you can find the most suitable distributional fit for your data and make predictions. Weibull, lognormal, Fréchet, extreme value, and other common distributions used in reliability and survival analysis are available. Figure 1.1 shows available distributions and an example of a probability plot comparing Weibull and Fréchet distributions.

**Figure 1.1 Distributional Fits and Comparisons**
Introduction

Life distribution analysis is an important part of assessing and improving product reliability and quality. It is useful for determining material characteristics, predicting reliability, assessing effects of design modifications, comparing manufacturers, assessing reliability in the field, verifying advertising claims, and predicting warranty costs. Life distribution, or life data analysis, involves analyzing the time-to-failure of a product, component, or system. Specialized statistical methods are necessary for analyzing reliability data since many of the observations can be censored (times at failure are not known).

Reliability measures include hazard or failure rates, quantiles, and failure probabilities. Both nonparametric and parametric models of the data can be important. While nonparametric methods are useful to describe the curve of a distribution, parametric models are used to describe the distribution more concisely and provide smooth estimates of failure-time distributions. Parametric models are also useful for extrapolation (in time) to the lower or upper tails of a distribution. Although the normal distribution is sometimes used to describe reliability data, other distributions are more frequently used. Some commonly used distributions include the Weibull, lognormal, loglogistic, smallest extreme value, gamma, and exponential.

Launching the Life Distribution Platform

To launch the Life Distribution platform:

1. Open the Fan.jmp sample data table (located in the Reliability folder).
2. Select Analyze > Reliability and Survival > Life Distribution.
3. Assign Time as Y, Time to Event.
4. Assign Censor as Censor.

Figure 1.2 shows the completed launch window. The Life Distribution platform requires only a time (Y) variable, which must be duration or survival times. The Censor, Failure Cause, Frequency, Label, and By variables are optional. The sort-order of the data is not important.

5. Click OK.
Launch Window Buttons

The Life Distribution launch window contains the following role buttons:

- **Y, Time to Event** is the only required variable. It contains the time to event or time to censoring. With interval censoring, specify two Y variables, where one Y variable gives the lower limit and the other Y variable gives the upper limit for each unit.

- **Censor** By default, censor column values of 1 indicate censored observations; all other observations coded differently from 1 are treated as uncensored. If your censored values are coded with values other than 1, enter your code for censored values in the entry field to the right of Censor Code. Values not entered into this box are treated as uncensored.

- **Failure Cause** is used to specify the column containing multiple failure causes. This column is particularly useful for estimating competing causes. Either labeled or numerically coded values can be used for failure events in the Failure Cause column.

- **Freq** is for a column whose values are the frequencies or counts of observations for each row when there are multiple units recorded.

- **Label** is used when the events have identifiers other than the row number. The column containing the identifiers is given the Label role when the platform is launched so that the event plot uses the Label as event labels.

- **By** is used to perform a separate analysis for each level of a classification or grouping variable.

You can also use profile likelihood confidence intervals for the parameters instead of the default Wald confidence intervals. Click on the drop-down menu under Select Confidence Interval Method and select Profile.

Working with the Life Distribution Report Window

Figure 1.3 shows the report window. The report window contains three main outline titles:

- “Event Plot,” p. 5
- “Compare Distributions,” p. 6
- “Statistics,” p. 11
Note: If your data set contains no failures (only right-censored data) you see a statement in the report window that the nonparametric estimate cannot be calculated. When no failures are indicated, nonparametric estimates are not shown on the probability plot in the report window.

**Event Plot**

The Event Plot is shown by clicking on the disclosure button for Event Plot in the report. Figure 1.4 shows an example that uses the Fan jmp sample data table. The x-axis of the Event Plot shows time measurements. The y-axis shows the individual units.
For each individual unit in the plot, note the different line styles:

- A solid line not followed by a dashed line and that is marked by an x at the end of the line indicates an event failure. Line 62, in Figure 1.4, shows an example of an event failure.
- A dashed line connecting to the right of a solid line with a left triangle indicates a right-censored event.
- Time measurements (taken from the Y, Time to Event column) are shown as dots.

Default labels for the x- and y-axes can be edited by clicking the default label. Time and Label are the default names given by JMP for the x- and y-axes, respectively. In Figure 1.4, Unit (representing each row of the data) is substituted for Label.

**Compare Distributions**

The Compare Distributions section of the report contains a column of check boxes and a column of radio buttons, each corresponding to a particular distribution. (See Figure 1.5.) There is also a graph of Probability by lifetime, or Time, for this example. The nonparametric estimates are shown with their confidence interval. The scales of the x- and y-axes are set to linear when the Nonparametric radio button is clicked.
Figure 1.5 Compare Distributions of Fan.jmp

Changing the Scale

Suppose you are interested in viewing these estimates on a different scale.¹ For example, suppose we want to see these estimates on a Fréchet scale. Click the radio button to the right of Fréchet to change the y-axis to a Fréchet probability distribution. (See Figure 1.6.)

¹. This functionality of using different scales is sometimes referred to as drawing the distribution on different types of “probability paper”.

Using a Fréchet scale, the nonparametric estimates approximate a straight line, indicating that a Fréchet fit might be reasonable.

**Fitting a Fréchet Distribution**

To fit a Fréchet distribution, click the check box to the left of Fréchet under Distribution. A Fréchet cumulative distribution function (CDF), with corresponding confidence intervals, is drawn on the graph. (See Figure 1.7.)
This same distribution can be seen with a linear scale in the Distribution Profiler. (See Figure 1.8.)

The Distribution Profiler works like the profilers in other platforms. See the *Modeling and Multivariate Methods* book. The vertical line can be dragged to change the value on the x-axis. Corresponding values on the y-axis show the estimated value and confidence bands for the selected distribution.
Fitting Multiple Parametric Distributions

To fit additional parametric distributions, click their corresponding check boxes. (See Figure 1.9.) The added distributions show up as lines or curves in the main graph and in the Distribution Profiler. Dragging the vertical line in the Distribution Profiler shows estimates and confidence intervals for the selected distributions.

Figure 1.9 Fan Data with Fréchet and Loglogistic Distributions in a Fréchet-Scaled Plot

Extrapolating Data

To extrapolate into other ranges of the data, drag the endpoints of the axes, or double-click them to bring up a window for entering minimum and maximum values. For example, Figure 1.10 shows the y-axis going from 0 to 1, and the x-axis extended to 30000. In addition, the scale is set to Nonparametric, but the Lognormal and Weibull distributions are shown. Since the scale is set to Nonparametric, the linear scale is used for both the x- and y-axes. Dots shown on the plot correspond to the midpoint estimates.
Statistics

Figure 1.11 shows the default Statistics report, with the Summary of Data and Nonparametric Estimate tables revealed by clicking the disclosure buttons. Since, by default, no distribution has been selected from the launch window, only nonparametric statistics are available. Statistics are appended to the report window after you select one or more distributions from either the check boxes under Distribution in the Compare Distributions report section, or from Fit All Distributions or Fit All Non-negative from the red-triangle menu of the Life Distribution outline title. (See “Parametric Statistics Report,” p. 15.)

The Summary of Data table includes the number of observations, the number of uncensored values (failure events), and the number of right-censored values for the time-to-event data. The Nonparametric Estimate table includes both midpoint and Kaplan-Meier estimates. Ninety-five percent pointwise confidence intervals are shown for the estimates. By default, the Life Distribution platform uses the midpoint estimates of the step function to construct probability plots. Kaplan-Meier estimates are shown instead, when the Show Points option is disengaged. When you toggle off the Show Points option in the red-triangle menu of the platform, the midpoint estimates are removed and replaced by Kaplan-Meier estimates. Figure 1.12 shows the probability plot with Kaplan-Meier estimates.

When you select parametric distributions by clicking the check boxes under the Compare Distributions outline title, or if you select Fit All Distributions or Fit All Non-negative from the red-triangle menu, additional statistics are appended to the Statistics report section. “Parametric Statistics Report,” p. 15, provides more detail.
Figure 1.11 Default Statistics for the Nonparametric Distribution

Summary of Data
- Observation Used: 70
- Uncensored Values: 12
- Right-Censored Values: 58

Nonparametric Estimates

<table>
<thead>
<tr>
<th>Start</th>
<th>Midpoint Estimate</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Lower 95% (Non)</th>
<th>Upper 95% (Non)</th>
<th>Kaplan-Meier Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>450.00</td>
<td>0.0091429</td>
<td>0.0092977</td>
<td>0.0094414</td>
<td>0.0094903</td>
<td>0.0095488</td>
<td>0.001426571</td>
</tr>
<tr>
<td>1150.00</td>
<td>0.0215036</td>
<td>0.0217287</td>
<td>0.0218958</td>
<td>0.0219406</td>
<td>0.0219756</td>
<td>0.0219756</td>
</tr>
<tr>
<td>1150.00</td>
<td>0.0307041</td>
<td>0.0404238</td>
<td>0.1257705</td>
<td>0.1267385</td>
<td>0.1267892</td>
<td>0.1267892</td>
</tr>
<tr>
<td>1150.00</td>
<td>0.0500372</td>
<td>0.0219353</td>
<td>0.1458632</td>
<td>0.1462507</td>
<td>0.1463573</td>
<td>0.1463573</td>
</tr>
<tr>
<td>1150.00</td>
<td>0.0666596</td>
<td>0.1956815</td>
<td>0.2698438</td>
<td>0.2756944</td>
<td>0.2756944</td>
<td>0.2756944</td>
</tr>
<tr>
<td>2070.00</td>
<td>0.0395974</td>
<td>0.0415665</td>
<td>0.1912092</td>
<td>0.2037521</td>
<td>0.2043744</td>
<td>0.2043744</td>
</tr>
<tr>
<td>2070.00</td>
<td>0.1004148</td>
<td>0.0592812</td>
<td>0.2138194</td>
<td>0.2153434</td>
<td>0.2153434</td>
<td>0.2153434</td>
</tr>
<tr>
<td>2070.00</td>
<td>0.1105295</td>
<td>0.0865649</td>
<td>0.2374119</td>
<td>0.2374119</td>
<td>0.2374119</td>
<td>0.2374119</td>
</tr>
<tr>
<td>3450.00</td>
<td>0.1300129</td>
<td>0.0670113</td>
<td>0.2417317</td>
<td>0.2417317</td>
<td>0.2417317</td>
<td>0.2417317</td>
</tr>
<tr>
<td>4600.00</td>
<td>0.1663169</td>
<td>0.0584465</td>
<td>0.2855328</td>
<td>0.2855328</td>
<td>0.2855328</td>
<td>0.2855328</td>
</tr>
<tr>
<td>4600.00</td>
<td>0.1886741</td>
<td>0.1132539</td>
<td>0.3412313</td>
<td>0.3412313</td>
<td>0.3412313</td>
<td>0.3412313</td>
</tr>
<tr>
<td>4600.00</td>
<td>0.2497232</td>
<td>0.1401882</td>
<td>0.5168187</td>
<td>0.5168187</td>
<td>0.5168187</td>
<td>0.5168187</td>
</tr>
</tbody>
</table>

Figure 1.12 Kaplan-Meier Estimates for the Nonparametric Distribution

Life Distribution

Event Plot

Compare Distributions
Understanding Platform Options

The following menu options are found by clicking on the red triangle of the Life Distribution report window:

**Fit All Distributions** fits all distributions and shows the best fit. The criteria for choosing the best distributions can be changed with the **Comparison Criterion** option.

**Fit All Non-negative** behaves differently depending on the data:
- if the data have negative values, it has no effect.
- if the data have zero values, it fits all four zero-inflated (ZI) distributions, including the ZI lognormal, ZI Weibull, ZI loglogistic, and the ZI Fréchet distributions.
- if the above two cases do not apply, it fits the lognormal, loglogistic, Fréchet, Weibull, and generalized gamma (GenGamma) distributions.

**Show Points** toggles the data points on and off in the probability plot under Compare Distributions. The points shown are the midpoint estimates of the step function and are shown in the plots by default. If this option is unchecked, the step functions (Kaplan-Meier estimates) are shown instead. If the Nonparametric check box under Compare Distributions is unchecked, this option is automatically turned on, and the confidence interval for the step function is not shown.

**Show Survival Curve** toggles between showing the failure probability and the survival curves on the Compare Distributions plot.

**Show Quantile Functions** shows or hides the Quantile Profiler for the selected distribution(s).

**Show Hazard Functions** shows or hides the Hazard Profiler for the selected distribution(s).

**Show Statistics** shows or hides an outline node that shows parameter estimates, profilers, a covariance matrix, and a Loglikelihood Contour for each distribution. (See the section “Parametric Statistics Report,” p. 15 for details.)

**Tabbed Report** toggles the layout of the platform report. Initially, the report is a conventional outline, with results stacked.

**Show Confidence Area** toggles the shaded confidence regions in the plots.

**Interval Type** is the type of confidence intervals shown on the Compare Distributions plot for the Nonparametric distribution fit.

**Change Confidence Level** produces a window that lets you change the confidence level for the entire platform. All plots and reports update accordingly.

**Comparison Criterion** produces a window that lets you select a formal comparison criterion for an analysis. Table 1.1 outlines the available criteria.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Formulaa</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2loglikelihood</td>
<td>Not applicable</td>
<td>Minus two times the natural log of the likelihood function evaluated at the best-fit parameter estimates</td>
</tr>
</tbody>
</table>
The comparison criterion you select should be based on knowledge and history of the data as well as personal preference. Burnham and Anderson (2004) discuss using AICc and BIC for model selection. Akaike (1974) discusses AICc for model selection. For AICc and BIC, models with smaller values are better.

### Navigating Report Details

As you select new distributions by clicking the check box beside distributions in the Compare Distributions report section, or if you select Fit All Distributions or Fit All Non-negative from the red-triangle menu of the Life Distribution outline title, statistics for each requested distribution are appended to the report window. Appended statistics include model criteria comparisons, parameter estimates, a covariance matrix, and distribution, quantile, hazard, and density profilers for each distribution. Report layout options can be modified with the Tabbed Report option selected.

### Profilers

For each distribution you select, four types of profilers appear: Distribution, Quantile, Hazard, and Density Profilers are shown in the Statistics outline for each distribution. The profilers in the Life Distribution platform work like the profilers in other platforms. See the Modeling and Multivariate Methods book. With each profiler, the vertical line can be dragged to change the value on the x-axis. Corresponding values on the y-axis show the estimated value and confidence bands for the selected distribution.

The Distribution and Density Profilers display the cumulative distribution function (cdf) and probability density function (pdf) on the y-axis, respectively, with the time-to-event variable displayed on the x-axis.

The Quantile Profiler displays quantile values on the y-axis and cumulative probabilities on the x-axis. The Hazard Profiler displays hazard values on the y-axis and quantile values on the x-axis. Selecting the options Show Quantile Functions or Show Hazard Functions from the red-triangle menu of the platform shows a quantile plot or hazard plot, respectively, for the selected distributions. Profilers are placed vertically for selected distributions. See Figure 1.13 for an example showing the Quantile Profiler, with the Lognormal and Weibull distributions selected.

---

**Table 1.1 Comparison Criteria (Continued)**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Formula&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>$\text{BIC} = -2\log\text{likelihood} + k\ln(n)$</td>
<td>Schwarz’s Bayesian Information Criterion</td>
</tr>
<tr>
<td>AICc</td>
<td>$\text{AICc} = -2\log\text{likelihood} + 2k\left(\frac{n}{n-k-1}\right)$</td>
<td>Corrected Akaike’s Information Criterion</td>
</tr>
<tr>
<td>AICc</td>
<td>$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}$</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> $k =$ The number of estimated parameters in the model; $n =$ The number of observations in the data set.
Parametric Statistics Report

The Statistics report consists of a summary of all analyses, and a collection of reports on individual analyses. The Model Comparisons section, presented first, tabulates -2loglikelihood, corrected AIC, and BIC statistics for each distribution previously selected. (See Figure 1.14.) For this example, a comparison of all possible distributions is shown by clicking on the option Fit All Distributions in the platform menu. By default, models are compared using the corrected AIC, or AICc, where the models with the lower values are considered better. The Model Comparisons table sorts the models by the best (lower) score or parametric fit, with the “better-fitting” models shown at the top of the table.

---

Figure 1.13  Quantile Profiler Showing Lognormal and Weibull Distributions
Note: The Model Comparison table is sorted based on the **Comparison Criterion** option, found under the red-triangle menu for Life Distribution.

For each distribution, there is an outline node that shows parameter estimates, a covariance matrix, the confidence intervals, summary statistics, and profilers. Using the red-triangle menu (Figure 1.15), you can save probability, quantile, and hazard estimates. You can also select the Likelihood Contour plot for each parametric distribution with two parameters. The Fix Parameters option enables you to fix the distribution parameters, and investigate the constrained distribution using the profilers. The Custom Estimation option allows you to estimate specific failure probabilities and quantiles, using both Wald and Profile confidence intervals. Figure 1.16 shows the Weibull distribution report section, with the Likelihood Contour plot.
Chapter 1

Lifetime Distribution
Navigating Report Details

17

Figure 1.16 Estimates, Profilers, and Contour for Weibull Distribution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>StdErr</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>location</td>
<td>15.177</td>
<td>0.446</td>
<td>12.284</td>
<td>18.068</td>
<td>-7.5logLikelihood</td>
</tr>
<tr>
<td>scale</td>
<td>0.845</td>
<td>0.239</td>
<td>0.376</td>
<td>1.416</td>
<td>AICc</td>
</tr>
<tr>
<td>Weibull a</td>
<td>2829.846</td>
<td>1225.129</td>
<td>1056.070</td>
<td>5653.448</td>
<td>BC</td>
</tr>
<tr>
<td>Weibull b</td>
<td>1.059</td>
<td>0.707</td>
<td>2.163</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: For the Weibull distribution, the Fix Parameter option provides for the Weibayes method.

The Tabbed Report Option

The Tabbed Report option provides a way for you to alter the layout of the Life Distribution platform. A series of tabs is shown for each of the platform options previously chosen. An advantage of this report is that you can concentrate on just one option at a time by clicking a tab of interest. Content from the other tabs remains hidden until you select them. For example, using the Fan.jmp data, select Show Hazard Functions from the Life Distribution drop-down menu. Then, click on the Tabbed Report option in the platform menu, and click on the Hazard Profiler tab. Figure 1.17 shows a hazard plot and profilers with the Lognormal and Weibull distributions selected.
Mixed Censoring Example

The Fan example uses data with right-censoring. The next example shows a mixed-censoring data set, with left, right, and interval censoring. To launch the platform:

1. Open the Microprocessor Data.jmp sample data table.
2. Select Analyze > Reliability and Survival > Life Distribution.
3. Select start time and end time as Y, Time to Event.
4. Select count as Freq.

Figure 1.18 shows the completed launch window.
5. Click **OK**.

Figure 1.19 shows the initial report.

Click on the disclosure button for the Event Plot and stretch the graphic by clicking and dragging inside the plot window to see the Microprocessor Data.jmp Event Plot. (See Figure 1.20.)
Note the differences in the Event plot and Probability Plot between this example and the right-censored Fan.jmp sample data, shown previously in Figure 1.4. Microprocessor Data.jmp includes right-censoring, left censoring, and interval censoring. This Event Plot includes frequency values for each type of censoring, since a Freq column was specified in the launch window. Notice also that:

- Right-censored values are indicated by a solid line, a left triangle, and a dashed line to the right of the left triangle.
- Left-censored values are indicated by a dashed line followed by a right triangle.
- Interval-censored values are shown by a solid line, a left triangle, a dashed line, and a right triangle.

**Illustration of Censoring Differences**

In order to better illustrate the censoring differences, a data set was fabricated to better illustrate censoring differences. To launch the example:

1. Open the Censor Labels.jmp sample data table.
2. Select Analyze > Reliability and Survival > Life Distribution.
3. Select start time and end time as Y, Time to Event.
4. Select count as Freq.
5. Select Censor as Censor.
6. Click OK.
7. Click the disclosure button for the Event Plot.
8. Stretch the plot, vertically and horizontally. (See Figure 1.21.)
Figure 1.21 Event Plot with Fabricated Data to Illustrate Censoring Types

Figure 1.22 shows the probability plot for the Microprocessor Data with the Loglogistic scale chosen. Note that for right-censored or complete data, the Kaplan-Meier method of estimation is used; interval, left, or mixed-censored data use the Turnbull estimation method. Compare this with the previous example in which the data consists of only failures and right-censored observations. (See “Fan Nonparametric Estimates with a Fréchet Probability Scale,” p. 8.)
Estimating Competing Causes

Many products, systems, or components have more than one failure cause, or failure mode. It is important to distinguish among the different causes of failure for such units to better understand how to improve product reliability and quality. Suppose a system or unit has more than one failure mode, and the different modes are independent. The failure times can be modeled by estimating a failure time for each failure mode. For each failure mode analysis, observations that do not fall into that mode are treated as censored. Meeker and Escobar (1998, chap. 15) discuss product reliability for two or more failure causes.

An example of a product with multiple failure causes can be found in Blenders.jmp, in the Reliability subfolder. To launch the Life Distribution platform:

1. Open the Blenders.jmp sample data table.
2. Select Analyze > Reliability and Survival > Life Distribution.
4. Select Causes as Failure Cause.
5. Select Censor as Censor.

The default value for Censor Code is 1. If your Censor column contains any other value for censored observations, enter that value in the entry field to the right of Censor Code. Observations coded differently from the entry in this field are treated as uncensored.
6. Select Individual Best as the Distribution.
   This is one of the menu options available under Failure Distribution by Cause.

7. Select AICc as the Comparison Criterion.
   This is an additional menu selection that is available only when Individual Best is selected as the Distribution.
   Figure 1.23 shows the completed launch window.

8. Click OK.

---

**Figure 1.23** Blender Example Competing Causes Launch Window

![Launch Window Image](image)

---

**Additional Life Distribution Launch Options**

**By, Freq, and Label Variables**

You can also analyze the sample data, Blenders.jmp, using the optional grouping variable by assigning Group to the By role. Group, in this case, is either Automatic or Manual. For this example, the By grouping variable is not used.

Freq and Label Variables, if available in your data set, can also be assigned. See “Launch Window Buttons,” p. 4.

**Confidence Interval Method**

Confidence interval method options for any of the distributional choices are Wald or Profile. Either option is selected from the Select Confidence Interval Method section of the launch window. Wald is the default confidence interval method.

**Failure Distribution by Cause**

When you select a column of your data set to be included in the Failure Cause role of the Life Distribution launch window, additional options become available. These options (located in the lower left corner) are listed under Failure Distribution by Cause. Menu options are available for selecting a failure distribution,
and a check box is available to indicate the censoring value used in the **Failure Cause** column. If you select **Individual Best** as your failure distribution, you also have the option of selecting a comparison criterion.

Six menu options are available for fitting the failure distribution. **Lognormal** is the default distribution. Other choices include **Weibull**, **Loglogistic**, **Fréchet**, **Individual Best**, and **Manual Pick**. If you select **Individual Best** as your failure distribution, JMP chooses the best distributional fit based on your criterion choice, which is selected from the **Comparison Criterion** menu. Criterion options include **AICc**, **BIC**, and **-2Loglikelihood**. See Table 1.1 for criteria formulas and descriptions. Using your criterion of choice, JMP selects the best fit among the Weibull, lognormal, loglogistic, and Fréchet distributions for each cause. The **Manual Pick** option allows you to select from one to all four of the fits for your data, for each of the failure modes.

Notice that the launch window contains a check box and an entry field next to Censor Indicator in Failure Cause Column. The check box is checked when some of the observations are not failure events and a **Censor** column is not included in the data set. Besides checking this box, you enter the value representing censored values in your **Failure Cause** column into the entry field. Figure 1.24 shows the check box and the entry field. For **Blenders.jmp**, you do not need to check this option since a **Censor** column is already included in the data set. (The check box beside Censor Indicator in Failure Cause Column could be checked and the entry field left blank, since the censored values in the **Causes** column in Blenders.jmp are blank.)

---

**Figure 1.24** Censor Indicator Check Box and Entry Field for Failure Cause Column

---

**Navigating the Competing Cause Report Window**

The Competing Cause report window contains three sections:

- “**Cause Combination,**” p. 24
- “**Statistics,**” p. 25
- “**Individual Causes,**” p. 26

**Tabbed Report** is selected from the red triangle of the Competing Cause outline title to separately view the three sections of the report.

**Cause Combination**

**Cause Combination** is the first report section shown and is the first tab in the **Tabbed Report** option. Figure 1.25 shows the probability plot for each of the causes. Curves for each failure mode are shown using a linear probability scale, by default. In addition to the linear scale, radio buttons on the left of the report are available for you to select and view linearized curves with four different parametric probability scales. The Fréchet distribution is chosen as the best overall fit for the data, based on AICc. This can also be seen by
selecting the different probability scales and viewing the colored curves in the plot. These curves become straight lines when the Fréchet probability scale is selected. (See Figure 1.26.)

Figure 1.25 Blender Data Competing Causes Curves with Linear Probability Scale

Figure 1.26 Blender Data Competing Causes with Fréchet Probability Scale

Statistics

The second section of the report is labeled Statistics and includes the Cause Summary and the Distribution, Quantile, Hazard, and Density Profilers for the chosen distribution. (See Figure 1.27.)
The Cause Summary report section lists six columns. The Cause column presents either labels of causes or an indicator for the censoring. The second column indicates whether the cause has enough failure events for inference—the number of events must be greater than or equal to two. If the cause has previously been selected for omission, this column also indicates omission of that cause. The Counts column lists the number of failures for each cause. The Distribution column specifies the chosen distribution for the individual causes. Finally, the location and scale columns display the parametric estimates for each cause.

The Distribution, Quantile, Hazard, and Density profilers are also listed under the Cause Summary. As with other platforms, you can explore various perspectives of your data with these profilers. (See the Modeling and Multivariate Methods book for more information.)

**Individual Causes**

The third section of the Competing Cause report window is labeled Individual Causes and consists of Life Distribution reports for each individual cause of failure. Eight modes of failure where there were two or more failure events are reported for the blender sample data. Each Individual Causes report is opened by clicking on the disclosure button to the left of the Failure Cause label. Frequency counts are shown in the title bar for each failure cause. Figure 1.28 shows a probability plot and Distribution Profiler with the DS Weibull distribution and the Nonparametric probability scale selected for the power switch failure cause.
A useful feature of this platform is that you can omit one or more of the competing causes. This becomes important when making decisions about which product improvements or process modifications to make, or when it is determined that a particular cause is no longer relevant. Click on the Cause Combination tab. Figure 1.29 shows the competing cause curves when the four lowest frequency count causes are omitted by checking the Omit check boxes to the left of the Cause label. Notice how the curve for the overall data (Aggregated) has now shifted slightly downward and to the right of the plot. In general, shifting of the curve can show that the product’s lifetime can be improved with the omission of certain failure causes. Statistics for the whole model are reevaluated and can readily be assessed with the selection of omitted causes.
Menu Options
As shown earlier with the Life Distribution platform, several options are available from the red-triangle menu for each individual failure cause. Quantile functions, hazard functions, statistics, a tabbed report, and options to change the confidence level and comparison criterion are available. Figure 1.30 shows the menu options.

Statistics
Statistics for each individual cause are shown by clicking the disclosure button for the individual cause. Model comparisons, parameter estimates, and distribution, quantile, hazard, and density profilers are shown, by default, in the report window. The Summary of Data, Nonparametric Estimate, and Covariance Matrix tables for each failure cause are shown by clicking the disclosure buttons.
Figure 1.31 shows the Model Comparisons table for the power switch competing cause. The comparison criteria indicate that the DS Weibull distribution is the most appropriate fit (lowest AICc score) for this competing cause.

**Figure 1.31 Model Comparison Statistics for Power Switch Failure Mode**

<table>
<thead>
<tr>
<th>Distribution</th>
<th>AICc</th>
<th>-2LogLikelihood</th>
<th>BIC</th>
<th>Convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEV</td>
<td>269.8176</td>
<td>283.6328</td>
<td>277.0242</td>
<td>Failed: Cannot Calculate Objective at Starting Value</td>
</tr>
<tr>
<td>DS Weibull</td>
<td>272.5381</td>
<td>287.2574</td>
<td>290.7507</td>
<td></td>
</tr>
<tr>
<td>DS Logistic</td>
<td>278.1383</td>
<td>292.8692</td>
<td>296.3617</td>
<td></td>
</tr>
<tr>
<td>DS Frechet</td>
<td>279.4329</td>
<td>293.1539</td>
<td>296.5565</td>
<td></td>
</tr>
<tr>
<td>Frechet</td>
<td>280.1339</td>
<td>293.9554</td>
<td>297.3580</td>
<td></td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>280.8704</td>
<td>294.5912</td>
<td>298.9939</td>
<td></td>
</tr>
<tr>
<td>Threshold Frechet</td>
<td>282.2463</td>
<td>295.9672</td>
<td>300.3702</td>
<td>Failed: Cannot Decrease Objective Function</td>
</tr>
<tr>
<td>Lognormal</td>
<td>294.4472</td>
<td>308.0695</td>
<td>311.4720</td>
<td>Failed: Cannot Decrease Objective Function</td>
</tr>
<tr>
<td>Loglogistic</td>
<td>294.1330</td>
<td>307.9443</td>
<td>311.3466</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>285.8969</td>
<td>300.5182</td>
<td>303.9207</td>
<td></td>
</tr>
<tr>
<td>Threshold Logistic</td>
<td>286.5353</td>
<td>300.1566</td>
<td>303.5591</td>
<td>Failed: Cannot Decrease Objective Function</td>
</tr>
<tr>
<td>Exponential</td>
<td>286.6012</td>
<td>300.2177</td>
<td>303.6203</td>
<td></td>
</tr>
<tr>
<td>Threshold Weibull</td>
<td>297.6410</td>
<td>308.2620</td>
<td>311.6646</td>
<td>Failed: Cannot Decrease Objective Function</td>
</tr>
<tr>
<td>Log Generalized Gamma</td>
<td>291.0410</td>
<td>304.6620</td>
<td>308.0656</td>
<td>Failed: Cannot Decrease Objective Function</td>
</tr>
<tr>
<td>LEV</td>
<td>311.6033</td>
<td>325.2254</td>
<td>328.6289</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>311.6231</td>
<td>325.7259</td>
<td>329.1295</td>
<td></td>
</tr>
<tr>
<td>Logistic</td>
<td>323.2200</td>
<td>336.8420</td>
<td>340.2456</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.32 shows the DS Weibull location and scale parameter estimates, standard errors, confidence intervals, and criteria scores.

**Figure 1.32 DS Weibull Location and Scale Parameters for Power Switch Failure Mode**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated Value</th>
<th>Std Error</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>5.6474</td>
<td>0.1201</td>
<td>5.3945</td>
<td>5.9003</td>
<td>-2LogLikelihood</td>
</tr>
<tr>
<td>Scale</td>
<td>8.4945</td>
<td>0.8444</td>
<td>6.8511</td>
<td>10.1389</td>
<td>AICc</td>
</tr>
<tr>
<td>p</td>
<td>0.22210</td>
<td>0.053857</td>
<td>0.114181</td>
<td>0.330048</td>
<td>BIC</td>
</tr>
<tr>
<td>Weibull Location</td>
<td>67.4085</td>
<td>4.860546</td>
<td>58.05573</td>
<td>76.76137</td>
<td>AICc</td>
</tr>
<tr>
<td>Weibull Scale</td>
<td>2.45989</td>
<td>0.169659</td>
<td>2.12928</td>
<td>2.89050</td>
<td>BIC</td>
</tr>
</tbody>
</table>

**Statistical Details**

This section provides details for the distributional fits in the Life Distribution platform. Meeker and Escobar (1998, chaps. 2-5) is an excellent source of theory, application, and discussion for both the nonparametric and parametric details that follow.

The parameters of all distributions, unless otherwise noted, are estimated via maximum likelihood estimates (MLEs). The only exceptions are the threshold distributions. If the smallest observation is an exact failure,
then this observation is treated as interval-censored with a small interval. The parameter estimates are the MLEs estimated from this slightly modified data set. Without this modification, the likelihood can be unbounded, so an MLE might not exist. This approach is similar to that described in Meeker and Escobar (1998, p. 275), except that only the smallest exact failure is censored. This is the minimal change to the data that guarantees boundedness of the likelihood function.

Two methods exist in the Life Distribution platform for calculating the confidence intervals of the distribution parameters. These methods are labeled as Wald or Profile (profile-likelihood) and can be selected in the launch window for the Life Distribution platform. Profile confidence intervals are used as the default setting. The confidence intervals for the cumulative distribution function (cdf) are calculated by computing Wald confidence intervals on the standardized variable and then transforming the intervals to the cdf scale (Nelson, 1982, pp. 332-333 and pp. 346-347). The confidence intervals given in the other graphs and profilers are transformed Wald intervals (Meeker and Escobar, 1998, chap. 7). Joint confidence intervals for the parameters of a two-parameter distribution are shown in the loglikelihood contour plots. They are based on approximate likelihood ratios for the parameters (Meeker and Escobar, 1998, chap. 8).

Nonparametric Fit

A nonparametric fit uses a Kaplan-Meier estimator for data with no censoring (failures only) and for data where the observations consist of both failures and right-censoring. The nonparametric fit uses the Turnbull estimator if mixed censoring is detected.

Parametric Distributions

Many parametric distributions are available for assessing reliability data. The parametric distributions used in the Life Distribution platform are discussed in this section. Formulas for each of the parametric distributions are shown.

**Note:** Parameter estimates for each of the distributions are parameterized in the JMP output as location and scale. Parameter estimates for threshold distributions also include a threshold estimate. Location corresponds to $\mu$, scale corresponds to $\sigma$, and threshold corresponds to $\gamma$.

**Lognormal**

Lognormal distributions are used commonly for failure times when the range of the data is several powers of ten. This distribution is often considered as the multiplicative product of many small positive identically independently distributed random variables. It is reasonable when the log of the data values appears normally distributed. Examples of data appropriately modeled by the lognormal distribution include hospital cost data, metal fatigue crack growth, and the survival time of bacteria subjected to disinfectants. The pdf curve is usually characterized by strong right-skewness. The lognormal pdf and cdf are

$$f(x; \mu, \sigma) = \frac{1}{x \sigma \phi_{\text{nor}}} \left(\frac{\log(x) - \mu}{\sigma}\right), \quad x > 0$$

$$F(x; \mu, \sigma) = \Phi_{\text{nor}} \left(\frac{\log(x) - \mu}{\sigma}\right),$$
where

\[ \phi_{\text{nor}}(z) = \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{z^2}{2} \right) \]

and

\[ \Phi_{\text{nor}}(z) = \int_{-\infty}^{z} \phi_{\text{nor}}(w) \, dw \]

are the pdf and cdf, respectively, for the standardized normal, or nor(μ = 0, σ = 1) distribution.

**Weibull**

The Weibull distribution can be used to model failure time data with either an increasing or a decreasing hazard rate. It is used frequently in reliability analysis because of its tremendous flexibility in modeling many different types of data, based on the values of the shape parameter, β. This distribution has been successfully used for describing the failure of electronic components, roller bearings, capacitors, and ceramics. Various shapes of the Weibull distribution can be revealed by changing the values of both the scale parameter, α, and the shape parameter, β (See Fit Distribution in the Advanced Univariate Analysis chapter, p. 61.) The Weibull pdf and cdf are commonly represented as

\[ f(x; \alpha, \beta) = \frac{\beta}{\alpha^\beta} x^{(\beta - 1)} \exp\left[ -\left( \frac{x}{\alpha} \right)^\beta \right] \quad \text{for } x > 0, \alpha > 0, \beta > 0 \]

\[ F(x; \alpha, \beta) = 1 - \exp\left[ -\left( \frac{x}{\alpha} \right)^\beta \right] \]

where α is a scale parameter, and β is a shape parameter. The Weibull distribution is particularly versatile since it reduces to an exponential distribution when β = 1. An alternative parameterization commonly used in the literature and in JMP is to use σ as the scale parameter and μ as the location parameter. These are easily converted to an α and β parameterization by

\[ \alpha = \exp(\mu) \]

and

\[ \beta = \frac{1}{\sigma} \]

The pdf and the cdf of the Weibull distribution are also expressed as a log-transformed smallest extreme value distribution, or sev, using a location scale parameterization, with μ = log(α) and σ = 1/β,

\[ f(x; \mu, \sigma) = \frac{1}{x\sigma} \Phi_{\text{sev}} \left[ \frac{\log(x) - \mu}{\sigma} \right] \quad \text{for } x > 0, \sigma > 0 \]

\[ F(x; \mu, \sigma) = \Phi_{\text{sev}} \left[ \frac{\log(x) - \mu}{\sigma} \right] \]
where
\[ \phi_{sev}(z) = \exp\left[z - \exp(z)\right] \]
and
\[ \Phi_{sev}(z) = 1 - \exp[-\exp(z)] \]
are the pdf and cdf, respectively, for the standardized smallest extreme value (μ = 0, σ = 1) distribution.

**Loglogistic**

The pdf of the loglogistic distribution is similar in shape to the lognormal distribution but has heavier tails. It is often used to model data exhibiting non-monotonic hazard functions, such as cancer mortality and financial wealth. The loglogistic pdf and cdf are

\[
\begin{align*}
    f(x; \mu, \sigma) &= \frac{1}{x\sigma} \phi_{logis}\left[\frac{\log(x) - \mu}{\sigma}\right] \\
    F(x; \mu, \sigma) &= \Phi_{logis}\left[\frac{\log(x) - \mu}{\sigma}\right],
\end{align*}
\]

where
\[ \phi_{logis}(z) = \frac{\exp(z)}{[1 + \exp(z)]^2} \]
and
\[ \Phi_{logis}(z) = \frac{\exp(z)}{[1 + \exp(z)]} = \frac{1}{1 + \exp(-z)} \]
are the pdf and cdf, respectively, for the standardized logistic or logis distribution (μ = 0, σ = 1).

**Fréchet**

The Fréchet distribution is known as a log-largest extreme value distribution or sometimes as a Fréchet distribution of maxima when it is parameterized as the reciprocal of a Weibull distribution. This distribution is commonly used for financial data. The pdf and cdf are

\[
\begin{align*}
    f(x; \mu, \sigma) &= \exp\left[-\exp\left(-\frac{\log(x) - \mu}{\sigma}\right)\right]\exp\left(-\frac{\log(x) - \mu}{\sigma}\right)\frac{1}{x\sigma} \\
    F(x; \mu, \sigma) &= \exp\left[-\exp\left(-\frac{\log(x) - \mu}{\sigma}\right)\right]
\end{align*}
\]

and are more generally parameterized as
Chapter 1

Lifetime Distribution

Statistical Details

where

\[ f(x; \mu, \sigma) = \frac{1}{x \sigma} \phi_{\text{lev}} \left[ \frac{\log(x) - \mu}{\sigma} \right] \]

\[ F(x; \mu, \sigma) = \Phi_{\text{lev}} \left[ \frac{\log(x) - \mu}{\sigma} \right] \]

where

\[ \phi_{\text{lev}}(z) = \exp[-\log(-z)] \]

and

\[ \Phi_{\text{lev}}(z) = \exp[-\log(-z)] \]

are the pdf and cdf, respectively, for the standardized largest extreme value, or lev(\(\mu = 0, \sigma = 1\)) distribution.

**Normal**

The normal distribution is the most widely used distribution in most areas of statistics because of its relative simplicity and the ease of applying the central limit theorem. However, it is rarely used in reliability. It is most useful for data where \(\mu > 0\) and the coefficient of variation (\(\sigma/\mu\)) is small. Because the hazard function increases with no upper bound, it is particularly useful for data exhibiting wearout failure. Examples include incandescent light bulbs, toaster heating elements, and mechanical strength of wires. The pdf and cdf are

\[ f(x; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \phi_{\text{nor}} \left( \frac{x - \mu}{\sigma} \right), \quad -\infty < x < \infty \]

\[ F(x; \mu, \sigma) = \Phi_{\text{nor}} \left( \frac{x - \mu}{\sigma} \right) \]

where

\[ \phi_{\text{nor}}(z) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{z^2}{2} \right) \]

and

\[ \Phi_{\text{nor}}(z) = \int_{-\infty}^{z} \phi_{\text{nor}}(w) \, dw \]

are the pdf and cdf, respectively, for the standardized normal, or nor(\(\mu = 0, \sigma = 1\)) distribution.
**Smallest Extreme Value (SEV)**

This non-symmetric (left-skewed) distribution is useful in two cases. The first case is when the data indicate a small number of weak units in the lower tail of the distribution (the data indicate the smallest number of many observations). The second case is when $\sigma$ is small relative to $\mu$, because probabilities of being less than zero, when using the SEV distribution, are small. The smallest extreme value distribution is useful to describe data whose hazard rate becomes larger as the unit becomes older. Examples include human mortality of the aged and rainfall amounts during a drought. This distribution is sometimes referred to as a Gumbel distribution. The pdf and cdf are

$$f(x;\mu,\sigma) = \frac{1}{\sigma} \phi_{\text{sev}}\left(\frac{x-\mu}{\sigma}\right), \quad -\infty < \mu < \infty, \quad \sigma > 0$$

$$F(x;\mu,\sigma) = \Phi_{\text{sev}}\left(\frac{x-\mu}{\sigma}\right)$$

where

$$\phi_{\text{sev}}(z) = \exp[z - \exp(z)]$$

and

$$\Phi_{\text{sev}}(z) = 1 - \exp[-\exp(z)]$$

are the pdf and cdf, respectively, for the standardized smallest extreme value, or sev ($\mu = 0, \sigma = 1$) distribution.

**Logistic**

The logistic distribution has a shape similar to the normal distribution, but with longer tails. Logistic regression models for a binary or ordinal response are often used to model life data when negative failure times are not an issue. The pdf and cdf are

$$f(x;\mu,\sigma) = \frac{1}{\sigma} \phi_{\text{logis}}\left(\frac{x-\mu}{\sigma}\right), \quad -\infty < \mu < \infty \text{ and } \sigma > 0.$$

$$F(x;\mu,\sigma) = \Phi_{\text{logis}}\left(\frac{x-\mu}{\sigma}\right)$$

where

$$\phi_{\text{logis}}(z) = \frac{\exp(z)}{[1 + \exp(z)]^2}$$

and

$$\Phi_{\text{logis}}(z) = \frac{\exp(z)}{[1 + \exp(z)]} = \frac{1}{1 + \exp(-z)}$$
are the pdf and cdf, respectively, for the standardized logistic or logis distribution ($\mu = 0$, $\sigma = 1$).

**Largest Extreme Value (LEV)**

This right-skewed distribution can be used to model failure times if $\sigma$ is small relative to $\mu > 0$. This distribution is not commonly used in reliability but is useful for estimating natural extreme phenomena, such as a catastrophic flood heights or extreme wind velocities. The pdf and cdf are

$$f(x; \mu, \sigma) = \frac{1}{\sigma} \phi_{\text{lev}}\left(\frac{x - \mu}{\sigma}\right), \quad -\infty < x < \infty \quad \text{and} \quad \sigma > 0.$$  

$$F(x; \mu, \sigma) = \Phi_{\text{lev}}\left(\frac{x - \mu}{\sigma}\right)$$

where

$$\phi_{\text{lev}}(z) = \exp[-z - \exp(-z)]$$

and

$$\Phi_{\text{lev}}(z) = \exp[-\exp(-z)]$$

are the pdf and cdf, respectively, for the standardized largest extreme value, or $\text{lev}(\mu = 0, \sigma = 1)$ distribution.

**Exponential**

Both one- and two-parameter exponential distributions are used in reliability. The pdf and cdf for the two-parameter exponential distribution are

$$f(x; \theta, \gamma) = \frac{1}{\theta} \exp\left(-\frac{x - \gamma}{\theta}\right), \quad \theta > 0.$$  

$$F(x; \theta, \gamma) = 1 - \exp\left(-\frac{x - \gamma}{\theta}\right)$$

where $\theta$ is a scale parameter and $\gamma$ is both the threshold and the location parameter. Reliability analysis frequently uses the one-parameter exponential distribution, with $\gamma = 0$. The exponential distribution is useful for describing failure times of components exhibiting wearout far beyond their expected lifetimes. This distribution has a constant failure rate, which means that for small time increments, failure of a unit is independent of the unit’s age. The exponential distribution should not be used for describing the life of mechanical components that can be exposed to fatigue, corrosion, or short-term wear. This distribution is, however, appropriate for modeling certain types of robust electronic components. It has been used successfully to describe the life of insulating oils and dielectric fluids (Nelson, 1990, p. 53).
Extended Generalized Gamma (GenGamma)

The extended generalized gamma distribution can include many other distributions as special cases, such as the generalized gamma, Weibull, lognormal, Fréchet, gamma, and exponential. It is particularly useful for cases with little or no censoring. This distribution has been successfully modeled for human cancer prognosis. The pdf and cdf are

\[
f(x; \mu, \sigma, \lambda) = \begin{cases} \frac{\lambda}{x^\lambda} \phi_{lg} \left[ \lambda \omega + \log \left( \frac{\lambda}{\omega} \right) \right] & \text{if } \lambda 
eq 0 \\ \frac{1}{x^\lambda} \phi_{nor}(\omega) & \text{if } \lambda = 0 \end{cases}
\]

\[
F(x; \mu, \sigma, \lambda) = \begin{cases} \Phi_{lg} \left[ \lambda \omega + \log \left( \frac{\lambda}{\omega} \right) \right] & \text{if } \lambda > 0 \\ \Phi_{nor}(\omega) & \text{if } \lambda = 0 \\ 1 - \Phi_{lg} \left[ \lambda \omega + \log \left( \frac{\lambda}{\omega} \right) \right] & \text{if } \lambda < 0 \end{cases}
\]

where \( x > 0, \omega = \left[ \log(x) - \mu \right]/\sigma \), and

\[-\infty < \mu < \infty, \quad -12 < \lambda < 12, \quad \text{and } \sigma > 0.\]

Note that

\[
\phi_{lg}(z; \kappa) = \frac{1}{\Gamma(\kappa)} \exp\left[ \kappa z - \exp(z) \right]
\]

\[
\Phi_{lg}(z; \kappa) = \frac{\Gamma_{1}\left[ \exp(z); \kappa \right]}{\Gamma(\kappa)}
\]

are the pdf and cdf, respectively, for the standardized log-gamma variable and \( \kappa > 0 \) is a shape parameter.

The standardized distributions above are dependent upon the shape parameter \( \kappa \). Meeker and Escobar (chap. 5) give a detailed explanation of the extended generalized gamma distribution.

**Note:** In JMP, the shape parameter, \( \lambda \), for the generalized gamma distribution is bounded between [-12,12] to provide numerical stability.
Threshold Distributions are log-location-scale distributions with threshold parameters. Some of the
distributions above are generalized by adding a threshold parameter, denoted by $\gamma$. The addition of this
threshold parameter shifts the beginning of the distribution away from 0. Threshold parameters are
sometimes called shift, minimum, or guarantee parameters since all units survive the threshold. Note that
while adding a threshold parameter shifts the distribution on the life or time axis, the shape and spread of
the distribution are not affected. Threshold distributions are useful for fitting moderate to heavily shifted
distributions. The general forms for the pdf and cdf of a log-location-scale threshold distribution are

$$f(x; \mu, \sigma, \gamma) = \frac{1}{\sigma(x - \gamma)} \phi\left[\frac{\log(x - \gamma) - \mu}{\sigma}\right], \quad x > \gamma$$

$$F(x; \mu, \sigma, \gamma) = \Phi\left[\frac{\log(x - \gamma) - \mu}{\sigma}\right]$$

where $\phi$ and $\Phi$ are the pdf and cdf, respectively, for the specific distribution. Examples of specific threshold
distributions are shown below for the Weibull, lognormal, Fréchet, and loglogistic distributions, where,
respectively, the sev, nor, lev, and logis pdfs and cdfs are appropriately substituted.

**TH Weibull**

The pdf and cdf of the three-parameter Weibull distribution are

$$f(x; \mu, \sigma, \gamma) = \frac{1}{(x - \gamma)\sigma} \phi_{\text{sev}}\left[\frac{\log(x - \gamma) - \mu}{\sigma}\right], \quad x > \gamma, \sigma > 0$$

$$F(x; \mu, \sigma, \gamma) = \Phi_{\text{sev}}\left[\frac{\log(x - \gamma) - \mu}{\sigma}\right] = 1 - \exp\left[-\left(\frac{x - \gamma}{\alpha}\right)^\beta\right], \quad x > \gamma$$

where $\mu = \log(\alpha)$, and $\sigma = 1/\beta$ and where

$\phi_{\text{sev}}(z) = \exp[z - \exp(z)]$

and

$\Phi_{\text{sev}}(z) = 1 - \exp[-\exp(z)]$

are the pdf and cdf, respectively, for the standardized smallest extreme value, or sev ($\mu = 0, \sigma = 1$)
distribution.

**TH Lognormal**

The pdf and cdf of the three-parameter lognormal distribution are

$$f(x; \mu, \sigma, \gamma) = \frac{1}{\sigma(x - \gamma)} \phi_{\text{nor}}\left[\frac{\log(x - \gamma) - \mu}{\sigma}\right], \quad x > \gamma$$

$$F(x; \mu, \sigma, \gamma) = \Phi_{\text{nor}}\left[\frac{\log(x - \gamma) - \mu}{\sigma}\right]$$
\[ F(x; \mu, \sigma, \gamma) = \Phi_{\text{nor}} \left( \frac{\log(x - \gamma) - \mu}{\sigma} \right) \]

where

\[ \phi_{\text{nor}}(z) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{z^2}{2} \right) \]

and

\[ \Phi_{\text{nor}}(z) = \int_{-\infty}^{z} \phi_{\text{nor}}(w) \, dw \]

are the pdf and cdf, respectively, for the standardized normal, or nor(\(\mu = 0, \sigma = 1\)) distribution.

**TH Fréchet**

The pdf and cdf of the three-parameter Fréchet distribution are

\[ f(x; \mu, \sigma, \gamma) = \frac{1}{\sigma(x - \gamma)} \phi_{\text{lev}} \left( \frac{\log(x - \gamma) - \mu}{\sigma} \right), \quad x > \gamma \]

\[ F(x; \mu, \sigma, \gamma) = \Phi_{\text{lev}} \left( \frac{\log(x - \gamma) - \mu}{\sigma} \right) \]

where

\[ \phi_{\text{lev}}(z) = \exp \left[ -z - \exp(-z) \right] \]

and

\[ \Phi_{\text{lev}}(z) = \exp \left[ -\exp(-z) \right] \]

are the pdf and cdf, respectively, for the standardized largest extreme value, or lev(\(\mu = 0, \sigma = 1\)) distribution.

**TH Loglogistic**

The pdf and cdf of the three-parameter loglogistic distribution are

\[ f(x; \mu, \sigma, \gamma) = \frac{1}{\sigma(x - \gamma)} \phi_{\text{logis}} \left( \frac{\log(x - \gamma) - \mu}{\sigma} \right), \quad x > \gamma \]

\[ F(x; \mu, \sigma, \gamma) = \Phi_{\text{logis}} \left( \frac{\log(x - \gamma) - \mu}{\sigma} \right) \]

where

\[ \phi_{\text{logis}}(z) = \frac{\exp(z)}{(1 + \exp(z))^2} \]
and

\[ \phi_{\text{logis}}(z) = \frac{\exp(z)}{1 + \exp(z)} = \frac{1}{1 + \exp(-z)} \]

are the pdf and cdf, respectively, for the standardized logistic or logis distribution \((\mu = 0, \sigma = 1)\).

**Distributions for Defective Subpopulations**

In reliability experiments, there are times when only a fraction of the population has a particular defect leading to failure. Since all units are not susceptible to failure, using the regular failure distributions is inappropriate and may produce misleading results. Use the DS distribution options to model failures that occur on only a subpopulation. The following DS distributions are available:

- DS Lognormal
- DS Weibull
- DS Loglogistic
- DS Frechet

**Zero-Inflated Distributions**

Zero-inflated distributions are used when some proportion \((p)\) of the data fail at \(t = 0\). When the data contain more zeros than expected by a standard model, the number of zeros is inflated. When the time-to-event data contain zero as the minimum value in the Life Distribution platform, four zero-inflated distributions are available. These distributions include:

- Zero Inflated Lognormal (ZI Lognormal)
- Zero Inflated Weibull (ZI Weibull)
- Zero Inflated Loglogistic (ZI Loglogistic)
- Zero Inflated Fréchet (ZI Fréchet)

The pdf and cdf for zero-inflated distributions are

\[
f(t) = \left[ (1 - p) \frac{1}{\sigma} \right] \phi \left[ \frac{\log(t) - \mu}{\sigma} \right] \\
F(t) = p + (1 - p) \Phi \left[ \frac{\log(t) - \mu}{\sigma} \right]
\]

where

- \(p\) is the proportion of zero data values,
- \(t\) is the time of measurement for the lifetime event,
- \(\mu\) and \(\sigma\) are estimated by calculating the usual maximum likelihood estimations after removing zero values from the original data,
- \(\phi(z)\) and \(\Phi(z)\) are the density and cumulative distribution function, respectively, for a standard distribution. For example, for a Weibull distribution,
\( \phi(z) = \exp(z - \exp(z)) \) and \( \Phi(z) = 1 - \exp(-\exp(z)) \).

See Lawless (2003, p 34) for a more detailed explanation of using zero-inflated distributions. Substitute 
\( p = 1 - \theta \) and \( S_1(t) = 1 - \Phi(t) \) to obtain the form shown above.

See Tobias and Trindade (1995, p 232) for additional information on reliability distributions. This reference gives the general form for mixture distributions. Using the parameterization in Tobias and Trindade, the form above can be found by substituting \( \alpha = p \), \( F_d(t) = 1 \), and \( F_N(t) = \Phi(t) \).
The Fit Life by X Platform helps you analyze lifetime events when only one factor is present. You can access the Fit Life by X platform by selecting it from the Reliability and Survival menu.

You can choose to model the relationship between the event and the factor using various transformations. Available transformation options include: Arrhenius (Celsius, Fahrenheit, and Kelvin), Voltage, Linear, Log, Logit, Location, and Location and Scale.

Using the Fit Life by X platform, you can also create a Custom transformation of your data. (See “Using the Custom Relationship for Fit Life by X,” p. 63.) You can even specify No Effect as an option. You also have the flexibility of comparing different distributions at the same factor level and comparing the same distribution across different factor levels. (See Figure 2.1.)
## Contents

- Introduction to Accelerated Test Models ................................................. 43
- Launching the Fit Life by X Platform Window ........................................... 43
- Platform Options ......................................................................................... 45
- Navigating the Fit Life by X Report Window ............................................. 46
  - Scatterplot ................................................................................................. 46
  - Nonparametric Overlay ............................................................................. 49
  - Wilcoxon Group Homogeneity Test ......................................................... 50
  - Comparisons ............................................................................................. 50
  - Results ....................................................................................................... 54
- Viewing Profilers and Surface Plots ............................................................. 61
- Using the Custom Relationship for Fit Life by X ....................................... 63
Introduction to Accelerated Test Models

The Fit Life by X platform provides the tools needed for accelerated life-testing analysis. Accelerated tests are routinely used in industry to provide failure-time information about products or components in a relatively short-time frame. Common accelerating factors include temperature, voltage, pressure, and usage rate. Results are extrapolated to obtain time-to-failure estimates at lower, normal operating levels of the accelerating factors. These results are used to assess reliability, detect and correct failure modes, compare manufacturers, and certify components.

The Fit Life by X platform includes many commonly used transformations to model physical and chemical relationships between the event and the factor of interest. Examples include transformation using Arrhenius relationship time-acceleration factors and Voltage-acceleration mechanisms. Linear, Log, Logit, Location, Location and Scale, and Custom acceleration models are also included in this platform.

Meeker and Escobar (1998, p. 495) offer a strategy for analyzing accelerated lifetime data:

1. Examine the data graphically. One useful way to visualize the data is by examining a scatterplot of the time-to-failure variable versus the accelerating factor.
2. Fit distributions individually to the data at different levels of the accelerating factor. Repeat for different assumed distributions.
3. Fit an overall model with a plausible relationship between the time-to-failure variable and the accelerating factor.
4. Compare the model in Step 3 with the individual analyses in Step 2, assessing the lack of fit for the overall model.
5. Perform residual and various diagnostic analyses to verify model assumptions.
6. Assess the plausibility of the data to make inferences.

Launching the Fit Life by X Platform Window

This example uses Devalt.jmp, from Meeker and Escobar (1998), and can be found in the Reliability folder of the sample data. It contains time-to-failure data for a device at accelerated operating temperatures. No time-to-failure observation is recorded for the normal operating temperature of 10 degrees Celsius; all other observations are shown as time-to-failure or censored values at accelerated temperature levels of 40, 60, and 80 degrees Celsius.

To launch the platform window:
1. Open the Devalt.jmp sample data table.
2. Select Analyze > Reliability and Survival > Fit Life by X.
3. Select Hours as Y, Time to Event.
4. Select Temp as X.
5. Select Censor as Censor.
6. Select Weight as Freq.
7. Select **Weibull** as the Distribution.
   
   The launch window allows you to specify only one distribution at a time, and includes the **Weibull**, **Lognormal**, **Loglogistic**, and **Fréchet** distributions. (Lognormal is the default setting.)

8. Click the check box to the left of **Nested Model Tests**.
   
   Checking this box appends a nonparametric overlay plot, nested model tests, and a multiple probability plot to the report window.

9. Select **Arrhenius Celsius** as the Relationship. (Arrhenius Celsius is the default setting.)
   
   Figure 2.2 shows the completed launch window with Relationship and Distribution menu options.

---

**Figure 2.2  Fit Life by X Launch Window**

10. Click **OK**. Figure 2.3 shows the Fit Life by X report window.

**Note:** See “Using the Custom Relationship for Fit Life by X,” p. 63, if you are using the Custom relationship for your model.
Platform Options

The following menu options are accessed by clicking the red triangle of the Fit Life by X outline title in the report window:

**Fit Lognormal** fits a lognormal distribution to the data.

**Fit Weibull** fits a Weibull distribution to the data.
Fit Loglogistic fits a loglogistic distribution to the data.
Fit Frechet fits a Fréchet distribution to the data.
Fit All Distributions fits the lognormal, Weibull, loglogistic, and Fréchet distributions, simultaneously, to the data.

Set Time Acceleration Baseline allows you to enter a baseline value for the explanatory variable of the acceleration factor in a popup window.

Change Confidence Level allows you to enter a desired confidence level, for the plots and statistics, in a popup window. The default confidence level is 0.95.

Tabbed Report allows you to specify how you want the report window displayed. Two options are available: Tabbed Overall Report and Tabbed Individual Report. Tabbed Individual Report is checked by default.

Show Surface Plot toggles the surface plot for the distribution on and off in the individual distribution results section of the report. The surface plot is shown in the Distribution, Quantile, Hazard, and Density sections for the individual distributions, and it is on by default.

Show Points toggles the data points on and off in the Nonparametric Overlay plot and in the Multiple Probability Plots. The points are shown in the plots by default. If this option is unchecked, the step functions are shown instead.

Navigating the Fit Life by X Report Window

The initial report window includes the following outline nodes:

- “Scatterplot,” p. 46
- “Nonparametric Overlay,” p. 49
- “Wilcoxon Group Homogeneity Test,” p. 50
- “Comparisons,” p. 50

(Distribution, Quantile, Hazard, Density, and Acceleration Factor profilers, along with criteria values under Comparison Criterion can be viewed and compared.)

- “Results,” p. 54

(Parametric estimates, covariance matrices, and nested model tests can be examined and compared for each of the selected distributions.)

Scatterplot

The Scatterplot of the lifetime event versus the explanatory variable is shown at the top of the report window. For the Devalt data, the Scatterplot shows Hours versus Temp. Table 2.1 indicates how each type of failure is represented on the Scatterplot in the report window.
Navigating the Fit Life by X Report Window

Using Scatterplot Options

Density curves and quantile lines for each group can be specified by clicking on the red-triangle menu for the scatterplot. You can select the **Show Density Curve** option to display the density curves. If the **Location** or the **Location and Scale** model is fit, or if **Nested Model Tests** is checked in the launch window, then the density curves for all of the given explanatory variable levels are shown. You can select the **Add Density Curve** option, where you can specify the density curve you want, one at a time, by entering any value within the range of the accelerating factor.

After the curves have been created, the **Show Density Curve** option toggles the curves on and off the plot. Similarly, you can specify which quantile lines you want by selecting **Add Quantile Line**, where you enter three quantiles of interest, at a time. You can add more quantiles by continually selecting the **Add Quantile Line**. Default quantile values are 0.1, 0.5, and 0.9. Invalid quantile values, like missing values, are ignored. If desired, you can enter just one quantile value, leaving the other entries blank. Figure 2.4 shows the initial scatterplot with density curve and quantile line menu options; Figure 2.5 shows the resulting scatterplot with the **Show Density Curve** and **Add Quantile Line** options selected.

The default view of the scatterplot incorporates the transformation scale. Turn off this option by selecting **Use Transformation Scale**.

### Table 2.1 Scatterplot Representation for Failure and Censored Observations

<table>
<thead>
<tr>
<th>Event</th>
<th>Scatterplot Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>failure</td>
<td>dots</td>
</tr>
<tr>
<td>right-censoring</td>
<td>upward triangles</td>
</tr>
<tr>
<td>left-censoring</td>
<td>downward triangles</td>
</tr>
<tr>
<td>interval-censoring</td>
<td>downward triangle on top of an upward triangle, connected by a solid line</td>
</tr>
</tbody>
</table>
Figure 2.4 Scatterplot of Hours versus Temp with Density Curve and Quantile Line Options

Figure 2.5 Resulting Scatterplot with Density Curve and Quantile Line Options Specified
This plot shows the density curves and the quantile lines for the various Temp levels for the Weibull distribution. You can also view density curves across all the levels of Temp for the Lognormal, Loglogistic, and Fréchet distributions. These distributions can be selected one at a time or can be viewed simultaneously by checking the boxes to the left of the desired distribution name(s).

You can also remove density curves and quantile lines, as desired, by selecting either Remove Density Curve or Remove Quantile Line in the drop-down menu under Scatterplot. (See Figure 2.6.) Density curve values previously entered are shown in the Remove Density Curve window and quantile values previously entered are shown in the Remove Quantile Line window. Curves and lines are removed by checking the appropriate check box.

**Figure 2.6 Scatterplot Remove Menus**

![Scatterplot Remove Menus](image)

**Nonparametric Overlay**

The Nonparametric Overlay plot is the second item shown in the report window (after Scatterplot). Figure 2.7 shows this plot. Differences among groups can readily be detected by examining this plot. You can view these differences for Hours on different scales. You can also change the interval type between Simultaneous and Pointwise.
Wilcoxon Group Homogeneity Test

For this example, the Wilcoxon Group Homogeneity Test, shown in Figure 2.7, indicates that there is a difference among groups. The high ChiSquare value and low p-value are consistent with the differences seen among the Temp groups in the Nonparametric Overlay plot.

Figure 2.7  Nonparametric Overlay Plot and Wilcoxon Test for Devalt.jmp

Comparisons

The Comparisons report section, shown in Figure 2.8, includes six tabs:

- Distribution
- Quantile
- Hazard
- Density
- Acceleration Factor
- Comparison Criterion
Using Profilers

The first five tabs show profilers for the selected distributions. Curves shown in the profilers correspond to both the time-to-event and explanatory variables. Figure 2.8 shows the Distribution Profiler for the Weibull, lognormal, loglogistic, and Fréchet distributions.

Figure 2.8 Distribution Profiler

Comparable results are obtained for the Quantile, Hazard, and Density tabs. The Distribution, Quantile, Hazard, Density, and Acceleration Factor Profilers behave similarly to the Prediction Profiler in other platforms. For example, the vertical lines of Temp and Hours can be dragged to see how each of the distribution values change with temperature and time. For a detailed explanation of the Prediction Profiler, see the Modeling and Multivariate Methods book.

Understanding the Acceleration Factor

Clicking the Acceleration Factor tab displays the Acceleration Factor Profiler for the time-to-event variable for each specified distribution. For this example, Fit All Distributions is selected from the red-triangle menu in the Fit Life by X outline title. The baseline value for the explanatory variable can be modified by selecting
Set Time Acceleration Baseline from the red-triangle menu of the Fit Life by X outline title and entering the desired value. Figure 2.9 shows the Acceleration Factor Profiler for each distribution.

**Figure 2.9** Acceleration Factor Profiler for Devalt.jmp

The Acceleration Factor Profiler allows you to estimate time-to-failure for accelerated test conditions when compared with the baseline condition and a parametric distribution assumption. The interpretation of a time-acceleration plot is generally the ratio of the $p^{th}$ quantile of the baseline condition to the $p^{th}$ quantile of the accelerated test condition. For the distributions shown in this platform, quantile values are not needed, since the scales are the same across all levels.

**Note:** No Acceleration Factor Profiler is produced if the explanatory variable is discrete, if the explanatory variable is treated as discrete, or if a customized formula does not use a unity scale factor.

Using the Quantile Profiler for Extrapolation

Suppose that the data are represented by a Weibull distribution. From viewing the Weibull Acceleration Factor Profiler in Figure 2.9, you see that the acceleration factor at 45 degrees Celsius is 17.42132 for a baseline temperature of 10 degrees Celsius. Click the Quantile tab under the Weibull Results to see the Quantile Profiler for the Weibull distribution. Click and drag the vertical line in the probability plot so that...
Chapter 2

Lifetime Distribution II
Navigating the Fit Life by X Report Window

the probability reads 0.2. From viewing Figure 2.10, where the Probability is set to 0.2, you find that the quantile for the failure probability of 0.2 at 45 degrees Celsius is 6257.179 hours. So, at 10 degrees Celsius, you can expect that 20 percent of the units fail by 6257.179*17.42132 = 109008 hours.

Figure 2.10 Weibull Quantile Profiler for Devalt.jmp

Comparison Criterion

The Comparison Criterion tab shows the -2Loglikelihood, AICc, and BIC criteria for the distributions of interest. Figure 2.11 shows these values for the Weibull, lognormal, loglogistic, and Fréchet distributions. Distributions providing better fits to the data are shown at the top of the Comparison Criterion table.

Figure 2.11 Comparison Criterion Report Tab

This table suggests that the lognormal and loglogistic distributions provide the best fits for the data, since the lowest criteria values are seen for these distributions. For a detailed explanation of the criteria, see Table 1.1 “Comparison Criteria,” p. 13 in the “Lifetime Distribution” chapter.
Results

The Results portion of the report window shows detailed statistics and prediction profilers that are larger than those shown in the Comparisons report section. Separate result sections are shown for each selected distribution.

Statistical results, diagnostic plots, and Distribution, Quantile, Hazard, Density, and Acceleration Factor Profilers are included for each of your specified distributions. The Custom Estimation tab allows you to estimate specific failure probabilities and quantiles, using both Wald and Profile interval methods. When the Box-Cox Relationship is selected on the platform launch dialog, the Sensitivity tab appears. This tab shows how the Loglikelihood and B10 Life change as a function of Box-Cox lambda.

Statistics

For each parametric distribution, there is a Statistics outline node that shows parameter estimates, a covariance matrix, the confidence intervals, summary statistics, and a Cox-Snell Residual P-P plot. You can save probability, quantile, and hazard estimates by selecting any or all of these options from the red-triangle menu of the Statistics title bar for each parametric distribution. The estimates and the corresponding lower and upper confidence limits are saved as columns in your data table. Figure 2.12 shows the save options available for any parametric distribution.

Working with Nested Models

Nested Model Tests are included, if this option was checked in the launch window of the platform. Figure 2.13 shows Weibull Statistical results, Nested Model Tests, and Diagnostic plots for Devalt.jmp. Separate Location and Scale, Separate Location, and Regression analyses results are shown by default for the Nested Model Tests. Regression parameter estimates and the location parameter formula are shown under the Estimates outline title, by default.

The Diagnostics plots for the No Effect model can be displayed by checking the box to the left of No Effect under the Nested Model Tests outline title. The red-triangle menu of the Statistics outline title provides options for you to save probability estimates, quantile estimates, and density estimates to your data table.

If the Nested Model Tests option was not checked in the launch window of the Fit Life by X platform, then the Separate Location and Scale, and Separate Location models are not assessed. In this case, estimates are given for the regression model for each distribution you select, and the Cox-Snell Residual P-P Plot is the only diagnostic plot.
Figure 2.13 Weibull Distribution Nested Model Tests for DevAlt.jmp Data

The Multiple Probability Plots you see in Figure 2.13 are used to validate the distributional assumption for the different levels of the accelerating variable. If the line for each level does not run through the data points for that level, the distributional assumption might not hold. See Meeker and Escobar (1998, sec. 19.2.2) for a discussion of multiple probability plots.

The Cox-Snell Residual P-P Plot is used to validate the distributional assumption for the data. If the data points deviate far from the diagonal, then the distributional assumption might be violated. See Meeker and Escobar (1998, sec. 17.6.1) for a discussion of Cox-Snell residuals.
The Nested Model Tests include statistics and diagnostic plots for the Separate Location and Scale, Separate Location, Regression, and No Effect models. To obtain results for each of the models, independently of the other models, click the underlined model of interest (listed under the Nested Model Tests outline title) and then uncheck the check boxes for the other models. Nested models are described in Table 2.2. Separate Location and Scale, Separate Location, Regression, and No Effect models, using a Weibull distribution for Devalt.jmp, are shown in Figure 2.14, Figure 2.15, Figure 2.16, and Figure 2.17, respectively.

<table>
<thead>
<tr>
<th>Nested Models</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separate Location and Scale</td>
<td>assumes that the location and scale parameters are different for all levels of the explanatory variable and is equivalent to fitting the distribution by the levels of the explanatory variable. The Separate Location and Scale Model has multiple location parameters and multiple scale parameters.</td>
<td>Figure 2.14</td>
</tr>
<tr>
<td>Separate Location</td>
<td>assumes that the location parameters are different, but the scale parameters are the same for all levels of the explanatory variable. The Separate Location Model has multiple location parameters and only one scale parameter.</td>
<td>Figure 2.15</td>
</tr>
<tr>
<td>Regression</td>
<td>is the default model shown in the initial Fit Life by X report window.</td>
<td>Figure 2.16</td>
</tr>
<tr>
<td>No Effect</td>
<td>assumes that the explanatory variable does not affect the response and is equivalent to fitting all of the data values to the selected distribution. The No Effect Model has one location parameter and one scale parameter.</td>
<td>Figure 2.17</td>
</tr>
</tbody>
</table>
**Figure 2.14** Separate Location and Scale Model with the Weibull Distribution for Devalt.jmp Data
Figure 2.15 Separate Location Model with the Weibull Distribution for Devalt.jmp Data
Figure 2.16 Regression Model with the Weibull Distribution for Devalt.jmp Data
Appending Diagnostics Plots

Checking the check box under Diagnostics (to the left of the model name in the report window) appends only the diagnostic plots for that model to the report window. Clicking the underlined model name under the Nested Model Tests outline title in the report window yields a new and separate report window for that
model. Figure 2.18 shows the appended diagnostic plots when the check boxes under Diagnostics are checked for the **Regression**, **Separate Location**, and the **Separate Location and Scale** models.

---

**Figure 2.18** Weibull Distribution Diagnostic Plots for Multiple Models

---

You can see, from Figure 2.18, that side-by-side comparisons of the diagnostic plots provide a visual comparison for the validity of the different models. The red-triangle menu on the Cox-Snell Residual P-P Plot has an option called **Save Residuals**.

---

**Viewing Profilers and Surface Plots**

In addition to a statistical summary and diagnostic plots, the Fit Life by X report window also includes profilers and surface plots for each of your specified distributions. To view the Weibull time-accelerating factor and explanatory variable profilers, click the **Distribution** tab under Weibull Results. To see the surface plot, click the disclosure button to the left of the Weibull outline title (under the profilers). Figure 2.19 shows the Weibull Distribution Profiler and surface plot for *Devalt.jmp*. The profilers and surface plot behave similarly to other platforms. See the *Modeling and Multivariate Methods* book.
The report window also includes a tab labeled **Acceleration Factor**. Clicking the **Acceleration Factor** tab shows the Acceleration Factor Profiler. This profiler is an enlargement of the Weibull plot shown under the **Acceleration Factor** tab in the Comparisons section of the report window. Figure 2.20 shows the Acceleration Factor Profiler for the Weibull distribution of **Devalt.jmp**. The baseline level for the explanatory variable can be modified by selecting the **Set Time Acceleration Baseline** option in the red-triangle menu of the Fit Life by X outline title.
Figure 2.20 Weibull Acceleration Factor Profiler for Devalt.jmp

Using the Custom Relationship for Fit Life by X

If you want to use a custom transformation to model the relationship between the lifetime event and the accelerating factor, use the Custom option. This option is found in the drop-down menu under Relationship in the launch window. Comma delimited rows of your design matrix are entered into the entry fields for the location (μ) and scale (σ) parameters.

For example, to create a quadratic model with Log(Temp) for the Weibull location parameter and a log-linear model with Log(Temp) for the Weibull scale parameter:

1. Open the Devalt.jmp sample data table.
2. Select Analyze > Reliability and Survival > Fit Life by X.
3. Select Hours as Y, Time to Event, Temp as X, Censor as Censor, and Weight as Freq.
4. Select Custom as the Relationship from the drop-down menu.
5. In the entry field for μ, enter 1, log(:Temp), log(:Temp)^2.
   (The 1 indicates that an intercept is included in the model.)
6. In the entry field for σ, enter 1, log(:Temp).
7. Click the check box for Use Exponential Link.
8. Select Weibull as the Distribution.

Figure 2.21 shows the completed launch window using the Custom option.
Note: The Nested Model Tests check box is not checked for non-constant scale models. Nested Model test results are not supported for this option.

9. Click OK.

Figure 2.21 Custom Relationship Specification in Fit Life by X Launch Window

Figure 2.22 shows the location and scale transformations, which are subsequently created and included at the bottom of the Estimates report section. Analysis proceeds similarly to the previous example, where the Arrhenius Celsius Relationship was specified.

Figure 2.22 Weibull Estimates and Formulas for Custom Relationship
Recurrence Analysis analyzes event times like the other Reliability and Survival platforms, but the events can recur several times for each unit. Typically, these events occur when a unit breaks down, is repaired, and then put back into service after the repair. The units are followed until they are ultimately taken out of service. Similarly, recurrence analysis can be used to analyze data from continuing treatments of a long-term disease, such as the recurrence of tumors in patients receiving treatment for bladder cancer. The goal of the analysis is to obtain the MCF, the mean cumulative function, which shows the total cost per unit as a function of time. Cost can be just the number of repairs, or it can be the actual cost of repair.
Contents

Recurrence Data Analysis ......................................................... 67
   Launching the Platform .......................................................... 67
Examples ................................................................. 68
   Valve Seat Repairs Example .................................................. 68
   Bladder Cancer Recurrences Example ................................. 71
Fit Model .......................................................... 73
Recurrence Data Analysis

Recurrent event data involves the cumulative frequency or cost of repairs as units age. In JMP, the Recurrence Analysis platform analyzes recurrent events data.

The data for recurrence analysis have one row for each observed event and a closing row with the last observed age of a unit. Any number of units or systems may be included. In addition, these units or systems may include any number of recurrences.

Launching the Platform

To launch the platform, select Analyze > Reliability and Survival > Recurrence Analysis. The dialog, shown in Figure 3.1 contains the following roles.

- **Y, Age at Event** contains the unit’s age at the time of an event, or at the time of reference for the end of service. This is a required numeric column.
- **Grouping** is an optional column if you want to produce separate MCF estimates for different groups, identified by this column.
- **Event Cause** is used to specify multiple failure modes.
- **Cost, End-of-Service** is a column that must contain one of the following:
  - a 1, indicating an event has occurred (a unit failed or was repaired, replaced, or adjusted).
  - a cost for the event (the cost of the repair, replacement, or adjustment).
  - a zero, indicating that the unit went out-of-service, or is no longer being studied. All units (each System ID) must have one row with a zero for this column, with the Y, Age at Event column containing the final observed age.

If costs are given here, the MCF is a mean cumulative cost per unit as a function of age. If indicators (1’s) are used here, then the MCF is the mean cumulative count of events per unit as a function of age.

- **Label, System ID** is a required column identifying the unit for each event and censoring age.
If each unit does not have exactly one last observed age in the table (where the Cost column cell is zero), then JMP gives an error message to that effect.

**Note:** Cost indicators for Recurrence Analysis are the reverse of censor indicators seen in Life Distribution or Survival Analysis. For the cost variable, the value of 1 indicates an event, such as repair; the value of 0 indicates that the unit is no longer in service. For the censor variable, the value of 1 indicates censored values, and the value of 0 indicates the event or failure of the unit (non-censored value).

**Examples**

**Valve Seat Repairs Example**

A typical unit might be a system, such as a component of an engine or appliance. For example, consider the sample data table Engine Valve Seat.jmp, which records valve seat replacements in locomotive engines. See Meeker and Escobar (1998, p. 395) and Nelson (2003). A partial listing of this data is shown in Figure 3.2. The EngineID column identifies a specific locomotive unit. Age is time in days from beginning of service to replacement of the engine valve seat. Note that an engine can have multiple rows with its age at each replacement and its cost, corresponding to multiple repairs. Here, Cost=0 indicates the last observed age of a locomotive.

![Figure 3.2 Partial Engine Valve Seat Data Table](image)

Complete the launch dialog as shown previously in Figure 3.1.

When you click OK, the Recurrence platform shows the reports in Figure 3.3 and Figure 3.4. The MCF plot shows the sample mean cumulative function. For each age, this is the nonparametric estimate of the mean cumulative cost or number of events per unit. This function goes up as the units get older and total costs grow. The plot in Figure 3.3 shows that about 580 days is the age that averages one repair event.
The event plot in Figure 3.4 shows a time line for each unit. There are markers at each time of repair, and each line extends to that unit’s last observed age. For example, unit 409 was last observed at 389 days and had three valve replacements.
Options

The following options are included in the platform drop-down menu:

- **MCF Plot**  toggles on and off the MCF plot.
- **MCF Confid Limits**  toggles on and off the lines corresponding to the approximate 95% confidence limits of the MCF.
- **Event Plot**  toggles on and off the Event plot.
- **Plot MCF Differences**  If you have a grouping variable, this option will create a plot of the difference of MCF’s, including a 95% confidence interval for that difference. The MCF’s are significantly different where the confidence interval lines do not cross the zero line. This option is available only when you specify a grouping variable.
MCF Plot Each Group produces an MCF plot for each level of the grouping variable. This option is available only when you specify a grouping variable. This option can be used to get an MCF Plot for each unit if the Label, System ID variable is also specified as the Grouping variable.

Fit Model is used to fit models for the Recurrence Intensity and Cumulative functions. See “Fit Model,” p. 73.

### Bladder Cancer Recurrences Example

The sample data file Bladder Cancer.jmp contains data on bladder tumor recurrences from the Veteran’s Administration Co-operative Urological Research Group. See Andrews and Herzberg (1985, table 45). All patients presented with superficial bladder tumors which were removed upon entering the trial. Each patient was then assigned to one of three treatment groups: placebo pills, pyridoxine (vitamin B6) pills, or periodic chemotherapy with thiotepa. The following analysis of tumor recurrence explores the progression of the disease, and whether there is a difference among the three treatments.

Launch the platform with the options shown in Figure 3.5.

**Figure 3.5** Bladder Cancer Launch Dialog

Figure 3.6 shows the MCF plots for the three treatments.
Note that all three of the MCF curves are essentially straight lines. The slopes (rates of recurrence) are therefore constant over time, implying that patients do not seem to get better or worse as the disease progresses.

To examine if there are differences among the treatments, select the **Plot MCF Differences** command from the platform drop-down menu to get the following plots.
To determine whether there is a statistically significant difference between treatments, examine the confidence limits on the differences plot. If the limits do not include zero, the treatments are convincingly different. The graphs in Figure 3.7 show there is no significant difference among the treatments.

**Fit Model**

The Fit Model option is used to fit models for the Recurrence Intensity and Cumulative functions. There are four models available for describing the intensity and cumulative functions. You can fit the models with constant parameters, or with parameters that are functions of effects.
Select Fit Model from the platform red-triangle menu to produce the Recurrence Model Specification window shown in Figure 3.8.

**Figure 3.8 Recurrence Model Specification**

You can select one of four models, with the following Intensity and Cumulative functions:

**Power Nonhomogeneous Poisson Process**

\[ I(t) = \left( \frac{1}{\theta} \right) \left( \frac{t}{\theta} \right)^{\beta - 1} \]

\[ C(t) = \left( \frac{1}{\theta} \right)^\beta \]

**Proportional Intensity Poisson Process**

\[ I(t) = \delta t^{\delta - 1} e^{\gamma t} \]

\[ C(t) = \delta e^{\gamma t} \]

**Loglinear Nonhomogeneous Poisson Process**

\[ I(t) = e^{\gamma t} + \delta t \]

\[ C(t) = \frac{I(t) - I(0)}{\delta} = \frac{e^{\gamma t} + \delta t - e^{\gamma}}{\delta} \]
Homogeneous Poisson Process

\[ I(t) = e^{\theta t} \]

\[ C(t) = t e^{\theta t} \]

where \( t \) is the age of the product.

To fit the models with constant parameters, do not include any Scale Effects or Shape Effects. If you include an effect, the model parameter will be a function of the effect.

Click Run Model to fit the model and see the model report (Figure 3.9).

Figure 3.9 Model Report

The report has the following options on the red-triangle menu:

- **Profiler** launches the Profiler showing the Intensity and Cumulative functions.
- **Effect Marginals** evaluates the parameter functions for each level of the categorical effect, holding other effects at neutral values. This helps you see how different the parameter functions are between groups. This is available only when you specify categorical effects.
- **Test Homogeneity** tests if the shape parameter is equal to 1. This option is not available for the Homogeneous Poisson Process model.
- **Effect Likelihood Ratio Test** produces a test for each effect in the model. This option is available only if there are effects in the model.
- **Specific Intensity and Cumulative** computes the intensity and cumulative values associated with particular time and effect values. The confidence intervals are profile-likelihood.
- **Specific Time for Cumulative** computes the time associated with particular age and number of recurrences.
- **Save Intensity Formula** saves the Intensity formula to the data table.
- **Save Cumulative Formula** saves the Cumulative formula to the data table.
- **Remove Fit** removes the model report.
Using the Degradation platform, you can analyze degradation data to predict pseudo failure times. These pseudo failure times can then be analyzed by other reliability platforms to estimate failure distributions. Both linear and non-linear degradation paths can be modeled. You can also perform stability analysis, which is useful when setting product expiration dates.

**Figure 4.1** Example of Degradation Analysis
Contents

Overview of the Degradation Platform ........................................... 79
Launching the Degradation Platform ........................................... 79
The Degradation Report ............................................................. 80
Model Specification ................................................................. 82
  Simple Linear Path ............................................................... 82
  Nonlinear Path ........................................................................ 84
Inverse Prediction ......................................................................... 91
Prediction Graph ........................................................................ 92
Platform Options ......................................................................... 93
Model Reports ............................................................................ 96
  Model Lists ............................................................................... 96
  Reports ................................................................................... 97
Destructive Degradation .............................................................. 98
Stability Analysis ......................................................................... 101
Overview of the Degradation Platform

In reliability analyses, the primary objective is to model the failure times of the product under study. In many situations, these failures occur because the product degrades (weakens) over time. But, sometimes failures do not occur. In these situations, modeling the product degradation over time is helpful in making predictions about failure times.

The Degradation platform can model data that follows linear or nonlinear degradation paths. If a path is nonlinear, transformations are available to linearize the path. If linearization is not possible, you can specify a nonlinear model.

You can also use the Degradation platform to perform stability analysis. Three types of linear degradation models are fit, and a failure time is estimated. Stability analysis is used in setting product expiration dates.

Launching the Degradation Platform

To launch the Degradation platform, select Analyze > Reliability and Survival > Degradation. Figure 4.2 shows the Degradation launch window using the GaAs Laser.jmp data table (located in the Reliability folder).

Table 4.1 describes features of the Degradation launch window.

Table 4.1 Explanation of Degradation Launch Window

<table>
<thead>
<tr>
<th>Role</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y, Response</td>
<td>Assign the column with degradation measurements.</td>
</tr>
<tr>
<td>Time</td>
<td>Assign the column containing the time values.</td>
</tr>
</tbody>
</table>
To produce the report shown in Figure 4.3, follow the steps below using the `GaAs Laser.jmp` data table.

1. Open the `GaAs Laser.jmp` data table in the Reliability folder of Sample Data.
2. Select Analyze > Reliability and Survival > Degradation.
4. Select Hours and click Time.
5. Select Unit and click Label, System ID.

### Table 4.1 Explanation of Degradation Launch Window (Continued)

<table>
<thead>
<tr>
<th>Role</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Assign a covariate variable.</td>
</tr>
<tr>
<td>Label, System ID</td>
<td>Assign the column that designates the unit IDs.</td>
</tr>
<tr>
<td>Freq</td>
<td>Assign a column giving a frequency for each row.</td>
</tr>
<tr>
<td>Censor</td>
<td>Assign a column that designates if a unit is censored.</td>
</tr>
<tr>
<td>By</td>
<td>Assign a variable to produce an analysis for each level of the variable.</td>
</tr>
<tr>
<td>Application</td>
<td>Select one of the following analysis methods:</td>
</tr>
<tr>
<td></td>
<td><strong>Repeated Measures Degradation</strong> is used to perform linear or nonlinear degradation analysis. This option does not allow for censoring. If your data involves censoring, use the Destructive Degradation option.</td>
</tr>
<tr>
<td></td>
<td><strong>Stability Test</strong> is used to perform a stability analysis for setting product expiration dates. For more information about stability analyses, see &quot;Stability Analysis,&quot; p. 101.</td>
</tr>
<tr>
<td></td>
<td><strong>Destructive Degradation</strong> is used if units are destroyed during the measurement process, or if you have censored data. For more information, see &quot;Destructive Degradation,&quot; p. 98.</td>
</tr>
<tr>
<td>Censor Code</td>
<td>Specify the value in the Censor column that designates censoring.</td>
</tr>
<tr>
<td>Upper Spec Limit</td>
<td>Assign an upper spec limit.</td>
</tr>
<tr>
<td>Lower Spec Limit</td>
<td>Assign a lower spec limit.</td>
</tr>
<tr>
<td>Censoring Time</td>
<td>Assign a censoring value.</td>
</tr>
</tbody>
</table>
6. Click OK.

**Figure 4.3 Initial Degradation Report**

![Graph showing degradation analysis and residuals by hours.](image-url)
The platform automatically fits a default model. The report includes the following items:

- An overlay plot of the Y, Response variable versus the Time variable. In this example, the plot is of Current versus Hours. The Overlay plot red triangle menu has the Save Estimates option, which creates a new data table containing the estimated slopes and intercepts for all units.

- The Model Specification outline. For more details, see “Model Specification,” p. 82.

- The Residual Plot tab. There is a single residual plot with all the units overlaid, and a separate residual plot for each unit. The Save Residuals option on the red triangle menu saves the residuals of the current model to a new data table.

- The Inverse Prediction tab. For more details, see “Inverse Prediction,” p. 91.

- The Prediction Graph tab. For more details, see “Prediction Graph,” p. 92.

**Model Specification**

You can use the Model Specification outline to specify the model that you want to fit to the degradation data. There are two types of Model Specifications:

- **Simple Linear Path** is used to model linear degradation paths, or nonlinear paths that can be transformed to linear. For details, see “Simple Linear Path,” p. 82.

- **Nonlinear Path** is used to model nonlinear degradation paths, especially those that cannot be transformed to linear. For details, see “Nonlinear Path,” p. 84.

To change between the two specifications, use the Degradation Path Style submenu from the platform red triangle menu.

**Simple Linear Path**

To model linear degradation paths, select Degradation Path Style > Simple Linear Path from the platform red triangle menu.

Use the Simple Linear Path Model specification to specify the form of the linear model that you want to fit to the degradation path. You can model linear paths, or nonlinear paths that can be transformed to linear. See Figure 4.4.

**Figure 4.4 Simple Linear Path Model Specification**
Table 4.2 describes the options for the Simple Linear Path specification.

**Table 4.2 Simple Linear Path Options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Use this menu to specify the form of the intercept.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Different</strong> fits a different intercept for each ID.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Common in Group</strong> fits the same intercept for each ID in the same level of the X variable, and different intercepts between levels.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Common</strong> fits the same intercept for all IDs.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Zero</strong> restricts the intercept to be zero for all IDs.</td>
</tr>
<tr>
<td>Slope</td>
<td>Use this menu to specify the form of the slope.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Different</strong> fits a different slope for each ID.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Common in Group</strong> fits the same slope for each ID in the same level of the X variable, and different slopes between levels.</td>
</tr>
<tr>
<td></td>
<td>- <strong>Common</strong> fits the same slope for all IDs.</td>
</tr>
<tr>
<td>&lt;Y, Response&gt; Transformation</td>
<td>If a transformation on the Y variable can linearize the degradation path, select the transformation here. For details about the Custom option, see “Custom Transformations,” p. 83.</td>
</tr>
<tr>
<td>&lt;Time&gt; Transformation</td>
<td>If a transformation for the Time variable can linearize the degradation path, select the transformation here. For details about the Custom option, see “Custom Transformations,” p. 83.</td>
</tr>
<tr>
<td>Reset Axes to Linear</td>
<td>Click this button to return the Overlay plot axes to their initial settings.</td>
</tr>
<tr>
<td>Generate Report for Current Model</td>
<td>Creates a report for the current model settings. This includes a Model Summary report, and Estimates report giving the parameter estimates. For more information, see “Model Reports,” p. 96.</td>
</tr>
</tbody>
</table>

**Custom Transformations**

If you need to perform a transformation that is not given, use the Custom option. For example, to transform the response variable using \( \exp(-x^2) \), enter the transformation as shown in the Scale box in Figure 4.5. Also, enter the inverse transformation in the Inverse Scale box.

**Note:** JMP automatically attempts to solve for the inverse transformation. If it can solve for the inverse, it automatically enters it in the Inverse Scale box. If it cannot solve for the inverse, you must enter it manually.
Name the transformation using the text box. When finished, click the **Use & Save** button to apply the transformation. Select a transformation from the menu if you have created multiple custom transformations. Click the **Delete** button to delete a custom transformation.

**Nonlinear Path**

To model nonlinear degradation paths, select **Degradation Path Style > Nonlinear Path** from the platform red triangle menu. This is useful if a degradation path cannot be linearized using transformations, or if you have a custom nonlinear model that you want to fit to the data.

To facilitate explaining the Nonlinear Path Model Specification, open the **Device B.jmp** data table. The data consists of power decrease measurements taken on 34 units, across four levels of temperature. Follow these steps:

1. Open the **Device B.jmp** data table in the Reliability folder of Sample Data.
2. Select **Analyze > Reliability and Survival > Degradation**.
3. Select **Power Drop** and click **Y, Response**.
4. Select **Hours** and click **Time**.
5. Select **Degrees C** and click **X**.
6. Select **Device** and click **Label, System ID**.
7. Click **OK**.

Figure 4.6 shows the initial overlay plot of the data.
The degradation paths appear linear for the first several hundred hours, but then start to curve. To fit a nonlinear model, select **Degradation Path Style > Nonlinear Path** from the platform red triangle menu to show the Nonlinear Path Model Specification outline. See Figure 4.7.

The first step to create a model is to select one of the options on the menu initially labeled Empty:

- For details about Reaction Rate models, see “Reaction Rate Models,” p. 86.
- For details about Constant Rate models, see “Constant Rate Models,” p. 86.
For details about using a Prediction Column, see “Prediction Columns,” p. 87.

**Reaction Rate Models**

The Reaction Rate option is applicable when the degradation occurs from a single chemical reaction, and the reaction rate is a function of temperature only.

Select **Reaction Rate** from the menu shown in Figure 4.7. The Setup window prompts you to select the temperature scale, and the baseline temperature. The baseline temperature is used to generate initial estimates of parameter values. The baseline temperature should be representative of the temperatures used in the study.

![Figure 4.8 Unit and Baseline Selection](image)

For this example, select Celsius as the Temperature Unit. Click **OK** to return to the report. For details about all the features for Model Specification, see “Model Specification Details,” p. 88.

**Constant Rate Models**

The Constant Rate option is applicable for modeling degradation paths that are linear with respect to time (or linear with respect to time after transforming the response or time), and where the reaction rate is a function of temperature only.

Select **Constant Rate** from the menu shown in Figure 4.7. The Constant Rate Model Settings window prompts you to enter transformations for the response, rate, and time.
Once a selection is made for the Rate Transformation, the Rate Formula appears in the lower left corner as shown in Figure 4.9.

After all selections are made, click OK to return to the report. For details about all the features for Model Specification, see "Model Specification Details," p. 88.

**Prediction Columns**

The Prediction Column option enables you to use a custom model that is stored in a data table column. The easiest approach is to create the model column before launching the Degradation platform. Alternatively, you can create the model column from within the Degradation platform if you want to use one of the built-in models (accomplished by using the Model Library button shown in Figure 4.10).

For details about how to create a model and store it as a column, see the Nonlinear Regression chapter of the *Modeling and Multivariate Methods* book.

Select **Prediction Column** from the menu shown in Figure 4.7. The Model Specification outline changes to prompt you to select the column that contains the model.
At this point, do one of three things:

- If the model that you want to use already exists in a column of the data table, select the column here, and then click **OK**. You are returned to the Nonlinear Path Model Specification. For details about all the features for that specification, see “Model Specification Details,” p. 88.

- If the model that you want to use does not already exist in the data table, and you want to use one of the built-in models, click the **Model Library** button. For details about using the Model Library button, see the Nonlinear Regression chapter of the *Modeling and Multivariate Methods* book. After the model is created, relaunch the Degradation platform and return to the column selection shown in Figure 4.10. Select the column that contains the model, and then click **OK**. You are returned to the Nonlinear Path Model Specification. For details about all the features for that specification, see “Model Specification Details,” p. 88.

- If the model that you want to use does not already exist in the data table, and you do not want to use one of the built-in models, then you are not ready to use this model specification. First create the model (see the Nonlinear Regression chapter of the *Modeling and Multivariate Methods* book), and then relaunch the Degradation platform and return to the column selection shown in Figure 4.10. Select the column containing the model, and then click **OK**. You are returned to the Nonlinear Path Model Specification. For details about all the features for that specification, see “Model Specification Details,” p. 88.

**Model Specification Details**

After you select one of the model types from the menu shown in Figure 4.7 and supply the required information, you are returned to the Nonlinear Path Model Specification. See Figure 4.11 for the Model Specification that you get after clicking **OK** in Figure 4.8.
A model is now shown in the script box that uses the Parameter statement. Initial values for the parameters are estimated from the data. For complete details about creating models that use parameters, see the Nonlinear Regression chapter in the *Modeling and Multivariate Methods* book. A nicely formatted view of the model is shown below the row of buttons.

If desired, type in the text box to name the model. For this example, use the name “Device RR”. After that, click the Use & Save button to enter the model and activate the other buttons and features. See Figure 4.12.
The Fit Model button is used to fit the model to the data.

The Fit by System ID is used to fit the model to every level of Label, System ID.

The Optimization Settings button is used to change the optimization settings.

The Delete button is used to delete a model from the model menu.


The optimization method menu provides three choices for the optimization method (Newton, QuasiNewton BFGS, and QuasiNewton SR1). For details about the methods, see the Nonlinear Regression chapter of the Modeling and Multivariate Methods book.

The initial parameter values are shown at the bottom, along with sliders for visualizing how changes in the parameters affect the model. To do so, first select Graph Options > Show Fitted Lines from the platform red-triangle menu to show the fitted lines on the plot. Then move the parameter sliders to see how changes affect the fitted lines. To compute the optimal values for the parameters, click the Fit Model or Fit by System ID button.

To fix a value for a parameter, check the box under Fixed for the parameter. When fixed, that parameter is held constant in the model fitting process.
Inverse Prediction

Use the Inverse Prediction tab to predict the time when the $Y$ variable will reach a specified value. These times are sometime called pseudo failure times. Figure 4.13 shows the Inverse Prediction tab.

Enter either the Lower or Upper Spec Limit. Generally, if your $Y$ variable decreases over time, then enter an Upper Spec Limit. If the $Y$ variable increases over time, then enter a Lower Spec Limit.

For the GaAs Laser example, enter 10 for the Lower Spec Limit and click Go. A plot is produced showing the estimated times until the units reach a 10% increase in operating current. See Figure 4.14.
The Inverse Prediction red triangle menu has the following options:

- **Save Crossing Time** saves the pseudo failure times to a new data table. The table contains a Life Distribution or Fit Life by X script that can be used to fit a distribution to the pseudo failure times. When one of the Inverse Prediction Interval options is enabled, the table also includes the intervals.

- **Set Upper Spec Limit** is used to set the upper spec limit.

- **Set Lower Spec Limit** is used to set the lower spec limit.

- **Set Censoring Time** is used to set the censoring time.

- **Inverse Prediction Interval** is used to show confidence or prediction intervals for the pseudo failure times on the Inverse Prediction plot. When intervals are enabled, the intervals are also included in the data table that gets created when using the Save Crossing Time option.

- **Inverse Prediction Alpha** is used to specify the alpha level used for the intervals.

- **Inverse Prediction Side** is used to specify one or two sided intervals.

---

**Prediction Graph**

Use the Prediction Graph tab to predict the $Y$ variable for a specified Time value. Figure 4.15 shows the Prediction Plot tab.

---

**Figure 4.15 Prediction Plot Tab**

For the GaAs Laser example, no data was collected after 4000 hours. If you want to predict the percent increase in operating current after 5000 hours, enter 5000 and click Go. A plot is produced showing the estimated percent decrease after 5000 hours for all the units. See Figure 4.16.
The Prediction Plot red triangle menu has the following options:

- **Save Predictions** saves the predicted Y values to a data table. When one of the Longitudinal Prediction Interval options is enabled, the table also includes the intervals.

- **Longitudinal Prediction Interval** is used to show confidence or prediction intervals for the estimated Y on the Prediction Plot. When intervals are enabled, the intervals are also included in the data table that gets created when using the Save Predictions option.

- **Longitudinal Prediction Time** is used to specify the time value for which you want to predict the Y.

- **Longitudinal Prediction Alpha** is used to specify the alpha level used for the intervals.

### Platform Options

The Degradation red triangle menu provides the option that are described in Table 4.3.
### Table 4.3 Degradation Platform Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path Definition</td>
<td>The $Y$ variable at a given time is assumed to have a distribution. You can model the mean, location parameter, or median of that distribution.</td>
</tr>
<tr>
<td></td>
<td><strong>Mean Path</strong> is used to model the mean.</td>
</tr>
<tr>
<td></td>
<td><strong>Location Parameter Path</strong> is used to model the location parameter.</td>
</tr>
<tr>
<td></td>
<td><strong>Median Path</strong> is used to model the median of the distribution.</td>
</tr>
<tr>
<td></td>
<td>When the Location Parameter or Median Path option is selected, a menu appears in the Model Specification. Select the distribution of the response from that menu. See Figure 4.17.</td>
</tr>
<tr>
<td>Degradation Path Style</td>
<td>Provides options for selecting the style of degradation path to fit.</td>
</tr>
<tr>
<td></td>
<td><strong>Simple Linear Path</strong> is used to fit linear degradation paths, and nonlinear paths that can be transformed to linear. For more information, see “Simple Linear Path,” p. 82.</td>
</tr>
<tr>
<td></td>
<td><strong>Nonlinear Path</strong> is used to fit nonlinear degradation paths. For more information, see “Nonlinear Path,” p. 84.</td>
</tr>
</tbody>
</table>
Graph Options

- **Connect Data Markers**: shows or hides lines connecting the points on the Overlay plot.
- **Show Fitted Lines**: shows or hides the fitted lines on the Overlay plot.
- **Show Residual Plot**: shows or hides the residual plot.
- **Show Inverse Prediction Plot**: shows or hides the inverse prediction plot.
- **Show Curve Interval**: shows or hides the confidence intervals on the fitted lines on the Overlay plot.
- **Curve Interval Alpha**: enables you to change the alpha used for the confidence interval curves.
- **Show Median Curves**: shows or hides median lines on the plot when the Path Definition is set to Location Parameter Path.
- **Show Legend**: shows or hides a legend for the markers used on the Overlay plot.
- **No Tab List**: shows or hides the Residual Plot, Inverse Prediction, and Prediction Graph in tabs or in stacked reports.

Prediction Settings

- **Upper Spec Limit**: is used to specify the upper spec limit.
- **Lower Spec Limit**: is used to specify the lower spec limit.
- **Censoring Time**: is used to set the censoring time.
- **Baseline**: is used to specify the normal use conditions for an X variable when modeling nonlinear degradation paths.
- **Inverse Prediction**: is used to specify interval type, alpha level, and one or two-sided intervals for inverse prediction. To do inverse prediction, you must also specify the lower or upper spec limit.
  
  For more information about inverse prediction, see "Inverse Prediction," p. 91.

- **Longitudinal Prediction**: is used to specify the Time value, interval type, and alpha level for longitudinal prediction.
  
  For more information about longitudinal prediction, see “Prediction Graph,” p. 92.
Model Reports

When the Generate Report for Current Model button is clicked, summary reports in two places:

- An entry is added to the Model List report. See “Model Lists,” p. 96 for more details.
- An entry is added to the Reports report. See “Reports,” p. 97 for more details.

Model Lists

The Model List report gives summary statistics and other options for every fitted model. Figure 4.18 shows an example of the Model List with summaries for three models. Table 4.4 gives information about the Model List report.

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applications</td>
<td>Provides options for further analysis of the degradation data.</td>
</tr>
<tr>
<td>Generate Pseudo Failure Data</td>
<td>creates a data table giving the predicted time each unit crosses the specification limit. The table contains a Life Distribution or Fit Life by X script that can be used to fit a distribution to the pseudo failure times.</td>
</tr>
<tr>
<td>Test Stability</td>
<td>performs stability analysis. For more information, see “Stability Analysis,” p. 101.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file.</td>
</tr>
<tr>
<td>Script All By-Groups</td>
<td>Provides options similar to those on the Script menu. This is available if a By variable is specified.</td>
</tr>
</tbody>
</table>

---

**Figure 4.18** Model List

<table>
<thead>
<tr>
<th>Display</th>
<th>Model Type</th>
<th>Report</th>
<th>Nparm</th>
<th>LN(Log10(Beta))</th>
<th>AICc</th>
<th>BIC</th>
<th>SFE</th>
<th>DF</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Simple Linear Path</td>
<td>30</td>
<td>-118.209</td>
<td>-110.395</td>
<td>-10.065</td>
<td>2.39694</td>
<td>225</td>
<td>Intercept Different, Slope Different, Y Linear, X Linear</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Simple Linear Path</td>
<td>16</td>
<td>-115.57</td>
<td>-91.284</td>
<td>-26.908</td>
<td>1.04936</td>
<td>238</td>
<td>Intercept Common, Slope Different, Y Linear, X Linear</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Simple Linear Path</td>
<td>2</td>
<td>758.148</td>
<td>760.195</td>
<td>787.226</td>
<td>209.647</td>
<td>253</td>
<td>Intercept Common, Slope Common, Y Linear, X Linear</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4

Degradation

Model Reports

The Reports report gives details about each model fit. The report includes a Model Summary report, and an Estimate report.

### Table 4.4 Model List Report

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display</td>
<td>Select the model that you want represented in the Overlay plot, Residual Plot, Inverse Prediction plot, and Prediction Graph.</td>
</tr>
<tr>
<td>Model Type</td>
<td>Gives the type of path, either linear or nonlinear.</td>
</tr>
<tr>
<td>Report</td>
<td>Select the check boxes to display the report for a model. For more details about the reports, see “Reports,” p. 97.</td>
</tr>
<tr>
<td>Nparm</td>
<td>Gives the number of parameters estimated for the model.</td>
</tr>
<tr>
<td>-2LogLikelihood</td>
<td>Gives -2xloglikelihood.</td>
</tr>
<tr>
<td>AICc</td>
<td>Gives the corrected Akaike Criterion.</td>
</tr>
<tr>
<td>BIC</td>
<td>Gives the Bayesian Information Criterion.</td>
</tr>
<tr>
<td>SSE</td>
<td>Gives the error sums-of-squares for the model.</td>
</tr>
<tr>
<td>DF</td>
<td>Gives the error degrees-of-freedom.</td>
</tr>
<tr>
<td>Description</td>
<td>Gives a description of the model.</td>
</tr>
</tbody>
</table>

### Table 4.5 Model Summary Report

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;Y, Response&gt; Scale</td>
<td>Gives the transformation on the response variable.</td>
</tr>
<tr>
<td>&lt;Time&gt; Scale</td>
<td>Gives the transformation on the time variable.</td>
</tr>
<tr>
<td>SSE</td>
<td>Gives the error sums-of-squares.</td>
</tr>
<tr>
<td>Nparm</td>
<td>Gives the number of parameters estimated for the model.</td>
</tr>
<tr>
<td>DF</td>
<td>Gives the error degrees-of-freedom.</td>
</tr>
<tr>
<td>RSquare</td>
<td>Gives the r-square.</td>
</tr>
<tr>
<td>MSE</td>
<td>Gives the mean square error.</td>
</tr>
</tbody>
</table>
To measure a product characteristic, sometimes the product must be destroyed. For example, when measuring breaking strength, the product is stressed until it breaks. The regular degradation analysis no longer applies in these situations. To handle these situations, select Destructive Degradation from the Application menu on the platform launch window.

For an example of destructive degradation, open the Adhesive Bond B.jmp data table in the Reliability folder of Sample Data. The data consists of measurements on the strength of an adhesive bond. The product is stressed until the bond breaks, and the required breaking stress is recorded. Because units at normal use conditions are unlikely to break, the units were tested at several levels of an acceleration factor. There is interest in estimating the proportion of units with a strength below 40 Newtons after 260 weeks (5 years) at use conditions of 25° C. Follow the steps below to do the destructive degradation analysis.

1. Open the Adhesive Bond B.jmp data table in the Reliability folder of Sample Data.
2. Select Rows > Clear Row States to clear the excluded rows.
5. Select Weeks and click Time.
6. Select Degrees C and click X.
7. Select Status and click Censor.
8. Type Right in the Censor Code box. This is the value in the censor column that identifies censored data.
9. Select Destructive Degradation from the Application menu.

### Table 4.6 Estimate Report

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Gives the name of the parameter.</td>
</tr>
<tr>
<td>Estimate</td>
<td>Gives the estimate of the parameter.</td>
</tr>
<tr>
<td>Std Error</td>
<td>Gives the standard error of the parameter estimate.</td>
</tr>
<tr>
<td>t Ratio</td>
<td>Gives the t statistic for the parameter, computed as Estimate/Std Error.</td>
</tr>
<tr>
<td>Prob&gt;</td>
<td>t</td>
</tr>
</tbody>
</table>

### Destructive Degradation

To measure a product characteristic, sometimes the product must be destroyed. For example, when measuring breaking strength, the product is stressed until it breaks. The regular degradation analysis no longer applies in these situations. To handle these situations, select Destructive Degradation from the Application menu on the platform launch window.

For an example of destructive degradation, open the Adhesive Bond B.jmp data table in the Reliability folder of Sample Data. The data consists of measurements on the strength of an adhesive bond. The product is stressed until the bond breaks, and the required breaking stress is recorded. Because units at normal use conditions are unlikely to break, the units were tested at several levels of an acceleration factor. There is interest in estimating the proportion of units with a strength below 40 Newtons after 260 weeks (5 years) at use conditions of 25° C. Follow the steps below to do the destructive degradation analysis.

1. Open the Adhesive Bond B.jmp data table in the Reliability folder of Sample Data.
2. Select Rows > Clear Row States to clear the excluded rows.
5. Select Weeks and click Time.
6. Select Degrees C and click X.
7. Select Status and click Censor.
8. Type Right in the Censor Code box. This is the value in the censor column that identifies censored data.
9. Select Destructive Degradation from the Application menu.
Note there is no variable assigned to the Label, System ID role. That role is used in regular degradation analysis when the same unit is measured multiple times. In destructive degradation, each unit is measured once, so each row of the data table corresponds to a different unit, and there is no need for an ID variable.

10. Click OK.

11. Select Lognormal from the distribution menu (under Location Parameter Path Specification).

12. From the platform red triangle menu, select Degradation Path Style > Nonlinear Path.

13. Select Constant Rate from the model type menu.

14. Select the following transformations:
   
   - No Transformation for Path Transformation.
   
   - Arrhenius Celsius for Rate Transformation.
   
   - Sqrt for Time Transformation.

15. Click OK.

16. Click Use & Save.

17. Click Fit Model. The fitted lines for the model are shown in Figure 4.20.
18. Select **Generate Report for Current Model**.
19. At the bottom of the report in the Profiler, enter the following values:
   - 40 for Newtons
   - 260 for Weeks
   - 25 for Degrees C

![Figure 4.21 Distribution Profiler Results](image)

The predicted proportion of units below 40 Newtons after 260 weeks is 0.000151, with a confidence interval of 0.00001 to 0.00153.
Stability Analysis

Stability analysis is used in setting product expiration dates. Three types of linear degradation models are fit, and an expiration date is estimated using the best model. The three models are the following:

- Different slopes and different intercepts for the batches.
- Common slopes and different intercepts for the batches.
- Common slopes and common intercepts for the batches.

The first model is tested against the second, and the second is tested against the third. All tests use a significance level of 0.25.

For example, consider the Stability.jmp data table. The data consists of product concentration measurements on four batches. A concentration of 95 is considered the end of the product's usefulness. Use the data to establish an expiration date for the new product.

To perform the stability analysis, do the following steps:

1. Open the Stability.jmp data table in the Reliability folder of Sample Data.
2. Select Analyze > Reliability and Survival > Degradation.
3. Select Concentration (mg/Kg) and click Y, Response.
4. Select Time and click Time.
5. Select Batch Number and click Label, System ID.
6. Select Stability Test from the Application menu.
7. Click OK.
8. Enter a lower spec limit of 95.
9. Click OK.

A portion of the initial report is shown in Figure 4.22.

---

The test for equal slopes has a p-value of 0.8043. Because this is larger than a significance level of 0.25, the test is not rejected, and you conclude the degradation slopes are equal between batches.
The test for equal intercepts has a p-value of <.0001. Because this is smaller than a significance level of 0.25, the test is rejected, and you conclude the intercepts are different between batches.

Since the test for equal slopes was not rejected, and the test for equal intercepts was rejected, the best model is the one with Different Intercepts and Common Slope. This model is the one selected in the report, and has an estimated expiration date of 23.475.
Survival data contain duration times until the occurrence of a specific event and are sometimes referred to as event-time response data. The event is usually failure, such as the failure of an engine or death of a patient. If the event does not occur before the end of a study for an observation, the observation is said to be censored.

The Reliability and Survival submenu accesses six kinds of survival analysis.

This chapter focuses on univariate survival, the fourth item on the Reliability and Survival Menu. Survival calculates estimates of survival functions using the product-limit (Kaplan-Meier) method for one or more groups of either right-censored or complete data. (Complete data have no censored values.) This platform gives an overlay plot of the estimated survival function for each group and for the whole sample. JMP also computes the log rank and Wilcoxon statistics to test homogeneity between groups. Diagnostic plots and fitting options are available for exponential, Weibull, and lognormal survival distributions. An analysis of competing causes for the Weibull model is also available. Interval censoring is supported by Turnbull estimates.

Life Distribution, Fit Life by X, and Recurrence Analysis are discussed in the “Lifetime Distribution” chapter, the “Lifetime Distribution II” chapter, and the “Recurrence Analysis” chapter. The “Reliability and Survival Analysis II” chapter gives details of both Fit Parametric Survival and Fit Proportional Hazards.
## Contents

Introduction to Survival Analysis ................................................. 105

Univariate Survival Analysis .................................................... 105
  Selecting Variables for Univariate Survival Analysis ......................... 106
  Example: Fan Reliability ................................................... 106
  Example: Rats Data ......................................................... 109
  Overview of the Univariate Survival Platform .................................. 109
  Statistical Reports for the Univariate Analysis ................................ 110
  Platform Options ........................................................... 112
  Fitting Distributions ........................................................ 116
  Interval Censoring ............................................................ 119
  WeiBayes Analysis ................................................................ 121

Estimation of Competing Causes .................................................. 121
  Omitting Causes ............................................................... 123
  Saving Competing Causes Information ........................................ 123
  Simulating Time and Cause Data ............................................... 123
Introduction to Survival Analysis

Survival data need to be analyzed with specialized methods for two reasons:

1. The survival times usually have specialized non-normal distributions, like the exponential, Weibull, and lognormal.
2. Some of the data could be censored—you don’t know the exact survival time but you know that it is greater than the specified value. This is called right-censoring. Right-censoring happens when the study ends without all the units failing, or when a patient has to leave the study before it is finished. The censored observations cannot be ignored without biasing the analysis.

The elements of a survival model are

- a time indicating how long until the unit (or patient) either experienced the event or was censored. Time is the model response ($Y$).
- a censoring indicator that denotes whether an observation experienced the event or was censored. JMP uses the convention that the code for a censored unit is 1 and the code for a non-censored event is zero.
- explanatory variables if a regression model is used.
- If interval censoring is needed, then two $y$ variables hold the lower and upper limits bounding the event time.

Common terms used for reliability and survival data include lifetime, life, survival, failure-time, time-to-event, and duration.

Univariate Survival Analysis

To do a univariate survival analysis, choose Analyze > Reliability and Survival and select Survival from its submenu.

After you complete the launch dialog, shown below, and click OK, the Survival and Reliability command produces product-limit (also called Kaplan-Meier) survival estimates, exploratory plots with optional parameter estimates, and a comparison of survival curves when there is more than one group.

Figure 5.1 Survival Launch Window
Selecting Variables for Univariate Survival Analysis

The Survival platform requires only a time (Y) variable, which must be duration or survival times. The censor, grouping, and frequency variables are optional. The sort-order of the data doesn’t matter.

- **Y, Time to Event** is the only required variable, which contains the time to event or time to censoring. If you have interval censoring, then you specify two Y variables, the lower and upper limits.
- **Grouping** is for a column to classify the data into groups, which are fit separately.
- **Censor** is the column that identifies censored values. The value that identifies censoring should be entered in the Censor Code box. This column can contain more than two distinct values, as long as all censored rows have the value entered in the Censor Code box, and non-censored rows have a value other than what is in the Censor Code box.
- **Freq** is for a column whose values are the frequencies of observations for each row when there are multiple units recorded.
- **By** is used to perform a separate analysis for each level of a classification or grouping variable.

Example: Fan Reliability

The failure of diesel generator fans was studied by Nelson (1982, p. 133) and Meeker and Escobar (1998, appendix C1). A partial listing of the data is shown in Figure 5.2. You may open the data set by clicking on Fan.jmp from the sample data directory in the Reliability/Survival subfolder.

**Figure 5.2 Fan Data**

<table>
<thead>
<tr>
<th></th>
<th>Time</th>
<th>Censor</th>
<th>Exponential</th>
<th>Weibull</th>
<th>Extreme value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>459</td>
<td>0</td>
<td>10.2014442</td>
<td>1985</td>
<td>1.96405344</td>
</tr>
<tr>
<td>2</td>
<td>459</td>
<td>1</td>
<td>0.51502161</td>
<td>459</td>
<td>0.16922454</td>
</tr>
<tr>
<td>3</td>
<td>1159</td>
<td>0</td>
<td>10.3048336</td>
<td>1159</td>
<td>1.28329941</td>
</tr>
<tr>
<td>4</td>
<td>1159</td>
<td>0</td>
<td>10.3048336</td>
<td>1159</td>
<td>1.28329941</td>
</tr>
<tr>
<td>5</td>
<td>1550</td>
<td>1</td>
<td>0.05439460</td>
<td>1560</td>
<td>0.57380193</td>
</tr>
<tr>
<td>6</td>
<td>1900</td>
<td>0</td>
<td>10.3205112</td>
<td>1900</td>
<td>1.19850348</td>
</tr>
<tr>
<td>7</td>
<td>1550</td>
<td>1</td>
<td>0.0578333</td>
<td>1560</td>
<td>0.51087967</td>
</tr>
<tr>
<td>8</td>
<td>1859</td>
<td>1</td>
<td>0.06445244</td>
<td>1859</td>
<td>0.60057967</td>
</tr>
<tr>
<td>9</td>
<td>1859</td>
<td>1</td>
<td>0.06445244</td>
<td>1859</td>
<td>0.60057967</td>
</tr>
<tr>
<td>10</td>
<td>1859</td>
<td>1</td>
<td>0.06445244</td>
<td>1859</td>
<td>0.60057967</td>
</tr>
<tr>
<td>11</td>
<td>1859</td>
<td>1</td>
<td>0.06445244</td>
<td>1859</td>
<td>0.60057967</td>
</tr>
<tr>
<td>12</td>
<td>1859</td>
<td>1</td>
<td>0.06445244</td>
<td>1859</td>
<td>0.60057967</td>
</tr>
<tr>
<td>13</td>
<td>2030</td>
<td>1</td>
<td>0.07072349</td>
<td>2030</td>
<td>0.74679527</td>
</tr>
<tr>
<td>14</td>
<td>2030</td>
<td>1</td>
<td>0.07072349</td>
<td>2030</td>
<td>0.74679527</td>
</tr>
</tbody>
</table>

After launching **Analyze > Reliability and Survival > Survival**, specify **Time as Y, Time to Event** and **Censor as Censor**. Also, check the checkbox for **Plot Failure instead of Survival**, since it is more conventional to show a failure probability plot instead of its reverse (a survival probability plot). The completed dialog is shown in Figure 5.3.
Figure 5.3 Fan Launch Dialog

Figure 5.4 shows the Failure plot. Notice the increasing failure probability as a function of time.

Figure 5.4 Fan Initial Output
Usually, the next step is to explore distributional fits, such as a Weibull model, using the Plot and Fit options for that distribution.

**Figure 5.5** Weibull Output for Fan Data

Since the fit is reasonable and the Beta estimate is near 1, you can conclude that this looks like an exponential distribution, which has a constant hazard rate. The Fitted Distribution Plots option produces three views of each distributional fit. Plots of the Weibull fit are shown in Figure 5.6.

**Figure 5.6** Fitted Distribution Plots
Example: Rats Data

An experiment was undertaken to characterize the survival time of rats exposed to a carcinogen in two treatment groups. The data are in the Rats.jmp table found in the sample data folder. The days variable is the survival time in days. Some observations are censored. The event in this example is death. The objective is to see if rats in one treatment group live longer (more days) than rats in the other treatment group.

Figure 5.7  Launch Window for Rats Data

Use the Survival launch dialog to assign columns to the roles as shown in the example dialog above.

Overview of the Univariate Survival Platform

The Survival platform computes product-limit (Kaplan-Meier) survival estimates for one or more groups. It can be used as a complete analysis or is useful as an exploratory analysis to gain information for more complex model fitting.

The Kaplan-Meier Survival platform

- shows a plot of the estimated survival function for each group and, optionally, for the whole sample.
- calculates and lists survival function estimates for each group and for the combined sample.
- optionally displays exponential, Weibull, and lognormal diagnostic failure plots to graphically check the appropriateness of using these distributions for further regression modeling. Parameter estimates are available on request.
- computes the Log Rank and generalized Wilcoxon Chi-square statistics to test homogeneity of the estimated survival function across groups.
- optionally performs analysis of competing causes, prompting for a cause of failure variable, and estimating a Weibull failure time distribution for censoring patterns corresponding to each cause.

Initially, the Survival platform displays overlay step plots of estimated survival functions for each group as shown in Figure 5.8. A legend identifies groups by color and line type.

Tables beneath the plot give summary statistics and quantiles for survival times, list the estimated survival time for each observation computed within groups and survival times computed from the combined sample. When there is more than one group, statistical tests are given that compare the survival curves.
Statistical Reports for the Univariate Analysis

Initial reports are the Summary table and Quantiles table (shown in Figure 5.9). The Summary table shows the number of failed and number of censored observations for each group (when there are groups) and for the whole study, and the mean and standard deviations adjusted for censoring. For computational details on these statistics, see the *SAS/Stat User’s Guide* (2001).

The quantiles table shows time to failure statistics for individual and combined groups. These include the median survival time, with upper and lower 95% confidence limits. The median survival time is the time (number of days) at which half the subjects have failed. The quartile survival times (25% and 75%) are also included.
When there are multiple groups, the Tests Between Groups table, shown below, gives statistical tests for homogeneity among the groups. Kalbfleisch and Prentice (1980, chap. 1), Hosmer and Lemeshow (1999, chap. 2), and Klein and Moeschberger (1997, chap. 7) discuss statistics and comparisons of survival curves.

**Figure 5.10 Tests Between Groups**

- **Test** names two statistical tests of the hypothesis that the survival functions are the same across groups.
- **Chi-Square** gives the Chi-square approximations for the statistical tests.
  - The Log-Rank test places more weight on larger survival times and is more useful when the ratio of hazard functions in the groups being compared is approximately constant. The hazard function is the instantaneous failure rate at a given time. It is also called the mortality rate or force of mortality.
  - The Wilcoxon test places more weight on early survival times and is the optimum rank test if the error distribution is logistic (Kalbfleisch and Prentice, 1980).
- **DF** gives the degrees of freedom for the statistical tests.
- **Prob>ChiSq** lists the probability of obtaining, by chance alone, a Chi-square value greater than the one computed if the survival functions are the same for all groups.
Figure 5.11 shows an example of the product-limit survival function estimates for one of the groups.

### Figure 5.11 Example of Survival Estimates Table

<table>
<thead>
<tr>
<th>Group 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>days</td>
</tr>
<tr>
<td>0.060</td>
</tr>
<tr>
<td>142.060</td>
</tr>
<tr>
<td>159.060</td>
</tr>
<tr>
<td>163.060</td>
</tr>
<tr>
<td>188.060</td>
</tr>
<tr>
<td>204.060</td>
</tr>
<tr>
<td>205.060</td>
</tr>
<tr>
<td>222.060</td>
</tr>
<tr>
<td>233.060</td>
</tr>
<tr>
<td>239.060</td>
</tr>
<tr>
<td>239.060</td>
</tr>
<tr>
<td>261.060</td>
</tr>
<tr>
<td>260.060</td>
</tr>
<tr>
<td>265.060</td>
</tr>
<tr>
<td>283.060</td>
</tr>
<tr>
<td>244.060</td>
</tr>
</tbody>
</table>

**Note:** When the final time recorded is a censored observation, the report indicates a *biased* mean estimate. The biased mean estimate is a lower bound for the true mean.

### Platform Options

All of the options on the red triangle menu alternately hide or display information. The following list summarizes these options:

- **Survival Plot** displays the overlaid survival plots for each group, as shown in Figure 5.8.
- **Failure Plot** displays the overlaid failure plots (proportion failing over time) for each group in the tradition of the Reliability literature. A Failure Plot reverses the y-axis to show the number of failures rather than the number of survivors. The difference is easily seen in an example. Both plots from the Rats.jmp data table appear in Figure 5.12.

Note that **Failure Plot** replaces the **Reverse Y Axis** command found in older versions of JMP (which is still available in scripts).
Figure 5.12  Survival Plot and Failure Plot of the Rats data

Plot Options is a submenu that contains the following options. Note that the first five options (Show Points, Show Kaplan Meier, Show Combined, Show Confid Interval, Show Simultaneous CI) and the last two options (Fitted Survival CI, Fitted Failure CI) pertain to the initial survival plot, while the other five (Midstep Quantile Points, Connect Quantile Points, Fitted Quantile, Fitted Quantile CI Lines, Fitted Quantile CI Shaded) only pertain to the distributional plots.
Show Points hides or shows the sample points at each step of the survival plot. Failures are shown at the bottom of the steps, and censorings are indicated by points above the steps.

Show Kaplan Meier hides or shows the Kaplan-Meier curves.

Show Combined displays the survival curve for the combined groups in the Survival Plot.

Show Confid Interval shows the pointwise 95% confidence bands on the survival plot for groups and for the combined plot when it is displayed with the Show Combined option.

When the plot has the Show Points and Show Combined options in effect, the survival plot for the total or combined sample shows as a gray line and the points show at the plot steps of each group.

Show Simultaneous CI toggles the simultaneous confidence bands for all groups on the plot.

Meeker and Escobar (1998, chap. 3) discuss pointwise and simultaneous confidence intervals and the motivation for simultaneous confidence intervals in survival analysis.

Midstep Quantile Points changes the plotting positions to use the modified Kaplan-Meier plotting positions, which are equivalent to taking mid-step positions of the Kaplan-Meier curve, rather than the bottom-of-step positions. A consequence of this is that there is an additional point in the plot. This option is recommended, so by default it is turned on.

Connect Quantile Points toggles the lines in the plot on and off. By default, this option is on.

Fitted Quantile toggles the straight-line fit on the fitted Weibull, lognormal, or Exponential Quantile plot.

Fitted Quantile CI Lines toggles the 95% confidence bands for the fitted Weibull, lognormal, or Exponential Quantile plot.

Fitted Quantile CI Shaded toggles the display of the 95% confidence bands for a fit as a shaded area or dashed lines.

Fitted Survival CI toggles the confidence intervals (on the survival plot) of the fitted distribution.

Exponential Plot when checked, plots the cumulative exponential failure probability by time for each group. Lines that are approximately linear empirically indicate the appropriateness of using an exponential model for further analysis. In Figure 5.15, the lines for Group 1 and Group 2 in the Exponential Plot are curved rather than straight, which indicate that the exponential distribution is not appropriate for this data.

Weibull Plot plots the cumulative Weibull failure probability by log(time) for each group. A Weibull plot that has approximately parallel and straight lines indicates a Weibull survival distribution model might be appropriate to use for further analysis.

Weibull Fit produces the linear fit to the Weibull cumulative distribution function in the Weibull plot and two popular forms of Weibull estimates shown in the Extreme Value Parameter Estimates table and the Weibull Parameter Estimates tables (Figure 5.15). The Alpha parameter is the 63.2 percentile of the failure-time distribution. The Extreme-value table shows a different parameterization of the same fit, where Lambda = ln(Alpha) and Delta = 1/Beta.
Lognormal Plot plots the cumulative lognormal failure probability by log(time) for each group. A lognormal plot that has approximately parallel and straight lines indicates a lognormal distribution is appropriate to use for further analysis.

Lognormal Fit produces the linear fit to the lognormal cumulative distribution function in the lognormal plot and the LogNormal Parameter Estimates table shown in Figure 5.15. Mu and Sigma correspond to the mean and standard deviation of a normally distributed natural logarithm of the time variable.

For the exponential, Weibull, and lognormal Fits, if you hold down the Shift key, click on the red triangle of the Product-Limit Survival Fit menu, and then click on the desired fit, the following dialog box appears (LogNormal Fit dialog box is shown).

Figure 5.13 Specify Split Window

You may set the confidence level for the limits, the constrained value for theta (in the case of an exponential fit) sigma (in the case of a lognormal fit) or beta (in the case of a Weibull fit) and request a Confidence Contour Plot for the Weibull and lognormal fits. For details on using constrained values, see “WeiBayes Analysis,” p. 121. An example of a contour plot is shown in Figure 5.14.

Figure 5.14 Confidence Contour Plot

Fitted Distribution Plots is available in conjunction with the fitted distributions to show three plots corresponding to the fitted distributions: Survival, Density, and Hazard. No plot will appear if you haven’t done a Fit command. An example is shown in the next section.
**Competing Causes** prompts for a column in the data table that contains labels for causes of failure. Then, for each cause, the estimation of the Weibull model is performed using that cause to indicate a failure event and other causes to indicate censoring. The fitted distribution is shown by the dashed line in the Survival Plot.

**Estimate Survival Probability** brings up a dialog allowing you to enter up to ten time values. The survival probabilities are estimated for the entered times.

**Estimate Time Quantile** brings up a dialog allowing you to enter up to ten survival probabilities. A time quantile is estimated for each entered probability.

**Save Estimates** creates a data table containing survival and failure estimates, along with confidence intervals, and other distribution statistics.

### Fitting Distributions

For each of the three distributions JMP supports, there is a plot command and a fit command. Use the plot command to see if the event markers seem to follow a straight line—which they will tend to do when the distributional fit is suitable for the data. Then, use the fit commands to estimate the parameters.

---

**Figure 5.15** Exponential, Weibull, and Lognormal Plots and Tables

<table>
<thead>
<tr>
<th>Group</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Failed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>μ</td>
<td>5.4751174</td>
<td>5.2603771</td>
<td>5.6904241</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>σ</td>
<td>0.2255161</td>
<td>0.1762498</td>
<td>0.3371056</td>
<td>19</td>
</tr>
<tr>
<td>1</td>
<td>μ</td>
<td>5.3775422</td>
<td>5.2041823</td>
<td>5.6607605</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>σ</td>
<td>0.1722861</td>
<td>0.1305759</td>
<td>0.2598467</td>
<td>17</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Group</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Failed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>α</td>
<td>5.5125419</td>
<td>5.4701805</td>
<td>5.575066</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>β</td>
<td>0.260114</td>
<td>0.1452447</td>
<td>0.284876</td>
<td>19</td>
</tr>
<tr>
<td>1</td>
<td>b</td>
<td>5.406089</td>
<td>5.3715601</td>
<td>5.442106</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.1940889</td>
<td>0.1203634</td>
<td>0.284876</td>
<td>19</td>
</tr>
</tbody>
</table>
The following table shows what to plot that makes a straight line fit for that distribution:

**Table 5.1** Straight Line Fits for Distribution

<table>
<thead>
<tr>
<th>Distribution Plot</th>
<th>X Axis</th>
<th>Y Axis</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>time</td>
<td>-log(S)</td>
<td>slope is 1/theta</td>
</tr>
<tr>
<td>Weibull</td>
<td>log(time)</td>
<td>log(-log(S))</td>
<td>slope is beta</td>
</tr>
<tr>
<td>Lognormal</td>
<td>log(time)</td>
<td>Probit(1-S)</td>
<td>slope is 1/sigma</td>
</tr>
</tbody>
</table>

**Note:** S = product-limit estimate of the survival distribution

The exponential distribution is the simplest, with only one parameter, which we call theta. It is a constant-hazard distribution, with no memory of how long it has survived to affect how likely an event is. The parameter theta is the expected lifetime.

The Weibull distribution is the most popular for event-time data. There are many ways in which different authors parameterize this distribution (as shown in Table 5.2 “Various Weibull parameters in terms of JMP’s alpha and beta,” p. 118). JMP reports two parameterizations—labeled the lambda-delta extreme value parameterization, and the Weibull alpha-beta parameterization. The alpha-beta parameterization is used in the Reliability literature. See Nelson (1990). Alpha is interpreted as the quantile at which 63.2% of the units fail. Beta is interpreted as follows: if beta>1, the hazard rate increases with time; if beta<1, the hazard rate decreases with time; and if beta=1, the hazard rate is constant—meaning it is the exponential distribution.

The lognormal distribution is also very popular. This is the distribution where if you take the log of the values, the distribution is normal. If you want to fit data to a normal distribution, you can take the exp() of it and analyze it as lognormal, as is done later in the Tobit example.

The option **Fitted Distribution Plots** shows the fitted distributions. Survival, Density, and Hazard plots are shown for the exponential, Weibull, and lognormal distributions. The plots share the same axis scaling so that the distributions can be easily compared.
Figure 5.16 Fitted Distribution Plots for Three Distributions

These plots can be transferred to other graphs through the use of graphic scripts. To copy the graph, context-click on the plot to be copied and select Customize. Highlight the desired script and copy it to the clipboard. On the destination plot, context-click and select Customize. Add a new script and paste the script from the clipboard into the window that results.

Table 5.2 Various Weibull parameters in terms of JMP's alpha and beta

<table>
<thead>
<tr>
<th>JMP Weibull</th>
<th>alpha</th>
<th>beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wayne Nelson</td>
<td>alpha=alpha</td>
<td>beta=beta</td>
</tr>
<tr>
<td>Meeker and Escobar</td>
<td>eta=alpha</td>
<td>beta=beta</td>
</tr>
<tr>
<td>Tobias and Trindade</td>
<td>c=alpha</td>
<td>m=beta</td>
</tr>
<tr>
<td>Kececioglu</td>
<td>eta=alpha</td>
<td>beta=beta</td>
</tr>
<tr>
<td>Hosmer and Lemeshow</td>
<td>exp(X beta)=alpha</td>
<td>lambda=beta</td>
</tr>
<tr>
<td>Blishke and Murthy</td>
<td>beta=alpha</td>
<td>alpha=beta</td>
</tr>
</tbody>
</table>
Interval Censoring

With interval censored data, you only know that the events occurred in some time interval. The Turnbull method is used to obtain non-parametric estimates of the survival function.

In this example from Nelson (1990, p. 147), microprocessor units are tested and inspected at various times and the failed units are counted. Missing values in one of the columns indicate that you don’t know the lower or upper limit, and therefore the event is left or right censored, respectively. The data may be found in the sample data files at Microprocessor Data.jmp, and are shown in Figure 5.17.

<table>
<thead>
<tr>
<th>Kalbfleisch and Prentice</th>
<th>lambda = 1/alpha</th>
<th>p = beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>JMP Extreme Value</td>
<td>lambda = log(alpha)</td>
<td>delta = 1/beta</td>
</tr>
<tr>
<td>Meeker and Escobar s.e.v.</td>
<td>mu = log(alpha)</td>
<td>sigma = 1/beta</td>
</tr>
</tbody>
</table>

Table 5.2 Various Weibull parameters in terms of JMP’s alpha and beta (Continued)

Figure 5.17 Microprocessor Data

When you launch the Survival platform, specify the lower and upper time limits as two Y columns, count as Freq and check Plot Failure instead of Survival, as shown in Figure 5.18.
The resulting Turnbull estimates are shown. Turnbull estimates may have gaps in time where the survival probability is not estimable, as seen here between, for example, 6 and 12, 24 and 48, 48 and 168 and so on.

At this point, select a distribution to see its fitted estimates—in this case, a Lognormal distribution is fit and is shown in Figure 5.19. Notice that the failure plot shows very small failure rates for these data.
WeiBayes Analysis

JMP can constrain the values of the Theta (Exponential), Beta (Weibull), and Sigma (LogNormal) parameters when fitting these distributions. This feature is needed in WeiBayes situations, discussed in Abernethy (1996), such as

- where there are few or no failures,
- there are existing historical values for beta, and
- there is still a need to estimate alpha.

With no failures, the standard technique is to add a failure at the end and the estimates would reflect a kind of lower bound on what the alpha value would be, rather than a real estimate. This feature allows for a true estimation.

To use this feature, hold down the Shift key, click on the red triangle of the Product-Limit Survival Fit menu, and then click on the desired fit. You may then enter a constrained value for the parameter as prompted: theta for the exponential fit; beta for the Weibull fit; and sigma for the lognormal fit.

Estimation of Competing Causes

Sometimes there are multiple causes of failure in a system. For example, suppose that a manufacturing process has several stages and the failure of any stage causes a failure of the whole system. If the different causes are independent, the failure times can be modeled by an estimation of the survival distribution for each cause. A censored estimation is undertaken for a given cause by treating all the event times that are not from that cause as censored observations.

Nelson (1982) discusses the failure times of a small electrical appliance that has a number of causes of failure. One group (Group 2) of the data is in the JMP data table Appliance.jmp sample data, in the Reliability subfolder.

To specify the analysis you only need to enter the time variable (Time Cycles) in the Survival dialog. Then use the Competing Causes menu command, which prompts you to choose a column in the data table to label the causes of failure. For this example choose Cause Code as the label variable.

The survival distribution for the whole system is just the product of the survival probabilities. The Competing Causes table gives the Weibull estimates of Alpha and Beta for each failure cause. It is shown with the hazard plot in Figure 5.21.
Reliability and Survival Analysis

Estimation of Competing Causes

In this example, most of the failures were due to cause 9. Cause 1 occurred only once and couldn't produce good Weibull estimates. Cause 15 happened for very short times and resulted in a small beta and large alpha. Recall that alpha is the estimate of the 63.2% quantile of failure time, which means that causes with early failures often have very large alphas; if these causes do not result in early failures, then these causes do not usually cause later failures.

Figure 5.21 Competing Causes Plot and Table

<table>
<thead>
<tr>
<th>Cause Code</th>
<th>a</th>
<th>\beta</th>
<th>Number Failed</th>
<th>Number censored</th>
<th>LogLikelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.2126784</td>
<td>2.7412046</td>
<td>3</td>
<td>33</td>
<td>-31.193224</td>
</tr>
<tr>
<td>2</td>
<td>20.3137419</td>
<td>1.2865828</td>
<td>2</td>
<td>34</td>
<td>-23.517167</td>
</tr>
<tr>
<td>15</td>
<td>22.0947540</td>
<td>0.7207386</td>
<td>7</td>
<td>26</td>
<td>-73.239040</td>
</tr>
<tr>
<td>9</td>
<td>5,340.2266</td>
<td>1.8276496</td>
<td>17</td>
<td>18</td>
<td>-161.06173</td>
</tr>
<tr>
<td>10</td>
<td>41,503.3281</td>
<td>1.0717523</td>
<td>2</td>
<td>34</td>
<td>-23.618585</td>
</tr>
<tr>
<td>15</td>
<td>26758.67984</td>
<td>3.3491086</td>
<td>2</td>
<td>34</td>
<td>-10.815072</td>
</tr>
</tbody>
</table>

Figure 5.22 shows the Fit Y by X plot of Time Cycles by Cause Code with the quantiles option in effect. This plot gives an idea of how the alphas and betas relate to the failure distribution.

Figure 5.22 Fit Y by X Plot of Time Cycles by Cause Code with Box Plots
Omitting Causes

If cause 9 was corrected, how would that affect the survival due to the remaining causes?

The popup menu icon on the Competing Causes title bar accesses the menu shown here. The Omit Causes command prompts you to select one or more cause values to omit. Survival estimates are then recalculated without the omitted cause(s).

Resulting survival plots can be shown (and toggled on and off) by clicking on the red-triangle platform menu and selecting Survival Plot. Survival plots with all competing causes and without cause 9 are shown in Figure 5.23. Note that the survival rate (as shown by the dashed line) without cause 9 doesn't improve much until 2,000 cycles, but then becomes much better and remains improved—even after 10,000 cycles.

Figure 5.23 Survival Plots with Omitted Causes

Saving Competing Causes Information

The Competing Causes table popup menu has commands to save estimates and to save fail-by-cause coordinates.

The Save Cause Coordinates command adds a new column to the current table called log(–log(Surv)). This information is often used to plot against the time variable, with a grouping variable, such as the code for type of failure.

Simulating Time and Cause Data

The Competing Causes table popup menu contains the Simulate command.

This command asks you to specify a sample size and then creates a new data table containing Time and Cause information from the Weibull distribution as estimated by the data.
Fit Parametric Survival launches a regression platform that fits a survival distribution scaled to a linear model. The distributions to choose from are Weibull, lognormal, exponential, Fréchet, and loglogistic. The regression uses an iterative maximum likelihood method. Accelerated Failure models are one kind of regression model. This chapter also shows how to fit parametric models using the Nonlinear platform.

Fit Proportional Hazards launches a regression platform that uses Cox’s proportional hazards method to fit a linear model. Proportional hazards models are popular regression models for survival data with covariates. See Cox (1972). This model is semiparametric; the linear model is estimated, but the form of the hazard function is not. Time-varying covariates are not supported.

Analyze > Fit Model also accesses the Parametric Regression and Proportional Hazard survival techniques as fitting personalities in the Fit Model dialog.
Contents

Parametric Regression Survival Fitting .......................................................... 127
  Example: Computer Program Execution Time ............................................. 127
Launching the Fit Parametric Survival Platform ............................................. 130
Options and Reports ....................................................................................... 132
  Example: Arrhenius Accelerated Failure Log-Normal Model ....................... 133
  Example: Interval-Censored Accelerated Failure Time Model .................... 138
  Example: Right-Censored Data; Tobit Model ......................................... 140
Proportional Hazards Model ......................................................................... 141
  Example: One Nominal Effect with Two Levels ....................................... 141
  Example: Multiple Effects and Multiple Levels .................................... 144
Nonlinear Parametric Survival Models ......................................................... 148
  Loss Formulas for Survival Distributions ................................................. 148
  Tobit Model Example ............................................................................. 152
  Fitting Simple Survival Distributions .................................................... 154
Parametric Regression Survival Fitting

If you have survival times which can be expressed as a function of one or more variables, you need a regression platform that fits a linear regression model but takes into account the survival distribution and censoring. You can do this kind of analysis with the Fit Parametric Survival command in the Survival submenu, or use the Fit Model fitting personality called Parametric Survival.

Example: Computer Program Execution Time

The data table Comptime.jmp, from Meeker and Escobar (1998, p. 434), is data on the analysis of computer program execution time whose lognormal distribution depends on the regressor Load. It is found in the Reliability subfolder of the sample data.

To begin the analysis, select Analyze > Reliability and Survival > Fit Parametric Survival. When the launch dialog appears, select ExecTime as the Time to Event and add Load as an Effect in the model. Also, change the Distrib from the default Weibull to LogNormal. The completed dialog should appear as in Figure 6.2.
Figure 6.2 Computing Time Dialog

When there is only one regressor, a plot of the survival quantiles for three survival probabilities are shown as a function of the regressor.
Time quantiles, as described on page 438 of Meeker and Escobar, are desired for when 90% of jobs are finished under a system load of 5. Select the **Estimate Time Quantile** command, which brings up a dialog as shown in Figure 6.4. Enter 5 as the **Load**, and 0.1 as the **Survival Prob**.
Figure 6.4 Estimate Time Quantile Dialog

Click Go to produce the quantile estimates and a confidence interval.

Figure 6.5 Estimates of Time Quantile

This estimates that 90% of the jobs will be done by 571 seconds of execution time under a system load of 5.

Launching the Fit Parametric Survival Platform

As an example to illustrate the components of this platform, use the VA Lung Cancer.jmp table from the Sample Data Directory folder. See Kalbfleisch and Prentice (1980). The response, Time, is the survival time in days of a group of lung cancer patients. The regressors are age in years (Age) and time in months from diagnosis to entry into the trial (Diag Time). Censor = 1 indicates censored values.

The Fit Parametric Survival survival command launches the Fit Model dialog shown in Figure 6.6. It is specialized for survival analysis with buttons that label the time (Time to Event) and censor (Censor) variables. Also, the fitting personality shows as Parametric Survival and an additional popup menu lets you choose the type of survival distribution (Weibull, LogNormal, Exponential, Fréchet, and Loglogistic) that you think is appropriate for your data.

The Fit Parametric Survival window contains the following role buttons:
**Time to Event** contains the time to event or time to censoring. With interval censoring, specify two $Y$ variables, where one $Y$ variable gives the lower limit and the other $Y$ variable gives the upper limit for each unit.

**Censor** non-zero values in the censor column indicate censored observations. (The coding of censored values is usually equal to 1.) Uncensored values must be coded as 0.

**Freq** is for a column whose values are the frequencies or counts of observations for each row when there are multiple units recorded.

**Cause** is used to specify the column containing multiple failure causes. This column is particularly useful for estimating competing causes. A separate parametric fit is performed for each cause value. Failure events can be coded with either numeric or categorical (labels) values.

**By** is used to perform a separate analysis for each level of a classification or grouping variable.

To launch the Fit Parametric Survival platform:

1. Open the sample data, VA Lung Cancer.jmp.
2. Select **Analyze > Reliability and Survival > Fit Parametric Survival**.
3. Select Time as **Time to Event**.
4. Select Age and Diag Time as model effects.
5. Select censor as **Censor**.
6. Click **Run Model**.
Options and Reports

The following parametric survival options (see Figure 6.7) are available from the red-triangle menu of the report:

- **Likelihood Ratio Tests**: produces tests that compare the log likelihood from the fitted model to one that removes each term from the model individually. The likelihood-ratio test is appended to the text reports.

- **Confidence Intervals**: calculates a profile-likelihood 95% confidence interval on each parameter and lists them in the Parameter Estimates table.

- **Correlation of Estimates**: produces a correlation matrix for the model effects with each other and with the parameter of the fitting distribution.

- **Covariance of Estimates**: produces a covariance matrix for the model effects with each other and with the parameter of the fitting distribution.

- **Estimate Survival Probability**: brings up a dialog where you specify regressor values and one or more time values. JMP then calculates the survival and failure probabilities with 95% confidence limits for all possible combinations of the entries.

- **Estimate Time Quantile**: brings up a dialog where you specify regressor values and one or more survival values. It then calculates the time quantiles and 95% confidence limits for all possible combinations of the entries.

- **Residual Quantile Plot**: shows a plot with the residuals on the x-axis and the Kaplan-Meier estimated quantiles on the y-axis. In cases of interval censoring, the midpoint is used. The residuals are the simplest form of Cox-Snell residuals, which convert event times to a censored standard Weibull or other standard distribution.

- **Save Residuals**: creates a new column to hold the residuals.

- **Distribution Profiler**: displays the response surfaces of the failure probability versus individual explanatory and response variables. The vertical line on each plot can be dragged to change the value on the x-axis. Corresponding y-axis values show the estimated failure probability values based upon the selected distribution. As with profilers in other platforms, one or more factors may be locked while varying the vertical line on all remaining factors. To lock a factor, move the vertical line to the desired value and Alt-click inside the plot. Then, check the box beside Lock Factor Setting in the popup dialog box.

- **Quantile Profiler**: displays the response surfaces of the response variable versus the explanatory and the failure probability. The vertical line on each plot can be dragged to change the value on the x-axis. Corresponding y-axis values show the estimated response variable values based upon the selected distribution. This response option allows you to assess the response at varying quantiles.
Save Probability Formula saves the estimated probability formula to a new column in the data table.

Save Quantile Formula saves the estimated quantile formula to a new column in the data table. Selecting this option displays a popup dialog, asking you to enter a probability value for the quantile of interest.

Figure 6.7 Parametric Survival Model Reports

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.9153406</td>
<td>0.67173</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0124267</td>
<td>0.0114403</td>
</tr>
<tr>
<td>Day</td>
<td>-0.0199189</td>
<td>0.0103906</td>
</tr>
</tbody>
</table>


Example: Arrhenius Accelerated Failure Log-Normal Model

The Devalt.jmp data set (also known as the Device A data) is described by Meeker and Escobar (1998, p. 493) as originating from Hooper and Amster (1990). The data set is found in the Reliability sample data folder. A partial listing of the data set is in Figure 6.8.

See the section “Nonlinear Parametric Survival Models,” p. 148, for further details on the statistical model.
Units are stressed by heating in order to make them fail soon enough to obtain enough failures to fit the distribution.

Use the Bivariate platform to see a plot of hours by temperature using the log scale for time. See Meeker and Escobar (1998, fig. 19.1). This can be accomplished manually (Analyze > Fit Y by X with Hours as Y, Response and Temp as X, Factor) or by running the Bivariate script on the left side of the data table.
Use the univariate survival platform to produce a LogNormal plot of the data for each temperature. To do so, select **Analyze > Reliability and Survival > Survival** with Hours as Y, Time to Event, Censor as Censor, Temp as Grouping, and Weight as Freq. From the resulting report, click on the red-triangle menu of the Product-Limit Survival Fit title bar and select **LogNormal Plot** and **LogNormal Fit**. Alternatively, run the Survival script attached to the data table. Either method produces the plot shown in Figure 6.10.

**Figure 6.9** Bivariate Plot of Hours by log Temp

**Figure 6.10** Lognormal Plot
Then, fit one model using a regressor for temperature. The regressor $x$ is the Arrhenius transformation of temperature calculated by the formula stored in the $x$ column of the data table:

$$\frac{11805}{(\text{Temp} + 273.15)}$$

$\text{Hours}$ is the failure or censoring time. The lognormal distribution is fit to the distribution of this time. To do so, select Analyze > Reliability and Survival > Fit Parametric Survival, assign $\text{Hours}$ as Time to Event, $x$ as a model effect, Censor as Censor and Weight as Freq. Also, change the Distribution type to LogNormal.

After clicking Run, the result shows the regression fit of the data. If there is only one regressor and it is continuous, then a plot of the survival as a function of the regressor is shown, with lines at 0.1, 0.5, and 0.9 survival probabilities. If the regressor column has a formula in terms of one other column, as in this case, the plot is done with respect to the inner column. In this case the regressor was the column $x$, but the plot is done with respect to $\text{Temp}$, of which $x$ is a function.
Finally, we illustrate how to get estimates of survival probabilities extrapolated to a temperature of 10 degrees celsius for the times 10000 and 30000 hours. Select the Estimate Survival Probability command, and enter the following values into the dialog. The Arrhenius transformation of 10 degrees is 40.9853, the regressor value.
Figure 6.12 Estimating Survival Probabilities

After clicking Go, the report shows the estimates and a confidence interval. (See Figure 6.13.)

Figure 6.13 Survival Probabilities

Example: Interval-Censored Accelerated Failure Time Model

Continuing with another example from Meeker and Escobar (1998, p. 508), ICdevice02.jmp shows data in which failures were found to have happened between inspection intervals. The data, found in the Reliability sample data folder, is illustrated in Figure 6.14.
The model uses two $y$-variables, containing the upper and lower bounds on the failure times. Right-censored times are shown with missing upper bounds. To perform the analysis, select Analyze > Reliability and Survival > Fit Parametric Survival with both HoursL and HoursU as Time to Event, Count as Freq, and x as an effect in the model. The resulting regression has a plot of time by degrees.

<table>
<thead>
<tr>
<th></th>
<th>HoursL</th>
<th>HoursU</th>
<th>Status</th>
<th>Count</th>
<th>DegreesC</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1536</td>
<td>*</td>
<td>Right</td>
<td>50</td>
<td>150</td>
<td>27.4252626</td>
</tr>
<tr>
<td>2</td>
<td>1536</td>
<td>*</td>
<td>Right</td>
<td>50</td>
<td>175</td>
<td>25.8653475</td>
</tr>
<tr>
<td>3</td>
<td>96</td>
<td>*</td>
<td>Right</td>
<td>50</td>
<td>200</td>
<td>24.5271066</td>
</tr>
<tr>
<td>4</td>
<td>384</td>
<td>788</td>
<td>Interval</td>
<td>1</td>
<td>250</td>
<td>22.1829303</td>
</tr>
<tr>
<td>5</td>
<td>788</td>
<td>1536</td>
<td>Interval</td>
<td>3</td>
<td>250</td>
<td>22.1829303</td>
</tr>
<tr>
<td>6</td>
<td>1536</td>
<td>2504</td>
<td>Interval</td>
<td>5</td>
<td>250</td>
<td>22.1829303</td>
</tr>
<tr>
<td>7</td>
<td>2504</td>
<td>*</td>
<td>Right</td>
<td>41</td>
<td>250</td>
<td>22.1829303</td>
</tr>
<tr>
<td>8</td>
<td>192</td>
<td>384</td>
<td>Interval</td>
<td>4</td>
<td>300</td>
<td>20.2477536</td>
</tr>
<tr>
<td>9</td>
<td>384</td>
<td>788</td>
<td>Interval</td>
<td>27</td>
<td>300</td>
<td>20.2477536</td>
</tr>
<tr>
<td>10</td>
<td>788</td>
<td>1536</td>
<td>Interval</td>
<td>16</td>
<td>300</td>
<td>20.2477536</td>
</tr>
<tr>
<td>11</td>
<td>1536</td>
<td>*</td>
<td>Right</td>
<td>3</td>
<td>300</td>
<td>20.2477536</td>
</tr>
</tbody>
</table>
Example: Right-Censored Data; Tobit Model

The Tobit model is normal, truncated at zero. However, you can take the exponential of the response and set up the intervals for a right-censored problem. In Tobit2.jmp, found in the Reliability folder of the sample data, run the attached script to estimate the lognormal.
Proportional Hazards Model

The proportional hazards model is a special semiparametric regression model proposed by D. R. Cox (1972) to examine the effect of explanatory variables on survival times. The survival time of each member of a population is assumed to follow its own hazard function.

Proportional hazards is nonparametric in the sense that it involves an unspecified arbitrary baseline hazard function. It is parametric because it assumes a parametric form for the covariates. The baseline hazard function is scaled by a function of the model’s (time-independent) covariates to give a general hazard function. Unlike the Kaplan-Meier analysis, proportional hazards computes parameter estimates and standard errors for each covariate. The regression parameters (β) associated with the explanatory variables and their standard errors are estimated using the maximum likelihood method. A conditional risk ratio (or hazard ratio) is also computed from the parameter estimates.

The survival estimates in proportional hazards are generated using an empirical method. See Lawless (1982). They represent the empirical cumulative hazard function estimates, \( H(t) \), of the survivor function, \( S(t) \), and can be written as \( S_0 = e^{x \beta} \), with the hazard function

\[
H(t) = \sum_{j: t_j < t} \frac{d_j}{\sum_{l \in R_j} e^{x_l \beta}}
\]

When there are ties in the response, meaning there is more than one failure at a given time event, the Breslow likelihood is used.

**Example: One Nominal Effect with Two Levels**

The following example uses the Rats.jmp sample data. To define a proportional hazard survival model, choose Fit Proportional Hazards from the Reliability and Survival submenu.

This launches the Fit Model dialog for survival analysis, with Proportional Hazard showing as the Fitting Personality. Alternatively, you can use the Fit Model command and specify the Proportional Hazard fitting personality (Figure 6.16).

The Fit Proportional Hazards window contains the following role buttons:

- **Time to Event** contains the time to event or time to censoring. With interval censoring, specify two Y variables, where one Y variable gives the lower limit and the other Y variable gives the upper limit for each unit.

- **Censor** non-zero values in the censor column indicate censored observations. (The coding of censored values is usually equal to 1.) Uncensored values must be coded as 0.

- **Freq** is for a column whose values are the frequencies or counts of observations for each row when there are multiple units recorded.

- **By** is used to perform a separate analysis for each level of a classification or grouping variable.

Using the Rats.jmp sample data, assign days as the Time to Event variable, Censor as Censor and Add Group to the model effects list. The next section describes the proportional hazard analysis results.
Reliability and Survival Analysis II
Proportional Hazards Model

Figure 6.16  Fit Model Dialog for Proportional Hazard Analysis

Statistical Reports for the Proportional Hazard Model

Finding parameter estimates for a proportional hazards model is an iterative procedure. When the fitting is complete, the report in Figure 6.17 appears. The Iteration History table lists iteration results occurring during the model calculations.

- The Whole Model table shows the negative of the natural log of the likelihood function (–LogLikelihood) for the model with and without the grouping covariate. Twice the positive difference between them gives a chi-square test of the hypothesis that there is no difference in survival time between the groups. The degrees of freedom (DF) are equal to the change in the number of parameters between the full and reduced models.

- The Parameter Estimates table gives the parameter estimate for Group, its standard error, and 95% upper and lower confidence limits. For the Rats.jmp sample data, there are only two levels in Group, so a confidence interval that does not include zero is evidence that there is an alpha-level significant difference between groups.

- The Effect Likelihood-Ratio Tests shows the likelihood-ratio chi-square test on the null hypothesis that the parameter estimate for the Group covariate is zero. Because Group only has two values, the test of the null hypothesis for no difference between the groups shown in the Whole Model Test table is the same as the null hypothesis that the regression coefficient for Group is zero.
Risk Ratios for One Nominal Effect with Two Levels

The Risk Ratios option is available from the red-triangle menu for Proportional Hazards Fit and shows the risk ratios for the effects. For this example, there is only one effect and there are only two levels for that effect. The risk ratio for Group 2 is compared with Group 1 and is shown in the Risk Ratios for Group table of the report window. (See Figure 6.18.) The risk ratio in this table is determined by computing the exponential of the parameter estimate for Group 2 and dividing it by the exponential of the parameter estimate for Group 1. The Group 1 parameter estimate is seen in the Parameter Estimates table. (See Figure 6.17.) The Group 2 parameter estimate is calculated by taking the negative value for the parameter estimate of Group 1. Reciprocal shows the value for 1/Risk Ratio.
For this example, the risk ratio for Group2/Group1 is calculated as
\[
\frac{\exp[-(-0.2979479)]}{\exp(-0.2979479)} = 1.8146558
\]
This risk ratio value suggests that the risk of death for Group 2 is 1.82 times higher than that for Group 1.

**Example: Multiple Effects and Multiple Levels**

This example uses a proportional hazards model for the sample data, VA Lung Cancer.jmp. The data were collected from a randomized clinical trial, where males with inoperable lung cancer were placed on either a standard or a novel (test) chemotherapy (Treatment). The primary interest of this trial was to assess if the treatment type has an effect on survival time, with special interest given to the type of tumor (Cell Type).

See Prentice (1973) and Kalbfleisch and Prentice (2002) for additional details regarding this data set. For the proportional hazards model, covariates include whether or not the patient had undergone previous therapy (Prior), the age of the patient (Age), the time from lung cancer diagnosis to beginning the study (Diag Time), and a general medical status measure (KPS). Age, Diag Time, and KPS are continuous measures and Cell Type, Treatment, and Prior are categorical (nominal) variables. The four nominal levels of Cell Type include Adeno, Large, Small, and Squamous.

This section illustrates the report window for a model with more than one effect and a nominal effect with more than two levels. This section also includes example calculations for risk ratios for a continuous effect and risk ratios for an effect that has more than two levels.

1. Open the sample data, VA Lung Cancer.jmp.
2. Select **Analyze > Reliability and Survival > Fit Proportional Hazards**.
3. Select Time as **Time to Event**.
4. Select censor as **Censor**.
5. Select Cell Type, Treatment, Prior, Age, Diag Time, and KPS as the model effects.

    Figure 6.19 shows the completed launch window.
6. Click Run.
7. Select Risk Ratios from the red-triangle menu of the Proportional Hazards Fit title bar in the report window.
8. Click the disclosure button on the Baseline Survival at mean title bar to close the plot, and click the disclosure button on Whole Model to close the report.

Figure 6.20 shows the resulting report window for this model.
Figure 6.20  Report Window for Proportional Hazards Model with Multiple Effects and Levels

<table>
<thead>
<tr>
<th>Censored/Iteration History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole Model</td>
</tr>
<tr>
<td>Parameter Estimates</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>StdError</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Type Aden</td>
<td>0.5719698</td>
<td>0.1844991</td>
<td>0.2065801</td>
<td>0.9361862</td>
</tr>
<tr>
<td>Cell Type Large</td>
<td>-0.2144755</td>
<td>0.1741977</td>
<td>-0.565182</td>
<td>0.136118</td>
</tr>
<tr>
<td>Cell Type Small</td>
<td>0.2496222</td>
<td>0.1629182</td>
<td>-0.0705952</td>
<td>0.563322</td>
</tr>
<tr>
<td>Treatment(Standard)</td>
<td>-0.1440979</td>
<td>0.1036051</td>
<td>-0.349894</td>
<td>0.057818</td>
</tr>
<tr>
<td>Pstat[No]</td>
<td>-0.0381833</td>
<td>0.1180936</td>
<td>-0.293533</td>
<td>0.197949</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0091498</td>
<td>0.0903342</td>
<td>-0.24647</td>
<td>0.919846</td>
</tr>
<tr>
<td>Diag Time</td>
<td>-0.0004976</td>
<td>0.0012515</td>
<td>-0.00222</td>
<td>0.001614</td>
</tr>
<tr>
<td>KPS</td>
<td>-0.0326171</td>
<td>0.0655502</td>
<td>-0.16442</td>
<td>-0.021034</td>
</tr>
</tbody>
</table>

| Covariance of Estimates |

Effect Likelihood Ratio Tests

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>ChiSquare</th>
<th>Prob&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Type</td>
<td>3</td>
<td>18.0963284</td>
<td>0.0012**</td>
</tr>
<tr>
<td>Treatment</td>
<td>1</td>
<td>1.8516187</td>
<td>0.1612</td>
</tr>
<tr>
<td>Prior</td>
<td>1</td>
<td>0.09631585</td>
<td>0.7562</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>0.03837359</td>
<td>0.6329</td>
</tr>
<tr>
<td>Diag Time</td>
<td>1</td>
<td>0.00051077</td>
<td>0.9320</td>
</tr>
<tr>
<td>KPS</td>
<td>1</td>
<td>0.48314103</td>
<td>&lt;0.001**</td>
</tr>
</tbody>
</table>

Baseline Survival at mean

Risk Ratios

<table>
<thead>
<tr>
<th>Unit Risk Ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>For unit change in regressor</td>
</tr>
<tr>
<td>Term</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Diag Time</td>
</tr>
<tr>
<td>KPS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Range Risk Ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>For change in regressor over entire range</td>
</tr>
<tr>
<td>Term</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Diag Time</td>
</tr>
<tr>
<td>KPS</td>
</tr>
</tbody>
</table>

Risk Ratios for Cell Type

<table>
<thead>
<tr>
<th>Level</th>
<th>Level1</th>
<th>Level2</th>
<th>Risk Ratio</th>
<th>Prob&gt;ChiSq</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large Aden</td>
<td>0.4544481</td>
<td>0.0036</td>
<td>0.2502748</td>
<td>0.823419</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Aden</td>
<td>0.7176217</td>
<td>0.2325</td>
<td>0.4199109</td>
<td>1.2410707</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Large</td>
<td>0.5796516</td>
<td>0.0825</td>
<td>0.4530393</td>
<td>2.6884796</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous Aden</td>
<td>0.3047391</td>
<td>&lt;0.001*</td>
<td>0.1693479</td>
<td>0.5501962</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous Large</td>
<td>0.8019836</td>
<td>0.1591</td>
<td>0.3841598</td>
<td>1.714599</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous Small</td>
<td>0.5428814</td>
<td>0.0014</td>
<td>0.2544471</td>
<td>0.721158</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adeno Large</td>
<td>2.0304712</td>
<td>0.0036</td>
<td>1.2133433</td>
<td>3.9890897</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adeno Small</td>
<td>1.3694918</td>
<td>0.2356</td>
<td>0.6692232</td>
<td>1.2119755</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large Small</td>
<td>0.8353297</td>
<td>0.0036</td>
<td>0.3715062</td>
<td>1.609085</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adeno Squamous</td>
<td>3.2014575</td>
<td>&lt;0.001*</td>
<td>1.8175235</td>
<td>5.4805784</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large Squamous</td>
<td>1.4262955</td>
<td>0.1691</td>
<td>0.9053577</td>
<td>2.662247</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Squamous</td>
<td>2.509727</td>
<td>0.0014</td>
<td>1.3902057</td>
<td>4.609094</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Risk Ratios for Treatment

<table>
<thead>
<tr>
<th>Level</th>
<th>Level1</th>
<th>Level2</th>
<th>Risk Ratio</th>
<th>Prob&gt;ChiSq</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>1.355418</td>
<td>0.1612</td>
<td>0.000171</td>
<td>2.306664</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>0.740815</td>
<td>0.1591</td>
<td>0.497596</td>
<td>1.122695</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Risk Ratios for Prior

<table>
<thead>
<tr>
<th>Level</th>
<th>Level1</th>
<th>Level2</th>
<th>Risk Ratio</th>
<th>Prob&gt;ChiSq</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>0.8303275</td>
<td>0.2562</td>
<td>0.5743091</td>
<td>1.670815</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.8303275</td>
<td>0.2562</td>
<td>0.5743091</td>
<td>1.670815</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Whole Model, Parameter Estimates, and Effect Likelihood Ratio Tests Tables

- The Whole Model table shows the negative of the natural log of the likelihood function (–LogLikelihood) for the model with and without the covariates. Twice the positive difference between them gives a chi-square test of the hypothesis that there is no difference in survival time among the effects. The degrees of freedom (DF) are equal to the change in the number of parameters between the full and reduced models. The low Prob>Chisq value (<.0001) indicates that there is a difference in survival time when at least one of the effects is included in the model.

- The Parameter Estimates table gives the parameter estimates for Cell Type, Treatment, Prior, Age, Diag Time, and KPS, the standard error for each estimate, and 95% upper and lower confidence limits for each estimate. A confidence interval for a continuous column that does not include zero indicates that the effect is significant. A confidence interval for a level in a categorical column that does not include zero indicates that the difference between the level and the average of all levels is not zero.

- The Effect Likelihood Ratio Tests table shows the likelihood ratio chi-square test on the null hypothesis that the parameter estimates for the effects of the covariates are zero. The Prob>ChiSq values indicate that KPS and at least one of the levels of Cell Type are significant, while Treatment, Prior, Age, and Diag Time effects are not significant.

Calculating Risk Ratios for Multiple Effects and Multiple Levels

The Risk Ratios option is available from the red-triangle menu of the Proportional Hazards Fit title bar of the report and shows the risk ratios for the effects. Figure 6.20 shows the Risk Ratios for the continuous effects (Age, Diag Time, KPS) and the nominal effects (Cell Type, Treatment, Prior). Of particular interest in this section, for illustration, is the continuous effect, Age, and the nominal effect with four levels (Cell Type) for the VA Lung Cancer.jmp sample data.

For continuous columns, unit risk ratios and range risk ratios are calculated. The Unit Risk Ratio is \( \exp(\beta) = \exp(-0.0085494) = 0.991487 \) and the Range Risk Ratio is \( \exp[\beta(x_{\text{Max}} - x_{\text{Min}})] = \exp(-0.0085494 \times 47) = 0.669099 \).

Risk Ratios for Cell Type

For categorical columns, risk ratios are shown in separate tables for each effect. For the nominal effect, Cell Type, all pairs of levels are calculated and are shown in the Risk Ratios for Cell Type table. Note that for a categorical variable with \( k \) levels, only \( k - 1 \) design variables, or levels, are used. In the Parameter Estimates table, parameter estimates are shown for only three of the four levels for Cell Type (Adeno, Large, and Small). The Squamous level is not shown, but it is calculated as the negative sum of the other estimates. Two example Risk Ratios for Cell Type calculations follow.
Large/Adeno = \exp(\beta_{\text{Large}}) / \exp(\beta_{\text{Adeno}}) = \exp(-0.2114757) / \exp(0.57719588) = 0.4544481

Squamous/Adeno = \exp[-(\beta_{\text{Adeno}} + \beta_{\text{Large}} + \beta_{\text{Small}})] / \exp(\beta_{\text{Adeno}})
= \exp[-(0.57719588 + (-0.2114757) + 0.24538322)] / \exp(0.57719588) = 0.3047391

Reciprocal shows the value for 1/Risk Ratio.

Nonlinear Parametric Survival Models

This section shows how to use the Nonlinear platform for survival models. You only need to learn the techniques in this section if:

- the model is nonlinear.
- you need a distribution other than Weibull, lognormal, exponential, Fréchet, or loglogistic.
- you have censoring that is not the usual right, left, or interval censoring.

With the ability to estimate parameters in specified loss functions, the Nonlinear platform becomes a powerful tool for fitting maximum likelihood models. See the *Modeling and Multivariate Methods* book for complete information about the Nonlinear platform.

To fit a nonlinear model when data are censored, you first use the formula editor to create a parametric equation that represents a loss function adjusted for censored observations. Then use the Nonlinear command in the *Analyze > Modeling* menu, which estimates the parameters using maximum likelihood.

As an example, suppose that you have a table with the variable *time* as the response. First, create a new column, *model*, that is a linear model. Use the calculator to build a formula for *model* as the natural log of *time* minus the linear model—that is, $\ln(\text{time}) - B_0 + B_1z$ where $z$ is the regressor.

Then, because the nonlinear platform minimizes the loss and you want to maximize the likelihood, create a loss function as the negative of the log-likelihood. The log-likelihood formula must be a conditional formula that depends on the censoring of a given observation (if some of the observations are censored).

Loss Formulas for Survival Distributions

The following formulas are for the negative log-likelihoods to fit common parametric models. Each formula uses the calculator `if` conditional function with the uncensored case of the conditional first and the right-censored case as the `else` clause. You can copy these formulas from tables in the `Loss Function Templates` folder in Sample Data and paste them into your data table. “Loglogistic Loss Function,” p. 149, shows the loss functions as they appear in the columns created by the formula editor.

Exponential Loss Function

The exponential loss function is shown in “Loglogistic Loss Function,” p. 149, where $\sigma$ represents the mean of the exponential distribution and $Time$ is the age at failure.

A characteristic of the exponential distribution is that the instantaneous failure rate remains constant over time. This means that the chance of failure for any subject during a given length of time is the same regardless of how long a subject has been in the study.
**Weibull Loss Function**

The Weibull density function often provides a good model for the lifetime distributions. You can use the Univariate Survival platform for an initial investigation of data to determine if the Weibull loss function is appropriate for your data.

There are examples of one-parameter, two-parameter, and extreme-value functions in the Loss Function Templates folder.

**Lognormal Loss Function**

The formula shown below is the lognormal loss function where \( \text{Normal Distribution}(\text{model}/\text{sigma}) \) is the standard normal distribution function. The hazard function has value 0 at \( t = 0 \), increases to a maximum, then decreases and approaches zero as \( t \) becomes large.

**Loglogistic Loss Function**

The loglogistic function has a symmetric density with mean 0 and slightly heavier tails than the normal density function. If \( Y \) is distributed as the logistic distribution, \( \exp(Y) \) is distributed as the loglogistic distribution. Once you have selected a loss function, choose the **Nonlinear** command and complete the dialog. If the response is included in the model formula, no \( Y \) variable is needed. The model is the prediction column and a loss function column is the loss column.

**Exponential Loss Function**

\[
\begin{align*}
\text{if} & \quad \text{Censor}=0 \Rightarrow \log\left(\frac{\text{sigma}}{\text{model}}\right) - \frac{\text{Time}}{\text{sigma}} \\
\text{else} & \quad \Rightarrow \left(\frac{\text{Time}}{\text{sigma}}\right) 
\end{align*}
\]

**Weibull Loss Function**

\[
\begin{align*}
\text{if} & \quad \text{Censor}=0 \Rightarrow \frac{\text{Model}}{\text{sigma}} \cdot \exp\left(\frac{-\text{Model}}{\text{sigma}}\right) - \log\left(\frac{\text{sigma}}{\text{model}}\right) \\
\text{else} & \quad \Rightarrow \exp\left(\frac{-\text{Model}}{\text{sigma}}\right) 
\end{align*}
\]

**Lognormal Loss Function**

\[
\begin{align*}
\text{if} & \quad \text{Censor}=0 \Rightarrow -0.5 \cdot \left(\frac{\text{Model}^2}{\text{sigma}}\right) - 0.5 \cdot \log\left(2^\pi\right) - \log\left(\frac{\text{sigma}}{\text{model}}\right) \\
\text{else} & \quad \Rightarrow \log\left(1 - \text{Normal Distribution}\left(\frac{\text{Model}}{\text{sigma}}\right)\right) 
\end{align*}
\]
Nonlinear Parametric Survival Models

Loglogistic Loss Function

\[
\text{Loss} = \begin{cases} 
\frac{\text{Model}}{\text{sigma}} - 2\cdot \log \left( 1 + \exp \left( \frac{\text{Model}}{\text{sigma}} \right) \right) - \log (\text{sigma}) 
\end{cases}
\]

Weibull Loss Function Example

This example uses the VA Lung Cancer.jmp table. Models are fit to the survival time using the Weibull, lognormal, and exponential distributions. Model fits include a simple survival model containing only two regressors, a more complex model with all the regressors and some covariates, and the creation of dummy variables for the covariate Cell Type to be included in the full model.

Open the data table VA Lung Cancer.jmp. The first model and all the loss functions have already been created as formulas in the data table. The model column has the formula

\[
\log(\text{Time}) - (b_0 + b_1\cdot\text{Age} + b_2\cdot\text{Diag Time})
\]

Initial Parameter Estimates

Nonlinear model fitting is often sensitive to the initial values you give to the model parameters. In this example, one way to find reasonable initial values is to first use the Nonlinear platform to fit only the linear model. When the model converges, the solution values for the parameters become the initial parameter values for the nonlinear model.

To do this select Analyze > Modeling > Nonlinear. Assign Model as X, Predictor Formula, and click OK to see the Nonlinear Control Panel, shown to the left in Figure 6.21. When you click Go, the platform computes the least squares parameter estimates for this model, as shown in the right-hand Control Panel.

Click Save Estimates in the Control Panel to set the parameter estimates in the column formulas to those estimated by this initial nonlinear fitting process.
Estimates with Respect to the Loss Function

The Weibull column has the Weibull formula previously shown. To continue with the fitting process, choose Nonlinear again, select Model as X, Predictor Formula column and Weibull loss as Loss. When you click OK, the Nonlinear Fit Control Panel on the left in Figure 6.22 appears. There is now the additional parameter called sigma in the loss function. Because it is in the denominator of a fraction, a starting value of 1 is reasonable for sigma. When using any loss function other than the default, the Loss is Neg LogLikelihood box on the Control Panel is checked by default. When you click Go, the fitting process converges as shown in the right-hand control panel.
The fitting process estimates the parameters by maximizing the negative log of the Weibull likelihood function. After the Solution table appears, you can click the Confidence Limits button on the Control Panel to get lower and upper 95% confidence limits for the parameters, as shown in the Solution table.

**Figure 6.23 Solution Report**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>ApproxStdErr</th>
<th>Lower CL</th>
<th>Upper CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>b0</td>
<td>5.61304465622</td>
<td>0.67171937</td>
<td>4.31160667</td>
<td>6.95494024</td>
</tr>
<tr>
<td>b1</td>
<td>-0.012428709</td>
<td>0.01114031</td>
<td>-0.0343969</td>
<td>0.00930037</td>
</tr>
<tr>
<td>b2</td>
<td>-0.010457622</td>
<td>0.00034899</td>
<td>-0.0204552</td>
<td>0.0095455</td>
</tr>
<tr>
<td>sigma</td>
<td>1.159937790</td>
<td>0.07023485</td>
<td>1.02336074</td>
<td>1.33040517</td>
</tr>
</tbody>
</table>

Solved By: Analytic NR

**Note:** Because the confidence limits are profile likelihood confidence intervals instead of the standard asymptotic confidence intervals, they can take time to compute.

You can also run the model with the predefined exponential and lognormal loss functions. Before you fit another model, reset the parameter estimates to the least squares estimates, as they may not converge otherwise. To reset the parameter estimates, select Revert to Original Parameters on the Nonlinear Fit drop-down menu.

### Tobit Model Example

You can also analyze left-censored data in the Nonlinear platform. For the left-censored data, zero is the censored value because it also represents the smallest known time for an observation. The tobit model is popular in economics for responses that must be positive or zero, with zero representing a censor point.

The Tobit2.jmp data table in the Reliability sample data folder can be used to illustrate a Tobit model. The response variable is a measure of the durability of a product and cannot be less than zero (Durable, is left-censored at zero). Age and Liquidity are independent variables. The table also includes the model and tobit loss function. The model in residual form is $durable-(b0+b1*age+b2*liquidity)$. Tobit Loss has the formula shown here.
Nonlinear Parametric Survival Models

\[
\begin{align*}
\text{if } \text{duration} == 0 & \Rightarrow \log(1 - \text{Normal Distribution}) \Rightarrow \text{Model} \Rightarrow \frac{\text{Sigma}}{} \\
\text{else} & \Rightarrow 0.5 \times \left( \frac{\text{Model}}{} \right)^2 \\
& \Rightarrow 0.5 \times \log(2 \times \pi) 
\end{align*}
\]
To run the model choose Analyze > Modeling > Nonlinear. Select Model as X, Predictor Formula and Tobit Loss as Loss, and click OK. When the Nonlinear Fit Control Panel appears, the model parameter starting values are set to (near) zero, the loss function parameter Sigma is set to 1, and the Loss is Neg LogLikelihood is checked. Click Go. If you also click Confidence Limits in the Control Panel after the model converges, you see the solution table shown here.

Figure 6.24 Solution Report

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>ApproxStdErr</th>
<th>Lower CL</th>
<th>Upper CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>b0</td>
<td>15.2771</td>
<td>16.0327</td>
<td>-22.0599</td>
<td>54.7012</td>
</tr>
<tr>
<td>b1</td>
<td>-0.1340</td>
<td>0.0190</td>
<td>-0.7386</td>
<td>0.32816</td>
</tr>
<tr>
<td>b2</td>
<td>-0.0451</td>
<td>0.0092</td>
<td>-0.1058</td>
<td>0.0153</td>
</tr>
<tr>
<td>Sigma</td>
<td>5.5892</td>
<td>1.7281</td>
<td>3.3181</td>
<td>11.5291</td>
</tr>
</tbody>
</table>

Fitting Simple Survival Distributions

The following examples show how to use maximum likelihood methods to estimate distributions from time-censored data when there are no effects other than the censor status. The Loss Function Templates folder has templates with formulas for exponential, extreme value, loglogistic, lognormal, normal, and one- and two-parameter Weibull loss functions. To use these loss functions, copy your time and censor values into the Time and censor columns of the loss function template.

To run the model, select Nonlinear and assign the loss column as the Loss variable. Because both the response model and the censor status are included in the loss function and there are no other effects, you do not need prediction column (model variable).

Exponential, Weibull and Extreme-Value Loss Function Examples

The Fan.jmp data table in the Reliability sample data folder illustrates the Exponential, Weibull, and Extreme value loss functions discussed in Nelson (1982). The data are from a study of 70 diesel fans that accumulated a total of 344,440 hours in service. The fans were placed in service at different times. The response is failure time of the fans or run time, if censored.

Here are the formulas for the loss functions as they appear in the formula editor.
Chapter 6

Nonlinear Parametric Survival Models

To use the exponential loss function, select **Nonlinear**, choose **Exponential** as the **Loss** function, and click **OK**. In the Nonlinear Fit Control Panel, enter 1 as the starting value for **Sigma**, and click **Go**. After the model converges, click **Confidence Limits** to see the results shown here.
To use the Weibull loss function with two parameters or the extreme value loss function, again select Nonlinear and choose the loss function you want. Use starting values of 1 for the alpha and beta parameters in the Weibull function and for delta and lambda in the extreme-value function. The results are shown above.

**Note:** Be sure to check the Loss is Neg LogLikelihood check box before you click Go.

**Lognormal Loss Function Example**

The Locomotive.jmp data in the Reliability sample data folder can be used to illustrate a lognormal loss. The lognormal distribution is useful when the range of the data is several powers of 10. The logNormal column in the table has the formula:
The lognormal loss function can be very sensitive to starting values for its parameters. Because the lognormal is similar to the normal distribution, you can create a new variable that is the log10 of Time and use Distribution to find the mean and standard deviation of this column. Then, use those values as starting values for the Nonlinear platform. In this example the mean of log10 of Time is 2.05 and the standard deviation is 0.15.

Run this example as described in the previous examples. Assign lognormal as the Loss function. In the Nonlinear Fit Control Panel give Mu and Sigma the starting values 2.05 and 0.15 and click Go. After the Solution is found, you can click Confidence Limits, on the Control Panel and see the table shown here.

Figure 6.26 Solution Report

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Approx StdErr</th>
<th>Lower CL</th>
<th>Upper CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mu</td>
<td>2.2222521686</td>
<td>0.04523491</td>
<td>2.1442676</td>
<td>2.32727262</td>
</tr>
<tr>
<td>Sigma</td>
<td>0.3063921639</td>
<td>0.04047563</td>
<td>0.24079986</td>
<td>0.46017465</td>
</tr>
</tbody>
</table>

The lognormal loss function can be very sensitive to starting values for its parameters. Because the lognormal is similar to the normal distribution, you can create a new variable that is the log10 of Time and use Distribution to find the mean and standard deviation of this column. Then, use those values as starting values for the Nonlinear platform. In this example the mean of log10 of Time is 2.05 and the standard deviation is 0.15.

Run this example as described in the previous examples. Assign lognormal as the Loss function. In the Nonlinear Fit Control Panel give Mu and Sigma the starting values 2.05 and 0.15 and click Go. After the Solution is found, you can click Confidence Limits, on the Control Panel and see the table shown here.

Figure 6.26 Solution Report

<table>
<thead>
<tr>
<th>Solution</th>
<th>Loss</th>
<th>DFE</th>
<th>Avg Loss</th>
<th>Avg Loss</th>
<th>Sqrt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>237.09307523</td>
<td>94</td>
<td>2.522268</td>
<td>1.5881648</td>
<td></td>
</tr>
</tbody>
</table>

Note: Remember to Save Estimates before requesting confidence limits.

The maximum likelihood estimates of the lognormal parameters are 2.2223 for Mu and 0.3064 for Sigma (in base 10 logs). The corresponding estimate of the median of the lognormal distribution is the antilog of 2.2223 (10^{2.2223}), which is approximately 167. This represents the typical life for a locomotive engine.
A variability chart plots the mean for each level of a second factor, with all plots side by side. Along with the data, you can view the mean, range, and standard deviation of the data in each category, seeing how they change across the categories. The analysis options assume that the primary interest is how the mean and variance change across the categories.

- A traditional name for this chart is a multivar chart, but because that name is not well known, we use the more generic term variability chart.
- A variability chart shows data side-by-side like the Oneway platform, but it has been generalized to handle more than one grouping column.
- A common use of a variability chart is for measurement systems analysis such as gauge R&R, which analyzes how much of the variability is due to operator variation (reproducibility) and measurement variation (repeatability). Gauge R&R is available for many combinations of crossed and nested models, regardless of whether the model is balanced.
- Just as a control chart shows variation across time in a process, a variability chart shows the same kind of variation across categories such as parts, operators, repetitions, and instruments.
- The Variability Chart platform can compute variance components. Several models of crossed and nested factors of purely random models are available.
- Attribute (multi-level) data can also be analyzed with this platform.

Figure 7.1 Example of a Variability Chart
## Contents

Variability Charts ...................................................... 161  
  Launch the Variability Platform .................................. 161  
  Variability Chart ..................................................... 162  
  Variability Platform Options ...................................... 163  
  Heterogeneity of Variance Tests .................................. 165

Variance Components .................................................. 168  
  Variance Component Method ......................................... 169

R&R Measurement Systems ........................................... 171  
  Gauge R&R Variability Report ....................................... 172  
  Misclassification Probabilities ..................................... 174  
  Bias ............................................................................. 174  
  Linearity Study ............................................................ 176

Discrimination Ratio Report ............................................ 177

Attribute Gauge Charts .................................................. 178  
  Data Organization ....................................................... 178  
  Launching the Platform ............................................... 178  
  Attribute Gauge Plots .................................................. 179  
  Agreement ....................................................................... 181  
  Effectiveness Report .................................................... 184
Variability Charts

A variability chart is built to study how a measurement varies across categories. Along with the data, you can view the mean, range, and standard deviation of the data in each category. The analysis options assume that the primary interest is how the mean and variance change across the categories.

A variability chart has the response on the y-axis and a multilevel categorized x-axis. The body of the chart can have the features illustrated in Figure 7.2.

Figure 7.2  Example of a Variability Chart

Launch the Variability Platform

The Variability /Gauge Chart command on the Graph menu displays the Variability Chart Launch dialog shown in Figure 7.3. You specify the classification columns that group the measurements in the X, Grouping list. If the factors form a nested hierarchy, specify the higher terms first. If it is a gauge study, specify operator first and then the part. Specify the measurement column in the Y, Response list. If you specify more than one Y column, there will be a separate variability chart for each response.

Specifying a standard or reference column that contains the “true” or known values for the measured part enables the Bias and Linearity Study options. Both of these options perform analysis on the differences between the observed measurement and the reference or standard value.

The following example uses 2 Factors Crossed.jmp, found in the Variability Data folder.
Variability Charts

Chapter 7

Variability Charts

Figure 7.3 The Variability/Gauge Chart Launch Dialog

**Chart Type** allows you to choose between a variability gauge analysis (for a continuous response) and an attribute gauge analysis (for a categorical response, usually "pass" or "fail"). The first half of this chapter describes variability gauge analysis. For details on attribute gauge analysis, see "Attribute Gauge Charts," p. 178.

**Model Type** allows you to choose the model type (Main Effect, Crossed, Nested, etc.).

**Options** allows you to specify the method for computing variance components (for more details see "Variance Component Method," p. 169), and the alpha level used by the platform.

**Variability Chart**

When you complete the Launch dialog and click OK, the variability chart and the standard deviation chart shown in Figure 7.4 appear by default. This variability chart shows three measurements taken by each operator for parts numbered 1 to 10, with maximum and minimum bars to show the range of measurements. The standard deviation chart plots the standard deviation of measurements taken on each part by each operator.
**Figure 7.4** Variability Charts for Two-Factors Crossed Data

Note: In other platforms, if a row is excluded, it still appears on a chart or plot. But, on variability charts, excluded rows are not shown on the charts. If all the data in a combination of X, Grouping variables is excluded, then that combination does not appear on the Variability Chart or Std Dev Chart.

**Variability Platform Options**

The platform popup menu lets you modify the appearance of the chart, perform Gauge R&R analysis and compute variance components.

- **Vertical Charts**  toggles between horizontal layout and vertical layout.
- **Variability Chart**  toggles the whole variability chart on or off.
- **Show Points**  shows the points for individual rows.
- **Show Range Bars**  shows the bar from the minimum to the maximum of each cell.
- **Show Cell Means**  shows the mean mark for each cell.
- **Connect Cell Means**  connects cell means within a group of cells.
- **Show Separators**  shows the separator lines between levels of the X, Grouping variables.
- **Show Group Means**  shows the mean for groups of cells as a horizontal solid line.
Variability Charts

Show Grand Mean shows the overall mean as a gray dotted line across the whole graph.
Show Grand Median shows the overall median as a blue dotted line across the whole graph.
Show Box Plots toggles box plots on and off.
Mean Diamonds turns the mean diamonds on and off. The confidence intervals use the within-group standard deviation for each cell.
XBar Control Limits draws lines at the UCL and LCL on the Variability chart.
Points Jittered adds some random noise to the plotted points so that coincident points do not plot atop one another.
Show Bias Line toggles the bias line (in the main variability chart) on and off.
Show Standard Mean shows the mean of the standard column. This option is available only when a variable is assigned to the Standard role on the platform launch window.
Variability Summary Report toggles a report that shows the mean, standard deviation, standard error of the mean, lower and upper 95% confidence intervals, and the minimum, maximum, and number of observations.
Std Dev Chart displays a separate graph that shows cell standard deviations across category cells.
Mean of Std Dev toggles a line at the mean standard deviation on the Std Dev chart.
S Control Limits toggles lines showing the LCL and UCL in the Std Dev chart.
Group Means of Std Dev toggles the mean lines on the Std Dev Charts.
Heterogeneity of Variance Tests performs a test for comparing variances across groups. For details, see “Heterogeneity of Variance Tests,” p. 165.
Variance Components estimates the variance components for a specific model. Variance components are computed for these models: nested, crossed, crossed then nested (three factors only), and nested then crossed (three factors only).
Gauge Studies interprets the first factors as grouping columns and the last as Part, and then it creates a gauge R&R report using the estimated variance components. (Note that there is also a Part field in the launch dialog). You are prompted to confirm a given k value to scale the results. You are also prompted for a tolerance interval or historical sigma, but these are optional and can be omitted.
Within this menu, you can request Discrimination Ratio, which characterizes the relative usefulness of a given measurement for a specific product. It compares the total variance of the measurement with the variance of the measurement error. Misclassification Probabilities show probabilities for rejecting good parts and accepting bad parts. Bias Report shows the average difference between the observed values and the standard. A graph of the average biases and a summary table are given for each X variable. Linearity Study performs a regression using the standard values as the X variable and the bias as the Y. This analysis examines the relationship between bias and the size of the part. Ideally, you want the slope to equal 0. A non-zero slope indicates your gauge performs differently with different sized parts. This option is only available when a standard variable is given.
AIAG Labels allows you to specify that quality statistics should be labeled in accordance with the AIAG standard, used extensively in automotive analyses.
A submenu for **Gauge RR Plots** lets you toggle **Mean Plots** (the mean response by each main effect in the model) and **Std Dev** plots. If the model is purely nested, the graphs are displayed with a nesting structure. If the model is purely crossed, interaction graphs are shown. Otherwise, the graphs plot at each effect independently.

**Figure 7.5** Gauge Mean plots for 2 Factors Crossed example

For the standard deviation plots, the red lines connect $\sqrt{\text{mean weighted variance}}$ for each effect.

**Script** has a submenu of commands available to all platforms that let you redo the analysis or save the JSL commands for the analysis to a window or a file.

The default condition of these options and others can be set by using preferences. To access the preferences dialog, select **File > Preferences** from the main JMP menu bar. After the dialog appears, click the **Platforms** icon on the left, and then select **Variability Charts** from the Platforms scroll menu.

**Heterogeneity of Variance Tests**

The **Heterogeneity of Variance Tests** option performs a test for comparing variances across groups. The test is an Analysis of Means for Variances (ANOMV) based method, which tests if any of the group standard deviations are different from the square root of the average group variance.
To be robust against non-normal data, the method uses a permutation simulation to compute decision limits. For complete details on the method, see Wludyka and Sa (2004). Because the method utilizes simulations, the decision limits may be slightly different each time the option is used. To obtain the same results each time, hold down Ctrl-Shift when selecting the option, and specify the same random seed.

As an example, open the 2 Factors Crossed.jmp data table, and follow the steps below:

1. Select Graph > Variability/Gauge Chart.
2. Assign Measurement to the Y, Response role.
3. Assign Operator and part# to the X, Grouping role.
4. Click OK.
5. Select Heterogeneity of Variance Tests from the platform red-triangle menu.
6. Select Crossed.
7. Click OK. Figure 7.7 shows the results.
For the Operator effect, all three levels exceed either the upper or lower decision limits. From this, you conclude that all three standard deviations are different from the square root of the average variance. For the part and interaction effects, none of the levels exceed the decision limits, so you conclude that none of the standard deviations are different from the square root of the average variance.
The red-triangle menus for the effect reports have the following options:

- **Set Alpha Level** is used for setting the alpha level for the test.
- **Show Summary Report** shows or hides a summary report for the test. The report gives the same values given in the plot.
- **Display Options** is used to show or hide the decision limits, shading, center line, and needles.

**Note:** The values given in the plots and the Summary Reports are not the group standard deviations, but the values used in performing the test.

### Variance Components

You can model the variation from measurement to measurement with a model, where the response is assumed to be a constant mean plus random effects associated with various levels of the classification. The exact model depends on how many new random values exist. For example, in a model where \( B \) is nested within \( A \), multiple measurements are nested within both \( B \) and \( A \), and there are \( na \times nb \times nw \) measurements. There are \( na \) random effects due to \( A \), \( na \times nb \) random effects due to each \( nb \) levels within \( A \), and \( na \times nb \times nw \) random effects due to each \( nw \) levels within \( B \) within \( A \), which can be written:

\[
y_{ijk} = u + Z_{ai} + Z_{bij} + Z_{wijk},
\]

The \( Z \)'s are the random effects for each level of the classification. Each \( Z \) is assumed to have mean zero and to be independent from all other random terms. The variance of the response \( y \) is the sum of the variances due to each \( z \) component:

\[
\text{Var}(y_{ijk}) = \text{Var}(Z_{ai}) + \text{Var}(Z_{bij}) + \text{Var}(Z_{wijk}).
\]

To request variance components, select **Variance Components** from the platform popup menu. If you ask for **Variance Components** estimates and you have not specified the type of model in the launch dialog, then you get a dialog like the one shown in Figure 7.8.

---

**Figure 7.8  Variance Component Dialog**

---

Table 7.1 “Models Supported by the Variability Charts Platform,” p. 169, shows the models supported and what the effects in the model would be.
Table 7.1 Models Supported by the Variability Charts Platform

<table>
<thead>
<tr>
<th>Model</th>
<th>Factors</th>
<th>Effects in the model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossed</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>A, B, A*B</td>
</tr>
<tr>
<td>Nested</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>A, B(A)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>A, B(A), C(A,B)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>A, B(A), C(A,B), D(A,B,C)</td>
</tr>
<tr>
<td>Crossed then Nested</td>
<td>3</td>
<td>A, B, A*B, C(A,B)</td>
</tr>
<tr>
<td>Nested then Crossed</td>
<td>3</td>
<td>A, B(A), C, A<em>C, C</em>B(A)</td>
</tr>
</tbody>
</table>

Variance Component Method

The platform launch dialog allows you to choose the method for computing variance components. Click the Analysis Settings button to get the following dialog.

Figure 7.9 Variance Component Options
Choose best analysis (EMS, REML, or Bayesian) is the default option. The logical flow of this option is described below:

- If the data are balanced, and if no variance components are negative, the EMS (expected mean squares) method is used to estimate the variance components.
- If the data are unbalanced, the REML (restricted maximum likelihood) method is used, unless a variance component is estimated to be negative, then the Bayesian method is used.
- If any variance component is estimated to be negative using the EMS method, the Bayesian method is used.
- If there is confounding in the variance components, then the bounded REML method is used, and any negative variance component estimates are set to zero.

Choose best analysis (EMS or REML) has the same logical flow as the first option, but never uses the Bayesian method, even for negative variance components. In that case, the bounded REML method is used and any negative variance component is forced to be 0.

Use REML analysis forces the platform to use the bounded REML method, even if the data are balanced. The bounded REML method can handle unbalanced data and forces any negative variance component to be 0.

Use Bayesian analysis forces the platform to use the Bayesian method. The Bayesian method can handle unbalanced data and forces all variance components to be positive and non-zero. If there is confounding in the variance components, then the bounded REML method is used, and any negative variance component estimates are set to zero.

The Maximum Iterations and Convergence Limit options only affect the REML method. The Number of Iteration Abscissas and Maximum Number of Function Evaluations options only affect the Bayesian method. Making these options more stringent increases the accuracy of results.

**Bayesian Method**

The Bayesian method leads to positive variance component estimates. The method implemented in JMP computes the posterior means using a modified version of Jeffreys’ prior. For details see Portnoy (1971) and Sahai (1974).

**Example**

The Analysis of Variance shows the significance of each effect in the model. The Variance Components report shows the estimates themselves. Figure 7.10 shows these reports after selecting the Crossed selection in the dialog.
R&R Measurement Systems

Measurement systems analysis is an important step in any quality control application. Before you study the process itself, you need to make sure that you can accurately and precisely measure the process. This generally means the variation due to measurement errors is small relative to the variation in the process. The instruments that take the measurements are called gauges, and the analysis of their variation is a gauge study. If most of the variation you see comes from the measuring process itself, then you aren’t reliably learning about the process. So, you do a measurement systems analysis, or gauge R&R study, to find out if the measurement system itself is performing well enough.

Gauge R&R results are available for all combinations of crossed and nested models, regardless of whether the model is balanced.

You collect a random sample of parts over the entire range of part sizes from your process. Select several operators randomly to measure each part several times. The variation is then attributed to the following sources:

- The process variation, from one part to another. This is the ultimate variation that you want to be studying if your measurements are reliable.
- The variability inherent in making multiple measurements—that is, repeatability. In Table 7.2 “Definition of Terms and Sums in Gauge R&R Analysis,” p. 172, this is called the within variation.
- The variability due to having different operators measure parts—that is, reproducibility.

A Gauge R&R analysis then reports the variation in terms of repeatability and reproducibility.
In the same way that a Shewhart control chart can identify processes that are going out of control over time, a variability chart can help identify operators, instruments, or part sources that are systematically different in mean or variance.

### Gauge R&R Variability Report

The Gauge R&R report shows measures of variation interpreted for a gauge study of operators and parts. When you select the Gauge Studies > Gauge RR option, the dialog shown here appears for you to change $K$, enter the tolerance for the process (which is the range of the specification limits, USL – LSL) and enter the historical sigma.

**Choose tolerance entry method** lets you choose the tolerance entry method.

- **Tolerance Interval** lets you enter the tolerance directly, where tolerance = USL – LSL.
- **LSL and/or USL** lets you enter the spec limits, then have JMP calculate the tolerance.

Note that the tolerance interval, spec limits, and historical sigma are optional. The Historical Mean is used for computing the tolerance range for one-sided spec limits, either USL-Historical Mean or Historical Mean-LSL. If no historical mean is entered, the grand mean is used.

**Figure 7.11 Enter/Verify Gauge R&R Specifications Window**
Also note that there is a platform preference (found in JMP’s preferences) that allows you to set the default $K$ that appears on this dialog.

In this example the report shows the statistics as a percentage of the tolerance interval (Upper Spec Limit minus Lower Spec Limit). The values are square roots of sums of variance components scaled by a value $k$, which is 6 in this example. Figure 7.12 shows the Gauge R&R report for the example shown previously, using the data in 2 Factors Crossed.jmp.

**Figure 7.12 Gauge R&R Report**

<table>
<thead>
<tr>
<th>Measurement Source</th>
<th>Variation ($\sqrt{\text{Sum of Var}}$) which is $6 \sqrt{\text{t}}$ of</th>
<th>% of Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatability (EV)</td>
<td>$0.3890052$ Equipment Variation $V(\text{Within})$</td>
<td></td>
</tr>
<tr>
<td>Reproducibility (AV)</td>
<td>$0.4550050$ Appraiser Variation $V(\text{Operator} \times \text{part})$</td>
<td></td>
</tr>
<tr>
<td>Operator</td>
<td>$0.6868194$ Operator $V(\text{Operator})$</td>
<td></td>
</tr>
<tr>
<td>OperatorPART</td>
<td>$0.6489223$ Operator-PART $V(\text{Operator} \times \text{part})$</td>
<td></td>
</tr>
<tr>
<td>Gauge R&amp;R (GR&amp;R)</td>
<td>$0.5990524$ Measurement Variation $V(\text{Within}) + V(\text{Operator} \times \text{part})$</td>
<td></td>
</tr>
<tr>
<td>Part variation (PV)</td>
<td>$1.0425529$ Part Variation $V(\text{part})$</td>
<td></td>
</tr>
<tr>
<td>Total Variation (TV)</td>
<td>$1.2015425$ Total Variation $V(\text{Within}) + V(\text{Operator} \times \text{part}) + V(\text{part})$</td>
<td></td>
</tr>
</tbody>
</table>

$6 \sqrt{\text{t}} = 49.2571$ % Gauge R&R = 100*(RRT/V)

<table>
<thead>
<tr>
<th>Precision to Part Variation = RRT/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Number of Distinct Categories = 1.41(PURR)</td>
</tr>
</tbody>
</table>

Using last column table for Part.

**Variance Components for Gauge R&R**

<table>
<thead>
<tr>
<th>Component</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauge R&amp;R</td>
<td>24.76</td>
</tr>
<tr>
<td>Reproducibility</td>
<td>14.44</td>
</tr>
<tr>
<td>Part-PART</td>
<td>17.34</td>
</tr>
</tbody>
</table>

Barrentine (1991) suggests the following guidelines for an acceptable RR percent (percent measurement variation):

**Table 7.3 Acceptable Variation**

<table>
<thead>
<tr>
<th>%</th>
<th>Acceptable Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10%</td>
<td>excellent</td>
</tr>
<tr>
<td>11% to 20%</td>
<td>adequate</td>
</tr>
<tr>
<td>21% to 30%</td>
<td>marginally acceptable</td>
</tr>
<tr>
<td>&gt; 30%</td>
<td>unacceptable</td>
</tr>
</tbody>
</table>

- If a tolerance interval is given on the Gauge specifications dialog, a new column appears in the Gauge R&R report called “% of Tolerance”. This column is computed as 100*(Variation/Tolerance). In addition, the Precision-to-Tolerance ratio is presented at the bottom of the report. It represents the proportion of the tolerance or capability interval that is lost due to gauge variability.
If a historical sigma is given on the Gauge specifications dialog, a new column appears in the Gauge R&R report, named “% Process”. This column is defined as \(100 \times \frac{\text{Variation}}{K \times \text{Historical Sigma}}\).

Number of distinct categories (NDC) is given in the summary table beneath the Gauge R&R report. NDC is defined as \(1.41 \times (\frac{\text{PV}}{\text{RR}})\), rounded down to the nearest integer.

**Misclassification Probabilities**

Due to measurement variation, good parts can be rejected, and bad parts can be accepted. This is called misclassification. To obtain estimates of misclassification probabilities, select **Gauge Studies > Misclassification Probabilities**. If you have not already done so, you are asked to select the model type and enter spec limits. Figure 7.13 shows an example for the 2 Factors Crossed.jmp data with spec limits of 0.5 and 1.1.

### Figure 7.13 Misclassification Probabilities Report

<table>
<thead>
<tr>
<th>Description</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Both good, is rejected)</td>
<td>0.0902</td>
</tr>
<tr>
<td>P(Both bad, is accepted)</td>
<td>0.2787</td>
</tr>
<tr>
<td>P(Part is good and is rejected)</td>
<td>0.0735</td>
</tr>
<tr>
<td>P(Part is bad and is accepted)</td>
<td>0.0255</td>
</tr>
<tr>
<td>P(Part is good)</td>
<td>0.5157</td>
</tr>
</tbody>
</table>

The first two are conditional probabilities, and the second two are joint probabilities. The fifth value is a marginal probability. The first four are probabilities of errors and decrease as the measurement variation decreases.

**Bias**

The **Gauge Studies > Bias Report** option shows a graph and summary table for each X variable, where the average bias, or differences between the observed values and the standard values, is given for each level of the X variable. A t-test for the bias is also given.

**Note:** The Bias option is only available when a Standard variable is given.

As an example, using the MSALinearity.Jmp data table,

- Choose **Graph > Variability/Gauge Chart**.
- Choose **Response** as the **Y, Response** variable.
- Choose **Standard** as the **Standard** variable.
- Choose **Part** as the **X, Grouping** variable.
- Click **OK**.
- From the platform menu, choose **Gauge Studies > Bias Report**.
The output in Figure 7.14 appears.

**Figure 7.14 Bias Report**

The bias (Response minus Standard) is calculated for every row. At the top is a histogram of the bias, along with a t-test testing if the average bias is equal to 0. On the bottom right is a table of average bias values for each part. To show confidence intervals for the bias, right-click in the table and select the options under the Columns submenu. Each of these bias averages is plotted on the graph along with the actual bias values for every part, so you can see the spread. In this example, Part number 1 is biased high and parts 4 and 5 are biased low.

The Measurement Bias Report node has the following options on the popup menu:

- **Confidence Intervals** calculates confidence intervals for the average bias for each part, and places marks on the Measurement Bias Report by Standard plot.

- **Measurement Error Graphs** produces plot and summary statistics of bias by part.
Linearity Study

The **Gauge Studies > Linearity Study** option performs a regression analysis using the standard variable as the X variable, and the bias as the Y. This analysis examines the relationship between bias and the size of the part. Ideally, you want a slope of 0. If the slope shows to be significantly different from zero, then you can conclude there is a significant relationship between the size of the part or variable measured as a standard, and the ability to measure.

**Note:** The **Linearity Study** option is only available when a Standard variable is given.

Following the example above, after creating the Gauge output using the MSALinearity.JMP data table,

- From the platform menu, choose **Gauge Studies > Linearity Study**.
- In the dialog prompting **Specify Process Variation**, enter 16.5368.

The following output should appear:

**Figure 7.15** Linearity Study

At the top of the report are bias summary statistics for each standard. Below that is an ANOVA table for testing if the slope of the line = 0. Below that are the parameters of the line, along with tests for the slope (linearity) and intercept (bias). The test for the intercept is only useful if the test on the slope fails to reject the hypothesis of slope = 0.

**Note:** The equation of the line is shown directly beneath the graph.
The Linearity Study node has the following options on the popup menu:

- **Set Alpha Level** allows you to change the alpha level used in the bias confidence intervals.

- **Linearity by Groups** produces separate linearity plots for each level of the X, Grouping variables specified on the platform launch dialog.

Here, you see that the slope is -0.131667, and the p-value associated with the test on the slope is quite small (<.0001). From this, you can conclude that there is a significant relationship between the size of the parts and the ability to measure them. Looking at the output, the smaller parts appear to have a positive bias and are measured larger than they actually are, whereas the larger parts have a negative bias, with measurement readings being smaller than the actual parts.

### Discrimination Ratio Report

The Gauge Studies > Discrimination Ratio option appends the Discrimination Ratio table to the Variability report. The discrimination ratio characterizes the relative usefulness of a given measurement for a specific product. It compares the total variance of the measurement, \( M \), with the variance of the measurement error, \( E \). The discrimination ratio is computed for all main effects, including nested main effects. The Discrimination Ratio, \( D \), is computed

\[
D = \sqrt{\frac{2(M)}{E}} - 1 = \sqrt{\frac{2(P)}{E}} + 1 = \sqrt{\frac{2(P)}{T-P}} + 1
\]

where

- \( M \) = estimated measurement variance
- \( P \) = estimated part variance
- \( E \) = estimated variance of measurement error
- \( T \) = estimated total variance

A rule of thumb is that when the Discrimination Ratio is less than 2, the measurement cannot detect product variation, so it would be best to work on improving the measurement process. A Discrimination Ratio greater than 4 adequately detects unacceptable product variation, implying a need for the improvement of the production process.
Attribute Gauge Charts

Attribute gauge analysis gives measures of agreement across responses (raters, for example) in tables and graphs summarized by one or more X grouping variables. Attribute data is data where the variable of interest has a finite number of categories. Typically, data will have only two possible results (ex: pass/fail).

Data Organization

Data should be in the form where each rater is in a separate column, since agreement and effectiveness are both computed on these variables. In other words, if you want to compare agreement among raters, each rater needs to be in a separate (Y) column.

Any other variables of interest, (part, instrument, rep, and so on) should appear stacked in one column each. An optional standard column may be defined, which is then used in the Effectiveness Report. An example data table, contained in the sample data folder as Attribute Gauge.jmp, is shown in Figure 7.17.

Figure 7.17  Attribute Data Example

Responses in the different Y columns may be character (Pass/Fail), numeric (0/1), or ordinal (low, medium, high).

Launching the Platform

To begin an attribute gauge analysis, select Graph > Variability/ Gauge Chart. For the Attribute Gauge.jmp example, fill in the dialog as shown in Figure 7.18.
Attribute Gauge Plots

In the plots that appear, by default the % Agreement is plotted, where % Agreement is measured by comparing all pairs of rater by replicate combinations, for each part.

The first plot in Figure 7.19 uses all X (Grouping) variables on the x-axis, while the second plot contains all Y variables on the x-axis (typically the rater). For the top plot,

\[
%\text{Agreement for grouping variable } j = \sum_{i=1}^{\text{number of levels}} \frac{\left( \text{number of responses for level } i \right)}{2 \times \left( \text{number of raters} \times \text{number of reps} \right)}
\]

For the bottom plot in Figure 7.19,

\[
%\text{Agreement for rater } k = \sum_{i=1}^{n} \left( \sum_{j=1}^{r_i} \frac{\text{number of uncounted matching levels for this rater } k \text{ within part } i \text{ for rep } j}{\sum_{i=1}^{n} \sum_{j=1}^{r_i} m_i \times r_i - j} \right)
\]

where

- \( n \) = number of subjects (grouping variables)
- \( r_i \) = number of reps for subject \( i \)
- \( m_i \) = number of raters for subject \( i \)
As an example of the calculations, consider the following table of data for three raters, each having three replicates for one subject.

**Table 7.4** Three replicates for raters A, B, and C

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Using this table,

\[
\text{% Agreement} = \frac{\binom{4}{2} + \binom{5}{2}}{\binom{9}{2}} = \frac{16}{36} = 0.444
\]

\[
\text{% Agreement [rater A]} = \% \text{ Agreement [rater B]} = \frac{4 + 3}{8 + 7} = \frac{10}{21} = 0.476\text{ and}
\]

\[
\text{% Agreement [rater C]} = \frac{4 + 3 + 2}{8 + 7 + 6} = \frac{9}{21} = 0.4286
\]
Agreement

The Agreement Report gives agreement summarized for every rater as well as overall agreement.

The Agreement Comparisons give Kappa statistics for each \( Y \) variable compared with all other \( Y \) variables. In other words, each rater is compared with all other raters.

The Agreement within Raters report shows the number of items that were inspected. The confidence intervals are Score confidence intervals, as suggested by Agresti and Coull, (1998). The Number Matched is defined as the sum of number of items inspected, where the rater agreed with him or herself on each inspection of an individual item. The Rater Score is Number Matched divided by Number Inspected.

The simple kappa coefficient is a measure of inter-rater agreement.

\[
\kappa = \frac{P_0 - P_e}{1 - P_e}
\]

where

\[
P_0 = \sum_i p_{ii}
\]

and

\[
P_e = \sum_i p_i p_{.i}
\]

Viewing the two response variables as two independent ratings of the \( n \) subjects, the kappa coefficient equals +1 when there is complete agreement of the raters. When the observed agreement exceeds chance agreement, the kappa coefficient is positive, with its magnitude reflecting the strength of agreement. Although unusual in practice, kappa is negative when the observed agreement is less than chance agreement. The minimum value of kappa is between -1 and 0, depending on the marginal proportions.

The asymptotic variance of the simple kappa coefficient is estimated by the following:

\[
\text{var} = \frac{A + B - C}{(1 - P_e)^2 n}
\]

where

\[
A = \sum_i p_{ii} \left[ 1 - (p_{.i} + p_{.i})(1 - \kappa) \right]
\]

\[
B = (1 - \kappa)^2 \sum_{i \neq j} \sum p_{ij} (p_{.i} + p_{.j})^2
\]

and

\[
C = \left[ \kappa - P_e (1 - \kappa) \right]^2
\]
The Kappa’s are plotted and the standard errors are also given.

**Note:** The Kappa statistic in the Attribute charts is given even when the levels of the variables are not the same.

Categorical Kappa statistics (Fleiss 1981) are found in the Agreement Across Categories report.

For

\[ n = \text{number of subjects reviewed} \]

\[ m_i = \text{number of ratings on subject } i (i = 1, 2, \ldots, n) \]. This includes responses for all raters, and repeat ratings on a part. For example, if subject \( i \) is measured 3 times by each of 2 raters, then \( m_i \) is 3x2=6.

\[ k = \text{number of categories in the response} \]

\[ x_{ij} = \text{number of ratings on subject } i (i = 1, 2, \ldots, n) \text{ into category } j (j = 1, 2, \ldots k) \]

The individual category Kappa is

\[ \hat{\kappa}_j = 1 - \frac{\sum_{i=1}^{n} x_{ij} (m_i - x_{ij})}{(\hat{p}_j \hat{q}_j) \sum_{i=1}^{n} m_i (m_i - 1)} \]

where

\[ \hat{p}_j = \frac{\sum_{i=1}^{n} x_{ij}}{\sum_{i=1}^{n} m_i} \]

\[ \hat{q}_j = 1 - \hat{p}_j \]

and the overall kappa is

\[ \hat{\kappa} = \frac{\sum_{j=1}^{k} \hat{p}_j \hat{q}_j \hat{\kappa}_j}{\sum_{j=1}^{k} \hat{p}_j \hat{q}_j} \]

The variance of \( \hat{\kappa}_j \) and \( \hat{\kappa} \) are

\[ \text{var}(\hat{\kappa}_j) = \frac{2}{nm(m-1)} \]

\[ \text{var}\hat{\kappa} = \frac{k}{\left( \sum_{j=1}^{k} \hat{p}_j \hat{q}_j \right)^2 \sum_{j=1}^{n} (m_i - 1)} \times \left[ \left( \sum_{j=1}^{k} \hat{p}_j \hat{q}_j \right)^2 - \sum_{j=1}^{k} \hat{p}_j \hat{q}_j \hat{q}_j \hat{q}_j \hat{p}_j \right] \]

The standard errors of \( \hat{\kappa}_j \) and \( \hat{\kappa} \) are only shown when there are an equal number of ratings per subject.
If a standard variable is given, an additional Kappa report is given that compares every rater with the standard.
Effectiveness Report

The Effectiveness Report appears when a standard variable is given.

**Figure 7.21** Effectiveness Report

<table>
<thead>
<tr>
<th>Rate</th>
<th>Correct(0)</th>
<th>Correct(1)</th>
<th>Total Correct</th>
<th>Incorrect(0)</th>
<th>Incorrect(1)</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>45</td>
<td>87</td>
<td>132</td>
<td>3</td>
<td>5</td>
<td>150</td>
</tr>
<tr>
<td>B</td>
<td>45</td>
<td>160</td>
<td>205</td>
<td>2</td>
<td>2</td>
<td>150</td>
</tr>
<tr>
<td>C</td>
<td>42</td>
<td>83</td>
<td>125</td>
<td>9</td>
<td>9</td>
<td>150</td>
</tr>
</tbody>
</table>

**Effectiveness**

<table>
<thead>
<tr>
<th>Rate</th>
<th>Effectiveness</th>
<th>Lower CI</th>
<th>Upper CI</th>
<th>Error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>94.6687</td>
<td>89.2298</td>
<td>97.7270</td>
<td>0.0533</td>
</tr>
<tr>
<td>B</td>
<td>96.1087</td>
<td>92.2548</td>
<td>99.6508</td>
<td>0.0333</td>
</tr>
<tr>
<td>C</td>
<td>90.0300</td>
<td>84.1508</td>
<td>95.6450</td>
<td>0.1000</td>
</tr>
<tr>
<td>Overall</td>
<td>93.7776</td>
<td>91.1542</td>
<td>95.6033</td>
<td>0.0522</td>
</tr>
</tbody>
</table>

**Misclassifications**

<table>
<thead>
<tr>
<th>Standard</th>
<th>Level</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

**Conformance Report**

<table>
<thead>
<tr>
<th>Rate</th>
<th>P(False Alarms)</th>
<th>P(Misses)</th>
<th>NonConform</th>
<th>Conform</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.0410</td>
<td>0.0525</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.0190</td>
<td>0.0825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.0962</td>
<td>0.1250</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Agreement Counts table gives cell counts on the number correct and incorrect for every level of the standard.

*Effectiveness* is defined as (# of correct decisions) / (Total # of opportunities for a decision).

This means that if rater A sampled every part four times, and on the sixth part, one of the decisions did not agree, the other three decisions would still be counted as correct decisions.

**Note:** This is different from the MSA 3rd edition, as all four opportunities for rater A by part 6 would be counted as incorrect. We feel including all inspections separately gives the user more information about the overall inspection process.

In the **Effectiveness Report**, 95% confidence intervals are given about the effectiveness. JMP is using Score Confidence Intervals. In recent literature, it has been demonstrated that score confidence intervals provide more sensible answers, particularly where observations lie near the boundaries (see Agresti and Coull, 1998).
The **Misclassifications** table shows the incorrect labeling with the y-axis representing the levels of the standard, or accepted reference value, and the x-axis containing the levels given by the raters.

The **Conformance Report** displays the probability of false alarms, and the probability of misses, where

- False Alarm = part is determined non-conforming, when it actually is conforming
- Miss = part is determined conforming, when it actually is not conforming.
- $P(\text{False Alarms}) = \frac{\text{Incorrect (PASS)}}{\text{Correct (PASS)} + \text{Incorrect (PASS)}}$.
- $P(\text{Miss}) = \frac{\text{Incorrect (FAIL)}}{\text{Correct (FAIL)} + \text{Incorrect (FAIL)}}$.

The Conformance Report has the following options on the red-triangle menu:

- **Change Conforming Category** is used to reverse the response category considered conforming.
- **Calculate Escape Rate** is used to calculate the Escape Rate, the probability that a non-conforming part will be produced and not detected. It is calculated as the probability that the process will produce a non-conforming part times the probability of a miss. You specify the probability that the process will produce a non-conforming part.

The conformance report is only displayed when the rating has two levels, like pass/fail or 0/1.

**Note:** Missing values are treated as a separate category in this platform. If missing values are removed, different calculations are performed than if the missing values are excluded. We recommend excluding all rows containing missing values.
Variability Charts
Attribute Gauge Charts
Capability analysis, used in process control, measures the conformance of a process to given specification limits. Using these limits, you can compare a current process to specific tolerances and maintain consistency in production. Graphical tools such as the goal plot and box plot give you quick visual ways of observing within-spec behaviors.

**Figure 8.1** Examples of Capability Analyses

### Specification

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
<th>Portion</th>
<th>% Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spec Target</td>
<td>377.0179</td>
<td>Above USL</td>
<td>2.4955</td>
</tr>
<tr>
<td>Upper Spec Limit</td>
<td>430.9453</td>
<td>Total Outside</td>
<td>3.019</td>
</tr>
</tbody>
</table>

### CPK Calculations

- **CPK**:
  - Capability: 1.33
  - Below LSL: 1.17
  - Above USL: 1.50

- **CPK**:
  - Capability: 1.33
  - Below LSL: 1.17
  - Above USL: 1.50

- **CPL**:
  - Capability: 1.33
  - Below LSL: 1.17
  - Above USL: 1.50

- **CPU**:
  - Capability: 1.33
  - Below LSL: 1.17
  - Above USL: 1.50

### Long Term Sigma

- **Sigma**:
  - Long Term: 1.33
  - Targets: 0.67
  - Standardized: 0.67

### Goal Plot

- **Goal**: 1.17
- **Actual**: 1.50
- ** Sigma**: 0.67
- **Percentiles**: 0.049, 0.951
- **Pareto**: 0.049, 0.951
Contents

Launch the Platform ....................................................... 189
Entering Limits ............................................................. 189
Capability Plots and Options ............................................ 191
    Goal Plot .................................................................. 191
    Capability Box Plots ................................................ 192
    Normalized Box Plots ............................................... 194
Individual Detail Reports .................................................. 194
Make Summary Table ...................................................... 195
Capability Indices Report .................................................. 196
Launch the Platform

We use CitySpecLimits.jmp, Cities.jmp, and Semiconductor Capability.jmp in the following examples. To launch the Capability platform, select **Capability** from the **Graph** menu. This presents you with the following dialog box.

**Figure 8.2** Capability Launch Window

Here, you select the variables that you want to analyze. After assigning the desired variables to **Y, Columns**, click **OK** to bring up the Specification Limits dialog. Columns selected in the launch dialog are listed here, with entry fields for the lower specification limit (LSL), target, and upper specification limit (USL).

**Entering Limits**

At this point, specification limits should be entered for each variable. Note that manually adding the limits at this point is only one of the available methods for entering them.

1. If the limits are already stored in a data table, they can be imported using the **Import Spec Limits** command.
2. You can enter the limits as Column Properties, thereby bypassing the spec limits dialog.
3. You can enter the limits on the dialog.
4. You can enter them using JSL.

**Using JSL**

Spec limits can be read from JSL statements or from a spec limits table.

As an example of reading in spec limits from JSL, consider the following code snippet, which places the spec limits inside a **Spec Limits()** clause.

```jsl
// JSL for reading in spec limits
Capability(
```
Using a Limits Data Table

A spec limits table can be in two different formats: wide or tall. Figure 8.3 shows an example of both types.

Figure 8.3 Tall (top) and Wide (bottom) Spec Limit Tables

<table>
<thead>
<tr>
<th></th>
<th>Column 1</th>
<th>_LSL</th>
<th>_Target</th>
<th>_USL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OZONE</td>
<td>0.075</td>
<td>0.15</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>CO</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>SO2</td>
<td>0.01</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>4</td>
<td>NO</td>
<td>0.01</td>
<td>0.025</td>
<td>0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>_LimitsKey</th>
<th>Max deg F Jam</th>
<th>OZONE</th>
<th>CO</th>
<th>SO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LSL</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Target</td>
<td>40</td>
<td>0.1</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>USL</td>
<td>*</td>
<td>0.029</td>
<td>0.1</td>
</tr>
</tbody>
</table>

A tall table has one row for each column analyzed in Capability, with four columns. The first holds the column names. The other three columns need to be named, _LSL, _USL, and _Target.

A wide table has one column for each column analyzed in Capability, with three rows plus a _LimitsKey column. In the _LimitsKey column, the three rows need to contain the identifiers _LSL, _USL, and _Target.

Either of these formats can be read using the Import Spec Limits command.

Using a Limits Table and JSL

There is no extra syntax needed to differentiate between the two table types when they are read using JSL. The following syntax works for either table. It places the spec limits inside an Import Spec Limits() clause.

```julia
// JSL for reading in a spec limits file
Capability(
    Y( :OZONE, :CO, :SO2, :NO ),
    Capability Box Plots( 1 ),
    Spec Limits(
        Import Spec Limits("<path>/filename.JMP"
    ));
```
**Saving Specification Limits**

After entering or loading the specification limits, you can save them to the data table using the *Save Spec Limits as Column Properties* command.

You can also save the specification limits to a new data table with the *Save Spec Limits in New Table* command.

---

**Capability Plots and Options**

By default, JMP shows a goal plot and capability box plots. Using the *Capability* pop-up menu, you can add Normalized box plots, capability indices, and a summary table, as well as display a capability report for each individual variable in the analysis. All platform options are described below.

**Goal Plot**

The Goal plot shows, for each variable, the spec-normalized mean shift on the x-axis, and the spec-normalized standard deviation on the y-axis. It is useful for getting a quick, summary view of how the variables are conforming to specification limits.

For each column with LSL, Target, and USL, these quantities are defined as

\[
\text{Mean Shift Normalized to Spec} = \frac{\text{Mean}(\text{Col}[i]) - \text{Target}}{\text{USL}[i] - \text{LSL}[i]}
\]

\[
\text{Standard Deviation Normalized to Spec} = \frac{\text{Standard Deviation}(\text{Col}[i])}{\text{USL}[i] - \text{LSL}[i]}
\]

To create the plot in Figure 8.4, open the *Semiconductor Capability.jmp* data, and run the attached script.

---

**Figure 8.4 Goal Plot**

[Image of a Goal Plot with CPK slider and CPK edit box]
By default, the CPK slider and number edit box is set to CPK = 1. This approximates a defect rate of 0.001. The red goal line represents the CPK shown in the edit box. To change the CPK value, move the slider or enter a number in the edit box. Points on the plot represent columns, not rows.

The shaded areas are described as follows. Let C represent the value shown in the CPK edit box.

- Points in the red area have CPK < C.
- Points in the yellow area have C < CPK < 2C.
- Points in the green area have 2C < CPK.

There is a preference for plotting PPK instead of CPK. When this is on, the slider is labeled with PPK.

The Goal Plot pop-up menu has the following commands:

- **Shade CPK Levels** shows or hides the CPK level shading.
- **Goal Plot Labels** shows or hides the labels on the points.
- **Defect Rate Contour** shows or hides a contour representing a defect rate you specify.

### Capability Box Plots

Capability box plots show a box plot for each variable in the analysis. The values for each column are centered by their target value and scaled by the difference between the specification limits. That is, for each column $Y_j$,

$$Y_j = \frac{Y_{ij} - T_j}{USL_j - LSL_j} \text{ with } T_j \text{ being the target}$$
Figure 8.5 Capability Box Plot

The left and right green lines, drawn at ±0.5, represent the $\text{LSL}_j$ and $\text{USL}_j$ respectively. This plot is useful for comparing variables with respect to their specification limits. For example, the majority of points for IVP1 are above its USL, while IVP2 has the majority of its points less than its target. PNP2 looks to be on target with all data points in the spec limits.

When a spec limit is missing for one or more columns, separate box plots are produced for those columns, with gray lines, as shown here. A note is given at the bottom of the plot that discusses the calculations used for the plot.

Figure 8.6 Note for Missing Spec Limits

Note: spec limit range was set to 2(P/USL Target), because of missing LSL.
Normalized Box Plots

When drawing Normalized box plots, JMP first standardizes each column by subtracting off the mean and dividing by the standard deviation. Next, quantiles are formed for each standardized column. The box plots are formed for each column from these standardized quantiles.

Figure 8.7 Normalized Box Plot

The green vertical lines represent the spec limits normalized by the mean and standard deviation. The gray vertical lines are drawn at ±0.5, since the data is standardized to a standard deviation of 1.

Individual Detail Reports

The Individual Detail Reports command shows a capability report for each variable in the analysis. This report is identical to the one from the Distribution platform, detailed in the Basic Analysis and Graphing book.
Figure 8.8 Individual Detail Report

This option makes a summary table that includes the variable’s name, its spec-normalized mean shift, and its spec-normalized standard deviation.

Figure 8.9 Summary Table

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean Shift Normalized to Spec</th>
<th>Std dev Normalized to Spec</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 EP2</td>
<td>-0.302772532</td>
<td>0.373085199224</td>
</tr>
<tr>
<td>2 NPN4</td>
<td>-0.084416444</td>
<td>0.1064029205</td>
</tr>
<tr>
<td>3 NF2</td>
<td>-0.0544158078</td>
<td>1.1799601131</td>
</tr>
<tr>
<td>4 NPN3</td>
<td>-0.056828799</td>
<td>0.0503418504</td>
</tr>
<tr>
<td>5 PNP4</td>
<td>0.003075872988</td>
<td>0.0566116219</td>
</tr>
<tr>
<td>6 NF1</td>
<td>1.360139792</td>
<td>0.5558989922</td>
</tr>
<tr>
<td>7 PNP3</td>
<td>0.0153977755</td>
<td>0.2609700335</td>
</tr>
<tr>
<td>8 NPN2</td>
<td>0.0569013305</td>
<td>0.0612288463</td>
</tr>
<tr>
<td>9 PNP2</td>
<td>0.007336131</td>
<td>0.0663496123</td>
</tr>
<tr>
<td>10 PNP1</td>
<td>0.009541422</td>
<td>0.2139687772</td>
</tr>
<tr>
<td>11 NPN1</td>
<td>-0.122294596</td>
<td>0.0968722163</td>
</tr>
</tbody>
</table>
Capability Indices Report

This option shows or hides a table showing each variable's LSL, USL, target, mean, standard deviation, Cp, CPK, and PPM. Optional columns for this report are Lower CI, Upper CI, CPM, CPL, CPU, Ppm Below LSL, and Ppm Above USL. To reveal these optional columns, right-click on the report and select the column names from the **Columns** submenu.

**Figure 8.10** Capability Indices

<table>
<thead>
<tr>
<th>Column</th>
<th>LSL</th>
<th>Target</th>
<th>USL</th>
<th>Mean</th>
<th>Std Deviation</th>
<th>CP</th>
<th>CPK</th>
<th>PPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP2</td>
<td>73.30508</td>
<td>76.2596</td>
<td>79.20811</td>
<td>74.47235</td>
<td>2.205364</td>
<td>0.445609</td>
<td>0.1759</td>
<td>314708.3</td>
</tr>
<tr>
<td>NIPN1</td>
<td>95.88567</td>
<td>105.8876</td>
<td>115.8896</td>
<td>104.189</td>
<td>2.126478</td>
<td>1.566373</td>
<td>1.3019</td>
<td>649369.7</td>
</tr>
<tr>
<td>IVP2</td>
<td>139.2004</td>
<td>142.3052</td>
<td>145.4099</td>
<td>138.2452</td>
<td>7.227184</td>
<td>0.141244</td>
<td>-0.0435</td>
<td>71591.5</td>
</tr>
<tr>
<td>NIPN3</td>
<td>97.31768</td>
<td>120.5047</td>
<td>144.2517</td>
<td>118.1352</td>
<td>2.364757</td>
<td>3.310698</td>
<td>2.9344</td>
<td>665e-13</td>
</tr>
<tr>
<td>PNP4</td>
<td>-54.4319</td>
<td>230.7356</td>
<td>531.9091</td>
<td>256.3756</td>
<td>32.60738</td>
<td>2.956975</td>
<td>2.8187</td>
<td>7.73e-16</td>
</tr>
<tr>
<td>JVP1</td>
<td>59.62007</td>
<td>68.41011</td>
<td>87.20015</td>
<td>72.78072</td>
<td>4.196328</td>
<td>0.30106</td>
<td>-0.5227</td>
<td>941949.5</td>
</tr>
<tr>
<td>PNP3</td>
<td>118.6778</td>
<td>130.2898</td>
<td>141.9018</td>
<td>137.6146</td>
<td>8.060762</td>
<td>0.638642</td>
<td>0.2358</td>
<td>240559.3</td>
</tr>
<tr>
<td>NIPN2</td>
<td>96.59381</td>
<td>113.749</td>
<td>130.9042</td>
<td>115.7471</td>
<td>2.100788</td>
<td>2.722029</td>
<td>2.4058</td>
<td>2.651e-7</td>
</tr>
<tr>
<td>PNP1</td>
<td>164.3695</td>
<td>281.0719</td>
<td>423.6463</td>
<td>313.0687</td>
<td>58.35335</td>
<td>0.757614</td>
<td>0.6659</td>
<td>82820.42</td>
</tr>
<tr>
<td>NIPN1</td>
<td>104.4129</td>
<td>118.1552</td>
<td>131.8535</td>
<td>114.7928</td>
<td>2.862101</td>
<td>1.72048</td>
<td>1.2097</td>
<td>48.27419</td>
</tr>
</tbody>
</table>
Chapter 9

Pareto Plots
The Pareto Plot Platform

A Pareto plot is a statistical quality improvement tool that shows frequency, relative frequency, and cumulative frequency of problems in a process or operation. It is a bar chart that displays severity (frequency) of problems in a quality-related process or operation. The bars are ordered by frequency in decreasing order from left to right, which makes a Pareto plot useful for deciding what problems should be solved first.

Launch the Pareto Plot platform with the Pareto Plot command in the Graph menu (or toolbar).

Figure 9.1 Examples of Pareto Charts
## Contents

*Pareto Plots* ................................................................. 199  
  Assigning Variable Roles .................................................. 199  
  Pareto Plot Platform Commands ........................................... 201  
  Options for Bars ............................................................ 202

*Launch Dialog Options* ............................................................. 204

*Threshold of Combined Causes* .................................................. 204

*One-Way Comparative Pareto Plot* ................................................. 206

*Two-Way Comparative Pareto Plot* ................................................. 208

*Defect Per Unit Analysis* ........................................................... 210

  Using Number of Defects as Sample Size .................................. 210
  Using a Constant Sample Size Across Groups ................................. 210
  Using a Non-Constant Sample Size Across Groups .......................... 212
Pareto Plots

The Pareto Plot command produces charts to display the relative frequency or severity of problems in a quality-related process or operation. A Pareto plot is a bar chart that displays the classification of problems arranged in decreasing order. The column whose values are the cause of a problem is assigned as $Y$ and is called the process variable. The column whose values hold the frequencies are assigned as $Freq$.

You can also request a comparative Pareto Plot, which is a graphical display that combines two or more Pareto Plots for the same process variable. Pareto Plot then produces a single graphical display with plots for each value in a column assigned the $X$ role, or combination of levels from two $X$ variables. Columns with the $X$ role are called classification variables.

The Pareto Plot command can chart a single $Y$ (process) variable with no $X$ classification variables, with a single $X$, or with two $X$ variables. The Pareto facility does not distinguish between numeric and character variables or between modeling types. All values are treated as discrete, and bars represent either counts or percentages. The following list describes the arrangement of the Pareto graphical display:

- A $Y$ variable with no $X$ classification variables produces a single chart with a bar for each value of the $Y$ variable.
- A $Y$ variable with one $X$ classification variable produces a row of Pareto plots. There is a plot for each level of the $X$ variable with bars for each $Y$ level.
- A $Y$ variable with two $X$ variables produces rows and columns of Pareto plots. There is a row for each level of the first $X$ variable and a column for each level of the second $X$ variable. The rows have a Pareto Plot for each value of the first $X$ variable, as described previously.

The following sections illustrate each of these arrangements.

Assigning Variable Roles

The Failure.jmp table (Figure 9.2) from the Quality Control sample data folder lists causes of failure during the fabrication of integrated circuits. The $N$ column in the table to the right lists the number of times each kind of defect occurred. It is a $Freq$ variable in the Pareto Launch dialog. For the raw data table, shown on the left (Figure 9.2), causes of failure are not grouped. The Pareto Plot command produces the same results from either of these tables. The following example uses the failure data with a frequency column.
When you select the **Pareto Plot** command, you see the Pareto Plot launch dialog shown in Figure 9.3. Select the *failure* column (causes of failure) as *Y, Cause*. It is the variable you want to inspect with Pareto plots. The *N* column in the data table is the *Freq* variable. When you click **OK**, you see the Pareto plot shown in Figure 9.4.

The left axis represents the count of failures, and the right axis represents the percent of failures in each category. For example, contamination accounts for 45% of the failures. The bars are in decreasing order with the most frequently occurring failure to the left. The curve indicates the cumulative failures from left to right. If you place the crosshairs from the **Tools** menu on the point above the oxide defect bar, the cumulative percent axis shows that contamination and oxide defect together account for 71% of the failures.

The type of scale and arrangement of bars are display options and are described in the next section. The options can be changed with the popup menu on the title bar of the window.
Pareto Plots Platform Commands

The popup menu on the Pareto plot title bar has commands that tailor the appearance of Pareto plots. It also has options in the Causes submenu that affect individual bars within a Pareto plot.

The following commands affect the appearance of the Pareto plot as a whole:

- **Percent Scale**  toggles between the count and percent left vertical axis display.
- **N Legend**  toggles the total sample size in the plot area.
- **Category Legend**  toggles between labeled bars and a separate category legend.
- **Pie Chart**  toggles between the bar chart and pie chart representation.
- **Reorder Horizontal, Reorder Vertical**  reorders grouped Pareto plots when there is one or more grouping variables.
- **Ungroup Plots**  allows a group of Pareto charts to be split up into separate plots.
- **Count Analysis**  lets you perform defect per unit analyses. See “Defect Per Unit Analysis,” p. 210 for a description of these commands.
- **Show Cum Percent Curve**  toggles the cumulative percent curve above the bars and the cumulative percent axis on the vertical right axis.
- **Show Cum Percent Axis**  toggles the cumulative percent axis on the vertical right axis.
- **Show Cum Percent Points**  toggles the points on the cumulative percent curve.
- **Label Cum Percent Points**  toggles the labels on the points on the cumulative curve.
- **Cum Percent Curve Color**  lets you change the color of the cumulative percent curve.
Causes has options that affect one or more individual chart bars. See “Options for Bars,” p. 202, for a description of these options.

Script is the standard JMP Script menu.

Figure 9.5 shows the effect of various options. The plot has the Percent Scale option off, shows counts instead of percents, and uses the Category Legend on its horizontal axis. The vertical count axis is rescaled and has grid lines at the major tick marks.

Figure 9.5 Pareto Plots with Display Options

Options for Bars

You can highlight a bar by clicking on it. Use Control-click (-click on the Macintosh) to select multiple bars that are not contiguous. When you select bars, you can access the commands on the platform menu that affect Pareto plot bars. They are found on the Causes submenu on the platform popup menu. These options are also available with a right-click (Control-click on the Macintosh) anywhere in the plot area. The following options apply to highlighted bars instead of to the chart as a whole.

- **Combine Causes** combines selected (highlighted) bars.
- **Separate Causes** separates selected bars into their original component bars.
- **Move to First** moves one or more highlighted bars to the left (first) position.
Move to Last moves one or more highlighted bars to the right (last) position.

Colors shows the colors palette for coloring one or more highlighted bars.

Markers shows the markers palette for assigning a marker to the points on the cumulative percent curve, when the Show Cum Percent Points command is in effect.

Label displays the bar value at the top of all highlighted bars.

The example Pareto plot on the left in Figure 9.7 is the default plot with a bar for each cause of failure. The example on the right shows combined bars. To combine bars, first select Causes > Combine to launch the Combine Causes dialog window. Complete the dialog as shown below.

![Figure 9.6 Combine Causes](image)

You can separate the highlighted bars into original categories with the Separate Causes option.

![Figure 9.7 Example of Combining Bars](image)
The plots in Figure 9.8 show the same data. The plot to the right results when you select the Pie Chart display option.

**Launch Dialog Options**

The Threshold of Combined Causes command is described in this section. The Per Unit Analysis command is described in the section “Defect Per Unit Analysis,” p. 210.

**Threshold of Combined Causes**

This option allows you to specify a threshold for combining causes by specifying a minimum Count or a minimum Rate. To specify the threshold, select the Threshold of Combined Causes option on the launch dialog, as shown in Figure 9.9.
You then select Tail % or Count in the drop-down menu that appears and enter the threshold value.

For example, using Failure.jmp, specifying a minimum count of 2 resulted in the following Pareto plot. All causes with counts 2 or fewer are combined into the final bar labeled 4 Others. (Compare to Figure 9.7 to see which causes were combined).
One-Way Comparative Pareto Plot

This section uses the contamination data called Failure2.jmp in the Quality Control sample data folder. This table records failures in a sample of capacitors manufactured before cleaning a tube in the diffusion furnace and in a sample manufactured after cleaning the furnace. For each type of failure, the variable clean identifies the samples with the values “before” or “after.” It is a classification variable and has the X role in the Pareto Plot launch dialog.

The grouping variable produces one Pareto plot window with side-by-side plots for each value of the X, Grouping variable, clean. You see the two Pareto plots in Figure 9.12.

These plots are referred to as the cells of a comparative Pareto plot. There is a cell for each level of the X (classification) variable. Because there is only one X variable, this is called a one-way comparative Pareto plot.
The horizontal and vertical axes are scaled identically for both plots. The bars in the first plot are in descending order of the $y$-axis values and determine the order for all cells.

The plots are arranged in alphabetical order of the classification variable levels. The levels (“after” and “before”) of the classification variable, clean, show above each plot. You can rearrange the order of the plots by clicking on the title (level) of a classification and dragging it to another level of the same classification. For example, the comparative cells of the clean variable shown in Figure 9.12 are logically reversed. The “after” plot is first and is in descending order. The “before” plot is second and is reordered to conform with the “after” plot. A comparison of these two plots shows a reduction in oxide defects after cleaning, but the plots would be easier to interpret if presented as the before-and-after plot shown in Figure 9.13.
Note that the order of the causes has been changed to reflect the order based on the first cell.

**Two-Way Comparative Pareto Plot**

You can study the effects of two classification variables simultaneously with a two-way comparative Pareto plot. Suppose that the capacitor manufacturing process from the previous example monitors production samples before and after a furnace cleaning for three days. The `Failure3.jmp` table has the column called `date` with values OCT 1, OCT 2, and OCT 3. To see a two-way array of Pareto plots, select the Pareto Plot command and add both `clean` and `date` to the X, Grouping list as shown in Figure 9.14.
Two grouping variables produce one Pareto plot window with a two-way layout of plots that show each level of both $X$ variables (Figure 9.15). The upper-left cell is called the *key cell*. Its bars are arranged in descending order. The bars in the other cells are in the same order as the key cell. You can reorder the rows and columns of cells with the **Reorder Vertical** or **Reorder Horizontal** options in the platform popup menu. The cell that moves to the upper-left corner becomes the new key cell and the bars in all other cells rearrange accordingly. You can also click-and-drag as before to rearrange cells.

You can click bars in the key cell and the bars for the corresponding categories highlight in all other cells. Use Control-click (⌘-click on the Macintosh) to select nonadjacent bars.

The Pareto plot shown in Figure 9.15 illustrates highlighting the *vital few*. In each cell of the two-way comparative plot, the bars representing the two most frequently occurring problems are selected. Contamination and Metallization are the two vital categories in all cells, but they appear to be less of a problem after furnace cleaning.

**Figure 9.15** Two-way Comparative Pareto Plot
Defect Per Unit Analysis

The Defect Per Unit analysis allows you to compare defect rates across groups. JMP calculates the defect rate as well as 95% confidence intervals of the defect rate. You may optionally specify a constant sample size on the launch dialog.

Although causes are allowed to be combined in Pareto plots, the calculations for these analyses do not change correspondingly.

Using Number of Defects as Sample Size

As an example of calculating the rate per unit, use the Failures.jmp sample data table (Note that this is not Failure.jmp, but Failures.jmp). After assigning Causes to Y, Cause and Count to Freq, click OK to generate a Pareto plot.

When the chart appears, select Count Analysis > Per Unit Rates from the platform drop-down menu to get the Per Unit Rates table shown in Figure 9.16.

Figure 9.16 Per Unit Rates Table

<table>
<thead>
<tr>
<th>Cause</th>
<th>Count</th>
<th>Rate</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contamination</td>
<td>110</td>
<td>0.4104</td>
<td>0.3373</td>
<td>0.4847</td>
</tr>
<tr>
<td>Oxide Defect</td>
<td>85</td>
<td>0.3208</td>
<td>0.2957</td>
<td>0.3493</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>18</td>
<td>0.0871</td>
<td>0.0398</td>
<td>0.1381</td>
</tr>
<tr>
<td>Silicon Defect</td>
<td>17</td>
<td>0.0634</td>
<td>0.0370</td>
<td>0.1015</td>
</tr>
<tr>
<td>Corrosion</td>
<td>16</td>
<td>0.0597</td>
<td>0.0341</td>
<td>0.0970</td>
</tr>
<tr>
<td>Metallization</td>
<td>11</td>
<td>0.0410</td>
<td>0.0295</td>
<td>0.0734</td>
</tr>
<tr>
<td>Doping</td>
<td>10</td>
<td>0.0373</td>
<td>0.0179</td>
<td>0.0665</td>
</tr>
<tr>
<td>Pooled Total</td>
<td>268</td>
<td>0.1428</td>
<td>0.1293</td>
<td>0.1610</td>
</tr>
</tbody>
</table>

Since there was no sample size entered on the launch dialog, the total number of defect counts across causes is used to calculate each rate and their 95% confidence interval.

Using a Constant Sample Size Across Groups

Using Failures.jmp, fill in the dialog as shown in Figure 9.17 and click OK. Note that checking Per Unit Analysis causes options to appear.
When the report appears, select **Count Analysis > Test Rates Across Groups**. This produces the analysis shown in the bottom of Figure 9.18.
The **Test Rates Across Groups** command tests (a likelihood-ratio chi square) whether the defects per unit (DPU) for each cause is the same across groups.

The **Test Rate Within Groups** command tests (a likelihood-ratio chi square) whether the defects per unit (DPU) across causes are the same within a group.

### Using a Non-Constant Sample Size Across Groups

To specify a unique sample size for a group, add rows to the data table for each group, specifying a special cause code (e.g., “size”) to designate the rows as size rows. For example, open `Failuresize.jmp`. Among the other causes (Oxide Defect, Silicon Defect, etc.) is a cause labeled size.

To conduct the analysis, fill in the Pareto launch dialog like the one shown in Figure 9.19. Be sure to type `size` as lower case.

![Figure 9.19 Non-Constant Sample Size Launch](image)

After clicking **OK**, select both **Per Unit Rates** and **Test Rates Across Groups**, found under **Count Analysis** in the platform drop-down. The resulting report is shown in Figure 9.20.
Note that the sample size of 101 is used to calculate the DPU for the causes in group A, while the sample size of 145 is used to calculate the DPU for the causes in group B.

If there are two group variables (say, Day and Process), Per Unit Rates lists DPU or rates for every combination of Day and Process for each cause. However, Test Rate Across Groups only tests overall differences between groups.
The Diagram platform is used to construct Ishikawa charts, also called fishbone charts, or cause-and-effect diagrams.

Figure 10.1 Example of an Ishikawa Chart

These charts are useful to organize the sources (causes) of a problem (effect), perhaps for brainstorming, or as a preliminary analysis to identify variables in preparation for further experimentation.
Contents

Preparing the Data for Diagramming .................................................. 217
Chart Types ....................................................................................... 218
Building a Chart Interactively .......................................................... 219
  Text Menu ....................................................................................... 220
  Insert Menu .................................................................................... 220
  Move Menu ..................................................................................... 222
Other Menu Options ........................................................................... 223
Drag and Drop .................................................................................. 223
Preparing the Data for Diagramming

To produce the diagram, begin with data in two columns of a data table.

**Figure 10.2** Ishikawa Diagram

![Ishikawa Diagram](image)

Some sample data (Montgomery, 1996) is shown in Figure 10.3, from the Ishikawa jmp sample data table.

**Figure 10.3** A Portion of the Ishikawa Sample Data

<table>
<thead>
<tr>
<th>Parent</th>
<th>Child</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defects in circuit ba</td>
<td>Inspection</td>
</tr>
<tr>
<td>Defects in circuit ba</td>
<td>Solder process</td>
</tr>
<tr>
<td>Defects in circuit ba</td>
<td>Raw card</td>
</tr>
<tr>
<td>Defects in circuit ba</td>
<td>Components</td>
</tr>
<tr>
<td>Defects in circuit ba</td>
<td>Component insertion</td>
</tr>
<tr>
<td>Inspection</td>
<td>Measurement</td>
</tr>
<tr>
<td>Test coverage</td>
<td>Inspector</td>
</tr>
<tr>
<td>Solder process</td>
<td>Flux</td>
</tr>
<tr>
<td>Solder process</td>
<td>Splatter</td>
</tr>
<tr>
<td>Solder process</td>
<td>Chain speed</td>
</tr>
<tr>
<td>Solder process</td>
<td>Temperature</td>
</tr>
<tr>
<td>Solder process</td>
<td>Wave pump</td>
</tr>
<tr>
<td>Temperature</td>
<td>Setup</td>
</tr>
</tbody>
</table>

Notice that the Parent value “Defects in circuit board” (rows 1–5) has many causes, listed in the Child column. Among these causes is “Inspection”, which has its own children causes listed in rows 6–8.

Examine the plot in Figure 10.2 to see “Defects in circuit board” as the center line, with its children branching off above and below. “Inspection” is one of these branches, which has its own branches for its child causes.
Select **Diagram** to bring up the launch dialog (Figure 10.4). Provide the columns that represent the **X**, **Parent** (Parent in the sample data) and the **Y**, **Child** (Child in the sample data).

**Figure 10.4** Diagram Launch Dialog

Including a variable in the **By** column produces separate diagrams for each value of the **By** variable. **Label** columns cause the text from them to be included in the nodes of the diagram.

**Chart Types**

The Diagram platform can draw charts of three types: Fishbone (Figure 10.5), Hierarchy (Figure 10.6), and Nested (Figure 10.7).

**Figure 10.5** Fishbone Chart

**Figure 10.6** Hierarchy Chart
To change the chart type, right-click (control-click on the Macintosh) on any node line in the chart. The nodes highlight as the mouse passes over them. Then, select the desired chart from the Change Type menu.

Building a Chart Interactively

Right click on any node in the chart to bring up a context menu (Figure 10.9) that allows a chart to be built piece-by-piece. You can edit new nodes into the diagram using context menus, accessible by right-clicking on the diagram itself. You can even create a diagram without a data table, which starts with the default diagram shown here.
Text Menu

There are two ways to change the appearance of text in a diagram:

- Right-click on a highlighted node in the chart. This brings up the menu shown in Figure 10.9 that has the following options:
  - **Font** brings up a dialog to select the font of the text.
  - **Color** brings up a dialog to select the color of the text.
  - **Rotate Left, Rotate Right, Horizontal** rotates the text to be horizontal, rotated 90 degrees left, or 90 degrees right (see Figure 10.10).

![Rotated Text Example](image)

- Right-click on a word in the chart. This brings up a smaller menu that has the following options:
  - **Font** brings up a dialog to select the font of the text.
  - **Font Color** brings up a dialog to select the color of the text.
  - **Rotate Text** rotates the text to be **Horizontal**, rotated 90 degrees **Left**, or 90 degrees **Right** (see Figure 10.10).
  - **Justify Text** brings up a dialog to justify the text left, center, or right.
  - **Hide** hides the text.
  - **Bullet Point** adds a bullet point to the left of the text.
  - **Set Wrap** brings up a dialog that allows you to set the text wrap width in pixels.

Insert Menu

The **Insert** menu allows you to insert items onto existing nodes.

- **Before** inserts a new node at the same level of the highlighted node. The new node appears before the highlighted node. For example, inserting “Child 1.5” before “Child 2” results in the following chart.
**Figure 10.11** Insert Before

![Insert Before Diagram]

*After* inserts a new node at the same level of the highlighted node. The new node appears after the highlighted node. For example, inserting “Child 3” after “Child 2” results in the following chart.

**Figure 10.12** Insert After

![Insert After Diagram]

*Above* inserts a new node at a level above the current node. For example, inserting “Grandparent” above “Parent” results in the following chart.

**Figure 10.13** Insert Above

![Insert Above Diagram]

*Below* inserts a new node at a level below the current node. For example, inserting “Grandchild 1” below “Child 2” results in the following chart.

**Figure 10.14** Insert Below

![Insert Below Diagram]
Move Menu

The Move menu allows you to customize the appearance of the diagram by giving you control over which side the branches appear on.

**First** moves the highlighted node to the first position under its parent. For example, after switching sides, telling “Child 2” to Move First results in the following chart.

![Figure 10.15 Move First](image)

**Last** moves the highlighted node to the last position under its parent.

**Other Side** moves the highlighted node to the other side of its parent line. For example, telling “Child 2” to switch sides results in the following chart.

![Figure 10.16 Move Other Side](image)

**Force Left** makes all horizontally-drawn elements appear to the left of their parent.

**Force Right** makes all horizontally-drawn elements appear to the right of their parent.

**Force Up** makes all vertically-drawn elements appear above their parent.

**Force Down** makes all vertically-drawn elements appear below their parent.

**Force Alternate** is the default setting, which draws siblings on alternate sides of the parent line.
Other Menu Options

- **Change Type** changes the entire chart type to Fishbone, Hierarchy, or Nested.
- **Uneditable** disables all other commands except Move and Change Type.
- **Text Wrap Width** brings up a dialog that allows you to specify the width of labels where text wrapping occurs.
- **Make Into Data Table** converts the currently highlighted node into a data table. Note that this can be applied to the whole chart by applying it to the uppermost level of the chart.
- **Close** is a toggle to alternately show or hide the highlighted node.
- **Delete** deletes the highlighted node and everything below it.

Drag and Drop

Nodes in a Diagram can be manipulated by drag and drop. Grab any outlined section of a Diagram and drag to a new location, using the highlighted bar as a guide to tell you where the node will appear after being dropped.

For example, the following picture shows the element Setup (initially a child of Temperature) being dragged to a position that makes it a child of Solder Process—in essence, equivalent to Flux, Splatter, and Chain Speed.
The next example shows two ways to make the Inspection tree a child of the Solder Process tree. Note that both drag operations result in the same final result.

These principles extend to nested and Hierarchical charts. The following example shows the two ways to move Temperature from its initial spot (under Moisture Content) to a new position, under Missing from reel.
Figure 10.20 Example of Dragging Elements

Drag here to make Temperature a sibling of Vendor and Setup.

Drag here to make Temperature a child of Missing from reel.

Both drag operations have the same result.
Ishikawa Diagrams
Building a Chart Interactively

Chapter 10
This chapter contains descriptions of options that apply to all control charts.

**Figure 11.1 Example of a Control Chart**
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Quality Control with Control Charts</td>
<td>229</td>
</tr>
<tr>
<td>The Control Chart Launch Dialog</td>
<td>229</td>
</tr>
<tr>
<td>- Process Information</td>
<td>231</td>
</tr>
<tr>
<td>- Chart Type Information</td>
<td>233</td>
</tr>
<tr>
<td>- Parameters</td>
<td>234</td>
</tr>
<tr>
<td>- Using Specified Statistics</td>
<td>235</td>
</tr>
<tr>
<td>Tailoring the Horizontal Axis</td>
<td>236</td>
</tr>
<tr>
<td>Display Options</td>
<td>236</td>
</tr>
<tr>
<td>- Single Chart Options</td>
<td>237</td>
</tr>
<tr>
<td>- Window Options</td>
<td>239</td>
</tr>
<tr>
<td>Tests for Special Causes</td>
<td>241</td>
</tr>
<tr>
<td>- Western Electric Rules</td>
<td>241</td>
</tr>
<tr>
<td>- Westgard Rules</td>
<td>244</td>
</tr>
<tr>
<td>Running Alarm Scripts</td>
<td>246</td>
</tr>
<tr>
<td>Saving and Retrieving Limits</td>
<td>247</td>
</tr>
<tr>
<td>Real-Time Data Capture</td>
<td>251</td>
</tr>
<tr>
<td>- Setting Up a Script in a Data Table</td>
<td>253</td>
</tr>
<tr>
<td>Excluded, Hidden, and Deleted Samples</td>
<td>253</td>
</tr>
</tbody>
</table>
Statistical Quality Control with Control Charts

Control charts are a graphical and analytic tool for deciding whether a process is in a state of statistical control and for monitoring an in-control process.

Control charts have the following characteristics:

- Each point represents a summary statistic computed from a subgroup sample of measurements of a quality characteristic.
- The vertical axis of a control chart is scaled in the same units as the summary statistic.
- The horizontal axis of a control chart identifies the subgroup samples.
- The center line on a Shewhart control chart indicates the average (expected) value of the summary statistic when the process is in statistical control.
- The upper and lower control limits, labeled UCL and LCL, give the range of variation to be expected in the summary statistic when the process is in statistical control.
- A point outside the control limits (or the V-mask of a CUSUM chart) signals the presence of a special cause of variation.
- **Graph > Control Chart** subcommands create control charts that can be updated dynamically as samples are received and recorded or added to the data table.

**Figure 11.2** Description of a Control Chart

The Control Chart Launch Dialog

When you select a Control Chart from the **Graph > Control Chart** menu (Figure 11.3), you see a Control Chart Launch dialog similar to the one in Figure 11.4. (The exact controls vary depending on the type of chart you choose.) Initially, the dialog shows three kinds of information:

- process information, for measurement variable selection
- chart type information
- limits specification.
Specific information shown for each section varies according to the type of chart you request.

Through interaction with the Launch dialog, you specify exactly how you want your charts created. The following sections describe the panel elements.
Process Information

The Launch dialog displays a list of columns in the current data table. Here, you specify the variables to be analyzed and the subgroup sample size.

Process

The Process role selects variables for charting.

- For variables charts, specify measurements as the process.
- For attribute charts, specify the defect count or defective proportion as the process. The data will be interpreted as counts, unless it contains non-integer values between 0 and 1.

**Note:** The rows of the table must be sorted in the order you want them to appear in the control chart. Even if there is a Sample Label variable specified, you still must sort the data accordingly.

Sample Label

The Sample Label role enables you to specify a variable whose values label the horizontal axis and can also identify unequal subgroup sizes. If no sample label variable is specified, the samples are identified by their subgroup sample number.

- If the sample subgroups are the same size, check the Sample Size Constant radio button and enter the size into the text box. If you entered a Sample Label variable, its values are used to label the horizontal axis.
- If the sample subgroups have an unequal number of rows or have missing values and you have a column identifying each sample, check the Sample Grouped by Sample Label radio button and enter the sample identifying column as the sample label.

For attribute charts (P-, NP-, C-, and U-charts), this variable is the subgroup sample size. In Variables charts, it identifies the sample. When the chart type is IR, a Range Span text box appears. The range span specifies the number of consecutive measurements from which the moving ranges are computed.

**Note:** The rows of the table must be sorted in the order you want them to appear in the control chart. Even if there is a Sample Label variable specified, you still must sort the data accordingly.

The illustration in Figure 11.5 shows an \( \bar{X} \)-chart for a process with unequal subgroup sample sizes, using the Coating.jmp sample data from the Quality Control sample data folder.
Phase

The Phase role enables you to specify a column identifying different phases, or sections. A phase is a group of consecutive observations in the data table. For example, phases might correspond to time periods during which a new process is brought into production and then put through successive changes. Phases generate, for each level of the specified Phase variable, a new sigma, set of limits, zones, and resulting tests. See “Phases,” p. 277 in the “Shewhart Control Charts” chapter for complete details of phases. For the Diameter.jmp data, found in the sample data, launch an XBar Control Chart. Then specify Diameter as Process, Day as Sample Label, Phase as Phase, and check the box beside S for an S control chart to obtain the two phases shown in Figure 11.6.
Figure 11.6 XBar and S Charts with Two Phases

Chart Type Information

Shewhart control charts are broadly classified as variables charts and attribute charts. Moving average charts and cusum charts can be thought of as special kinds of variables charts.
Figure 11.7 Dialog Options for Variables Control Charts

- **XBar** charts menu selection gives XBar, R-, and S check boxes.
- The **IR** menu selection has checkbox options for the Individual Measurement, Moving Range, and Median Moving Range charts.
- The uniformly weighted moving average (UWMA) and exponentially weighted moving average (EWMA) selections are special charts for means.
- The **CUSUM** chart is a special chart for means or individual measurements.
- **Presummarize** allows you to specify information on pre-summarized statistics.
- **P, NP, C, and U charts**, **Run Chart**, and **Levey-Jennings** charts have no additional specifications.

The types of control charts are discussed in “Shewhart Control Charts,” p. 255.

**Parameters**

You specify computations for control limits by entering a value for \( k \) (**K Sigma**), or by entering a probability for \( \alpha \) (**Alpha**), or by retrieving a limits value from the process columns' properties or a previously created Limits Table. Limits Tables are discussed in the section “Saving and Retrieving Limits,” p. 247, later in this chapter. There must be a specification of either **K Sigma** or **Alpha**. The dialog default for **K Sigma** is 3.

**K Sigma**

The **K Sigma** parameter option allows specification of control limits in terms of a multiple of the sample standard error. **K Sigma** specifies control limits at \( k \) sample standard errors above and below the expected value, which shows as the center line. To specify \( k \), the number of sigmas, click the radio button for **K Sigma** and enter a positive \( k \) value into the text box. The usual choice for \( k \) is 3, which is three standard errors.
deviations. The examples shown in Figure 11.8 compare the $\bar{X}$-chart for the Coating.jmp data with control lines drawn with $K \text{ Sigma} = 3$ and $K \text{ Sigma} = 4$.

**Figure 11.8** $K \text{ Sigma} = 3$ (left) and $K \text{ Sigma} = 4$ (right) Control Limits

---

**Alpha**

The **Alpha** parameter option specifies control limits (also called *probability limits*) in terms of the probability $\alpha$ that a single subgroup statistic exceeds its control limits, assuming that the process is in control. To specify alpha, click the **Alpha** radio button and enter the probability you want. Reasonable choices for $\alpha$ are 0.01 or 0.001. The **Alpha** value equivalent to a $K \text{ Sigma}$ of 3 is 0.0027.

**Using Specified Statistics**

After specifying a process variable, if you click the **Specify Stats** (when available) button on the Control Chart Launch dialog, a tab with editable fields is appended to the bottom of the launch dialog. This lets you enter historical statistics (statistics obtained from historical data) for the process variable. The Control Chart platform uses those entries to construct control charts. The example here shows 1 as the standard deviation of the process variable and 20 as the mean measurement.

**Figure 11.9** Example of Specify Stats
Note: When the mean is user-specified, it is labeled in the plot as $\mu_0$.

If you check the Capability option on the Control Chart launch dialog (see Figure 11.4), a dialog appears as the platform is launched asking for specification limits. The standard deviation for the control chart selected is sent to the dialog and appears as a Specified Sigma value, which is the default option. After entering the specification limits and clicking OK, capability output appears in the same window next to the control chart. For information on how the capability indices are computed, see the Basic Analysis and Graphing book.

**Tailoring the Horizontal Axis**

When you double-click the x-axis, the X Axis Specification dialog appears for you to specify the format, axis values, number of ticks, gridline and reference lines to display on the x-axis.

For example, the Pickles.JMP data lists measurements taken each day for three days. In this example, by default, the x-axis is labeled at every other tick. Sometimes this gives redundant labels, as shown to the left in Figure 11.10. If you specify a label at an increment of eight, with seven ticks between them, the x-axis is labeled once for each day, as shown in the chart on the right.

**Figure 11.10 Example of Labeled x-Axis Tick Marks**

---

**Display Options**

Control Charts have popup menus that affect various parts of the platform:

- The menu on the top-most title bar affects the whole platform window. Its items vary with the type of chart you select.
- There is a menu of items on the chart type title bar with options that affect each chart individually.
Single Chart Options

The popup menu of chart options appears when you click the icon next to the chart name, or context-click the chart space (right-mouse click on Windows or Control-click on the Macintosh). The CUSUM chart has different options that are discussed in “Cumulative Sum Control Charts,” p. 281.

**Box Plots** superimposes box plots on the subgroup means plotted in a Mean chart. The box plot shows the subgroup maximum, minimum, 75th percentile, 25th percentile, and median. Markers for subgroup means show unless you deselect the **Show Points** option. The control limits displayed apply only to the subgroup mean. The **Box Plots** option is available only for $\bar{X}$-charts. It is most appropriate for larger subgroup sample sizes (more than 10 samples in a subgroup).

**Needle** connects plotted points to the center line with a vertical line segment.

**Connect Points** toggles between connecting and not connecting the points.

**Show Points** toggles between showing and not showing the points representing summary statistics. Initially, the points show. You can use this option to suppress the markers denoting subgroup means when the **Box Plots** option is in effect.

**Connect Color** displays the JMP color palette for you to choose the color of the line segments used to connect points.

**Center Line Color** displays the JMP color palette for you to choose the color of the line segments used to draw the center line.

**Limits Color** displays the JMP color palette for you to choose the color of the line segments used in the upper and lower limits lines.

**Line Width** allows you to pick the width of the control lines. Options are Thin, Medium, or Thick.

**Point Marker** allows you to pick the marker used on the chart.

**Show Center Line** initially displays the center line in green. Deselecting **Show Center Line** removes the center line and its legend from the chart.

**Show Control Limits** toggles between showing and not showing the chart control limits and their legends.

**Tests** shows a submenu that enables you to choose which tests to mark on the chart when the test is positive. Tests apply only for charts whose limits are $3\sigma$ limits. Tests 1 to 4 apply to Mean, Individual and attribute charts. Tests 5 to 8 apply to Mean charts, Presummarize, and Individual Measurement charts only. If tests do not apply to a chart, the Tests option is dimmed. Tests apply, but will not appear for charts whose control limits vary due to unequal subgroup sample sizes, until the sample sizes become equal. These special tests are also referred to as the **Western Electric rules**. For more information on special causes tests, see **Tests for Special Causes** later in this chapter.

**Westgard Rules** are detailed below. See the text and chart on p. 244.

**Test Beyond Limits** flags as a “*” any point that is beyond the limits. This test works on all charts with limits, regardless of the sample size being constant, and regardless of the size of $k$ or the width of the limits. For example, if you had unequal sample sizes, and wanted to flag any points beyond the limits of an $r$-chart, you could use this command.
Statistical Control Charts
Display Options

**Show Zones** toggles between showing and not showing the zone lines. The zones are labeled A, B, and C as shown here in the Mean plot for weight in the Coating.jmp sample data. Control Chart tests use the zone lines as boundaries. The seven zone lines are set one sigma apart, centered on the center line.

**Figure 11.11 Show Zones**

**Shade Zones** toggles between showing and not showing the default green, yellow, and red colors for the three zone areas and the area outside the zones. Green represents the area one sigma from the center line, yellow represents the area two and three sigmas from the center line, and red represents the area beyond three sigma. Shades may be shown with or without the zone lines.

**Figure 11.12 Shade Zones**
**OC Curve** gives Operating Characteristic (OC) curves for specific control charts. OC curves are defined in JMP only for $\bar{X}$-, $P$-, $NP$-, $C$-, and $U$-charts. The curve shows how the probability of accepting a lot changes with the quality of the sample. When you choose the **OC Curve** option from the control chart option list, JMP opens a new window containing the curve, using all the calculated values directly from the active control chart. Alternatively, you can run an OC curve directly from the **Control** category of the JMP Starter. Select the chart on which you want the curve based, then a dialog prompts you for **Target**, **Lower Control Limit**, **Upper Control Limit**, $k$, **Sigma**, and **Sample Size**. You can also perform both single and double acceptance sampling in the same manner. To engage this feature, choose **View > JMP Starter > Control** (under Click Category) > **OC Curves**. A pop-up dialog box allows you to specify whether or not single or double acceptance sampling is desired. A second pop-up dialog is invoked, where you can specify acceptance failures, number inspected, and lot size (for single acceptance sampling). Clicking **OK** generates the desired OC curve.

**Window Options**

The popup menu on the window title bar lists options that affect the report window. The example menu shown here appears if you request **XBar** and **R** at the same time. You can check each chart to show or hide it.

---

**Figure 11.13: Report Options**

The specific options that are available depend on the type of control chart you request. Unavailable options show as grayed menu items.

The following options show for all control charts except Run Charts:

- **Show Limits Legend** shows or hides the Avg, UCL, and LCL values to the right of the chart.
- **Connect through Missing** connects points when some samples have missing values.
- **Capability** performs a Capability Analysis for your data. A popup dialog is first shown, where you can enter the Lower Spec Limit, Target, and Upper Spec Limit values for the process variable.
Figure 11.14 Capability Analysis Dialog

An example of a capability analysis report is shown in Figure 11.15 for Coating.jmp when the Lower Spec Limit is set as 16.5, the Target is set to 21.5, and the Upper Spec Limit is set to 23.

Figure 11.15 Capability Analysis Report for Coating.jmp

For additional information on Capability Analysis, see the Basic Analysis and Graphing book.

Save Sigma saves the computed value of sigma as a column property in the process variable column in the JMP data table.

Save Limits > in Column saves the computed values of sigma, center line, and the upper and lower limits as column properties in the process variable column in the JMP data table. These limits are later automatically retrieved by the Control Chart dialog and used in a later analysis.
Figure 11.16 Properties in the Column Info Window

Save Limits > in New Table saves all parameters for the particular chart type, including sigma and K Sigma, sample size, the center line, and the upper and lower control limits in a new JMP data table. These limits can be retrieved by the Control Chart dialog and used in a later analysis. See the section “Saving and Retrieving Limits,” p. 247 for more information.

Save Summaries creates a new data table that contains the sample number, sample label (if there is one), and statistic being plotted for each plot within the window.

Alarm Script displays a dialog for choosing or entering a script or script name and executes by either writing to the log or speaking whenever the tests for special causes is in effect and a point is out of range. See section “Tests for Special Causes,” p. 241 for more information. See “Running Alarm Scripts,” p. 246 for more information on writing custom Alarm Scripts.

Script has a submenu of commands available to all platforms that let you redo the analysis or save the JSL commands for the analysis to a window or a file.

Tests for Special Causes

The Tests option in the chart type popup menu displays a submenu for test selection. You can select one or more tests for special causes with the options popup menu. Nelson (1984) developed the numbering notation used to identify special tests on control charts.

If a selected test is positive for a particular sample, that point is labeled with the test number. When you select several tests for display and more than one test signals at a particular point, the label of the numerically lowest test specified appears beside the point.

Western Electric Rules

Western Electric rules are implemented in the Tests submenu. Table 11.1 “Description and Interpretation of Special Causes Tests,” p. 242 lists and interprets the eight tests, and Figure 11.18 illustrates the tests. The following rules apply to each test:

• The area between the upper and lower limits is divided into six zones, each with a width of one standard deviation.
• The zones are labeled A, B, C, C, B, A with zones C nearest the center line.
• A point lies in Zone B or beyond if it lies beyond the line separating zones C and B. That is, if it is more than one standard deviation from the center line.

• Any point lying on a line separating two zones lines is considered belonging to the outermost zone.

Note: All Tests and zones require equal sample sizes in the subgroups of nonmissing data.

Tests 1 through 8 apply to Mean ($\bar{X}$) and individual measurement charts. Tests 1 through 4 can also apply to $P$, $NP$, $C$, and $U$-charts.

Tests 1, 2, 5, and 6 apply to the upper and lower halves of the chart separately. Tests 3, 4, 7, and 8 apply to the whole chart.

See Nelson (1984, 1985) for further recommendations on how to use these tests.

Figure 11.17 Zones for Western Electric Rules

Table 11.1 Description and Interpretation of Special Causes Tests

<table>
<thead>
<tr>
<th>Test 1</th>
<th>One point beyond Zone A</th>
<th>detects a shift in the mean, an increase in the standard deviation, or a single aberration in the process. For interpreting Test 1, the $R$-chart can be used to rule out increases in variation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 2</td>
<td>Nine points in a row in a single (upper or lower) side of Zone C or beyond</td>
<td>detects a shift in the process mean.</td>
</tr>
<tr>
<td>Test 3</td>
<td>Six points in a row steadily increasing or decreasing</td>
<td>detects a trend or drift in the process mean. Small trends will be signaled by this test before Test 1.</td>
</tr>
<tr>
<td>Test 4</td>
<td>Fourteen points in a row alternating up and down</td>
<td>detects systematic effects such as two alternately used machines, vendors, or operators.</td>
</tr>
</tbody>
</table>
Table 11.1 Description and Interpretation of Special Causes Tests (Continued)

<table>
<thead>
<tr>
<th>Test</th>
<th>Description</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 5</td>
<td>Two out of three points in a row in Zone A or beyond</td>
<td>detects a shift in the process average or increase in the standard deviation. Any two out of three points provide a positive test.</td>
</tr>
<tr>
<td>Test 6</td>
<td>Four out of five points in Zone B or beyond</td>
<td>detects a shift in the process mean. Any four out of five points provide a positive test.</td>
</tr>
<tr>
<td>Test 7</td>
<td>Fifteen points in a row in Zone C, above and below the center line</td>
<td>detects stratification of subgroups when the observations in a single subgroup come from various sources with different means.</td>
</tr>
<tr>
<td>Test 8</td>
<td>Eight points in a row on both sides of the center line with none in Zones C</td>
<td>detects stratification of subgroups when the observations in one subgroup come from a single source, but subgroups come from different sources with different means.</td>
</tr>
</tbody>
</table>

Westgard Rules

Westgard rules are implemented under the \textit{Westgard Rules} submenu of the Control Chart platform. The different tests are abbreviated with the decision rule for the particular test. For example, \textit{1 2s} refers to a test where one point is two standard deviations away from the mean.

Because Westgard rules are based on sigma and not the zones, they can be computed without regard to constant sample size.

**Table 11.2 Westgard Rules**

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rule 1 2s</strong></td>
<td>2s is commonly used with Levey-Jennings charts, where control limits are set 2 standard deviations away from the mean. The rule is triggered when any one point goes beyond these limits.</td>
</tr>
<tr>
<td><strong>Rule 1 3s</strong></td>
<td>3s refers to a rule common to Levey-Jennings charts where the control limits are set 3 standard deviations away from the mean. The rule is triggered when any one point goes beyond these limits.</td>
</tr>
<tr>
<td><strong>Rule 2 2s</strong></td>
<td>2s is triggered when two consecutive control measurements are farther than two standard deviations from the mean.</td>
</tr>
<tr>
<td><strong>Rule 4s</strong></td>
<td>4s is triggered when one measurement in a group is two standard deviations above the mean and the next is two standard deviations below.</td>
</tr>
<tr>
<td><strong>Rule 4 1s</strong></td>
<td>1s is triggered when four consecutive measurements are more than one standard deviation from the mean.</td>
</tr>
</tbody>
</table>
Running Alarm Scripts

If you want to run a script that alerts you when the data fail one or more tests, you can run an Alarm Script. As an Alarm Script is invoked, the following variables are available, both in the issued script and in subsequent JSL scripts:

- qc_col is the name of the column
- qc_test is the test that failed
- qc_sample is the sample number
- qc_firstRow is the first row in the sample
- qc_lastRow is the last row in the sample

Example 1: Automatically writing to a log

One way to generate automatic alarms is to make a script and store it with the data table as a Data Table property named QC Alarm Script. To automatically write a message to the log whenever a test fails,

- Run the script below to save the script as a property to the data table,
- Run a control chart,
- Turn on the tests you're interested in. If there are any samples that failed, you'll see a message in the log.

```javascript
CurrentData Table() << Set Property("QC Alarm Script", 
  Write(match( 
    QC_Test, 1, "One point beyond zone A", 
    2, "Nine points in a row in zone C or beyond", 
    3, "Six points in a row Steadily increasing or decreasing", 
    4, "Fourteen points in a row alternating up and down", 
    5, "Two out of three points in a row in Zone A or beyond", 
    6, "Four out of five points in a row in Zone B or beyond", 
    7, "Fifteen points in a row in Zone C", 
    8, "Eight points in a row on both sides of the center line with none in Zone C" )));
```
**Example 2: Running a chart with spoken tests**

With the Coating.JMP data table open, submit the following script:

```
Control Chart(Alarm Script(Speak(match(
    QC_Test,1, "One point beyond Zone A",
    QC_Test,2, "Nine points in a row in zone C or beyond",
    QC_Test,5, "Two out of three points in a row in Zone A or beyond"))),
Sample Size( :Sample), Ksigma(3), Chart Col( :Weight,
Xbar(Test 1(1), Test 2(1), Test 5(1)), R));
```

You can have either of these scripts use any of the JSL alert commands such as `Speak`, `Write` or `Mail`.

**Note:** Under Windows, in order to have sound alerts you must install the Microsoft Text-to-Speech engine, which is included as an option with the JMP product installation.

---

**Saving and Retrieving Limits**

JMP can use previously established control limits for control charts:

- upper and lower control limits, and a center line value
- parameters for computing limits such as a mean and standard deviation.

The control limits or limit parameter values must be either in a JMP data table, referred to as the *Limits Table* or stored as a column property in the process column. When you specify the **Control Chart** command, you can retrieve the Limits Table with the **Get Limits** button on the Control Chart launch dialog.

The easiest way to create a Limits Table is to save results computed by the Control Chart platform. The **Save Limits** command in the popup menu for each control chart automatically saves limits from the sample values. The type of data saved in the table varies according to the type of control chart in the analysis window. You can also use values from any source and create your own Limits Table. All Limits Tables must have

- a column of special key words that identify each row
- a column for each of the variables whose values are the known standard parameters or limits. This column name must be the same as the corresponding process variable name in the data table to be analyzed by the Control Chart platform.

You can save limits in a new data table or as properties of the response column. When you save control limits using the **in New Table** command, the limit key words written to the table depend on the current chart types displayed. A list of limit key words and their associated control chart is shown in Table 11.3 “Limits Table Keys with Appropriate Charts and Meanings,” p. 249.
The data table shown next is the data table created when you Save Limits using the Clips1.jmp data analysis. The report window showed an Individual Measurement chart and a Moving Range chart, with a specified standard deviation of 0.2126. Note that there is a set containing a center line value and control limits for each chart. The rows with values _Mean, _LCL, and _UCL are for the individual measurement chart. The values _AvgR, _LCLR, and _UCLR are for the Moving Range chart. If you request these kinds of charts again using this Limits Table, the Control Chart platform identifies the appropriate limits from key words in the _LimitsKey column.

Note that values for _KSigma, _Alpha, and _Range Span can be specified in the Control Chart Launch dialog. JMP always looks at the values from the dialog first. Values specified in the dialog take precedence over those in an active Limits Table.

The Control Chart command ignores rows with unknown key words and rows marked with the excluded row state. Except for _Range Span, _KSigma, _Alpha, and _Sample Size, any needed values not specified are estimated from the data.

As an aid when referencing Table 11.3 “Limits Table Keys with Appropriate Charts and Meanings,” p. 249, the following list summarizes the kinds of charts available in the Control Chart platform:

Run charts

Variables charts are

- \( \bar{X} \)-chart (Mean)
- \( R \)-chart (range)
- \( S \)-chart (standard deviation)
- IM chart (individual measurement)
- MR chart (moving range)
- UWMA chart (uniformly weighted moving average)
- EWMA chart (exponentially weighted moving average)
- CUSUM chart (cumulative sum)
- Levey-Jennings chart (Mean)
Attribute charts are

- \( \bar{P} \)-chart (proportion of nonconforming or defective items in a subgroup sample)
- \( NP \)-chart (number of nonconforming or defective items in a subgroup sample)
- \( C \)-chart (number of nonconformities or defects in a subgroup sample)
- \( U \)-chart (number of nonconforming or defects per unit).

**Table 11.3 Limits Table Keys with Appropriate Charts and Meanings**

<table>
<thead>
<tr>
<th>Key Words</th>
<th>For Chars</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>_Sample Size</td>
<td>( \bar{X}, R, S, \bar{P}, NP, C, U, ) UWMA, EWMA, CUSUM</td>
<td>fixed sample size for control limits; set to missing if the sample size is not fixed. If specified in the Control Chart launch dialog, fixed sample size is displayed.</td>
</tr>
<tr>
<td>_Range Span</td>
<td>IM, MR</td>
<td>specifies the number ( 2 \leq n \leq 25 ) of consecutive values for computation of moving range</td>
</tr>
<tr>
<td>_Span</td>
<td>UWMA</td>
<td>specifies the number ( 2 \leq n \leq 25 ) of consecutive subsample means for computation of moving average</td>
</tr>
<tr>
<td>_Weight</td>
<td>EWMA</td>
<td>constant weight for computation of EWMA</td>
</tr>
<tr>
<td>_KSigma</td>
<td>All</td>
<td>multiples of the standard deviation of the statistics to calculate the control limits; set to missing if the limits are in terms of the alpha level</td>
</tr>
<tr>
<td>_Alpha</td>
<td>All</td>
<td>Type I error probability used to calculate the control limits; used if multiple of the standard deviation is not specified in the launch dialog or in the Limits Table</td>
</tr>
<tr>
<td>_Std Dev</td>
<td>( \bar{X}, R, S, IM, MR, ) UWMA, EWMA, CUSUM</td>
<td>known process standard deviation</td>
</tr>
<tr>
<td>_Mean</td>
<td>( \bar{X}, IM, UWMA, EWMA, CUSUM</td>
<td>known process mean</td>
</tr>
<tr>
<td>_U</td>
<td>( C, U )</td>
<td>known average number of nonconformities per unit</td>
</tr>
<tr>
<td>_P</td>
<td>( NP, P )</td>
<td>known value of average proportion nonconforming</td>
</tr>
</tbody>
</table>
Table 11.3 Limits Table Keys with Appropriate Charts and Meanings  (Continued)

<table>
<thead>
<tr>
<th>Key Words</th>
<th>For Chars</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>_LCL, _UCL</td>
<td>(\bar{X}), IM, (P), (NP), (C), (U)</td>
<td>lower and upper control limit for Mean Chart, Individual Measurement chart, or any attribute chart</td>
</tr>
<tr>
<td>_AvgR</td>
<td>(R), MR</td>
<td>average range or average moving range</td>
</tr>
<tr>
<td>_LCLR, _UCLR</td>
<td>(R), MR</td>
<td>lower control limit for (R) or MR chart</td>
</tr>
<tr>
<td>_AvgS, _LCLS, _UCLS</td>
<td>S-Chart</td>
<td>upper and lower control limits for (S)-chart</td>
</tr>
<tr>
<td>_Head Start</td>
<td>CUSUM</td>
<td>head start for one-sided scheme</td>
</tr>
<tr>
<td>_Two Sided, _Data Units</td>
<td>CUSUM</td>
<td>type of chart</td>
</tr>
<tr>
<td>_H, _K</td>
<td>CUSUM</td>
<td>alternative to alpha and beta; K is optional</td>
</tr>
<tr>
<td>_Delta, _Beta</td>
<td>CUSUM</td>
<td>Absolute value of the smallest shift to be detected as a multiple of the process standard deviation or standard error probability and are available only when _Alpha is specified.</td>
</tr>
<tr>
<td>_AvgR_PreMeans, _AvgR_PreStdDev</td>
<td>IM, MR</td>
<td>Mean, upper and lower control limits based on pre-summarized group means or standard deviations.</td>
</tr>
<tr>
<td>_LCLR_PreMeans, _LCLR_PreStdDev</td>
<td></td>
<td></td>
</tr>
<tr>
<td>_UCLR_PreMeans, _UCLR_PreStdDev</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 11.3 Limits Table Keys with Appropriate Charts and Meanings**

<table>
<thead>
<tr>
<th>Key Words</th>
<th>For Chars</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>_LCL, _UCL</td>
<td>(\bar{X}), IM, (P), (NP), (C), (U)</td>
<td>lower and upper control limit for Mean Chart, Individual Measurement chart, or any attribute chart</td>
</tr>
<tr>
<td>_AvgR</td>
<td>(R), MR</td>
<td>average range or average moving range</td>
</tr>
<tr>
<td>_LCLR, _UCLR</td>
<td>(R), MR</td>
<td>lower control limit for (R) or MR chart</td>
</tr>
<tr>
<td>_AvgS, _LCLS, _UCLS</td>
<td>S-Chart</td>
<td>upper and lower control limits for (S)-chart</td>
</tr>
<tr>
<td>_Head Start</td>
<td>CUSUM</td>
<td>head start for one-sided scheme</td>
</tr>
<tr>
<td>_Two Sided, _Data Units</td>
<td>CUSUM</td>
<td>type of chart</td>
</tr>
<tr>
<td>_H, _K</td>
<td>CUSUM</td>
<td>alternative to alpha and beta; K is optional</td>
</tr>
<tr>
<td>_Delta, _Beta</td>
<td>CUSUM</td>
<td>Absolute value of the smallest shift to be detected as a multiple of the process standard deviation or standard error probability and are available only when _Alpha is specified.</td>
</tr>
<tr>
<td>_AvgR_PreMeans, _AvgR_PreStdDev</td>
<td>IM, MR</td>
<td>Mean, upper and lower control limits based on pre-summarized group means or standard deviations.</td>
</tr>
</tbody>
</table>
Real-Time Data Capture

In JMP, real-time data streams are handled with a *DataFeed* object set up through JMP Scripting Language (JSL) scripts. The *DataFeed* object sets up a concurrent thread with a queue for input lines that can arrive in real time, but are processed during background events. You set up scripts to process the lines and push data on to data tables, or do whatever else is called for. Full details for writing scripts are in the *JMP Scripting Language Guide*.

**The Open Datafeed Command**

To create a DataFeed object, use the `Open DataFeed` function specifying details about the connection, in the form

```
feedname = Open DataFeed( options... );
```

For example, submit this to get records from com1 and just list them in the log.

```
feed = OpenDataFeed(
    Connect( Port("com1:"), Baud(9600), DataBits(7)),
    SetScript(print(feed<<getLine));
);
```

This command creates a scriptable object and starts up a thread to watch a communications port and collect lines. A reference to the object is returned, and you need to save this reference by assigning it to a global variable. The thread collects characters until it has a line. When it finishes a line, it appends it to the line queue and schedules an event to call the *On DataFeed* handler.

**Commands for Data Feed**

The scriptable *DataFeed* object responds to several messages. To send a message in JSL, use the `<<` operator, aimed at the name of the variable holding a reference to the object.

To give it a script or the name of a global holding a script:

```
feedName << Set Script(script or script name);
```

To test *DataFeed* scripts, you can send it lines from a script:

```
feedName << Queue Line (character expression);
```

For the *DataFeed* script to get a line from the queue, use this message:

```
feedName << GetLine;
```

To get a list of all the lines to empty the queue, use this:

```
lineListName = feedName << GetLines;
```

To close the *DataFeed*, including the small window:

```
feedName << Close;
```

To connect to a live data source:

```
feedName << Connect(port specification);
```
where the port specifications inside the `Connect` command are as follows. Each option takes only one argument, but they are shown below with the possible arguments separated by `|` with the default value shown first. The last three options take boolean values that specify which control characters are sent back and forth to the device indicating when it is ready to get data. Usually, at most, one of these three is used:

```plaintext
Port( "com1:" | "com2:" | "lpt1:" |...),
Baud( 9600 | 4800 | ...),
Data Bits( 8 | 7 ),
Parity( None | Odd | Even ),
Stop Bits( 1 | 0 | 2 ),
DTR_DSR( 0 | 1 ), // DataTerminalReady
RTS_CTS( 0 | 1 ), // RequestToSend/ClearToSend
XON_XOFF( 1 | 0 )
```

The `Port` specification is needed if you want to connect; otherwise, the object still works but is not connected to a data feed.

To disconnect from the live data source:

```plaintext
feedName << Disconnect;
```

To stop and later restart the processing of queued lines, either click the respective buttons, or submit the equivalent messages:

```plaintext
feedName << Stop;
feedName << Restart;
```

**Operation**

The script is specified with `Set Script`.

```plaintext
feedName << Set Script(myScript);
```

Here `myScript` is the global variable that you set up to contain the script to process the data feed. The script typically calls `Get DataFeed` to get a copy of the line, and does whatever it wants. Usually, it parses the line for data and adds it to some data table. In the example below, it expects to find a three-digit long number starting in column 11; if it does, it adds a row to the data table in the column called `thickness`.

```plaintext
myScript= Expr(
    line = feed<<Get Line;
    if (Length(line)>=14,
        x = Num(SubString(line,11,3));
        if (x!=.,
            CurrentDataTable()<<Add Row({thickness=x}));)
```

**Setting up a script to start a new data table**

Here is a sample script that sets up a new data table and starts a control chart based on the data feed:

```plaintext
// make a data table
dt = NewTable("Gap Width");
// make a new column and setup control chart properties
dc = dt<<NewColumn("gap",Numeric,
```
Chapter 11
Statistical Control Charts

Excluded, Hidden, and Deleted Samples

SetProperty("Control Limits",
{XBar(Avg(20),LCL(19.8),UCL(20.2))},
SetProperty("Sigma", 0.1));

// make the data feed
feed = OpenDatafeed();
feedScript = expr(
  line = feed<<get line;
  z = Num(line);
  Show(line,z); // if logging or debugging
  if (!IsMissing(z), dt<<AddRow({:gap = z}));
);
feed<<SetScript(feedScript);
// start the control chart
Control Chart(SampleSize(5),KSigma(3),ChartCol(gap,XBar,R));
// either start the feed:
// feed<<connect("com1:",Port("com1:"),Baud(9600));
// or test feed some data to see it work:
 For(i=1,i<20,i++,
  feed<<Queue Line(Char(20+RandomUniform()*1)));

Setting Up a Script in a Data Table

In order to further automate the production setting, you can put a script like the one above into a data table property called On Open, which is executed when the data table is opened. If you further marked the data table as a template style document, a new data table is created each time the template table is opened.

Excluded, Hidden, and Deleted Samples

The following table summarizes the effects of various conditions on samples and subgroups:

<table>
<thead>
<tr>
<th>Description</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>All rows of the sample are excluded before creating the chart.</td>
<td>Sample is not included in the calculation of the limits, but it appears on the graph.</td>
</tr>
<tr>
<td>Sample is excluded after creating the chart.</td>
<td>Sample is included in the calculation of the limits, and it appears in the graph. Nothing will change on the output by excluding a sample with the graph open.</td>
</tr>
<tr>
<td>Sample is hidden before creating the chart.</td>
<td>Sample is included in the calculation of the limits, but does not appear on the graph.</td>
</tr>
<tr>
<td>Sample is hidden after creating the chart.</td>
<td>Sample is included in the calculation of the limits, but does not appear on the graph. The sample marker will disappear from the graph, the sample label will still appear on the axis, but limits remain the same.</td>
</tr>
</tbody>
</table>
Excluded, Hidden, and Deleted Samples

Some additional notes:

- Hide operates only on the rowstate of the first observation in the sample. For example, if the second observation in the sample is hidden, while the first observation is not hidden, the sample will still appear on the chart.

- An exception to the exclude/hide rule: Tests for Special Causes can flag if a sample is excluded, but will not flag if a sample is hidden.

- Because of the specific rules in place (see Table 11.4), the control charts do not support the Automatic Recalc script.

Table 11.4 Excluded, Hidden, and Deleted Samples (Continued)

| All rows of the sample are both excluded and hidden before creating the chart. | Sample is not included in the calculation of the limits, and it does not appear on the graph. |
| All rows of the sample are both excluded and hidden after creating the chart. | Sample is included in the calculation of the limits, but does not appear on the graph. The sample marker will disappear from the graph, the sample label will still appear on the axis, but limits remain the same. |
| Data set is subsetted with Sample deleted before creating chart. | Sample is not included in the calculation of the limits, the axis will not include a value for the sample, and the sample marker does not appear on the graph. |
| Data set is subsetted with Sample deleted after creating chart. | Sample is not included in the calculation of the limits, and does not appear on the graph. The sample marker will disappear from the graph, the sample label will still be removed from the axis, the graph will shift, and the limits will change. |
Shewhart Control Charts

Variables and Attribute Control Charts

Control charts are a graphical and analytic tool for deciding whether a process is in a state of statistical control.

The concepts underlying the control chart are that the natural variability in any process can be quantified with a set of control limits and that variation exceeding these limits signals a special cause of variation. Out-of-control processes generally justify some intervention to fix a problem to bring the process back in control.

Shewhart control charts are broadly classified into control charts for variables and control charts for attributes. Moving average charts and cumulative sum (Cusum) charts are special kinds of control charts for variables.

The Control Chart platform in JMP implements a variety of control charts:

- Run Chart
- $\bar{X}$-, $R$-, and $S$-Charts
- Individual and Moving Range Charts
- $P$-, $NP$-, $C$-, and $U$-Charts
- UWMA and EWMA Charts
- CUSUM Charts
- Presummarize, Levey-Jennings, and Multivariate Control Charts
- Phase Control Charts for $\bar{X}$-, $R$-, $S$, $IR$-, $P$-, $NP$-, $C$-, $U$-, Presummarize, and Levey-Jennings Charts.

This platform is launched by the Control Chart command in the Graph menu, by the toolbar or JMP Starter, or through scripting.

One feature special to Control Charts, different from other platforms in JMP, is that they update dynamically as data is added or changed in the table.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shewhart Control Charts for Variables</td>
<td>257</td>
</tr>
<tr>
<td>XBar-, R-, and S- Charts</td>
<td>257</td>
</tr>
<tr>
<td>Run Charts</td>
<td>262</td>
</tr>
<tr>
<td>Individual Measurement Charts</td>
<td>262</td>
</tr>
<tr>
<td>Presummarize Charts</td>
<td>265</td>
</tr>
<tr>
<td>Moving Average Charts</td>
<td>267</td>
</tr>
<tr>
<td>Uniformly Weighted Moving Average (UWMA) Charts</td>
<td>267</td>
</tr>
<tr>
<td>Exponentially Weighted Moving Average (EWMA) Charts</td>
<td>269</td>
</tr>
<tr>
<td>Shewhart Control Charts for Attributes</td>
<td>271</td>
</tr>
<tr>
<td>P- and NP-Charts</td>
<td>272</td>
</tr>
<tr>
<td>U-Charts</td>
<td>274</td>
</tr>
<tr>
<td>C-Charts</td>
<td>275</td>
</tr>
<tr>
<td>Levey-Jennings Charts</td>
<td>277</td>
</tr>
<tr>
<td>Phases</td>
<td>277</td>
</tr>
<tr>
<td>Example</td>
<td>278</td>
</tr>
<tr>
<td>JSL Phase Level Limits</td>
<td>280</td>
</tr>
</tbody>
</table>
Shewhart Control Charts for Variables

Control charts for variables are classified according to the subgroup summary statistic plotted on the chart:

- $\bar{X}$-charts display subgroup means (averages)
- $R$-charts display subgroup ranges (maximum – minimum)
- $S$-charts display subgroup standard deviations
- Run charts display data as a connected series of points.

The IR selection gives two additional chart types:

- Individual Measurement charts display individual measurements
- Moving Range charts display moving ranges of two or more successive measurements.

XBar-, R-, and S- Charts

For quality characteristics measured on a continuous scale, a typical analysis shows both the process mean and its variability with a mean chart aligned above its corresponding $R$- or $S$-chart. Or, if you are charting individual measurements, the individual measurement chart shows above its corresponding moving range chart.

Example: $\bar{X}$- and R-Charts

The following example uses the Coating.jmp data in the Quality Control sample data folder (taken from the ASTM Manual on Presentation of Data and Control Chart Analysis). The quality characteristic of interest is the Weight column. A subgroup sample of four is chosen. An $\bar{X}$-chart and an $R$-chart for the process are shown in Figure 12.1.

To replicate this example,

- Choose the Graph > Control Chart > XBar command.
- Note the selected chart types of XBar and R.
- Specify Weight as the Process variable.
- Specify Sample as the Sample Label.
- Click OK.

Alternatively, you can also submit the following JSL for this example:

```
Control Chart(Sample Size( :Sample), KSigma(3), Chart Col( :Weight, XBar, R));
```

Sample six indicates that the process is not in statistical control. To check the sample values, click the sample six summary point on either control chart. The corresponding rows highlight in the data table.

**Note:** If an $S$ chart is chosen with the $\bar{X}$-chart, then the limits for the $\bar{X}$-chart are based on the standard deviation. Otherwise, the limits for the $\bar{X}$-chart are based on the range.
You can use **Fit Y by X** for an alternative visualization of the data. First, change the modeling type of **Sample** to Nominal. Specify the interval variable **Weight** as **Y, Response** and the nominal variable **Sample** as **X, Factor**. Select the **Quantiles** option from the Oneway Analysis drop-down menu. The box plots in Figure 12.2 show that the sixth sample has a small range of high values.
Control Limits for $\bar{X}$- and R-charts

JMP generates control limits for $\bar{X}$- and R-charts as follows.

LCL for $\bar{X}$ chart = $\bar{X}_{w} - \frac{k\hat{\sigma}}{n_{i}}$

UCL for $\bar{X}$ chart = $\bar{X}_{w} + \frac{k\hat{\sigma}}{\sqrt{n_{i}}}$

LCL for $R$-chart = max($d_{2}(n)\hat{\sigma} - kd_{3}(n)\hat{\sigma}$, 0)

UCL for $R$-chart = $d_{2}(n)\hat{\sigma} + kd_{3}(n)\hat{\sigma}$

Center line for $R$-chart: By default, the center line for the $i^{th}$ subgroup, where $k$ is the number of subgroups, indicates an estimate of the expected value of $R_{i}$, which is computed as $d_{2}(n)\hat{\sigma}$, where $\hat{\sigma}$ is an estimate of $\sigma$. If you specify a known value ($\sigma_{0}$) for $\sigma$, the central line indicates the value of $d_{2}(n)\sigma_{0}$. Note that the central line varies with $n_{i}$.

The standard deviation of an $\bar{X}$/R chart is estimated by

$$\hat{\sigma} = \frac{\sum_{i=1}^{N} R_{i}}{d_{2}(n_{1}) + \ldots + d_{2}(n_{N})}$$

where

$\bar{X}_{w}$ = weighted average of subgroup means

$\hat{\sigma}$ = process standard deviation

$n_{i}$ = sample size of $i^{th}$ subgroup

$d_{2}(n)$ is the expected value of the range of $n$ independent normally distributed variables with unit standard deviation

$d_{3}(n)$ is the standard error of the range of $n$ independent observations from a normal population with unit standard deviation

$N$ is the number of subgroups for which $n_{i} \geq 2$
Example: $\bar{X}$- and $S$-charts with Varying Subgroup Sizes

This example uses the same data as example 1, Coating.jmp, in the Quality Control sample data folder. This time the quality characteristic of interest is the Weight 2 column. An $\bar{X}$-chart and an $S$ chart for the process are shown in Figure 12.3.

To replicate this example,
- Choose the Graph > Control Chart > XBar command.
- Select the chart types of XBar and $S$.
- Specify Weight 2 as the Process variable.
- Specify the column, Sample as the Sample Label variable.
- The Sample Size option should automatically change to Sample Grouped by Sample Label.
- Click OK.

Alternatively, you can also submit the following JSL for this example:

```
Control Chart(Sample Size( :Sample), KSigma(3), Chart Col( :Weight 2, XBar, S));
```

Figure 12.3 $\bar{X}$ and $S$ charts for Varying Subgroup Sizes
Weight 2 has several missing values in the data, so you may notice the chart has uneven limits. Although, each sample has the same number of observations, samples 1, 3, 5, and 7 each have a missing value.

**Note:** Although they will turn on and appear checked, no zones or tests will appear on the chart until all samples are equally sized, as neither are valid on charts with unequally sized samples. If the samples change while the chart is open and they become equally sized, and the zone and/or test option is selected, the zones and/or tests will be applied immediately and appear on the chart.

### Control Limits for \( \bar{X} \)- and \( S \)-Charts

JMP generates control limits for \( \bar{X} \)- and \( S \)-charts as follows.

- **LCL for \( \bar{X} \) chart**
  \[
  \bar{X}_{W} - \frac{\hat{k}\sigma}{\sqrt{n_i}}
  \]

- **UCL for \( \bar{X} \) chart**
  \[
  \bar{X}_{W} + \frac{\hat{k}\sigma}{\sqrt{n_i}}
  \]

- **LCL for \( S \) chart**
  \[
  \max\left( c_4(n_i)\hat{\sigma} - k\hat{c}_4(n_i)\hat{\sigma}, 0 \right)
  \]

- **UCL for \( S \) chart**
  \[
  c_4(n_i)\hat{\sigma} + k\hat{c}_4(n_i)\hat{\sigma}
  \]

**Center line for \( S \)-chart:** By default, the center line for the \( i \)th subgroup, where \( k \) is equal to the number of subgroups, indicates an estimate of the expected value of \( s_i \), which is computed as \( c_4(n_i)\hat{\sigma} \), where \( \hat{\sigma} \) is an estimate of \( \sigma \). If you specify a known value \( (\sigma_0) \) for \( \sigma \), the central line indicates the value of \( c_4(n_i)\sigma_0 \). Note that the central line varies with \( n_i \).

The estimate for the standard deviation in an \( \bar{X} / S \) chart is

\[
\hat{\sigma} = \frac{\hat{s}_1}{c_4(n_1)} + \cdots + \frac{\hat{s}_n}{c_4(n_N)}
\]

where

- \( \bar{X}_{W} \) = weighted average of subgroup means
- \( \sigma \) = process standard deviation
- \( n_i \) = sample size of \( i \)th subgroup
- \( c_4(n) \) is the expected value of the standard deviation of \( n \) independent normally distributed variables with unit standard deviation.
Shewhart Control Charts

Shewhart Control Charts for Variables

$c_5(n)$ is the standard error of the standard deviation of $n$ independent observations from a normal population with unit standard deviation.

$N$ is the number of subgroups for which $n_i \geq 2$

$s_i$ is the sample standard deviation of the $i^{th}$ subgroup

Run Charts

Run charts display a column of data as a connected series of points. The following example is a Run chart for the Weight variable from Coating.jmp.

**Figure 12.4 Run Chart**

When you select the **Show Center Line** option in the Run Chart drop-down, a line is drawn through the center value of the column. The center line is determined by the **Use Median** setting of the platform drop-down. When **Use Median** is selected, the median is used as the center line. Otherwise, the mean is used. When saving limits to a file, both the overall mean and median are saved.

Run charts can also plot the group means when a sample label is given, either on the dialog or through a script.

Individual Measurement Charts

**Individual Measurement** Chart Type displays individual measurements. Individual Measurement charts are appropriate when only one measurement is available for each subgroup sample.

**Moving Range** Chart Type displays moving ranges of two or more successive measurements. Moving ranges are computed for the number of consecutive measurements you enter in the Range Span box. The default range span is 2. Because moving ranges are correlated, these charts should be interpreted with care.
Example: Individual Measurement and Moving Range Charts

The Pickles.jmp data in the Quality Control sample data folder contains the acid content for vats of pickles. Because the pickles are sensitive to acidity and produced in large vats, high acidity ruins an entire pickle vat. The acidity in four vats is measured each day at 1, 2, and 3 PM. The data table records day, time, and acidity measurements. The dialog in Figure 12.5 creates Individual Measurement and Moving Range charts with date labels on the horizontal axis.

To complete this example,

- Choose the Graph > Control Chart > IR command.
- Select both Individual Measurement and Moving Range chart types.
- Specify Acid as the Process variable.
- Specify Date as the Sample Label variable.
- Click OK.

Alternatively, you can also submit the following JSL for this example:

```jsl
Control Chart(Sample Label( :Date), GroupSize(1), KSigma(3), Chart Col( :Acid, Individual Measurement, Moving Range));
```

The individual measurement and moving range charts shown in Figure 12.6 monitor the acidity in each vat produced.
**Note:** A Median Moving Range chart can also be evaluated. If you choose a Median Moving Range chart and an Individual Measurement chart, the limits on the Individual Measurement chart use the Median Moving Range as the sigma, rather than the Average Moving Range.

**Figure 12.6** Individual Measurement and Moving Range Charts for Pickles Data

**Control Limits for Individual Measurement, Moving Range, and Median Moving Range Charts**

LCL for Individual Measurement Chart = $\bar{X} - k\hat{\sigma}$

UCL for Individual Measurement Chart = $\bar{X} + k\hat{\sigma}$

LCL for Moving Range Chart = $\max\left(\frac{d_2(n)\hat{\sigma}}{\hat{d}_3(n)} - k\frac{d_3(n)\hat{\sigma}}{\hat{d}_4(n)}, 0\right)$

UCL for Moving Range Chart = $\frac{d_2(n)\hat{\sigma}}{\hat{d}_4(n)} + k\frac{d_3(n)\hat{\sigma}}{\hat{d}_4(n)}$

LCL for Median Moving Range Chart = $\max(0, \text{Avg} - (k\text{Std Dev}\cdot d_3(n)))$

UCL for Median Moving Range Chart = $\text{Avg} + (k\text{Std Dev}\cdot d_3(n))$
The standard deviation for Individual Measurement and Moving Range charts is estimated by
\[
\sigma = \frac{R}{d_2(n)}
\]
and the standard deviation for Median Moving Range charts is estimated by
\[
\text{Std Dev} = \frac{\text{MMR}}{d_4(n)}
\]
where
\[
\bar{X} = \text{the mean of the individual measurements}
\]
\[
\bar{R} = \text{the mean of the non-missing moving ranges computed as } [(R_1 + R_2 + \ldots + R_N)/(N+1-n)]
\]
\[
\sigma = \text{the process standard deviation}
\]
\[
k = \text{the number of standard deviations}
\]
\[
\text{MMR} = \text{Center Line (Avg) for Median Moving Range chart}
\]
\[
d_2(n) = \text{expected value of the range of } n \text{ independent normally distributed variables with unit standard deviation}
\]
\[
d_3(n) = \text{standard error of the range of } n \text{ independent observations from a normal population with unit standard deviation.}
\]
\[
d_4(n) = \text{expected value of the range of a normally distributed sample of size nsample.}
\]

**Presummarize Charts**

If your data consist of repeated measurements of the same process unit, then you will want to combine these into one measurement for the unit. Pre-summarizing is not recommended unless the data have repeated measurements on each process or measurement unit.

**Presummarize** summarizes the process column into sample means and/or standard deviations, based either on the sample size or sample label chosen. Then it charts the summarized data based on the options chosen in the launch dialog. Optionally, you can append a capability analysis by checking the appropriate box in the launch dialog.

**Example: Presummarize Chart**

For an example, using the Coating.jmp data table,

- Choose the **Graph > Control Chart > Presummarize** command.
- Choose **Weight** as the Process variable and **Sample** as the Sample Label.
- In the dialog check both **Individual on Group Means** and **Moving Range on Group Means**. The **Sample Grouped by Sample Label** button is automatically selected when you choose a Sample Label variable.
- Click **OK**.
Figure 12.7 Presummarize Dialog

Figure 12.8 Example of Charting Presummarized Data
Although the points for $\bar{X}$- and $S$-charts are the same as the Individual on Group Means and Individual on Group Std Devs charts, the limits are different because they are computed as Individual charts.

Another way to generate the presummarized charts, with the Coating.jmp data table,

- Choose **Tables > Summary**.
- Assign **Sample** as the Group variable, then **Mean(Weight)** and **Std Dev(Weight)** as **Statistics**.
- Click **OK**.
- Select **Graph > Control Chart > IR**.
- Select **Mean(Weight)** and **Std Dev(Weight)** as Process variables.
- Click **OK**.

The resulting charts match the presummarized charts.

When using Presummarize charts, you can select either **On Group Means** or **On Group Std Devs** or both. Each option will create two charts (an Individual Measurement, also known as an X chart, and a Moving Range chart) if both IR chart types are selected.

The **On Group Means** options compute each sample mean, then plot the means and create an Individual Measurement and a Moving Range chart on the means.

The **On Group Std Devs** options compute each sample standard deviation, then plot the standard deviations as individual points and create an Individual Measurement and a Moving Range chart on the standard deviations.

**Moving Average Charts**

The control charts previously discussed plot each point based on information from a single subgroup sample. The Moving Average chart is different from other types because each point combines information from the current sample and from past samples. As a result, the Moving Average chart is more sensitive to small shifts in the process average. On the other hand, it is more difficult to interpret patterns of points on a Moving Average chart because consecutive moving averages can be highly correlated (Nelson 1983).

In a Moving Average chart, the quantities that are averaged can be individual observations instead of subgroup means. However, a Moving Average chart for individual measurements is not the same as a control (Shewhart) chart for individual measurements or moving ranges with individual measurements plotted.

**Uniformly Weighted Moving Average (UWMA) Charts**

Each point on a Uniformly Weighted Moving Average (UWMA) chart, also called a Moving Average chart, is the average of the $w$ most recent subgroup means, including the present subgroup mean. When you obtain a new subgroup sample, the next moving average is computed by dropping the oldest of the previous $w$ subgroup means and including the newest subgroup mean. The constant, $w$, is called the span of the moving average, and indicates how many subgroups to include to form the moving average. The larger the span ($w$), the smoother the UWMA line, and the less it reflects the magnitude of shifts. This means that larger values of $w$ guard against smaller shifts.
Example: UWMA Charts

As an example, consider Clips1.jmp. The measure of interest is the gap between the ends of manufactured metal clips. To monitor the process for a change in average gap, subgroup samples of five clips are selected daily and a UWMA chart with a moving average span of three is examined. To see this chart, complete the Control Chart launch dialog as shown in Figure 12.9, submit the JSL or follow the steps below.

```
Control Chart(Sample Size(5), KSigma(3), Moving Average Span(3), Chart Col(:Gap, UWMA));
```

- Choose the Graph > Control Chart > UWMA command.
- Change the Moving Average Span to 3.
- Choose Gap as the Process variable.
- Click OK.

![Figure 12.9 Specification for UWMA Charts of Clips1.jmp Data](image)

The result is the chart in Figure 12.10. The point for the first day is the mean of the five subgroup sample values for that day. The plotted point for the second day is the average of subgroup sample means for the first and second days. The points for the remaining days are the average of subsample means for each day and the two previous days.

The average clip gap appears to be decreasing, but no sample point falls outside the 3σ limits.
Control Limits for UWMA Charts

Control limits for UWMA charts are computed as follows. For each subgroup $i$,

$$
LCL_i = \bar{X}_w - k \frac{\hat{\sigma}}{\min(i, w)} \sqrt{\frac{1 + \frac{1}{n_{i-1}} + \cdots + \frac{1}{n_{\max(i-w, 0)}}}{\frac{1}{n_i} + \frac{1}{n_{i-1}} + \cdots + \frac{1}{n_{\max(i-w, 0)}}}}
$$

$$
UCL_i = \bar{X}_w + k \frac{\hat{\sigma}}{\min(i, w)} \sqrt{\frac{1 + \frac{1}{n_{i-1}} + \cdots + \frac{1}{n_{\max(i-w, 0)}}}{\frac{1}{n_i} + \frac{1}{n_{i-1}} + \cdots + \frac{1}{n_{\max(i-w, 0)}}}}
$$

where

- $w$ is the span parameter (number of terms in moving average)
- $n_i$ is the sample size of the $i^{th}$ subgroup
- $k$ is the number of standard deviations
- $\bar{X}_w$ is the weighted average of subgroup means
- $\hat{\sigma}$ is the process standard deviation

Exponentially Weighted Moving Average (EWMA) Charts

Each point on an Exponentially Weighted Moving Average (EWMA) chart, also referred to as a Geometric Moving Average (GMA) chart, is the weighted average of all the previous subgroup means, including the mean of the present subgroup sample. The weights decrease exponentially going backward in time. The weight ($0 < \text{weight} \leq 1$) assigned to the present subgroup sample mean is a parameter of the EWMA chart. Small values of weight are used to guard against small shifts.
Example: EWMA Charts

Using the Clips1.jmp data table, submit the JSL or follow the steps below.

```
Control Chart(Sample Size(5), KSigma(3), Weight(0.5), Chart Col( :Gap, EWMA));
```
- Choose the Graph > Control Chart > EWMA command.
- Change the Weight to 0.5.
- Choose Gap as the Process variable.
- Leave the Sample Size Constant as 5.
- Click OK.

The figure below shows the EWMA chart for the same data seen in Figure 12.10. This EWMA chart was generated for weight = 0.5.

![EWMA Chart](image)

Control Limits for EWMA Charts

Control limits for EWMA charts are computed as follows.

\[
LCL = \overline{X}_w - k\hat{\sigma}_r \sqrt{\frac{\sum_{j=0}^{i-1} (1-r)^{2j}}{n_{i-j}}}
\]

\[
UCL = \overline{X}_w + k\hat{\sigma}_r \sqrt{\frac{\sum_{j=0}^{i-1} (1-r)^{2j}}{n_{i-j}}}
\]

where

- \( r \) is the EWMA weight parameter \( (0 < r \leq 1) \)
- \( x_{ij} \) is the \( j \)th measurement in the \( i \)th subgroup, with \( j = 1, 2, 3, ..., n_i \)
Chapter 12

Shewhart Control Charts for Attributes

In the previous types of charts, measurement data was the process variable. This data is often continuous, and the charts are based on theory for continuous data. Another type of data is count data, where the variable of interest is a discrete count of the number of defects or blemishes per subgroup. For discrete count data, attribute charts are applicable, as they are based on binomial and poisson models. Since the counts are measured per subgroup, it is important when comparing charts to determine whether you have a similar number of items in the subgroups between the charts. Attribute charts, like variables charts, are classified according to the subgroup sample statistic plotted on the chart:

**Table 12.1 Determining which Attribute Chart to use**

<table>
<thead>
<tr>
<th>Each item is judged as either conforming or non-conforming</th>
<th>For each item, the number of defects is counted</th>
</tr>
</thead>
<tbody>
<tr>
<td>The subgroups are a constant size</td>
<td>The subgroups vary in size</td>
</tr>
<tr>
<td>The subgroups are a constant size</td>
<td>The subgroups vary in size</td>
</tr>
<tr>
<td><strong>NP-chart</strong></td>
<td><strong>P-chart</strong></td>
</tr>
<tr>
<td><strong>C-chart</strong></td>
<td><strong>U-chart</strong></td>
</tr>
</tbody>
</table>

- $P$-charts display the proportion of nonconforming (defective) items in subgroup samples which can vary in size. Since each subgroup for a $P$-chart consists of $N$ items, and an item is judged as either conforming or nonconforming, the maximum number of nonconforming items in a subgroup is $N$.
- $NP$-charts display the number of nonconforming (defective) items in constant sized subgroup samples. Since each subgroup for a $NP$-chart consists of $N_i$ items, and an item is judged as either conforming or nonconforming, the maximum number of nonconforming items in subgroup $i$ is $N_i$.
- $C$-charts display the number of nonconformities (defects) in a subgroup sample that usually consists of one inspection unit.
- $U$-charts display the number of nonconformities (defects) per unit in subgroup samples that can have a varying number of inspection units.

**Note:** If adding a column property for sigma in any of or all of the attribute control charts, the value needs to be equal to the proportion since the column property is not control-chart specific.

**Note:** For attribute charts, specify the defect count or defective proportion as the Process variable. The data will be interpreted as counts, unless it contains non-integer values between 0 and 1.
**P- and NP-Charts**

**Example: NP-Charts**

The Washers.jmp data in the Quality Control sample data folder contains defect counts of 15 lots of 400 galvanized washers. The washers were inspected for finish defects such as rough galvanization and exposed steel. If a washer contained a finish defect, it was deemed nonconforming or defective. Thus, the defect count represents how many washers were defective for each lot of size 400. To replicate this example, follow these steps or submit the JSL script below:

- Choose the **Graph > Control Chart > NP** command.
- Choose **# defective** as the **Process** variable.
- Change the **Constant Size** to 400.
- Click OK.

```julia
Control Chart(Sample Size(400), KSigma(3), Chart Col( :Name("# defective"), NP));
```

The example here illustrates an **NP-chart** for the number of defects.

---

**Example: P-Charts**

Again, using the Washers.jmp data, we can specify a sample size variable, which would allow for varying sample sizes.

**Note:** This data contains all constant sample sizes. Follow these steps or submit the JSL script below:

- Choose the **Graph > Control Chart > P** command.
- Choose **Lot** as the **Sample Label** variable.
- Choose **# defective** as the **Process** variable.
Choose Lot Size as the Sample Size variable.

Click OK.

Control Chart(Sample Label( :Lot), Sample Size( :Lot Size), K Sigma(3), Chart Col(Name("# defective"), P))

The chart shown here illustrates a P-chart for the proportion of defects.

Figure 12.13 P-Chart

Note that although the points on the chart look the same as the NP-chart, the y-axis, Avg and limits are all different since they are now based on proportions.

Control Limits for P- and NP- Charts

The lower and upper control limits, LCL and UCL, respectively, are computed as follows.

\[
P\text{-chart LCL} = \max(\hat{p} - k \sqrt{\hat{p}(1 - \hat{p})/n_i}, 0) \\
P\text{-chart UCL} = \min(\hat{p} + k \sqrt{\hat{p}(1 - \hat{p})/n_i}, 1) \\
NP\text{-chart LCL} = \max(n_i \hat{p} - k \sqrt{n_i \hat{p}(1 - \hat{p})}, 0) \\
NP\text{-chart UCL} = \min(n_i \hat{p} + k \sqrt{n_i \hat{p}(1 - \hat{p})}, n_i)
\]

where

\[
\hat{p} = \frac{n_1 p_1 + \ldots + n_N p_N}{n_1 + \ldots + n_N} = \frac{X_1 + \ldots + X_N}{n_1 + \ldots + n_N}
\]

\(n_i\) is the number of items in the \(i^{th}\) subgroup

\(k\) is the number of standard deviations
Shewhart Control Charts
Shewhart Control Charts for Attributes

U-Charts

The Braces.jmp data in the Quality Control sample data folder records the defect count in boxes of automobile support braces. A box of braces is one inspection unit. The number of boxes inspected (per day) is the subgroup sample size, which can vary. The $U$-chart, shown here, is monitoring the number of brace defects per subgroup sample size. The upper and lower bounds vary according to the number of units inspected.

Note: When you generate a $U$-chart, and select Capability, JMP launches the Poisson Fit in Distribution and gives a Poisson-specific capability analysis.

Figure 12.14 $U$-Chart

Example: U-Charts

To replicate this example, follow these steps or submit the JSL below.

- Open the Braces.jmp data in the Quality Control sample data folder.
- Choose the Graph > Control Chart > U command.
- Choose # defects as the Process variable.
- Choose Unit size as the Unit Size variable.
- Choose Date as the Sample Label.
- Click OK.

Control Chart(Sample Label( :Date), Unit Size( :Unit size), K Sigma(3), Chart Col( :Name("# defects"), U));
Control Limits on U-charts

The lower and upper control limits, LCL and UCL, are computed as follows

\[
\begin{align*}
  \text{LCL} &= \max(\bar{u} - k\sqrt{\bar{u}/n_i}, 0) \\
  \text{UCL} &= \bar{u} + k\sqrt{\bar{u}/n_i}
\end{align*}
\]

The limits vary with \( n_i \).

- \( u \) is the expected number of nonconformities per unit produced by process
- \( u_i \) is the number of nonconformities per unit in the \( i \)th subgroup. In general, \( u_i = c_i/n_i \)
- \( c_i \) is the total number of nonconformities in the \( i \)th subgroup
- \( n_i \) is the number of inspection units in the \( i \)th subgroup
- \( \bar{u} \) is the average number of nonconformities per unit taken across subgroups. The quantity \( \bar{u} \) is computed as a weighted average

\[
\bar{u} = \frac{n_1u_1 + \ldots + n_Nu_N}{n_1 + \ldots + n_N} = \frac{c_1 + \ldots + c_N}{n_1 + \ldots + n_N}
\]

\( N \) is the number of subgroups

C-Charts

C-charts are similar to U-charts in that they monitor the number of nonconformities in an entire subgroup, made up of one or more units. However, they require constant subgroup sizes. C-charts can also be used to monitor the average number of defects per inspection unit.

**Note:** When you generate a C-chart, and select Capability, JMP launches the Poisson Fit in Distribution and gives a Poisson-specific capability analysis.

Example: C-Charts

In this example, a clothing manufacturer ships shirts in boxes of ten. Prior to shipment, each shirt is inspected for flaws. Since the manufacturer is interested in the average number of flaws per shirt, the number of flaws found in each box is divided by ten and then recorded. To replicate this example, follow these steps or submit the JSL below.

- Open the Shirts.jmp data in the Quality Control sample data folder.
- Choose the Graph > Control Chart > C command.
- Choose # Defects as the Process variable.
- Choose Box Size as the Sample Size.
- Choose Box as the Sample Label.
Control Limits on C-Charts

The lower and upper control limits, LCL and UCL, are computed as follows.

\[
\begin{align*}
LCL &= \max(n_i \bar{u} - k \sqrt{n_i \bar{u}}, 0) \\
UCL &= n_i \bar{u} + k \sqrt{n_i \bar{u}}
\end{align*}
\]

The limits vary with \(n_i\).

\(u\) is the expected number of nonconformities per unit produced by process

\(u_i\) is the number of nonconformities per unit in the \(i^{th}\) subgroup. In general, \(u_i = c_i / n_i\)

\(c_i\) is the total number of nonconformities in the \(i^{th}\) subgroup

\(n_i\) is the number of inspection units in the \(i^{th}\) subgroup

\(\bar{u}\) is the average number of nonconformities per unit taken across subgroups. The quantity \(\bar{u}\) is computed as a weighted average

\[
\bar{u} = \frac{n_1 u_1 + \ldots + n_N u_N}{n_1 + \ldots + n_N} = \frac{c_1 + \ldots + c_N}{n_1 + \ldots + n_N}
\]

\(N\) is the number of subgroups
Levey-Jennings Charts

Levey-Jennings charts show a process mean with control limits based on a long-term sigma. The control limits are placed at $3\sigma$ distance from the center line.

The standard deviation, $s$, for the Levey-Jennings chart is calculated the same way standard deviation is in the Distribution platform, i.e.

$$s = \sqrt{\frac{\sum_{i=1}^{N} \left( y_i - \bar{y} \right)^2}{N-1}}$$

**Figure 12.16** Levey Jennings Chart

Phases

A *phase* is a group of consecutive observations in the data table. For example, phases might correspond to time periods during which a new process is brought into production and then put through successive changes. Phases generate, for each level of the specified Phase variable, a new sigma, set of limits, zones, and resulting tests.

On the dialog for $\bar{X}$-, $R$-, $IR$-, $P$-, $NP$-, $C$-, $U$-, Presummarize, and Levey-Jennings charts, a *Phase* variable button appears. If a phase variable is specified, the phase variable is examined, row by row, to identify to which phase each row belongs.

Saving to a limits file reveals the sigma and specific limits calculated for each phase.
Example

Open Diameter.JMP, found in the Quality Control sample data folder. This data set contains the diameters taken for each day, both with the first prototype (phase 1) and the second prototype (phase 2).

- Select Graph > Control Chart > XBar.
- Choose DIAMETER as the Process, DAY as the Sample Label, and Phase as the Phase.
- Select S and XBar.
- Click OK.

Figure 12.17  Launch Dialog for Phases

The resulting chart has different limits for each phase.
Figure 12.18 Phase Control Chart

Note: The sigma was calculated using the standard deviation.
JSL Phase Level Limits

The JSL syntax for setting the phase level limits in control charts is specific. The following example illustrates setting the limits for the different phases of Diameter.jmp:

```julia
Control Chart(
    Phase( :Phase ),
    Sample Size( :DAY ),
    KSigma(3),
    Chart Col( 
        :DIAMETER,
        XBar( 
            Phase Level("1", Sigma(0.29), Avg(4.3), LCL(3.99), UCL(4.72)),
            Phase Level("2", Sigma(0.21), Avg(4.29), LCL(4), UCL(4.5))),
        R( 
            Phase Level("1"),
            Phase Level("2"))
    )
);
```
Cusum control charts are used when it is important to detect that a process has wandered away from a specified process mean. Although Shewhart $\bar{X}$-charts can detect if a process is moving beyond a two- or three-sigma shift, they are not effective at spotting a one-sigma shift in the mean. They will still appear in control because the cumulative sum of the deviations will wander farther and farther away from the specified target, and a small shift in the mean will appear very clearly and much sooner.

**Figure 13.1  Example of a Cumulative Sum Chart**
Contents

Cumulative Sum (Cusum) Charts ................................................................. 283
Launch Options for Cusum Charts ......................................................... 284
Cusum Chart Options ........................................................................... 286
Formulas for CUSUM Charts ................................................................. 290
Notation ................................................................................................. 290
One-Sided CUSUM Charts ................................................................. 290
Two-Sided Cusum Schemes ............................................................... 291
Cumulative Sum (Cusum) Charts

Cumulative Sum (Cusum) charts display cumulative sums of subgroup or individual measurements from a target value. Cusum charts are graphical and analytical tools for deciding whether a process is in a state of statistical control and for detecting a shift in the process mean.

JMP cusum charts can be one-sided, which detect a shift in one direction from a specified target mean, or two-sided to detect a shift in either direction. Both charts can be specified in terms of geometric parameters ($h$ and $k$ described in Figure 13.2); two-sided charts allow specification in terms of error probabilities $\alpha$ and $\beta$.

To interpret a two-sided Cusum chart, you compare the points with limits that compose a V-mask. A V-mask is formed by plotting V-shaped limits. The origin of a V-mask is the most recently plotted point, and the arms extended backward on the x-axis, as in Figure 13.2. As data are collected, the cumulative sum sequence is updated and the origin is relocated at the newest point.

**Figure 13.2 Illustration of a V-Mask for a Two-Sided Cusum Chart**

Shifts in the process mean are visually easy to detect on a cusum chart because they produce a change in the slope of the plotted points. The point where the slope changes is the point where the shift occurs. A condition is out-of-control if one or more of the points previously plotted crosses the upper or lower arm of the V-mask. Points crossing the lower arm signal an increasing process mean, and points crossing the upper arm signal a downward shift.

There are major differences between cusum charts and other control (Shewhart) charts:

- A Shewhart control chart plots points based on information from a single subgroup sample. In cusum charts, each point is based on information from all samples taken up to and including the current subgroup.
On a Shewhart control chart, horizontal control limits define whether a point signals an out-of-control condition. On a cusum chart, the limits can be either in the form of a V-mask or a horizontal decision interval.

The control limits on a Shewhart control chart are commonly specified as $3\sigma$ limits. On a cusum chart, the limits are determined from average run length, from error probabilities, or from an economic design.

A cusum chart is more efficient for detecting small shifts in the process mean. Lucas (1976) comments that a V-mask detects a $1\sigma$ shift about four times as fast as a Shewhart control chart.

**Launch Options for Cusum Charts**

When you choose Graph > Control Chart > Cusum, the Control Chart launch dialog appears, including appropriate options and specifications as shown here.

**Figure 13.3 Cusum Chart Launch Options**

See “Parameters,” p. 234 in the “Statistical Control Charts” chapter for a description of $K\text{Sigma}$ and Alpha. The following items pertain only to cusum charts:

**Two Sided**

Requests a two-sided cusum scheme when checked. If it is not checked, a one-sided scheme is used and no V-mask appears. If an $H$ value is specified, a decision interval is displayed.

**Data Units**

Specifies that the cumulative sums be computed without standardizing the subgroup means or individual values so that the vertical axis of the cusum chart is scaled in the same units as the data.

**Note:** Data Units requires that the subgroup sample size be designated as constant.

**Beta**

Specifies the probability of failing to discover that the specified shift occurred. Beta is the probability of a Type II error and is available only when you specify Alpha.
Chapter 13
Cumulative Sum Control Charts

Chapter 13
Cumulative Sum Control Charts

Cumulative Sum (Cusum) Charts

H

H is the vertical distance between the origin for the V-mask and the upper or lower arm of the V-mask for a two-sided scheme. When you click H, the Beta entry box is labeled K. You also enter a value for the increase in the lower V-mask per unit change on the subgroup axis (see Figure 13.2). For a one-sided scheme, H is the decision interval. Choose H as a multiple of the standard error.

Specify Stats

Specify Stats appends the panel shown here to the Control Charts launch dialog, which lets you enter the process variable specifications.

Figure 13.4 Specify Process Variables

Target is the target mean (goal) for the process or population. The target mean must be scaled in the same units as the data.

Delta specifies the absolute value of the smallest shift to be detected as a multiple of the process standard deviation or of the standard error, depending on whether the shift is viewed as a shift in the population mean or as a shift in the sampling distribution of the subgroup mean, respectively. Delta is an alternative to the Shift option (described next). The relationship between Shift and Delta is given by

\[ \delta = \frac{\Delta}{(\sigma / (\sqrt{n}))} \]

where \( \delta \) represents Delta, \( \Delta \) represents the shift, \( \sigma \) represents the process standard deviation, and \( n \) is the (common) subgroup sample size.

Shift is the minimum value you want to detect on either side of the target mean. You enter the shift value in the same units as the data, and you interpret it as a shift in the mean of the sampling distribution of the subgroup mean. You can choose either Shift or Delta.

Sigma specifies a known standard deviation, \( \sigma_0 \), for the process standard deviation, \( \sigma \). By default, the Control Chart platform estimates sigma from the data. You can use Sigma instead of the Alpha option on the Control Charts launch dialog.

Head Start specifies an initial value for the cumulative sum, \( S_0 \), for a one-sided cusum scheme (\( S_0 \) is usually zero). Enter Head Start as a multiple of standard error.
Cusum Chart Options

Cusum charts have these options (in addition to standard chart options):

- **Show Points** shows or hides the sample data points.
- **Connect Points** connects the sample points with a line.
- **Mask Color** displays the JMP color palette for you to select a line color for the V-mask.
- **Connect Color** displays the JMP color palette for you to select a color for the connect line when the **Connect Points** option is in effect.
- **Center Line Color** displays the JMP color palette for you to select a color for the center line.
- **Show Shift** shows or hides the shift you entered, or center line.
- **Show V Mask** shows or hides the V-mask based on the parameters (statistics) specified in the Cusum Control Charts launch window.
- **Show Parameters** displays a Parameters table (see Figure 13.9) that summarizes the Cusum charting parameters.
- **Show ARL** displays the average run length (ARL) information.

**Example 1. Two-Sided Cusum Chart with V-mask**

To see an example of a two-sided cusum chart, open the Oil1 Cusum.jmp file from the Quality Control sample data folder. A machine fills 8-ounce cans of two-cycle engine oil additive. The filling process is believed to be in statistical control. The process is set so that the average weight of a filled can, $\mu_0$, is 8.10 ounces. Previous analysis shows that the standard deviation of fill weights, $\sigma_0$, is 0.05 ounces.

Subgroup samples of four cans are selected and weighed every hour for twelve hours. Each observation in the Oil1 Cusum.jmp data table contains one value of weight along with its associated value of hour. The observations are sorted so that the values of hour are in increasing order. The Control Chart platform assumes that the data are sorted in increasing order.

A two-sided cusum chart is used to detect shifts of at least one standard deviation in either direction from the target mean of 8.10 ounces.

To create a Cusum chart for this example,

- Choose the **Graph > Control Chart > CUSUM** command.
- Click the **Two Sided** check box if it is not already checked.
- Specify **weight** as the Process variable.
- Specify **hour** as the Sample Label.
- Click the **H** radio button and enter 2 into the text box.
- Click **Specify Stats** to open the Known Statistics for CUSUM Chart tab.
- Set **Target** to the average weight of 8.1.
- Enter a **Delta** value of 1.
- Set **Sigma** to the standard deviation of 0.05.
The finished dialog should look like the one in Figure 13.5. Alternatively, you can bypass the dialog and submit the following JSL script:

```
Control Chart(Sample Size( :hour), H(2), Chart Col( :weight, CUSUM(Two sided(1), Target(8.1), Delta(1), Sigma(0.05)))));
```

**Figure 13.5** Dialog for Cusum Chart Example

When you click **OK**, the chart in Figure 13.6 appears.

**Figure 13.6** Cusum Chart for Oil1 Cusum.jmp Data
Cumulative Sum (Cusum) Charts

You can interpret the chart by comparing the points with the V-mask whose right edge is centered at the most recent point (\( \text{hour} = 12 \)). Because none of the points cross the arms of the V-mask, there is no evidence that a shift in the process has occurred.

A shift or out-of-control condition is signaled at a time \( t \) if one or more of the points plotted up to the time \( t \) cross an arm of the V-mask. An upward shift is signaled by points crossing the lower arm, and a downward shift is signaled by points crossing the upper arm. The time at which the shift occurred corresponds to the time at which a distinct change is observed in the slope of the plotted points.

The cusum chart automatically updates when you add new samples. The Cusum chart in Figure 13.7 is the previous chart with additional points. You can move the origin of the V-mask by using the grabber tool to click a point. The center line and V-mask adjust to reflect the process condition at that point.

Figure 13.7 Updated Cusum Chart for the Oil1 Cusum.jmp Data

---

Example 2. One-Sided Cusum Chart with No V-mask

Consider the data used in Example 1, where the machine fills 8-ounce cans of engine oil. Consider also that the manufacturer is now concerned about significant over-filling in order to cut costs, and not so concerned about under-filling. A one-sided Cusum Chart can be used to identify data approaching or exceeding the side of interest. Anything 0.25 ounces beyond the mean of 8.1 is considered a problem. To do this example,

- Choose the Graph > Control Chart > CUSUM command.
- Deselect the Two Sided check box.
- Specify weight as the Process variable.
- Specify hour as the Sample Label.
- Click the H radio button and enter 0.25 into the text box.
- Click Specify Stats to open the Known Statistics for CUSUM Chart tab.
- Set Target to the average weight of 8.1.
• Enter a Delta value of 1.
• Set Sigma to the standard deviation 0.05.

Alternatively, you can submit the following JSL script:

```
Control Chart(Sample Size( :hour), H(0.25), Show Limits Legend(0), Chart Col(:weight, CUSUM(Two Sided(0), Target(8.1), Delta(1), Sigma(0.05))));
```

The resulting output should look like the picture in Figure 13.8.

**Figure 13.8** One-Sided Cusum Chart for the Oil1 Cusum.jmp Data

Notice that the decision interval or horizontal line is set at the H-value entered (0.25). Also note that no V-mask appears with One-Sided Cusum charts.

The Show Parameters option in the Cusum chart popup menu shows the Parameters report in Figure 13.9. The parameters report summarizes the charting parameters from the Known Statistics for CUSUM Chart tab on the Control Chart launch dialog. An additional chart option, Show ARL, adds the average run length (ARL) information to the report. The average run length is the expected number of samples taken before an out-of-control condition is signaled:

• ARL(Delta), sometimes denoted ARL1, is the average run length for detecting a shift the size of the specified Delta
• ARL(0), sometimes denoted ARL0, is the in-control average run length for the specified parameters (Montgomery (1985)).

**Figure 13.9** Show Parameters and Show ARL Options
Formulas for CUSUM Charts

Notation

The following notation is used in these formulas:

- $\mu$ denotes the mean of the population, also referred to as the process mean or the process level.
- $\mu_0$ denotes the target mean (goal) for the population. Sometimes, the symbol $\bar{X}_0$ is used for $\mu_0$. See American Society for Quality Statistics Division (2004). You can provide $\mu_0$ as the sigma on the known statistics dialog or $r$ through JSL.
- $\sigma$ denotes the population standard deviation.
- $\sigma_0$ denotes a known standard deviation. You can provide $\sigma_0$ as the sigma on the known statistics dialog or through JSL.
- $\hat{\sigma}$ denotes an estimate of $\sigma$.
- $n$ denotes the nominal sample size for the cusum scheme.
- $\delta$ denotes the shift in $\mu$ to be detected, expressed as a multiple of the standard deviation. You can provide $\delta$ as the delta on the dialog or through JSL.
- $\Delta$ denotes the shift in $\mu$ to be detected, expressed in data units. If the sample size $n$ is constant across subgroups, then
  \[ \Delta = \delta \bar{X} = (\delta \sigma) / \sqrt{n} \]

Note that some authors use the symbol $D$ instead of $\Delta$. You can provide $\Delta$ as the delta on the dialog or through JSL.

One-Sided CUSUM Charts

Positive Shifts

If the shift $\delta$ to be detected is positive, the CUSUM computed for the $t^{th}$ subgroup is

\[ S_t = \max(0, S_{t-1} + (z_t - k)) \]

for $t = 1, 2, ..., n$, where $S_0 = 0$, $z_t$ is defined as for two-sided schemes, and the parameter $k$, termed the reference value, is positive. The cusum $S_t$ is referred to as an upper cumulative sum. Since $S_t$ can be written as

\[ \max \left\{ 0, S_{t-1} + \frac{\bar{X}_t - (\mu_0 + k \sigma \bar{X}_t)}{\sigma \bar{X}_t} \right\} \]

the sequence $S_t$ cumulates deviations in the subgroup means greater than $k$ standard errors from $\mu_0$. If $S_t$ exceeds a positive value $h$ (referred to as the decision interval), a shift or out-of-control condition is signaled.
Negative Shifts

If the shift to be detected is negative, the cusum computed for the \( t \)th subgroup is

\[
S_t = \max(0, S_{t-1} - (z_t + k))
\]

for \( t = 1, 2, \ldots, n \), where \( S_0 = 0 \), \( z_t \) is defined as for two-sided schemes, and the parameter \( k \), termed the reference value, is positive. The cusum \( S_t \) is referred to as a lower cumulative sum. Since \( S_t \) can be written as

\[
\max \left( 0, S_{t-1} - \frac{\bar{X}_t - (\mu_0 - k\sigma / \sqrt{n})}{\sigma / \sqrt{n}} \right)
\]

the sequence \( S_t \) cumulates the absolute value of deviations in the subgroup means less than \( k \) standard errors from \( \mu_0 \). If \( S_t \) exceeds a positive value \( h \) (referred to as the decision interval), a shift or out-of-control condition is signaled.

Note that \( S_t \) is always positive and \( h \) is always positive, regardless of whether \( \delta \) is positive or negative. For schemes designed to detect a negative shift, some authors define a reflected version of \( S_t \) for which a shift is signaled when \( S_t \) is less than a negative limit.

Lucas and Crosier (1982) describe the properties of a fast initial response (FIR) feature for CUSUM schemes in which the initial CUSUM \( S_0 \) is set to a “head start” value. Average run length calculations given by them show that the FIR feature has little effect when the process is in control and that it leads to a faster response to an initial out-of-control condition than a standard CUSUM scheme. You can provide head start values on the dialog or through JSL.

Constant Sample Sizes

When the subgroup sample sizes are constant (= \( n \)), it may be preferable to compute cusums that are scaled in the same units as the data. Cusums are then computed as

\[
S_t = \max(0, S_{t-1} + (\bar{X}_t - (\mu_0 + k\sigma / \sqrt{n})))
\]

for \( \delta > 0 \) and the equation

\[
S_t = \max(0, S_{t-1} - (\bar{X}_t - (\mu_0 - k\sigma / \sqrt{n})))
\]

for \( \delta < 0 \). In either case, a shift is signaled if \( S_t \) exceeds \( h' = h\sigma / \sqrt{n} \). Some authors use the symbol \( H \) for \( h' \).

Two-Sided Cusum Schemes

If the cusum scheme is two-sided, the cumulative sum \( S_t \) plotted for the \( t \)th subgroup is

\[
S_t = S_{t-1} + z_t
\]

for \( t = 1, 2, \ldots, n \). Here \( S_0 = 0 \), and the term \( z_t \) is calculated as

\[
z_t = (\bar{X}_t - \mu_0) / (\sigma / \sqrt{n}_t)
\]
where $\tilde{X}_i$ is the $i^{th}$ subgroup average, and $n_i$ is the $i^{th}$ subgroup sample size. If the subgroup samples consist of individual measurements $x_i$, the term $z_i$ simplifies to

$$z_i = (x_i - \mu_0)/\sigma$$

Since the first equation can be rewritten as

$$S_t = \sum_{i=1}^{t} z_i = \sum_{i=1}^{t} \frac{(\tilde{X}_i - \mu_0)}{\sigma \tilde{X}_i}$$

the sequence $S_t$ cumulates standardized deviations of the subgroup averages from the target mean $\mu_0$.

In many applications, the subgroup sample sizes $n_i$ are constant ($n_i = n$), and the equation for $S_t$ can be simplified.

$$S_t = \frac{1}{\sigma \bar{X}} \sum_{i=1}^{t} (\tilde{X}_i - \mu_0) = \frac{\sqrt{n}}{\sigma} \sum_{i=1}^{t} (\tilde{X}_i - \mu_0)$$

In some applications, it may be preferable to compute $S_t$ as

$$S_t = \sum_{i=1}^{t} (\tilde{X}_i - \mu_0)$$

which is scaled in the same units as the data. In this case, the procedure rescales the V-mask parameters $h$ and $k$ to $h' = h \sigma / \sqrt{n}$ and $k' = k \sigma / \sqrt{n}$, respectively. Some authors use the symbols $F$ for $k'$ and $H$ for $h'$.

If the process is in control and the mean $\mu$ is at or near the target $\mu_0$, the points will not exhibit a trend since positive and negative displacements from $\mu_0$ tend to cancel each other. If $\mu$ shifts in the positive direction, the points exhibit an upward trend, and if $\mu$ shifts in the negative direction, the points exhibit a downward trend.
Multivariate control charts address process monitoring problems where several related variables are of interest.

**Figure 14.1** Example of a Multivariate Control Chart

Note: UCL is calculated based on Alpha=0.05
Contents

Launch the Platform ......................................................... 295
Control Chart Usage ....................................................... 295
  Phase 1—Obtaining Targets ............................................. 295
  Phase 2—Monitoring the Process ...................................... 296
  Monitoring a Grouped Process ......................................... 297
Change-Point Detection ..................................................... 300
  Method ............................................................................ 300
  Example ......................................................................... 300
Platform Options ............................................................... 302
Statistical Details ............................................................... 302
  Ungrouped Data ............................................................ 302
  Grouped Data ............................................................... 303
  Additivity ................................................................. 304
  Change-Point Detection .................................................. 304
Launch the Platform

To generate a multivariate control chart, select Graph > Control Chart > Multivariate Control Chart.

**Figure 14.2 Multivariate Control Chart Launch Window**

- **Y, Columns** are the columns to be analyzed.
- **Subgroup** is a column that specifies group membership. Hierarchically, this group is within **Group**.
- **Group** is a column that specifies group membership at the highest hierarchical level.

In addition, there is a **Get Targets** button that allows you to pick a JMP table that contains historical targets for the process.

Control Chart Usage

There are two distinct phases in generating a multivariate control chart. Phase 1 is the use of the charts for establishing control. Historical limits are set in this stage, and initial tests show whether the process is in statistical control. The objective of phase 1 is to obtain an in-control set of observations so that control limits can be set for phase 2, which is the monitoring of future production.

Phase 1—Obtaining Targets

To illustrate the process of creating a multivariate control chart, we use data collected on steam turbines, taken from Mason and Young (2002). Historical data, stored in Steam Turbine Historical.jmp, and found in the Quality Control subfolder, is used to construct the initial chart.

Launch the platform and assign all continuous variables to the **Y, Columns** role. When you click **OK**, Figure 14.3 appears.
The process seems to be in reasonable statistical control, since there is only one out of control point. Therefore, we use it to establish targets. To do so, select **Save Target Statistics** from the platform menu. This creates a new data table containing target statistics for the process.

**Phase 2—Monitoring the Process**

With targets saved, we can create the multivariate control chart that monitors the process.

Open **Steam Turbine Current.jmp**, from the Quality Control subfolder, which holds recent observations from the process. Launch the Multivariate Control Chart platform, and again assign all variables to the **Y, Columns** role. This time, click the **Get Targets** button in the launch dialog, and select the **Steam Turbine Targets.jmp** table that was saved in phase 1. Click OK. By default, the T Square Chart, at the top of the
report, shows a UCL, which is calculated with an alpha level of 0.05. This alpha level may be easily changed by clicking on the red-triangle menu of the Multivariate Control Chart platform and selecting Set Alpha Level. Several alpha level options are provided, including 0.01, 0.05, 0.10, 0.50, and Other.... For example, to set the alpha level to 0.001, select Other.... Type 0.001 into the dialog window asking you to specify the alpha level for the upper control limit and click OK. A new T Square Chart is displayed with an UCL calculated using your specified alpha (0.001 in this example). Figure 14.5 shows the T Square Chart with the UCL based on an alpha level of 0.001.

As shown in Figure 14.5, out-of-control conditions occur at observations 2, 3, 4, 5, and 8. This result implies that these observations do not conform to the historical data of Steam Turbine Historical.jmp, and that the process should be further investigated.

**Monitoring a Grouped Process**

The workflow for monitoring a multivariate process with grouped data is similar to the one for ungrouped data. An initial control chart is used to create target statistics, and these statistics are used in monitoring the process.

As an example, open Aluminum Pins Historical.jmp, which monitors a process of manufacturing aluminum pins. Enter all the Diameter and Length variables as Y, Columns and subgroup as the Subgroup. After clicking OK, you see the chart shown in Figure 14.6.
Again, the process seems to be in statistical control, making it appropriate to create targets. Select **Save Target Statistics** and save the resulting table as *Aluminum Pins Targets.jmp*.

Now, open *Aluminum Pins Current.jmp* to see current values for the process. To monitor the process, launch the Multivariate Control Chart platform, specifying the columns as in phase 1. Click **Get Targets** and select the saved targets file to produce the chart shown in Figure 14.7, which also has the **Show Means** option selected. Notice that the **Principal Components** option is shown by default.
Figure 14.7  Grouped Multivariate Control Chart, Phase 2

![Multivariate Control Chart](image)

**Note:** UCL is calculated based on Alpha=0.05

**Principal Components of Target Data: on Covariances**

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Percent</th>
<th>Cum Percent</th>
<th>ChiSquare</th>
<th>DF</th>
<th>Prob&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0020</td>
<td>77.438</td>
<td>77.438</td>
<td>194.981</td>
<td>20.00</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>0.0004</td>
<td>14.112</td>
<td>91.548</td>
<td>90.046</td>
<td>14.00</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>0.0002</td>
<td>6.543</td>
<td>98.092</td>
<td>57.140</td>
<td>9.00</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>0.0000</td>
<td>1.116</td>
<td>99.201</td>
<td>7.721</td>
<td>5.00</td>
<td>0.1723</td>
</tr>
<tr>
<td>0.0000</td>
<td>0.442</td>
<td>99.643</td>
<td>0.227</td>
<td>2.00</td>
<td>0.6925</td>
</tr>
<tr>
<td>0.0000</td>
<td>0.357</td>
<td>100.000</td>
<td>0.000</td>
<td>0.00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Eigenvectors**

- Diameter2: 1.94391, 21.04343, 4.33901, 83.66331, 225.8578, 72.86281
- Diameter3: -2.05350, 24.72960, 17.2384, 95.30283, -115.819, -185.706
- Diameter4: -1.81865, 30.15928, 4.14866, -136.824, 58.89922, 97.6126
- Length1: 17.11879, 26.33745, 48.34854, 4.23224, 8.27883, 19.3736

**Note:** Eigenvectors were divided by square root of eigenvalues.

**Principal Components of Current Data: on Covariances**

<table>
<thead>
<tr>
<th>Group Means</th>
<th>Count</th>
<th>Diameter1</th>
<th>Diameter2</th>
<th>Diameter3</th>
<th>Diameter4</th>
<th>Length1</th>
<th>Length2</th>
</tr>
</thead>
<tbody>
<tr>
<td>subgroup</td>
<td>Diameter1</td>
<td>Diameter2</td>
<td>Diameter3</td>
<td>Diameter4</td>
<td>Length1</td>
<td>Length2</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>10.0960</td>
<td>9.9950</td>
<td>9.9950</td>
<td>14.9950</td>
<td>49.8250</td>
<td>60.0300</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>10.0960</td>
<td>9.9950</td>
<td>9.9950</td>
<td>14.9950</td>
<td>49.8250</td>
<td>60.0300</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>10.0960</td>
<td>9.9950</td>
<td>9.9950</td>
<td>14.9950</td>
<td>49.8250</td>
<td>60.0300</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>10.0960</td>
<td>10.0000</td>
<td>9.9950</td>
<td>14.9950</td>
<td>49.8000</td>
<td>59.9950</td>
</tr>
</tbody>
</table>
Change-Point Detection

The Change-Point Detection method is based upon the work of Sullivan and Woodall (2000). When the dataset is comprised of multivariate individual observations, a control chart can be developed to detect a shift in the mean vector, the covariance matrix, or both. This method partitions the data and calculates likelihood ratio statistics for a shift. These statistics are divided by the expected value for no shift and are then plotted by the row number. A Change-Point Detection plot readily shows the change point for a shift occurring at the maximized value of the test statistic.

Method

Suppose there are $m$ independent observations from a multivariate normal distribution of dimensionality $p$ such that

$$x_i \sim N_p(\mu_i, \Sigma_i), \quad i = 1, \ldots, m.$$  

where $x_i$ is an individual observation, and $N_p(\mu_i, \Sigma_i)$ represents a multivariate normally distributed mean vector and covariance matrix, respectively.

If a distinct change occurs in the mean vector, the covariance matrix, or both, after $m_1$ observations, all observations through $m_1$ possess the same mean vector and the same covariance matrix ($\mu_a, \Sigma_a$). Similarly, all ensuing observations, beginning with $m_1 + 1$, have the same mean vector and covariance matrix ($\mu_b, \Sigma_b$). If the data are from an in-control process, then $\mu_a = \mu_b$ and $\Sigma_a = \Sigma_b$ for all values of $m$, and the parameters of the in-control process can be estimated directly from the data.

A likelihood ratio test approach is used to determine changes or a combination of changes in the mean vector and covariance matrix. The likelihood ratio statistic is plotted for all possible $m_1$ values, and an appropriate Upper Control Limit (UCL) is chosen. The location (observation or row number) of the maximum test statistic value corresponds to the maximum likelihood location of only one shift, assuming that exactly one change (or shift) occurred. For technical details of this method, refer to the Statistical Details in the section "Change-Point Detection," p. 304.

Example

As an example of determining a possible change or shift in the data, open Gravel.jmp from the Quality Control subfolder in the Sample Data directory. This dataset can be found in Sullivan and Woodall (2000) and contains 56 observations from a European gravel production plant. The two columns of the dataset show the percent of the particles (by weight) that are large and medium in size. Select Graph > Control Chart > Multivariate Control Chart. Select Large and Medium as $Y$, Columns, and click OK. Select Change-Point Detection from the Multivariate Control Chart platform menu in the report. The resulting Change-Point Detection Plot is shown in Figure 14.8.
Control chart statistics for the Change-Point Detection plot are obtained by dividing the likelihood ratio statistic of interest (either a mean vector or a covariance matrix) by a normalizing factor. Plotted values above 1.0 indicate a possible shift in the data. The change point of the data occurs for the observation having the maximum test statistic value for the Change-Point Detection plot. For the Gravel.jmp data, at least one shift is apparent, with the change point occurring at observation 24 and the shift occurring immediately following observation 24. See “Change-Point Detection,” p. 304 in this chapter for technical details regarding shift changes.

A scatterplot matrix of the data, divided into two groups, is shown in Figure 14.9. This plot shows the shift in the sample mean vector. The first 24 observations are identified as the first group; the remaining observations are classified as the second group. The 95% prediction regions for the two groups have approximately the same size, shape, and orientation, visually indicating that the sample covariance matrices are similar.

Although the Scatterplot Matrix is created automatically when you select Change-Point Detection from the platform's menu, you may need to slightly alter the axes in order to see the density ellipses for the two groups, depending upon your data. This is done by clicking and dragging the axes, as needed. For this example, you can also obtain the plot shown in Figure 14.9 by clicking on the red-triangle menu for the Multivariate Control Chart in the Gravel.jmp data table and selecting Run Script.
Platform Options

The following options are available from the platform drop-down menu:

- **$T^2$ Chart** shows or hides the $T^2$ chart.
- **T Square Partitioned** allows you to specify the number of major principal components for $T^2$.
- **Set Alpha Level** sets the $\alpha$-level used to calculate the control limit. The default is $\alpha=0.05$.
- **Show Covariance** shows or hides the covariance report.
- **Show Correlation** shows or hides the correlation report.
- **Show Inverse Covariance** shows or hides the inverse covariance report.
- **Show Inverse Correlation** shows or hides the inverse correlation report.
- **Show Means** shows or hides the group means.
- **Save T Square** creates a new column in the data table containing $T^2$ values.
- **Save T Square Formula** creates a new column in the data table, and stores a formula that calculates the $T^2$ values.
- **Save Target Statistics** creates a new data table containing target statistics for the process.
- **Change-Point Detection** shows or hides a Change-Point Detection plot of test statistics by row number and indicates the row number where the change point appears.
- **Principal Components** shows or hides a report showing a scaled version of the principal components on the covariances. The components are scaled so that their sum is the $T^2$ value.
- **Save Principal Components** creates new columns in the data table that hold the scaled principal components.

Statistical Details

Ungrouped Data

The $T^2$ statistic is defined as

$$T^2 = (Y - \mu)'S^{-1}(Y - \mu)$$

where

- $S$ is the covariance matrix
- $\mu$ is the true mean
- $Y$ represents the observations
During Phase 1 (when you have not specified any targets), the upper control limit (UCL) is a function of the beta distribution. Specifically,

\[
UCL = \frac{(n-1)^2}{n} \beta\left(\frac{p}{2}, \frac{n-p-1}{2}\right)
\]

where

- \( p \) is number of variables
- \( n \) is the sample size

During phase 2, when targets are specified, the UCL is a function of the \( F \)-distribution, defined as

\[
UCL = \frac{p(n+1)(n-1)}{n(n-p)} F(\alpha, p, n-p)
\]

where

- \( p \) is number of variables
- \( n \) is the sample size for each subgroup

**Grouped Data**

The \( T^2 \) statistic is defined as

\[
T^2 = n(\bar{Y} - \mu)^T S^{-1} (\bar{Y} - \mu)
\]

where

- \( S \) is the covariance matrix
- \( \mu \) is the true mean
- \( \bar{Y} \) represents the observations

During Phase 1, the Upper Control Limit is

\[
UCL = \frac{p(m-1)(m-1)}{mn - m - p + 1} F(\alpha, p, mn - m - p + 1)
\]

where

- \( p \) is number of variables
- \( n \) is the sample size for each subgroup
- \( m \) is the number of subgroups
During Phase 2, the Upper Control Limit is

\[
UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F(\alpha, p, mn-m-p+1)
\]

where

- \( p \) is number of variables
- \( n \) is the sample size for each subgroup
- \( m \) is the number of subgroups

**Additivity**

When a sample of \( mn \) independent normal observations are grouped into \( m \) rational subgroups of size \( n \), the distance between the mean \( \bar{Y}_j \) of the \( j \)th subgroup and the expected value \( \mu \) is \( T_M^2 \). Note that the components of the \( T^2 \) statistic are additive, much like sums of squares. That is,

\[
T_A^2 = T_M^2 + T_D^2
\]

Let \( T_M^2 \) represent the distance from a target value,

\[
T_M^2 = n(\bar{Y}_j - \mu)^{T} S_p^{-1}(\bar{Y}_j - \mu)
\]

The internal variability is

\[
T_D^2 = \sum_{j=1}^{n} (Y_j - \bar{Y})^{T} S_p^{-1}(Y_j - \bar{Y})
\]

The overall variability is

\[
T_A^2 = \sum_{j=1}^{n} (\bar{Y}_j - \mu)^{T} S_p^{-1}(\bar{Y}_j - \mu)
\]

**Change-Point Detection**

The log of the likelihood function is maximized for the first \( m_1 \) observations:

\[
l_1 = -m_1 k_1 \log[2\pi] - m_1 \log \left| S_1 \right| - m_1 \bar{k}_1
\]

where \( \left| S_1 \right| \) is the maximum likelihood estimate of the covariance matrix for the first \( m_1 \) observations, and the rank of \( S_1 \) is defined as \( k_1 = \text{Min}[p, m_1 - 1] \), where \( p \) is the dimensionality of the matrix.

The log-likelihood function for the subsequent \( m_2 = m - m_1 \) observations is \( l_2 \), and is calculated similarly to \( l_0 \), which is the log-likelihood function for all \( m \) observations.
The sum \( l_1 + l_2 \) is the likelihood that assumes a possible shift at \( m_1 \), and is compared with the likelihood \( l_0 \), which assumes no shift. If \( l_0 \) is substantially smaller than \( l_1 + l_2 \), the process is assumed to be out of control.

The log-likelihood ratio, multiplied by two, is

\[
\text{lrt}[m_1] = l_1 + l_2 - l_0
\]

\[
\text{lrt}[m_1] = (m_1(p - k_1) + m_2(p - k_2))(1 + \log[2\pi])
\]

\[
+ m\log[|S|] - m_1\log\left[\frac{|S_1|}{k_1}\right] - m_2\log\left[\frac{|S_2|}{k_2}\right]
\]

and has a chi-squared distribution, asymptotically, with the degrees of freedom equal to \( p(p + 3)/2 \). Large log-likelihood ratio values indicate that the process is out-of-control.

Dividing the above equation (the log-likelihood ratio, multiplied by two) by its expected value, which is determined from simulation and by the UCL, yields an upper control limit of one on the control chart. Therefore, the control chart statistic becomes:

\[
y[m_1] = \frac{\text{lrt}[m_1]}{E[\text{lrt}[m_1]] \times \text{UCL}}
\]

and, after dividing by \( p(p + 3)/2 \), yields the expected value:

\[
\text{ev}[m,p,m_1] = a_p + m_1b_p \quad \text{if } m_1 < p + 1,
\]

\[
\text{ev}[m,p,m_1] = a_p + (m - m_1)b_p \quad \text{if } (m - m_1) < p + 1,
\]

\[
\text{ev}[m,p,m_1] = 1 + \frac{(m - 2p - 1)}{(m - p)(m - p - m_1)}, \quad \text{otherwise.}
\]

The intercept in the above equation is:

\[
a_p = \frac{-0.08684(p - 14.69)(p - 2.036)}{(p - 2)}
\]

and the slope is:

\[
b_p = \frac{0.1228(p - 1.839)}{(p - 2)}
\]

When \( p = 2 \), the value of \( \text{ev}[m,p,m_1] \) when \( m_1 = m_2 = 2 \) is 1.3505. Note that the above formulas are not accurate for \( p > 12 \) or \( m < (2p + 4) \). In such cases, simulation should be used.
With Phase 1 control charts, it is useful to specify the upper control limit (UCL) as the probability of a false out-of-control signal. An approximate UCL where the false out-of-control signal is roughly 0.05 and is dependent upon \( m \) and \( p \), is given as:

\[
UCL[m, p] = (3.338 - 2.115\log[p] + 0.8819(\log[p])^2 - 0.1382(\log[p])^3) + (0.6389 - 0.3518\log[p] + 0.01784(\log[p])^3)\log[m].
\]

The approximate control chart statistic is then given by:

\[
\hat{y}[m_1] = \frac{2\text{lrt}[m_1]}{\rho(\rho + 3)(\text{ev}[m, p, m_1 | UCL[m, p]])}
\]
References


References


Hsu, J. (1989), “Multiple Comparison Procedures” ASA Short Course notes, Columbus OH: Ohio State University.
References


Linnerud: see Rawlings (1988).


Myers, R.H. (1989), *Classical and Modern Regression with Applications*, Boston: PWS-KENT.

Miller, A.J. (1990), Subet Selection in Regression, New York: Chapman and Hall.


References


Umetrics (1995), Multivariate Analysis (3-day course), Winchester, MA.


Western Electric Company (1956), Statistical Quality Control Handbook, currently referred to as the AT&T Statistical Quality Control Handbook.


Institute Inc.


Designs.” *Journal of Quality Technology*, 33, 140-152.
Index

Quality and Reliability Methods

Symbols

_LimitsKey 190

Numerics

2 Factors Crossed.jmp 161
-2loglikelihood 13

A

Agreement Comparisons 181
Agreement Counts 184
Agreement Report 181
Agreement within raters 181
AIAG Labels 164
AICc 14
Aluminum Pins Historical.jmp 297
Appliance.jmp 121
Arrhenius model 133
Attribute data 178
Attribute gauge analysis 178
AttributeGage.jmp 178

B

Bayesian variance components 170
Bias Report 164, 174
biased mean estimate 112
bibliographic references 307–318
BIC 14
Bladder Cancer.jmp 71
Blenders.jmp 22
Braces.jmp 274
By role 4

C

Capability Analysis 265
with Control Charts 236
Capability Indices Report 196
Capability Platform 187
Category Legend 201–202
Cause 131
Cause role 200
Cause Summary 25
Causes 201–202
c-Chart 275
Censor 130, 141
censor data 103, 105, 148, 152, 154
Censor Indicator 24
Censor role 4, 106, 131, 141
CensorLabels.jmp 20
Change Confidence Level 13
Change Confidence Level 46
Change Type in Diagram 219
Chi Square 142
Chi-Square 111
citations 307–318
Clips1.jmp 268, 270
Coating.jmp 257, 260, 262, 265, 267
Colors 203
Combine Causes 202
comparative Pareto chart 199, 206
Compare Distributions 6
Comparison Criterion 13
Comparison Criterion 16
Competing Causes 22
Competing Causes 116, 121–123
comptime.jmp 127
conditional risk ratio 141
Confidence Intervals 132
Confidence Limits 152, 154–155, 157
Conformance Report 185
Connect Cell Means 163
Connect Quantile Points 114
Control Charts

c 275
Cusum 281–292
EWMA 269
Individual Measurement 262
Moving Average 267
Moving Range 262
Index

Gauge RR Plots 165
Gauge Std Dev 165
Gauge Studies 164
Get Targets 295
Go 150–151
Goal Plot Labels 192
Group Means of Std Dev 164
Grouping 67
Grouping role 106, 161, 206, 208

H
hazard function 141
also see Proportional Hazards
hazard ratio 141

I
icdevice02.jmp 138
Import Spec Limits 189
Import Spec Limits() 190
Individual Causes Report 27
Individual Measurement Chart 262
Interval-Censored model 138
interval-censoring 21
Ishikawa diagram 215
Iteration History report 142

K
Kaplan-Meier see Survival, Univariate
key cell 209

L
Label 203
Label Cum Percent Points 201
Label role 4
Label, System ID 67
Largest Extreme Value (LEV) 35
left-censor 152
left-censoring 21
Levey-Jennings 234
Life Distribution
  Statistical Details 29
  Lifetime Distribution Platform 1–39
Likelihood Ratio Tests 132
Limits Data Table 190
Linearity Study 164, 176
Locomotive.jmp 156
Logistic 34
Loglogistic 32
loglogistic loss function 149
Loglogistic with threshold 38
LogNormal 130
Lognormal 30
lognormal distribution 105
Lognormal Fit 115
lognormal loss function 149–152, 156–157
Lognormal Plot 115
Lognormal with threshold 37
Log-Rank 111
Loss 151, 154–155
loss column 149
loss function 148, 154
lower specification limit 172
LSL 172

M
Make Summary Table 195
Markers 203
maximum likelihood 148, 154, 157
MCF Confid Limits 70
MCF Plot 70
MCF Plot, Table 69
Mean Diamonds 164
Mean of Std Dev 164
Mean Plots 165
MicroprocessorData.jmp 18
Midstep Quantile Points 114
Misclassification Probabilities 174
Mixed Censoring Example 18
mortality rate 111
Move to First 202
Move to Last 203
Moving Average Chart 267
Moving Range Chart 262
MSALinearity.JMP 174, 176
multivar chart see Variability Chart platform
Multivariate Control Chart 295–304

N
N Legend 201
nested effect 159
Nonlinear 154
Nonlinear Fit platform
  example 152–156
  survival models 141–157
Nonlinear platform 125, 148–151, 154–155
  control panel 150, 157
  example 150–157
  survival models 148
Nonlinear Templates folder 154
Nonparametric 30
Nonparametric Estimates 8
Normal 33
  normal density function 149
  Normal Distribution 149
  normal distribution 149, 157
np-Chart 272
O
  Oil1 Cusum.jmp 286
  Omit Causes 123
  Omitted Causes 27
  one-way comparative Pareto chart 206
  Oneway platform 159
P
  Parameter Estimates table 142
  Parametric Regression 127–133
  Parametric Survival 127
  Pareto Chart platform 197–210
    before/after chart 206–208
    example 206–210
    launch 199–202
    options 202–204
    two-way comparative chart 208–210
  p-Chart 272
  Per Unit Rates 210
  Percent Scale 201–202
  Phase Detection 302
  Pickles.jmp 263
  Pie 204
  Pie Chart 201
  Plot MCF Differences 70, 72
  Plot Options 113
  Points Jittered 164
  prediction column 149, 154
  Predictor role 150–151, 154
  Pre-summarize 265
  Presummarize 234
  Presummarize Chart 265
  Principal Components 302
  Prob>ChiSq 111
  Process role 199
Product-Limit Survival Fit 110
  Proportional Hazards 141–148
Q
  Quantiles table 110
R
  R&R see Gauge R&R
  Rate 204
  Rats.jmp 109, 141
  R-Chart 257
  Recurrence 65, 67
  references 307–318
  relative frequency chart 197–210
  Reorder Horizontal 201, 209
  Reorder Vertical 201, 209
  Residual Quantile Plot 132
  Response role 161
  Reverse Y Axis command 112
  right-censoring 21, 105
  Run Charts 234
  Run Model 131
S
  S Control Limits 164
  Save Cause Coordinates 123
  Save Estimates 116, 150, 157
  Save Principal Components 302
  Save Residuals 132
  Save Spec Limits 191
  Save Spec Limits in New Table 191
  Save Summaries 241
  Save T Square 192
  Save T Square Formula 302
  Save Target Statistics 296, 302
  S-Chart 257
  Script 163, 165
  Semiconductor Capability.jmp 189
  Separate Causes 202–203
  Set Alpha Level 302
  Set Time Acceleration Baseline 46
  Shade Cpk Levels 192
  Shewhart Control Charts 255
  Shirts.jmp 275
  Show Bias Line 164
  Show Box Plots 164
  Show Cell Means 163
  Show Center Line 262
Weibull 130
Weibull distribution 105, 117
**Weibull Fit** 114
Weibull loss function 149–152, 154–156
Weibull Parameter Estimates table 114
**Weibull Plot** 114
Weibull with threshold 37
**Westgard Rules** 244
Whole Model Test table 142
Wilcoxon 111
X role 150–151, 154, 161, 206, 208
XBar Chart 257
**XBar Chart Limits** 164
Y role 3, 106, 161, 200
**Y, Age at Event** 67
Y, Time to Event role 4