Lean Six Sigma Black Belt

Using JMP Software

Written by
Will Mokczycki, Dr. Russell Boyles, Joan Ambrose, and Tracy Camp
Lean Six Sigma Master Black Belts
ETI Group

Presented by

ETI GROUP

Oregon: 503-484-5979
Washington: 425- 971-2990
Email: info@etigroupusa.com
Website: www.etigroupusa.com

This page intentionally left blank

Lean Six Sigma Black Belt, Volume II

Course outline with slide numbers

Tab 1. Statistical Analysis and Graphs

1. JMP Menu Maps 3
2. Basic Statistics and Statistical Graphics 5
3. Fitting and Using Distributions 35
4. Introduction to Life Data 71
5. Analyzing Life Data 79
6. Categorical MSA Without Standards 101
7. Comparing Populations — Continuous Y 117
8. Comparing Populations — Pass/fail Y 149

Tab 2. Regression

1. Introduction to Regression 179
2. Checking Model Adequacy 205
3. Using the Model: RMSE and Prediction Profiler 215
4. Introduction to the Prediction Profiler. 223
5. Multiple Regression 235
6. Dealing with Adequacy Issues 279
7. Simple Regression with Pass/Fail Y 305
8. Multiple Regression with Pass/Fail Y 319

This page intentionally left blank

Lean Six Sigma Black Belt, Volume II

Course outline with slide numbers

Tab 3. Design of Experiments

1. DOE Designed Experiments vs. File Cabinet Data. 327

2. One factor at a Time 343

3. DOE Terminology 353

4. Full Factorial Design 369

5. Statistical Assumptions 385

6. Statistical models 395

7. Design Principles 413

8. The Custom Design Process 425

9. Determining Sample Size for an Experiment 445

10. Screening Experiments 455

11. Trebuchet Exercise 491

12. Experiments with Hard to Change Factors 493

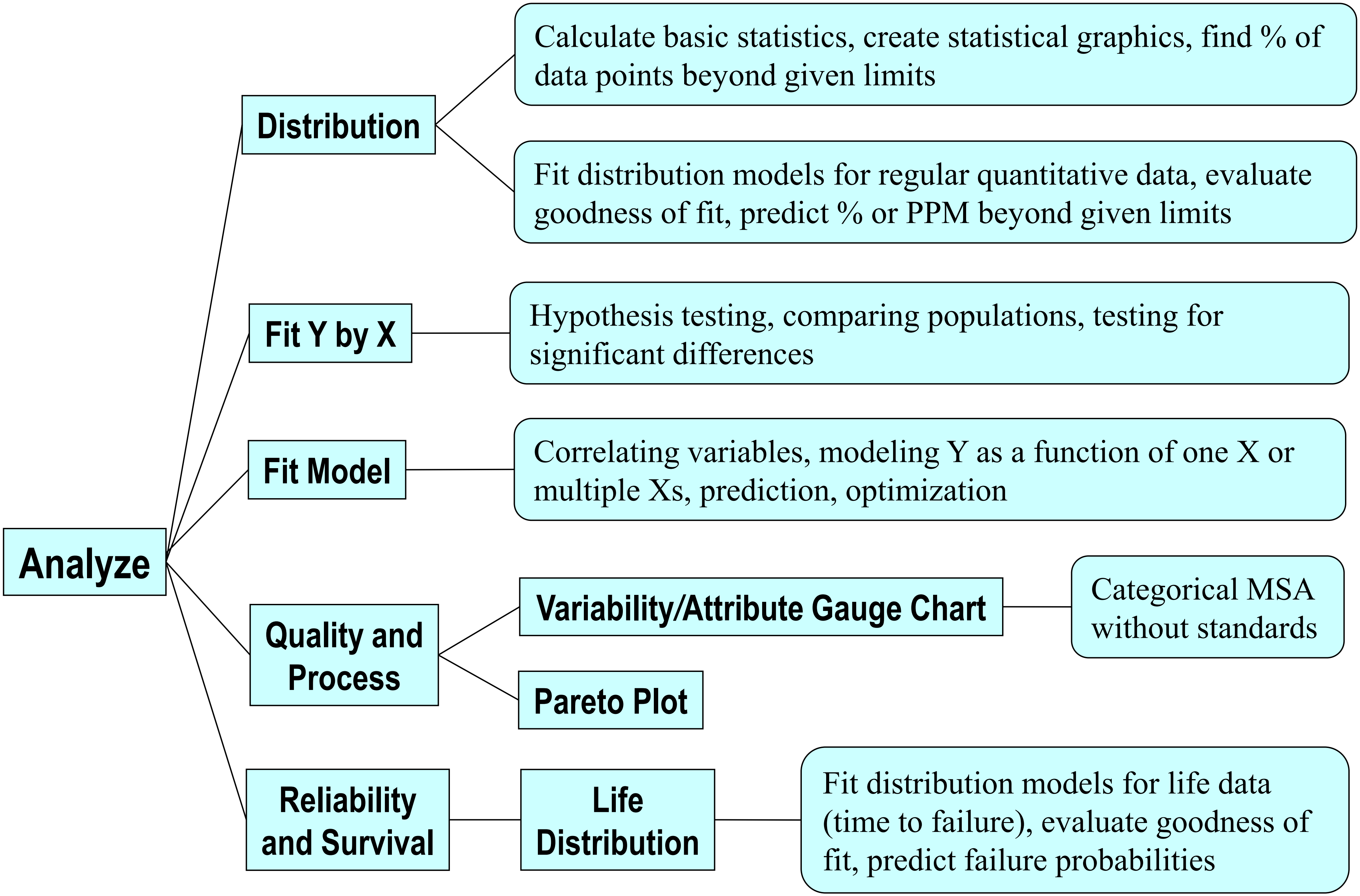
13. Multiple Response Optimization 509

This page left blank for pagination

Tab 1
Statistical Analysis Graphs

This Page Intentionally Left Blank

1 JMP menu map



JMP menu map (cont'd)

- **File**
  - **Preferences**
    - **Platforms** — Specify desired default settings for JMP analysis platforms
- **Tables**
  - **Summary** — Derive a smaller data table by calculating statistics over a subset of a larger data table
  - **Subset** — Extract a subset of a data table
  - **Sort** — Sort a data table by specified columns
  - **Stack** — Stack values from multiple columns into a single column
  - **Split** — Unstack values from a single column into multiple columns
- **DOE**
  - Calculate the required sample size for a designed experiment
  - Create the design matrix for a designed experiment
- **Graph**
  - **Overlay Plot** — Plot one or more data series in time sequence

2 Basic Statistics and Statistical Graphics

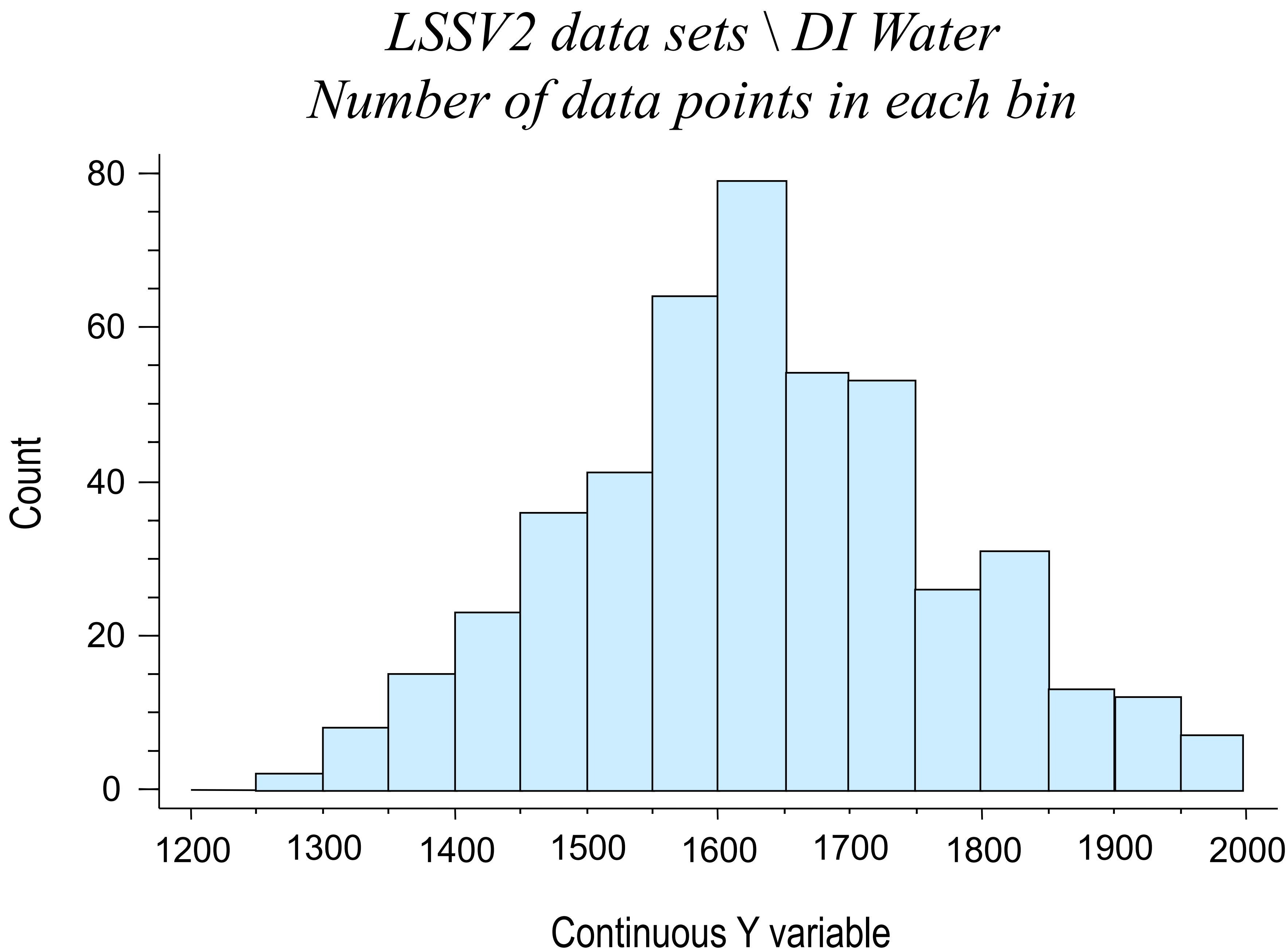
- Frequency histogram
- Cumulative distribution function
- Percentiles
- Box and whisker plot
- JMP distribution analysis
- Data validation
- Distribution analysis options
- Plotting data in time sequence
- Saving analyses and data tables

Y variables are characteristics of parts or transactions that determine customer satisfaction, or lack thereof. They provide the data from which project metrics are computed. In sections 2 and 3 we focus on *quantitative* Y variables. Examples include:

- Properties: physical, chemical, electrical, optical, . . .
- Distance, time, dimensions, cost, quantity
- Event counts (when there is not a discrete number of opportunities for the event to occur)

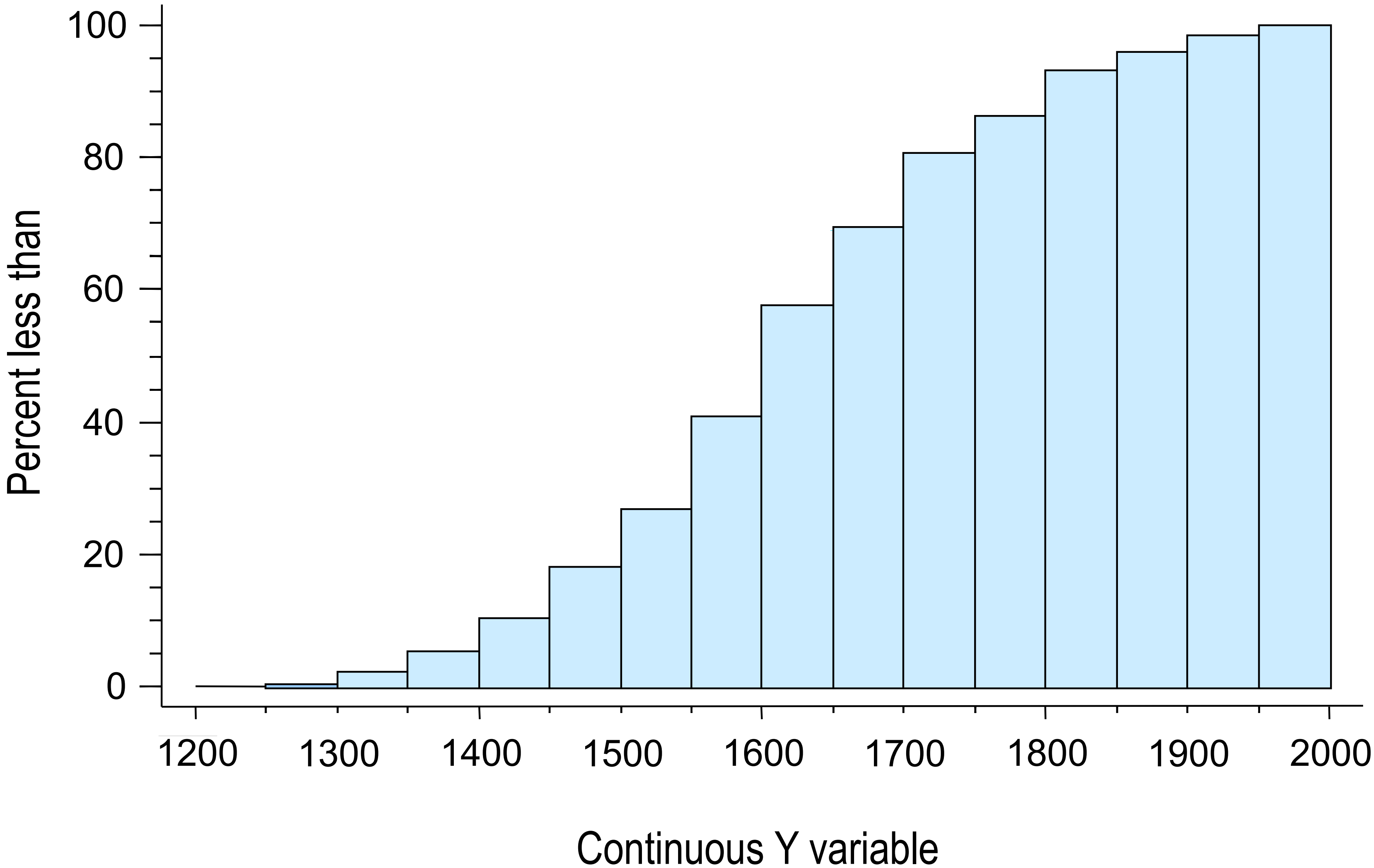
JMP uses the term *continuous* for quantitative variables, and often uses the term *nominal* for categorical variables.

Frequency histogram



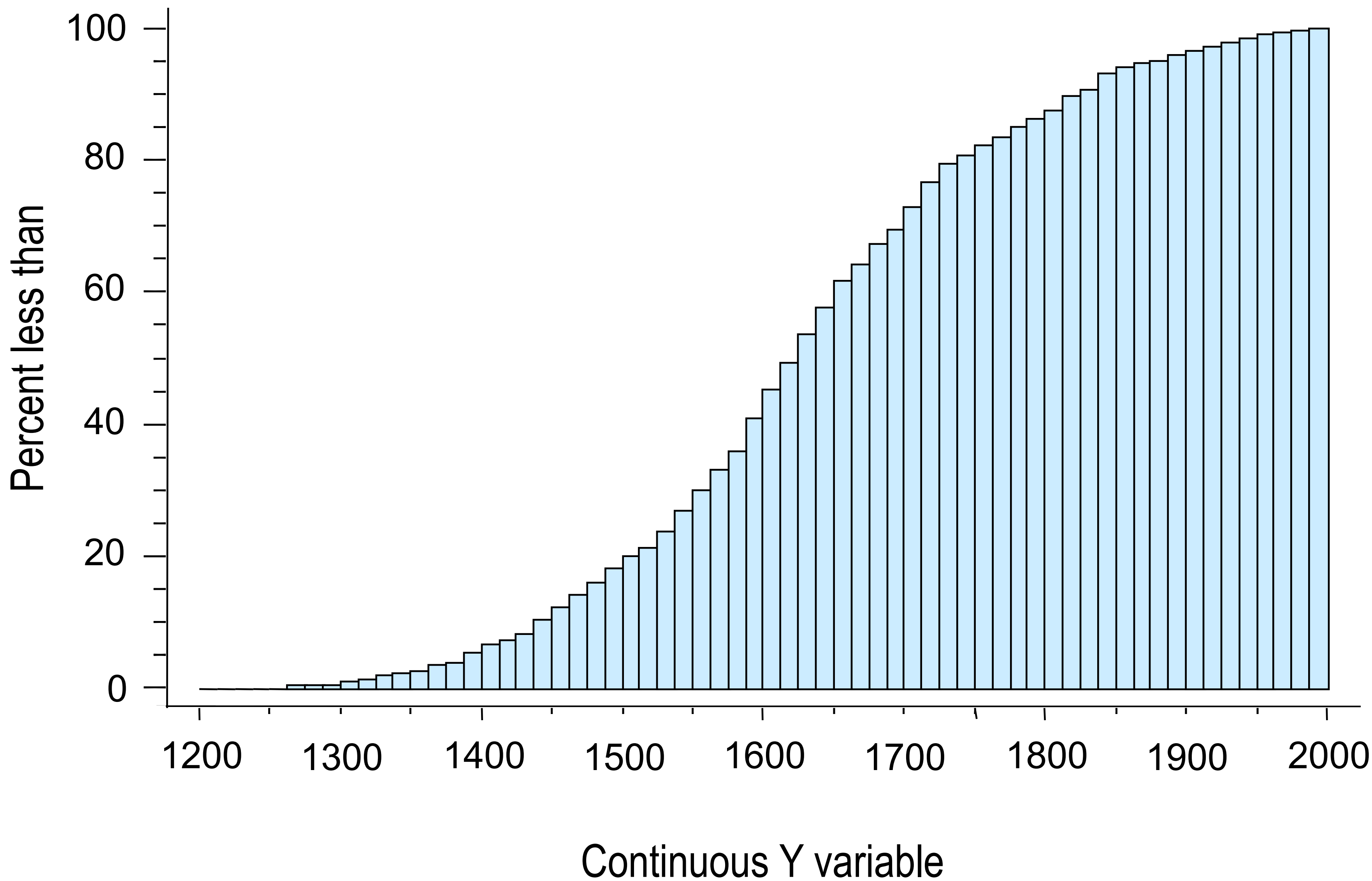
Cumulative percentage histogram

$$Percentage\ of\ data\ points\ \leq\ upper\ limit\ of\ each\ bin$$



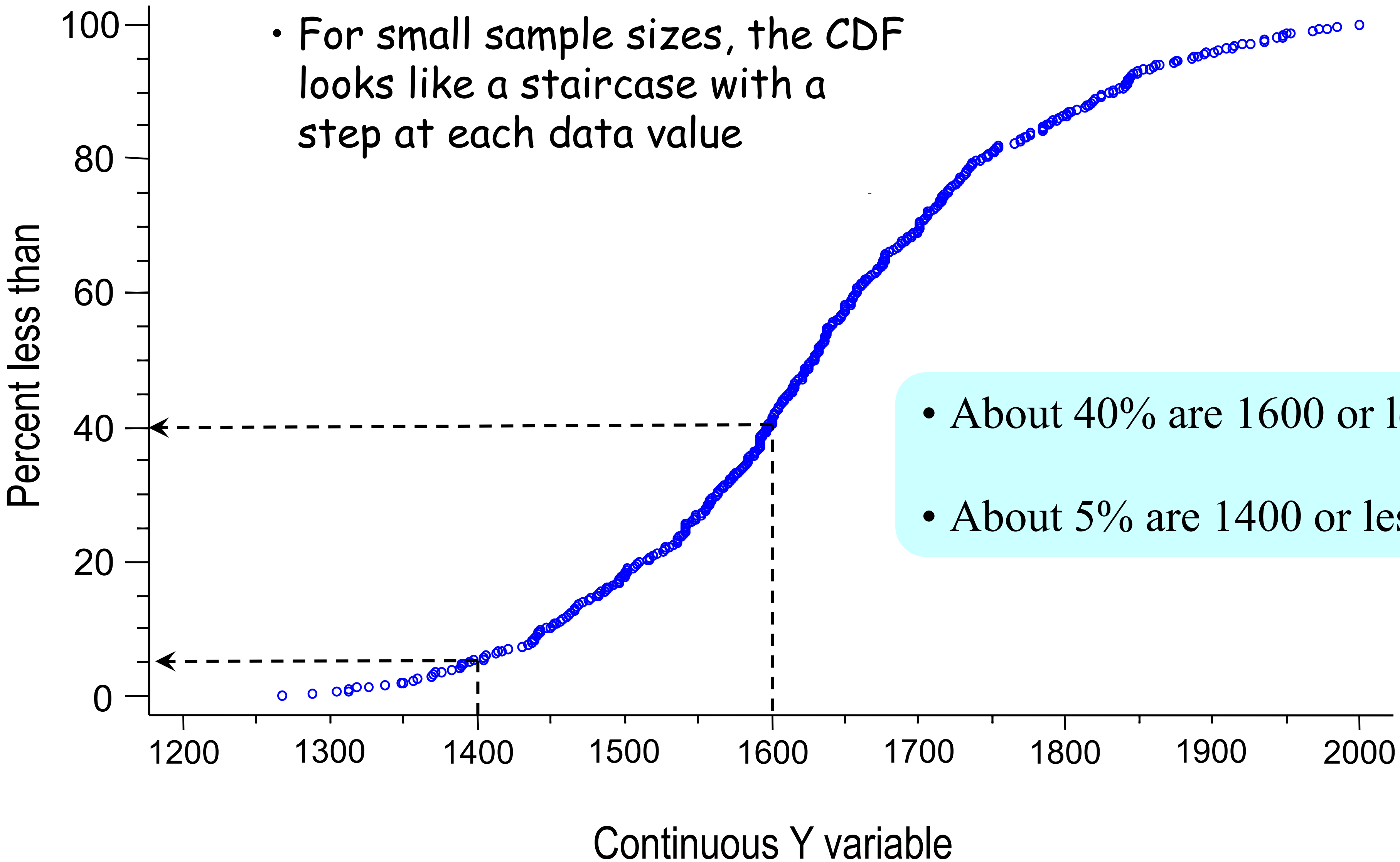
Cumulative percentage histogram (cont'd)

Made the bins smaller



Cumulative distribution function (CDF)

- Bins are so small they isolate individual data values
- For small sample sizes, the CDF looks like a staircase with a step at each data value



- About 40% are 1600 or less
- About 5% are 1400 or less

Percentiles

A *percentile* is a value that divides a population or data set into two groups, based on a stated percentage

10% are less than the **10th percentile**, 90% are greater

25% are less than the **25th percentile**, 75% are greater

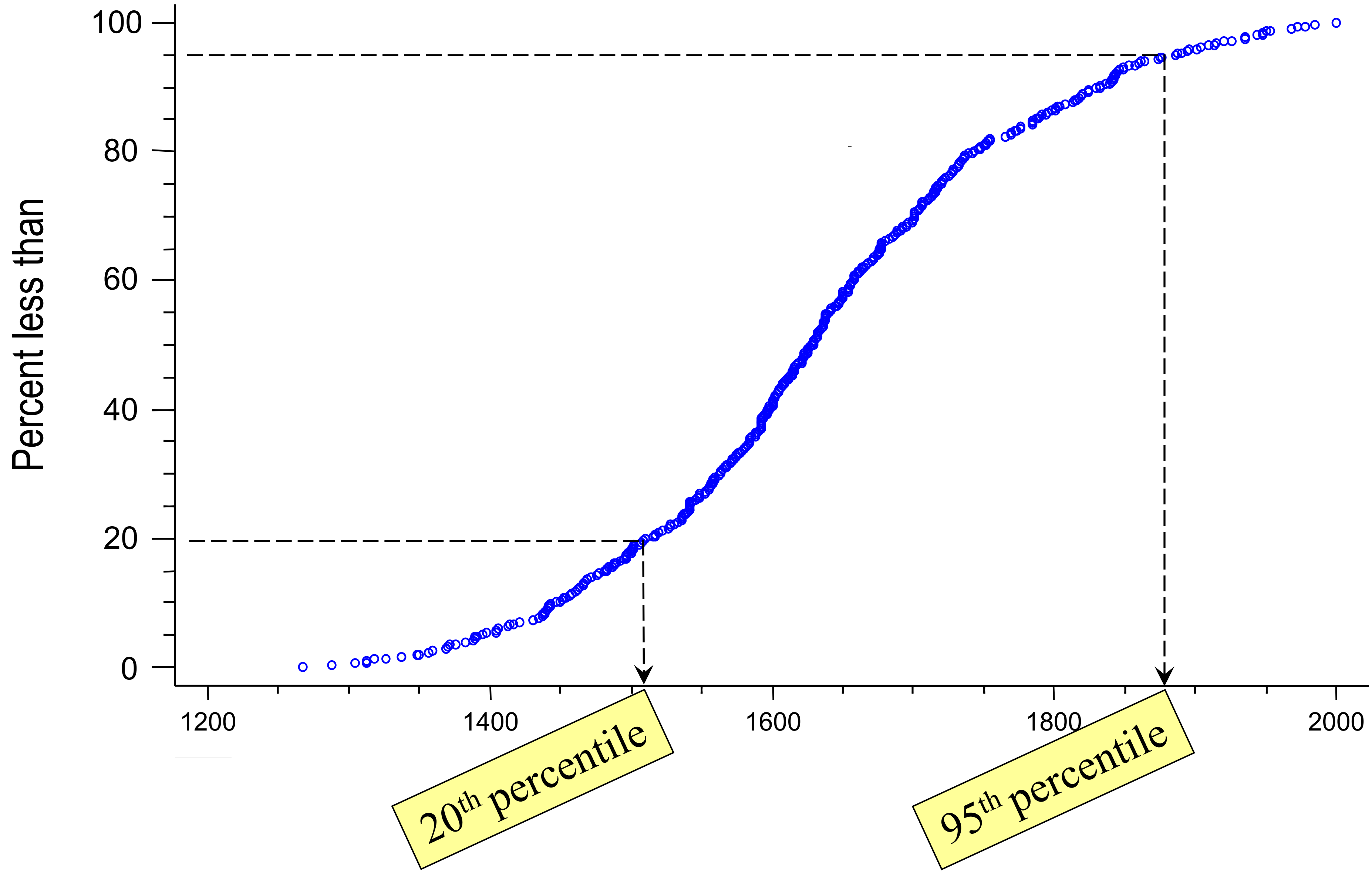
50% are less than the **50th percentile**, 50% are greater

75% are less than the **75th percentile**, 25% are greater

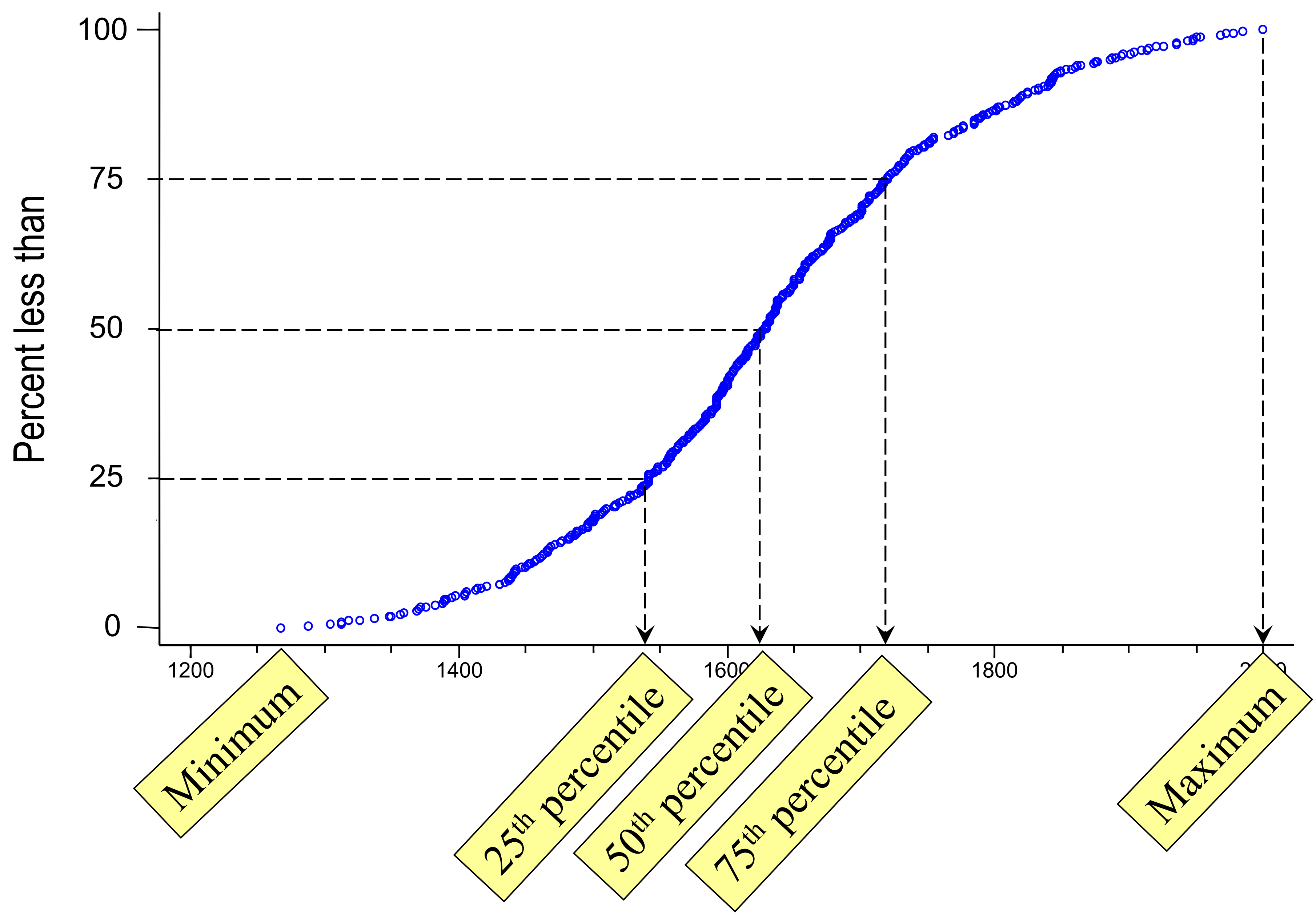
90% are less than the **90th percentile**, 10% are greater

Percentiles (cont'd)

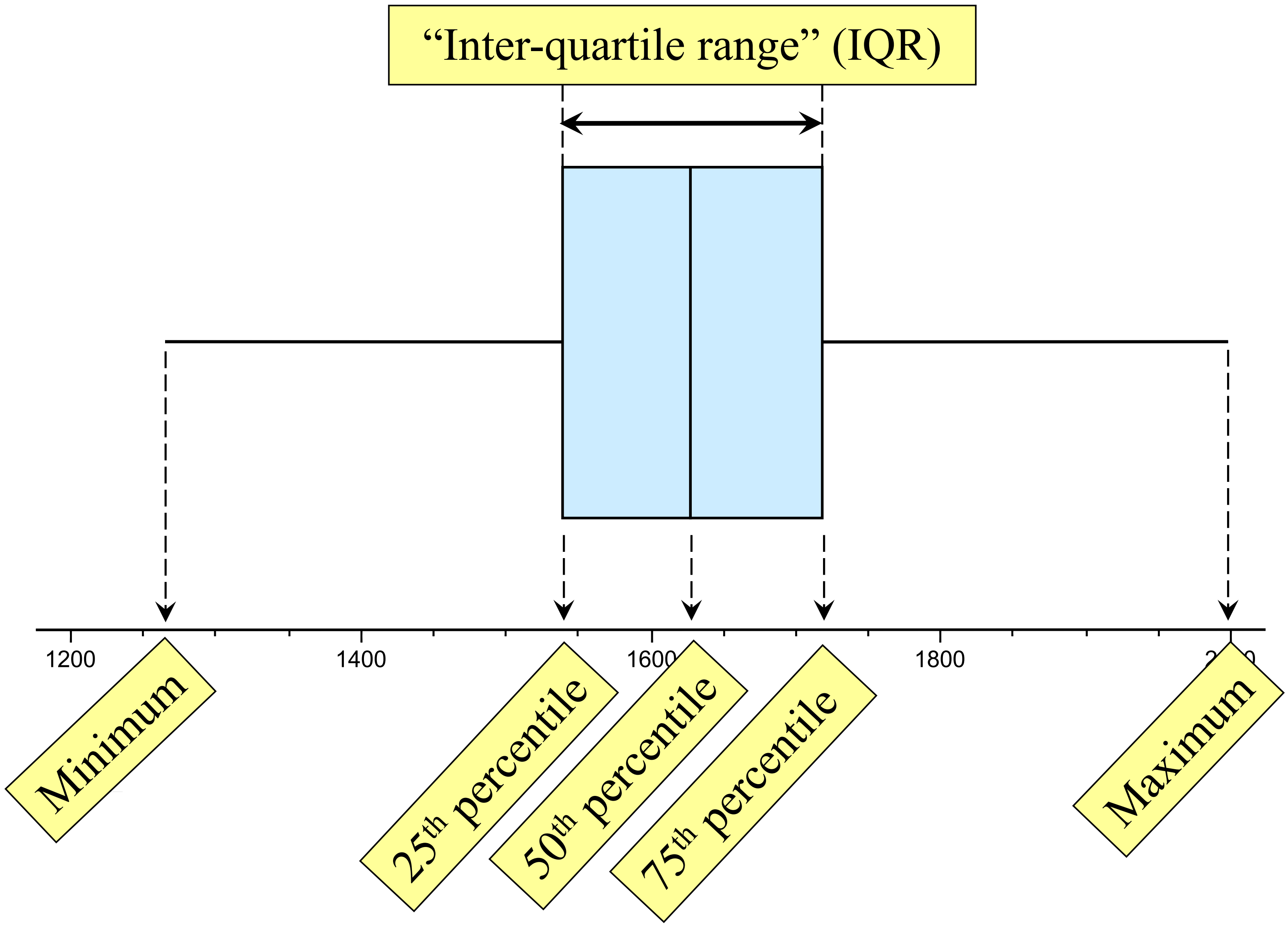
Illustration of 20th and 95th percentiles



Common percentile-based data summary

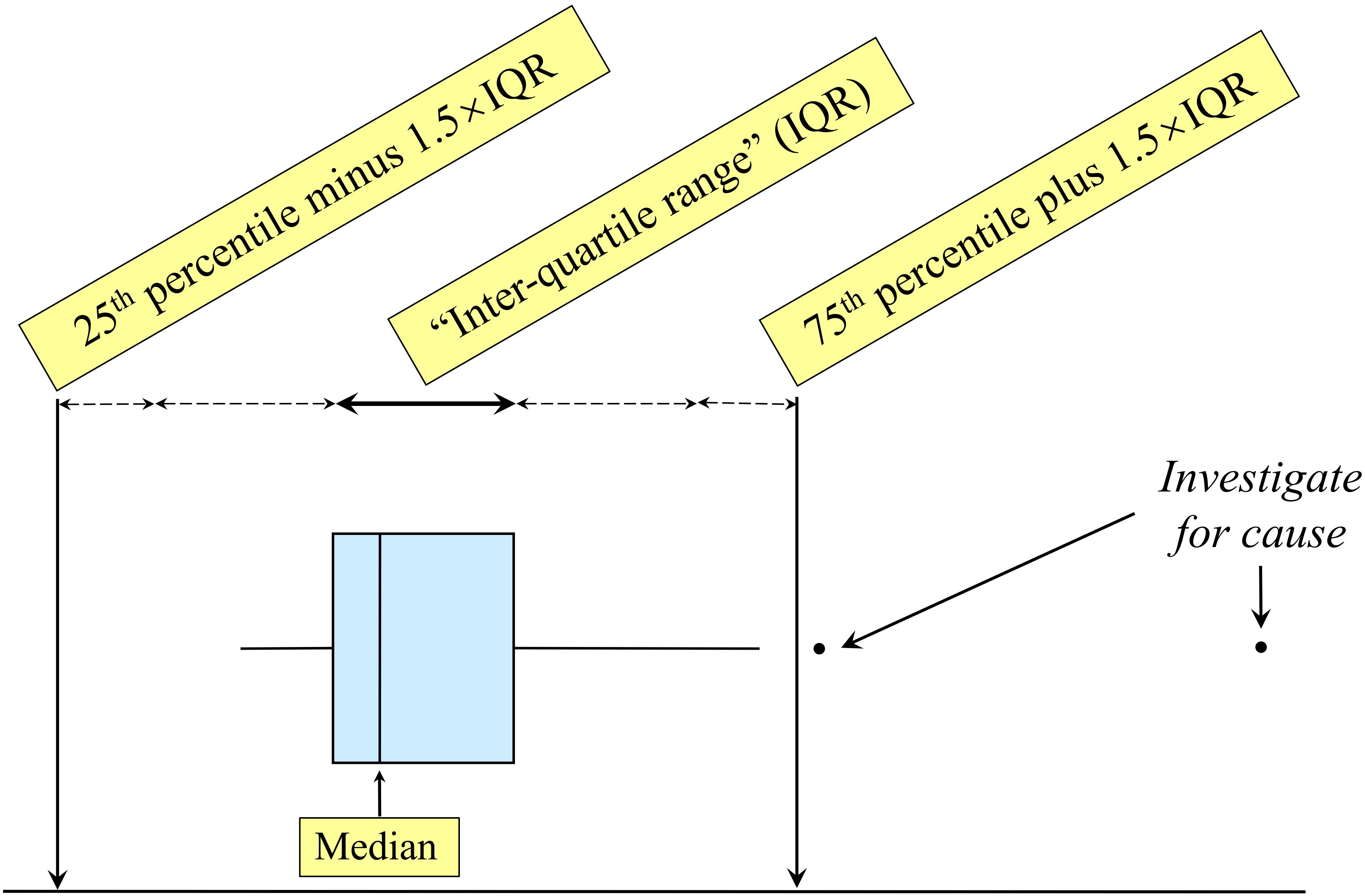


Box-and-whisker plot



"Whiskers" show the minimum and maximum data points, not including outliers (see next slide)

Rule for plotting points separately

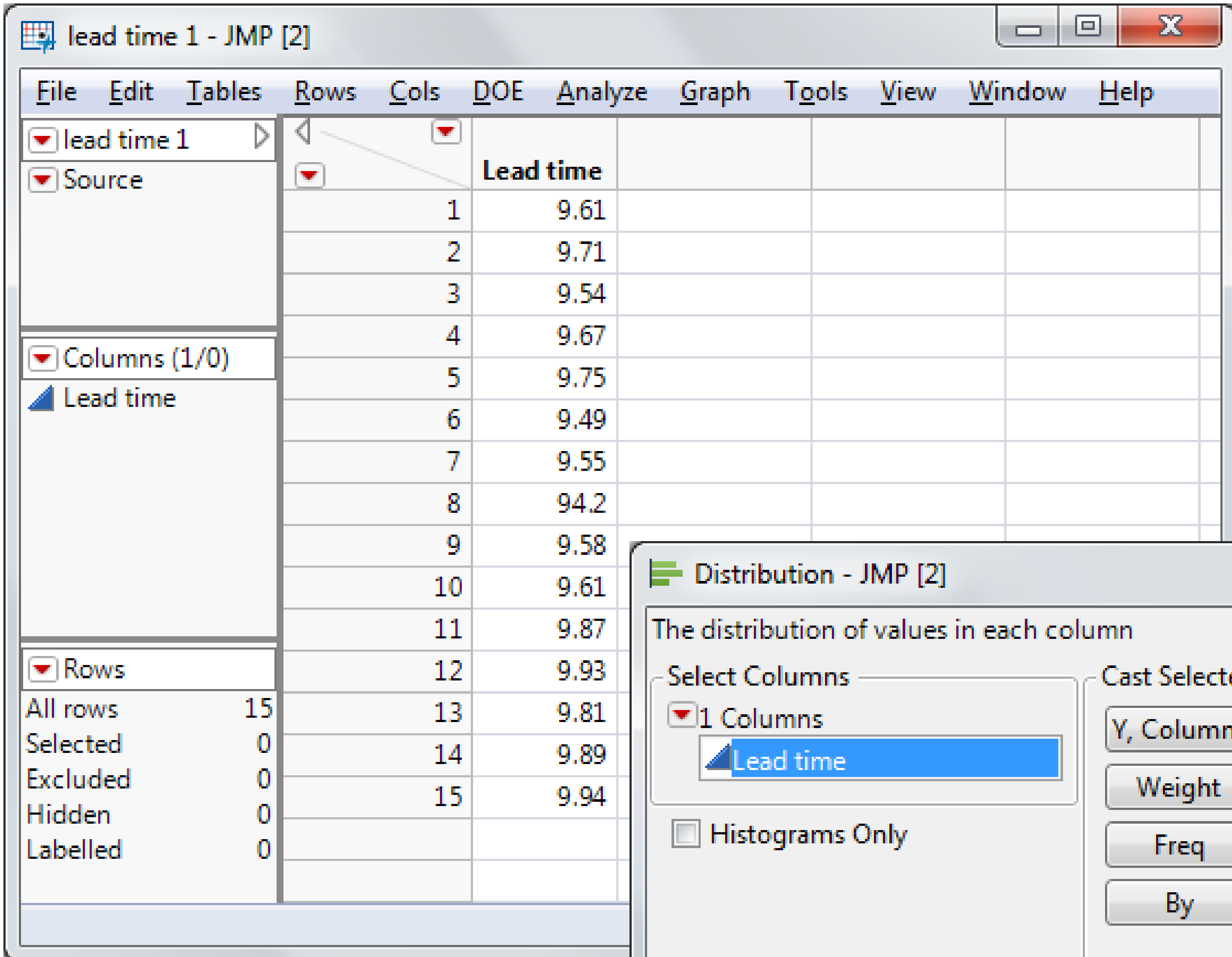


Ends of whiskers are determined by the highest and lowest data points that are inside the calculated ranges.

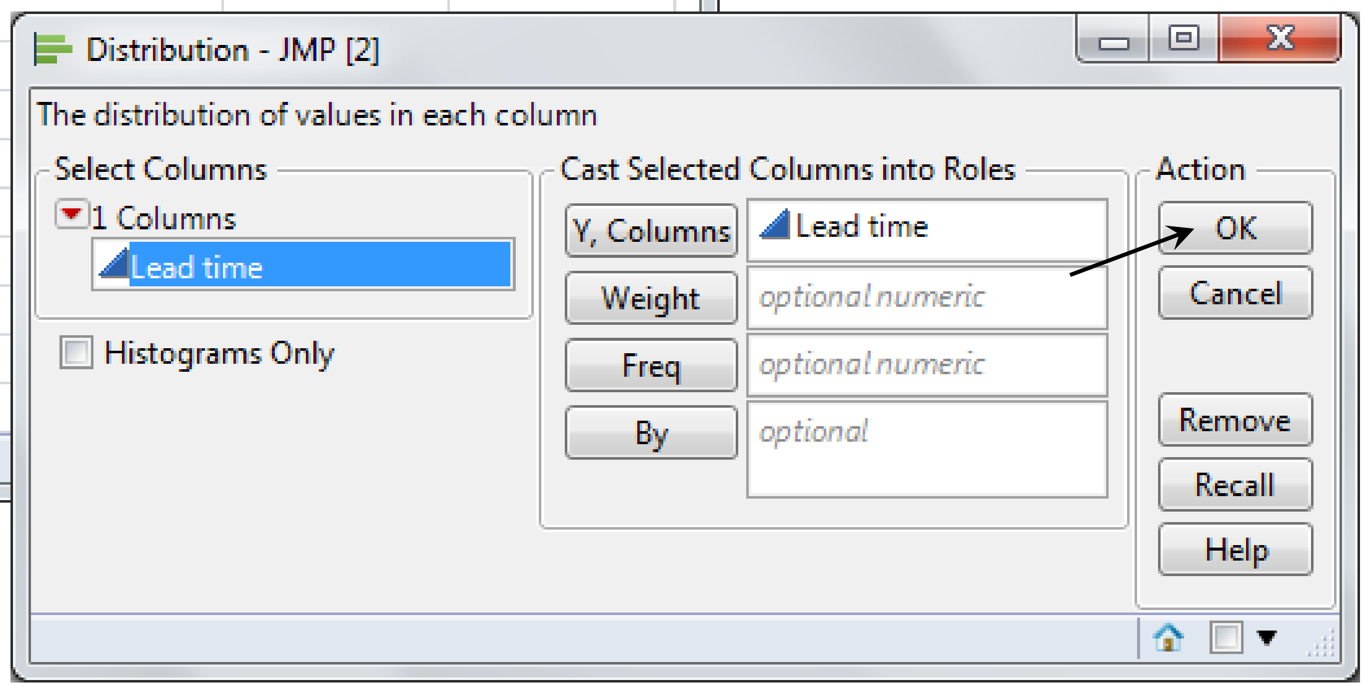
Points plotted separately are outliers, and should be investigated.

JMP distribution analysis

File → Open → All Files → *Data sets* \ *lead time 1* → Open → Import*

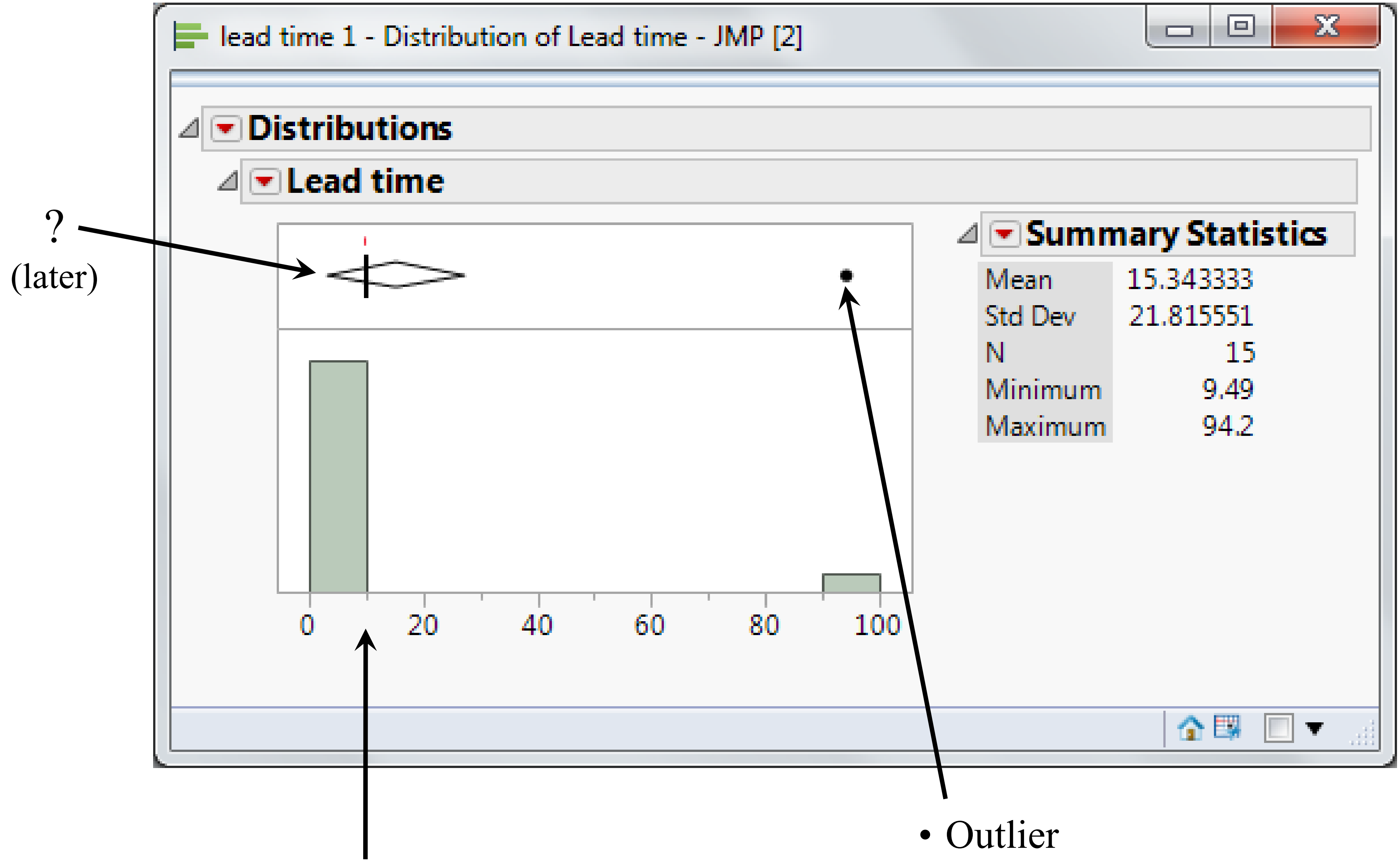


Analyze
↓
Distribution
↓



*Needed only for non-JMP files

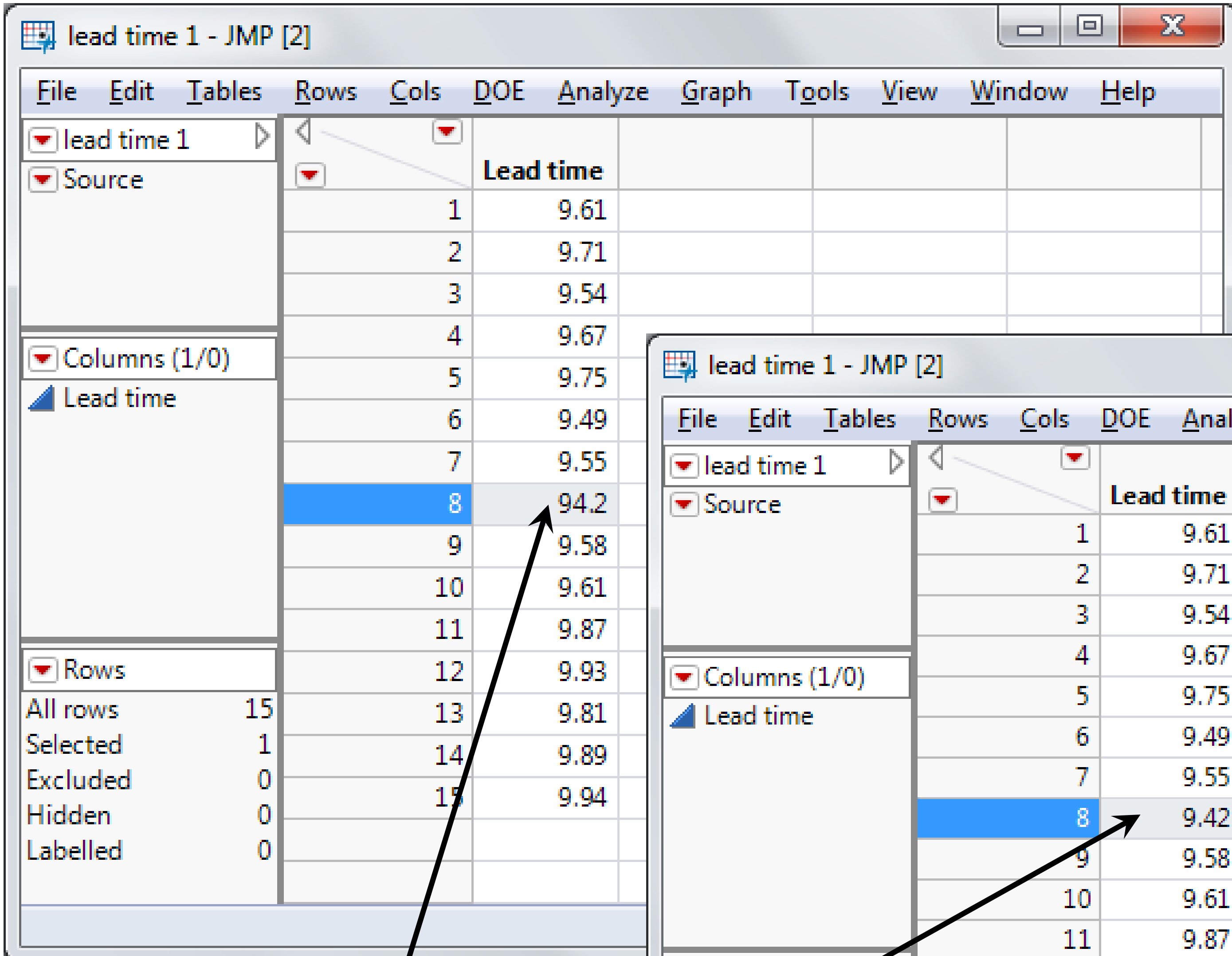
Data validation

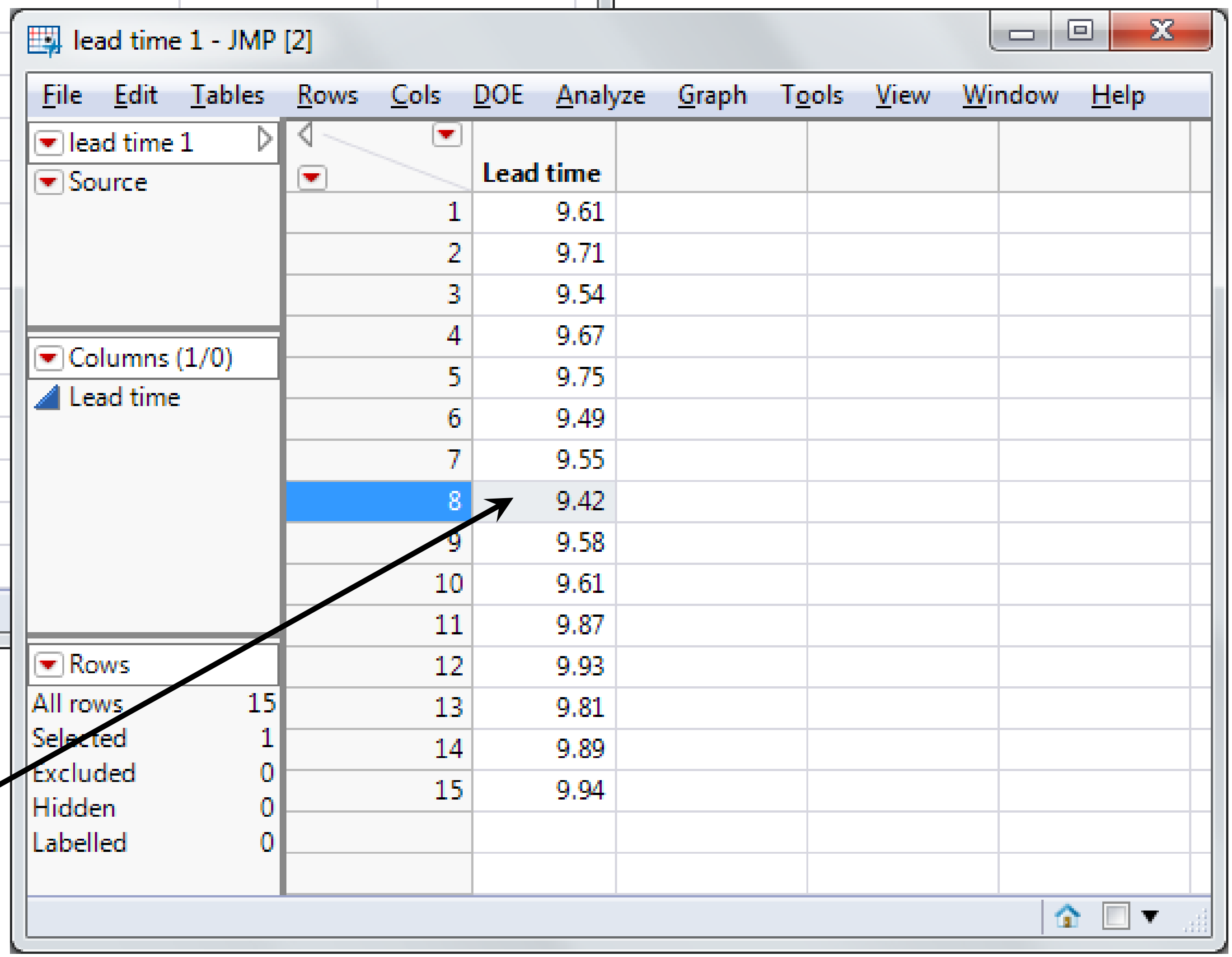


Frequency histogram

- Outlier
- Not always visible in the histogram
- Click on it
- Look in the data table

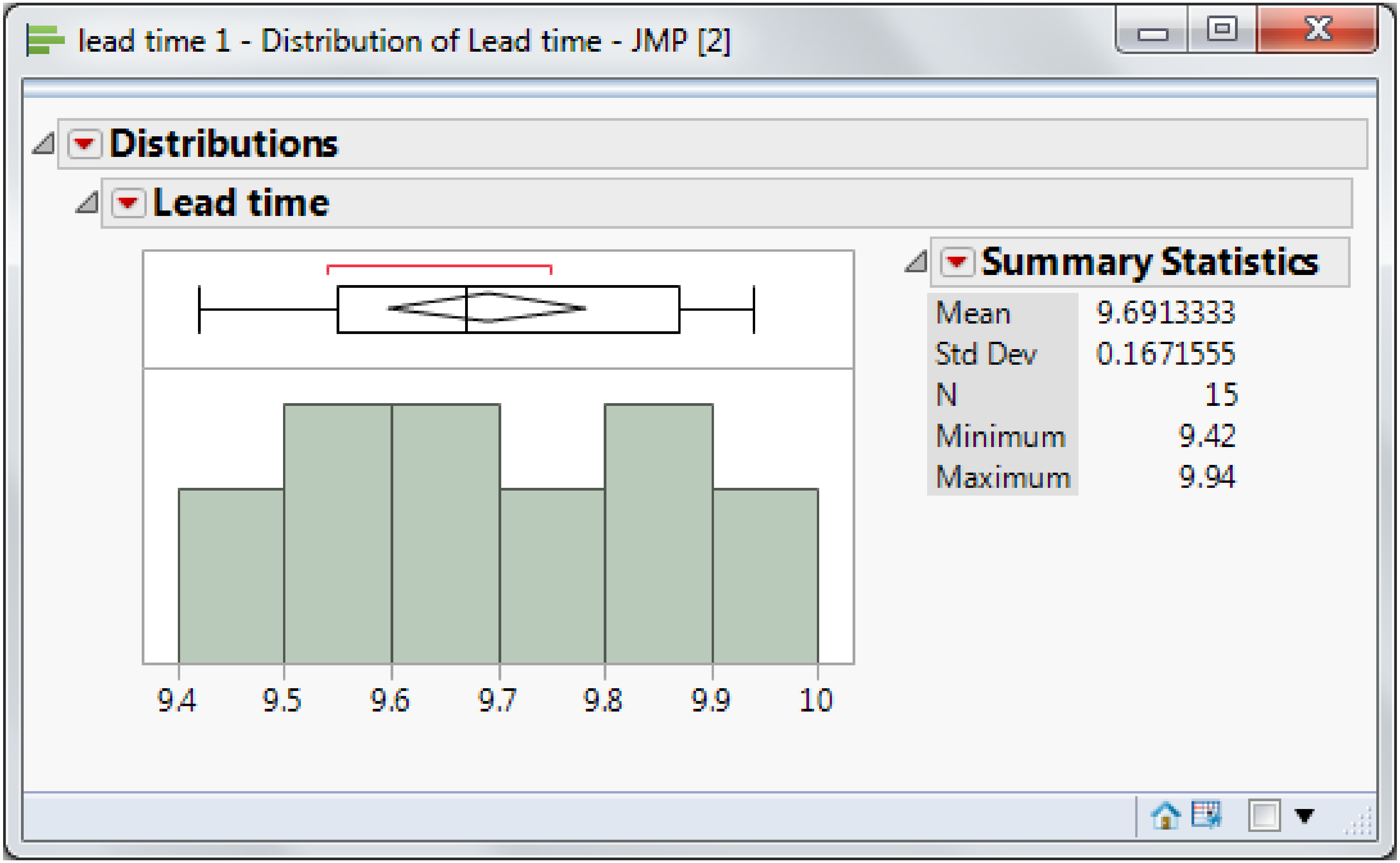
Data validation (cont'd)





- ✓ Data entry error
- ✓ Enter the correct value
- ✓ Go to next slide

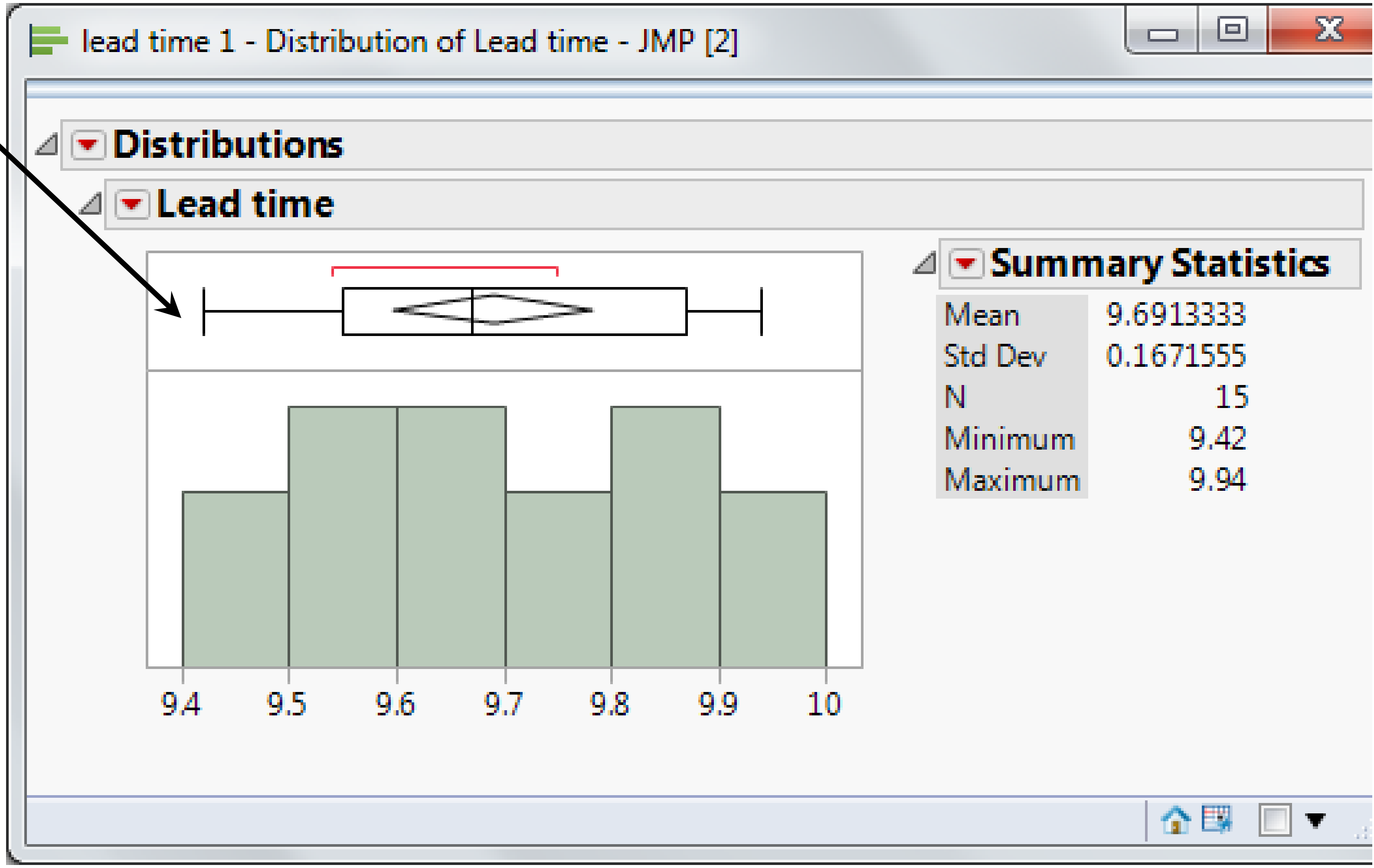
Distribution analysis with data correction



Note the change in the histogram and the summary statistics

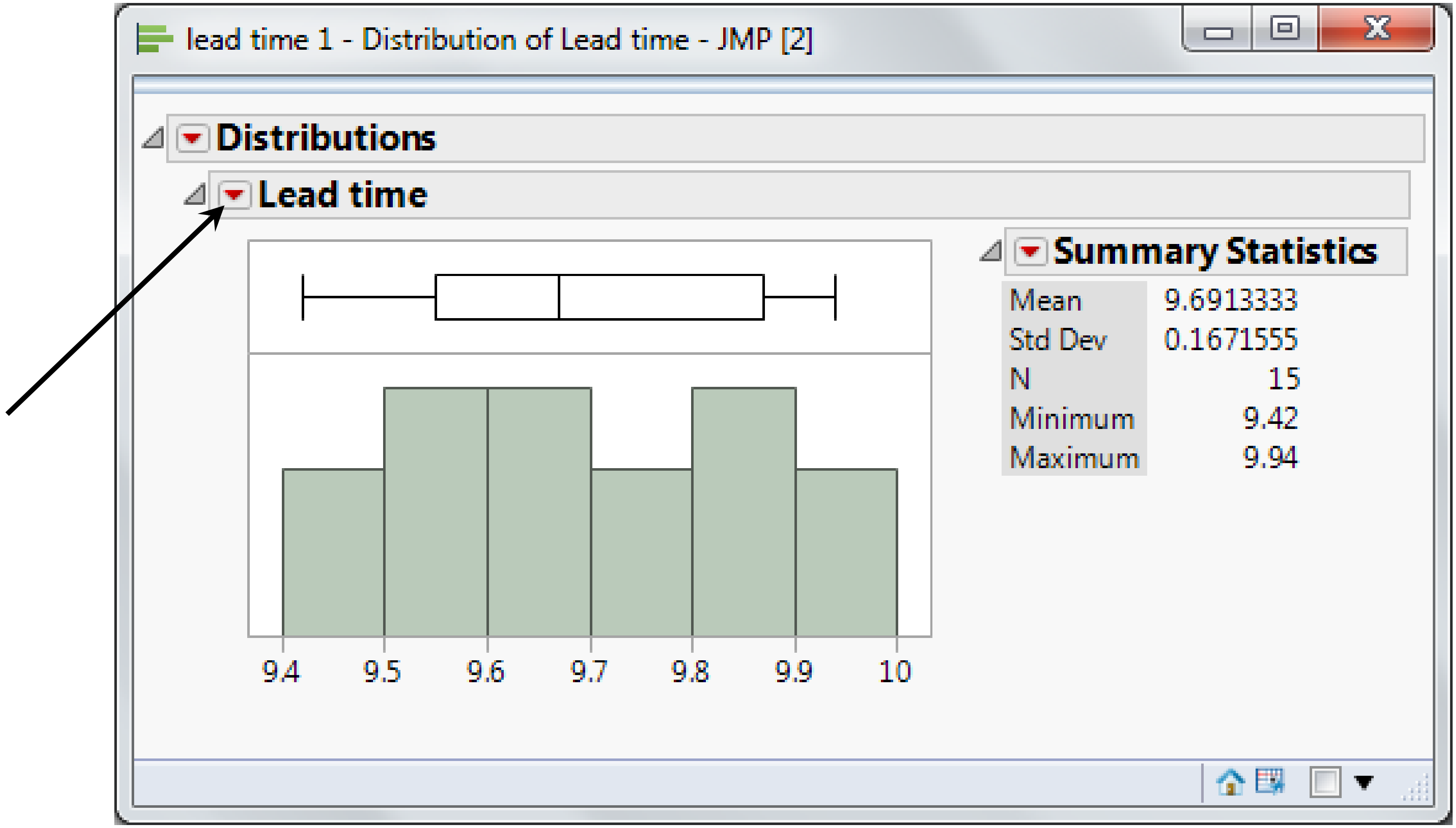
Cleaning up the box plot (optional)

- Right click in this space
- Select *Customize*
- Select *Box Plot*
- Uncheck *Confidence Diamond* and *Shortest Half* → *OK*



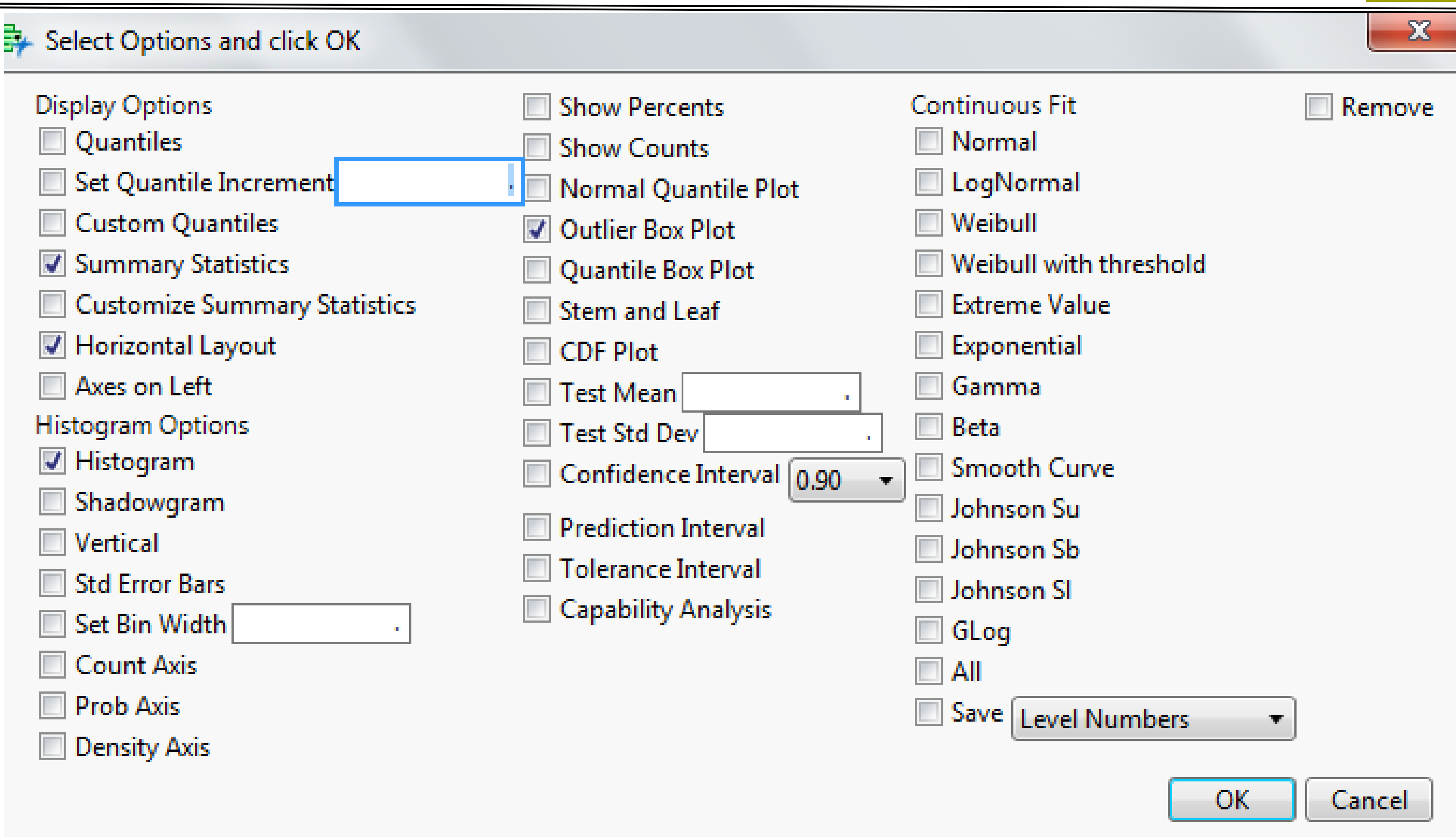
- What remains is the box and whisker plot
- JMP calls it *Outlier Box Plot* because its main purpose in this context is to show outliers

Distribution analysis options



- Click on the red triangle next to *Lead time* while holding down the *Alt* key
- This will show the default analysis options for the *Distribution* platform
- See next slide

Default analysis options (cont'd)

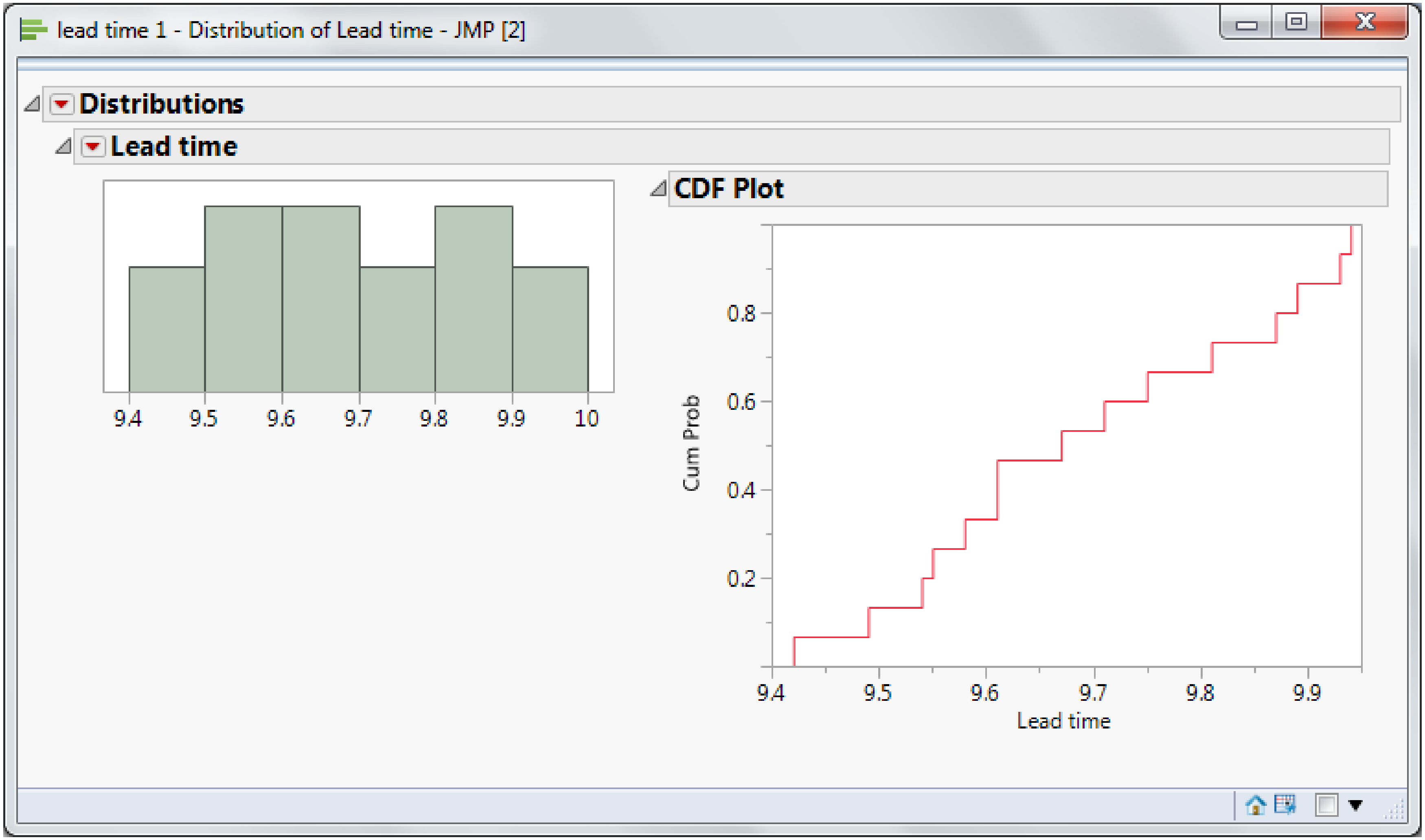


Just for practice:

Uncheck *Summary Statistics* and *Outlier Box Plot* → Check *CDF Plot* → OK

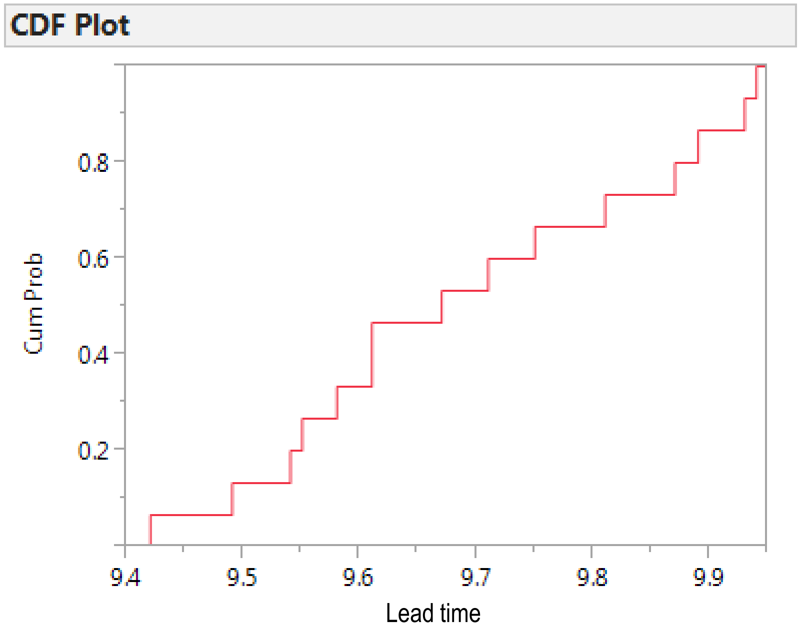
This can also be done by just clicking on the red triangle, but requires more steps.

Cumulative distribution function (CDF plot)

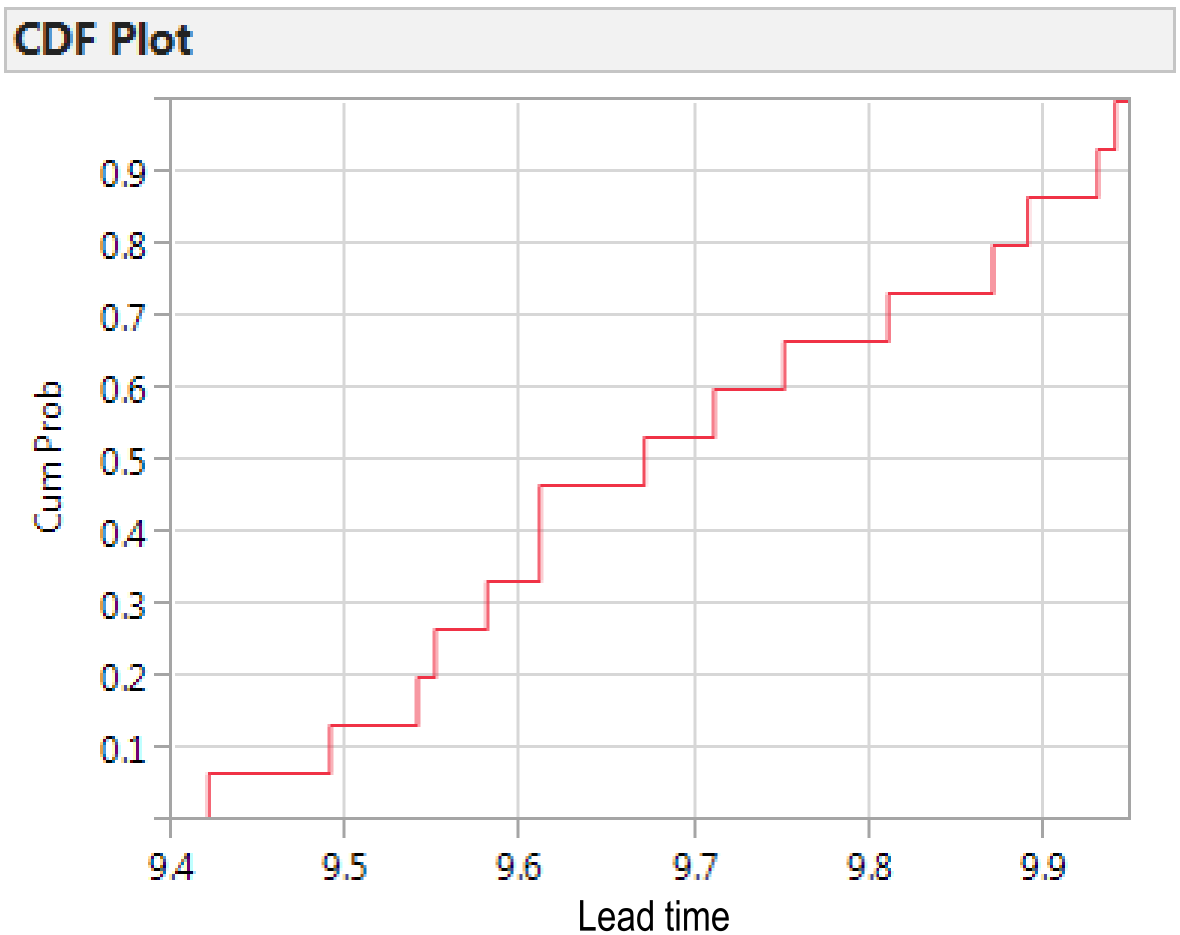


- Plots the proportion of data points ≤ each value in the data set
- The step size at each data value is usually 1/N, where N is the sample size
- If the same value occurs twice in the data set, the step size there is 2/N

Modifying JMP plots

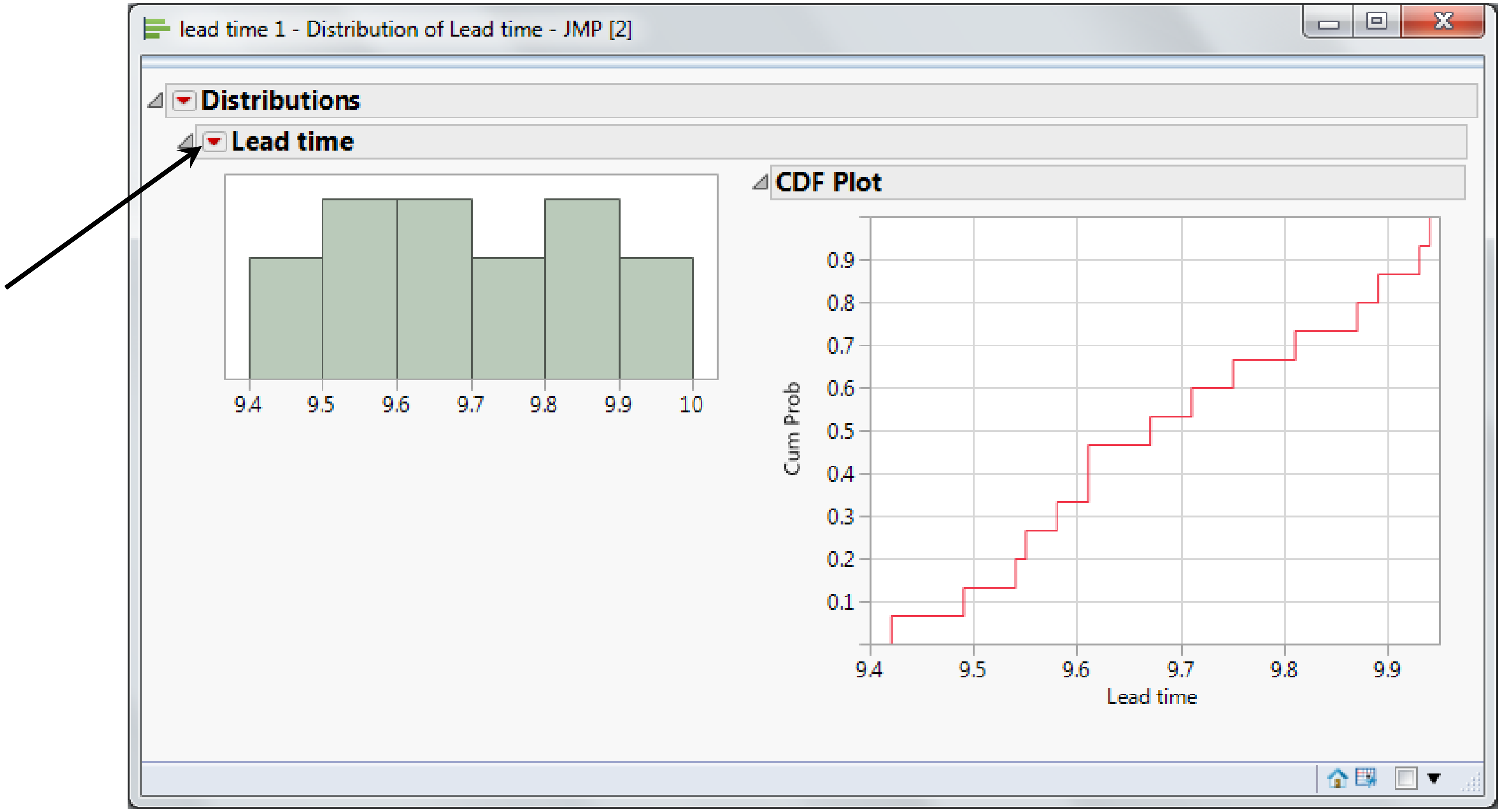


1. Double click on a number on the Y axis
   - change *Increment* to 0.1
   - check *Major Grid Lines*
   - uncheck *Minor Tick Mark*
   - Set Minimum to 0 and Maximum to 1
   - OK



2. Double click on a number on the X axis
   - check *Major Grid Lines*
   - uncheck *Minor Tick Mark*
   - OK

Calculating percentages

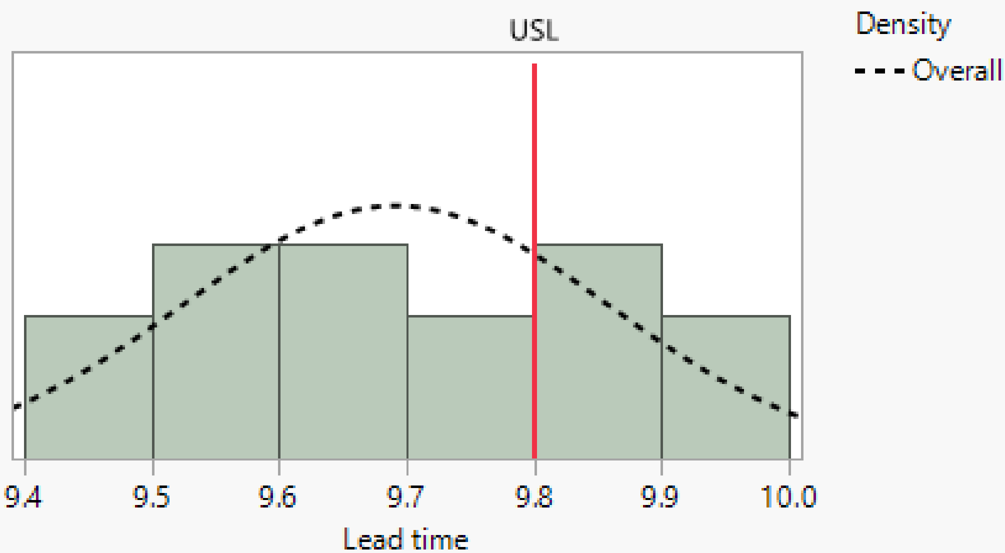


- Suppose we want to know the percentage of data points exceeding 9.8
- Click the *Lead time* red triangle → select *Process Capability*
- Enter 9.8 for the *Upper Spec Limit* → click OK

Percentages (cont'd)

Lead time Capability

Histogram



Overall Sigma Capability

Index	Estimate	Lower 95%	Upper 95%
Cpk	0.217	0.027	0.400
Cpu	0.217	0.027	0.400

Nonconformance

Portion	Observed %	Expected Within %	Expected Overall %
Above USL	33.3333	15.4871	25.7816
Total Outside	33.3333	15.4871	25.7816

Nonconformance shows:

- Observed percent out-of-spec
- Expected (predicted), based on the Normal distribution

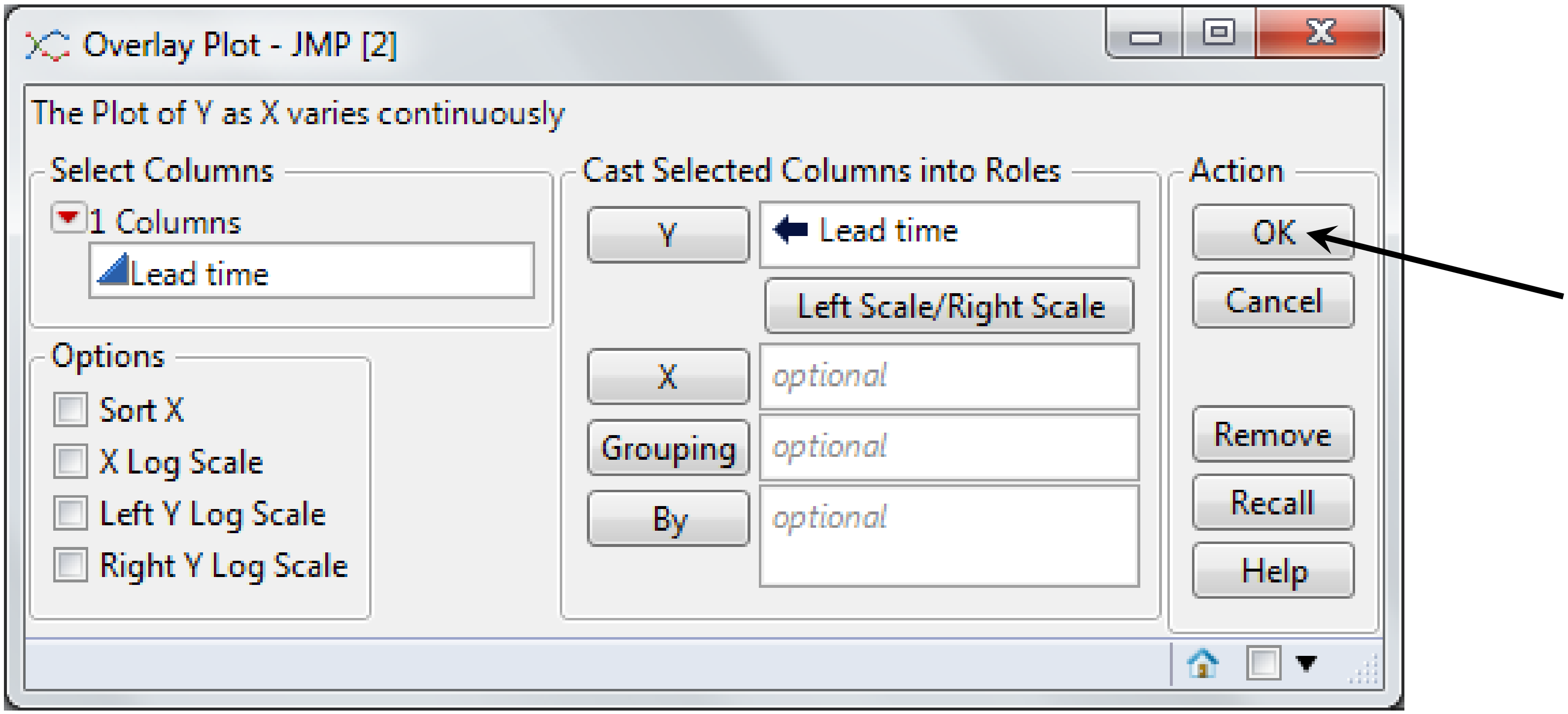
Capability indices are calculated:

- *Within Sigma Capability* can be used when small samples are collected, such as for an Xbar-R chart
- Turn this off by clicking on the red triangle next to Lead time Capability
- Turn off the Within curve on the histogram by clicking on the red triangle next to Histogram

We will cover distribution fitting in the next section

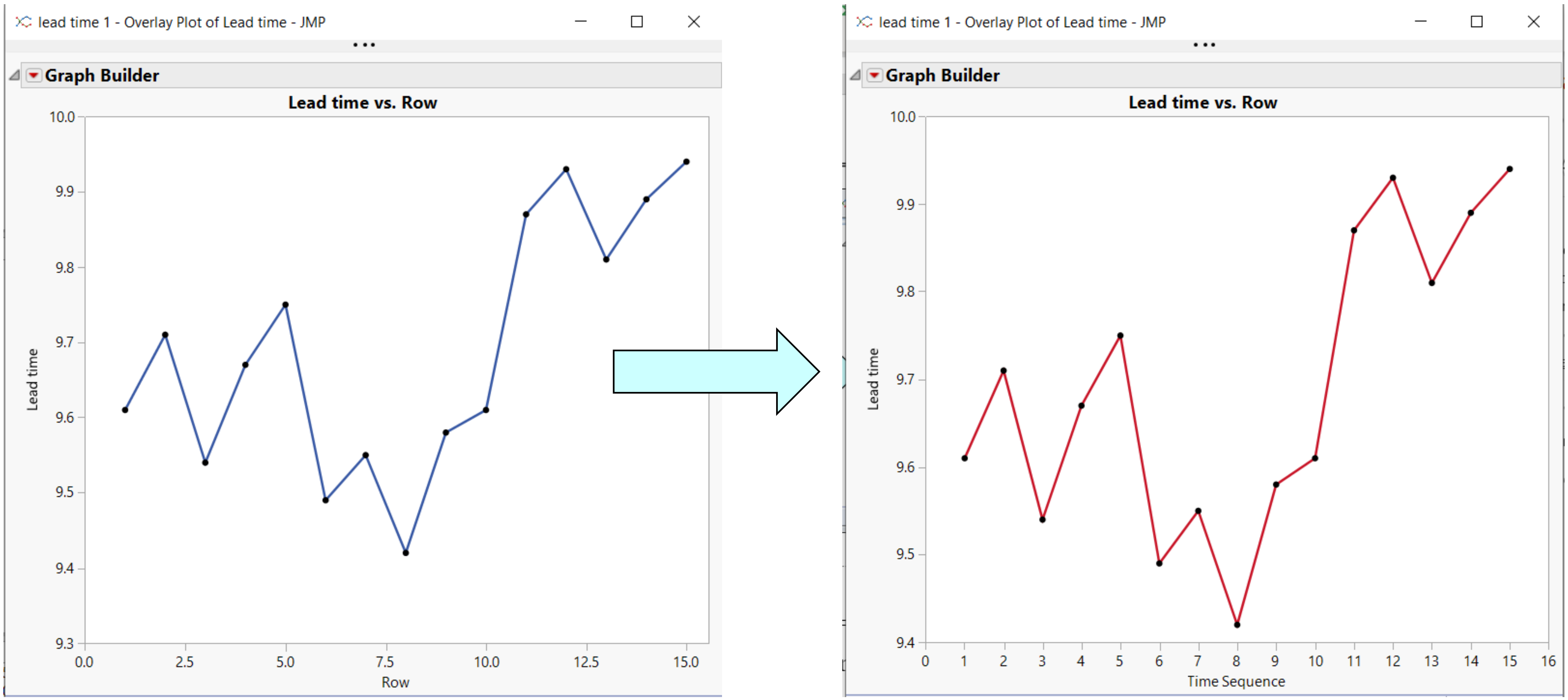
Plotting data in time sequence

Graph → Legacy → Overlay Plot



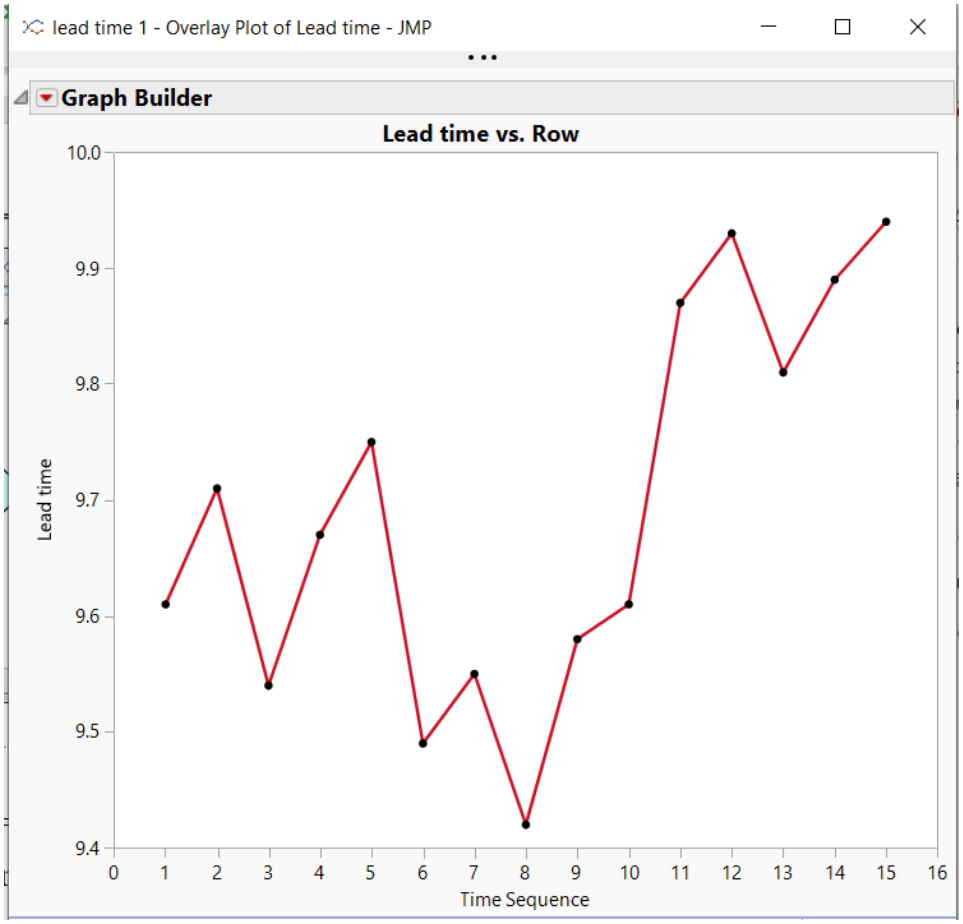
- You can have different left and right scales for plotting multiple Y variables
  - Cast both Y variables into Y
  - Select the one you want to display on the secondary (right) scale
  - Click Left Scale/Right Scale.
  - Arrows point to the Y-scale for each Y variable
- A date, time, or other sequencing variable could be cast into X

Overlay plot (cont'd)



- Modify the chart as follows:
  - Double Click X-Axis: Minimum = 0, Maximum = 16, Increment = 1, Dec = 0
  - Double Click on Y-Axis: Minimum = 9.4
  - Right Click on Chart: Customize > Line > Line Color > Red
  - Double Click on X-Axis Title: Change "Row" to "Time Sequence"

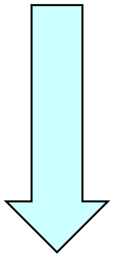
Overlay plot (cont'd)

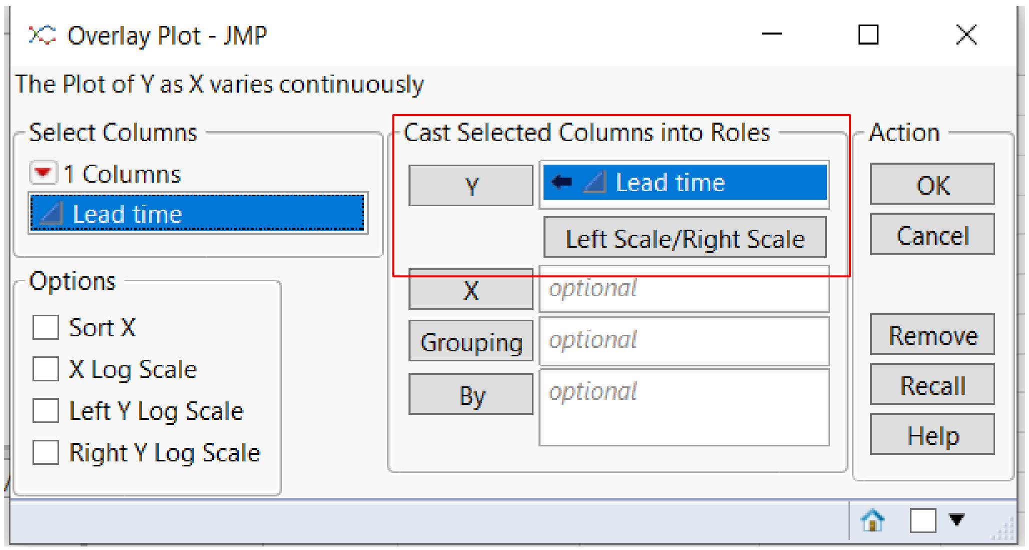


- Good way to look for assignable cause patterns versus their time sequence

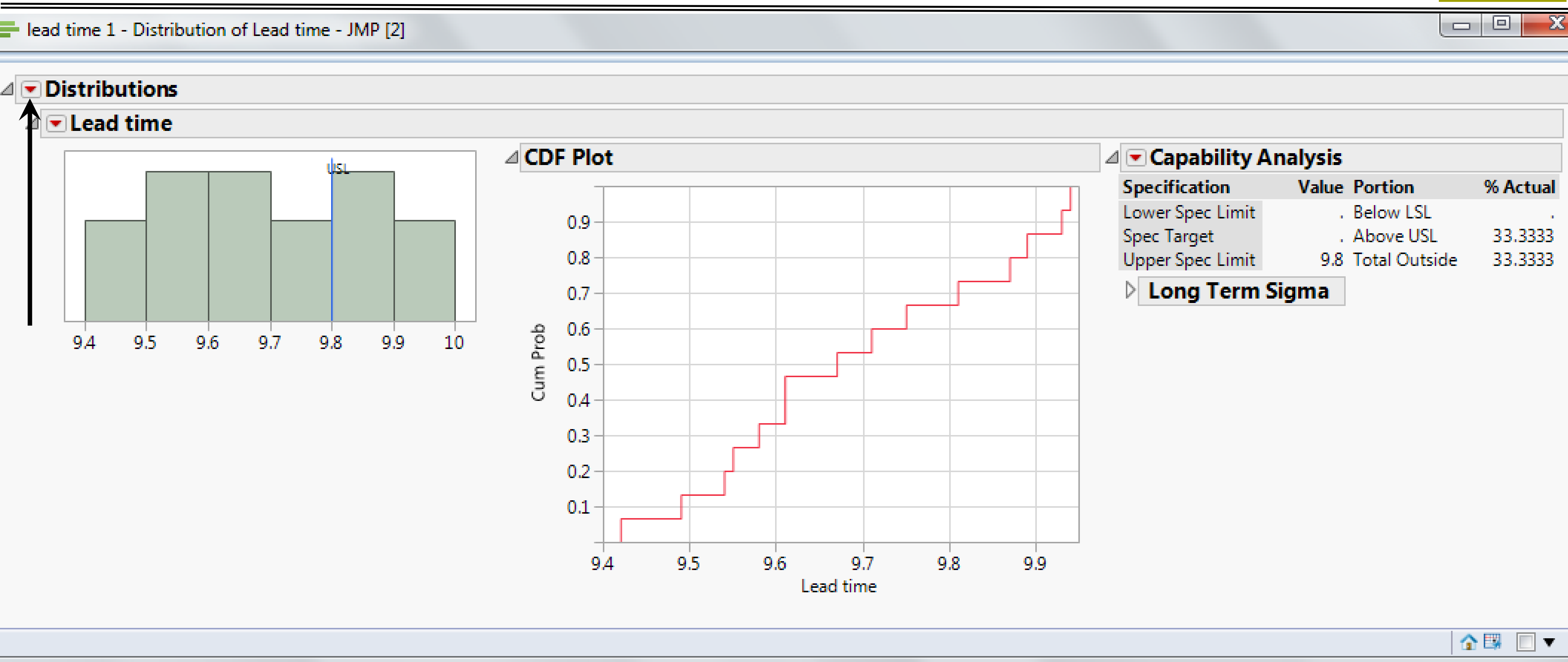
- Same as a line chart in Excel

- Overlay plot can be used to display different data sets on different Y-Axis



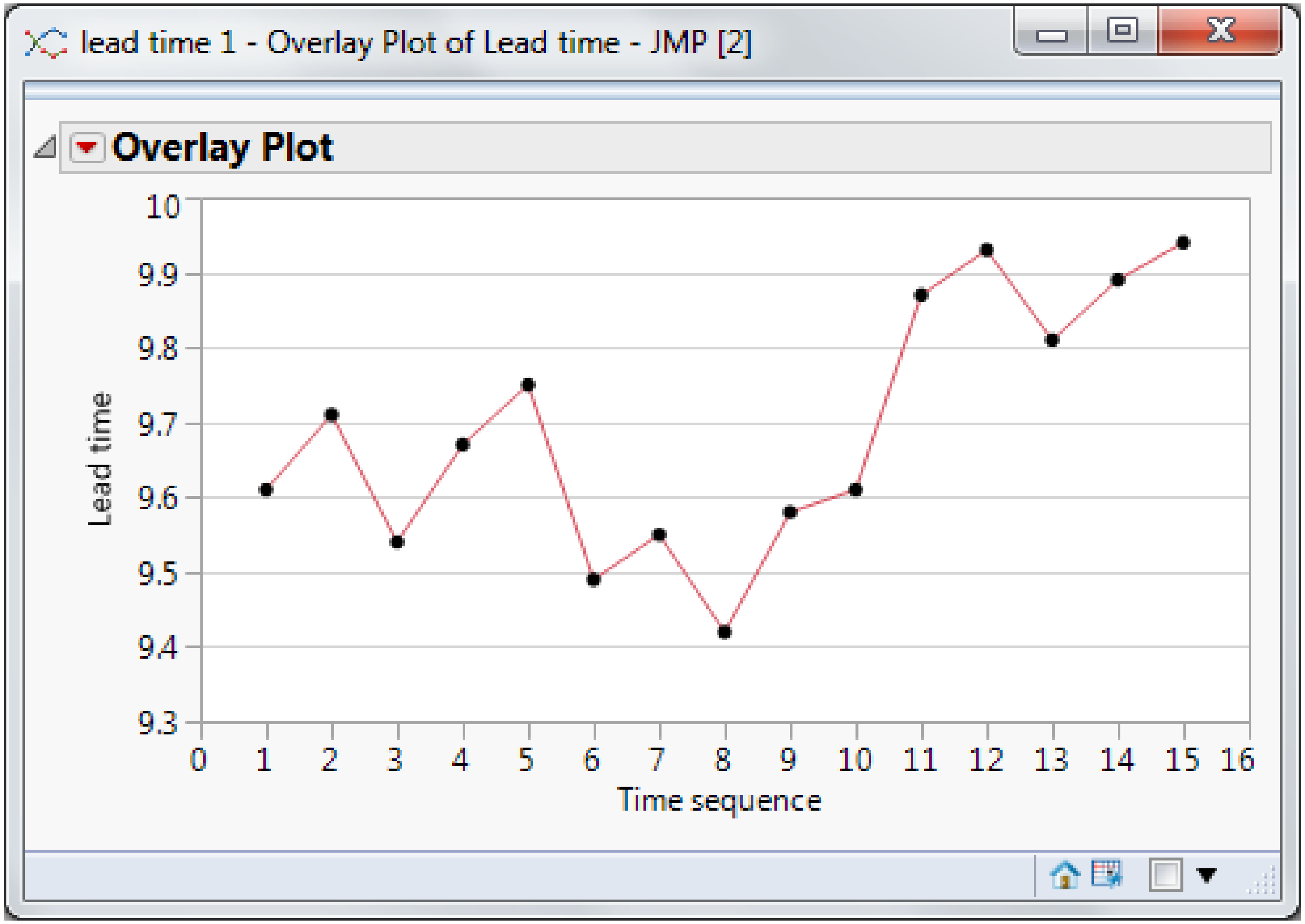


Saving your analyses and data table



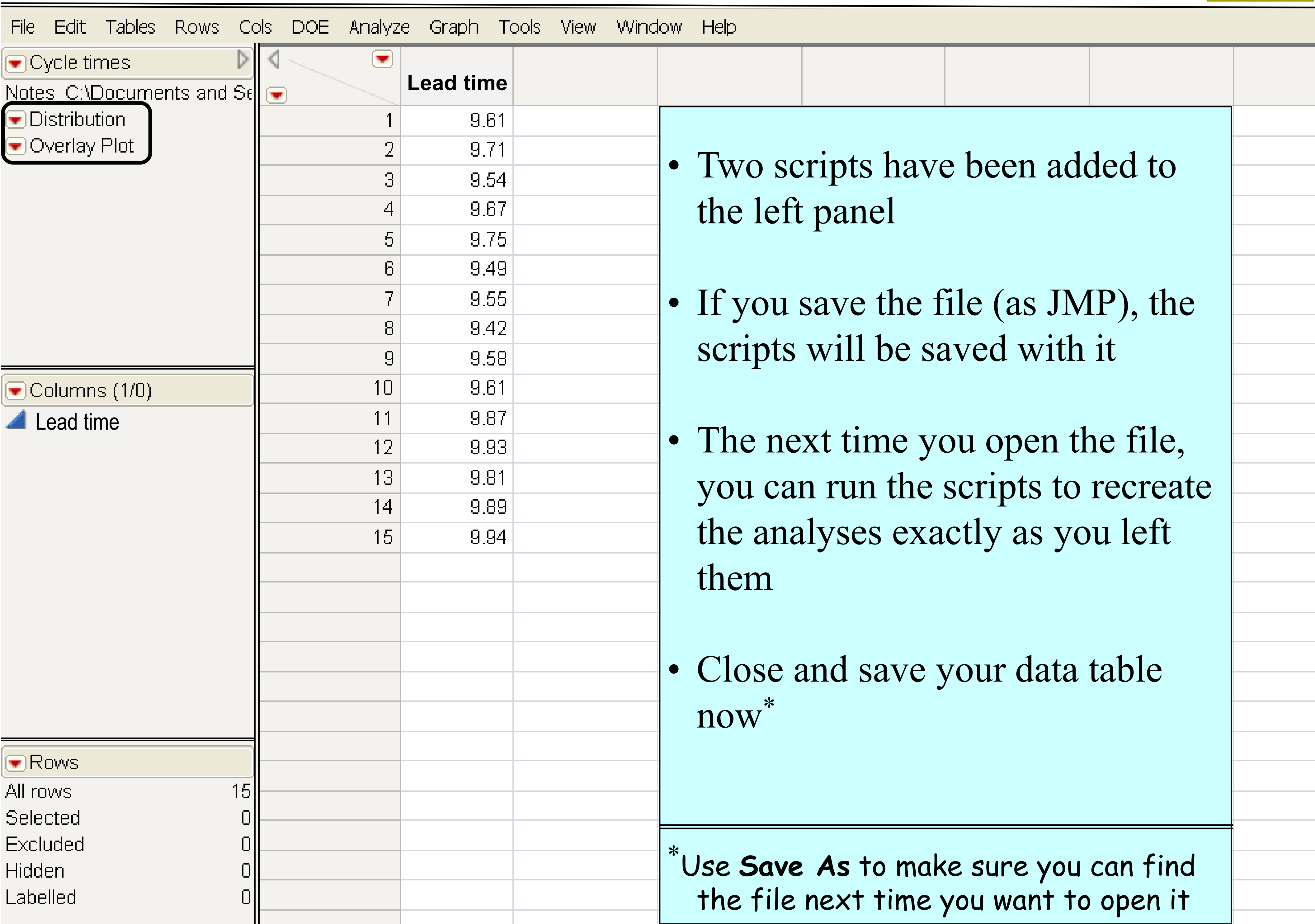
- Click on the thumbnail for the distribution analysis at the bottom of the data table
- Click the red triangle next to *Distributions*
- *Save Script* → *To Data Table* → Name: *Distribution* → OK

Saving things (cont'd)



- Click on the thumbnail for your overlay plot, click the red triangle next to *Overlay Plot*
- *Save Script* → *To Data Table* → Name: *Overlay Plot* → OK
- Go to your data table

Saving things (cont'd)



- Two scripts have been added to the left panel
- If you save the file (as JMP), the scripts will be saved with it
- The next time you open the file, you can run the scripts to recreate the analyses exactly as you left them
- Close and save your data table now*

*Use **Save As** to make sure you can find the file next time you want to open it

Exercise 2.1

Open *Data sets \ quotation process*. Perform the following data analysis tasks for the variable *TAT* (turnaround time).

(a) Run a distribution analysis. Note the large number of points plotted separately on the outlier box plot. This pattern is common with asymmetric “ski slope” distributions that pile up near zero. These points are *not* assignable causes, so they would not be investigated or removed.

(b) Record the average, standard deviation, sample size, minimum, maximum and median.

(c) Turn off the outlier box plot.

(d) Find the % of data points exceeding 3.

(e) Turn off the Within Sigma Capability.

(f) Save your analysis script. Close and save the data table.

Exercise 2.2

Data sets \ DI water. Perform the following data analysis tasks for the variable *Resistivity*.

(a) Create an overlay plot. You should see something that suggests bad data (stretch the graph if necessary). Use your mouse to draw a box around the suspicious data points. Right click in an uninhabited area of the plot, select *Row Hide and Exclude*.

(b) Run a distribution analysis. Record the average, standard deviation, sample size, minimum, and maximum.

(c) Turn off the outlier box plot.

(d) Find the % of data points falling below 1500.

(e) Turn off the Within Sigma Capability.

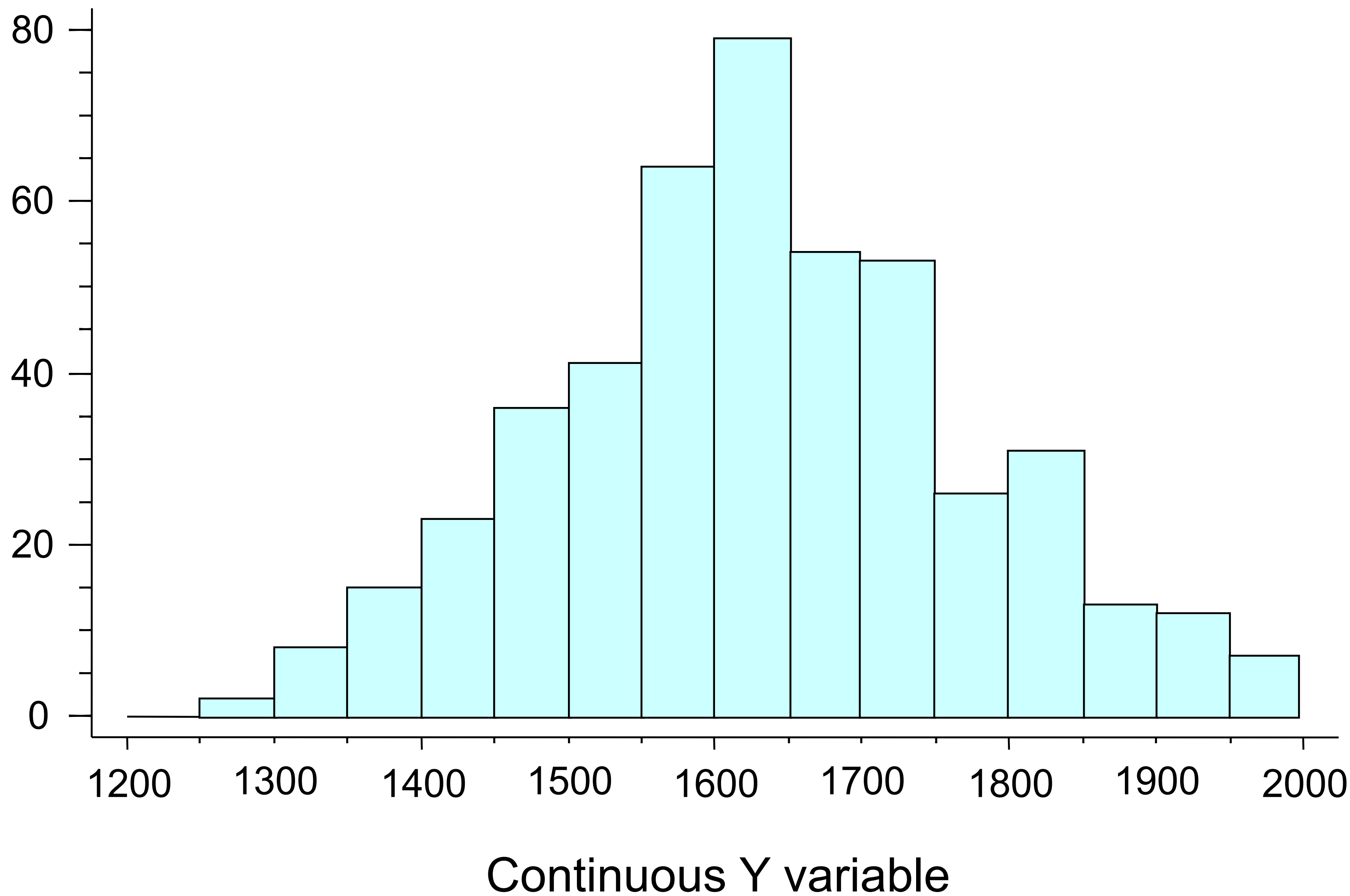
(f) Save your analysis scripts. Close and save the data table.

3 Fitting and Using Distributions

- Distribution curves
- Checking goodness of fit
- JMP examples
- Fitting and using the Normal distribution
- Fitting and using the Lognormal distribution
- Finding the best fitting distribution(s)
- Using the best fitting distributions(s)

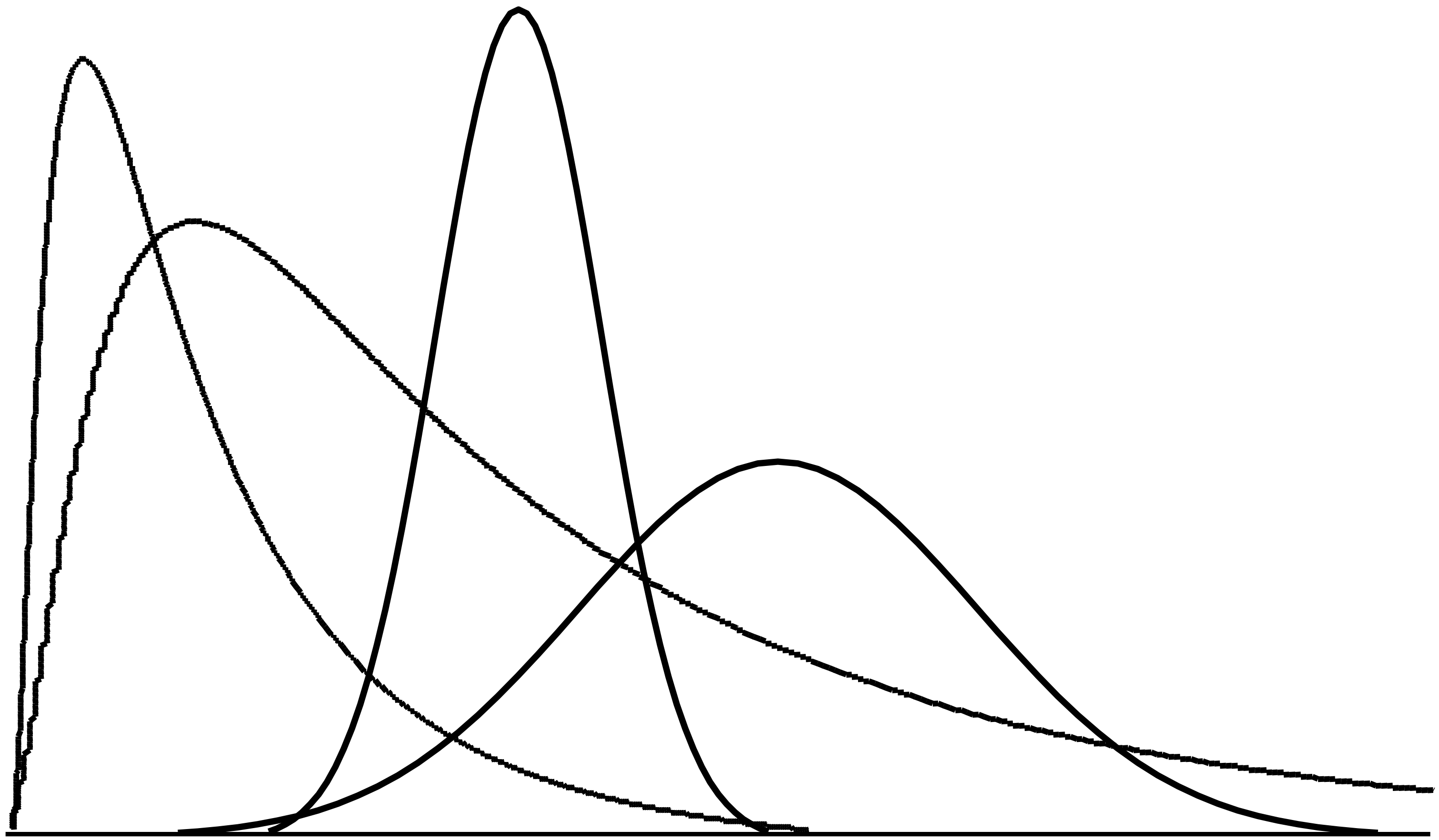
Frequency histogram

A description of the data



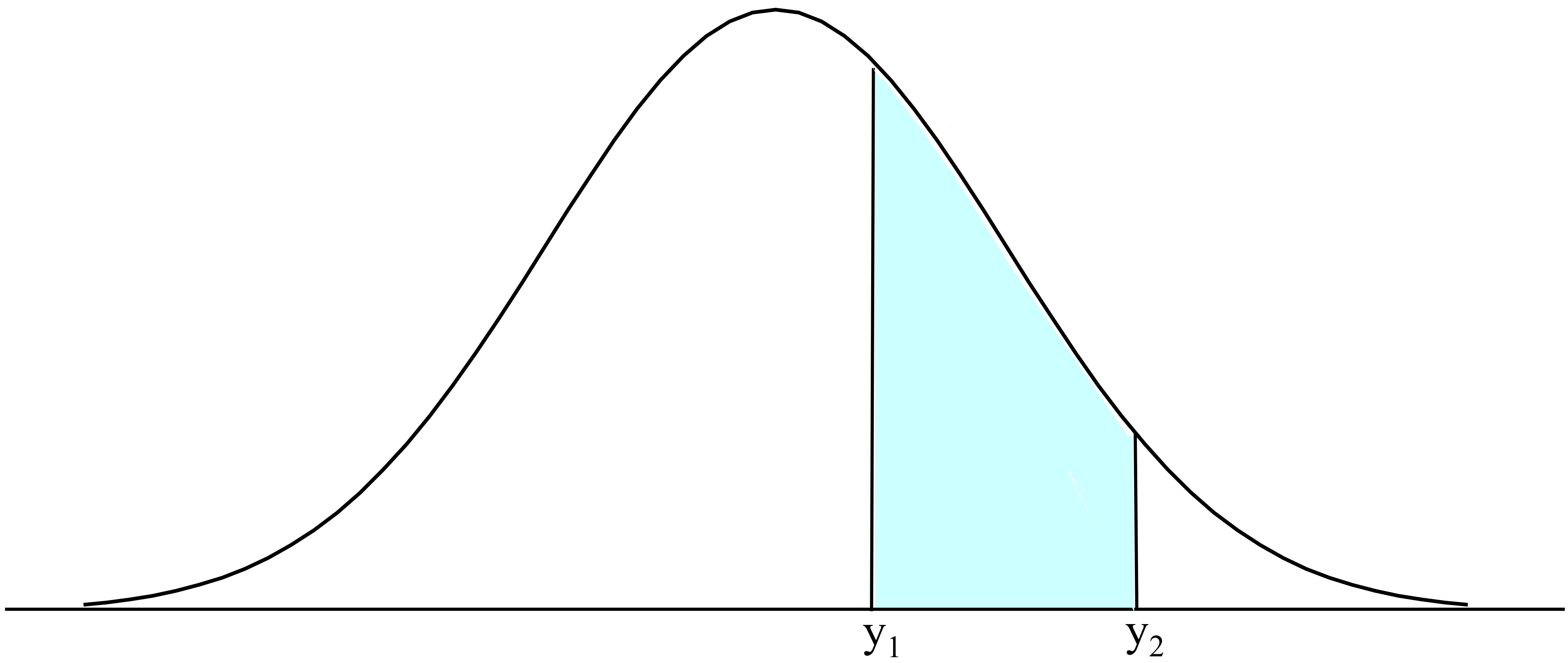
Distribution curves

Possible descriptions of the population



Distribution curves (cont'd)

$$\textit{Area under the curve between } y_1 \textit{ and } y_2$$
$$= \textit{ \% of the population with } y_1 < Y \leq y_2$$

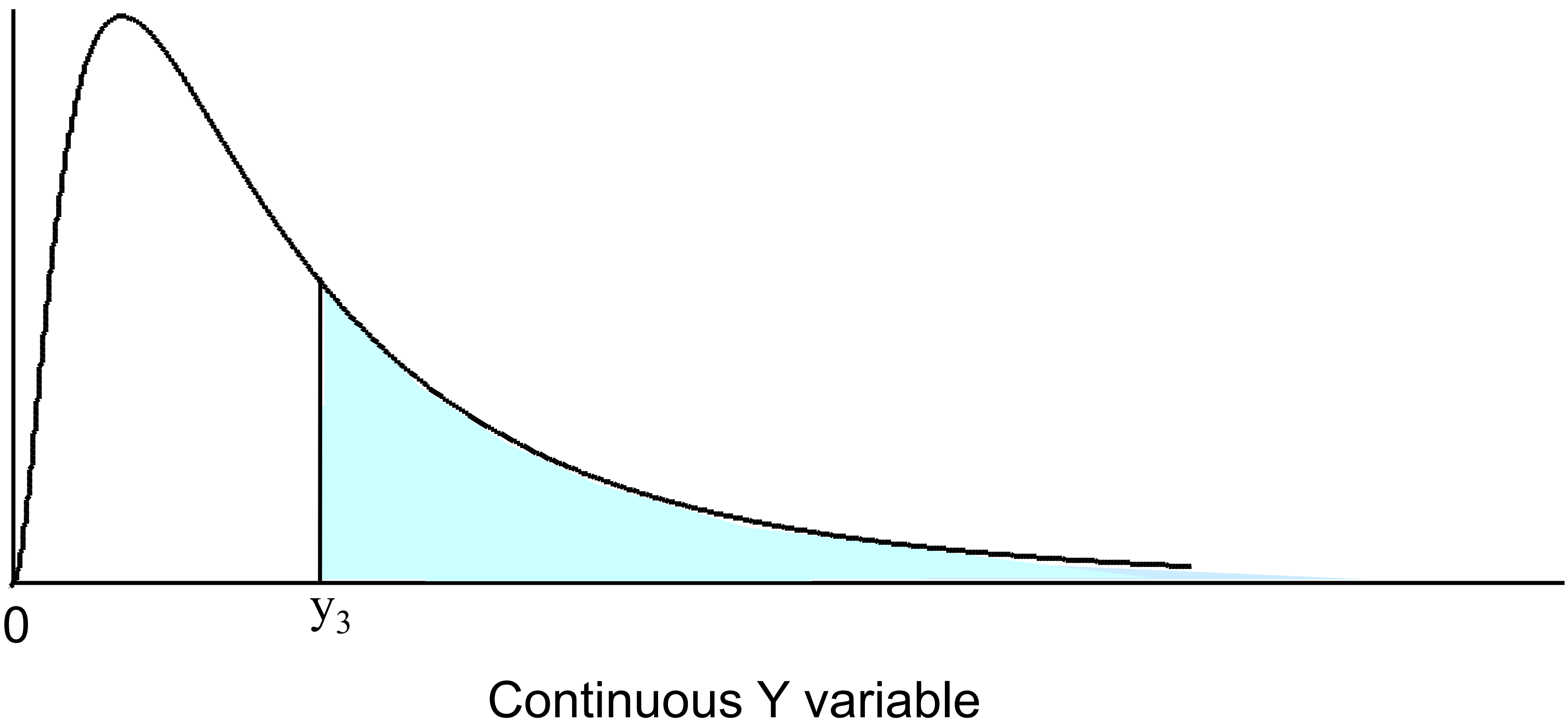


Continuous Y variable

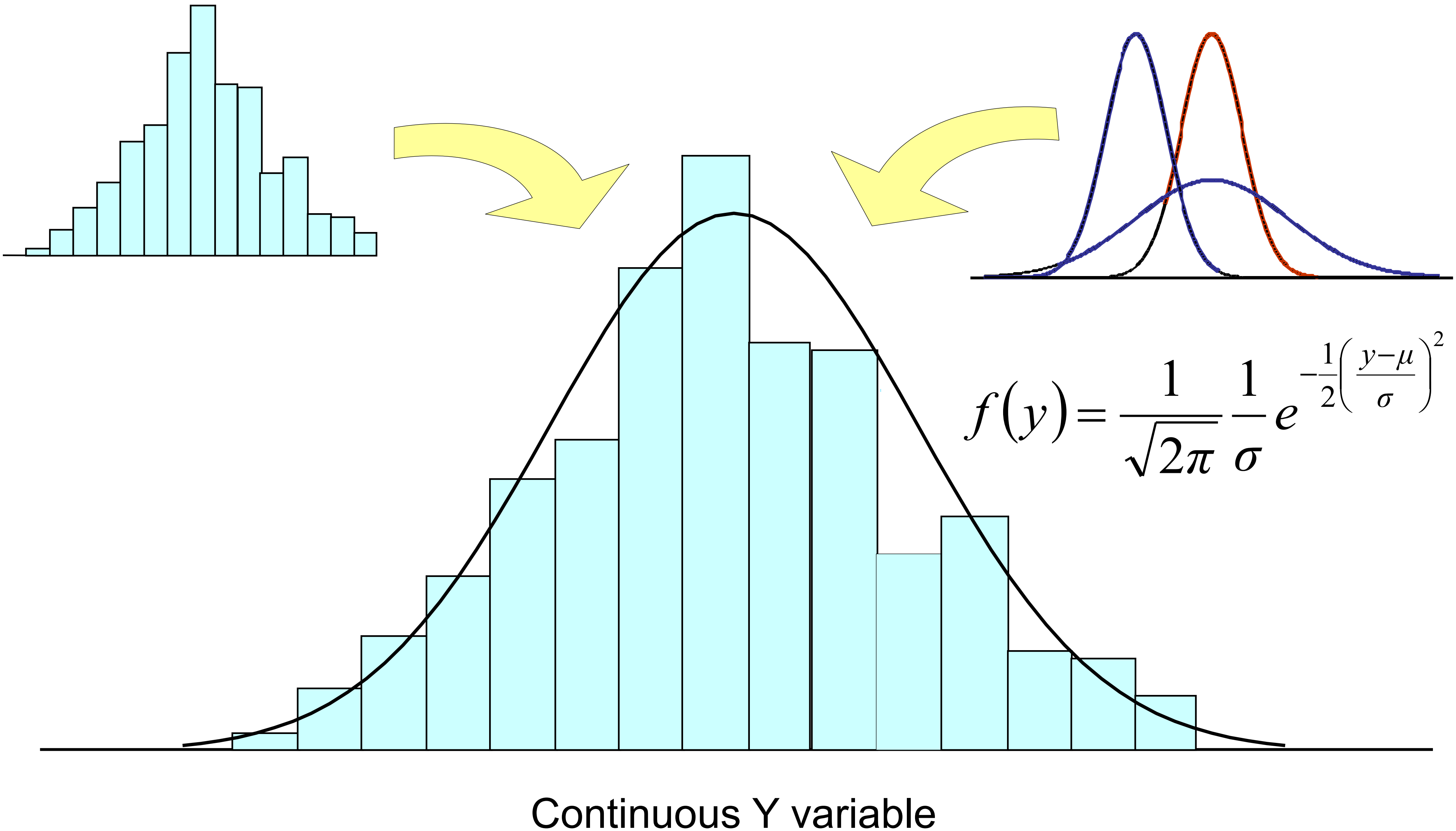
Distribution curves (cont'd)

Area under the curve to the right of y_3

$= $ *% of the population with* $Y > y_3$



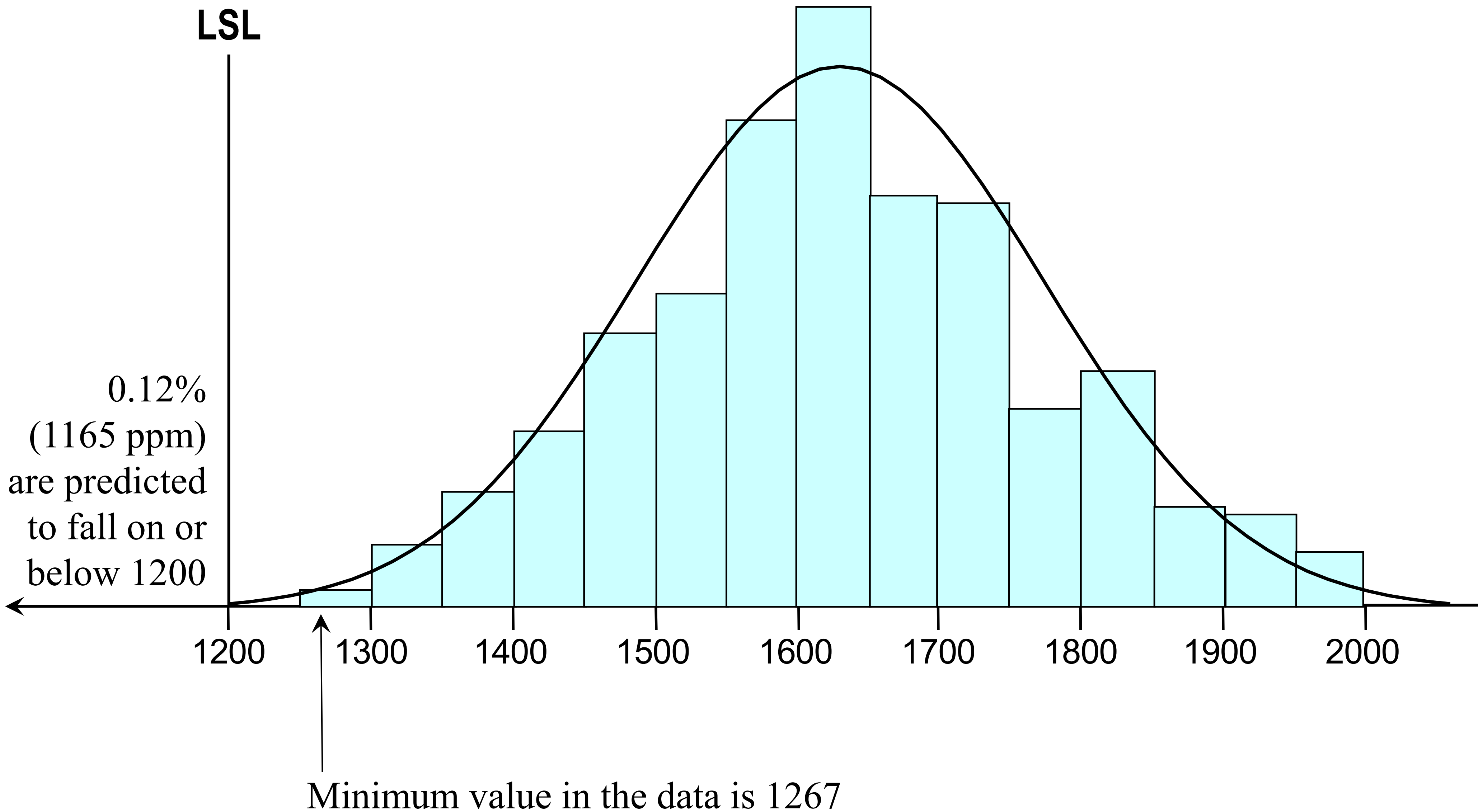
Fitting a distribution curve to the data



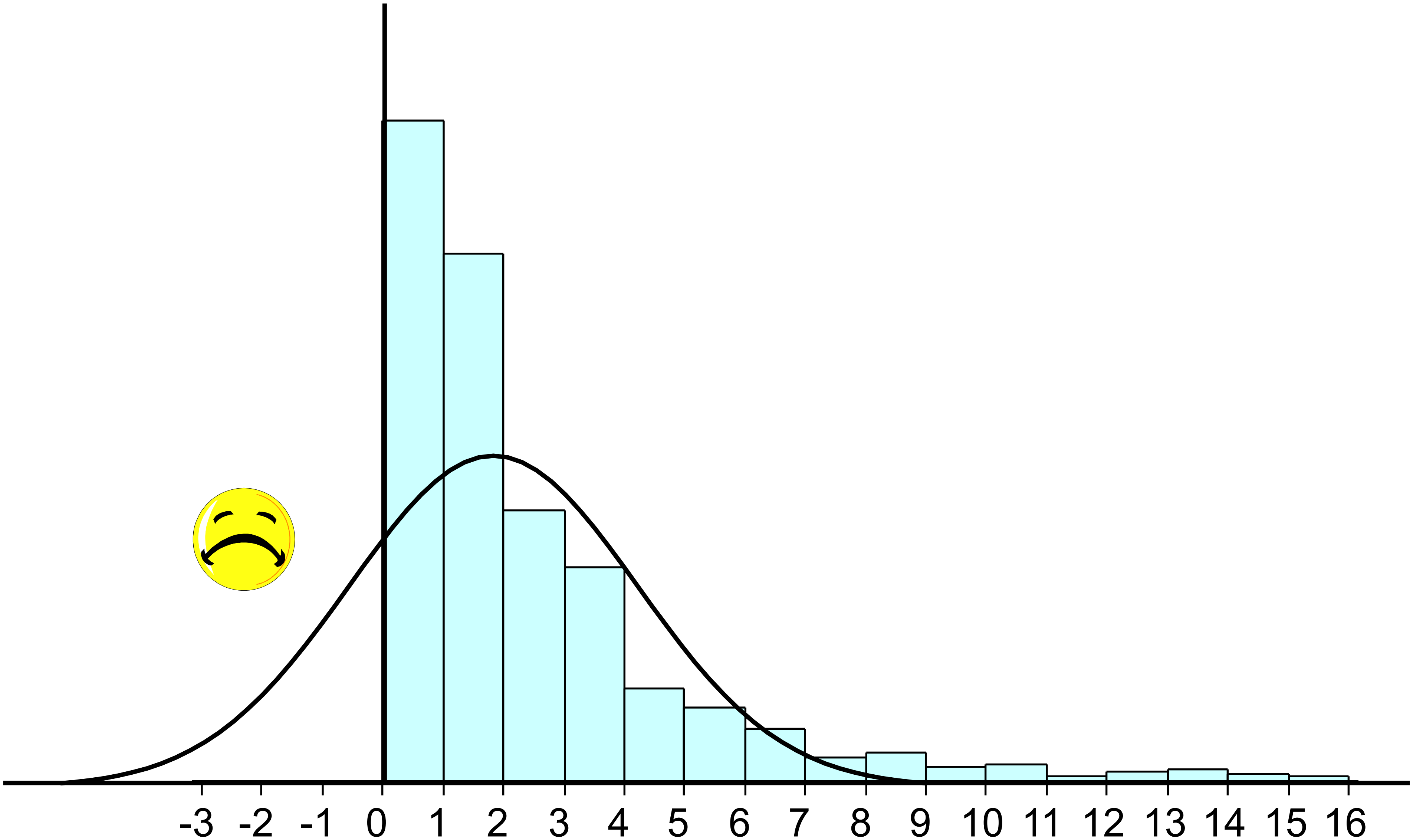
$$f(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

- The Normal curve depends only on μ and σ (population mean and std. dev.)
- Plug the sample mean and std. dev. into the formula in place of μ and σ

Distribution curves allow us to extrapolate . . .

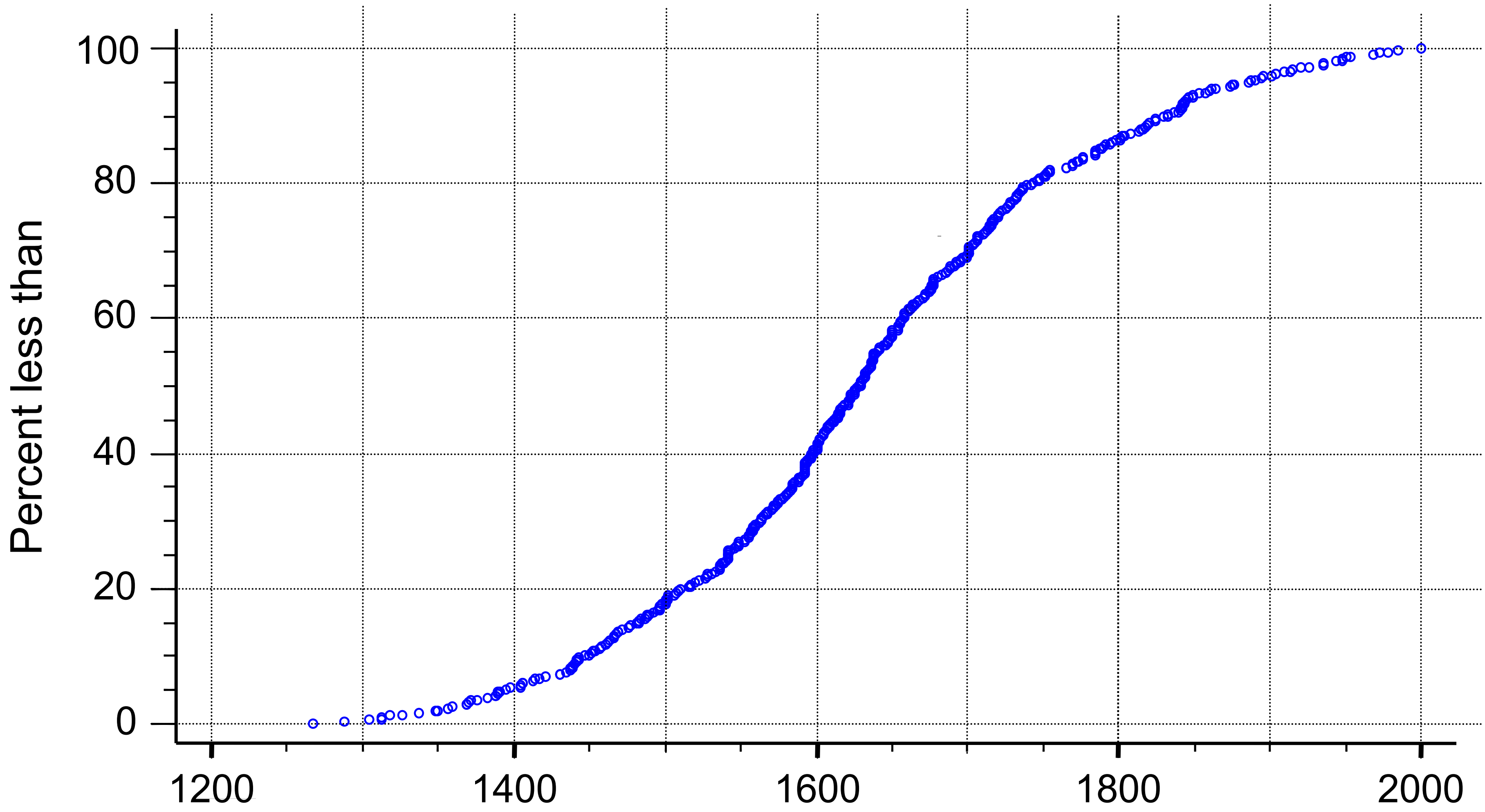


. . . but only if the distribution matches the data!



Checking goodness of fit

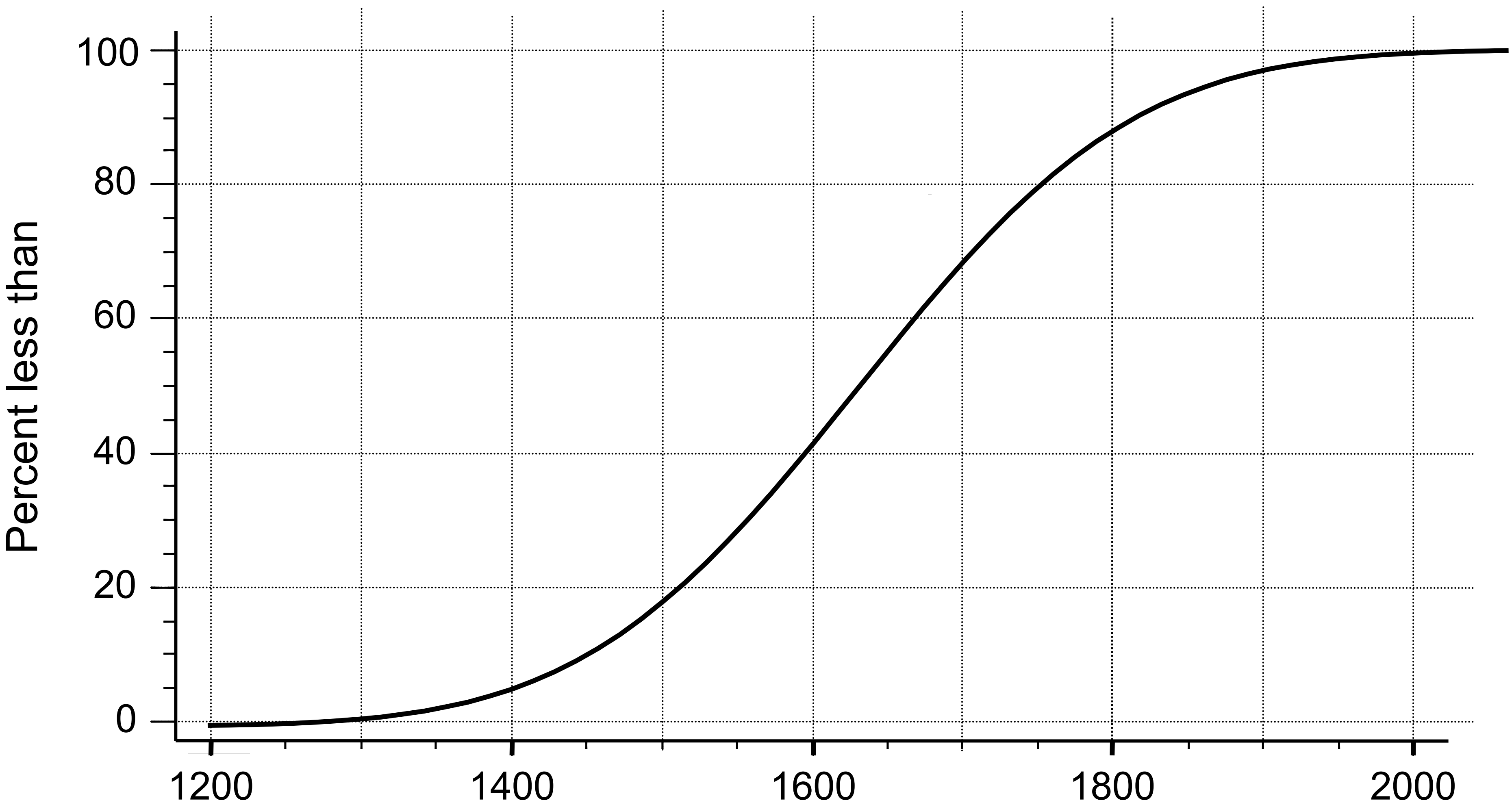
*Data CDF**



*Cumulative Distribution Function

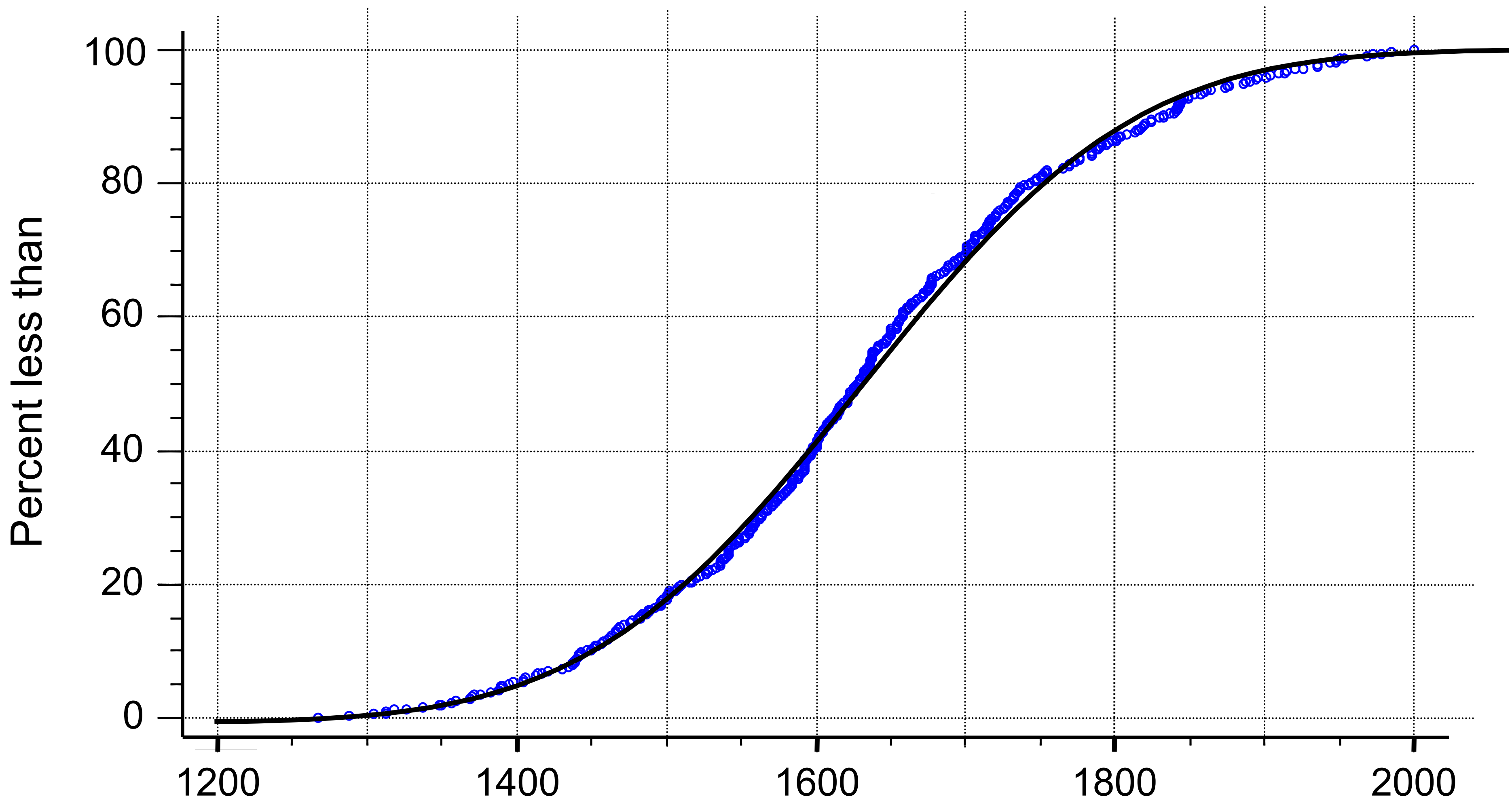
Checking goodness of fit (cont'd)

Best fitting population CDF (assuming Normal)



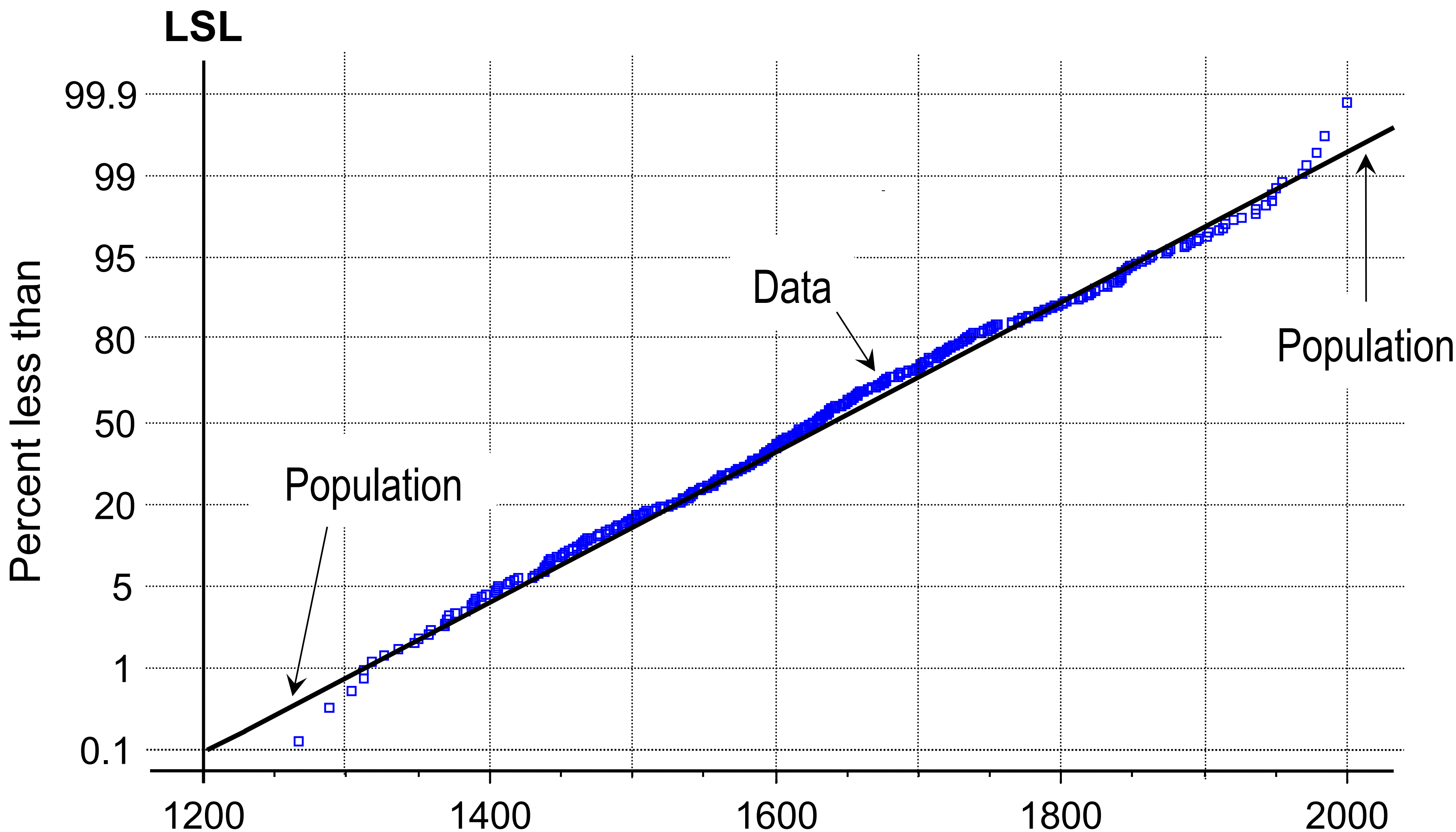
Checking goodness of fit (cont'd)

Data and population CDFs should match



Normal Quantile Plot (also known as Normal probability plot)

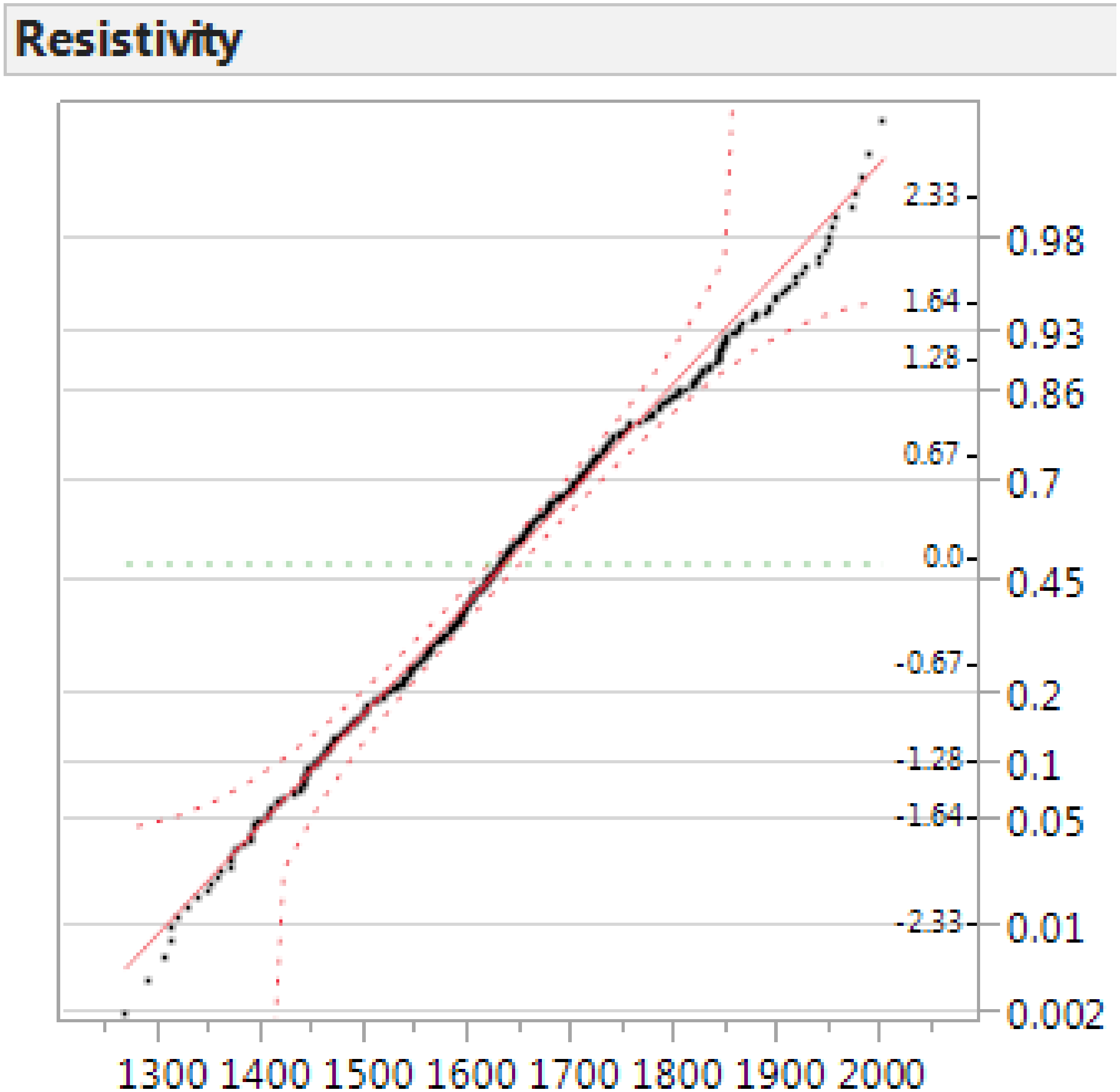
CDFs plotted on a Normal distribution scale



JMP example: Normal data

File → Open → Data sets → DI water → Open → Import

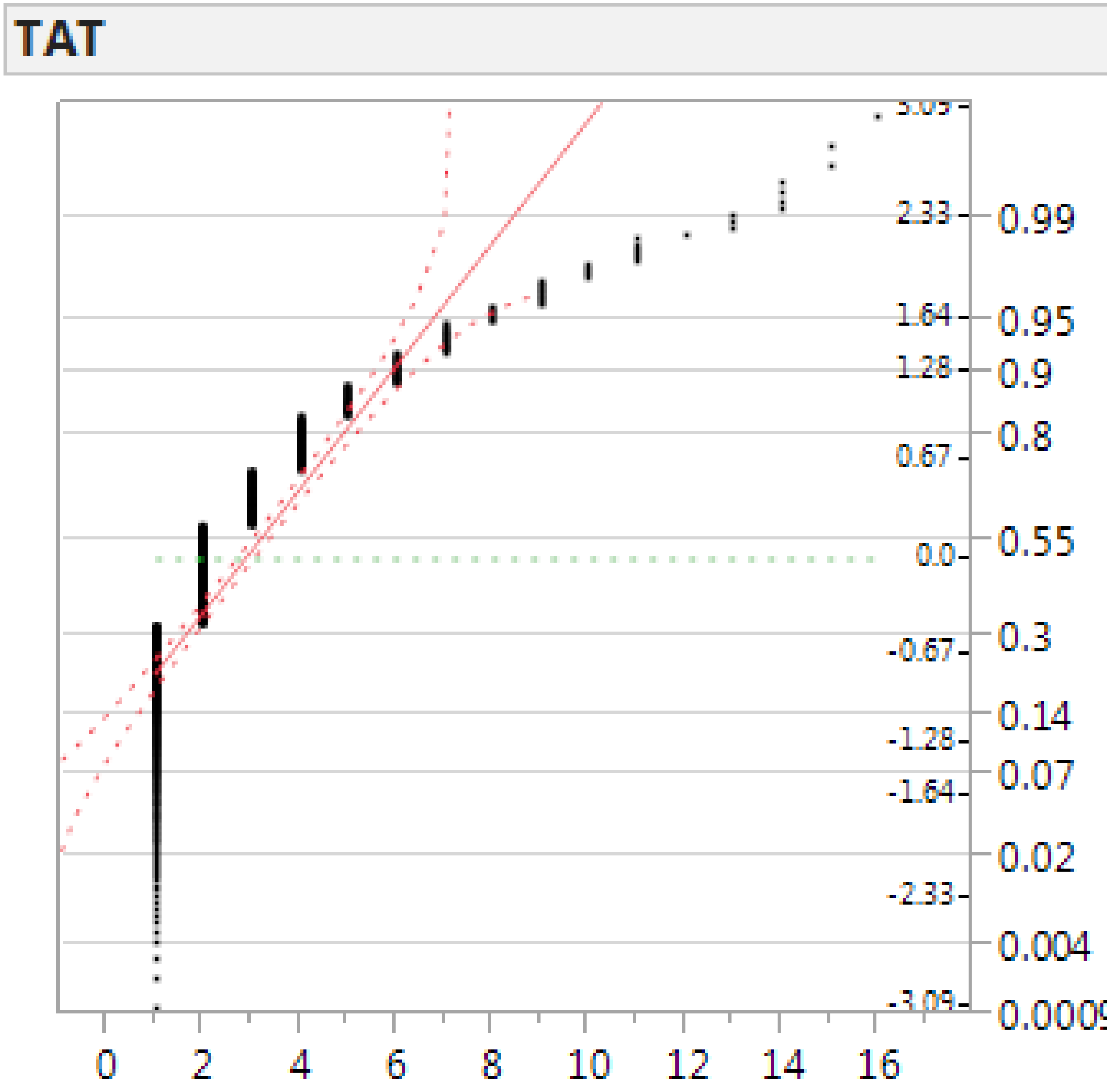
- Analyze → Distribution → *Resistivity* → Y, Columns → OK
- ▼ **Resistivity** → Normal Quantile Plot
- Fit is good – the points form a relatively straight line and stay within the hyperbolic band
    - It is common for the data to curve up a little at the top and down a little at the bottom of the Normal Quantile Plot
    - A curve throughout the graph indicates non-normal data
- Save the script to the data table
- File save as → *DI water.jmp*
- Leave the data table open



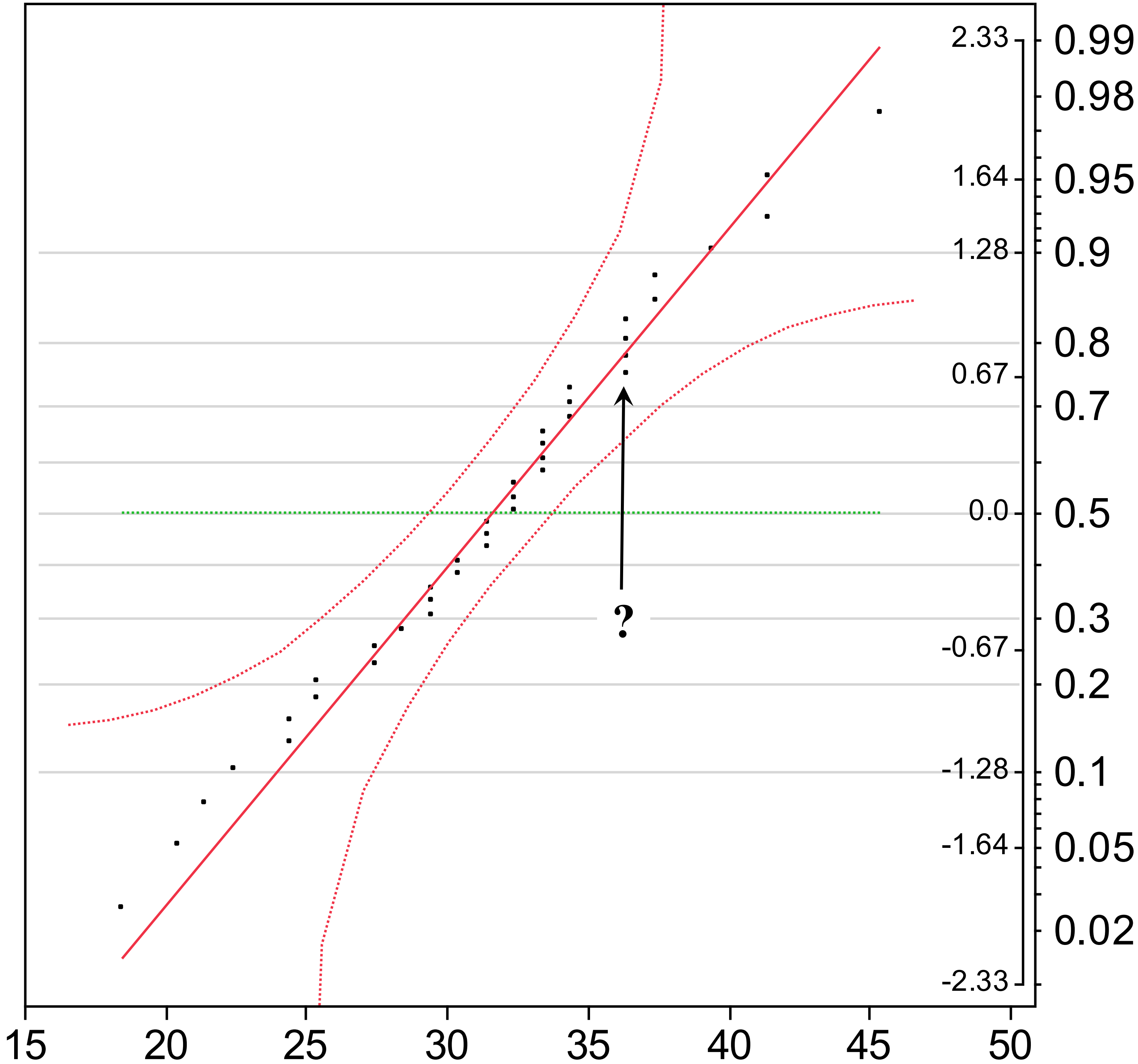
JMP example: non-Normal data

File → Open → Data sets → quotation process → Open → Import

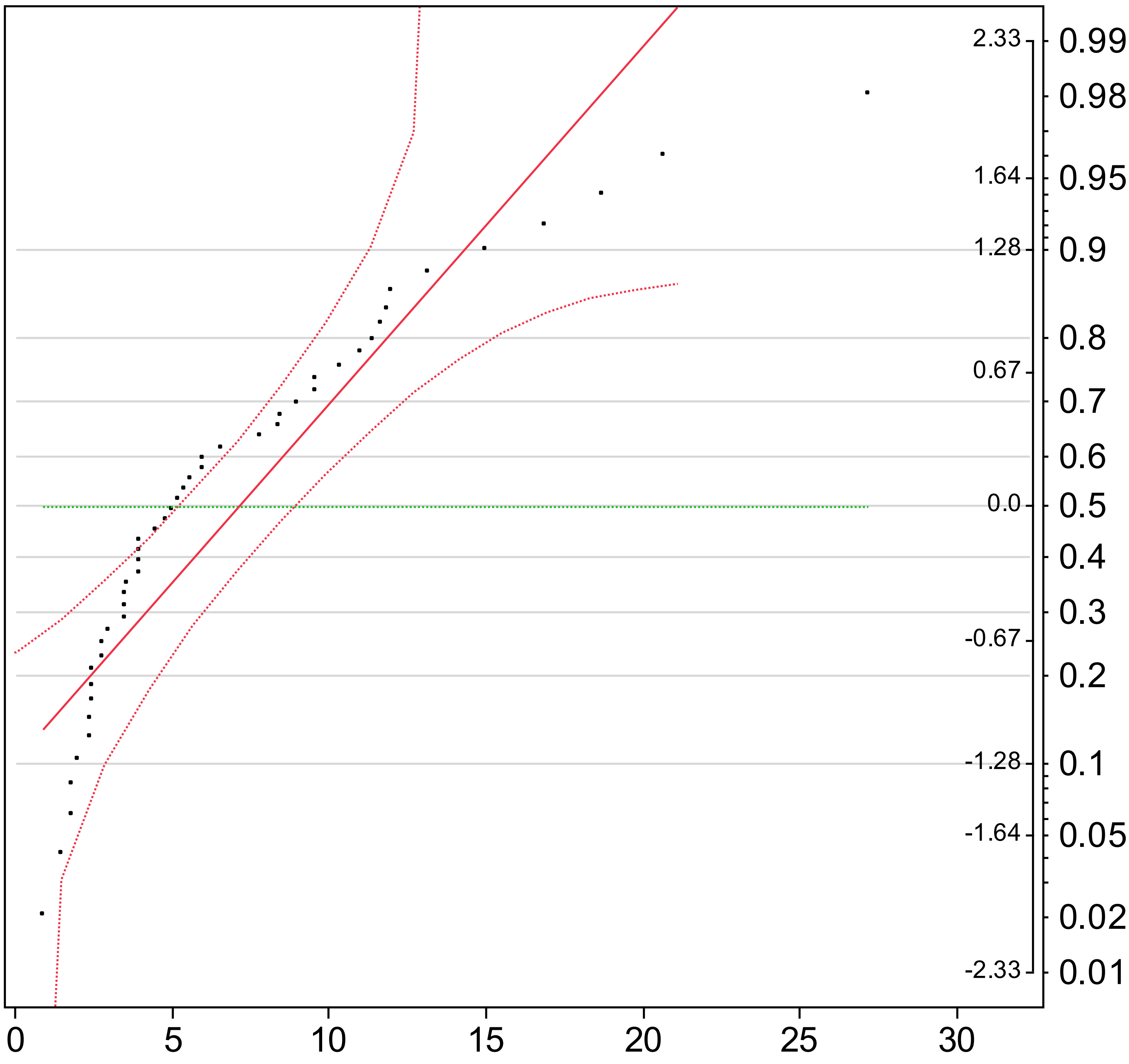
- Analyze → Distribution → Y, Columns → *TAT* → OK
- Distributions → Stack
- *TAT* → Normal Quantile Plot
- Fit is bad – the points do not follow the line and do not stay inside the hyperbolic band
- Save the script to the data table
- File save as → *quotation process.jmp*
- Close the data table



Is this data Normal?

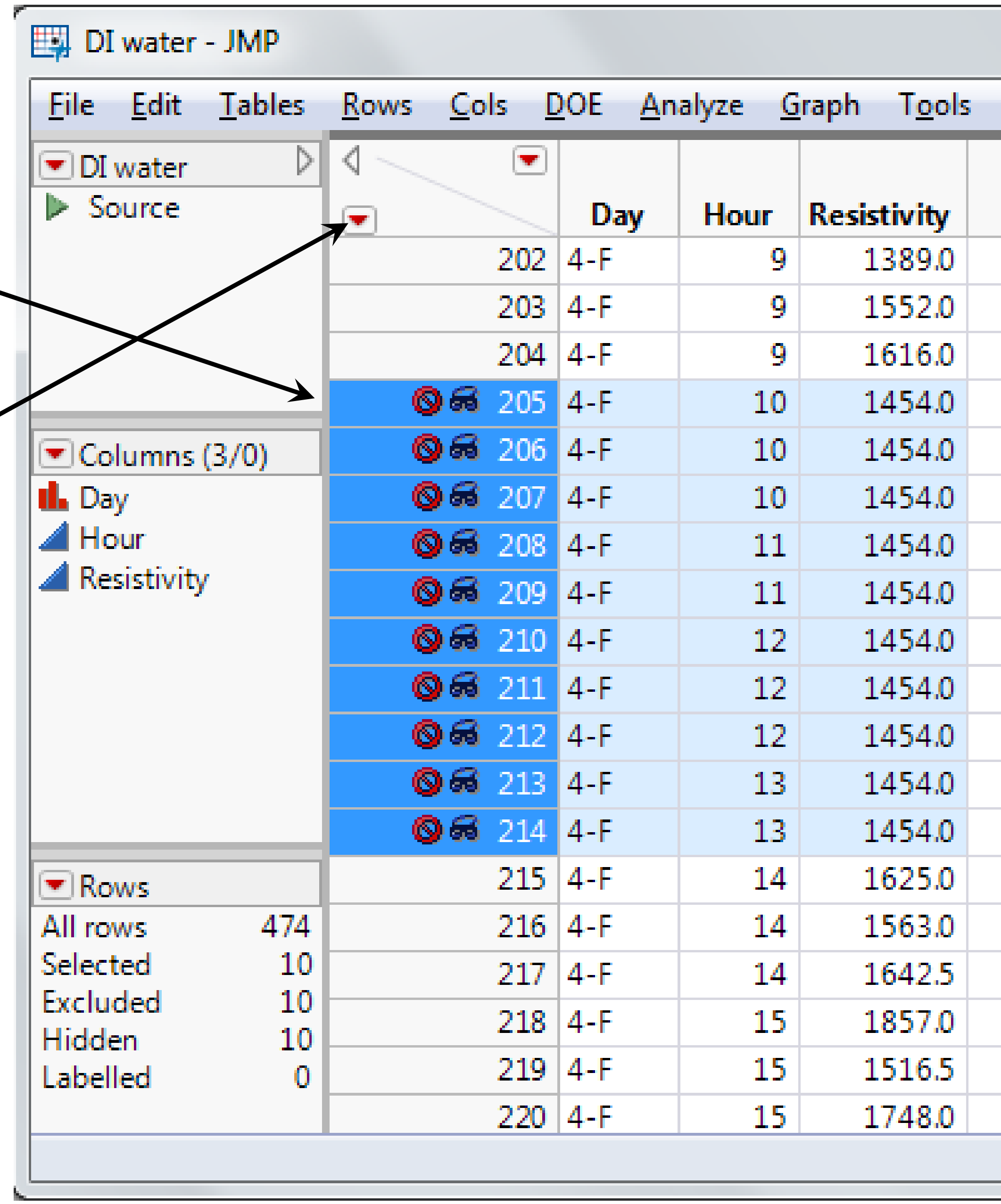


Is this data Normal?



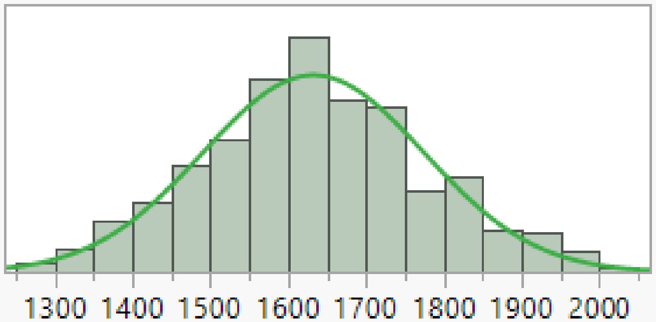
Fitting and using the Normal distribution

- Go to *DI water.jmp*
- The values of *Resistivity* in rows 205 to 214 are constant at 1454
- These are not true measurements, so we use the red triangle to hide and exclude the questionable values
- This reduces the sample size from 474 to 464
- Next slide:
  - Analyze → Distribution
  - Red Triangle → Continuous Fit → Fit Normal



Normal distribution (cont'd)

Resistivity



Compare Distributions

Summary Statistics

Mean	1632.5248
Std Dev	142.06717
N	464
Minimum	1267
Maximum	2000
Median	1628.75

Fitted Normal Distribution

Parameter		Estimate	Std Error	Lower 95%	Upper 95%
Location	μ	1632.5248	6.595303	1619.5982	1645.4513
Dispersion	σ	142.06717	0.3919007	141.30114	142.65648

Measures	
-2*LogLikelihood	5915.2214
AICc	5919.2474
BIC	5927.5011

Click on the *Fitted Normal Distribution* red triangle:

→ Select *Diagnostic Plots* → *QQ Plot*

→ Next slide

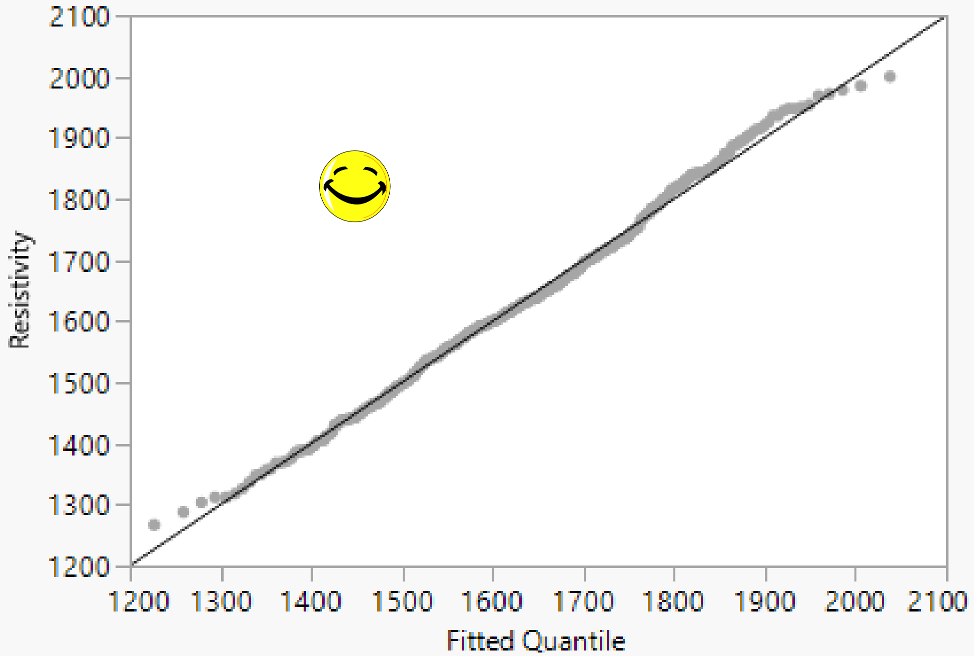
Normal distribution (cont'd)

Fitted Normal Distribution

Parameter		Estimate	Std Error	Lower 95%	Upper 95%
Location	μ	1632.5248	6.595303	1619.5982	1645.4513
Dispersion	σ	142.06717	0.3919007	141.30114	142.65648

Measures	
-2*LogLikelihood	5915.2214
AICc	5919.2474
BIC	5927.5011

QQ Plot



- The QQ Plot is similar to the Normal Quantile Plot
  - When the distribution is a good fit, the data will fall in a line on the plot

- Click on the *Fitted Normal Distribution* red triangle again:

  → Select *Process Capability*

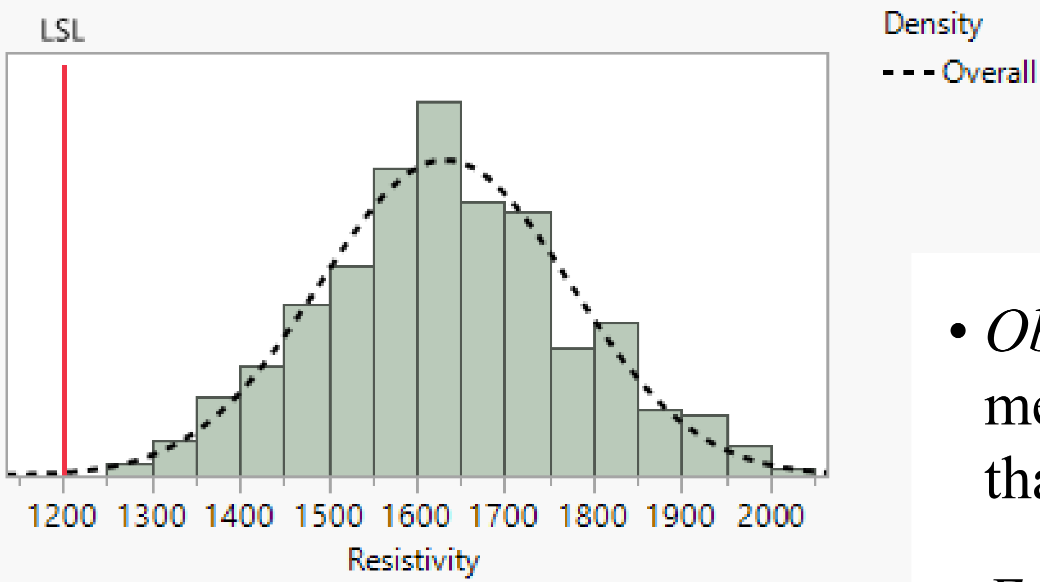
  → Enter 1200 for *Lower Spec Limit* → OK

  → Next slide

Normal distribution (cont'd)

Resistivity Capability

Histogram



Overall Sigma Capability

Index	Estimate	Lower 95%	Upper 95%
Cpk	1.015	0.943	1.087
Cpl	1.015	0.943	1.087

Nonconformance

Portion	Observed %	Expected Within %	Expected Overall %
Below LSL	0.0000	0.0693	0.1165
Total Outside	0.0000	0.0693	0.1165

- *Observed %* shows that none of the measurements in the data set are less than 1200
- *Expected Overall %* shows that 0.12% are predicted to fall below 1200 in the population (future production)
- Save script from the *Distributions* red triangle
- Save and close the data table

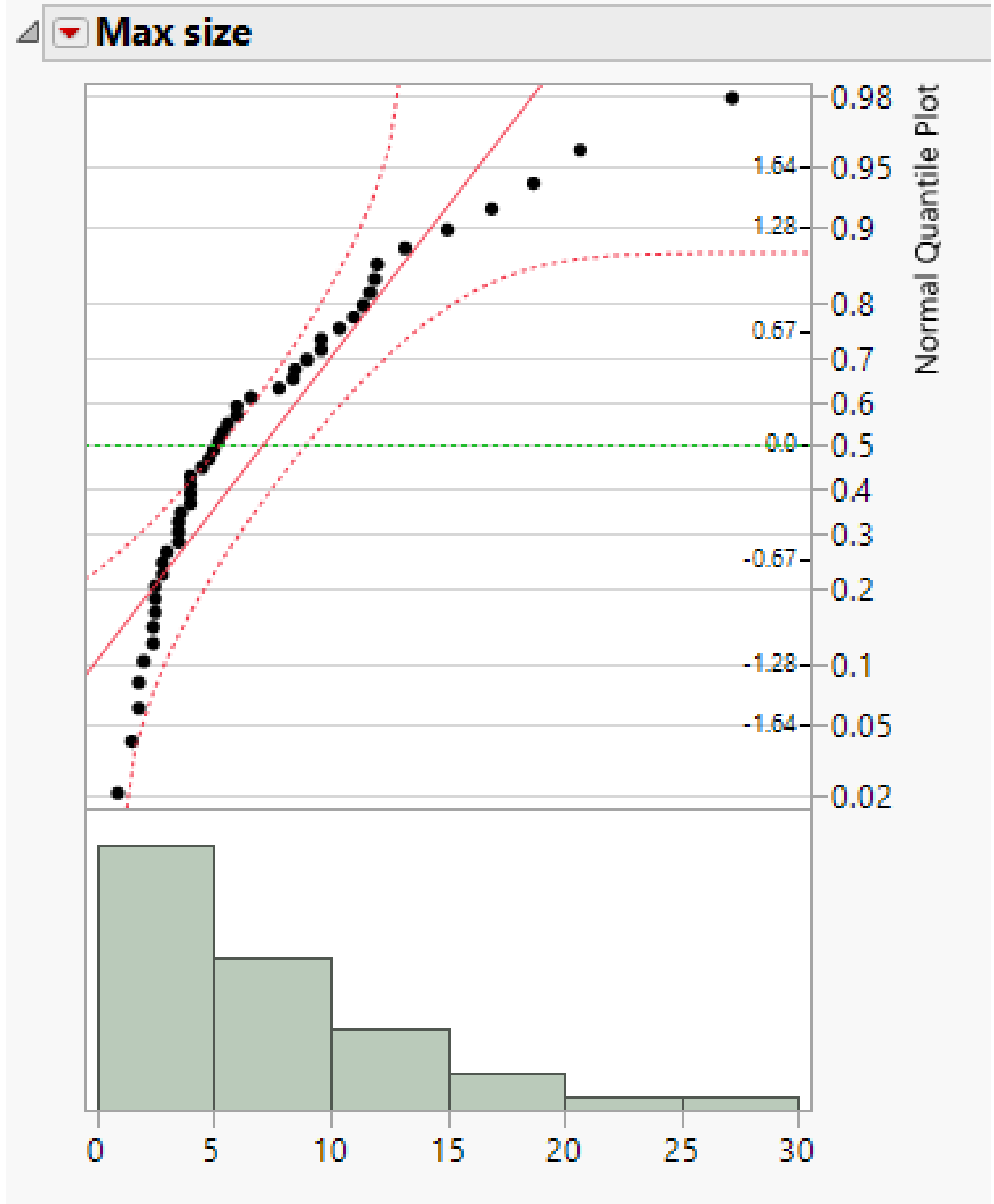
What if the Normal distribution isn't a good fit?

Steps for fitting a distribution to data:

1. Analyze → Distribution
   - Check Normal Quantile Plot—data in straight line indicates good fit
   - If uncertain: Continuous Fit → Fit Normal
   - ▼Fitted Normal Distribution → Goodness of Fit
   - Anderson-Darling p-value > 0.05 indicates good fit
2. If Normal not a good fit: Continuous Fit → Fit Lognormal
   - ▼Fitted Lognormal Distribution → Diagnostic Plots → QQ Plot
   - Data in a relatively straight line on the QQ Plot indicates good fit
   - If uncertain: ▼Fitted Lognormal Distribution → Goodness of Fit
   - Anderson-Darling p-value > 0.05 indicates good fit
3. If Lognormal is not a good fit: Continuous Fit → Fit All
   - Check QQ Plot and view the curve on the histogram to make sure that the fit makes sense for the data.
   - Standard distributions are Weibull, Gamma, Normal, Lognormal, Exponential, Cauchy.
   - JMP offers other specialty distributions that often don't apply, so use caution with them (Johnson Sb, SHASH, Normal 2 Mixture, etc.)

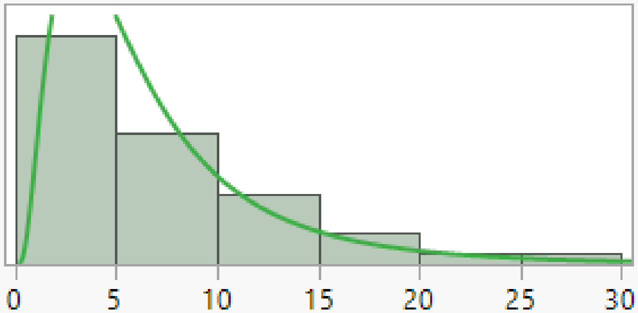
Fitting and using the Lognormal distribution

- *Data sets → number & size of defects*
- Analyze → Distribution → *Max size*
- *Max size* is not Normal
- The *LogNormal* distribution is the most common alternative
- Red triangle *Max Size* → Continuous Fit → Fit LogNormal
- Red triangle *Fitted Lognormal Dist* → Diagnostic Plots → QQ Plot



Lognormal distribution (cont'd)

Max size



Summary Statistics

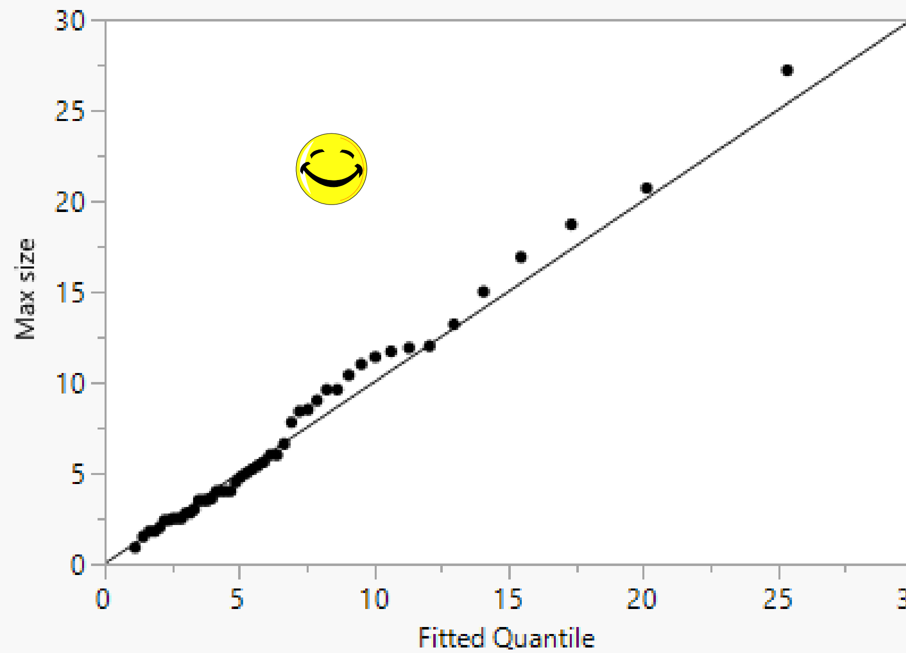
Mean	7.10625
Std Dev	5.6174654
N	48
Minimum	0.9
Maximum	27.2
Median	5.1

Fitted Lognormal Distribution

Parameter		Estimate	Std Error	Lower 95%	Upper 95%
Scale	μ	1.6799251	0.1096067	1.4607293	1.899121
Shape	σ	0.7593775	0.0775036	0.6295191	0.9408779

Measures	
-2*LogLikelihood	271.06631
AICc	275.33298
BIC	278.80872

QQ Plot



Click on the *Fitted LogNormal Distribution* red triangle

→ Select *Process Capability*

→ Enter 30 for the *Upper Spec Limit*

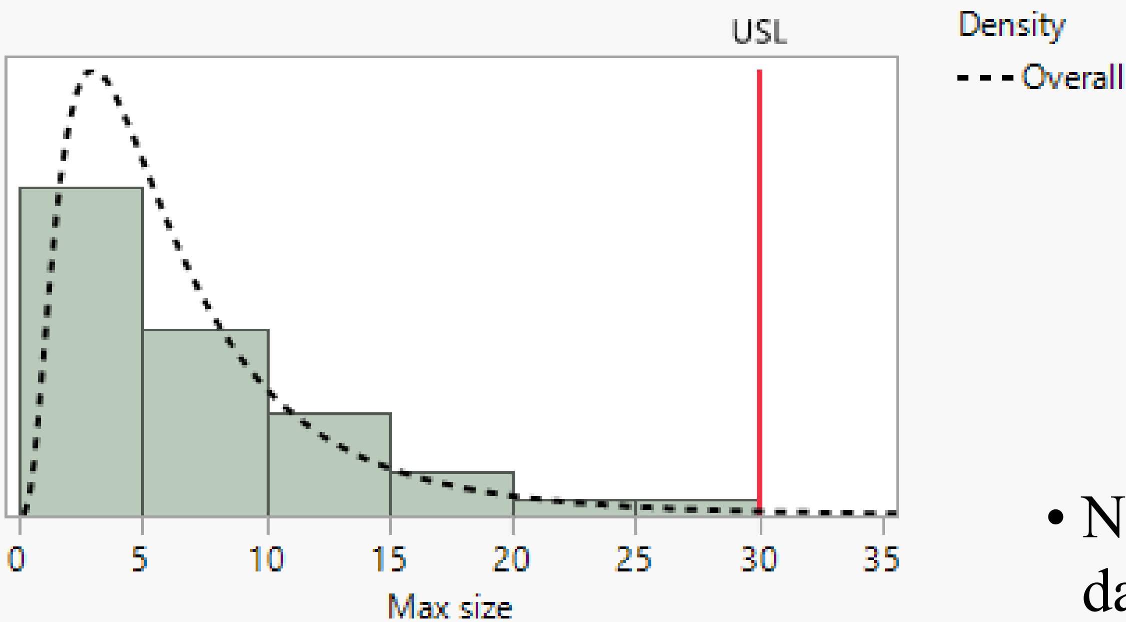
→ OK

Lognormal distribution (cont'd)

Max size(Lognormal) Capability

Nonnormal capability indices calculated with the Percentiles method.

Histogram



Overall Sigma Capability

Index	Estimate
Cpk	0.524
Cpu	0.524

Parameter Estimates

Parameter		Estimate
Scale	μ	1.6799251
Shape	σ	0.7593775

Nonconformance

Portion	Observed %	Expected Overall %
Above USL	0.0000	1.1705
Total Outside	0.0000	1.1705

- None of the measurements in the data set are greater than 30
- 1.17% are predicted to exceed 30 in the population (future production)
- Save script from the *Distributions* red triangle
- Save and close the data table

Finding the best-fitting distribution(s)

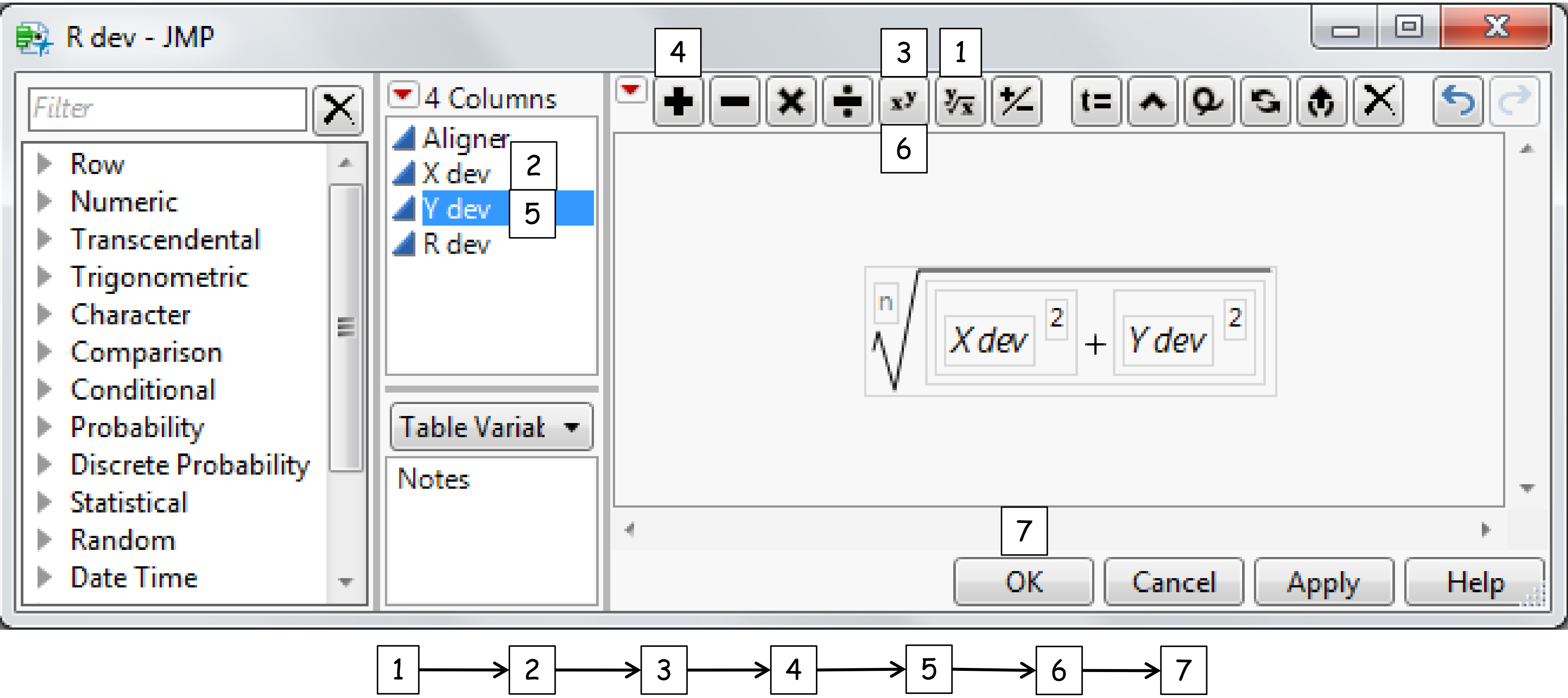
4/0 Cols, 678/0 Rows	Aligner	X dev	Y dev	R dev
1	1	-17	4	17.464249197
2	2	-7	6	9.2195444573
3	3	-10	-21	23.259406699
4	2	0	-1	1
5	2	-10	5	11.180339887
6	2	-7	0	7
7	3	-14	-15	20.518284529
8	2	-3	-17	17.262676502
9	2	-8	3	8.5440037453
10	2	-7	-8	10.630145813
11	1	-11	-6	12.529964086
12	2	-6	0	6
13	2	-7	5	8.602325267
14	3	-10	-5	11.180339887
15	2	-3	1	3.1622776602
16	2	-8	4	8.94427191
17	3	-16	-12	20
18	3	-16	-15	21.931712199
19	1	-14	3	14.317821063
20	2	-8	-8	11.313708499
21	3	-23	-2	23.086792761
22	3	-19	-15	24.207436874
23	2	-7	9	11.401754251
24	2	-10	0	10
25	2	-9	-5	10.295630141
26	1	-8	-11	13.601470509
27	2	-8	-3	8.5440037453
28	3	-16	0	16
29	1	-13	-21	24.69817807
30	3	-8	-4	8.94427191

If neither the Normal or Lognormal are a good fit to the data, you'll need to find a better option.

- *Data sets \ alignment process*
- Three similar alignment tools are used to attach orifice plates to computer chips. *Y dev* and *X dev* are the vertical and horizontal deviations from target in mils.
- The alignment specification applies to the radial deviation calculated from *X* and *Y*. See slide below for the calculation of *R dev*.
- Analyze → Distribution → *R dev*
- Remove:
  - ✓ Summary Statistics
  - ✓ Outlier Box Plot
- Red triangle (R Dev) → Continuous Fit → Fit All
- Go to slide 61 to see the results

Using the formula tool

Double click on the blank column header next to *Y dev*, click on *Column 4*, rename as *R dev*. Click on *Column Properties*, select *Formula, Edit Formula*. Use your mouse to create the formula for *R dev* as shown below.

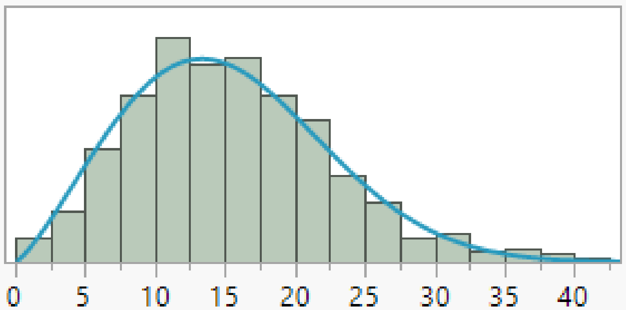


$$\sqrt{X\,dev^2 + Y\,dev^2}$$

Best-fitting distributions (cont'd)

Distributions

R dev



Compare Distributions

Show	Distribution	AICc	AICc Weight	BIC	-2*LogLikelihood
✓	Weibull	4543.338	0.9524	4552.3585	4539.3202
	Johnson Sb	4550.2829	0.0296	4568.3001	4542.2235
	SHASH	4551.4929	0.0161	4569.5101	4543.4335
	Normal 3 Mixture	4556.0329	0.0017	4591.9708	4539.8176
	Normal 2 Mixture	4559.8833	0.0002	4582.3898	4549.7941
	Gamma	4571.173	0	4580.1935	4567.1552
	Normal	4584.4949	0	4593.5154	4580.4771
	Lognormal	4676.7723	0	4685.7928	4672.7545
	Cauchy	4822.69	0	4831.7106	4818.6723
	Exponential	5055.8892	0	5060.4024	5053.8832

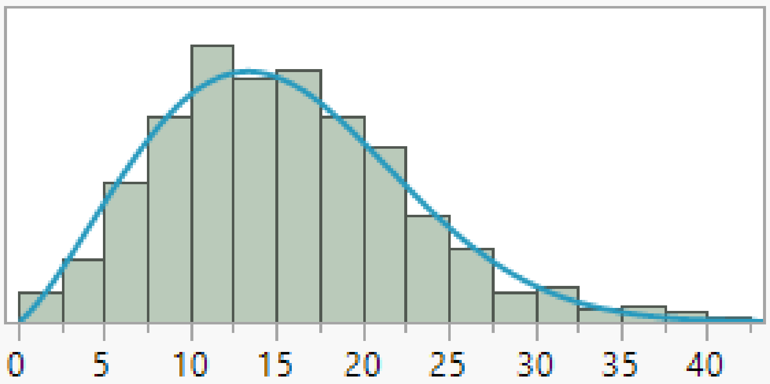
- Distributions are ranked by AICc ("Akaike Information Criterion corrected" – will call it AICc from now on)

- AICc is a measure of *lack* of fit
  - It helps us compare fit of models -- fit of distributions in this case
  - Smaller values indicate better model fit
  - AICc is not a hypothesis test—it doesn't tell you how well a model fits, only which is better

Double click
↓
Format
↓
Fixed Decimal (1)
↓
OK

Best-fitting distributions (cont'd)

R dev



Compare Distributions

Show	Distribution		AICc	AICc Weight	.2 .4 .6 .8
✓	Weibull		4543.3	0.9524	
	Johnson Sb		4550.3	0.0296	
	SHASH		4551.5	0.0161	
	Normal 3 Mixture		4556.0	0.0017	
	Normal 2 Mixture		4559.9	0.0002	
	Gamma		4571.2	0	
	Normal		4584.5	0	
	Lognormal		4676.8	0	
	Cauchy		4822.7	0	
	Exponential		5055.9	0	

- Distributions with the same AICc (rounded to the nearest tenth) have the same lack of fit (or equivalently, the same goodness of fit)
- The distribution with the *AICc Weight* closest to one is the better fit

Using the best-fitting distribution: Weibull

What % of future parts will have *R dev* > 40?

Compare Distributions

Show	Distribution		AICc	AICc Weight	.2 .4 .6 .8	BIC	-2*LogLikelihood
☑	Weibull		4543.338	0.9524		4552.3585	4539.3202
☐	Johnson Sb		4550.2829	0.0296		4568.3001	4542.2235
☐	SHASH		4551.4929	0.0161		4569.5101	4543.4335
☐	Normal 3 Mixture		4556.0329	0.0017		4591.9708	4539.8176
☐	Normal 2 Mixture		4559.8833	0.0002		4582.3898	4549.7941
☐	Gamma		4571.173	0		4580.1935	4567.1552
☐	Normal		4584.4949	0		4593.5154	4580.4771
☐	Lognormal		4676.7723	0		4685.7928	4672.7545
☐	Cauchy		4822.69	0		4831.7106	4818.6723
☐	Exponential		5055.8892	0		5060.4024	5053.8832

Fitted Weibull Distribution

Parameter		Estimate	Std Error	Lower 95%	Upper 95%
Scale	α	17.246152	0.3070044	16.650713	17.855926
Shape	β	2.2716665	0.0672977	2.1415545	2.4053358

Measures

-2*LogLikelihood	4539.3202
AICc	4543.338
BIC	4552.3585

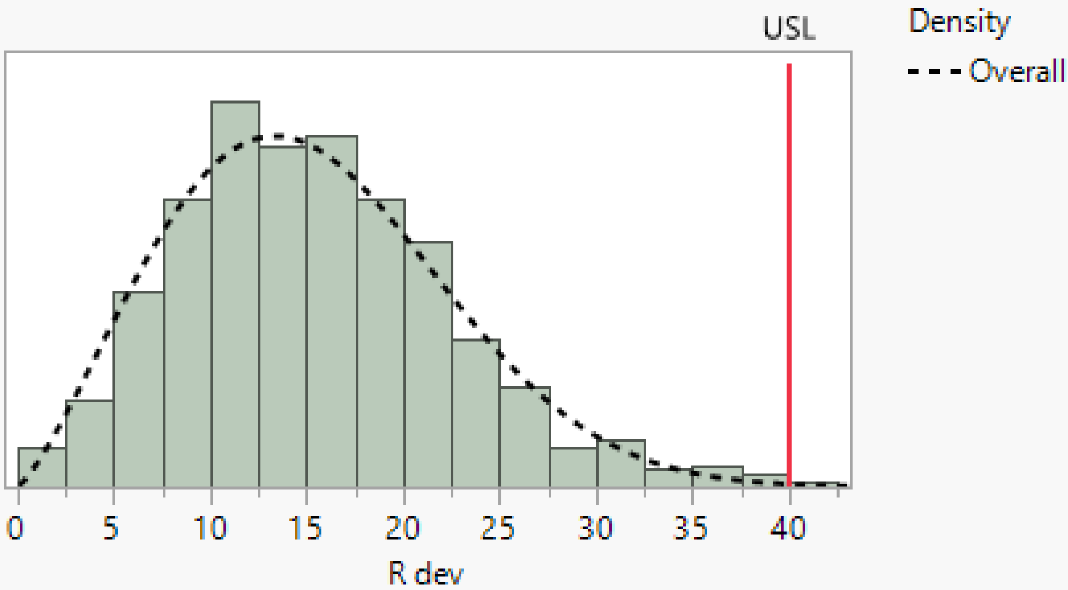
- Click on the *Fitted Weibull Distribution* red triangle
- Select *Process Capability*
- Enter 40 for *USL* → OK

Weibull fit (cont'd)

R dev(Weibull) Capability

Nonnormal capability indices calculated with the Percentiles method.

Histogram



Process Summary

USL	40
N	678
Sample Mean	15.28776
Sample Std Dev	7.097424

Overall Sigma Capability

Index	Estimate
Cpk	1.016
Cpu	1.016

Parameter Estimates

Parameter		Estimate
Scale	α	17.246152
Shape	β	2.2716665

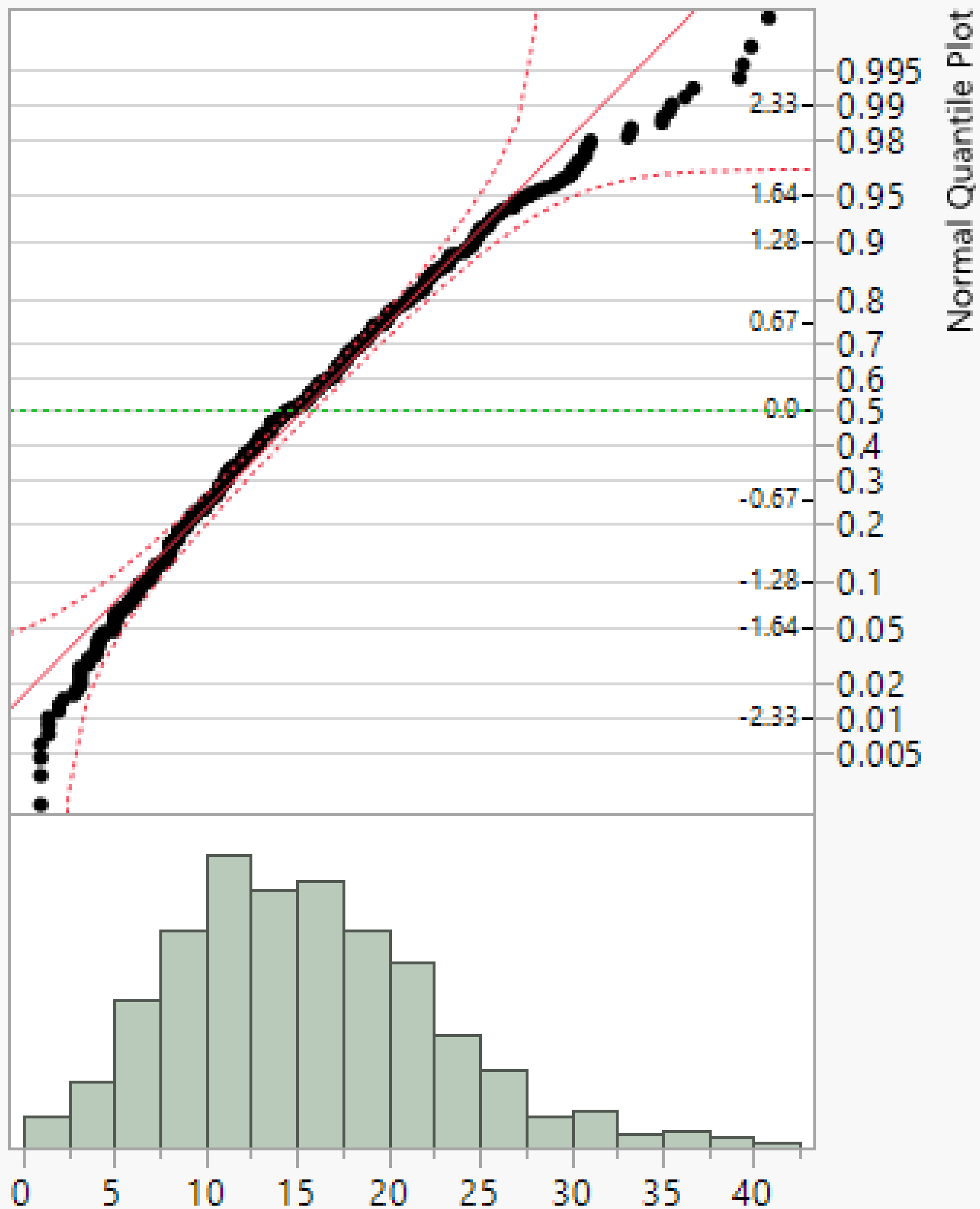
Nonconformance

Portion	Observed %	Expected Overall %
Above USL	0.1475	0.1158
Total Outside	0.1475	0.1158

- 0.15% of the data values exceed 40
- 0.12% are predicted to exceed 40 in the population (future production), based on estimates made using the Weibull distribution

What if we had assumed a Normal distribution?

R dev



Summary Statistics

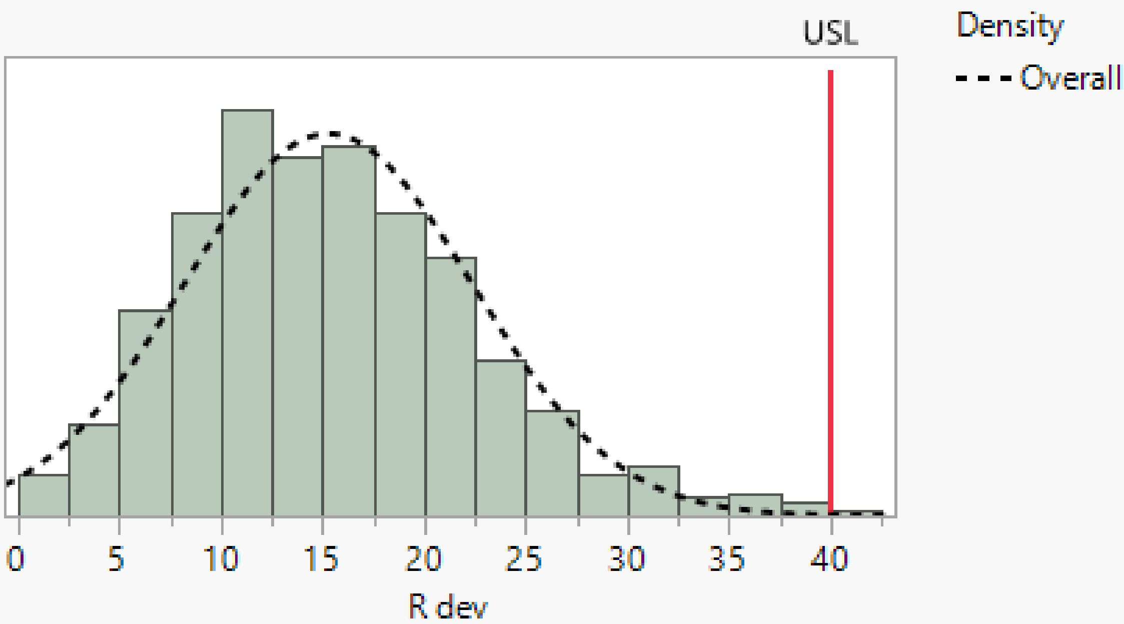
Mean	15.287761
Std Dev	7.0974238
N	678
Minimum	1
Maximum	40.804412
Median	14.422205

- The curve throughout this Normal Quantile Plot indicates that this is not a good fit

What if we had assumed a Normal distribution? (cont'd)

R dev Capability

Histogram



Process Summary

USL	40
N	678
Sample Mean	15.28776
Within Sigma	7.078084
Overall Sigma	7.097424
Stability Index	1.002732

Within sigma estimated by average moving range.

Overall Sigma Capability

Index	Estimate	Lower 95%	Upper 95%
Cpk	1.161	1.094	1.227
Cpu	1.161	1.094	1.227

Nonconformance

Portion	Observed %	Expected Within %	Expected Overall %
Above USL	0.1475	0.0240	0.0249
Total Outside	0.1475	0.0240	0.0249

We would have underestimated the future % defective:

	Expected % Defective
Weibull	0.12%
Normal	0.02%

Steps for fitting a distribution to data

If the Normal or Lognormal is a good fit, use it!

1. Analyze → Distribution
   - Check Normal Quantile Plot—data in straight line indicates good fit
   - If uncertain: Continuous Fit → Fit Normal
   - ▼Fitted Normal Distribution → Goodness of Fit
   - Anderson-Darling p-value > 0.05 indicates good fit
2. If Normal not a good fit: Continuous Fit → Fit Lognormal
   - ▼Fitted Lognormal Distribution → Diagnostic Plots → QQ Plot
   - Data in straight line on the QQ Plot indicates good fit
   - If uncertain: ▼Fitted Lognormal Distribution → Goodness of Fit
   - Anderson-Darling p-value > 0.05 indicates good fit
3. If Lognormal is not a good fit: Continuous Fit → Fit All
   - Check QQ Plot and view the curve on the histogram to make sure that the fit makes sense.
   - Standard distributions are Weibull, Gamma, Normal, Lognormal, Exponential, Cauchy.
   - JMP offers other specialty distributions that often don't apply, so use caution with them (Johnson Sb, SHASH, Normal 2 Mixture, etc.)

Exercise 3.1

Answer questions below. Save the analysis scripts, save and close the data tables. [When opening files, make sure JMP is looking for "All files" not "All JMP files."]

a) *Data sets \ quotation process,* variable TAT. What % of RFQs in the data set have TAT > 15?

b) What % (or PPM) of future RFQs will have TAT > 15?

c) *Data sets \ solution properties,* variable *SG coded*. What % of solution vials in the data set have *SG coded* > 50?

d) What % (or PPM) of future vials will have *SG coded* > 50?

e) *Data sets \ number and size of defects*, variable # *Defects*. What % of castings in the data set have more than 50 defects?

Exercise 3.1 (cont'd)

f) What % (or PPM) of future castings will have more than 50 defects?

g) *Data sets \ casting dimensions*, variable *Length*. What % of castings in the data set have length outside the interval [598, 602]?

h) What % (or PPM) of future castings will have lengths outside this interval?

i) *Data sets \ casting dimensions*, variable *Diam*. What % of castings in the data set have diameters outside the interval [49, 51]?

j) What % (or PPM) of future castings will have diameters outside this interval?

This Page Intentionally Left Blank

Life = elapsed time until the occurrence of some event

- Failure of an item on test
- Planned end of test
- Unplanned end of test
- Failure of an item in service
- Scheduled downtime

Definitions of "time"

- Seconds, minutes, hours
- Days, weeks, months
- Usage cycles, number of moves, distance

Life data (cont'd)

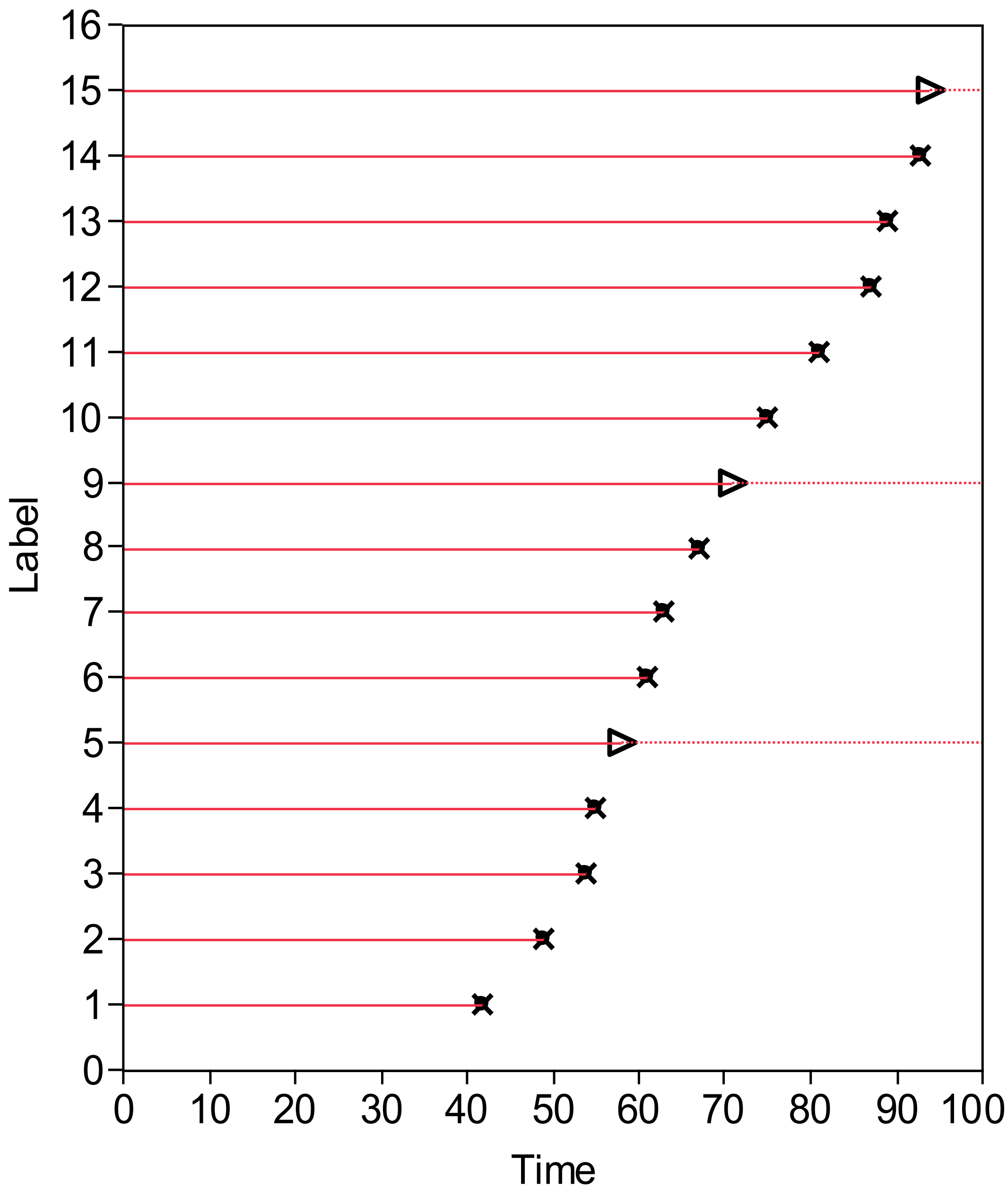
Usually there is one event of primary interest

- Usually, failure of an item

Other events may preempt the event of primary interest

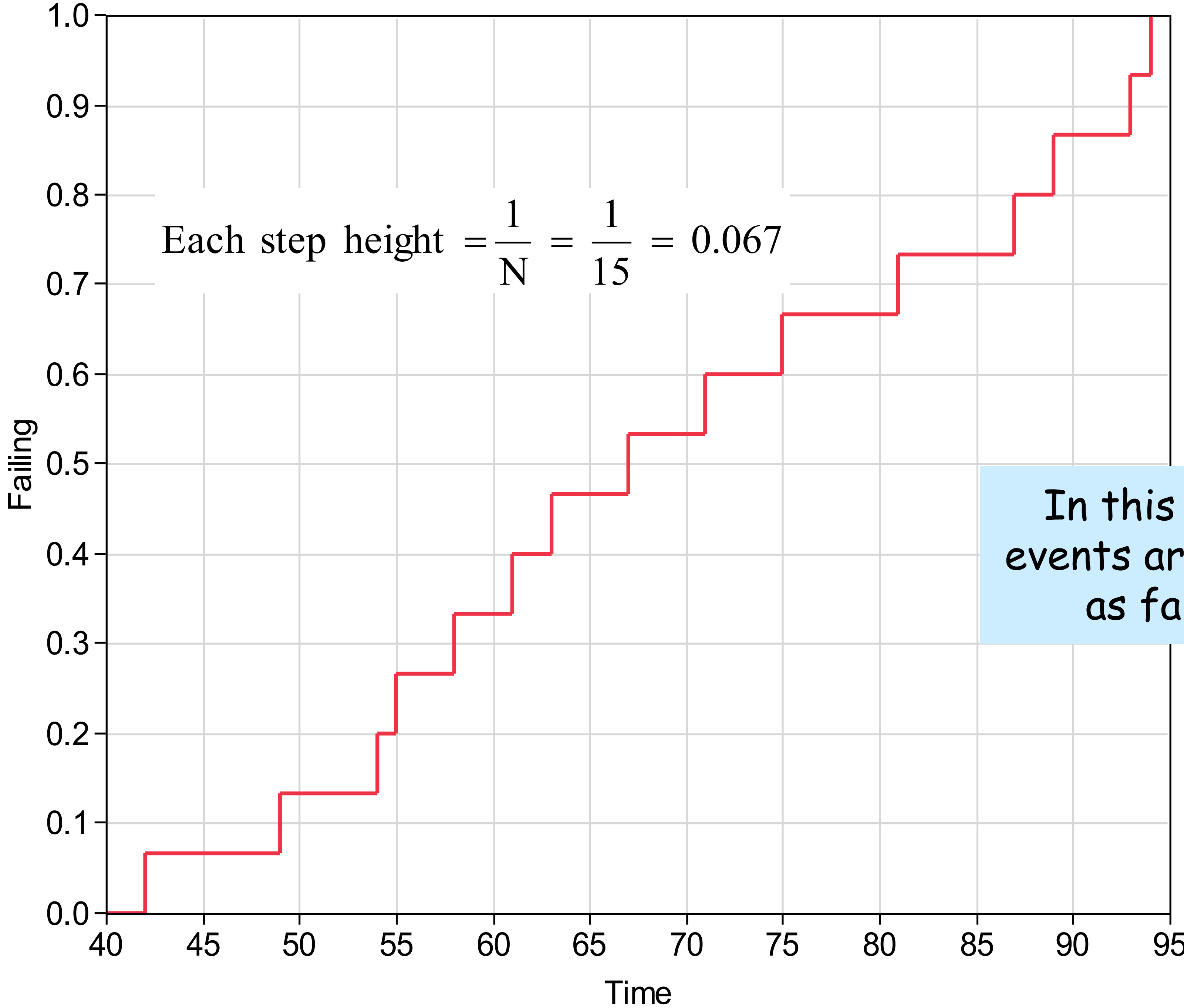
- Planned end of test
- Unplanned end of test
- These are called "suspensions"
- We say that the time to failure is "censored"

Data sets \ failures and suspensions.jmp



- 15 items were tested
- 12 failures (×)
- 3 suspensions (▷·······)
- This "event plot" distinguishes suspensions from failures and shows the event times
- If we don't distinguish suspensions from failures, the calculated failure probabilities will be biased upwards
- This will make our reliability look worse than it really is

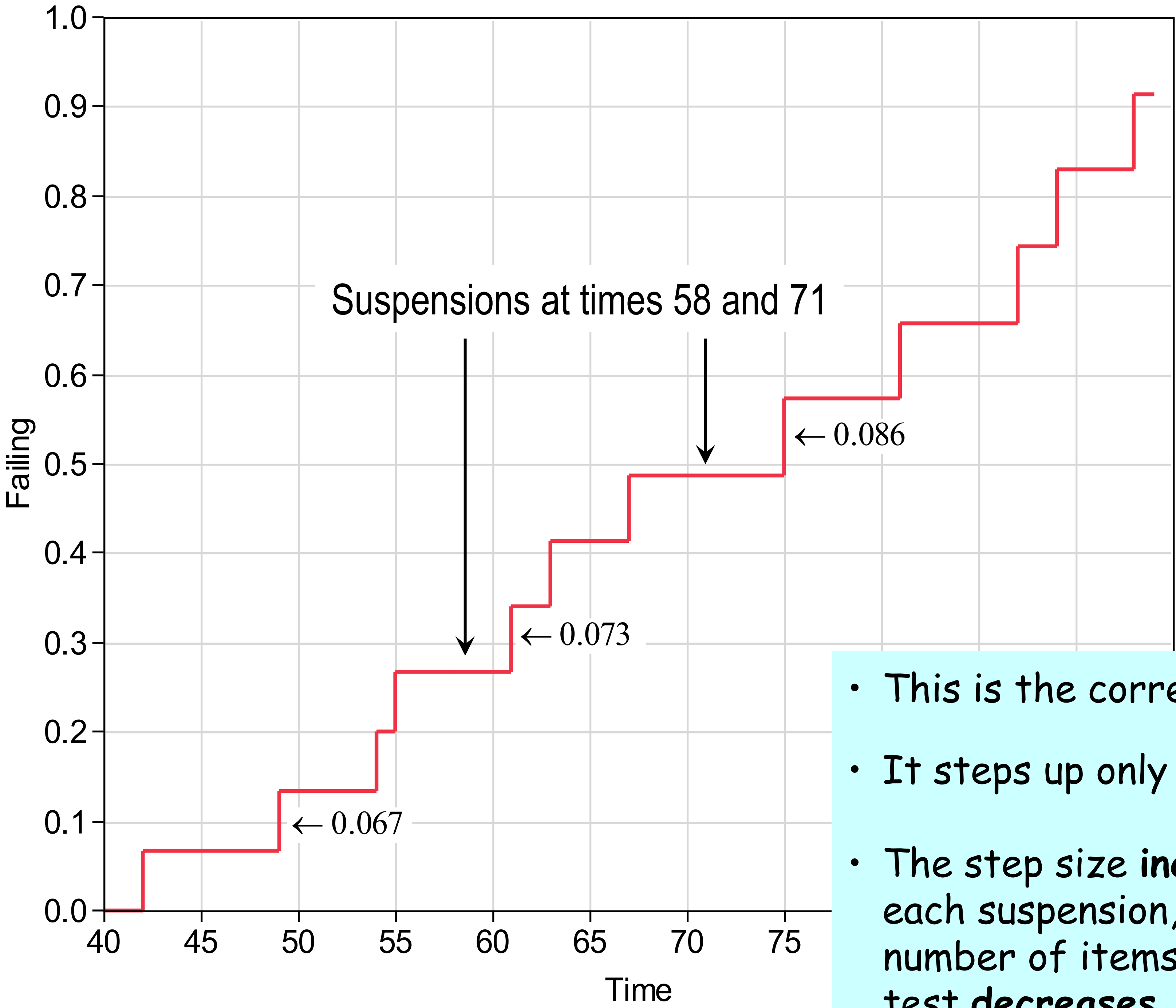
Cumulative distribution function (CDF)



$$\text{Each step height} = \frac{1}{N} = \frac{1}{15} = 0.067$$

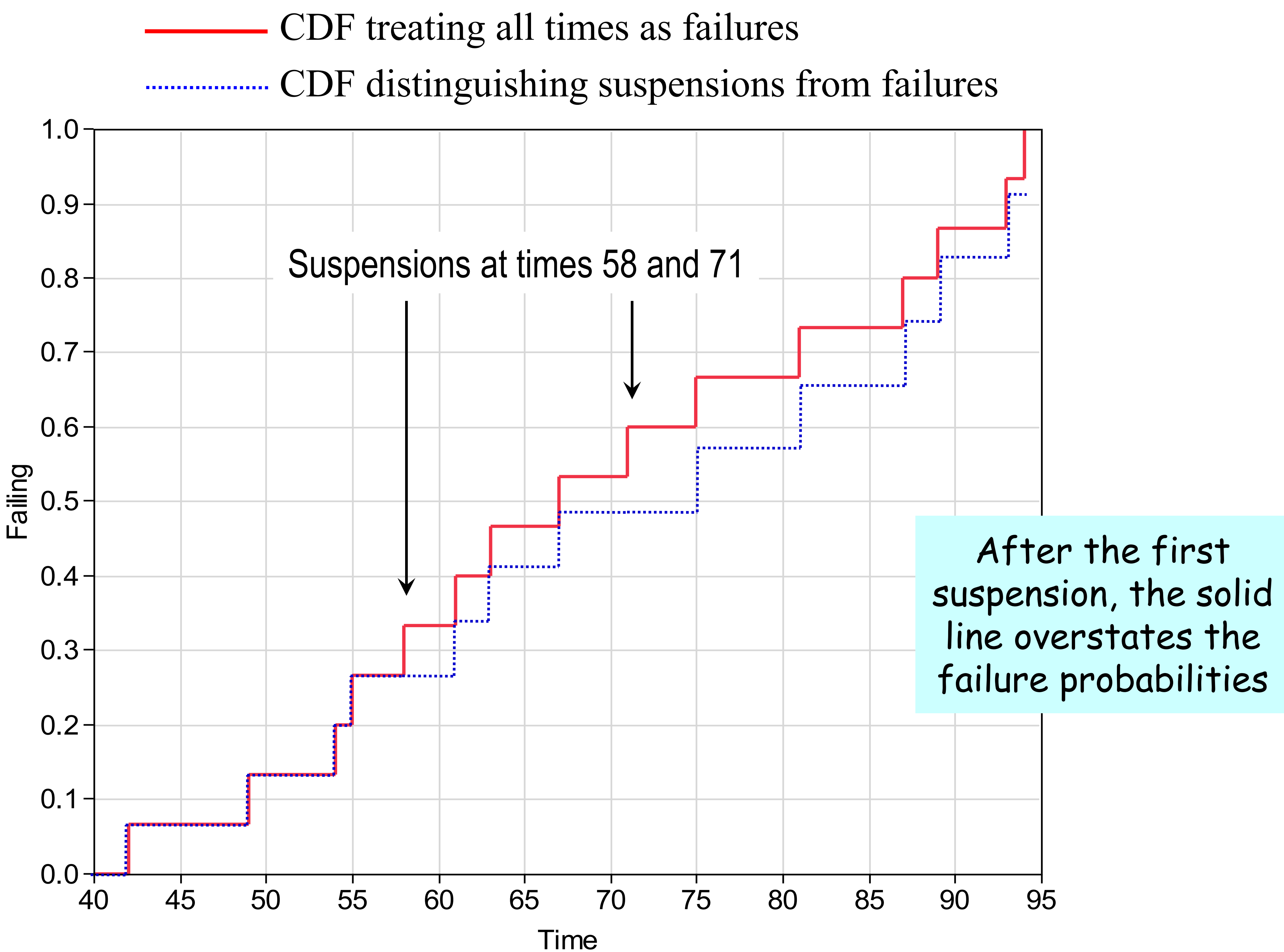
In this plot, all events are treated as failures

CDF distinguishing suspensions from failures

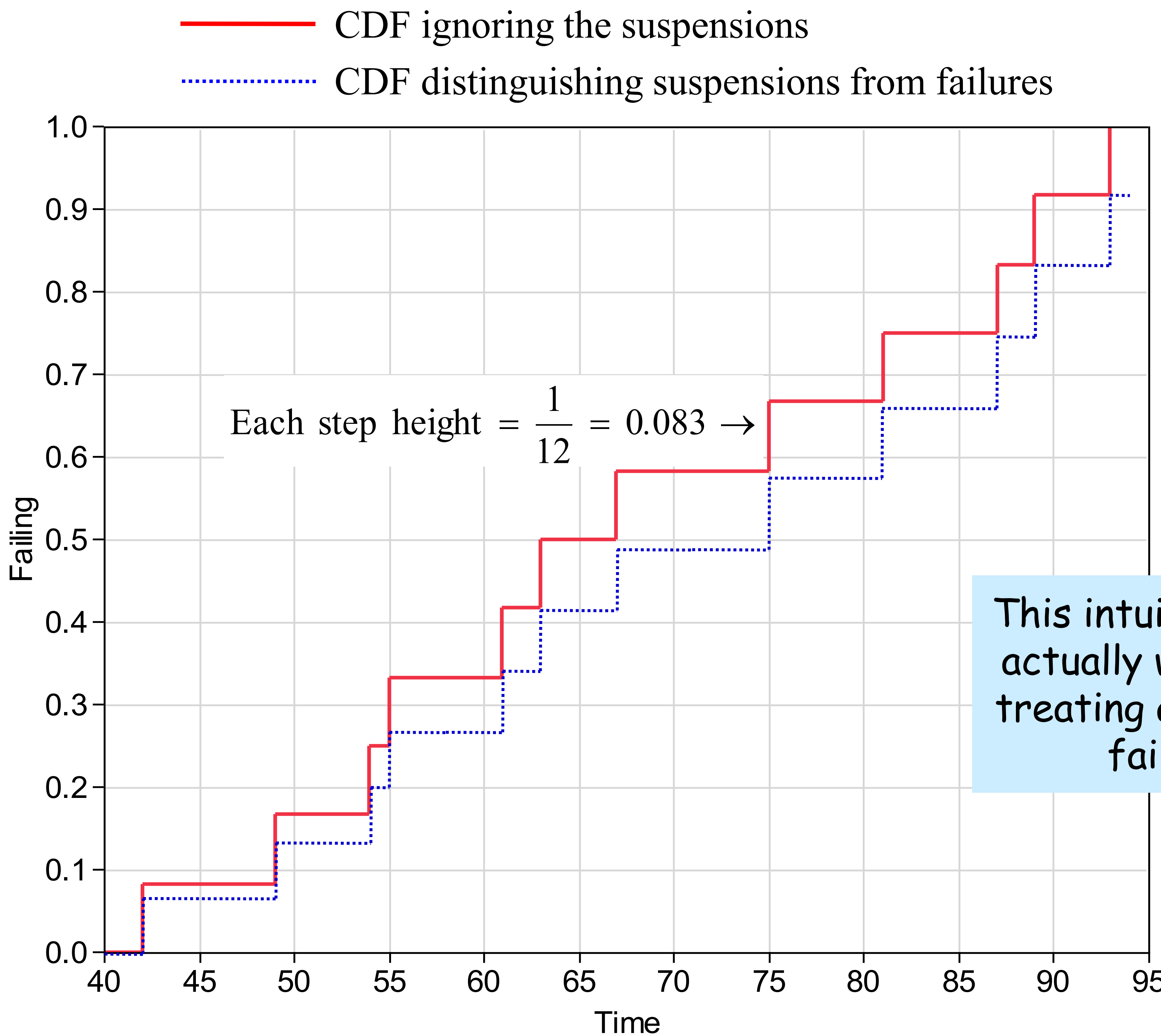


- This is the correct plot
- It steps up only at failure times
- The step size **increases** after each suspension, because the number of items remaining on test **decreases**

Overlay of CDFs



Can't we just ignore the suspensions?



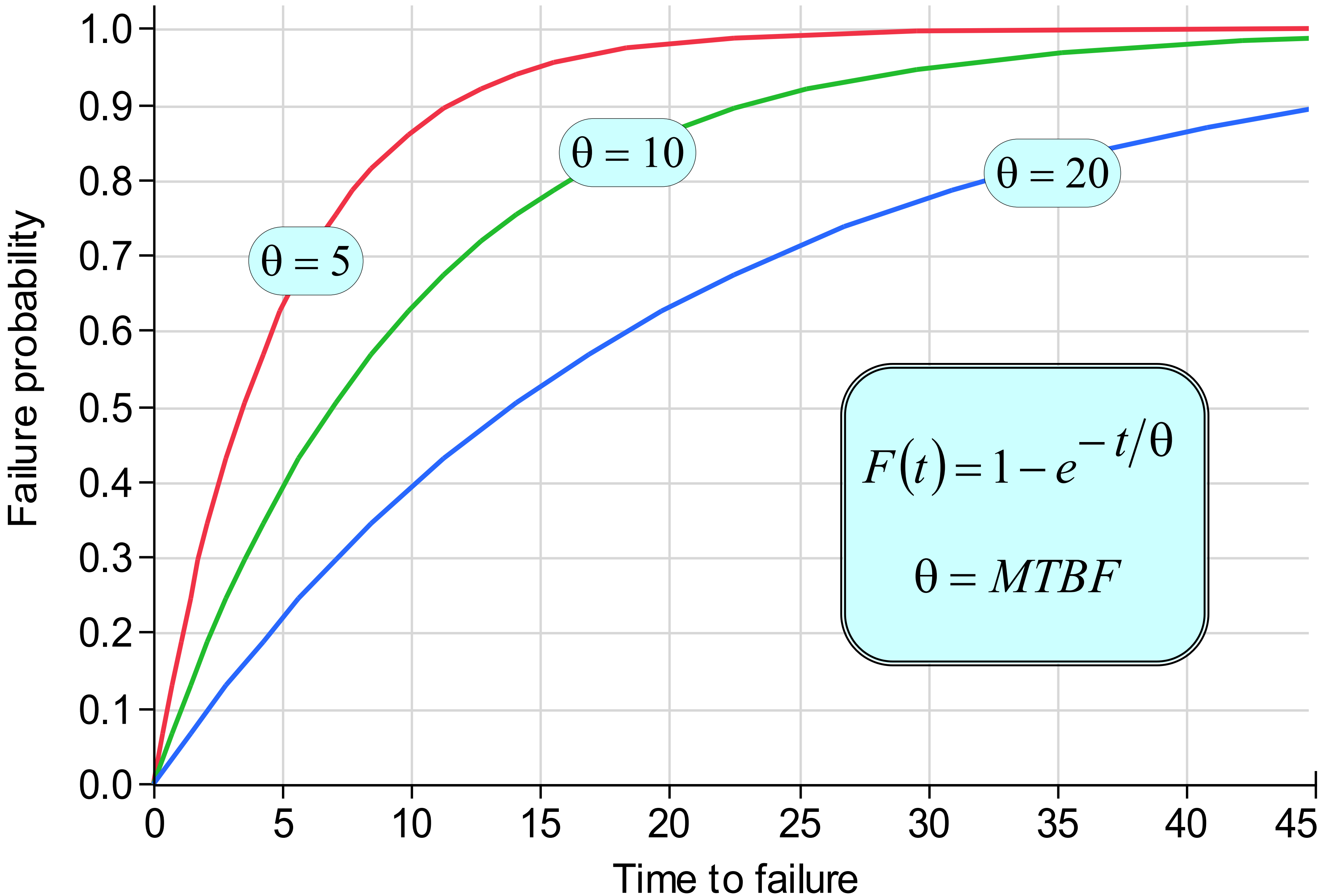
This intuitive idea is actually worse than treating all times as failures

This page intentionally left blank

5 Analyzing Life Data

- The Exponential distribution
- The Weibull distribution
- Fitting life distributions in JMP
- Finding and using the best fitting life distribution

Failure curves for the Exponential distribution



$$F(t) = 1 - e^{-t/\theta}$$

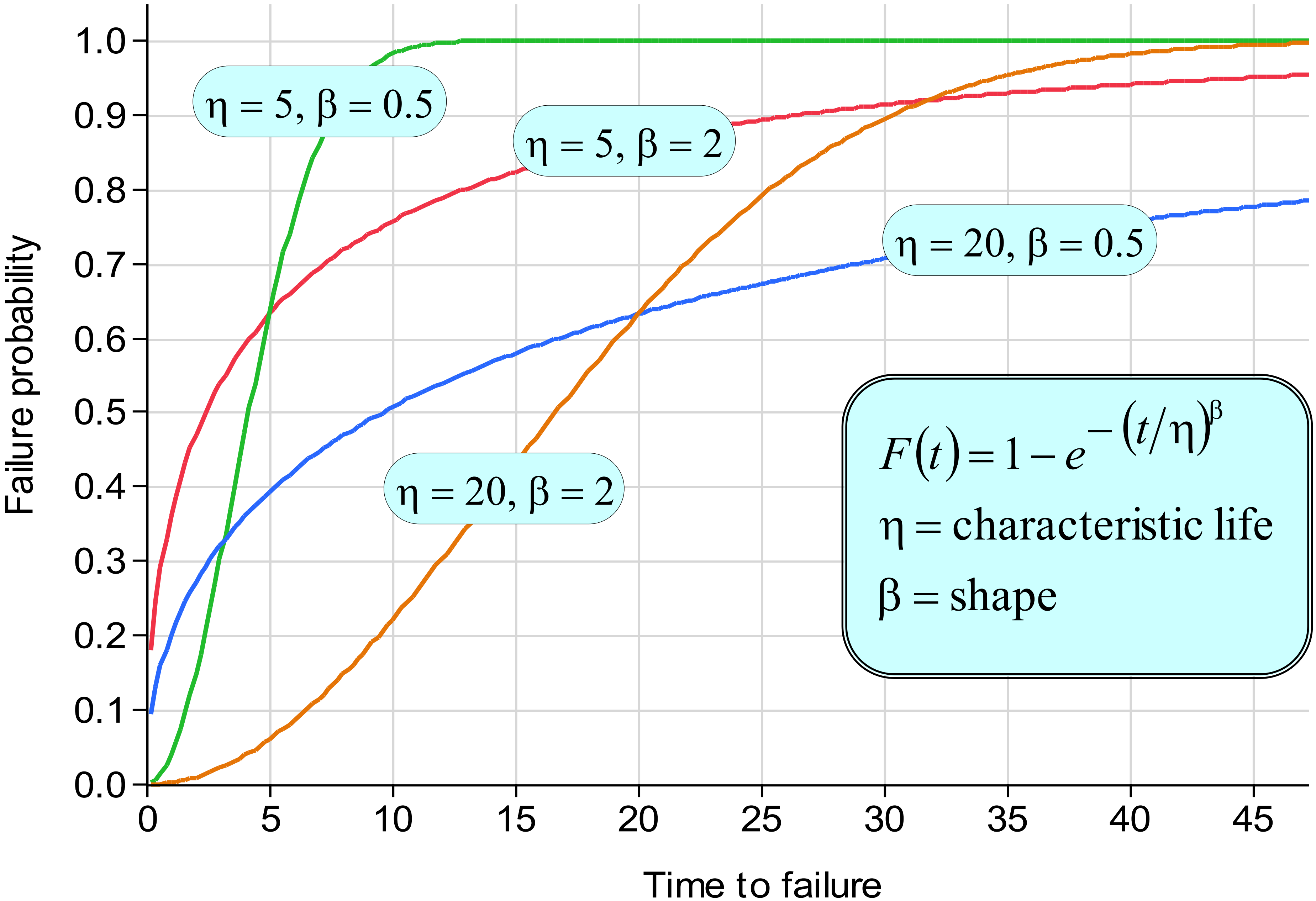
$$\theta = MTBF$$

The Exponential distribution is the simplest life distribution. It has only one parameter: the mean time between/before failure (MTBF). The Greek letter θ (theta) is often used to denote the population value of the MTBF.

Shown above are the *failure functions* $F(t)$ for three different Exponential distributions. $F(t)$ is the probability that an item will fail before time t.

The *reliability function* is defined as $R(t) = 1 - F(t)$. $R(t)$ is the probability that an item will survive beyond time t. The Exponential reliability function is given by $R(t) = \exp(-t/\theta)$.

Failure curves for the Weibull distribution



The Weibull distribution was introduced to the reliability engineering community in the 1950s by a man named Waloddi Weibull. Prior to that, most reliability work was based on the Exponential distribution. Due to its greater flexibility, the Weibull has become one of the most widely-used life distributions.

The Weibull distribution has two parameters: the *characteristic life* η (eta), and the *shape* β (beta). The characteristic life (η) has the same qualitative interpretation as the MTBF (θ). The shape parameter (β) determines which of two distinct failure modes are represented. When $\beta < 1$, we have a *burn-in* or *infant-mortality* failure mode. When $\beta > 1$, we have a *wear-out* failure mode. A Weibull distribution with $\beta = 1$ is identical to an Exponential distribution with $\theta = \eta$.

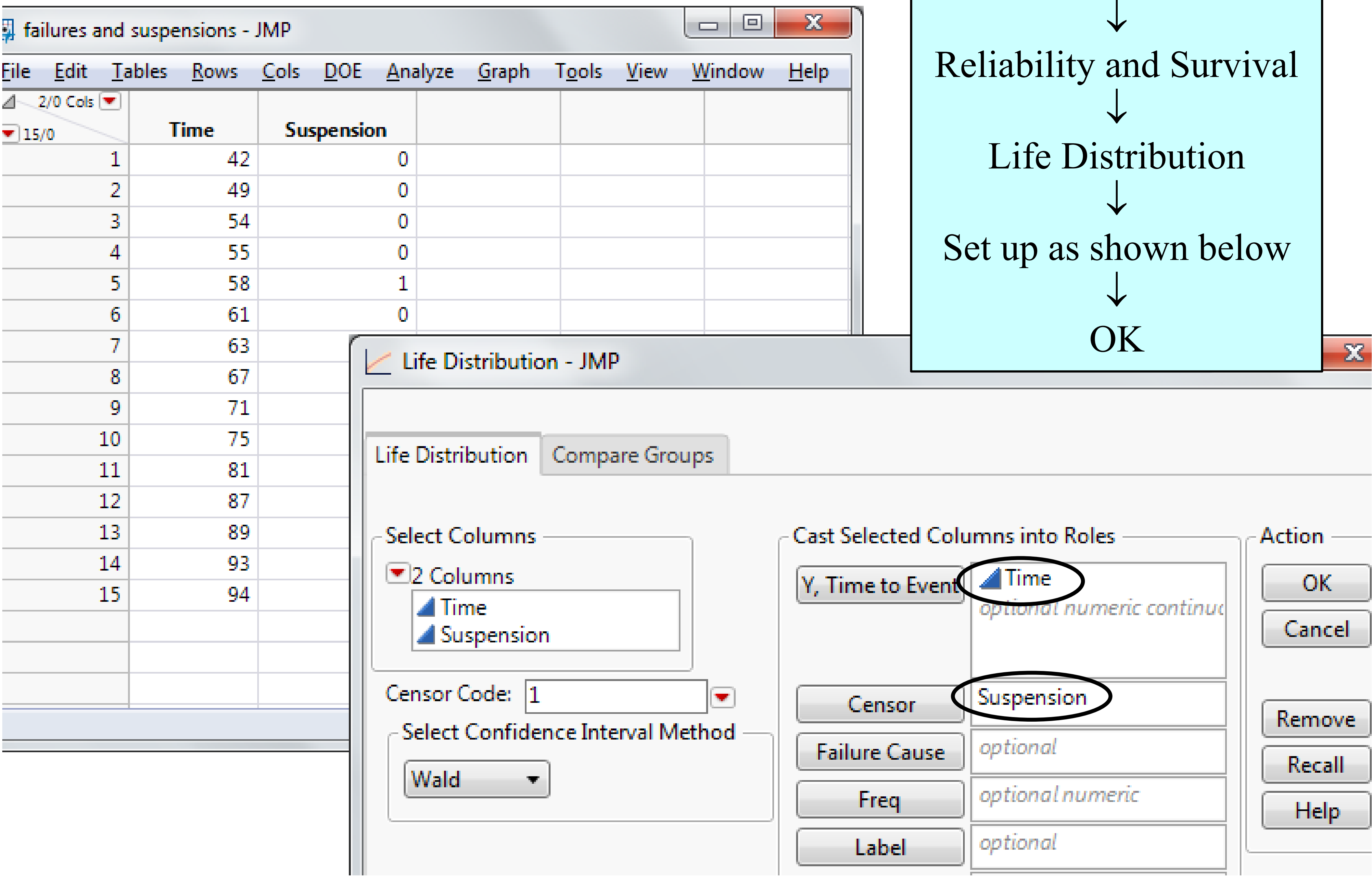
Shown above are failure functions $F(t)$ for four different Weibull distributions. $F(t)$ is the probability that an item will fail before time t.

The Weibull reliability function (probability that an item will survive beyond time t) is given by $R(t) = \exp[-(t/\eta)^{\beta}]$.

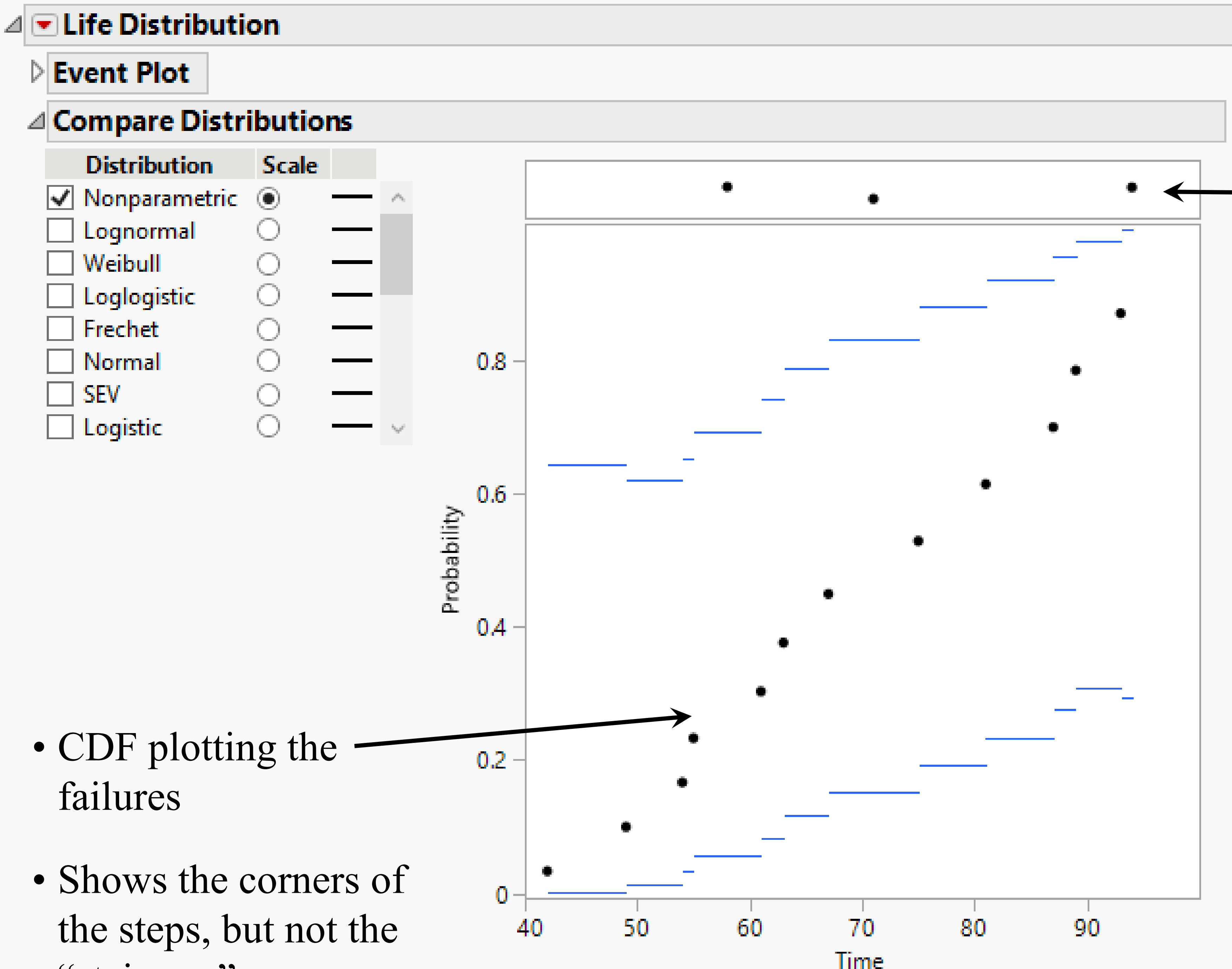
Fitting life distributions in JMP

Data sets \ failures and suspensions

Analyze
↓
Reliability and Survival
↓
Life Distribution
↓
Set up as shown below
↓
OK

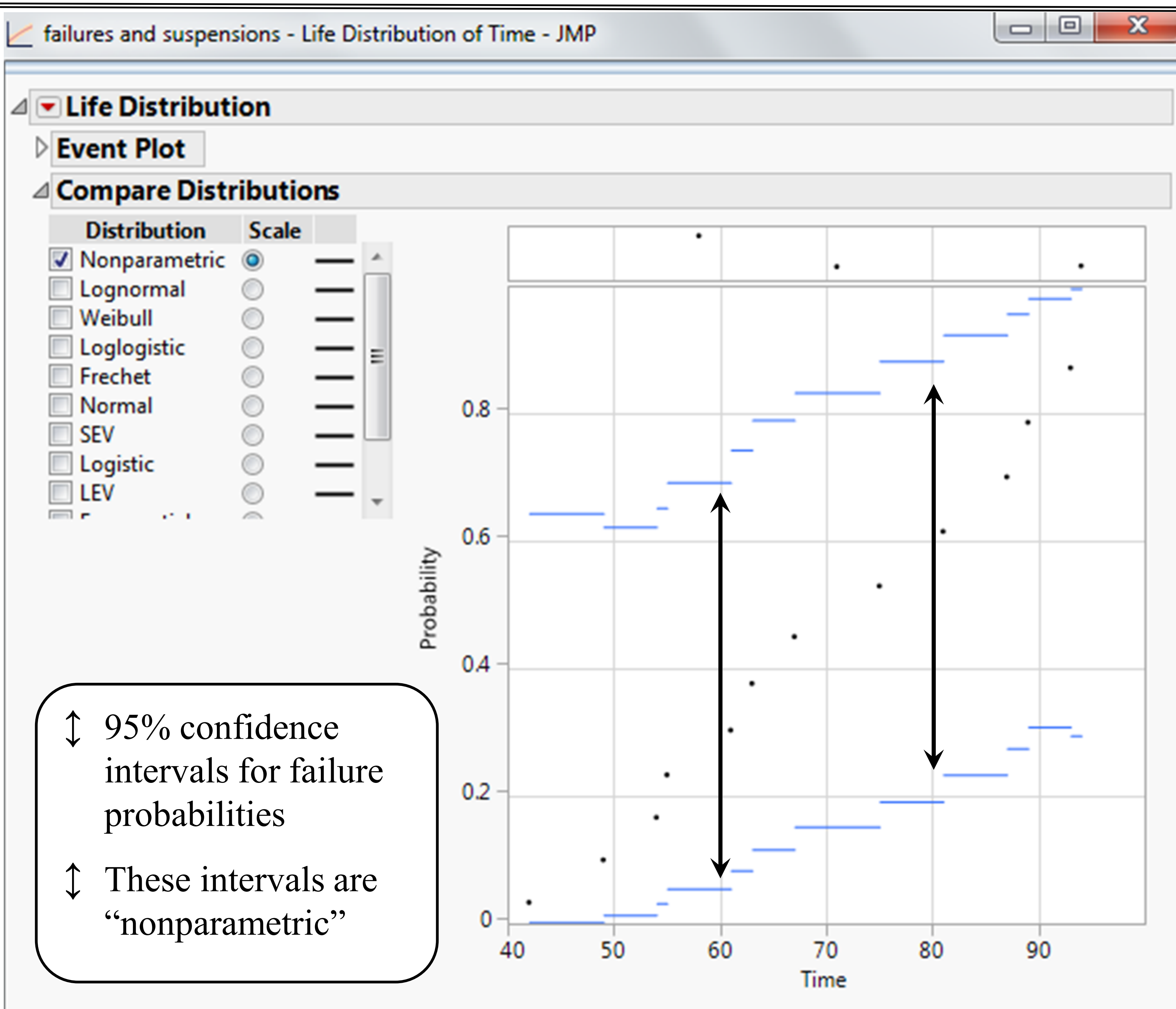


Fitting life distributions (cont'd)



- CDF plotting the failures
- Shows the corners of the steps, but not the "staircase"

Fitting life distributions (cont'd)



- 95% confidence intervals for failure probabilities
- These intervals are "nonparametric"

This analysis is referred to as *nonparametric*, meaning that it is not based on a statistical model (such as the ones listed on the left.) This is a good thing, because statistical models can be wrong. However, there are drawbacks:

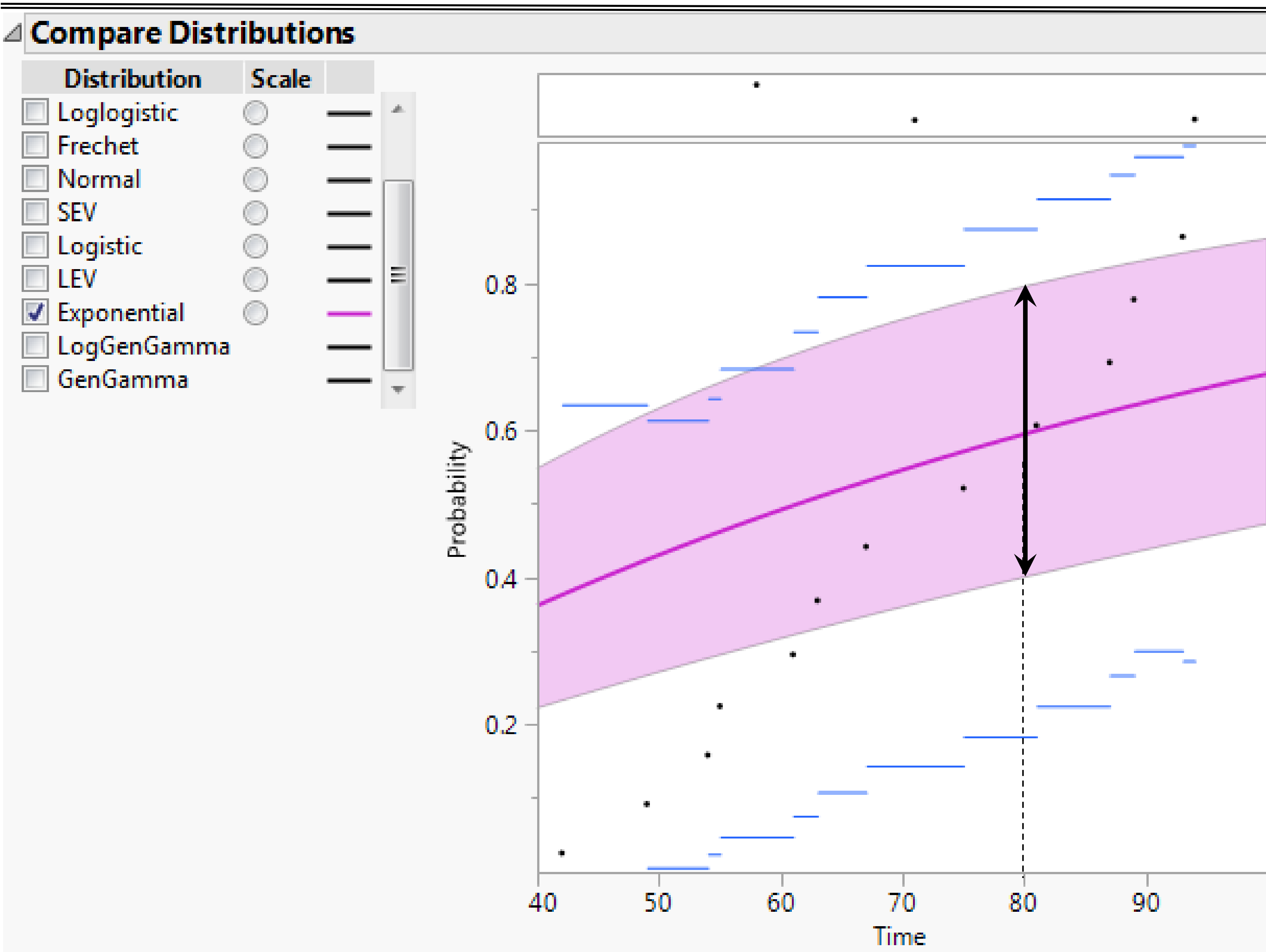
a) The nonparametric CDF is discontinuous.

b) Large numbers of failures are required to get margins of error small enough to be useful.

In practice, it is preferable to use a statistical model that fits the data well. This provides a continuous estimate of the failure function and smaller margins of error.

You can change the confidence level by selecting *Change Confidence Level* on the menu produced by the red triangle next to *Life Distribution*.

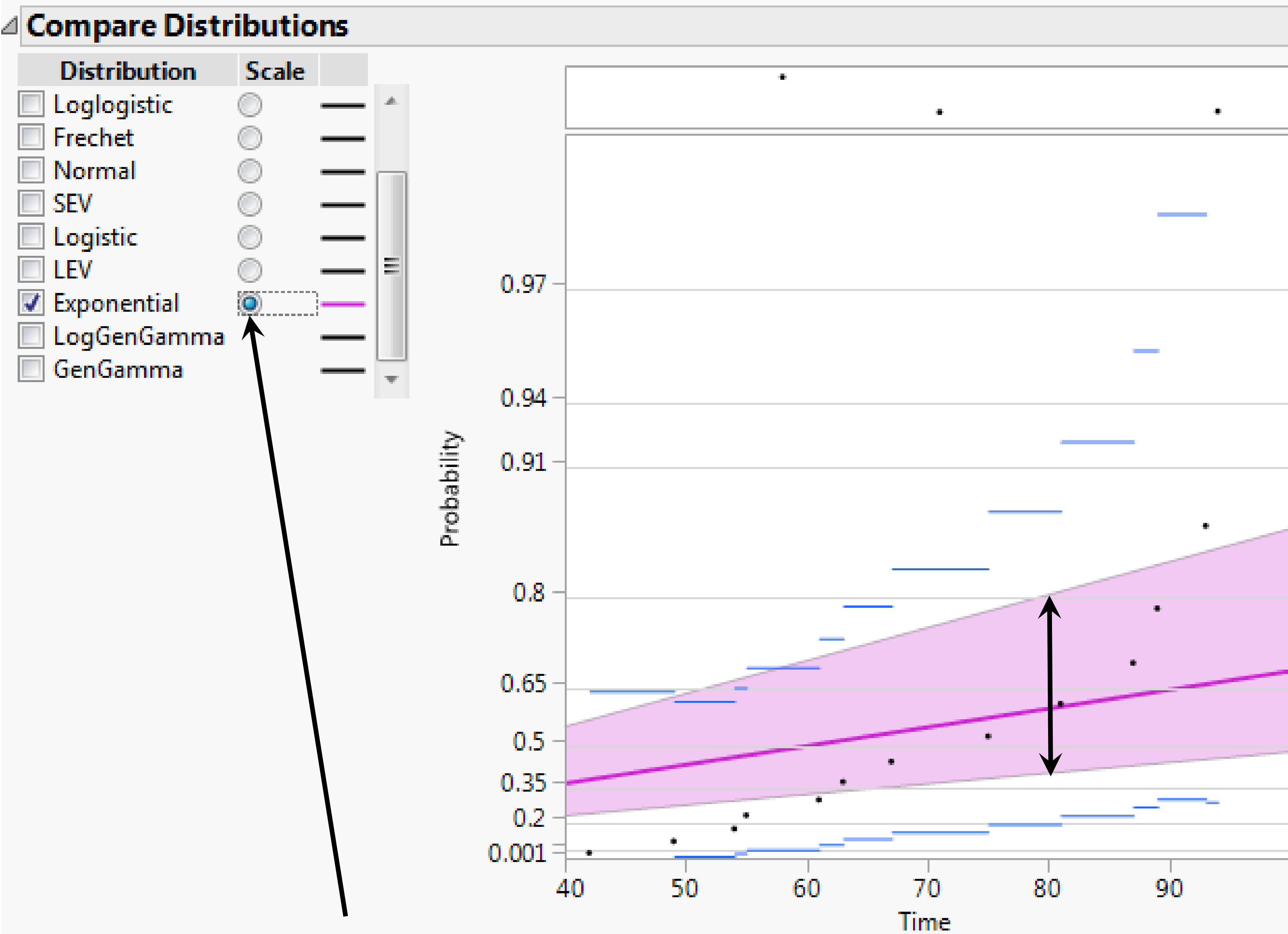
Exponential fit — linear probability scale



95% confidence interval for $F(80)$ based on the Exponential model

Bad fit – the Exponential failure curve doesn't match the data

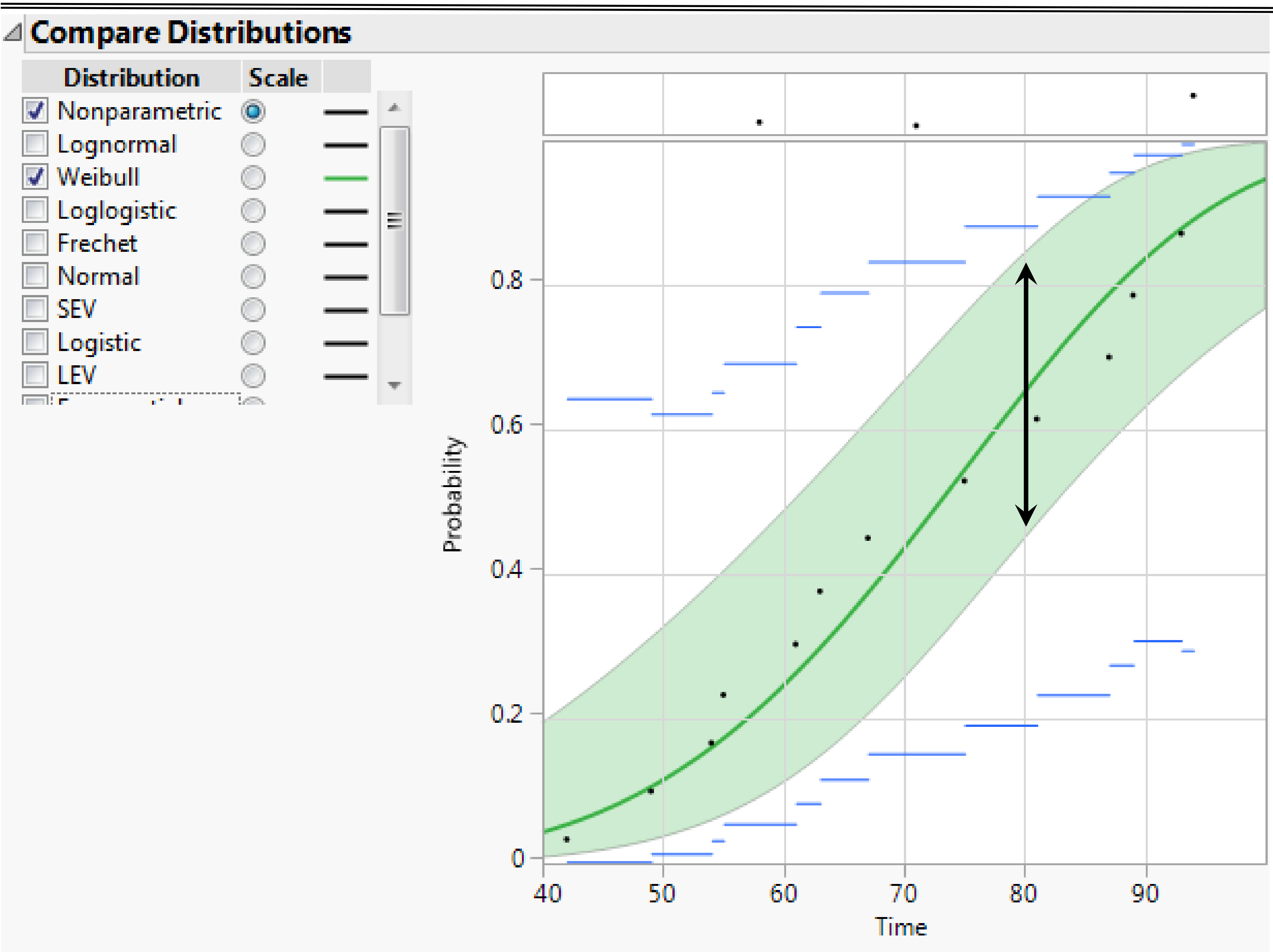
Exponential fit — Exponential probability scale



95% confidence interval for $F(80)$ based on the Exponential model

- The *Scale* button allows the failure curve to plot as a straight line
- This used to be the only way to plot failure curves

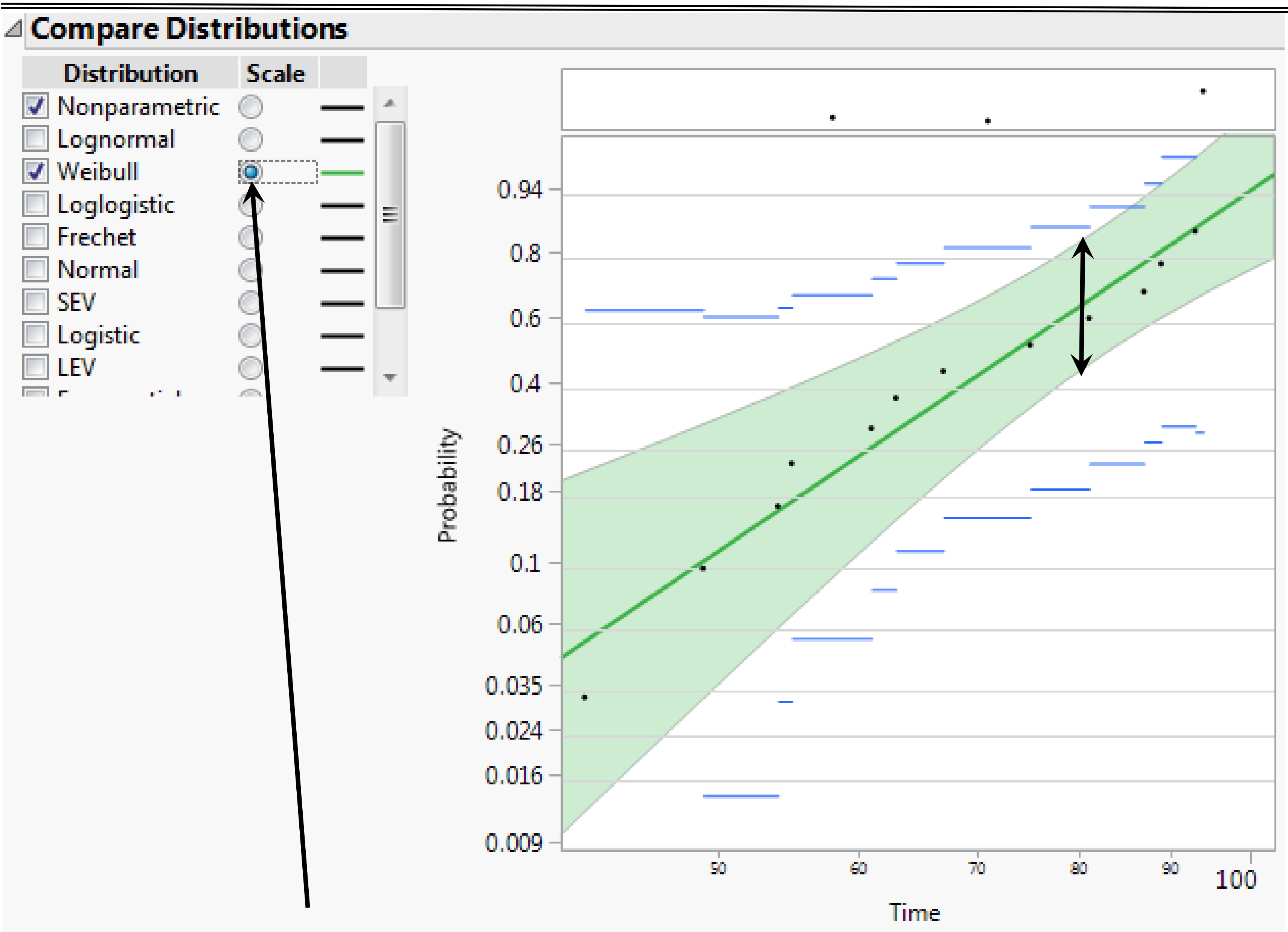
Weibull fit — linear probability scale



95% confidence interval for $F(80)$ based on the Weibull model

A better fit

Weibull fit — Weibull probability scale



95% confidence interval for $F(80)$ based on the Weibull model

- The *Scale* button allows the failure curve to plot as a straight line
- This used to be the only way to plot failure curves

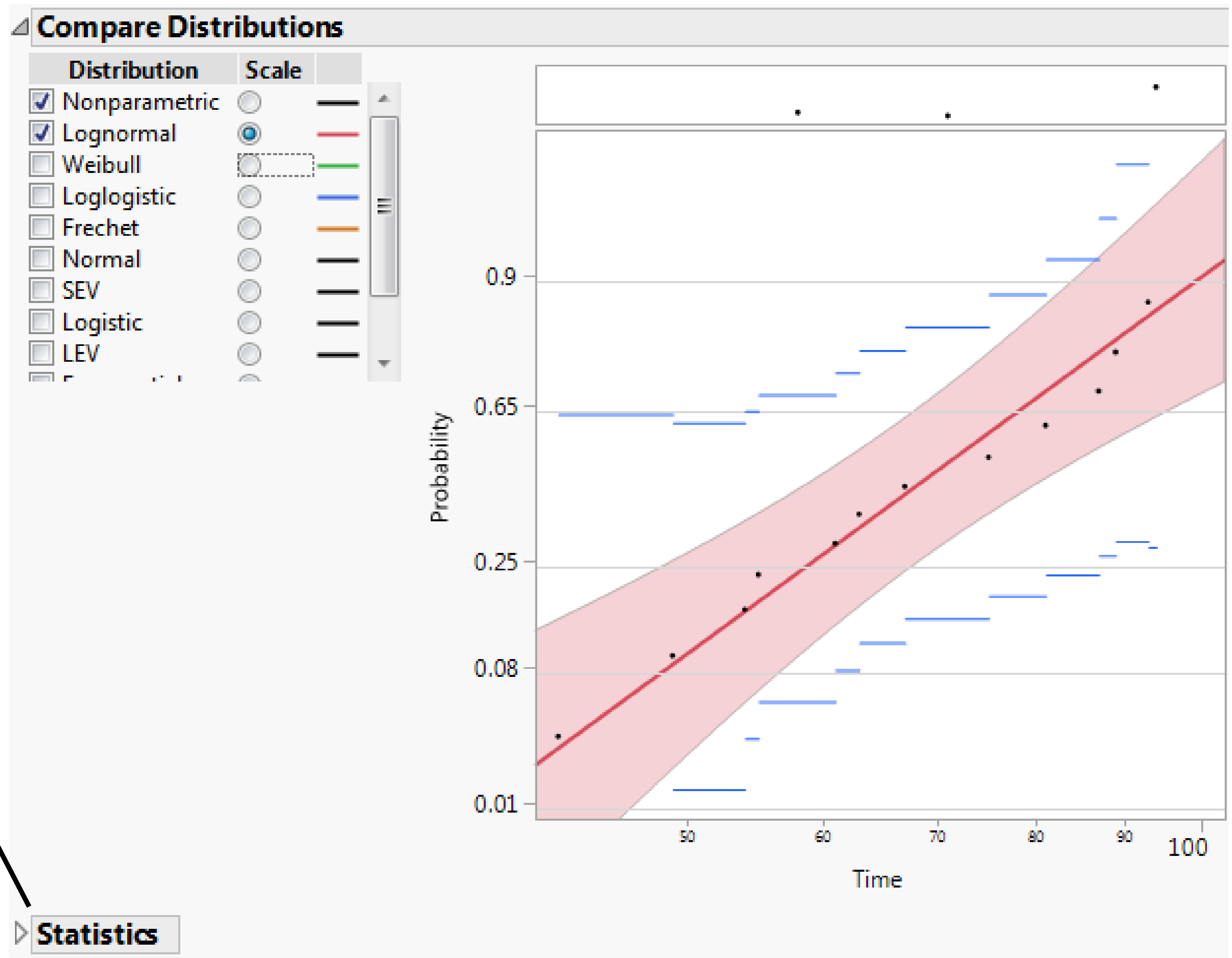
Finding and using the best fitting distribution

- Click the *Life Distribution* red triangle → Fit All Nonnegative*

- JMP plots the best fitting model on the corresponding probability scale

- In this case, *Lognormal* gives the best fit

- See next slide



*You can't have a negative time to failure!

Best fitting distribution (cont'd)

Statistics

Model Comparisons

Distribution	AICc	-2Loglikelihood	BIC
Lognormal	112.6	107.57926	112.99536
Weibull	112.8	107.81732	113.23342
Loglogistic	113.3	108.33193	113.74804
Frechet	113.8	108.75681	114.17291
Generalized Gamma	115.7	107.51791	115.64206
Exponential	133.4	131.06658	133.77463

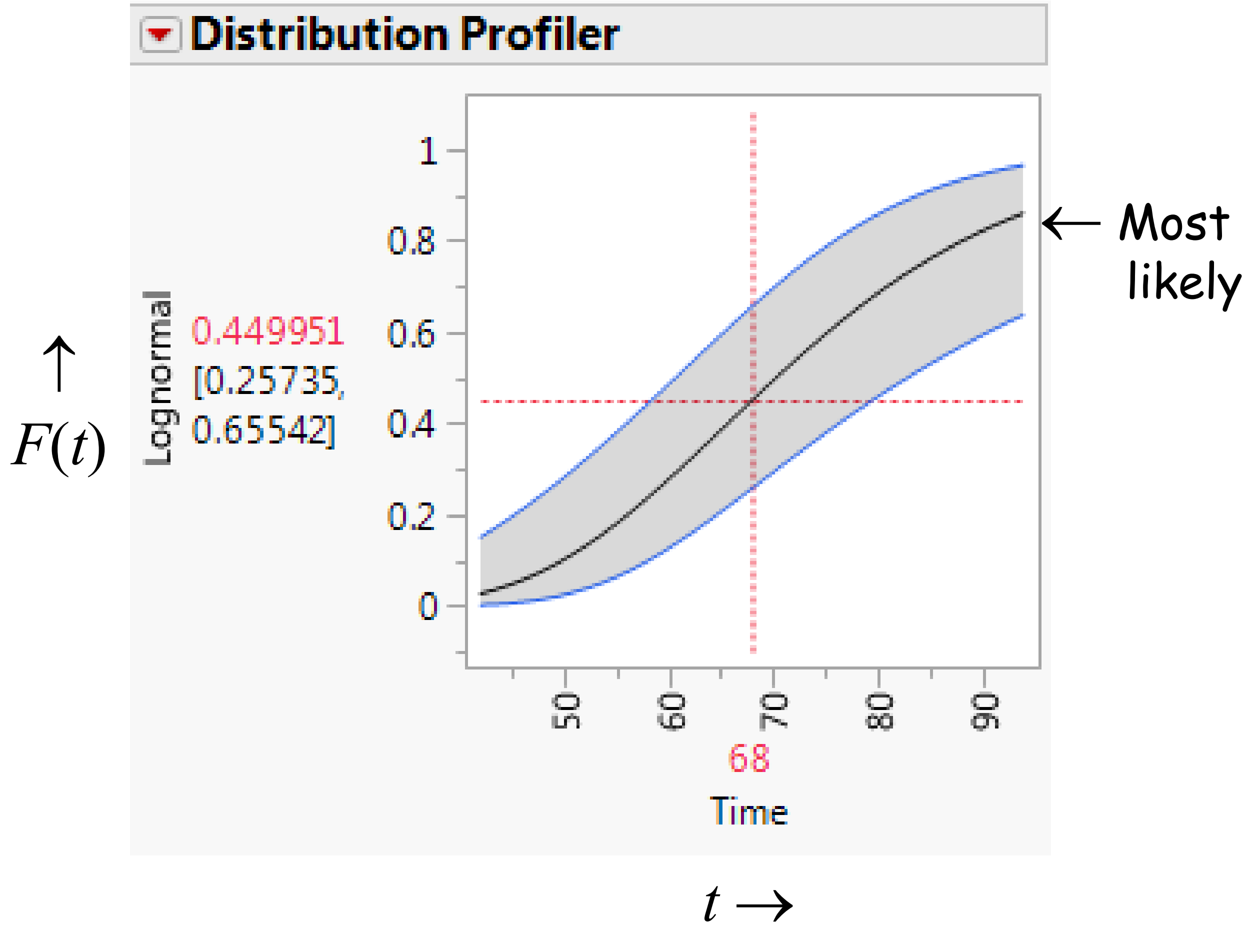
- As before, models are ranked by AIC (smaller is better)
- As before, round the AIC values to the nearest tenth
- In this case, *Lognormal* gives the best fit

The distribution profiler

- $F(t)$ is the probability that an item from this population will fail *before* time t

- The middle curve is the *most likely* value of $F(t)$

- For example, the most likely value of $F(68)$ is 0.45 (45%) (shown in red on the left side of the profiler)

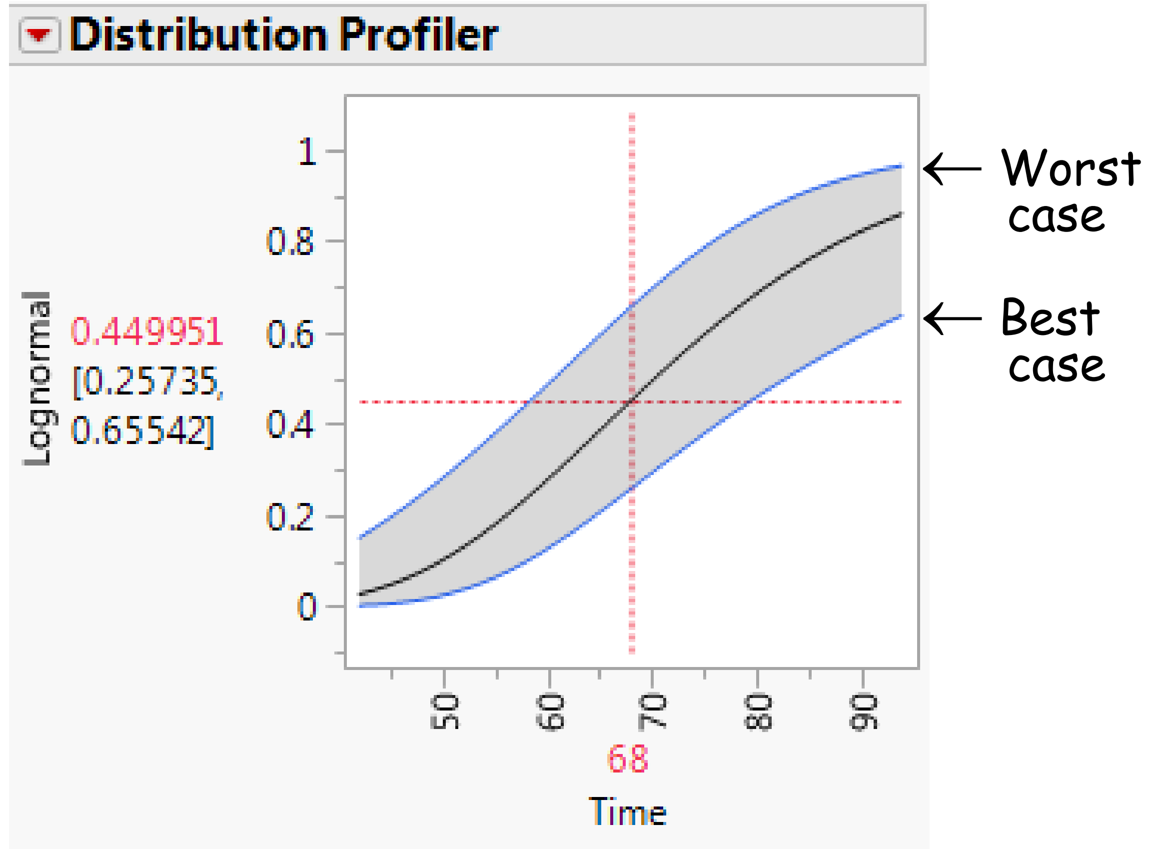


- The *reliability* function $R(t)$ is defined as $1 - F(t)$

- $R(t)$ is the probability that an item from this population will not fail until *after* time t

- For example, $R(68) = 0.55$ (55%)

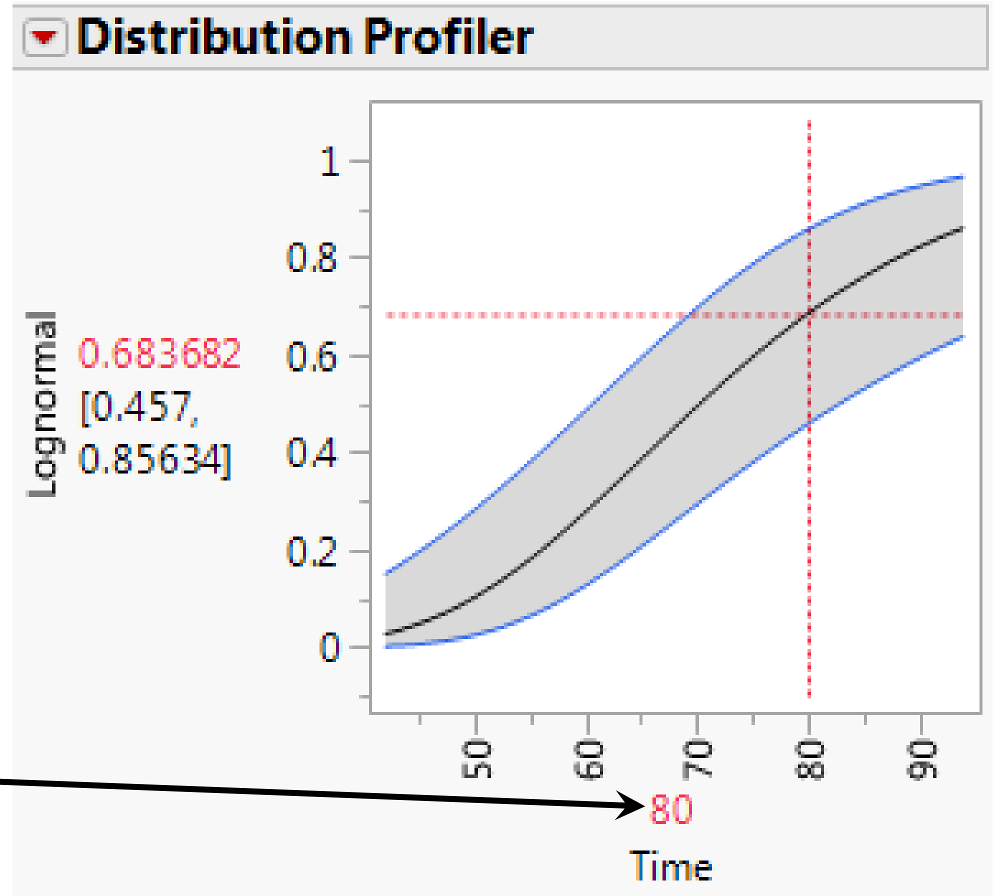
Distribution profiler (cont'd)



- The upper and lower curves give 95% confidence intervals for $F(t)$
- The upper curve gives the *worst case* value of $F(t)$*
- For example, the worst case value of $F(68)$ is 0.655 (65.5%)
- The lower curve gives the *best case* value of $F(t)$**
- For example, the best case value of $F(68) = 0.257$ (2.57%)

*For Engineering. **For Sales.

Distribution profiler (cont'd)



- Suppose we are interested in $F(80)$
- Change the value 68 to 80 (click and edit)
- The most likely value of $F(80)$ is 68.4
- The worst case value of $F(t)$ is 85.6%
- The best case value of $F(80)$ is 45.7%

Exercise 5.1

Data sets \ print life. The “time” to failure is *Pages*.

a) Identify the best fitting non-negative distribution. Use that distribution to answer the following questions.

b) What is the most likely value of $F(10{,}000)$?

c) With 95% confidence, what is the worst-case value of $F(10{,}000)$?

d) Save the analysis script, close and save the data table.

Exercise 5.2

Data sets \ probe reliability. The "time" to failure is *Hits*.

a) Identify the best fitting non-negative distribution. Use that distribution to answer the following questions.

b) What is the most likely value of $F(200)$?

c) With 95% confidence, what is the worst-case value of $F(200)$?

d) Save the analysis script, close and save the data table.

Exercise 5.3

Data sets \ field reliability. The time to failure is *Days in field*.

a) Identify the best fitting non-negative distribution. Use that distribution to answer the following questions.

b) What is the most likely value of $F(365)$?

c) With 95% confidence, what is the worst-case value of $F(365)$?

d) Save the analysis script, close and save the data table.

This page intentionally left blank

6 Categorical MSA Without Standards

- It is preferable to base nominal MSA on a set of items whose true status is known (standards)
- With standards, we can determine the probabilities of passing bad items and failing good ones
- Creating standards can be difficult and time consuming
- Lacking standards, "% agreement within and between appraisers" can serve as a proxy for "% agreement with standard"

Example 1

Data sets \ pass-fail no stds

msa pass-fail no stds
Notes C:\Documents and Se

Columns (5/0)
- Session
- Part
- Insp A
- Insp B
- Insp C

Rows
All rows 150

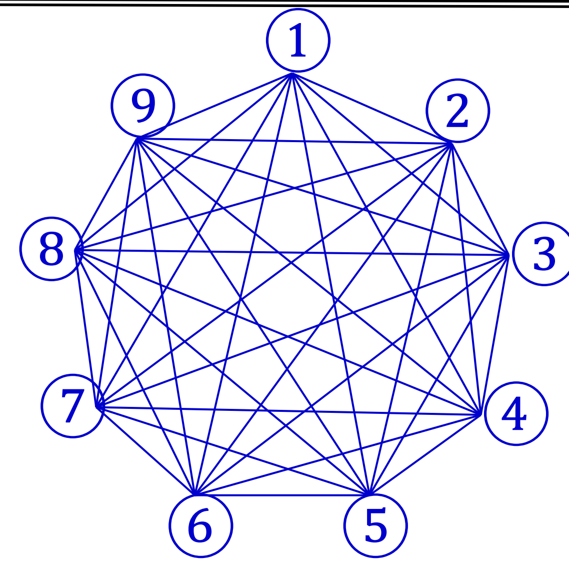
	Session	Part	Insp A	Insp B	Insp C
1	1	1	P	P	P
2	2	1	P	P	P
3	3	1	P	P	P
4	1	2	P	P	P
5	2	2	P	P	P
6	3	2	P	P	P
7	1	3	F	F	F
8	2	3	F	F	F
9	3	3	F	F	F
10	1	4	F	F	F
11	2	4	F	F	F
12	3	4	F	F	F
13	1	5	F	F	F
14	2	5	F	F	F
15	3	5	F	F	F
16	1	6	P	P	P
17	2	6	P	P	F
18	3	6	F	F	F
19	1	7	P	P	P
20	2	7	P	P	F
21	3	7	P	P	P
22	1	8	P	P	P
23	2	8	P	P	P
24	3	8	P	P	P
25	1	9	F	F	F
26	2	9	F	F	F
27	3	9	F	F	F
28	1	10	P	P	P

- 50 parts
- Appraisers A, B, C
- 3 inspections per part per appraiser
- *Part* is actually nominal, since part numbers are only identifiers without a numerical relationship. Change by:
  - Right click on next to Part and Select Nominal, or
  - Right click on field name "Part" > Column Info > Data Type = Character
  - Please be aware that JMP is occasionally inconsistent in its terminology

Agreement within & between appraisers

	Session	Part	Insp A	Insp B	Insp C
1	1	1	P	P	P
2	2	1	P	P	P
3	3	1	P	P	P
4	1	2	P	P	P
5	2	2	P	P	P
6	3	2	P	P	P
7	1	3	F	F	F
8	2	3	F	F	F
9	3	3	F	F	F
10	1	4	F	F	F
11	2	4	F	F	F
12	3	4	F	F	F
13	1	5	F	F	F
14	2	5	F	F	F
15	3	5	F	F	F
16	1	6	P	P	P
17	2	6	P	P	F
18	3	6	F	F	F
19	1	7	P	P	P
20	2	7	P	P	F
21	3	7	P	P	P
22	1	8	P	P	P
23	2	8	P	P	P
24	3	8	P	P	P
25	1	9	F	F	F
26	2	9	F	F	F
27	3	9	F	F	F
28	1	10	P	P	P
29	2	10	P	P	P
30	3	10	P	P	P
31	1	11	P	P	P
32	2	11	P	P	P
33	3	11	P	P	P
34	1	12	F	F	F
35	2	12	F	F	P
36	3	12	F	F	F

- 100% agreement

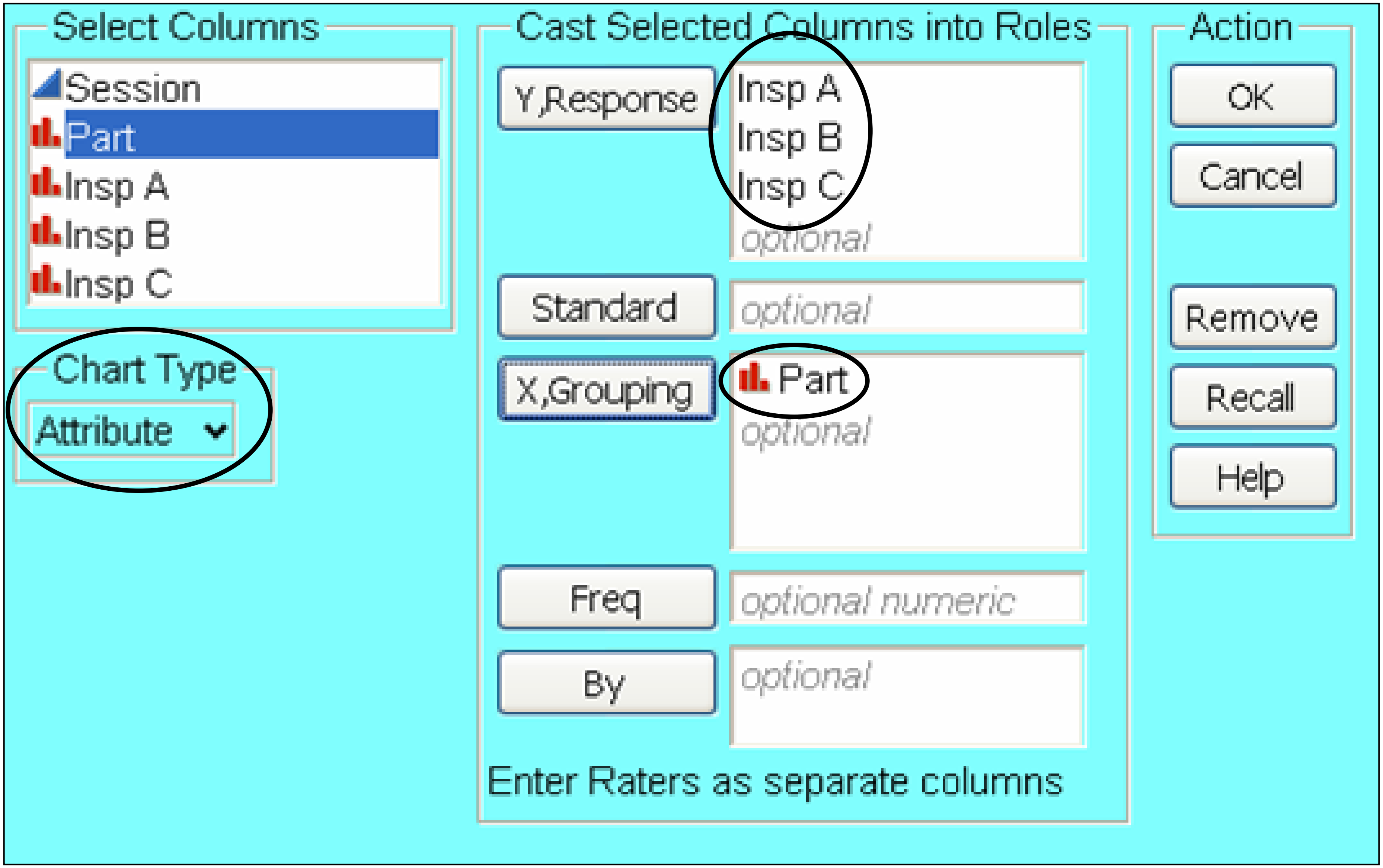


- 36 opportunities for pairwise agreement
- 16 pairwise agreements
- Agreement = 16/36 = 0.444

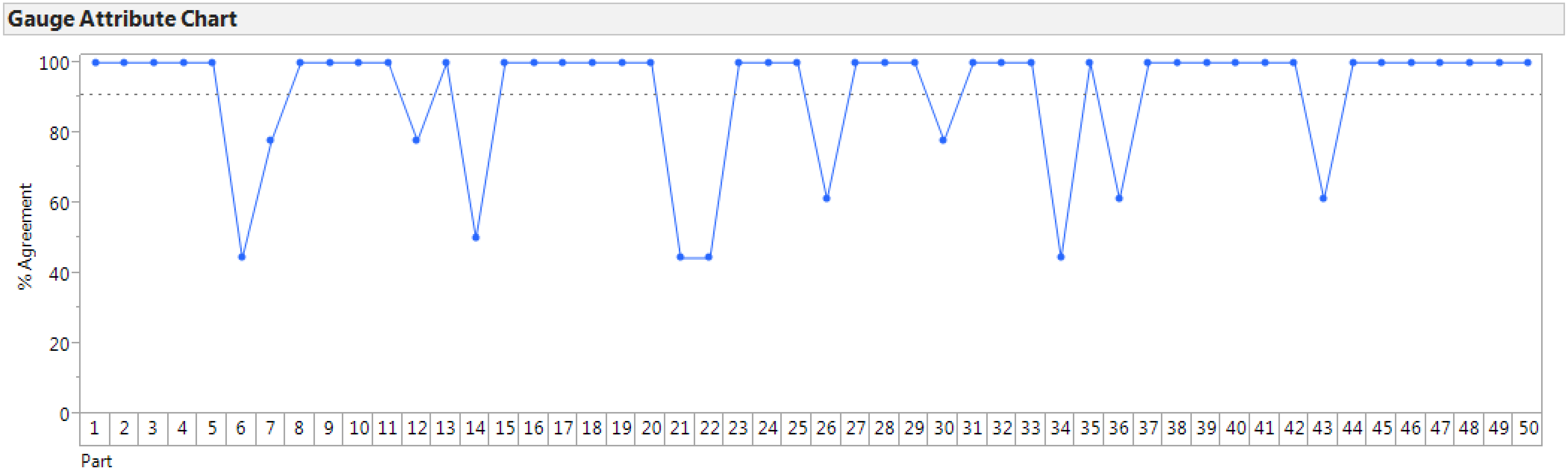
- 36 opportunities for pairwise agreement
- 8 pairwise disagreements
- Agreement = 28/36 = 0.778

Analyzing a categorical MSA without standards

Analyze → Quality and Process → Variability / Attribute Gauge Chart

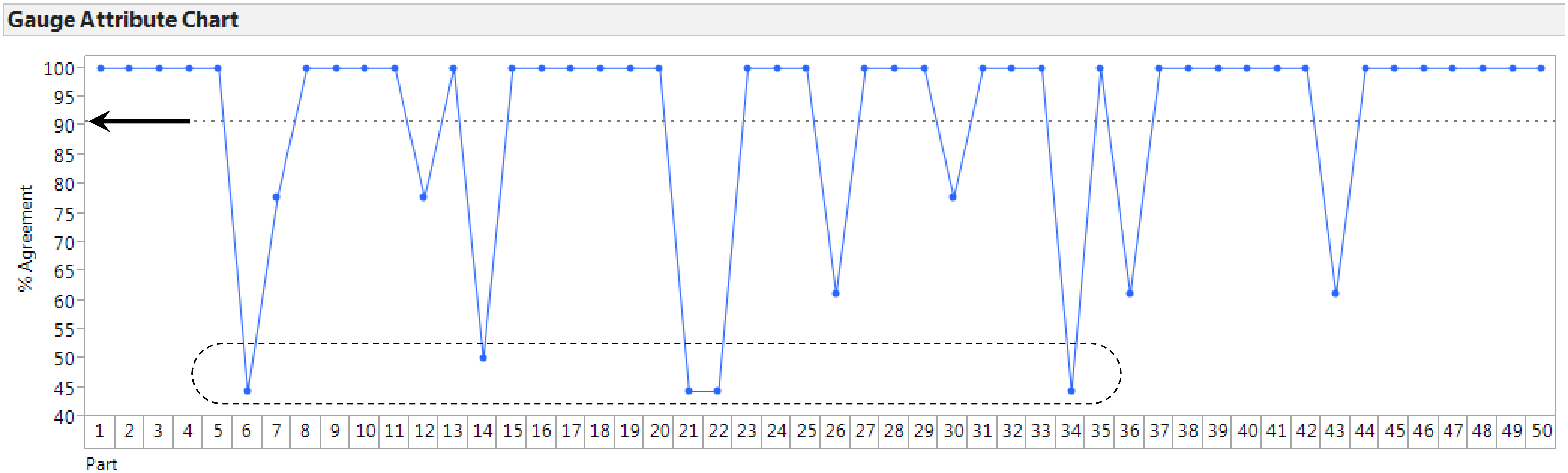


Agreement report



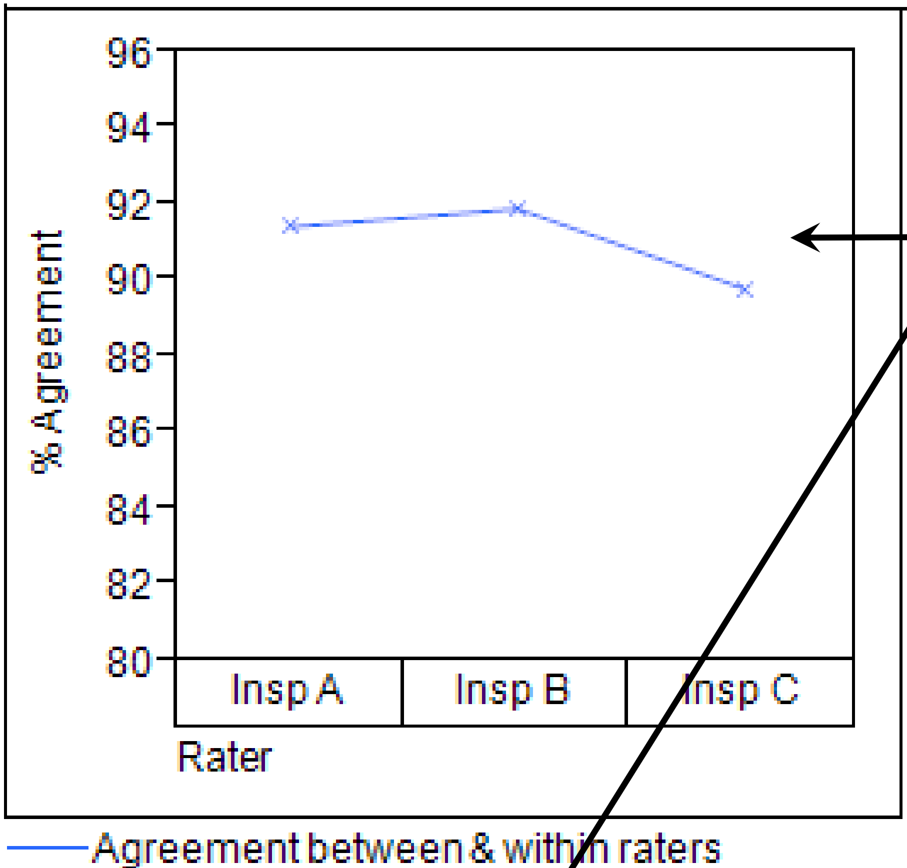
- Plot of the agreement percentages for the items in the study
- It is helpful to rescale the vertical axis
- See next slide

Agreement report (cont'd)



- The horizontal dotted line marks the "agreement grand mean"
- In this example, the agreement grand mean is a little over 90 (read off graph)
- Nowhere in the report is this number printed — bad JMP!
- If the agreement grand mean is too low, follow-up should focus on the items with the lowest % agreement
- There are no recognized standards for the agreement grand mean. A lower bound of 95% is fairly common. 99% is often used in applications involving safety.

Agreement report (cont'd)



Agreement Report

Rater	% Agreement	95% Lower CI	95% Upper CI
Insp A	91.4286	89.5082	93.0248
Insp B	91.9048	90.0502	93.4388
Insp C	89.8095	87.6057	91.6588

Number Inspected	Number Matched	% Agreement	95% Lower CI
50	39	78.000	64.758

- These are the agreement percentages for each appraiser
- The appraiser with the lowest percentage represents the greatest opportunity for improvement
- Sometimes the smallest % agreement among the appraisers is used as the metric

- Percentage of items for which agreement was 100%
- This should not be used as a metric

Save the script, close and save the data table.

Agreement Comparisons:
Each rater compared to all others, using Kappa statistics

$K \geq 0.9 \rightarrow$ Good measurement system
$K \leq 0.7 \rightarrow$ Bad measurement system
$0.7 \leq K \leq 0.9 \rightarrow$ Marginal measurement system

Agreement across Categories:
Agreement in classification corrected for the amount of agreement which would be expected by chance. Kappa assesses the agreement between a fixed number of raters when classifying items.

When $K = 1$, perfect agreement exists.
When $K = 0$, agreement is the same as would be expected by chance.
When $K < 0$, agreement is weaker than expected by chance; this rarely occurs and usually means that the appraisers have different definitions of the assigned categories.

Example 2

Data sets \ application rating no stds

application rating no stds
Notes C:\Documents and Se

Columns (4/0)
Application
Session
Appraiser
Rating

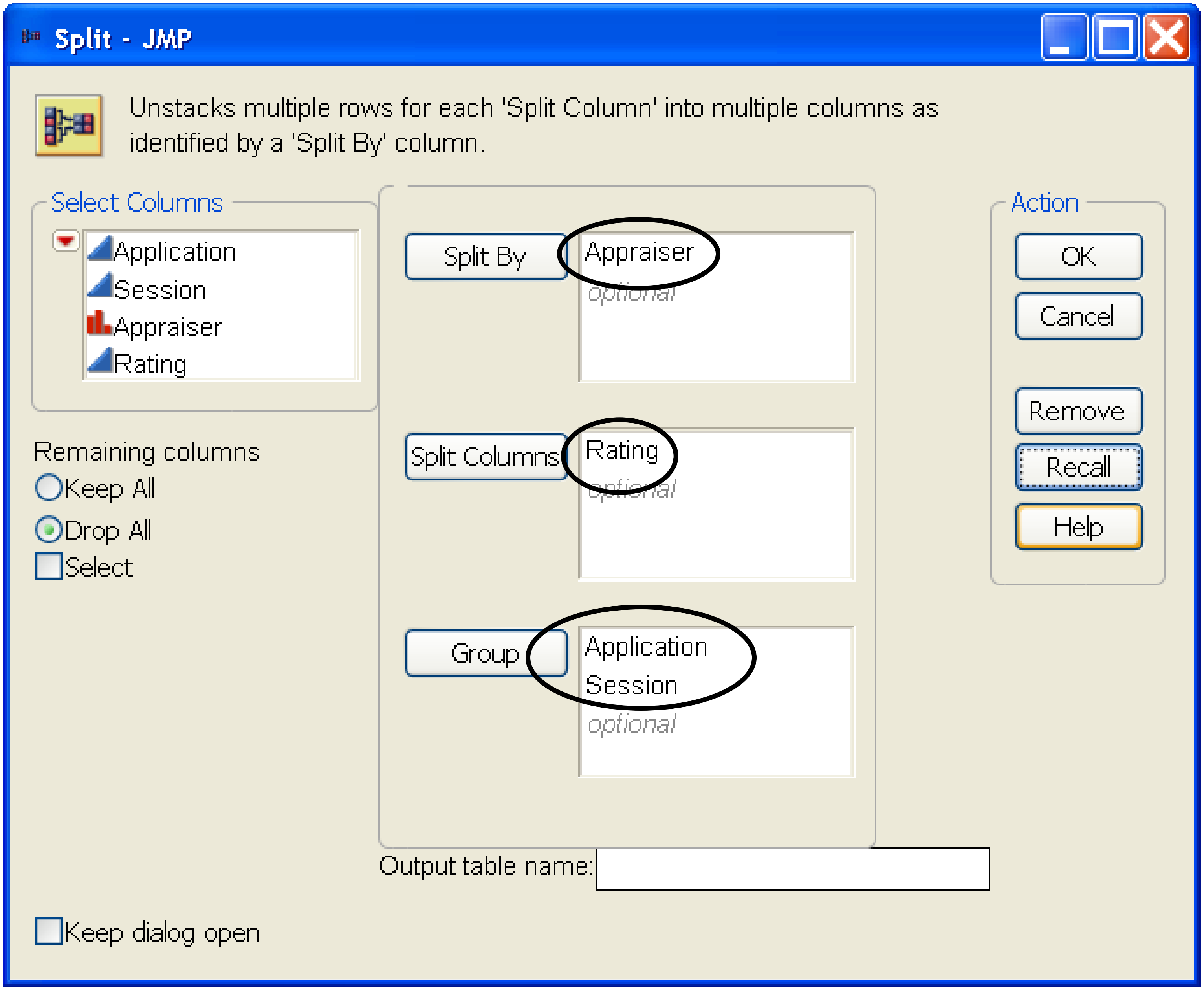
Rows
All rows 150

	Application	Session	Appraiser	Rating
1	1	1	Simpson	5
2	1	1	Montgomery	5
3	1	1	Holmes	5
4	1	1	Duncan	4
5	1	1	Hayes	5
6	2	1	Simpson	2
7	2	1	Montgomery	2
8	2	1	Holmes	2
9	2	1	Duncan	1
10	2	1	Hayes	2
11	3	1	Simpson	4
12	3	1	Montgomery	3
13	3	1	Holmes	3
14	3	1	Duncan	3
15	3	1	Hayes	3
16	4	1	Simpson	1
17	4	1	Montgomery	1
18	4	1	Holmes	1
19	4	1	Duncan	1
20	4	1	Hayes	1
21	5	1	Simpson	3
22	5	1	Montgomery	3
23	5	1	Holmes	3
24	5	1	Duncan	2
25	5	1	Hayes	3
26	6	1	Simpson	4
27	6	1	Montgomery	4

- 15 employment applications
- 5 appraisers
- 2 inspections per application per appraiser
- Five point scale, higher is better
- Change *Rating* to nominal
- For categorical MSA, we must *unstack* this data table

Unstacking a data table

Tables → Split



Example 2 in required format

Untitled 12
Source

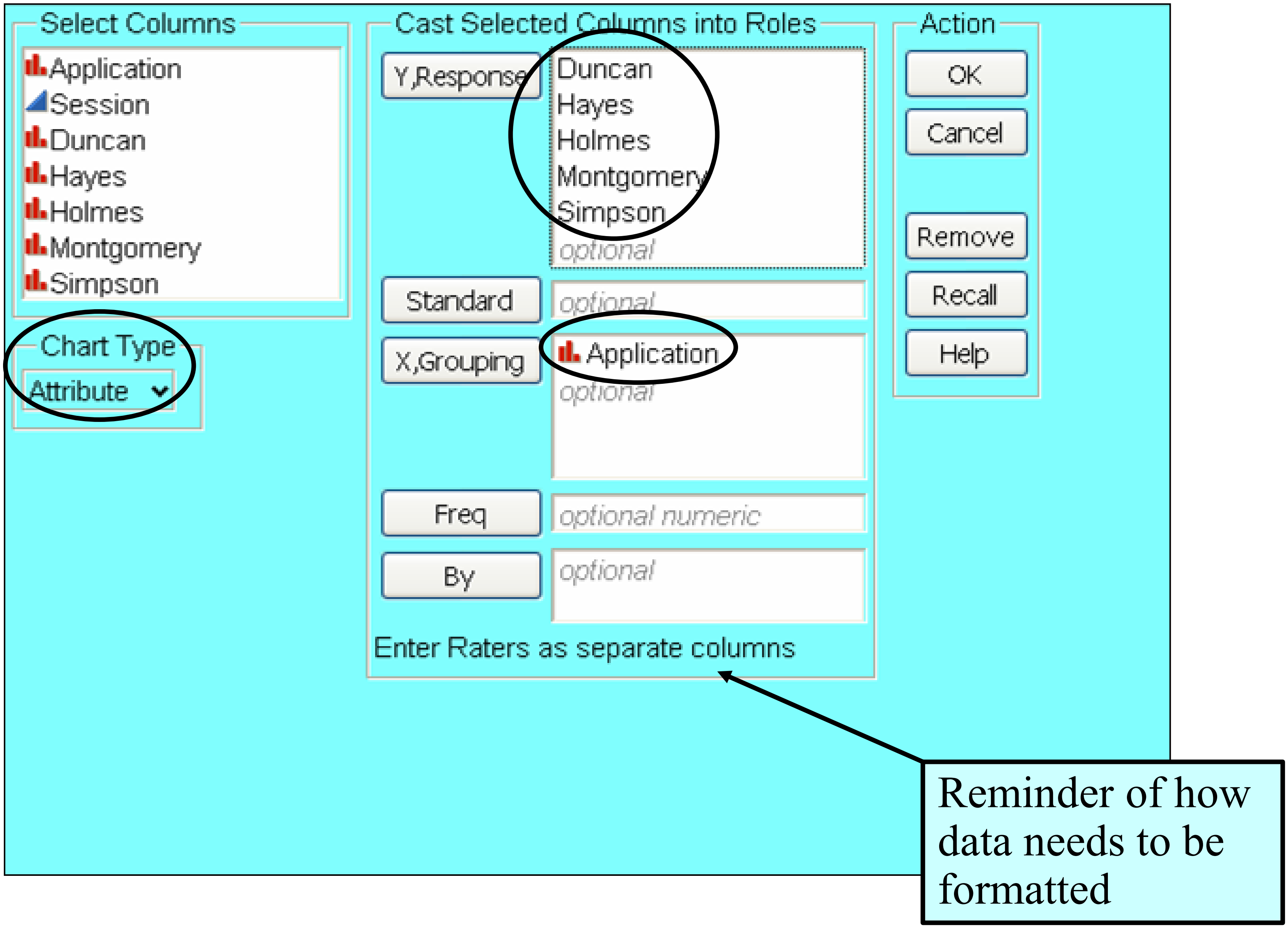
Columns (7/0)
Application
Session
Duncan
Hayes
Holmes
Montgomery
Simpson

Rows
All rows 30

	Application	Session	Duncan	Hayes	Holmes	Montgomery	Simpson
1	1	1	4	5	5	5	5
2	2	1	1	2	2	2	2
3	3	1	3	3	3	3	4
4	4	1	1	1	1	1	1
5	5	1	2	3	3	3	3
6	6	1	4	4	4	4	4
7	7	1	4	5	5	5	5
8	8	1	3	3	3	3	3
9	9	1	1	2	2	2	2
10	10	1	3	5	4	4	4
11	11	1	1	2	1	1	1
12	12	1	2	3	3	3	3
13	13	1	5	5	5	5	5
14	14	1	2	2	2	2	2
15	15	1	4	4	4	4	4
16	1	2	4	5	5	5	4
17	2	2	1	2	2	2	2
18	3	2	3	3	4	4	4
19	4	2	1	1	1	1	1
20	5	2	2	3	3	3	3
21	6	2	4	4	4	5	5
22	7	2	4	5	5	5	5
23	8	2	3	4	3	3	3
24	9	2	1	2	2	2	2
25	10	2	3	5	4	4	4
26	11	2	1	2	1	1	1
27	12	2	2	3	3	3	3
28	13	2	5	5	5	5	5

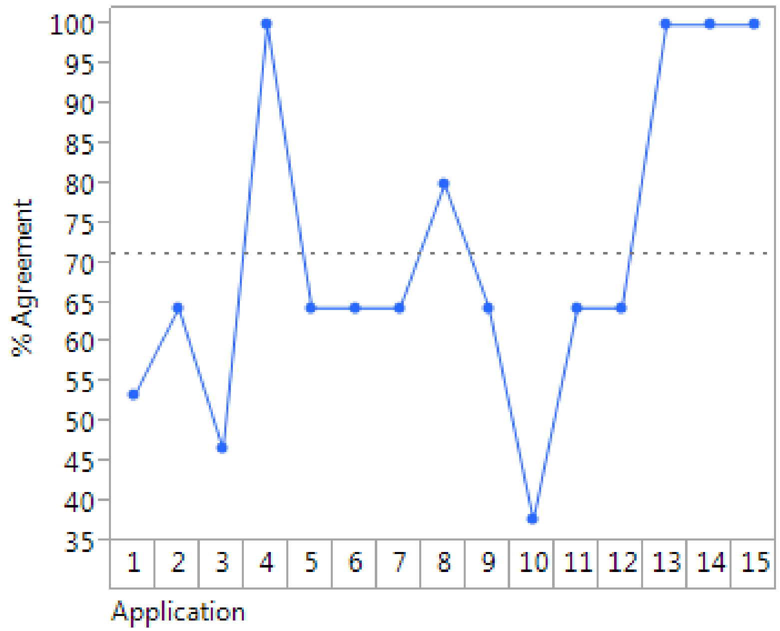
Example 2 (cont'd)

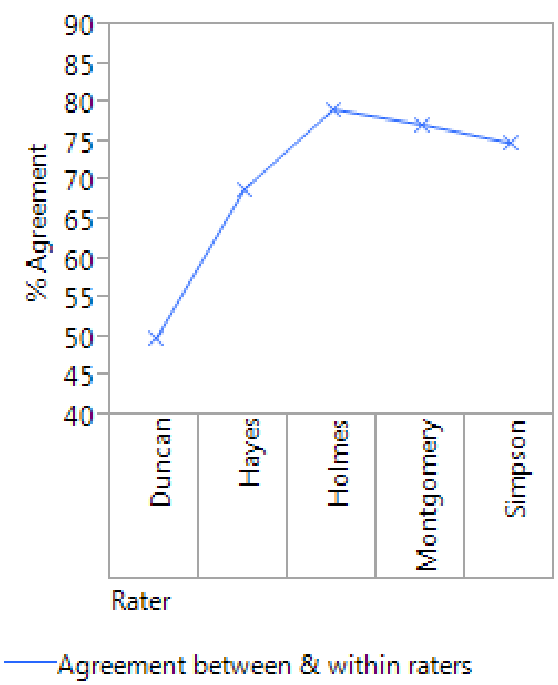
Analyze → Quality and Process → Variability / Attribute Gauge Chart



Example 2 (cont'd)

Gauge Attribute Chart





- The agreement grand mean is about 71 – way too low
- Follow-up: focus on application 1, 3 and 10
- Greatest opportunity for improvement: further training of Duncan and Hayes

Agreement Report

Rater	% Agreement	95% Lower CI	95% Upper CI
Duncan	49.8039	27.2673	72.4205
Hayes	69.0196	43.9053	86.3784
Holmes	79.2157	53.9935	92.5247
Montgomery	77.2549	51.9716	91.4246
Simpson	74.9020	49.5997	90.0500

Number Inspected	Number Matched	% Agreement	95% Lower CI	95% Upper CI
15	4	26.667	10.897	51.950

Save the analysis script to the data table, close and save the data table as:

application rating no stds unstacked

Exercise 6.1

Data sets \ print samples 1 no stds. In this study 3 appraisers inspected 18 print samples 3 times each.

a) Reformat the file as needed and run the analysis.

b) Record the approximate agreement grand mean.

c) Which sample(s) would be most useful in follow-up?

d) Of the 3 appraisers, which has the highest % agreement? What is the highest % agreement?

e) Save the script, close and save the data table as *print samples 1 no stds unstacked.*

Exercise 6.2

Data sets \ print samples 2 no stds. This is the follow-up study after the appraisers received additional training.

a) Reformat the file as needed and run the analysis.

b) Record the approximate agreement grand mean.

c) Of the 3 appraisers, which has the lowest % agreement? What is the lowest % agreement?

d) Save the script, close and save the data table as *print samples 2 no stds unstacked.*

7 Comparing Populations – Continuous Y

- Example of comparing populations
- Analysis of variance (ANOVA) for comparing populations
- Interpreting P-values
- Degrees of freedom for signal and noise
- ANOVA in JMP

Y variables are characteristics of parts, products or transactions that determine customer satisfaction, or lack thereof. They provide the data from which project metrics can be computed.

Comparison of statistical populations is equivalent to $Y = f(X)$ analysis where the X variable is categorical. The distinct values of the X variable define the populations or sub-populations to be compared.

JMP uses the term *continuous* for quantitative variables. Except in the DOE section, JMP uses the term *nominal* for categorical variables.

Example of comparing populations

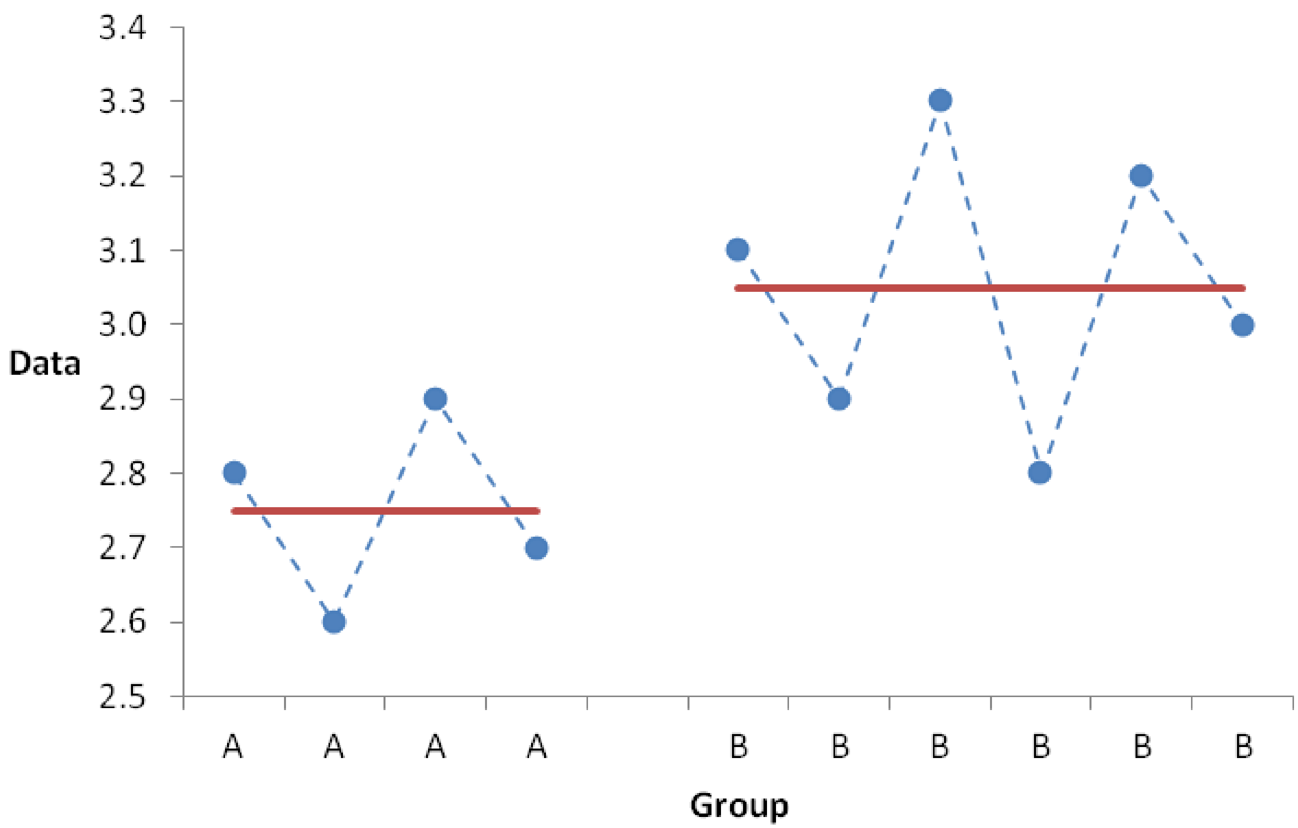
Data sets \ Anova 2 groups

Group	Data	Avg.	SD
A	2.8	2.75	0.129
A	2.6		
A	2.9		
A	2.7		
B	3.1	3.05	0.187
B	2.9		
B	3.3		
B	2.8		
B	3.2		
B	3.0		

- We have two groups of data
- Could be a *before/after* comparison
- Could be a *stratification* analysis

- The sample means for the two groups are different
- Is this enough to conclude that the *population* means are different?

Example (cont'd)



- Plotting the data is helpful, but it doesn't give a definitive answer

- How far apart do the sample means have to be before we can say the population means are different?

- How do we take into account the *scatter* around the means?

LSSV2 student files \ ANOVA two groups

B	C	D	E	F	G	H	I	J	K	L	M
Group		Data		Grand mean		Difference		Group		Error	
A		2.8		2.93		-0.13		-0.18		0.05	
A		2.6		2.93		-0.33		-0.18		-0.15	
A		2.9		2.93		-0.03		-0.18		0.15	
A		2.7		2.93		-0.23		-0.18		-0.05	
B		3.1	−	2.93	=	0.17	=	0.12	+	0.05	
B		2.9		2.93		-0.03		0.12		-0.15	
B		3.3		2.93		0.37		0.12		0.25	
B		2.8		2.93		-0.13		0.12		-0.25	
B		3.2		2.93		0.27		0.12		0.15	
B		3.0		2.93		0.07		0.12		-0.05	

ANOVA (1 of 6, cont'd)

This worksheet shows all the calculations used to determine, based on the data, whether or not the population means are different.

The first step is to calculate the *Difference* column by subtracting the grand mean from the *Data* column. The *Difference* is then decomposed into *Group* (the "signal") plus *Error* (the "noise").

The *Group* column captures the portion of total variation caused by the difference between the sample means.

The *Error* column captures the rest of the variation, variously called the *residual*, *unexplained*, or *noise* variation.

LSSV2 student files \ ANOVA two groups

Group	Data		Grand mean		Difference		Group		Error
A	2.8		2.93		-0.13		-0.18		0.05
A	2.6		2.93		-0.33		-0.18		-0.15
A	2.9		2.93		-0.03		-0.18		0.15
A	2.7		2.93		-0.23		-0.18		-0.05
B	3.1	−	2.93	=	0.17	=	0.12	+	0.05
B	2.9		2.93		-0.03		0.12		-0.15
B	3.3		2.93		0.37		0.12		0.25
B	2.8		2.93		-0.13		0.12		-0.25
B	3.2		2.93		0.27		0.12		0.15
B	3.0		2.93		0.07		0.12		-0.05
Degrees of freedom (DF)	10	−	1	=	9	=	1	+	8

ANOVA (2 of 6, cont'd)

The *Data* column consists of 10 mathematically independent quantities. We describe this by saying it has 10 *degrees of freedom* (DF).

The *Grand mean* column consists of 10 values, but they are all identical. This column has 1 DF.

The *Difference* column contains 10 values, but they are mathematically constrained to sum to 0. This column contains only 9 independent quantities, so it has 9 DF.

The *Group* column inherits the zero-sum constraint from the *Difference* column (it must sum to zero), and it consists of only 2 distinct values. This column contains only one independent quantity, so it has 1 DF.

The *Error* column has 8 DF, because DFs have to add up.

The DFs for *Group* and *Error* play a role in determining whether or not the population means are different.

LSSV2 student files \ *ANOVA two groups*

A	B	C	D	E	F	G	H	I	J	K	L	M
	Group		**Data**		**Grand mean**		**Difference**		**Group**		**Error**	
	A		2.8		2.93		-0.13		-0.18		0.05	
	A		2.6		2.93		-0.33		-0.18		-0.15	
	A		2.9		2.93		-0.03		-0.18		0.15	
	A		2.7		2.93		-0.23		-0.18		-0.05	
	B		3.1	−	2.93	=	0.17	=	0.12	+	0.05	
	B		2.9		2.93		-0.03		0.12		-0.15	
	B		3.3		2.93		0.37		0.12		0.25	
	B		2.8		2.93		-0.13		0.12		-0.25	
	B		3.2		2.93		0.27		0.12		0.15	
	B		3.0		2.93		0.07		0.12		-0.05	
Degrees of freedom (DF)			10	−	1	=	9	=	1	+	8	
Sum of squares (SS)			86.29	−	85.85	=	0.441	=	0.216	+	0.225	
Mean square (MS)			(*SS / DF*)				0.049		0.216		0.028	

ANOVA (3 of 6, cont'd)

The sum of squares (SS) is a measure of the magnitude of each column. It is the sum of the squares of the values in a column.

The sums of squares for the *Difference*, *Group*, and *Error* columns are usually much smaller than those of the *Data* and *Grand mean* columns.

The mean square (MS) is the statistically normalized measure (averaged, in a sense) of the magnitude of each column. It is the SS for a column divided by the DF for that column.

The mean squares for the *Data* and *Grand mean* columns play no role in determining whether or not the population means are different, so the MS is usually calculated only for the *Difference*, *Group*, and *Error* columns.

LSSV2 student files \ *ANOVA two groups*

Group	Data		Grand mean		Difference		Group		Error
A	2.8		2.93		-0.13		-0.18		0.05
A	2.6		2.93		-0.33		-0.18		-0.15
A	2.9		2.93		-0.03		-0.18		0.15
A	2.7		2.93		-0.23		-0.18		-0.05
B	3.1	−	2.93	=	0.17	=	0.12	+	0.05
B	2.9		2.93		-0.03		0.12		-0.15
B	3.3		2.93		0.37		0.12		0.25
B	2.8		2.93		-0.13		0.12		-0.25
B	3.2		2.93		0.27		0.12		0.15
B	3.0		2.93		0.07		0.12		-0.05
Degrees of freedom (DF)	10	−	1	=	9	=	1	+	8
Sum of squares (SS)	86.29	−	85.85	=	0.441	=	0.216	+	0.225
Mean square (MS)	*(SS / DF)*				0.049		0.216		0.028
F ratio	*(Group MS / Error MS)*								7.680

ANOVA (4 of 6, cont'd)

The *Group* MS measures the magnitude of the variation caused by the difference between the sample means.

The *Error* MS measures the magnitude of the variation caused by everything *except* the difference between the sample means.

The *F ratio* is the *Group* MS divided by *Error* MS. It is a signal-to-noise ratio.

The larger the F ratio, the stronger the evidence of a difference between the population means.

ANOVA (5 of 6)

A	B	C	D	E	F	G	H	I	J	K	L	M
	Group		**Data**		**Grand mean**		**Difference**		**Group**		**Error**	
	A		2.8		2.93		-0.13		-0.18		0.05	
	A		2.6		2.93		-0.33		-0.18		-0.15	
	A		2.9		2.93		-0.03		-0.18		0.15	
	A		2.7		2.93		-0.23		-0.18		-0.05	
	B		3.1	−	2.93	=	0.17	=	0.12	+	0.05	
	B		2.9		2.93		-0.03		0.12		-0.15	
	B		3.3		2.93		0.37		0.12		0.25	
	B		2.8		2.93		-0.13		0.12		-0.25	
	B		3.2		2.93		0.27		0.12		0.15	
	B		3.0		2.93		0.07		0.12		-0.05	
Degrees of freedom (DF)			10	−	1	=	9	=	1	+	8	
Sum of squares (SS)			86.29	−	85.85	=	0.441	=	0.216	+	0.225	
Mean square (MS)			*(SS / DF)*				0.049		0.216		0.028	
F ratio			*(Group MS / Error MS)*								7.680	
P value			*(Probability of an F ratio this large by chance alone)*								0.0242	

ANOVA (5 of 6, cont'd)

The *P-value* is a probability calculation based on the F ratio, the DF for the *Group* column, and the DF for the *Error* column.

Interpreting P-values

P-value	Evidence that populations are different or variables are correlated	Confidence level (CL)
1.00 – 0.15	None	None
0.15 – 0.05	Some	85% ≤ CL < 95%
0.05 – 0.01	Strong	95% ≤ CL < 99%
0.01 – 0.0001	Very strong	CL ≥ 99%

P-values (cont'd)

As shown above, the P-value has fixed reference values for interpretation.

The P value is inversely related to the F ratio:

- The smaller the P-value, the stronger the evidence of a difference between the population means.

If there are 3 or more groups, the interpretation is:

- The smaller the P-value, the stronger the evidence of one or more differences among the population means.

ANOVA (6 of 6)

A	B	C	D	E	F	G	H	I	J	K	L	M
	Group		**Data**		**Grand mean**		**Difference**		**Group**		**Error**	
	A		2.8		2.93		-0.13		-0.18		0.05	
	A		2.6		2.93		-0.33		-0.18		-0.15	
	A		2.9		2.93		-0.03		-0.18		0.15	
	A		2.7		2.93		-0.23		-0.18		-0.05	
	B		3.1	−	2.93	=	0.17	=	0.12	+	0.05	
	B		2.9		2.93		-0.03		0.12		-0.15	
	B		3.3		2.93		0.37		0.12		0.25	
	B		2.8		2.93		-0.13		0.12		-0.25	
	B		3.2		2.93		0.27		0.12		0.15	
	B		3.0		2.93		0.07		0.12		-0.05	
Degrees of freedom (DF)			10	−	1	=	9	=	1	+	8	
Sum of squares (SS)			86.29	−	85.85	=	0.441	=	0.216	+	0.225	
Mean square (MS)			*(SS / DF)*				0.049		0.216		0.028	
F ratio			*(Group MS / Error MS)*								7.680	
P value			*(Probability of an F ratio this large by chance alone)*								0.0242	
Root mean square (RMS)			*(Square root of MS)*				0.221				0.168	

ANOVA (6 of 6, cont'd)

The *Root Mean Square* (RMS) for a column is the square root of the MS for that column.

The RMS for the *Difference* column (0.221) is equal to the usual standard deviation of the data (STDEV function in Excel).

The RMS for the *Error* column (0.168) is the standard deviation of the noise variation (error, residual, unexplained, etc.).

JMP uses the term *Root Mean Square Error* (RMSE) for the RMS of the *Error* column.*

*Given that Statistics is a body of knowledge dedicated to quantifying and reducing variation, the variation in statistical terminology is appalling.

Degrees of freedom for comparing populations

N = total sample size

G = number of groups being compared

G – 1 = DF for the group column

N – G = DF for the error column

- The *Error* DF is more important than the *Group* DF
- It determines the accuracy of the predicted values
- Larger is better, 10 is OK, bare minimum is 5
- When DF is mentioned without a qualifier, it always means *Error* DF

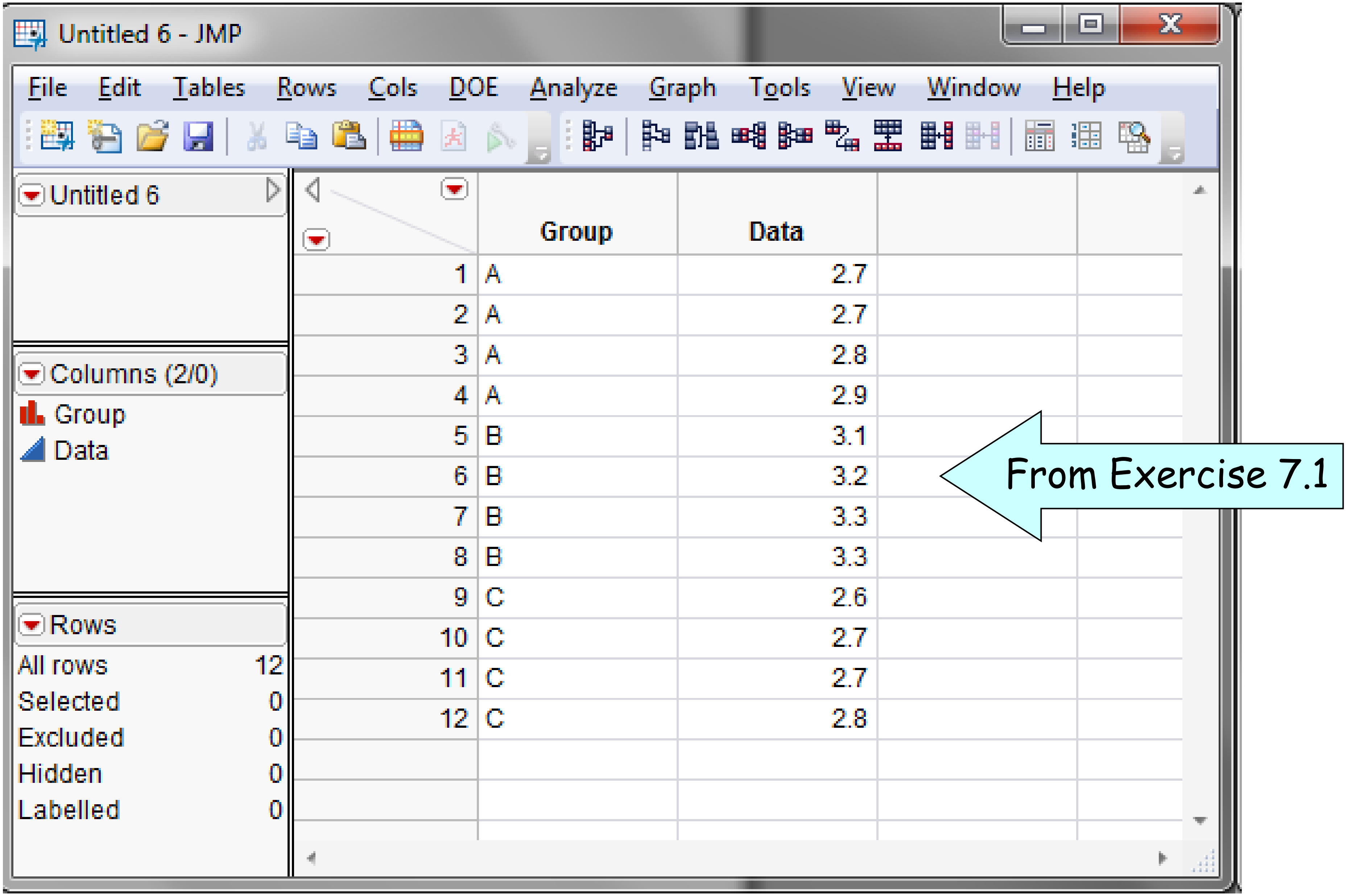
Exercise 7.1

LSSV2 student files \ ANOVA three groups. Enter the appropriate numbers and formulas into the white cells to produce an ANOVA for the data shown here.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1						Grand								
2		Group		Data		mean		Variance		Group		Error		
3		A		2.7										
4		A		2.7										
5		A		2.8										
6		A		2.9										
7		B		3.1										
8		B		3.2	–		=		=		+			
9		B		3.3										
10		B		3.3										
11		C		2.6										
12		C		2.7										
13		C		2.7										
14		C		2.8										
15	Degrees of freedom (DF)				–		=		=		+			
16	Sum of squares (SS)				–		=		=		+			
17	Mean square (MS)			*(SS / DF)*										
18	F ratio			(*Group MS / Error MS*)										
19	P value			*(Probability of getting an F ratio this large by chance alone)*										
20	Root mean square (RMS)			*(Square root of MS)*										

ANOVA in JMP

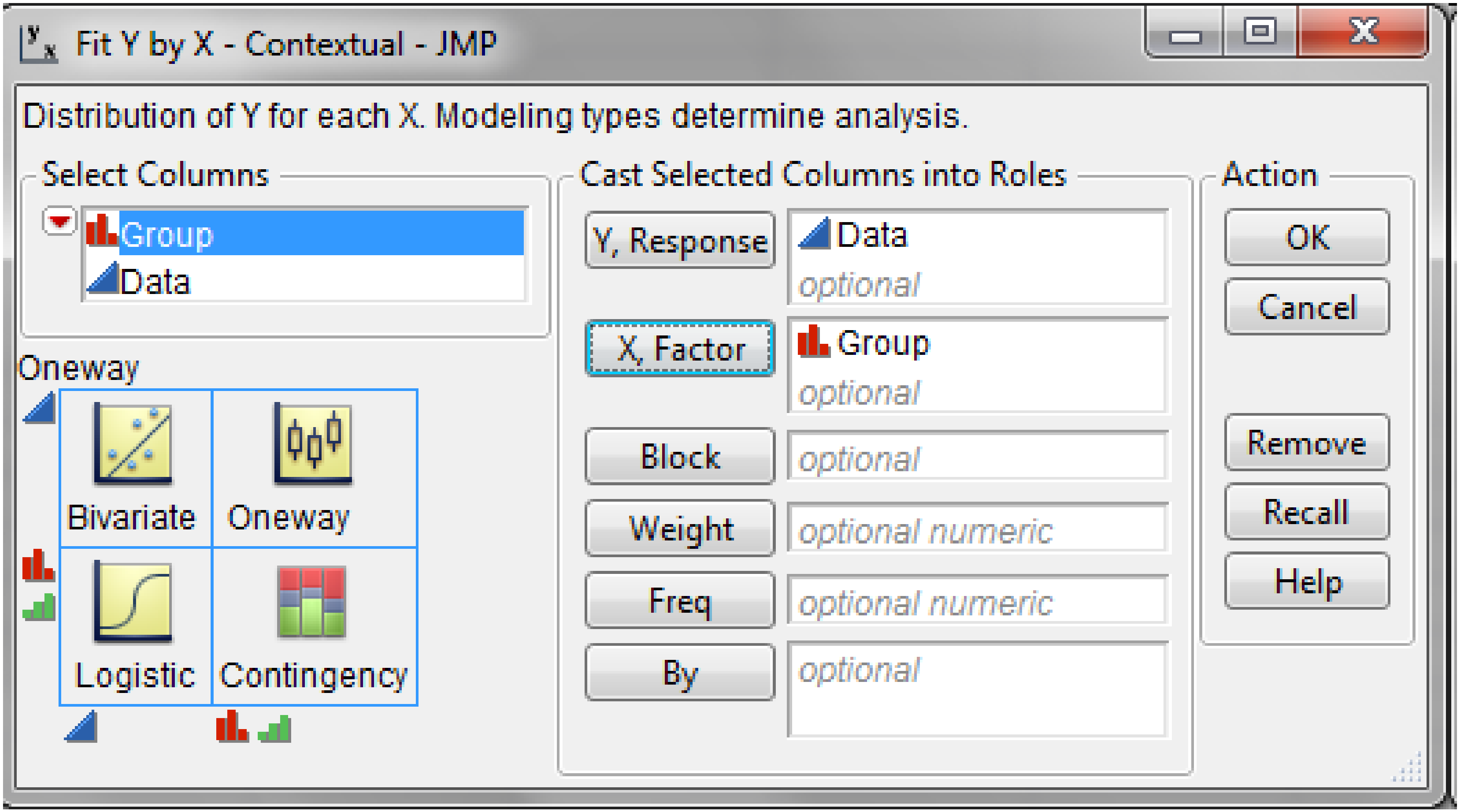
File → *New* → *Data Table* → Enter (or copy-paste) data as shown



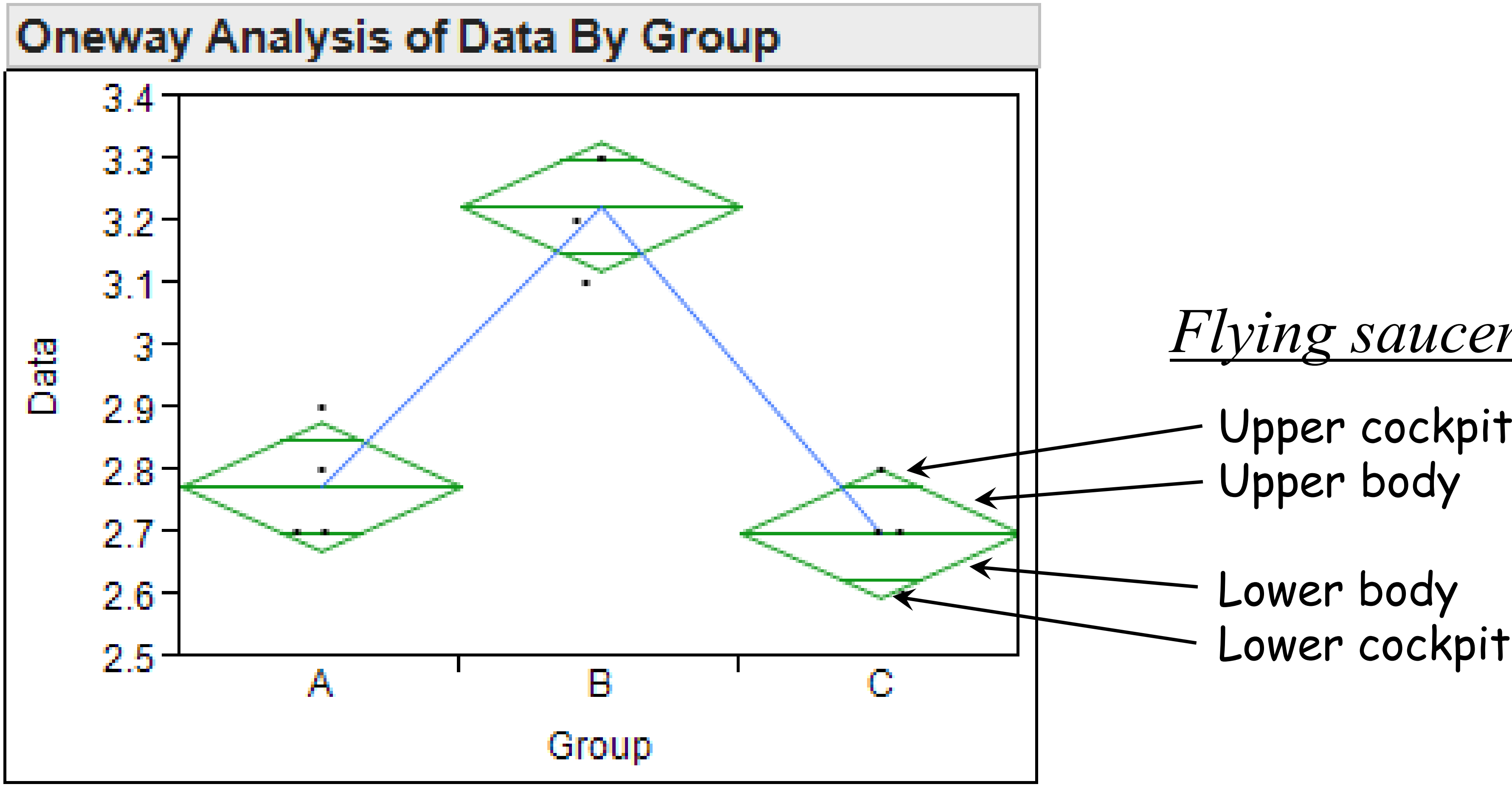
	Group	Data
1	A	2.7
2	A	2.7
3	A	2.8
4	A	2.9
5	B	3.1
6	B	3.2
7	B	3.3
8	B	3.3
9	C	2.6
10	C	2.7
11	C	2.7
12	C	2.8

ANOVA in JMP (cont'd)

Analyze → *Fit Y by X* → Set up as shown → OK



Explanation of "mean diamonds"

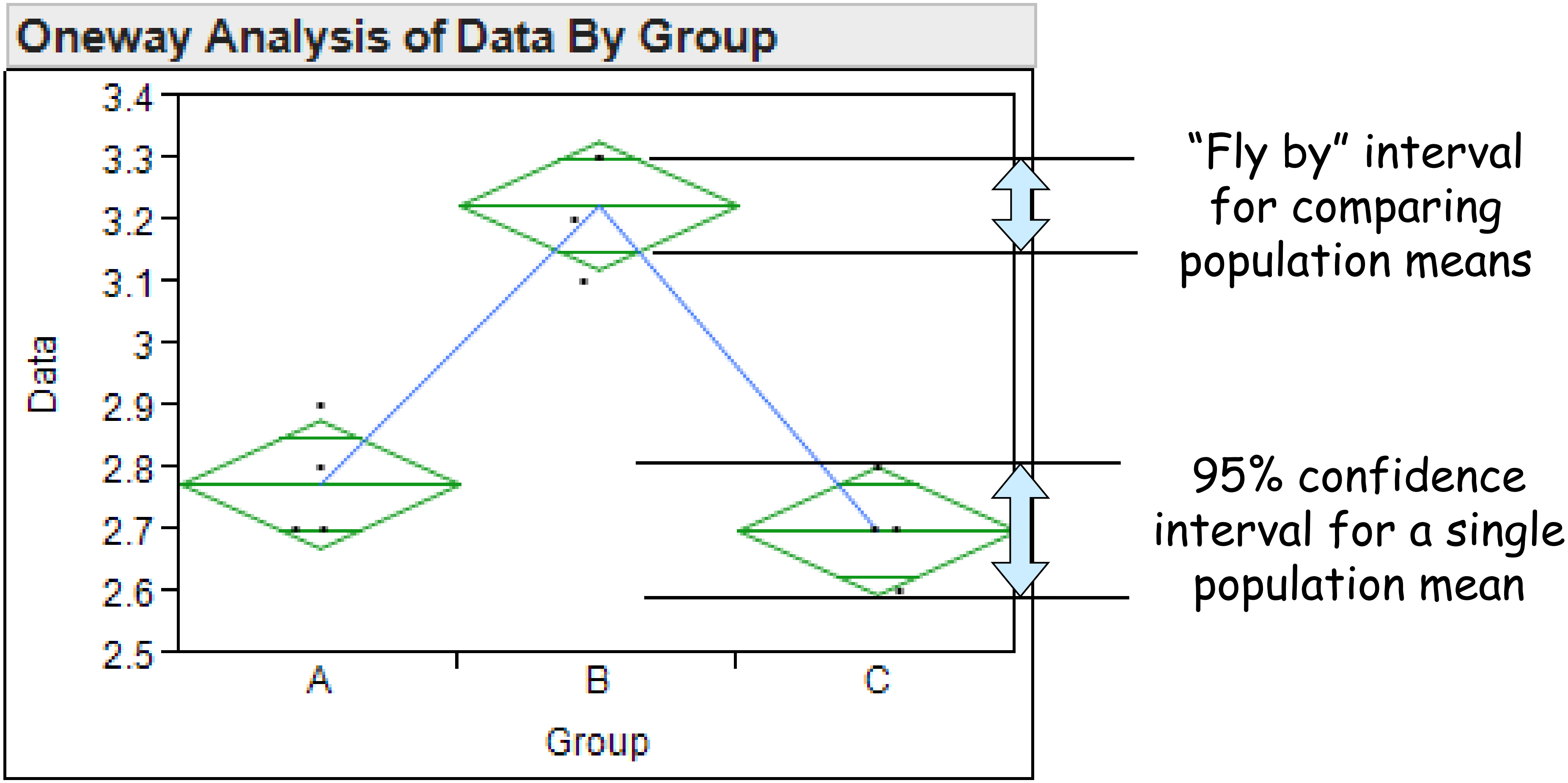


Population means are different (with 95% confidence)

⟷

Saucers can fly horizontally past each other with no contact between their bodies

Mean diamonds (cont'd)



Approx. formula for "fly by" interval: $\text{Sample mean} \pm \sqrt{2}\left(\text{RMSE}/\sqrt{\text{N}}\right)$

Approx. formula for 95% confidence interval: $\text{Sample mean} \pm 2\left(\text{RMSE}/\sqrt{\text{N}}\right)$

N = sample size for each group

Analysis details

Oneway Anova

Summary of Fit

Rsquare	0.895833
Adj Rsquare	0.872685
Root Mean Square Error	0.091287
Mean of Response	2.9
Observations (or Sum Wgts)	12

RMSE

- Standard deviation of the variation about the fitted line (error, residual, etc.)
- Smaller is better
- Has units of the Y variable

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Group	2	0.64500000	0.322500	38.7000	<.0001*
Error	9	0.07500000	0.008333		
C. Total	11	0.72000000			

Group → Regression

Prob > F → P-value

- Indicates whether any of the model terms in the regression are significant

Analysis details (cont'd)

Oneway Anova

Summary of Fit

Rsquare	0.895833
Adj Rsquare	0.872685
Root Mean Square Error	0.091287
Mean of Response	2.9
Observations (or Sum Wgts)	12

Adjusted R^2

Analysis of Variance

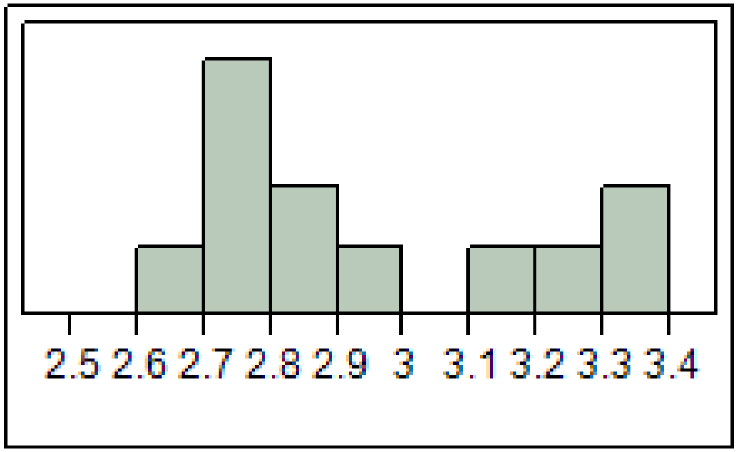
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Group	2	0.64500000	0.322500	38.7000	<.0001*
Error	9	0.07500000	0.008333		
C. Total	11	0.72000000			

- Proportion of the total variation in Y that is caused by ("explained by") variation in X
- Larger is better
- Unitless

How adjusted R^2 is calculated

Distributions

Data



Summary Statistics

Mean	2.9
Std Dev	0.2558409
N	12

STDEV → Total variation in the data

$$\text{Proportion of Y variation NOT caused by X} = \left(\frac{\text{RMSE}}{\text{STDEV}}\right)^2 = \left(\frac{0.091287}{0.2558409}\right)^2 = 0.127315$$

$$\text{Proportion of Y variation CAUSED by X} = 1 - \left(\frac{\text{RMSE}}{\text{STDEV}}\right)^2 = 0.872685 = \text{Adjusted R}^2$$

Exercise 7.2

Data sets \ number and size of defects. *Max size* is the area in square centimeters of the largest contiguous weld repair area on each casting. Smaller *Max size* is better.

a) Test for a difference between welders A and B with respect to *Max size*. Give the P value and interpret the result. (Ignore the *t Test* section of the output.)

b) Which welder represents best practice? What follow-up action should be taken?

c) Give the value and the units of the RMSE in this example.

d) The RMSE is meaningful only if each group has roughly the same amount of variation. Is this true in this case?

e) Save your analysis script to the data table, close and save the data table.

Exercise 7.3

Data sets \ quotation process. Supplier business units (BUs) receive requests for quote (RFQs) from customers. Account managers develop and submit the quotes. TAT is the turnaround around time in days. The shorter the TAT, the happier the customer.

a) Is the modeling type for BU correct? If not, change it to what it should be.

b) Test for differences among the BUs. Give the P value and interpret the result.

c) Use the “flying saucers” to determine which BUs represent best practice.

d) What follow-up action should be taken?

e) Save your analysis script to the data table, close and save the data table.

Exercise 7.4

Data sets \ alignment process. If the modeling type for *Aligner* is incorrect, change it to what it should be.

a) Test for differences among the three aligners with respect to *R dev*. Give the P-value and interpret the results.

b) Use the “flying saucers” to determine which aligner represents best practice. (Smaller *R dev* is better.)

c) What follow-up action should be taken?

d) Save your analysis script to the data table, close and save the data table.

Exercise 7.5

Data sets \ casting dimensions. We want to reduce variation in the length of cylindrical metal castings. The specification for *Length* is 600 ± 1.5. The wax patterns for these castings are molded on two machines A and B.

a) Test for differences between the molding machines with respect to *Length*. Give the P-value and interpret the result.

b) Use the "flying saucers" to determine which machine represents best practice? (It is helpful to draw a reference line at the nominal value. Right click on one of the numbers on the vertical axis, select *Axis Settings*, use the *Reference Lines* tool.)

c) What follow-up action should be taken?

d) Save your analysis script to the data table, but don't close the data table.

Exercise 7.5 (cont'd)

We also want to reduce variation in the diameter of the castings. The specification for *Diam* is 50 ± 0.75.

d) Test for differences between the molding machines with respect to *Diam*. Give the P-value and interpret the result.

e) Use the "flying saucers" to determine which machine represents best practice. (Draw a reference line at the nominal value.)

f) What follow-up action should be taken?

g) For each of the variables *Length* and *Diam*, a certain proportion of the total variation is caused by the difference between the machines. For which variable is this proportion highest?

h) Save your analysis script to the data table, close and save the data table.

8 Comparing Populations – Pass/fail Y

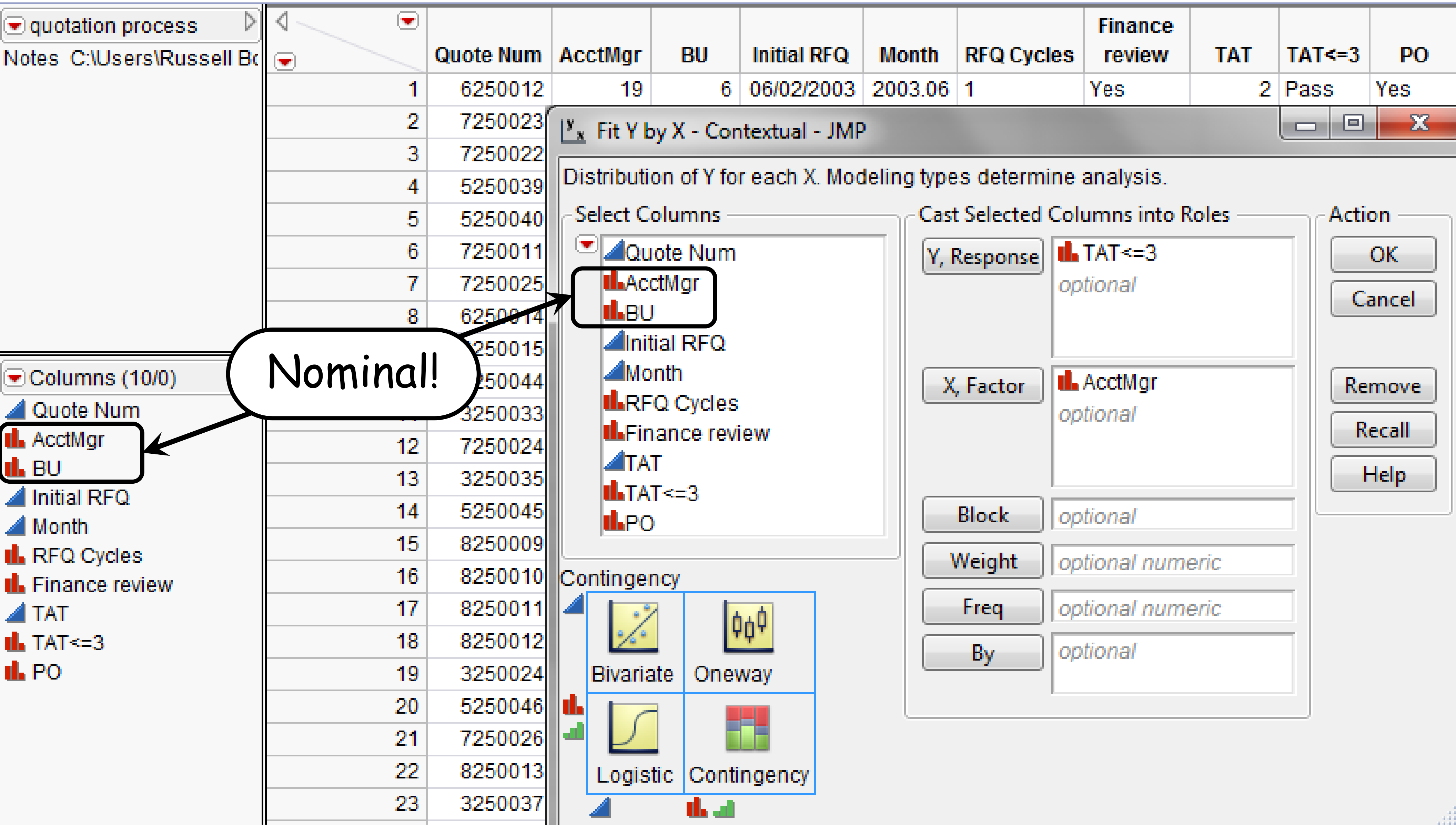
Raw data	One part or transaction per row
Tabulated data	Multiple parts or transactions per row

Raw data example

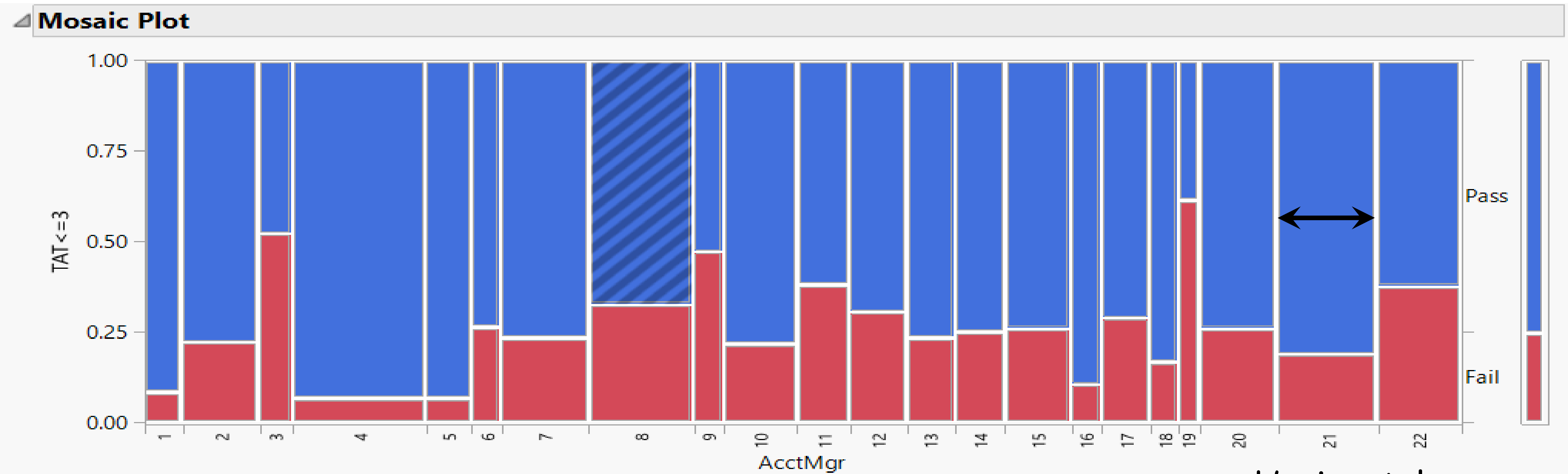
Data sets \ quotation process

We want to compare the account managers in terms of % late

Analyze → *Fit Y by X* → set up as shown → OK



“Mosaic plot” for pass/fail data



Horizontal dimension is proportional to sample size

Tests

N	DF	-LogLike	RSquare (U)
837	21	32.411285	0.0687

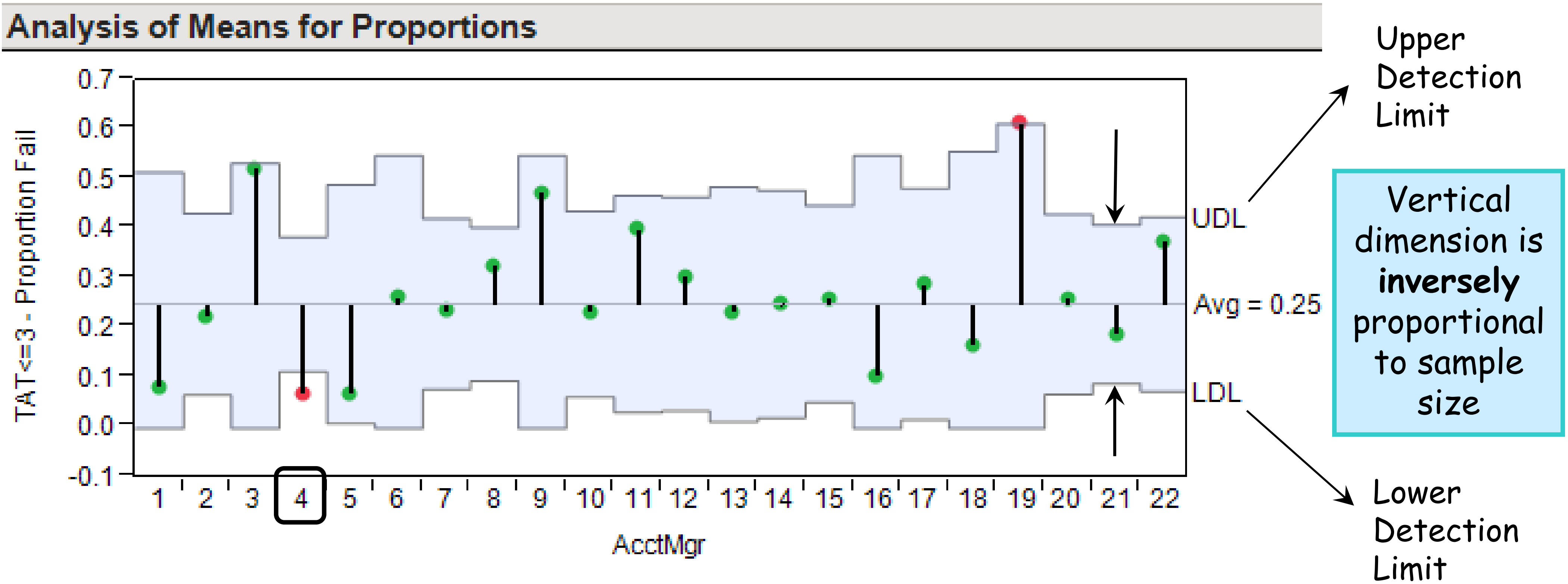
Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	64.823	<.0001*
Pearson	62.018	<.0001*

P-value

- Very strong evidence of differences among account managers
- Who represents best practice?

"Control chart" for pass/fail data

- Red triangle (Contingency Analysis) → *Analysis of Means for Proportions*



- "Flying saucers" are not available for pass/fail data
- Points outside the shaded region are significantly different from points inside
- *AcctMgr 4* represents best practice (lowest failure rate)
- Find out what *AcctMgr 4* is doing, make it the standard
- Save your analysis script to the data table, but don't close the data table

Exercise 8.1

a) Analyze *TAT<=3* as a function of *BU*. Give the P-value and interpret the result. Is there best practice? If so, where is it?

b) Analyze *PO* as a function of *BU*. Give the P-value and interpret the result. Is there best practice? If so, where is it?

c) Right click on the *PO* header in the data table. Select *Column Properties* → *Value Ordering* → *Reverse* → *OK*. This reverses the *Yes* and *No* positions on the *PO* axis. Most people focus on the *PO* hit rate rather than the miss rate.

d) Analyze *PO* hit rate as a function of *TAT<=3* . Give the P-value and interpret the result.

e) Save your scripts, close and save the data table.

Exercise 8.2

Data sets \ ATE data. If necessary, change the modeling types for part number (*P/N)* and *Tester*.

a) Test for a difference between the part numbers (*P/N*) with respect to *Result*. Give the P-value and interpret the results.

b) Test for differences among the testers with respect to *Result*. Give the P-value and interpret the results. If significant differences exist, describe them. If possible, suggest causes of the differences.

c) Test for differences among the *P/N-Tester* groupings with respect to *Result*. Give the P-value and interpret the results. If significant differences exist, describe them. If possible, suggest causes of the differences.

d) Save your scripts, close and save the data table.

Tabulated pass/fail data

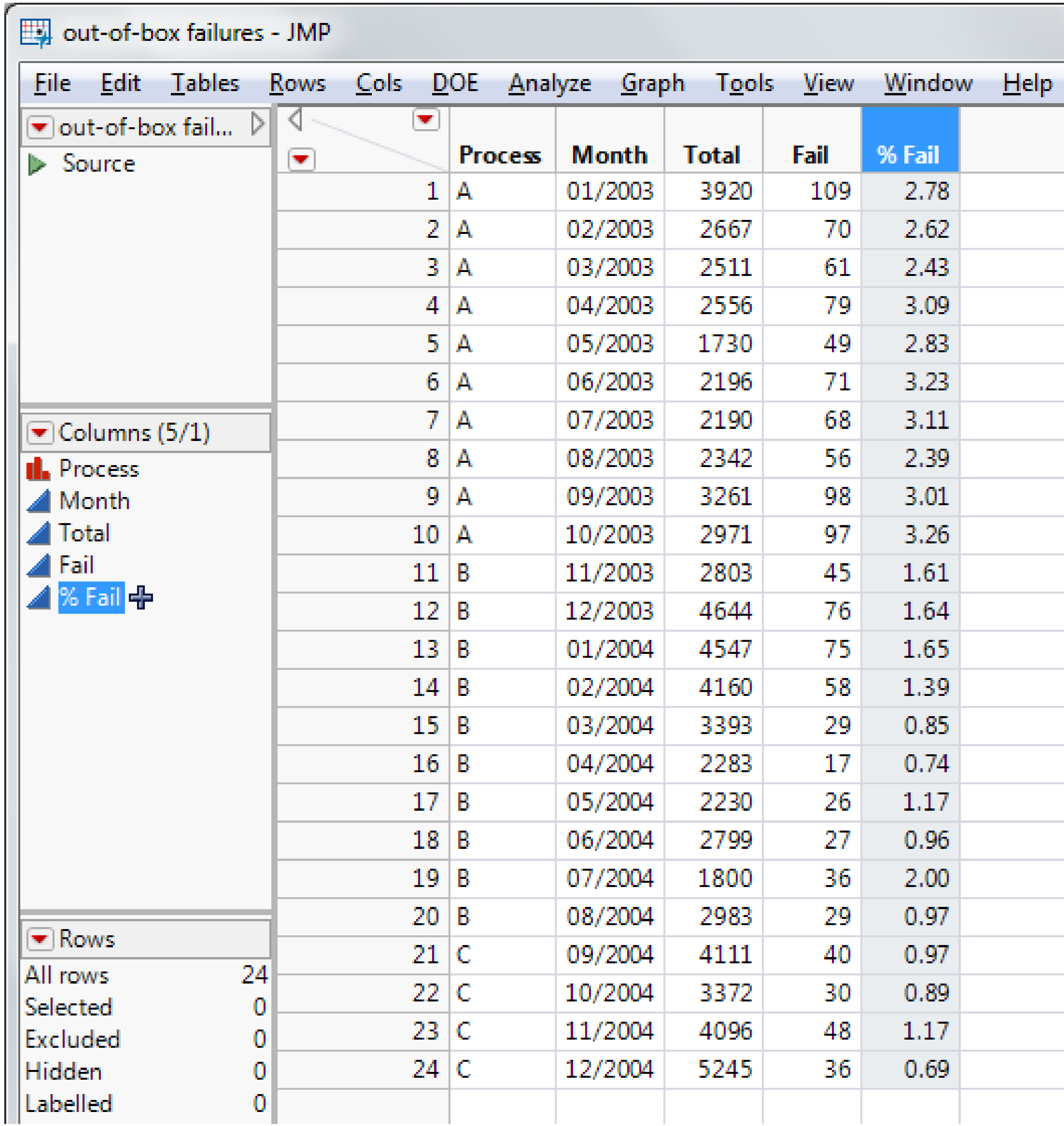
- Pass/fail data often comes in tabulated form
- Each row may represent a
  - ✓ Production lot
  - ✓ Work order
  - ✓ Time period
  - ✓ Machine
  - ✓ Work center
  - ✓ Part number . . .
- This format is perfect for plotting % defective
- However, it is the wrong format for comparing populations in JMP

Plotting % fail

1. Create a new column called *% Fail*
2. Define it by the formula

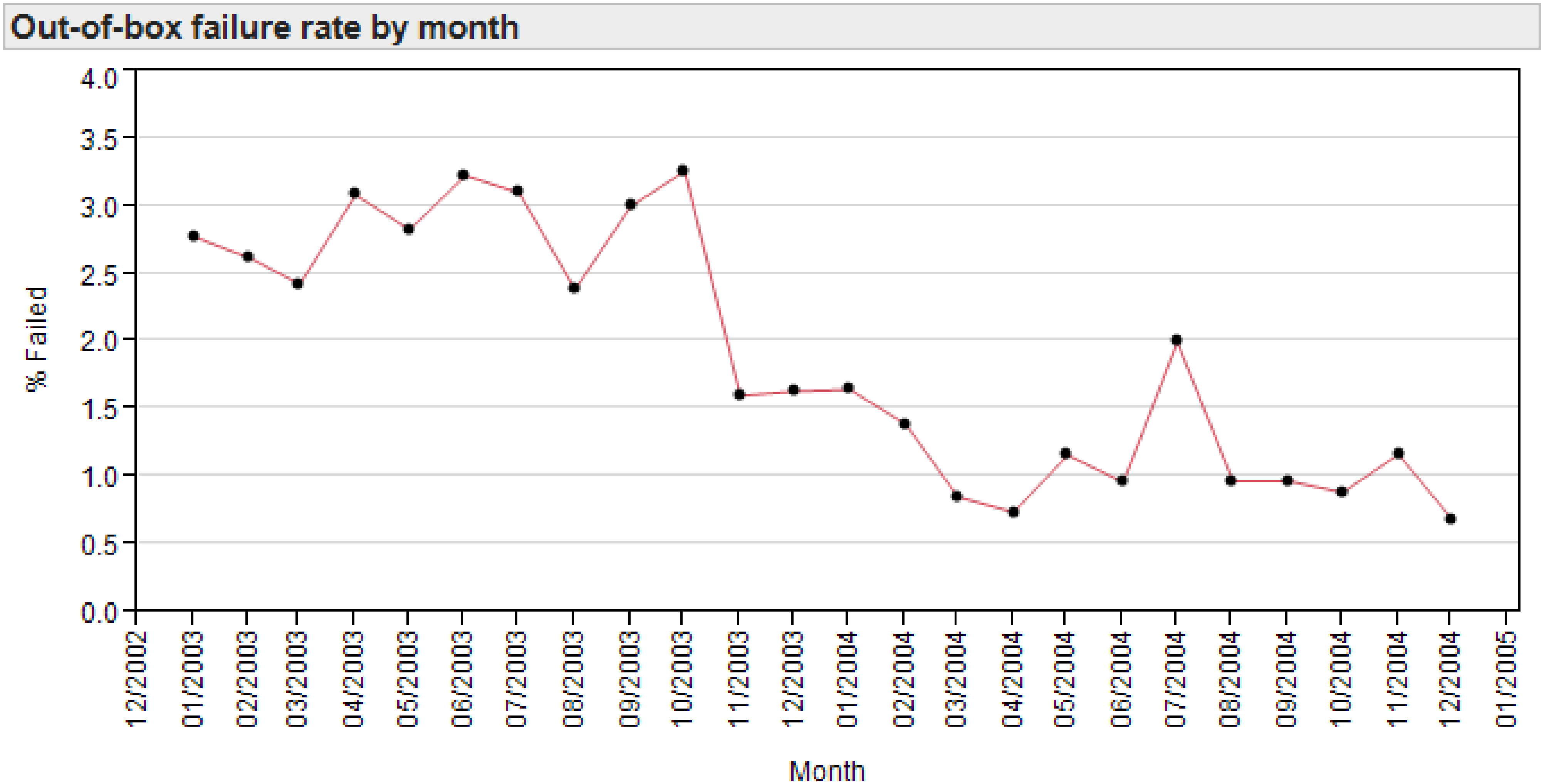
$$\left(\frac{\text{Fail}}{\text{Total}}\right)\cdot 100$$

3. To edit decimal places: Right click column → Column Info → Format to Fixed Decimal and Dec = 2
4. Use *Graph* → *Legacy* → *Overlay Plot* to create the plot on the next slide



	Process	Month	Total	Fail	% Fail
1	A	01/2003	3920	109	2.78
2	A	02/2003	2667	70	2.62
3	A	03/2003	2511	61	2.43
4	A	04/2003	2556	79	3.09
5	A	05/2003	1730	49	2.83
6	A	06/2003	2196	71	3.23
7	A	07/2003	2190	68	3.11
8	A	08/2003	2342	56	2.39
9	A	09/2003	3261	98	3.01
10	A	10/2003	2971	97	3.26
11	B	11/2003	2803	45	1.61
12	B	12/2003	4644	76	1.64
13	B	01/2004	4547	75	1.65
14	B	02/2004	4160	58	1.39
15	B	03/2004	3393	29	0.85
16	B	04/2004	2283	17	0.74
17	B	05/2004	2230	26	1.17
18	B	06/2004	2799	27	0.96
19	B	07/2004	1800	36	2.00
20	B	08/2004	2983	29	0.97
21	C	09/2004	4111	40	0.97
22	C	10/2004	3372	30	0.89
23	C	11/2004	4096	48	1.17
24	C	12/2004	5245	36	0.69

Plotting % fail (cont'd)



Reformatting for comparing populations

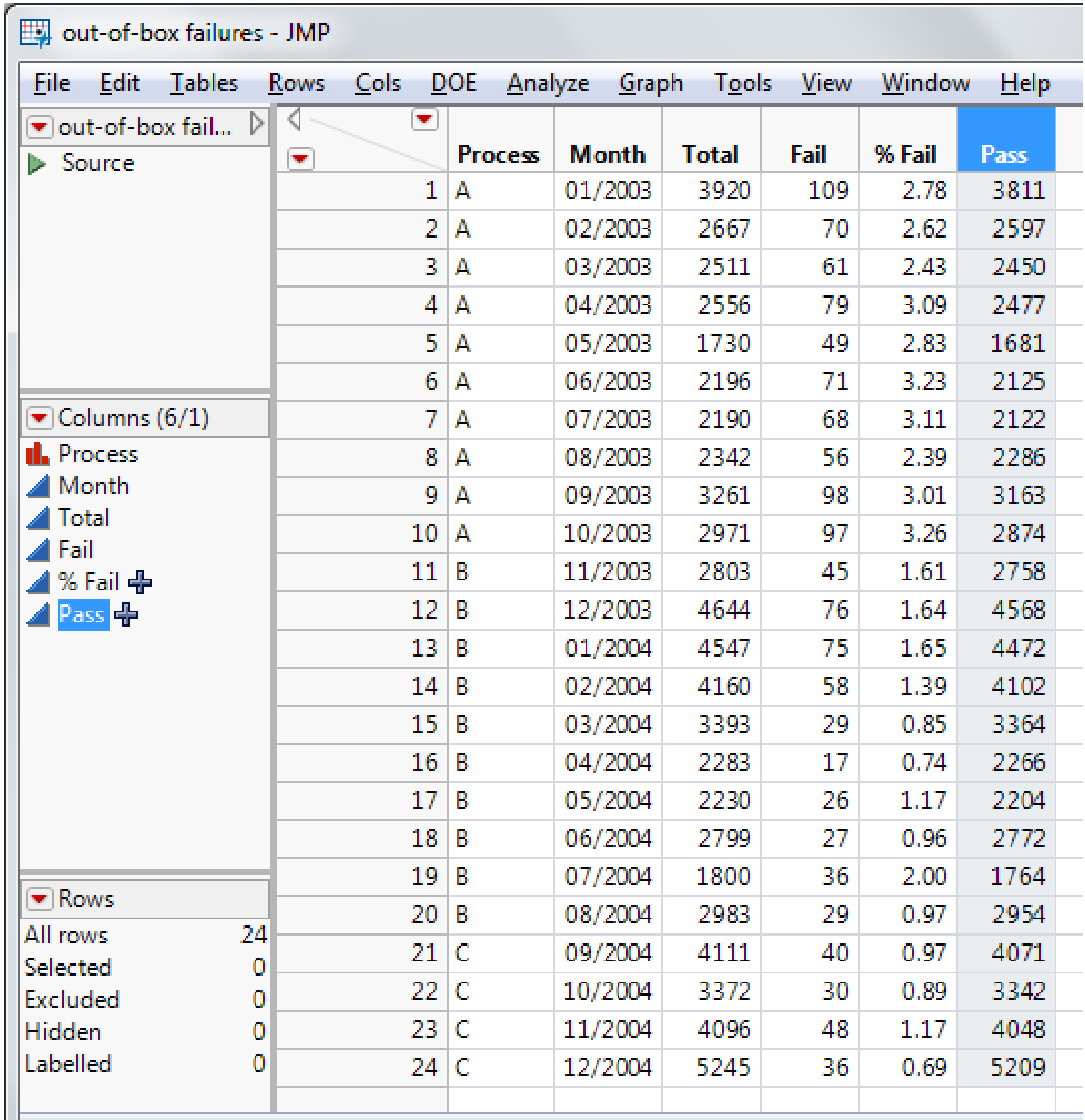
1. Create a new column called *Pass* defined by the formula

   Total – Fail

2. Go to *Tables* → *Stack*

3. Use *Fail* and *Pass* as the *Stack Columns*

4. See next slide



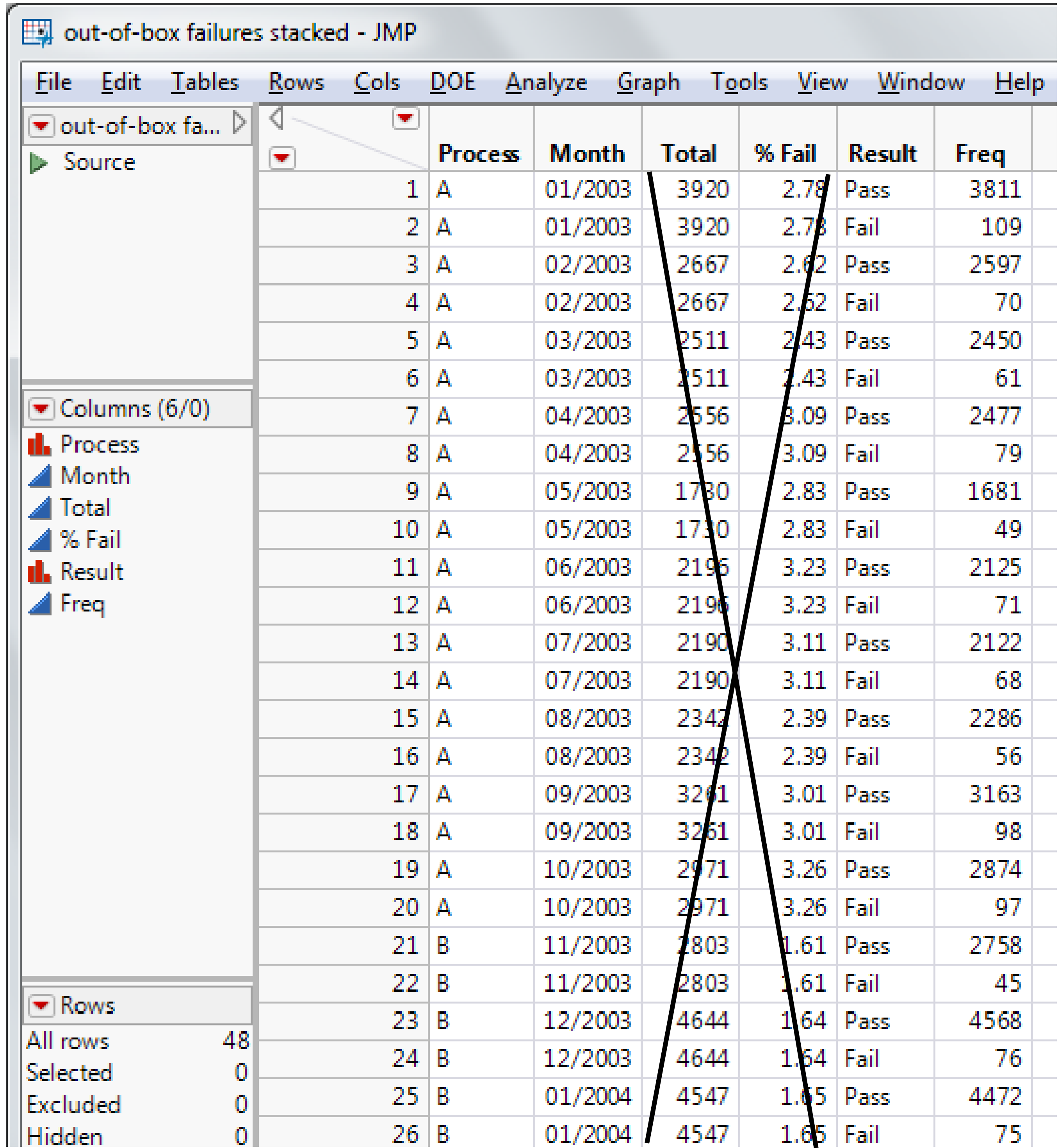
	Process	Month	Total	Fail	% Fail	Pass
1	A	01/2003	3920	109	2.78	3811
2	A	02/2003	2667	70	2.62	2597
3	A	03/2003	2511	61	2.43	2450
4	A	04/2003	2556	79	3.09	2477
5	A	05/2003	1730	49	2.83	1681
6	A	06/2003	2196	71	3.23	2125
7	A	07/2003	2190	68	3.11	2122
8	A	08/2003	2342	56	2.39	2286
9	A	09/2003	3261	98	3.01	3163
10	A	10/2003	2971	97	3.26	2874
11	B	11/2003	2803	45	1.61	2758
12	B	12/2003	4644	76	1.64	4568
13	B	01/2004	4547	75	1.65	4472
14	B	02/2004	4160	58	1.39	4102
15	B	03/2004	3393	29	0.85	3364
16	B	04/2004	2283	17	0.74	2266
17	B	05/2004	2230	26	1.17	2204
18	B	06/2004	2799	27	0.96	2772
19	B	07/2004	1800	36	2.00	1764
20	B	08/2004	2983	29	0.97	2954
21	C	09/2004	4111	40	0.97	4071
22	C	10/2004	3372	30	0.89	3342
23	C	11/2004	4096	48	1.17	4048
24	C	12/2004	5245	36	0.69	5209

Reformatting (cont'd)

6. Change the name of the *Data* column to *Freq* and the *Label* column to *Result*

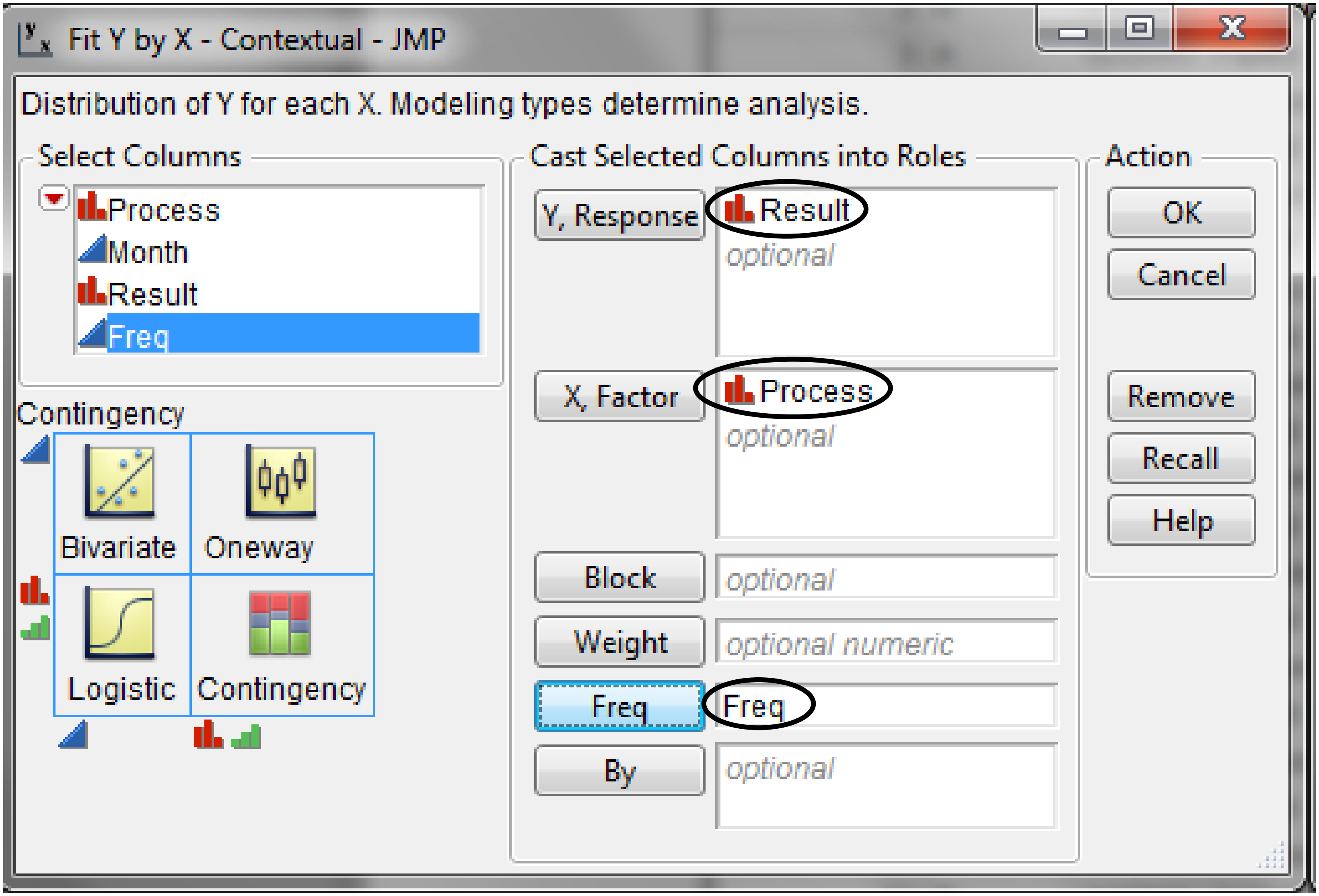
7. There are now two rows for each month. The *Total* and *% Fail* columns are no longer relevant, and may be deleted.

8. Save the new data table as *out-of-box failures stacked*



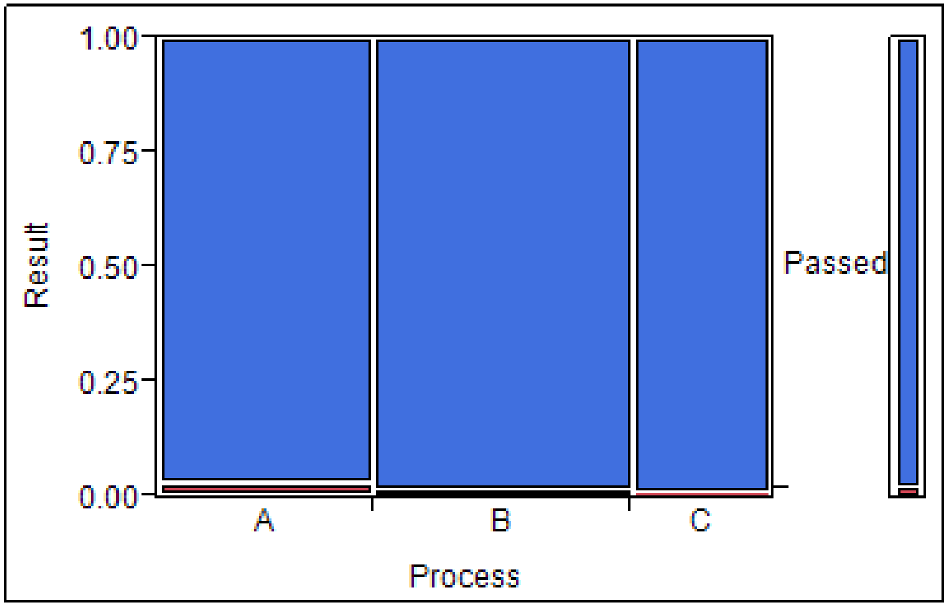
	Process	Month	Total	% Fail	Result	Freq
1	A	01/2003	3920	2.78	Pass	3811
2	A	01/2003	3920	2.78	Fail	109
3	A	02/2003	2667	2.62	Pass	2597
4	A	02/2003	2667	2.62	Fail	70
5	A	03/2003	2511	2.43	Pass	2450
6	A	03/2003	2511	2.43	Fail	61
7	A	04/2003	2556	3.09	Pass	2477
8	A	04/2003	2556	3.09	Fail	79
9	A	05/2003	1730	2.83	Pass	1681
10	A	05/2003	1730	2.83	Fail	49
11	A	06/2003	2196	3.23	Pass	2125
12	A	06/2003	2196	3.23	Fail	71
13	A	07/2003	2190	3.11	Pass	2122
14	A	07/2003	2190	3.11	Fail	68
15	A	08/2003	2342	2.39	Pass	2286
16	A	08/2003	2342	2.39	Fail	56
17	A	09/2003	3261	3.01	Pass	3163
18	A	09/2003	3261	3.01	Fail	98
19	A	10/2003	2971	3.26	Pass	2874
20	A	10/2003	2971	3.26	Fail	97
21	B	11/2003	2803	1.61	Pass	2758
22	B	11/2003	2803	1.61	Fail	45
23	B	12/2003	4644	1.64	Pass	4568
24	B	12/2003	4644	1.64	Fail	76
25	B	01/2004	4547	1.65	Pass	4472
26	B	01/2004	4547	1.65	Fail	75

Analyze → *Fit Y by X* → set up as shown → OK



Data analysis (cont'd)

Mosaic Plot



Contingency Table

Result

Count / Row %	Failed	Passed	
A	758 2.88	25586 97.12	26344
B	418 1.32	31224 98.68	31642
C	154 0.92	16670 99.08	16824
	1330	73480	74810

Process

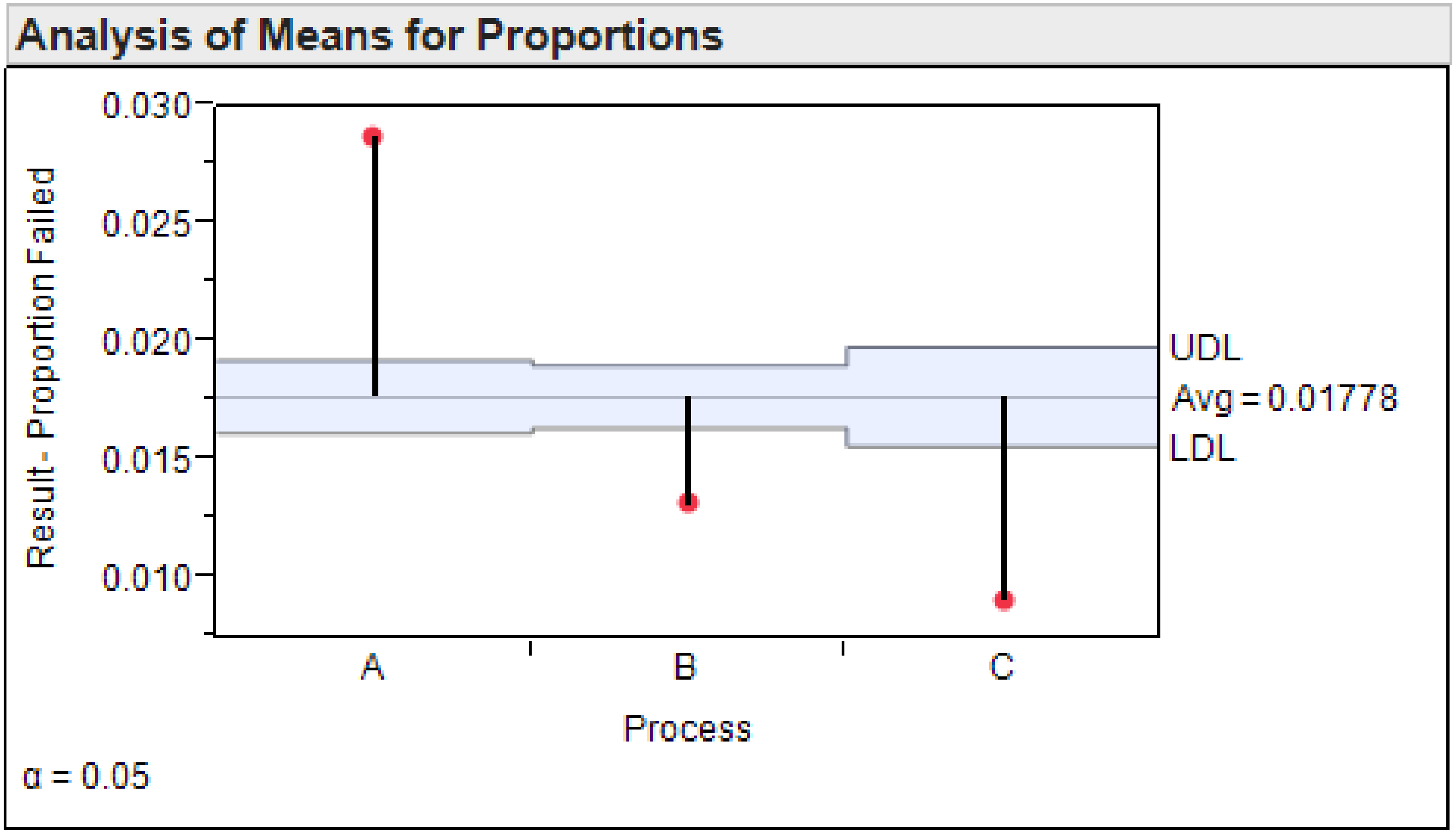
Tests

N	DF	-LogLike	RSquare (U)
74810	2	141.17363	0.0211

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	282.347	<.0001*
Pearson	291.850	<.0001*

- Very strong evidence that processes A, B, and C do not all have the same failure rate
- The mosaic plot does not help us determine where the differences are
- Click on the red triangle at the top of the analysis window
- Select *Analysis of Means for Proportions*
- See next slide

Data analysis (cont'd)



- This plot shows that Processes B and C are significant improvements over Process A
- It does not tell us whether or not C is a significant improvement over B
- Save your script, but don't close the data table.
- You may prefer to display the Result as Proportion Passed: Click on Red Triangle by Analysis of Means for Proportions and select Switch Response Level for Proportion

Exercise 8.3

a) Exclude the rows for process A.

b) Test for a difference between C and B. Give the P-value and interpret the result.

c) Close and save the data table. (No need to save the script again.)

Exercise 8.4

Data sets \ molding process - stratification.

a) Did JMP assign the correct modeling type for *Machine*?

b) Go to *Tables* → *Summary* → use *PN* as the *Group* variable → use *Machine* as the *Subgroup* variable → OK.

10/0	PN	N Rows	N(01)	N(02)	N(03)	N(09)	N(10)	N(11)	N(13)	N(14)	N(15)
1	GV0098	43	0	0	0	0	0	0	0	11	32
2	GV0101	31	0	0	0	30	0	0	0	0	1
3	GV0119	42	3	0	39	0	0	0	0	0	0
4	GV0129	89	0	0	0	0	0	0	0	88	1
5	GV0132	64	0	64	0	0	0	0	0	0	0
6	GY0251	37	0	0	0	0	0	17	20	0	0
7	GY0298	31	0	0	0	24	7	0	0	0	0
8	GY0306	53	0	0	0	0	0	27	26	0	0
9	GY0325	36	1	0	0	0	0	34	1	0	0
10	KU0041	84	83	1	0	0	0	0	0	0	0

c) Note that each part number runs on only one or two of the machines. A comparison of part numbers could be biased by differences among the machines, and a comparison of machines could be biased by differences among the part numbers. Because of this, we should use the concatenated variable *PN-Machine* as the X variable in the analysis.

Exercise 8.4 (cont'd)

d) Reformat the data for comparing populations (follow steps 1 through 7 in the worked example).

e) Test for significant differences among the *PN-Machine* groupings with respect to fraction defective. Give the P-value and interpret the results.

f) Which three *PN-Machine* groupings would be the best focus for an improvement project? (Hint: highest fractions defective.)

g) Save your script, save the data table as *molding process - stacked*, then close it.

This page intentionally left blank

Appendix: Reformatting Data for Pareto Analysis

- Data on defect types or failure reasons often is available only in tabulated form

- Each row may represent a production lot, work order, time period, machine work center, part number, . . . , or some combination thereof

- Common problem with tabulated data: wrong format for Pareto analysis

Big example: *molding process - Pareto*

Each row = Date, Machine, P/N, . . .

Total parts run = Good + Bad

	A	B	C	D	E	F	G	H	I	J
1	Date	Machine	P/N	Primary material	Primary lot #	Concentrate	Concen lot #	Regrind type	Parts palletized	Total defective
2	03-Apr-06	9	LSGV0101	CHEIL VE-1877S DrkGry	121642	NA	NA	25	120	7
3	03-Apr-06	11	LSGY0251	CHEIL VE-1877S Blk	101200	NA	NA	NA	300	17
4	03-Apr-06	15	LSGV0098	CHEIL HF1690H DrkGry	122930	NA	NA	8	372	18
5	04-Apr-06	2	LSGV0093	CHEIL VE-1877S DrkGry	121642	NA	NA	25	288	6
6	04-Apr-06	9	LSGV0101	CHEIL VE-1877S DrkGry	121642	NA	NA	25	600	2
7	04-Apr-06	11	LSGY0251	CHEIL VE-1877S Blk	101200	NA	NA	NA	690	33
8	04-Apr-06	13	LSGY0307	CHEIL HF1690H LtGry	133232	NA	NA	NA	160	8
9	04-Apr-06	15	LSGV0098	CHEIL HF1690H DrkGry	122930	NA	NA	8	624	0
10	05-Apr-06	2	LSGV0102	CHEIL VE-1877S DrkGry	121642	NA	NA	25	120	15
11	05-Apr-06	11	LSGY0251	CHEIL VE-1877S Blk	101200	NA	NA	NA	650	21
12	05-Apr-06	13	LSGY0252	CHEIL VE-1877S Blk	101200	NA	NA	NA	300	18
13	05-Apr-06	13	LSGY0307	CHEIL VE-1877S LtGry	133232	NA	NA	NA	160	0
14	05-Apr-06	14	LSGY0308	CHEIL HF1690H LtGry	133232	NA	NA	NA	240	25
15	05-Apr-06	15	LSGV0098	CHEIL HF1690H DrkGry	122930	NA	NA	8	336	17
16	06-Apr-06	2	LSGV0102	CHEIL VE-1877S DrkGry	121642	NA	NA	25	780	0
17	06-Apr-06	11	LSGY0251	CHEIL VE-1877S Blk	101200	NA	NA	NA	600	7
18	06-Apr-06	13	LSGY0252	CHEIL VE-1877S Blk	101200	NA	NA	NA	500	49
19	06-Apr-06	14	LSGV0130	CHEIL HF1690H DrkGry	122930	NA	NA	8	108	34
20	06-Apr-06	15	LSGV0099	CHEIL HF1690H DrkGry	122930	NA	NA	8	276	95
21	07-Apr-06	2	LSGV0102	CHEIL VE-1877S DrkGry	121642	NA	NA	25	300	0
22	07-Apr-06	2	LSGV0102	CHEIL VE-1877S DrkGry	121642	NA	NA	25	1020	5
23	07-Apr-06	11	LSGY0251	CHEIL VE-1877S Blk	101200	NA	NA	NA	360	6
24	07-Apr-06	13	LSGY0252	CHEIL VE-1877S Blk	387487	NA	NA	NA	200	16
25	07-Apr-06	13	LSGY0252	CHEIL VE-1877S Blk	387487	NA	NA	NA	700	7
26	07-Apr-06	14	LSGV0130	CHEIL HF1690H DrkGry	122930	NA	NA	8	72	0
27	07-Apr-06	14	LSGV0131	CHEIL HF1690H DrkGry	122930	NA	NA	8	120	17
28	07-Apr-06	15	LSGV0099	CHEIL HF1690H DrkGry	122930	NA	NA	8	180	0

Big example (cont'd)

Total defective × Cost per pc.

Counts for each type of defect

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	Cost per pc.	Total cost	Start-up	Sink	Flash	Weld line	Flow mark	Short shot	Warp	Burn marks	Silver	Gas marks	Color/ carbon	Oil	Broken part	Scratches	Bubbles
2	$2.89	$20.25	3	0	0	0	0	0	0	0	4	0	0	0	0	0	0
3	$5.08	$86.43	4	0	0	0	0	4	0	0	0	0	0	0	0	0	9
4	$11.10	$199.76	0	0	0	0	0	6	0	0	12	0	0	0	0	0	0
5	$2.69	$16.12	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	$2.89	$5.79	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
7	$5.08	$167.77	0	4	0	0	0	2	0	0	0	0	0	0	0	2	0
8	$3.55	$28.44	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	$11.10	$0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	$4.13	$62.00	6	6	0	0	0	3	0	0	0	0	0	0	0	0	0
11	$5.08	$106.76	0	17	0	0	0	3	0	0	0	0	0	0	0	0	1
12	$4.96	$89.28	8	0	0	0	0	0	0	0	0	0	0	0	0	1	9
13	$3.55	$0.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0
14	$8.97	$224.36	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0
15	$11.10	$188.66	0	0	0	0	0	12	0	0	5	0	0	0	0	0	0
16	$4.13	$0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	$5.08	$35.59	0	2	0	0	0	4	0	0	0	0	0	0	0	1	0
18	$4.96	$243.04	3	15	0	0	0	0	0	0	0	0	0	0	0	4	27
19	$10.33	$351.07	8	0	0	0	0	14	0	0	12	0	0	0	0	0	0
20	$14.19	$1,347.62	56	30	0	0	0	0	0	0	9	0	0	0	0	0	0
21	$4.13	$0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	$4.13	$20.67	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	$5.08	$30.50	4	0	0	0	0	0	0	0	0	0	0	0	0	0	2
24	$4.96	$79.36	0	14	0	0	0	0	0	0	0	0	0	0	0	1	1
25	$4.96	$34.72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
26	$10.33	$0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	$15.15	$257.56	8	0	0	0	0	0	0	0	1	0	0	8	0	0	0
28	$14.19	$0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

One of the things we would want from a data set like this is a Pareto breakdown of defect types by frequency of occurrence. For this, we need to calculate the total number of defective parts for each defect type. With the format shown above, we cannot do this by means of a pivot table. As an alternative, we could calculate the totals for the columns representing the defect types. However, compared to a pivot table, this method is extremely tedious for doing anything else, such as comparing Pareto breakdowns for stratifications of the data set.

Another thing we would want from a data set like this is a Pareto breakdown of defect types by total cost. It is not impossible to do this with the format shown above, but, once again, it would be extremely tedious compared to a pivot table.

Small example

Open *molding process - small* (in JMP)

7/0 Cols / 3/0 Rows	Total defective	Cost per pc.	Total cost	Start-up	Short shot	Silver	Bubbles
1	7	3	21	3	0	4	0
2	17	5	85	4	4	0	9
3	18	11	198	0	6	12	0

↑

This is what we have

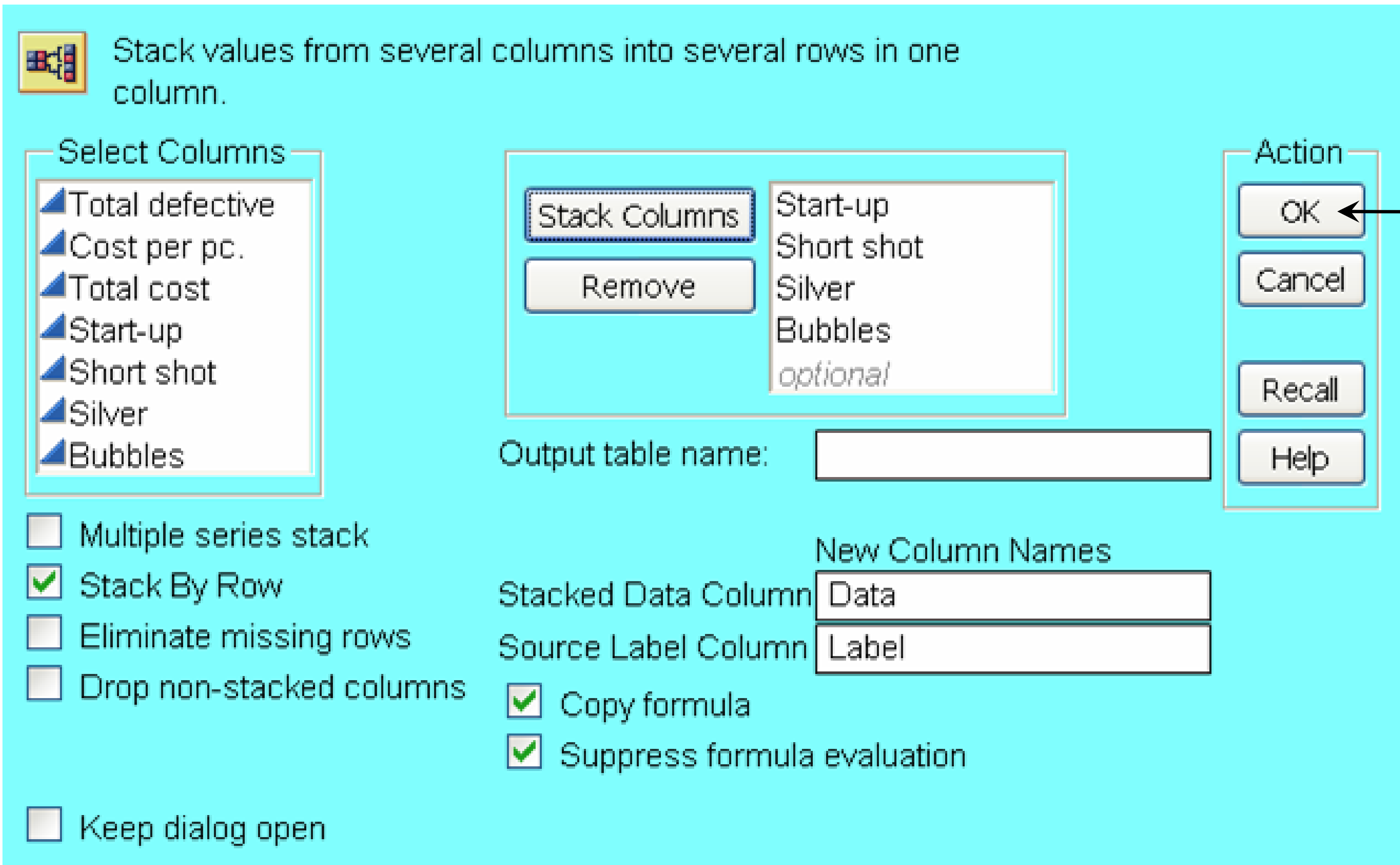
This is what we need →

4/0 Cols / 12/0	Cost per pc.	Defect	Freq	Total cost
1	3	Start-up	3	9
2	3	Short shot	0	0
3	3	Silver	4	12
4	3	Bubbles	0	0
5	5	Start-up	4	20
6	5	Short shot	4	20
7	5	Silver	0	0
8	5	Bubbles	9	45
9	11	Start-up	0	0
10	11	Short shot	6	66
11	11	Silver	12	132
12	11	Bubbles	0	0

→ How do we get there?

Stacking a data table

Tables → *Stack* → Select the defect columns as the *Stack Columns*



Editing the columns

5/0 Cols 12/0	Total defective	Cost per pc.	Total cost	Label	Data
1	7	3	21	Start-up	3
2	7	3	21	Short shot	0
3	7	3	21	Silver	4
4	7	3	21	Bubbles	0
5	17	5	85	Start-up	4
6	17	5	85	Short shot	4
7	17	5	85	Silver	0
8	17	5	85	Bubbles	9
9	18	11	198	Start-up	0
10	18	11	198	Short shot	6
11	18	11	198	Silver	12
12	18	11	198	Bubbles	0

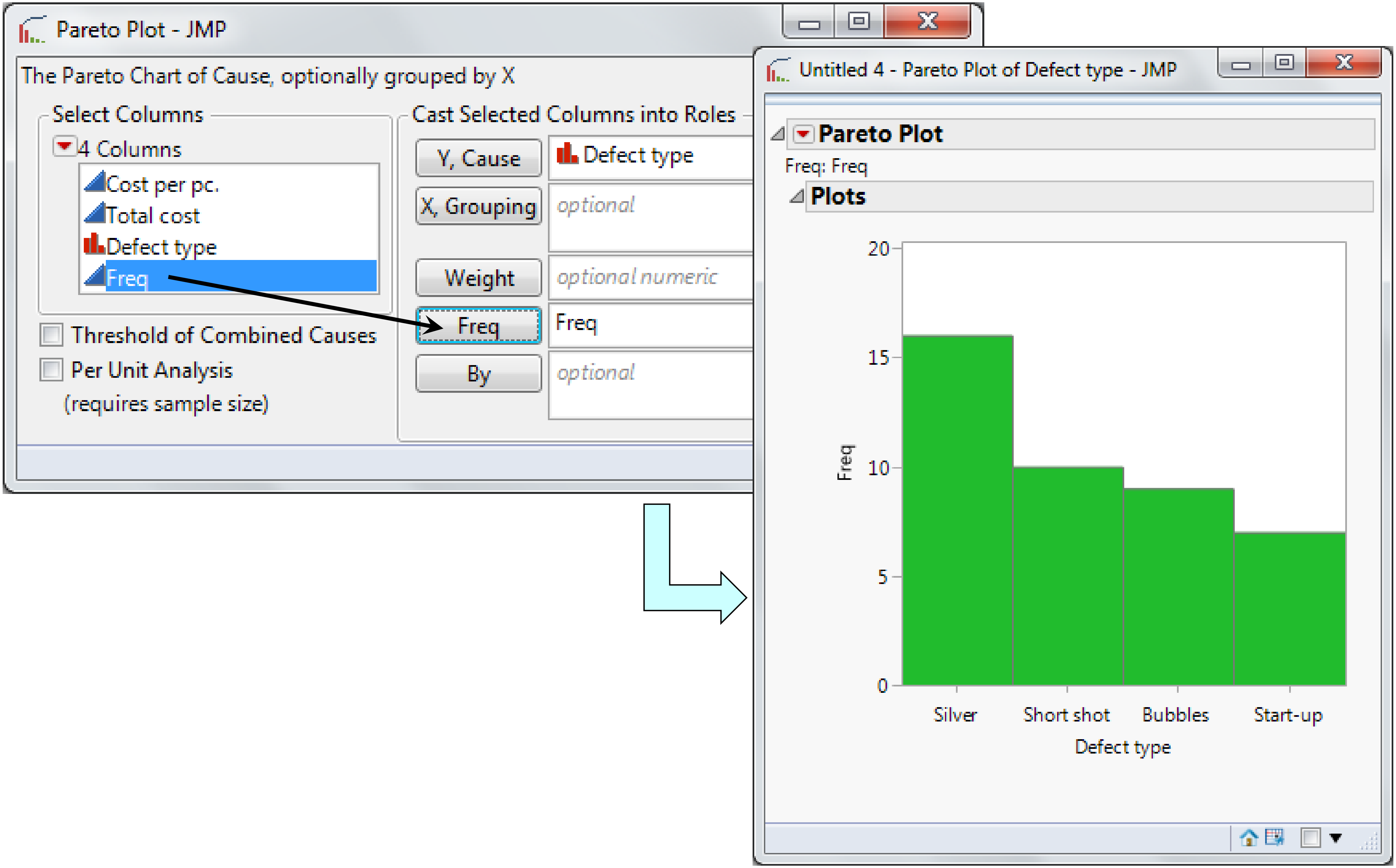
Total defective and *Total cost* are now incorrect row by row

1. Right-click on *Data*
2. Select *Column Info*
3. Rename as *Freq* → OK
4. Rename *Label* as *Defect type*
5. Delete *Total defective*
6. Right-click on *Total cost*
7. Select *Formula* → *Cost per pc.*Freq*
8. Save as *molding data small stacked.xls*

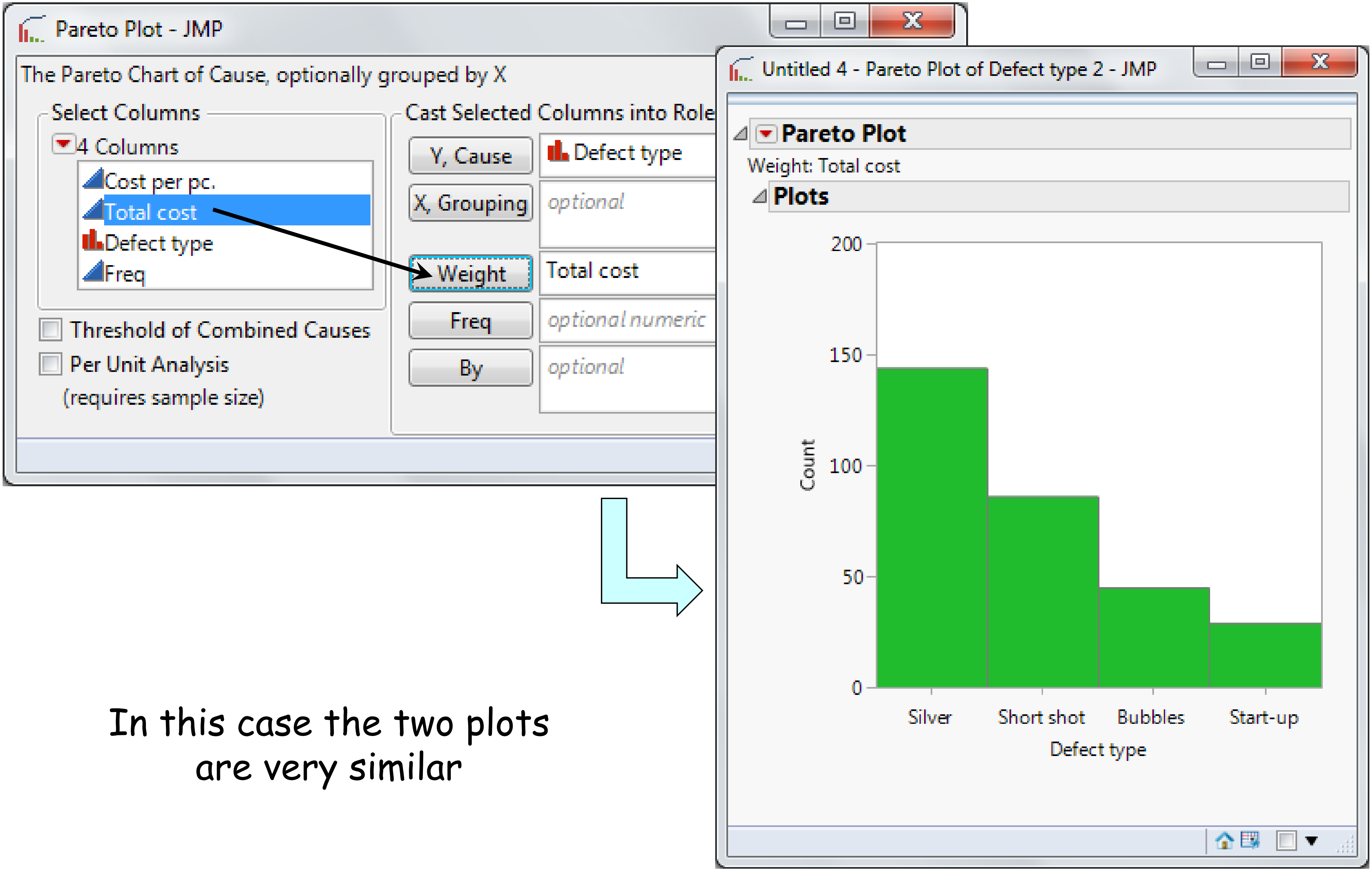
4/0 Cols 12/0	Cost per pc.	Total cost	Defect type	Freq
1	3	9	Start-up	3
2	3	0	Short shot	0
3	3	12	Silver	4
4	3	0	Bubbles	0
5	5	20	Start-up	4
6	5	20	Short shot	4
7	5	0	Silver	0
8	5	45	Bubbles	9
9	11	0	Start-up	0
10	11	66	Short shot	6
11	11	132	Silver	12
12	11	0	Bubbles	0

Pareto plot by frequency

Analyze → Quality and Process → Pareto Plot → set up as shown → OK

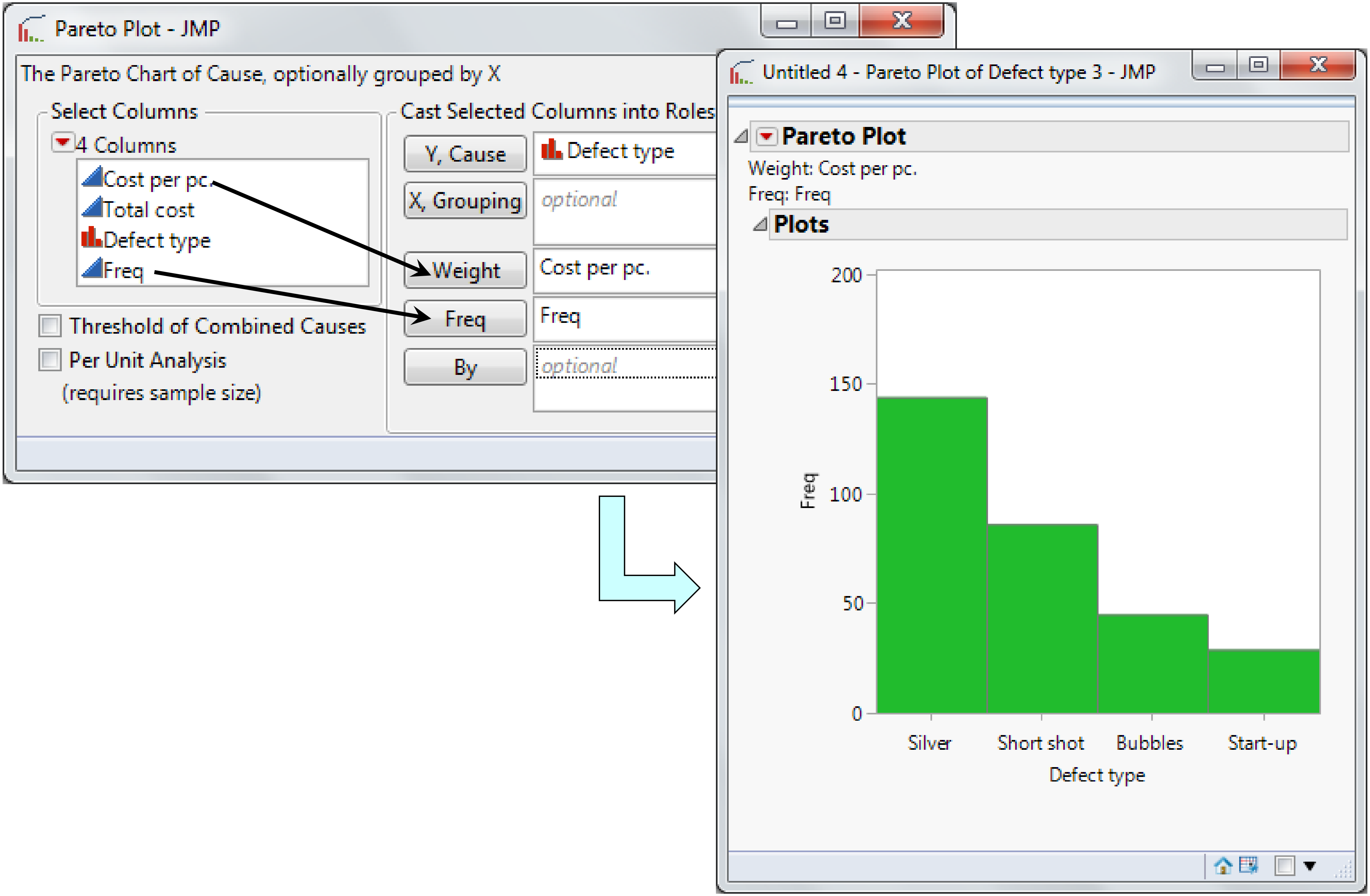


Pareto plot by total cost



In this case the two plots
are very similar

Cost Pareto without calculating the total cost column



Exercise: Appendix

Data sets \ molding process - Pareto.

Use the method described in this section to reformat the file for Pareto analysis. Save the reformatted file as *molding process - stacked*. Create Pareto plots of defect types by frequency of occurrence and total cost.

Tab 2
Regression

1 Introduction to Regression

Regression analysis is used to create an empirical model of the relationship between process inputs (x's) and outputs (y's).

- It is the method for analyzing designed experiments.
- It can also be used with historical data to help identify some factors for an experiment, or to develop an empirical model with that data.

Topics:

- Terminology
- Purposes of regression analysis
- Data collection for use in regression analysis
- The line of best fit
- Simple Regression

Terminology

- The term ***correlation*** is often used any time we speak of relating one variable to another
    - Correlation is a measure of the relationship
    - An input/output relationship between the two variables is not required (for example, two variables measured at the same point in a process)
    - As a result, unrelated things can be "correlated." Remember, correlation *does not* prove causation.
- ***Regression*** analysis yields a model equation of the input-output relationship, $Y = f(X)$, which can be useful in prediction
    - In the dataset, a series of inputs and their resulting output measures are aligned
    - Regression is used to investigate and model the relationship

The result of regression analysis is an empirical model, created from the data/observations, that can be used to:

- Understand and describe the relationship between Y and X's
- Predict Y from X's
- Determine best setting for X's (optimization)
- Reduce variation in Y by controlling X's

Regression analysis is only as good as the data used.

Three basic sources of data are:

- Historical data (data that exists in routine collection systems)
- An observational study (data collected from uncontrolled processes for a specific purpose)
- A designed experiment (data from structured and controlled tests)

Regression analysis is a very big statistical topic and is commonly the analysis type for data from all three sources listed above.

Designs of experiments (DOEs) is the best strategy for many problems we are trying to solve as it is constructed to eliminate many of the problems that exist with the first two sources. However, historical and observational data is often easier to get and can still give powerful insights, although care must be taken with the analysis and conclusions drawn.

Considerations when using historical data

Historical data is often plentiful and easily accessible.

- It may be useful in identifying some variables that are critical to our process

However, there are several potential issues in using it:

- Some relevant data is not available, such as values of critical x's that are not recorded as part of the on-going process
- Reliability of the data is often questionable, including data being missing or lost
- The nature of the data is not helpful in solving the problem, as in situations when an x variable is controlled, so its impact cannot be seen in the regression analysis
- Often, data is used in ways that were not intended, such as using available data as a surrogate for what was really needed

Caution: We will not be able to cover the many aspects of creating and validating regression models from historical data in this course. If you choose to do this, proceed with caution! Better yet, get additional help.

Considerations when using an observational study

In an observational study, we would observe the process, with as little interaction or disturbance as possible, in order to obtain the data.

- With adequate planning, an observational study can yield accurate, complete, reliable data
- These studies can lead to ideas on what might be impacting the process
- However, these studies often provide limited information about specific relationships of interest, such as the impact of a variable that is tightly controlled in normal operation

"Simple" regression

Simple linear regression refers to the case when there is only one regressor (variable) x used.

- In simple regression, the model equation is for a best-fit line

- The form of the model equation created is:

$$Y = b_0 + b_1 x_1 + error$$

  where b_0 is the intercept and b_1 is the slope of the line.

- This may remind you of your early algebra days, when you learned the equation for a line between two points:

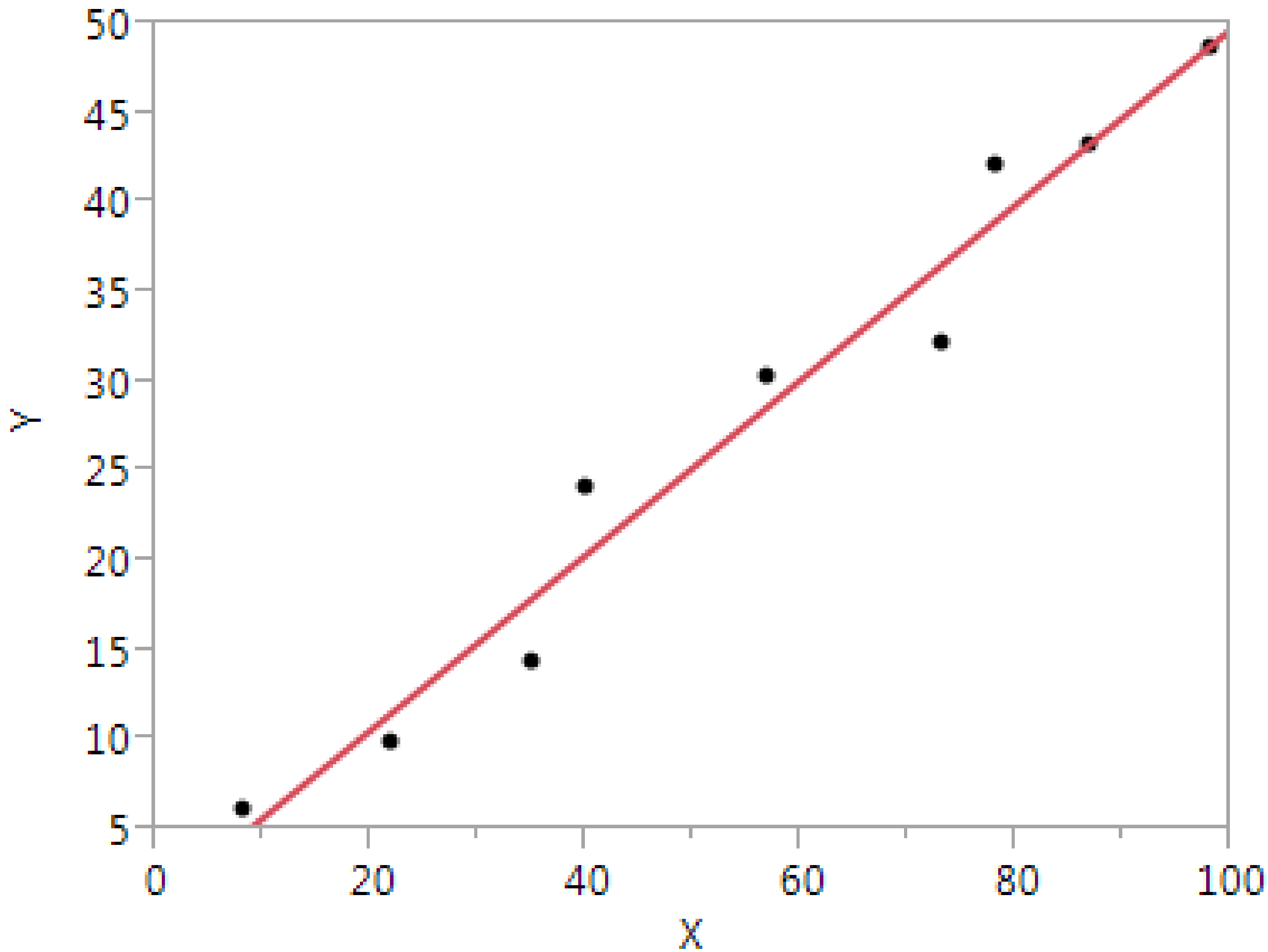
$$Y = mx + b$$

- Because there is variation (and more than two points to create the line), there will be scatter around the best-fit line determined by regression analysis.

Simple regression (cont'd)

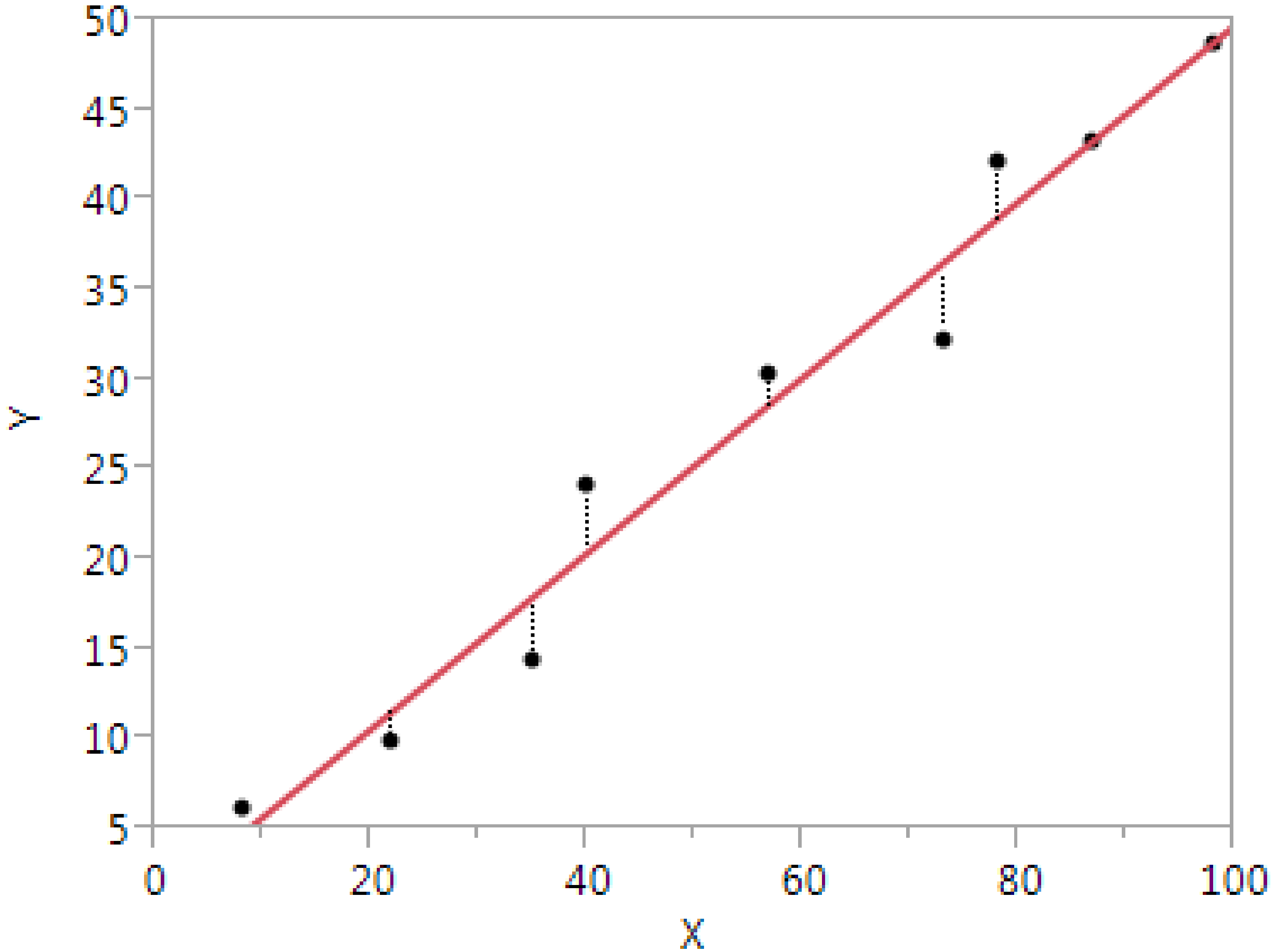
Intercept ↓ Slope ↓

$$Y = 0.8387 + 0.4891\,X + \text{“Error”}$$



The line of best fit

The best-fitting line is the one that minimizes the sum of the squared "errors"



The line of best fit (cont'd)

- **"Errors" are the vertical distances between each Y data value and the fitted line**
- The line of best fit is the one that minimizes the sum of the squared errors
- This is the simplest example of *least-squares model fitting*
- The fitted line is often referred to as the predicted Y value

LSSV2 student files \ ANOVA linear fit

Worksheet \ *Prediction & error 1*

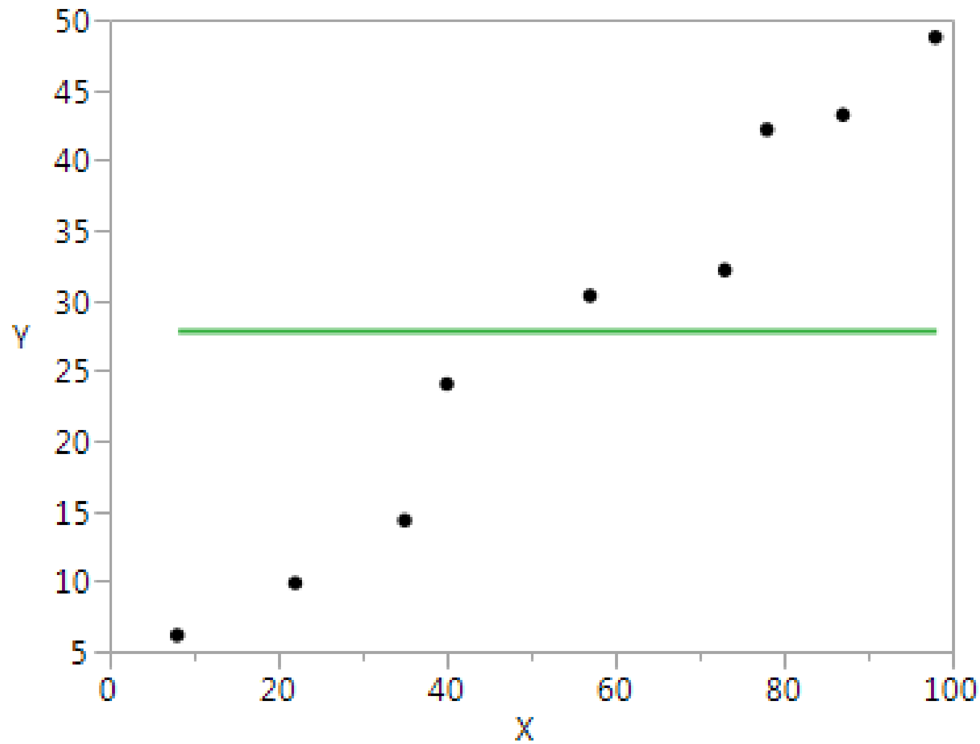
Y = 27.9033 + 0.0000 X

	X data	Y data		Prediction		Error
	8	6.16		27.90		-21.74
	22	9.88		27.90		-18.02
	35	14.35		27.90		-13.55
	40	24.06		27.90		-3.84
	57	30.34	=	27.90	+	2.44
	73	32.17		27.90		4.27
	78	42.18		27.90		14.28
	87	43.23		27.90		15.33
	98	48.76		27.90		20.86
Sum of squares (SS)		8901.3	=	7007.4	+	1893.9
Degrees of freedom (DF)		9	=	1	+	8
Root mean square error (RMSE)						15.39

Average Y	27.90
STDEV of Y	15.39

Finding the line of best fit (cont'd)

In this worksheet we ignore the X variable completely, and use the average value of Y as the prediction. This is just the calculation of the mean and standard deviation of the Y variable. (The values in cells I14 and E17 are the same.)



The sum of the squared errors (cell I12) can be dramatically reduced by using the X variable to "explain" more of the variation in the Y variable.

Finding the line of best fit (cont'd)

Worksheet \ *Prediction & error 2*

$$Y = 0.8387 + 0.4891\,X$$

	X data	Y data		Prediction		Error
	8	6.16		4.75		1.41
	22	9.88		11.60		-1.72
	35	14.35		17.96		-3.61
	40	24.06		20.40		3.66
	57	30.34	=	28.72	+	1.62
	73	32.17		36.54		-4.37
	78	42.18		38.99		3.19
	87	43.23		43.39		-0.16
	98	48.76		48.77		-0.01
Sum of squares (SS)		8901.3	=	8838.0	+	63.3
Degrees of freedom (DF)		9	=	2	+	7
Root mean square error (RMSE)						3.007

Average Y	27.90
STDEV of Y	15.39
Adjusted R square	0.962

Proportion of total Y variation caused by ("explained by") X variation

Degrees of freedom for regression

N = total sample size

G = number of parameters in the equation

= DF for the prediction column

N - G = DF for the error column

- The *Error* DF is more important than the *Prediction* DF
- It determines the accuracy of the predicted values
- When DF is mentioned without a qualifier, it usually means *Error* DF

Steps in Simple Regression

1. Run Analyze > Fit Model in JMP to investigate the relationship between y and x

2. Check the p-value for the fit to determine whether the regression is significant. If not, then no need to go further.

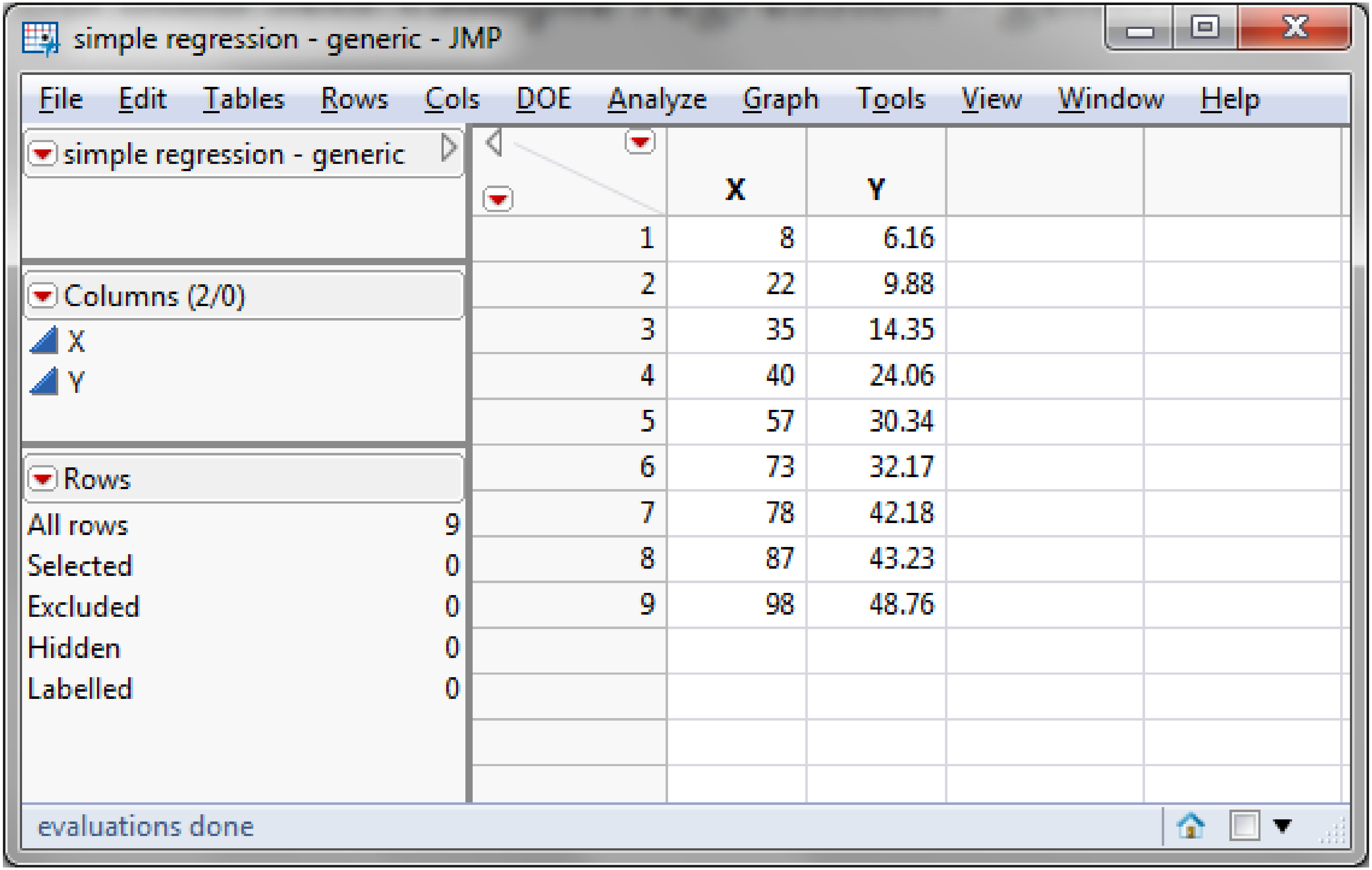
3. If the regression is significant, determine the strength of the relationship, using the $Adjusted\ R^2$

4. Check model adequacy by reviewing the residuals plots
   - Residual Normal Quantile Plot
   - Residual by Predicted Plot
   - Studentized Residuals (in run order)

We'll go through these steps and additional analysis details, for simple regression in the following example.

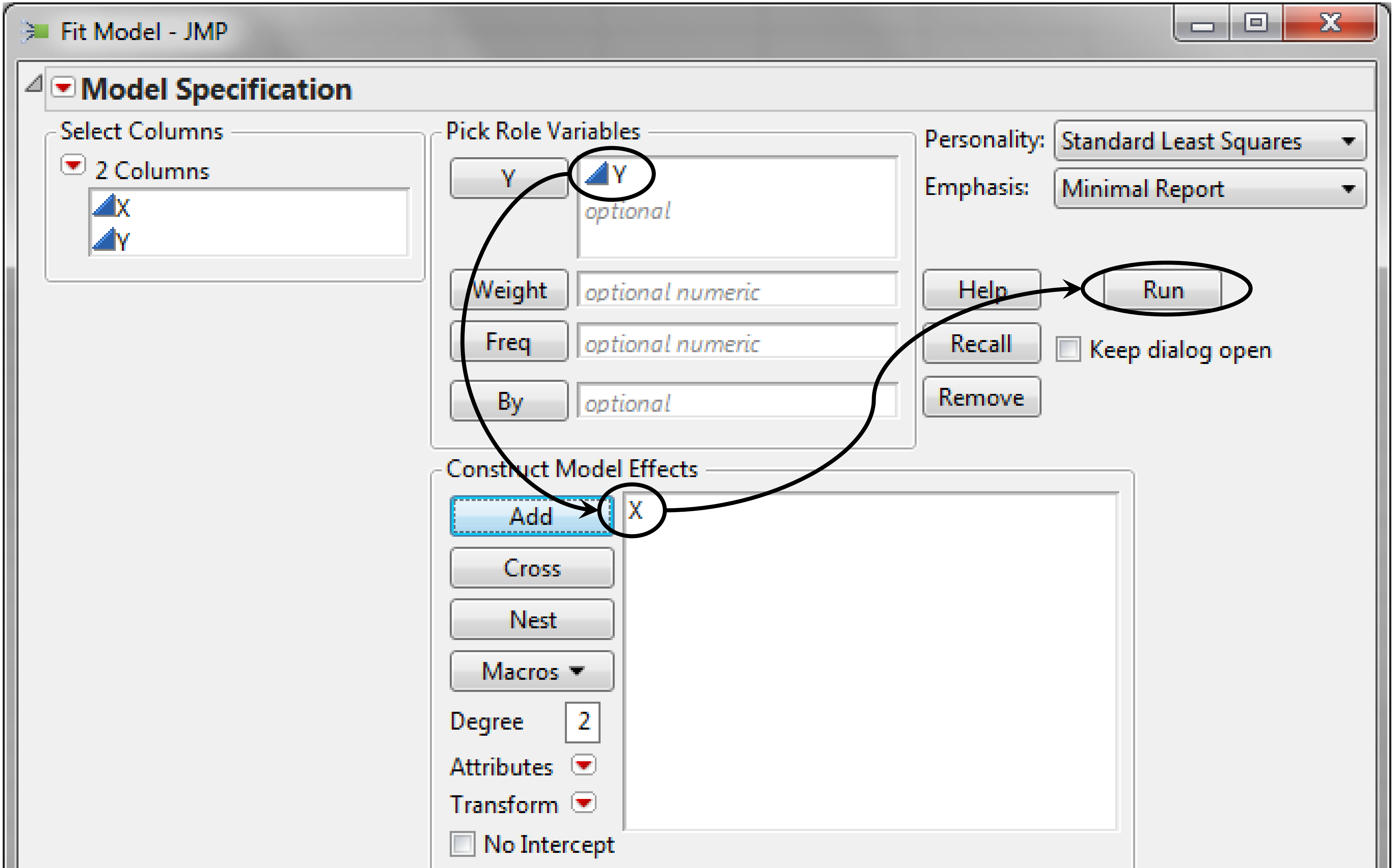
Simple Regression in JMP

Open: Data sets \ simple regression - generic



Simple Regression in JMP (cont'd)

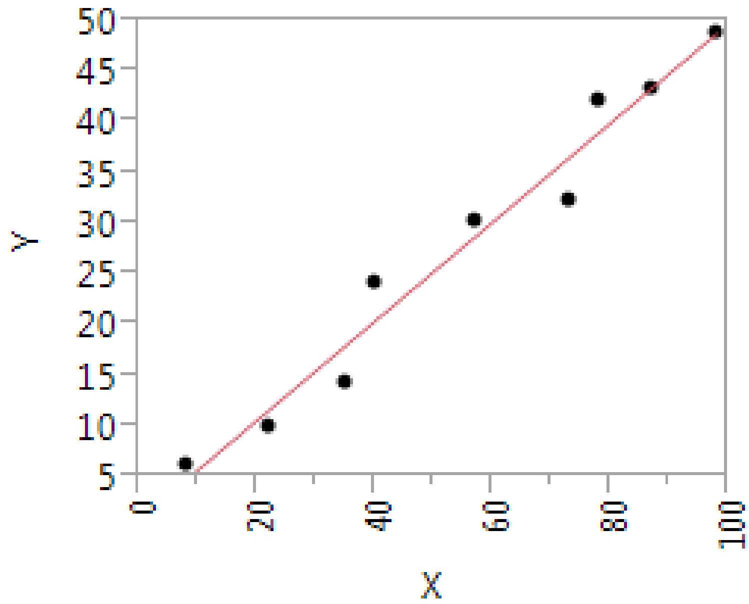
Analyze → *Fit Model* → Set up as shown → Run



Analysis details

Response Y

Regression Plot



Summary of Fit

RSquare	0.966581
RSquare Adj	0.961807
Root Mean Square Error	3.006984
Mean of Response	27.90333
Observations (or Sum Wgts)	9

- The **Root Mean Square Error (RMSE)** is the standard deviation of Y caused by factors other than X
- It can be thought of as the **standard deviation** about the fitted line (or model)
- Also known as the "error" or "residual" standard deviation
- Smaller is better

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1830.6557	1830.66	202.4624
Error	7	63.2937	9.04	Prob > F
C. Total	8	1893.9494		<.0001*

- **P-value** indicates whether the regression is significant
- This low p-value shows that it <u>is</u> significant

Analysis details (cont'd)

Summary of Fit

RSquare	0.966581
RSquare Adj	0.961807
Root Mean Square Error	3.006984
Mean of Response	27.90333
Observations (or Sum Wgts)	9

Analysis of Variance

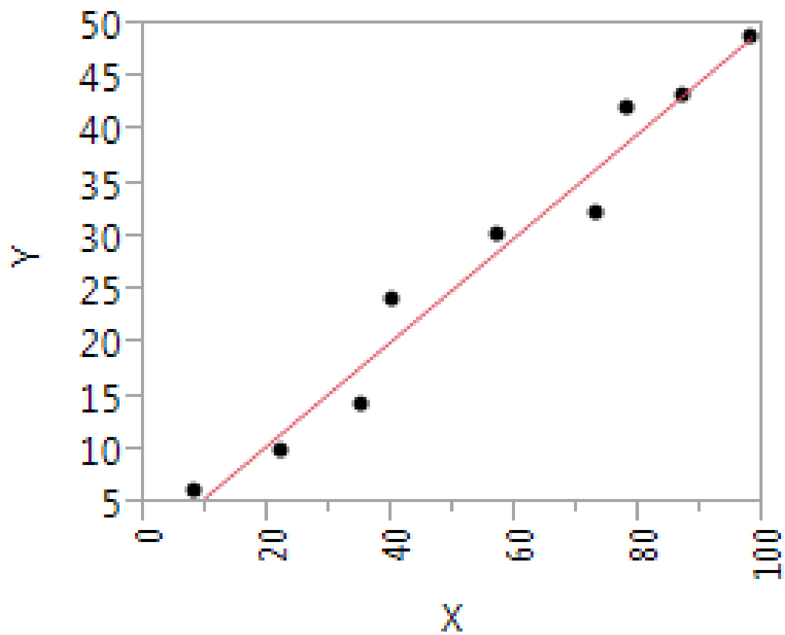
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1830.6557	1830.66	202.4624
Error	7	63.2937	9.04	Prob > F
C. Total	8	1893.9494		<.0001*

R^2
"Coefficient of Determination"

- Proportion of the variation in Y that is "explained by" variation in X.
- Varies from 0 to 1.
- Larger is better
- Unitless

Analysis details (cont'd)

Regression Plot



Summary of Fit

RSquare	0.966581
RSquare Adj	0.961807
Root Mean Square Error	3.006984
Mean of Response	27.90333
Observations (or Sum Wgts)	9

Analysis of Variance

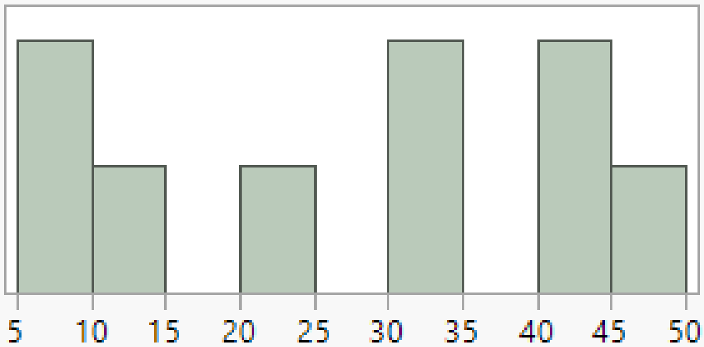
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1830.6557	1830.66	202.4624
Error	7	63.2937	9.04	Prob > F
C. Total	8	1893.9494		<.0001*

- **Adjusted R^2** also gives us the proportion of Y variation *explained by* the model (a line in simple regression)
- Varies from 0 to 1
- Larger is better
- **Always use the Adjusted R^2 value, not R^2**
- Adjusted R^2 takes the number of model terms into account and penalizes for including insignificant terms
- In this example, the simple regression model explains much of the variation in Y.

How R^2 and R^2_{Adj} are calculated

Distributions

Y



Summary Statistics

Mean	27.903333
Std Dev	15.386477
N	9
Minimum	6.16
Maximum	48.76
Median	30.34

Standard Deviation (STDEV) of the data set

$$R^2 = 1 - \frac{SS_{Error}}{SS_{Total}}$$

$$R^2_{Adj} = 1 - \frac{SS_{Error}/(n-p)}{SS_{Total}/(n-1)} = 1 - \left(\frac{RMSE}{STDEV}\right)^2$$

p = number of terms in the model (including the intercept)

n = sample size (number of measurements in the data set)

SS_{Total} is the sum of squares of the data (measurements in the data set)

SS_{Error} is the sum of squares of the Errors or residuals

We saw the sum of squares calculations earlier, in the ANOVA

Why use Adjusted R^2?

There is a potential problem with R^2:

- R^2 always increases when terms are added to a model, even when the terms are not significant
- **This is particularly a problem in multiple regression**, as it can lead to “overfitting,” giving false confidence in using the model, especially for prediction.
- Adjusted R^2 corrects for this by considering the number of terms in the model
- Adjusted R^2 can actually decrease if non-significant terms are added to a model

Adjusted R^2 is the recommended statistic for determining the proportion of variation in Y explained by the model

P-values for the ANOVA and individual model parameters

Red triangle next to *Response Y* → *Regression Reports* → *Parameter Estimates*

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1830.6557	1830.66	202.4624
Error	7	63.2937	9.04	Prob > F
C. Total	8	1893.9494		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.8386661	2.150023	0.39	0.7081
X	0.4891205	0.034375	14.23	<.0001*

- In regression of Y on a single X, the Analysis of Variance P-value is the same as the P-value for the slope of the line.
- The P-value for the slope of the line indicates the evidence of a correlation between Y and X.
- Significance of individual model terms are determined by testing whether their regression coefficient is equal to 0, using the t statistic. Hypotheses are:

$$H_0: b_i = 0$$
$$H_1: b_i \neq 0$$

- This is a test of the contribution of the model term, given the other terms in the model.

P-values for individual model parameters

Estimates and P-values for the slope and intercept

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.8386661	2.150023	0.39	0.7081
X	0.4891205	0.034375	14.23	<.0001*

Model: $\boldsymbol{Y = 0.84 + 0.50X + error}$

- In this example, the P-value for the slope of the line indicates very strong evidence of a correlation between Y and X.
- The P-value for the Intercept indicates that it is not significant.
  - ➢ **Best practice is to leave the Intercept in the model, whether or not the P-value indicates that it is significant**
  - o Regression equations are developed, and are only valid, over the region of the regressor variables (x's) contained in the data set
  - o Forcing the model to pass through (0, 0) by removing the intercept, can create problems in the region being modeled

Using Adjusted R^2 and p-values

Both the Adjusted R^2 and the p-values must be considered, in order to understand what has been learned in the analysis.

When the resulting model has:

- **High Adjusted R^2 and significant model term p-values,** this is ideal. Factors driving the response have been identified and the variation is largely explained. A decent model has been created.
- **Low Adjusted R^2 and significant model term p-values,** more work must be done. Some significant factors influencing the response have been identified, but the low Adjusted R^2 indicates that <u>other important factors exist</u>. These need to be found, for the model to be useful.
- **High R^2 and insignificant model terms**, this is usually due to the data violating the assumptions of the regression analysis. There is more information on this scenario in upcoming slides.
- **Low Adjusted R^2 and insignificant model terms,** no relationship between X and Y variables have been found. Usually this means that new ideas about which factors influence Y must be developed, although it can occasionally be due to missing higher order terms.

This slide intentionally left blank

2 Checking Model Adequacy

In least squares fit regression (continuous Y), the analysis methods used to calculate regressor coefficients and their p-values, depend on certain assumptions being met.

Assumptions:

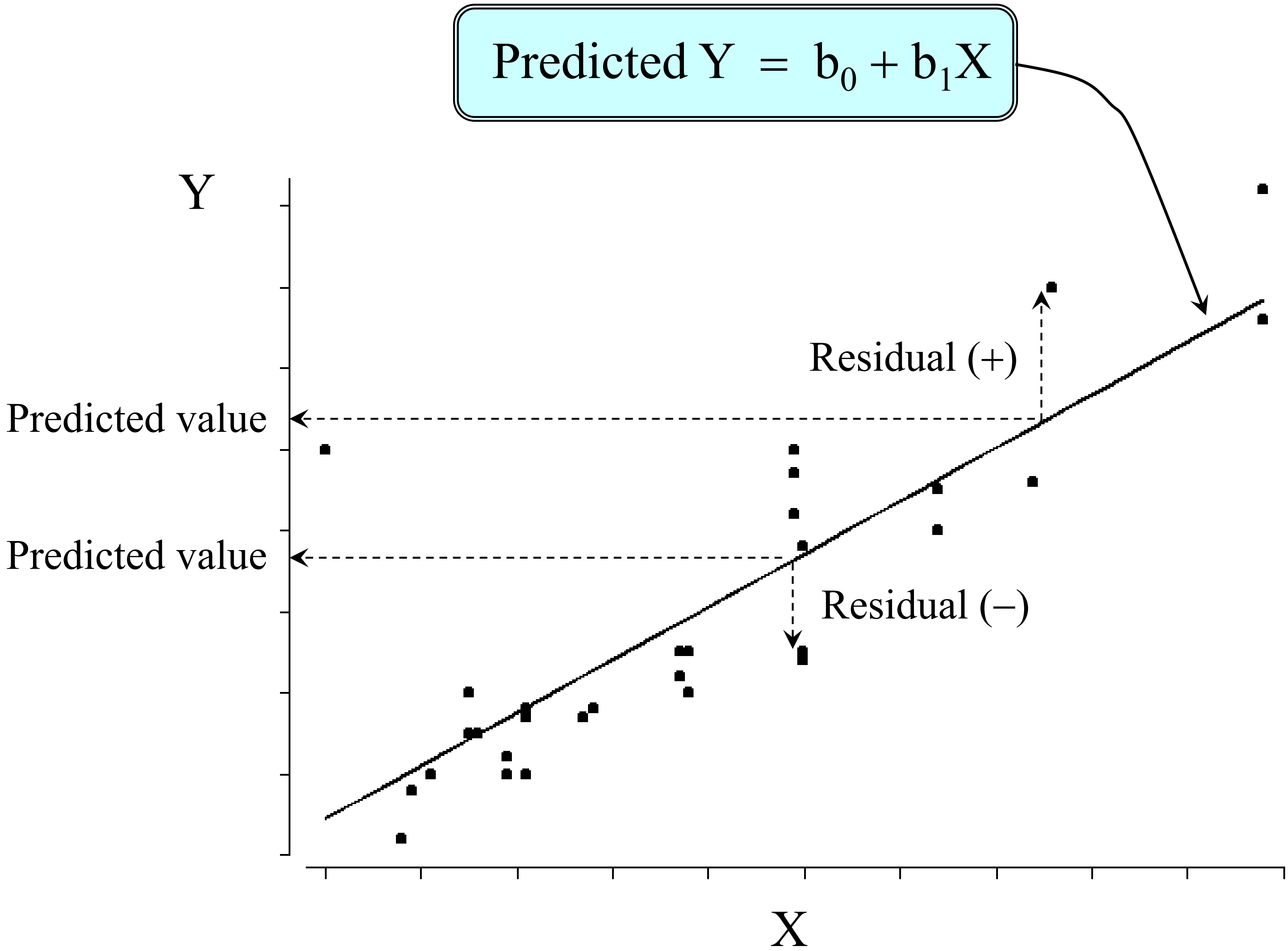
- Errors (residuals) are normally and independently distributed with mean zero and constant variance (σ^2)
- Observations are adequately described by the model

Whether performing regression from "file cabinet" data or analyzing the results of a designed experiment, these assumptions must be validated.

To validate that these assumptions have been met, the *residuals* are examined:

1. **Normal Probability Plot of Residuals**
   - Validate that the residuals are normally distributed
   - In JMP, this is the *Residual Normal Quantile Plot*
2. **Residuals vs. Predicted (or Fitted) Values**
   - Validate constant variance and mean 0
   - In JMP, this is the *Residual by Predicted Plot*
3. **Residuals vs. Run Order**
   - Verify independence of errors
   - There should be no patterns over the timeframe of the data
   - In JMP, the best graph to use is *Studentized Residuals*
   - The JMP data table must be in run order for *Studentized Residuals* to graph the residuals in run order

Residuals Review



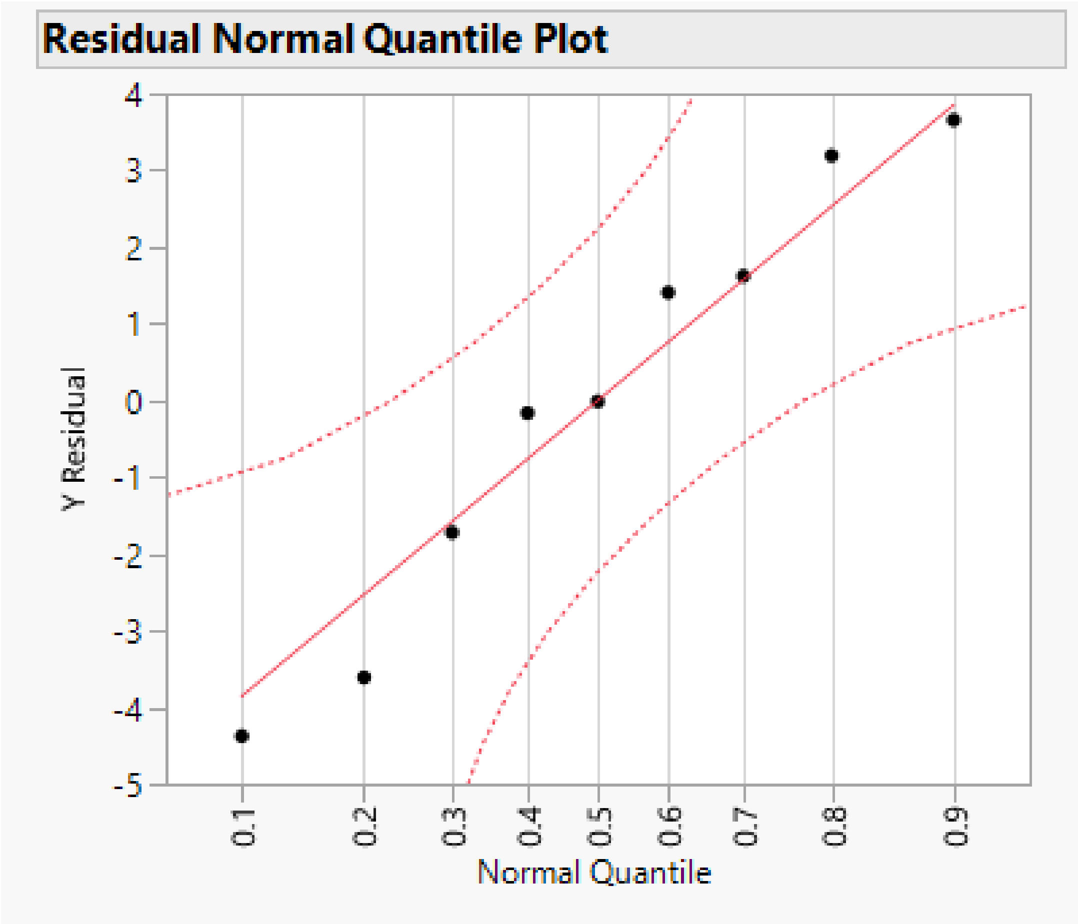
A fitted model, the equation generated during regression, gives the predicted mean value of the response variable as a function of the predictor variables. These predicted mean values are also called *predicted values*, or just *predicted* for short. The *residual value* is the data (observation) value minus the predicted value. Residual values are called *residuals* for short.

These terms are easiest to visualize in the simple linear model shown above. A predicted value is the fitted line evaluated at some X value. A residual is the difference between a measured (observed) Y value and the predicted value at the corresponding X.

Residuals contain information about the magnitude and direction of variability in the data relative to the fitted model.

- An unusually large residual might signal a measurement error, data entry error or some other type of outlier.
- A systematic trend or pattern in the residuals might signal an inadequacy in the fitted model.

Residual Analysis



In viewing the Residual Normal Quantile Plot for the *simple regression-generic*, we can see whether the residuals are normally distributed.

Notes

If residuals are normally distributed, the plot will be approximately a straight line.

Emphasis should be on the central values of the plot, rather than the ends

It is common for plots to bend upward at the high end and downward at the low end.

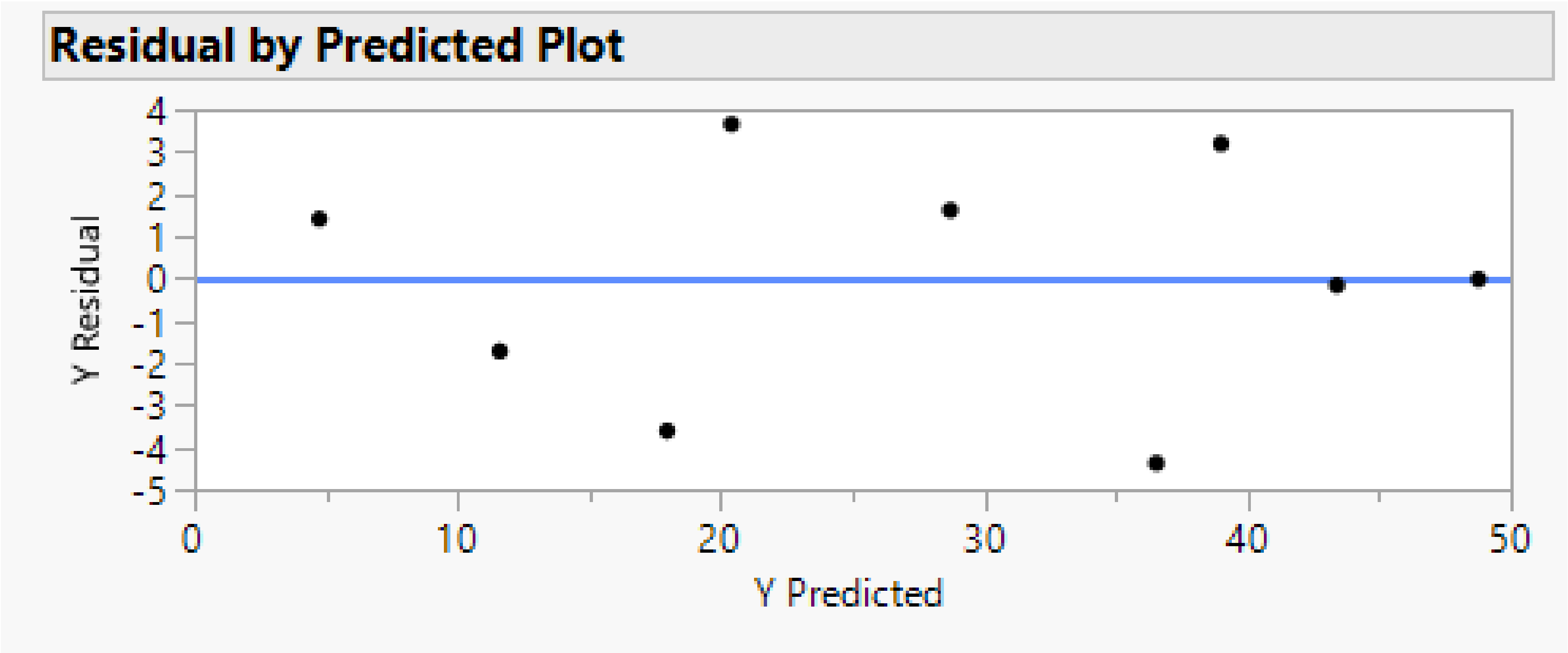
Small sample sizes, such as from experiments, often appear more non-normal

Use the "Fat Pencil" Rule: If a "fat pencil" placed over the central points would cover them on the plot, then the residuals are approximately normal (good enough). Hyperbolic bands displayed in JMP plots give these bounds.

A curve throughout the plot is a strong indication of non-normality. In this case, a transformation would be needed.

The plot above shows an error (residuals) distribution that is approximately normal, so it is not concerning.

Residual Analysis (cont'd)

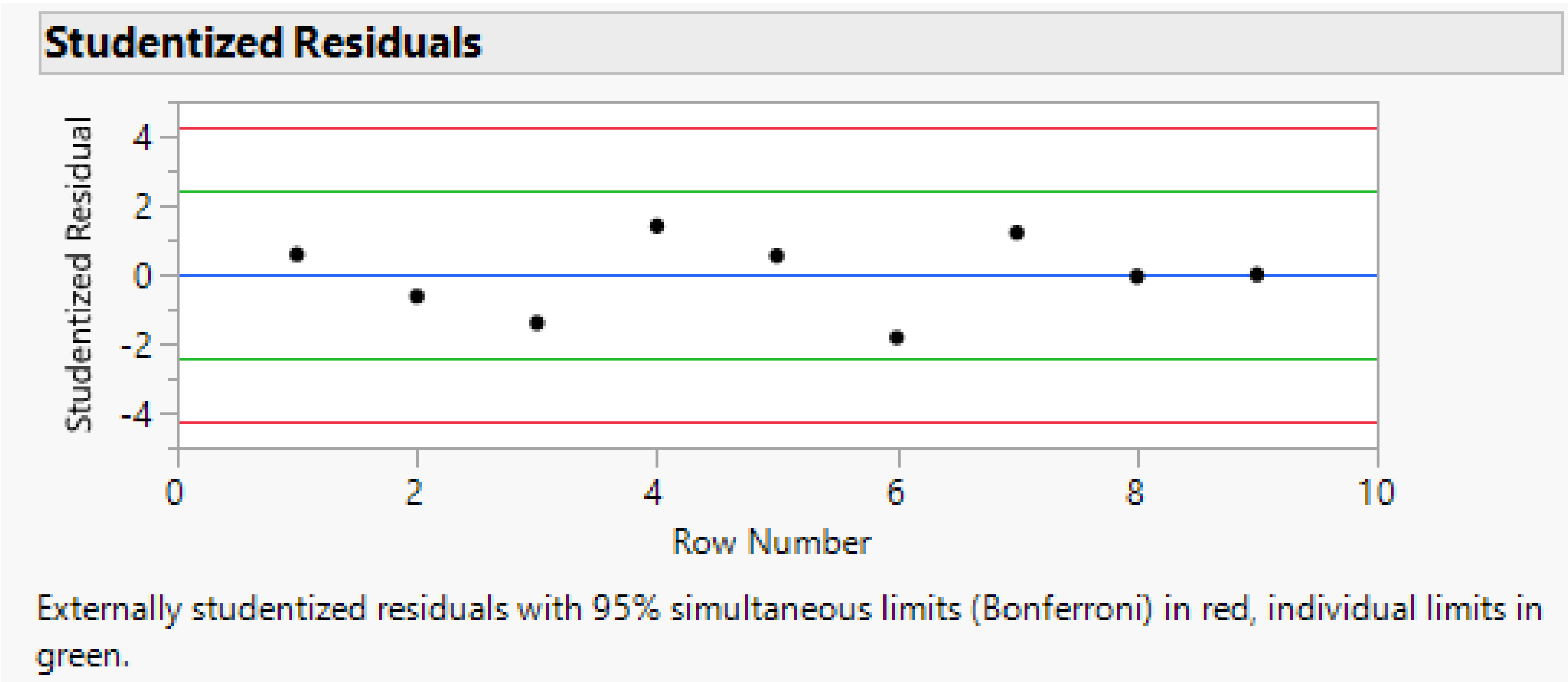


In viewing the Residual by Predicted Plot for the *simple regression-generic*, we can see whether the residuals have constant variance and mean 0.

Here the residuals are plotted against the predicted values. This is a good all-around diagnostic plot.

"Healthy" residuals look like random scatter around 0. There should be no obvious patterns. The amount of "scatter" or variance (how high and low the plot goes) should be consistent across the graph. This verifies the assumption of constant variance. If the variance is increasing or decreasing across the graph, a transformation is needed.

Residual Analysis (cont'd)



In viewing the Studentized Residuals for the *simple regression-generic*, the best form for checking residuals by run order, we can see whether there are any patterns over the timeframe of the data.

Note that the data table must be in run order for this plot.

Again, on this graph, healthy residuals look like a random scatter around 0.

Runs (points in a row) of positive-negative-positive-negative residuals indicate correlation between runs. This implies that the assumption of independence has been violated. **In designed experiments, randomization protects against this! Do it every time!**

This plot can also show a change in variance over the time span of the experiment. This could be due to increased skill as the experiment progresses, a process drift, operator fatigue, tool wear, etc. This type of problem would show as an increase or decrease in spread or "scatter" of the residuals across the graph. Increasing or decreasing variance indicates the need for a transformation.

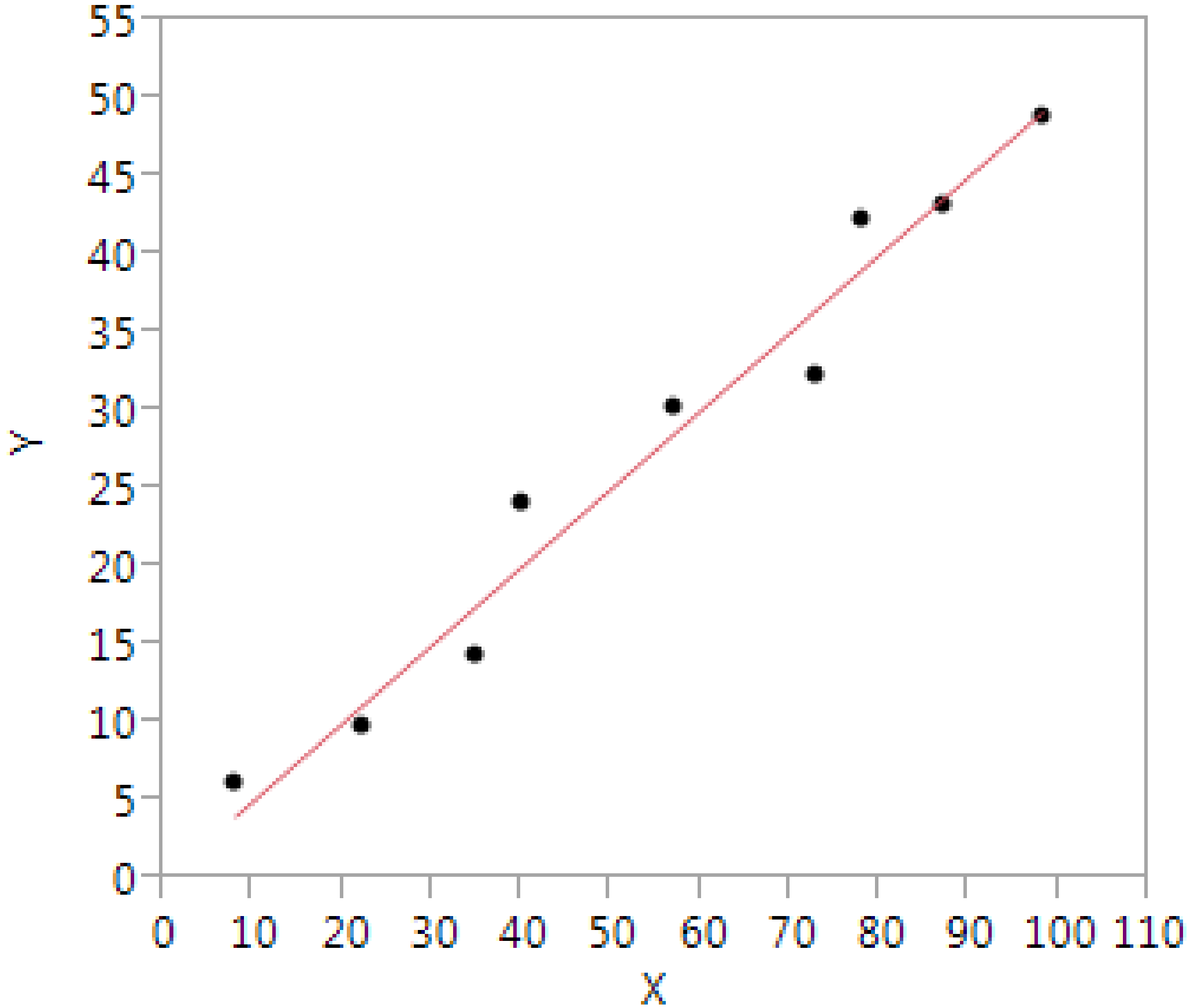
3 Using the Model: RMSE and Prediction Profiler

In this section, we'll see how we can:

- Use the Root Mean Square Error (RMSE) in predicting our future process variation,
- Use JMP's Prediction Profiler to help us optimize our process, and
- Estimate our future % defective, using the t distribution calculator.

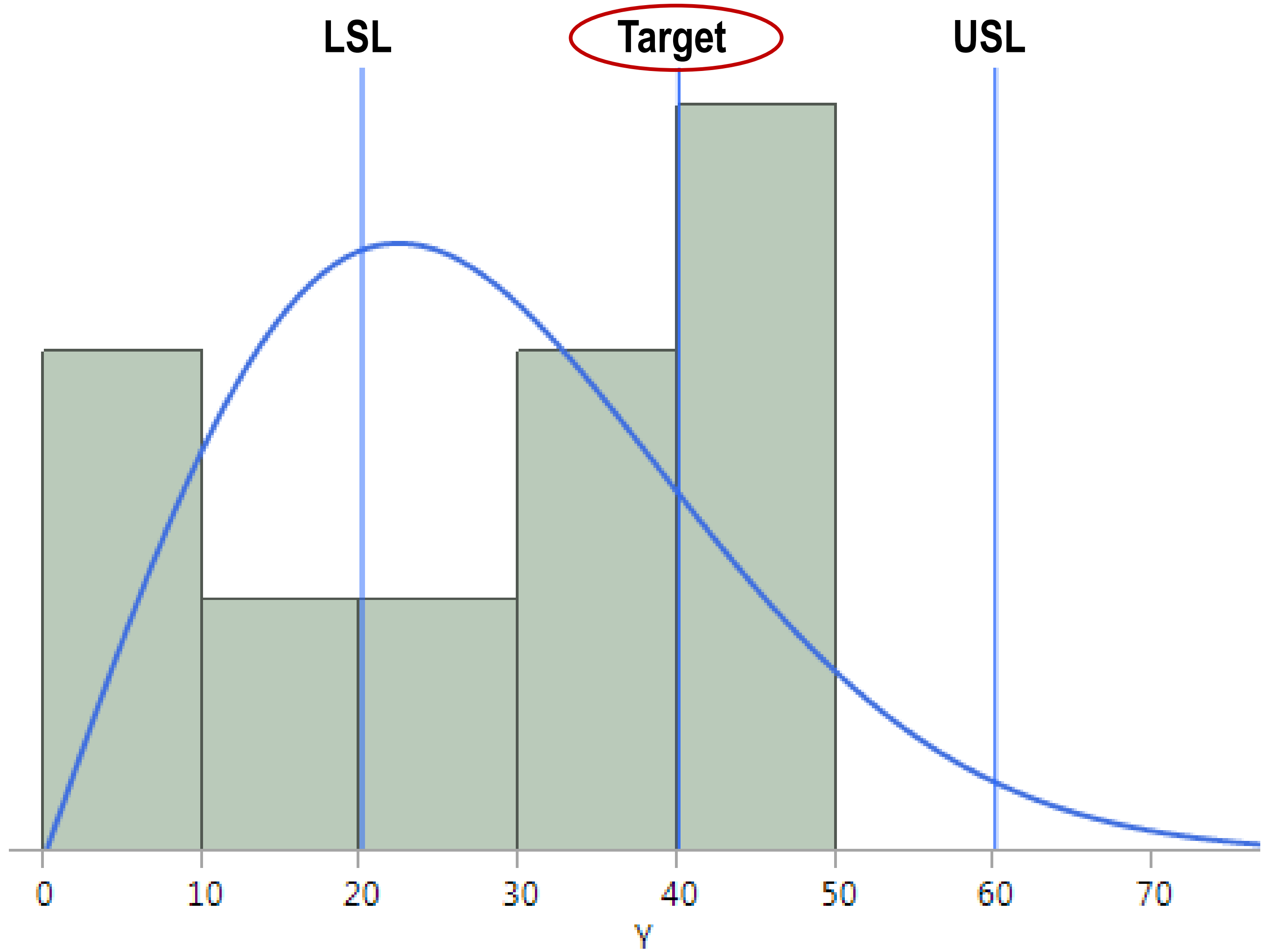
Using the Model (cont'd)

When Y is correlated with a controllable X variable,



how can we use the regression to improve the Y capability?

Using the Root Mean Square Error (RMSE)



Suppose we are not happy with our current process capability

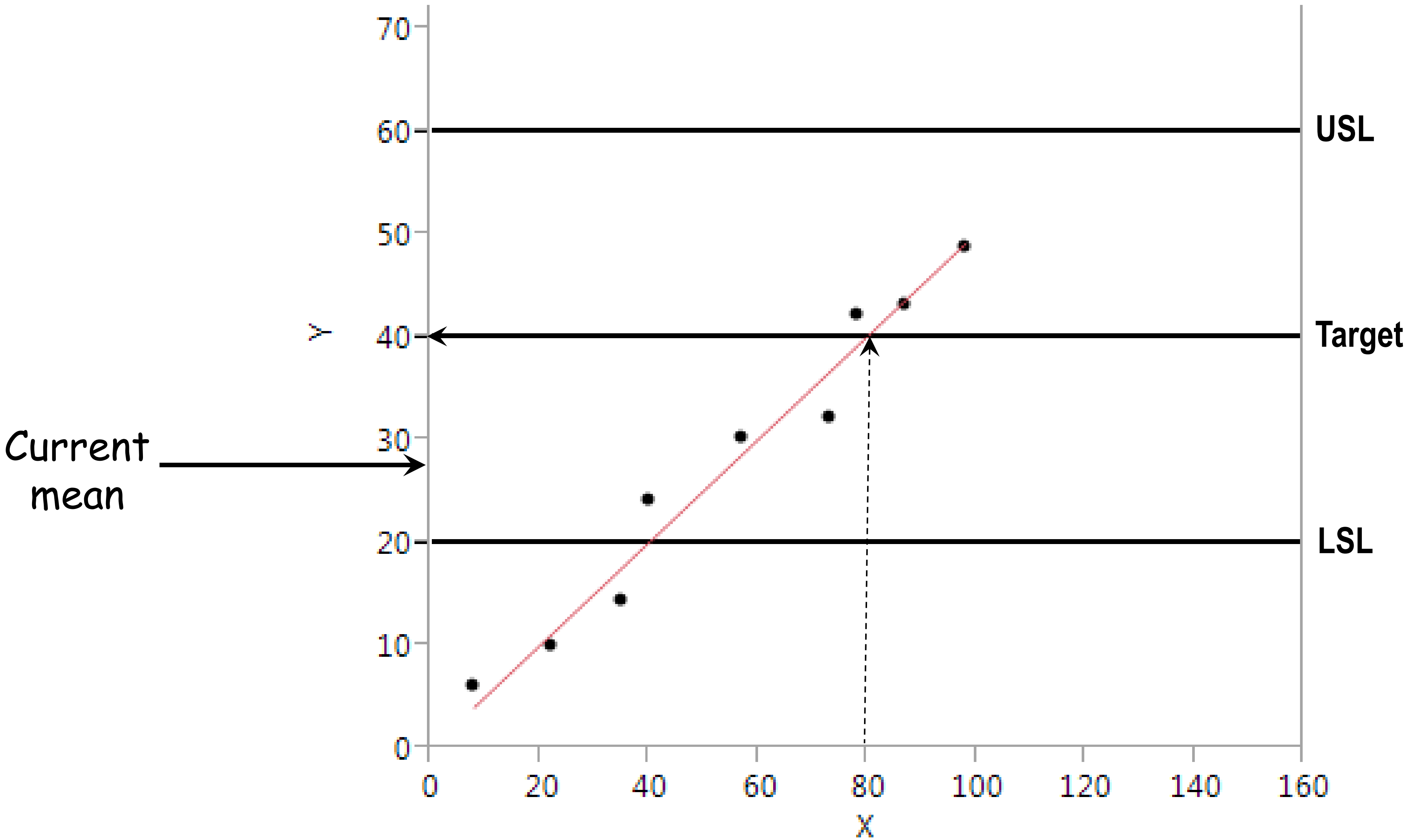
Mean = 27.9, Std dev = 15.4

Defective in the data: 33.3%

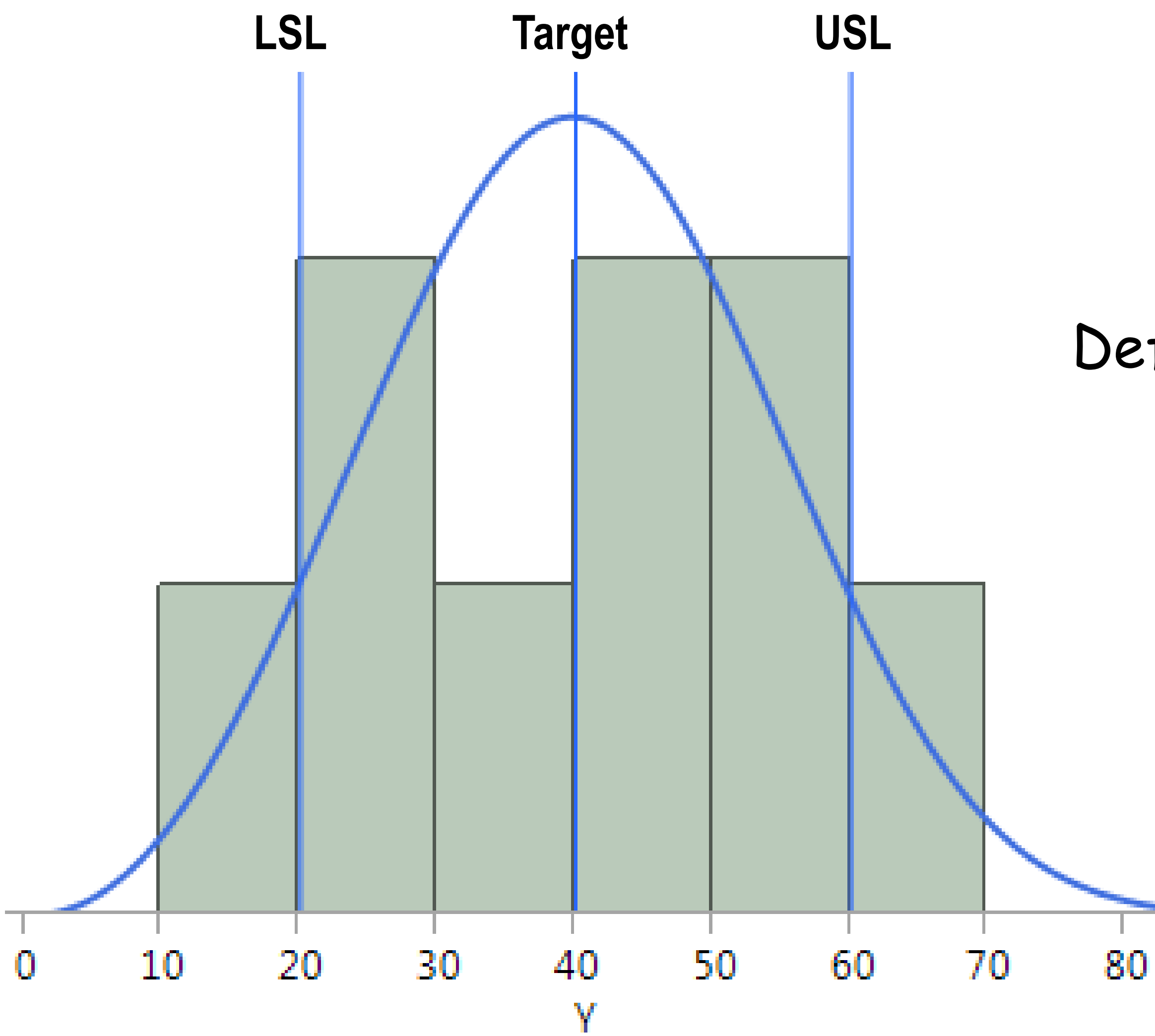
Predicted from distribution curve: 35.8%

RMSE (cont'd)

If we control X at 80, the mean will change from 27.9 to 40



RMSE (cont'd)



Mean = 40.0

Std dev = 15.4

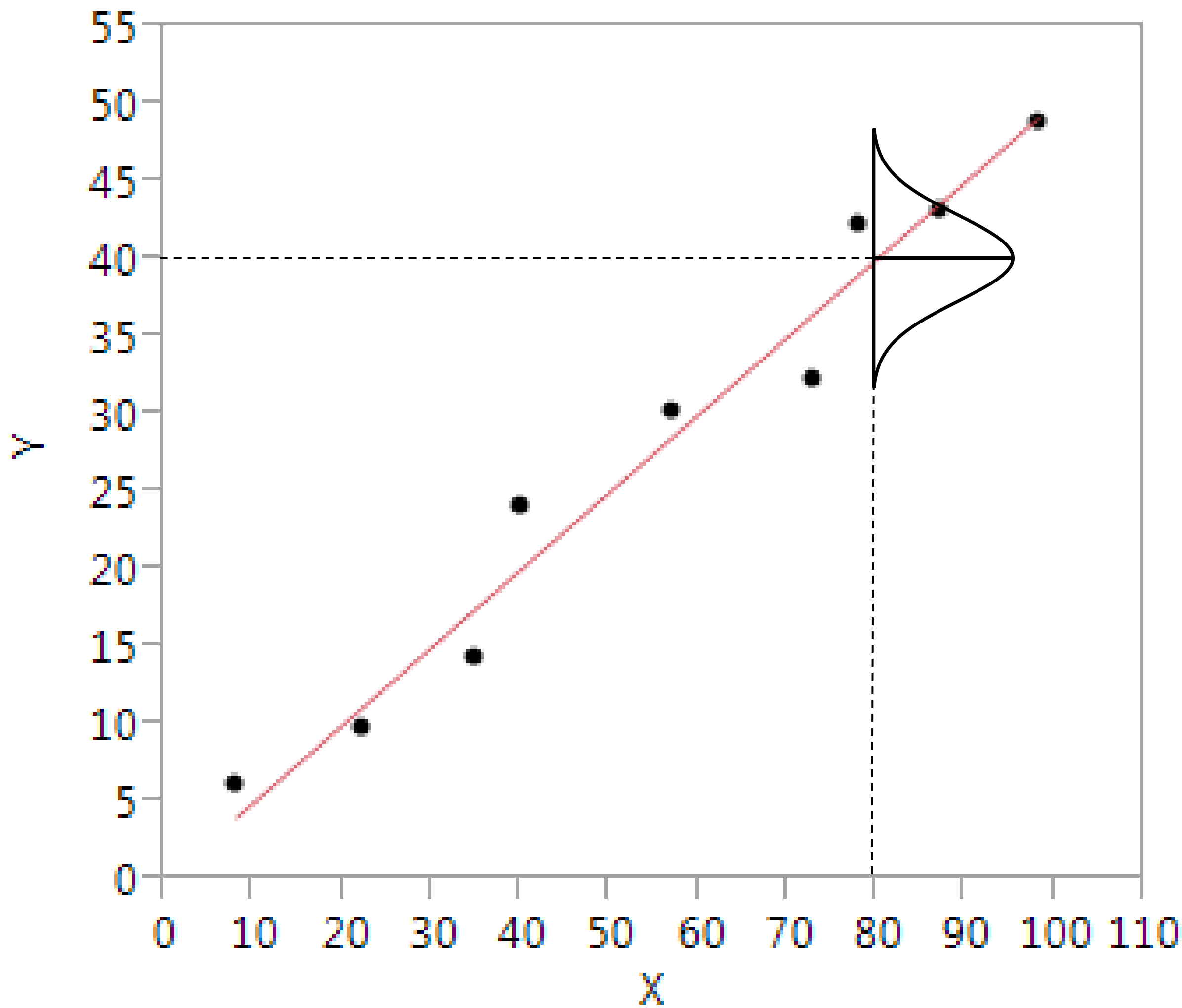
Defective in the data: 22.2%

Distribution curve: 15.9%

- Moving mean Y to the center of the spec range does reduce % defective
- Is the mean the only thing that changes when we control X at 80?

RMSE (cont'd)

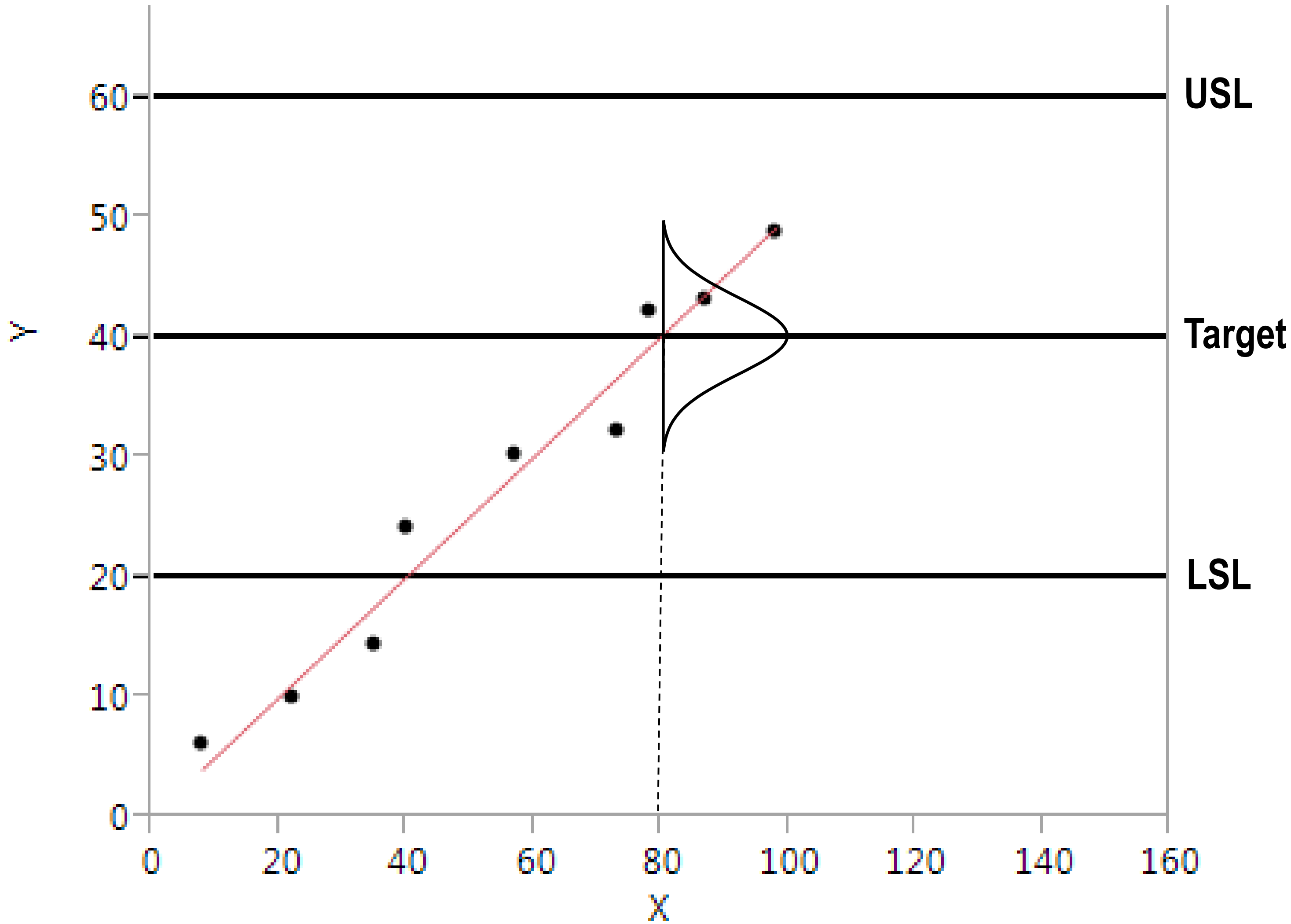
By definition, RMSE is the standard deviation of Y that would result from eliminating the variation in X



σ = RMSE
= 2.84

RMSE (cont'd)

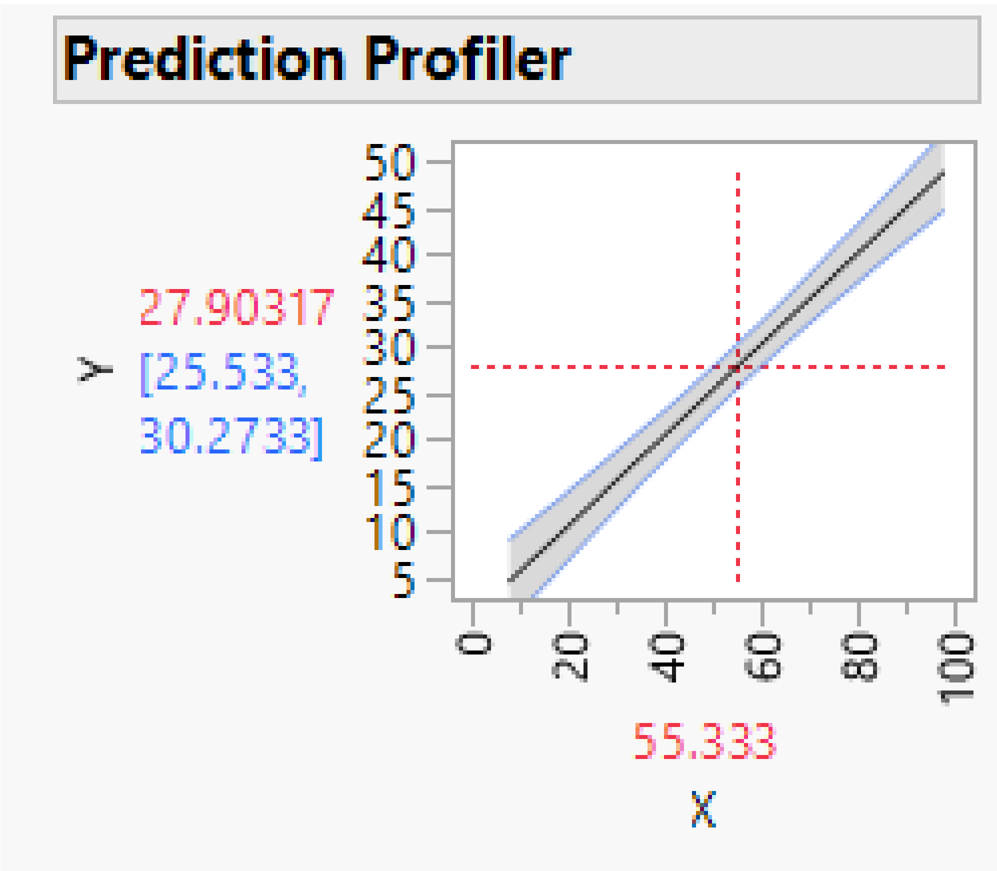
When we control X at 80, we don't just move the mean from 27.9 to 40 — we also reduce the standard deviation from 15.4 to 2.84 !



This slide intentionally left blank

4. Introduction to the Prediction Profiler

JMP's Prediction Profiler helps us use our regression model to make predictions and optimize our process.

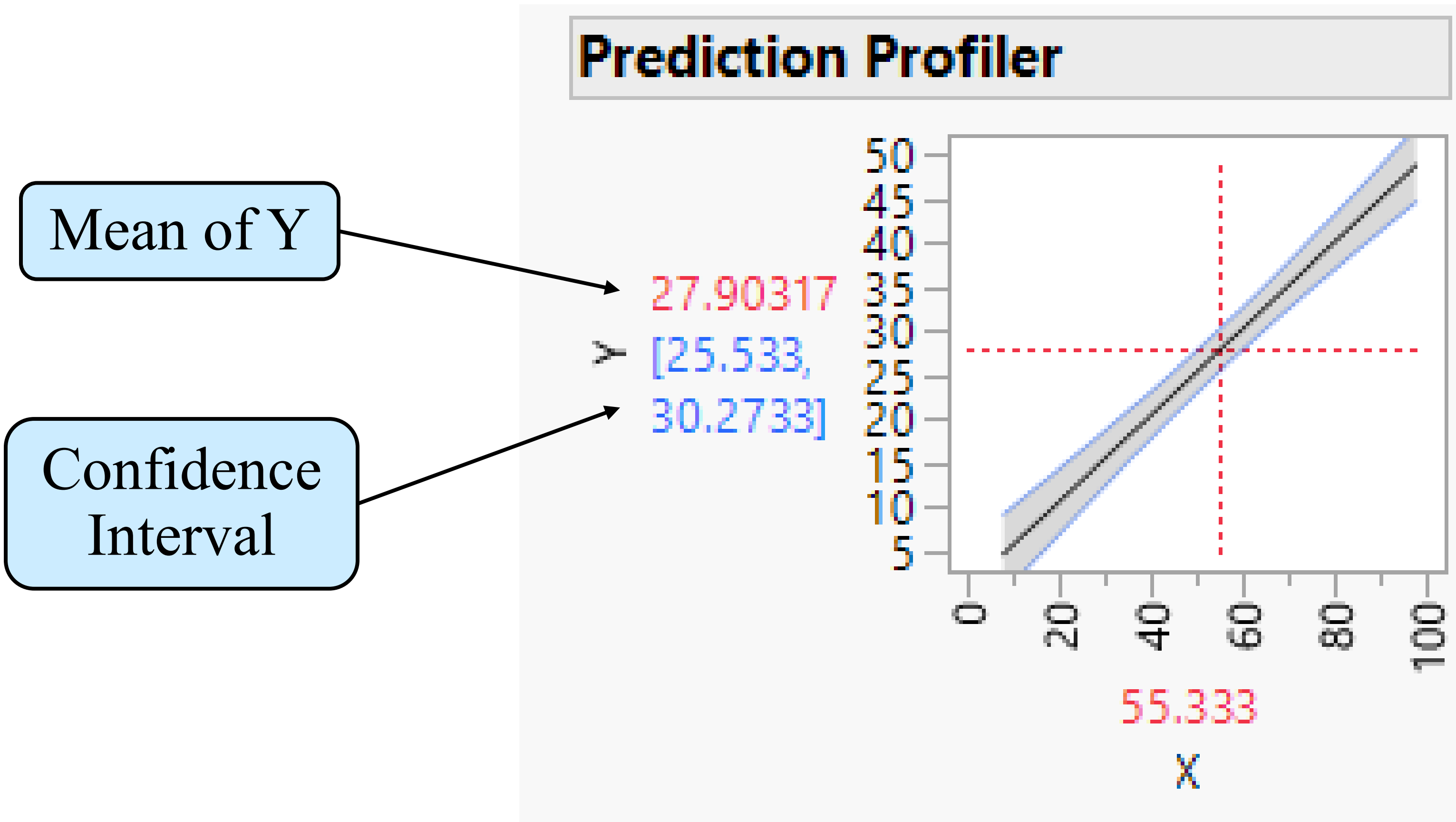


Follow these steps to access the prediction profiler:

- Analyze > Fit Model > Y = Y, Model Effects = X > Run > Red Triangle > Factor Profiling > Profiler

Introduction to the Prediction Profiler (cont'd)

JMP's Prediction Profiler helps us use our regression model to make predictions and optimize our process.

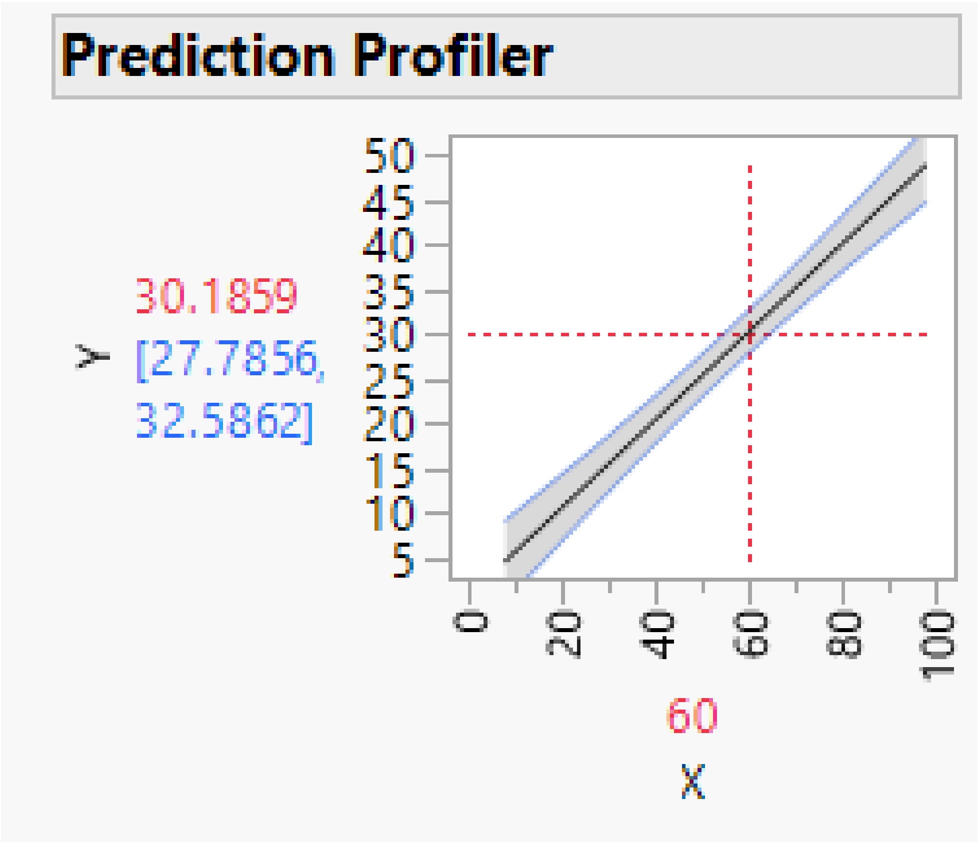


- Calculates predicted ***mean Y*** as a function of X
- Calculates **confidence intervals** for predicted **means**

Simple example of prediction of Mean Y

Continuing with the *simple regression-generic* data:

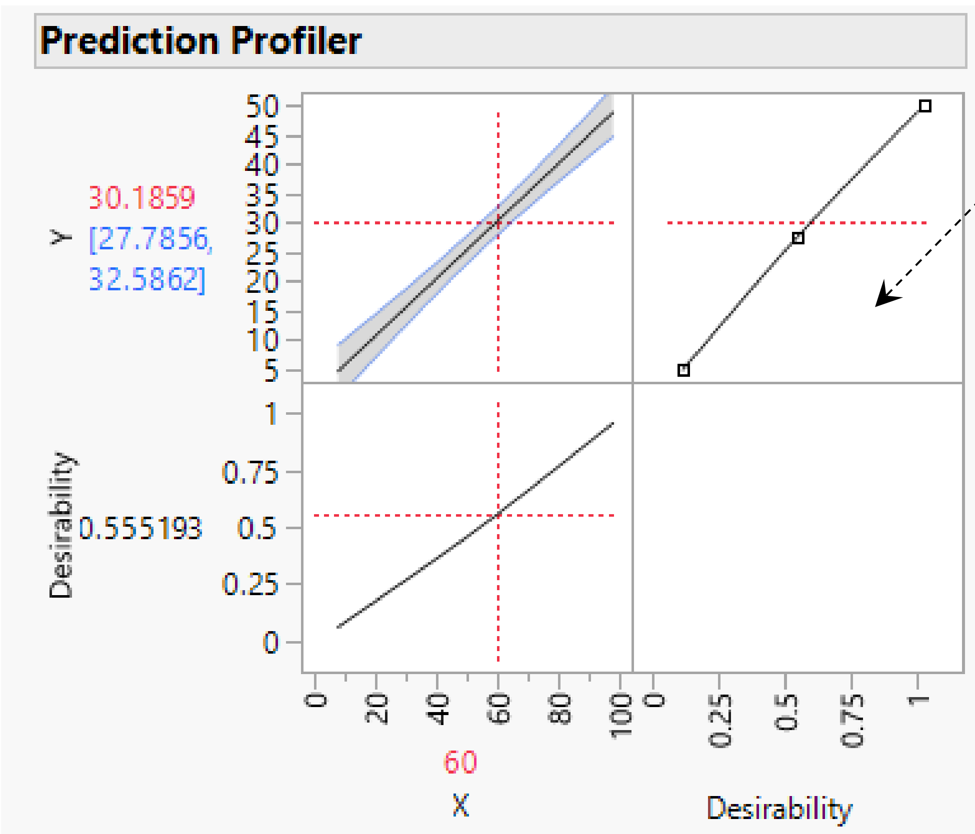
- Suppose we are interested in the predicted mean Y for X = 60
- Click on the 55.333, change it to 60



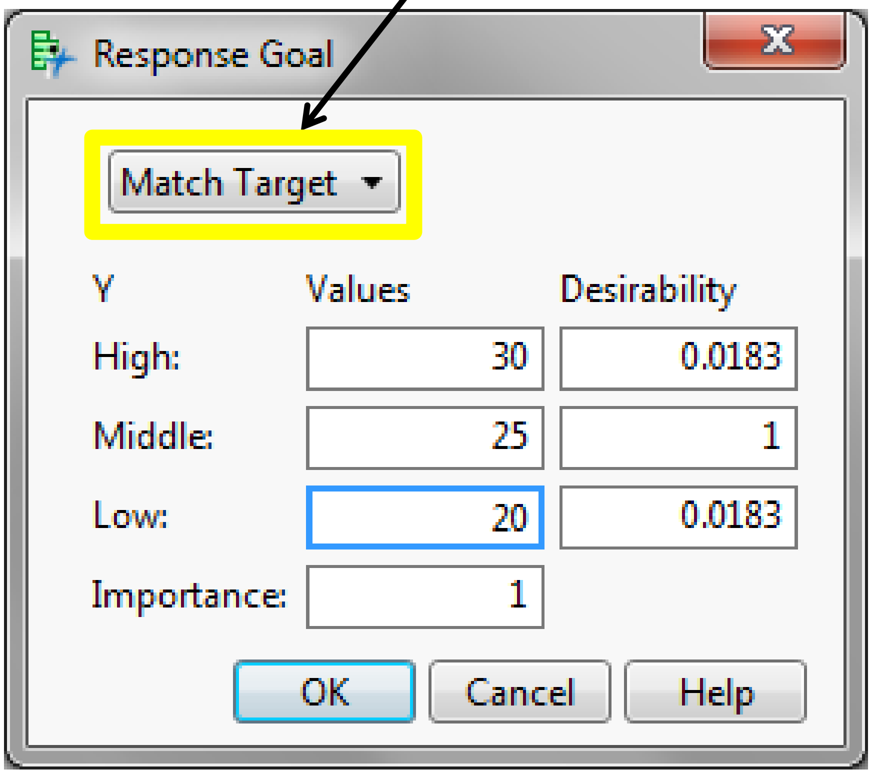
- Predicted mean Y (based on the data) is 30.19
- With 95% confidence, the population mean lies between 27.79 and 32.59

Simple example of optimization

- Suppose we want to find the X value that predicts a mean Y value of 25
- Red triangle next to *Prediction Profiler* → *Optimization and Desirability* → *Desirability Functions*

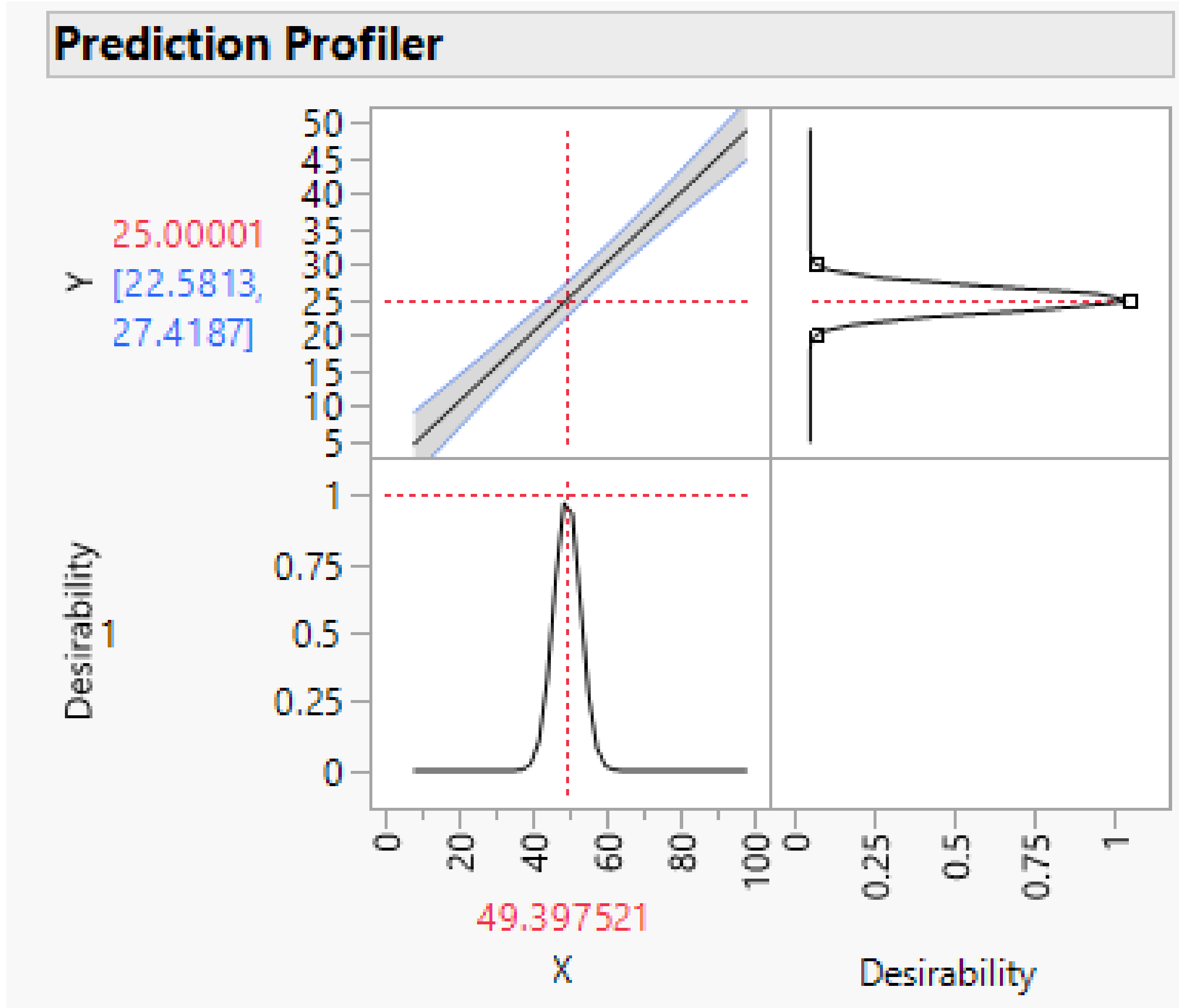


- Double click in here (don't touch the line plot)
- Modify the **Response Goal** dialog as shown below
- Click OK



Optimization (cont'd)

Red triangle
next to
Prediction Profiler
↓
Optimization and Desirability
↓
Maximize Desirability



- Predicted mean Y of 25 is achieved when X = 49.4
- With 95% confidence, this population mean lies between 22.6 and 27.4

Confidence Intervals and Prediction Intervals

- The ***95% Confidence Interval on the Mean Response*** gives the range which will contain the "true" mean, μ, 95% of the time
  - For a sample, the confidence interval is calculated:

$$\bar{Y} - t_{.025,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{Y} + t_{.025,n-1}$$

  - For a regression, calculation of the confidence interval is similarly structured, but considerably more complicated, involving matrix math.

- A ***95% Prediction Interval*** gives the range which will contain future individual response observations 95% of the time.
  - The prediction interval is wider than the confidence interval, because it is to contain individual measurements, not averages.
  - Calculation of this interval is complicated, involving matrix math.

Exercise 4.1

a) Continuing with *simple regression-generic,* find the X value that predicts a mean Y value of 35. Give the confidence limits for the predicted mean.

b) The overall standard deviation of Y is 15.39. The RMSE from the regression is 2.84. Which of these would be the standard deviation of Y if we controlled X to a constant value?

c) Save your script, close and save the data table.

Exercise 4.2

Data sets \ production vs capacity.

(a) Fit a regression for *Production qty* as a function of *Capacity utilized* (%) (using *Fit Model*, of course). Is there a correlation? Give the appropriate P-value and strength of evidence.

(b) For this exercise, we will not review the residuals plots. Use your model to find the capacity utilization level that predicts a mean daily production quantity of 3500. Give the confidence limits.

(c) The overall standard deviation of *Production qty* is 733.5 (not shown in Fit Model output—calculated in Distribution Platform). The RMSE from the analysis in (a) is 409.732. Which of these would be the standard deviation if capacity utilization was held constant?

(d) Save your scripts, close and save the data table.

Estimating Improved % Defective

Once we determine the level at which we want to control our x, we can use the root mean square error (RMSE) and other regression results to estimate the % defective in the improved process.

Remember that by definition, the RMSE is the standard deviation of the improved process, with x's held at desired levels.

The *t distribution calculator* helps us calculate the future % defective.

1. Enter the quantities in the YELLOW cells.

2. The other values are calculated for you.

LSL	20
USL	60
Mean	40
Standard deviation	3.006984
Degrees of freedom	7

	LSL	USL	Total
Population % out of spec	0.015	0.015	0.029
Population PPM out of spec	145.1	145.1	290.2

PPM defective = 290

These calculations can be sensitive to round-off error. Don't round off the mean and standard deviation when you enter them into the calculator.

Error DF from the Analysis of Variance

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1830.6557	1830.66	202.4624
Error	7	63.2937	9.04	Prob > F
C. Total	8	1893.9494		<.0001*

Exercise 4.3

Data sets \ production vs capacity.jmp.

In this process data, on 75% of the days production quantity fell below 3000.

Based on the best fit distribution, the Lognormal, the expected % of days that production quantity will fall below 3000 is 71.8%.

a) We found earlier that capacity utilization 52.1% gives a mean daily production quantity of 3500. The RMSE was 409.7, the error degrees of freedom was 34. Assuming 52.1% capacity utilization, use the *t distribution calculator* to find the predicted % of days on which production quantity will be less than 3000.

b) Save your scripts, close and save the data table.

Exercise 4.4

Open *Data sets \ outgassing process*. *Current* (the Y variable) is the current required to heat a filament to a target temperature. *Resist* (the X variable) is the electrical resistance of the filament. *Machine* is the processing unit. This example shows how to reduce % defective by separate optimization of each machine.

a) For this process, the % of *Current* data values that fall outside the interval (1.9, 2.1) is 8.87%.

b) Fit a regression for *Current* as a function of *Resist*, using *Machine* as the *By variable*. For each machine, give the RMSE, the error degrees of freedom, and the resistance that predicts a mean current of 2.

Machine	RMSE	DF	Resistance	% Outside
A				
B				
C				

c) Assuming we use the indicated resistance values, use the *t distribution calculator* to find for each machine the % of *Current* values predicted to fall outside the interval (1.9, 2.1).

d) Save your scripts, close and save the data table.

5 Multiple Regression

- Multiple regression model
- Examples
- Fitting regression models
- Interactive effects
- Predicted values and uncertainty
- Modeling and optimization

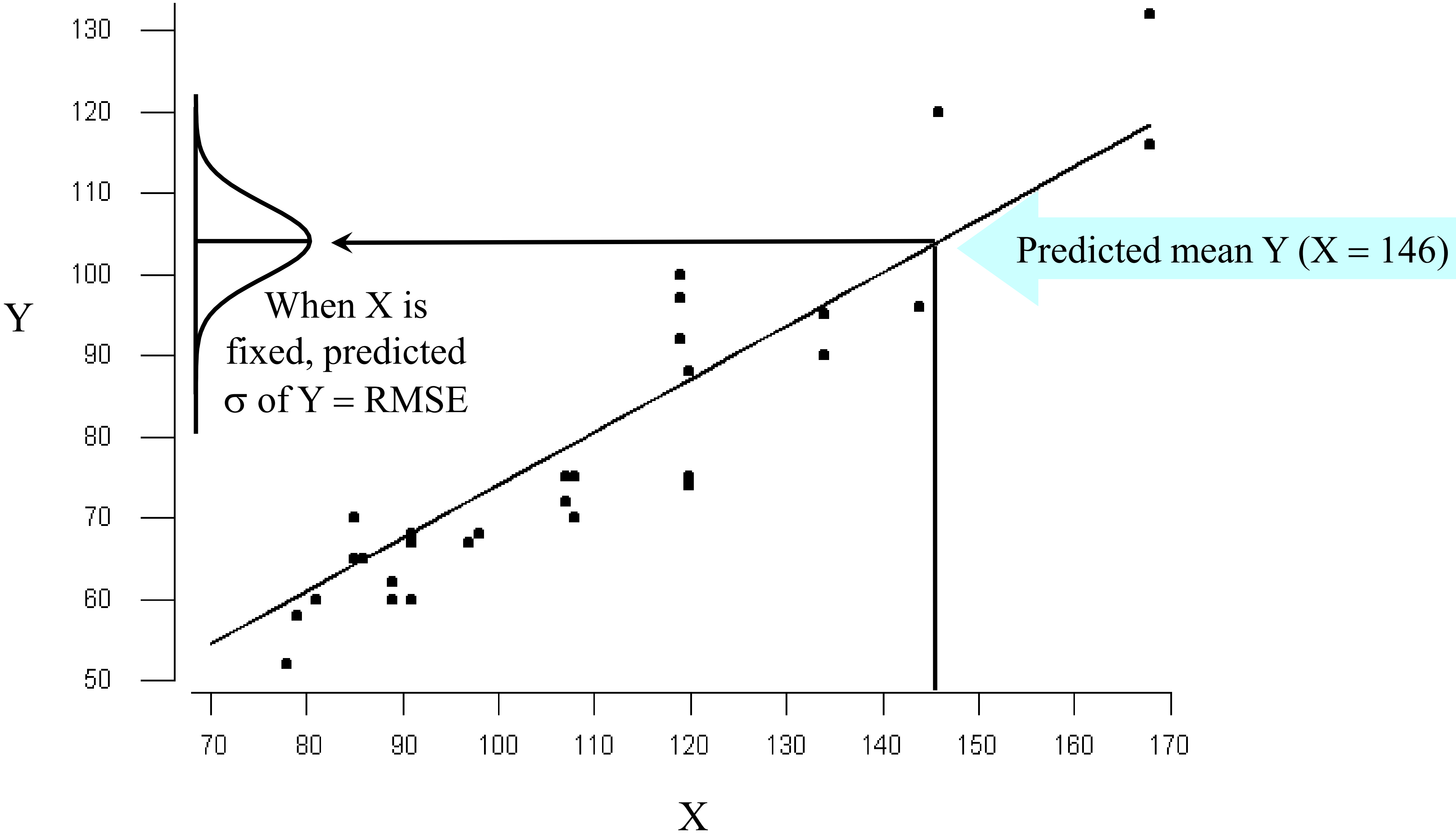
Multiple regression model

$$Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_kX_k + \text{“error”}$$

Y	$X_1, X_2, \ldots, X_k$	b_0	$b_1, b_2, \ldots, b_k$	“Error”
Dependent variable	Independent variables	Intercept	Regression coefficients	Residuals Mean = 0
Response variable	Explanatory variables	Parameter	Parameters	Standard deviation = σ (RMSE)
Output	Inputs			Distribution = Assumed to be Normal
	Predictors			
	Regressors			
	Factors (in DOE)			

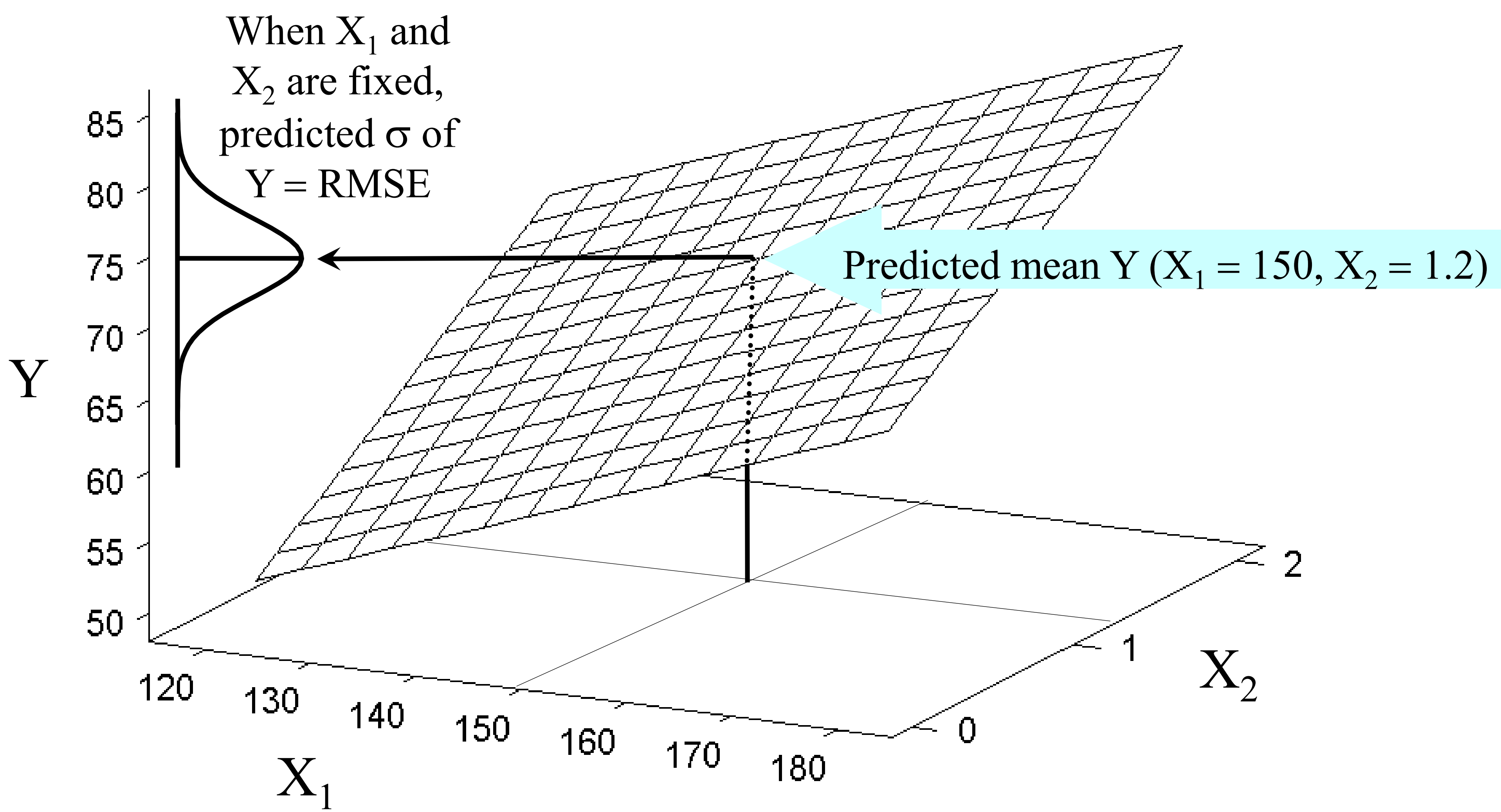
Model and error components, one X

$$Y = b_0 + b_1X + \text{“error”}$$



Model and error components, two Xs

$$Y \; = \; b_0 + b_1X_1 + b_2X_2 \; + \; \text{“error”}$$



Multiple regression examples

Y	X_1	X_2	X_3	X_4	X_5
Life of cutting tool	RPM	Tool type	Material	Feed rate	
MPG	Displace-ment	Horsepower	Weight		
Salary	Education	Experience	Performance	Seniority	Gender
Vending machine service time	Amount of product stocked	Distance from truck to machine			

Fill in examples of interest to you

Regression model equations

Y	X_1	X_2	X_3	X_4	X_5
MPG	Displacement (D)	Horsepower (H)	Weight (W)		

$$MPG = b_0 + b_1 D + b_2 H + b_3 W + \text{error}$$

Y	X_1	X_2	X_3	X_4	X_5
Bond strength	Temperature (T)	Dwell time (D)	$T \times D$	T^2	D^2

$$Bond = b_0 + b_1 T + b_2 D + b_3 TD + b_4 T^2 + b_5 D^2 + \text{error}$$

↑

Response surface model (RSM) with two continuous Xs.

TD is the interaction term for T and D, T^2 and D^2 show curvature.

Linearizing nonlinear models

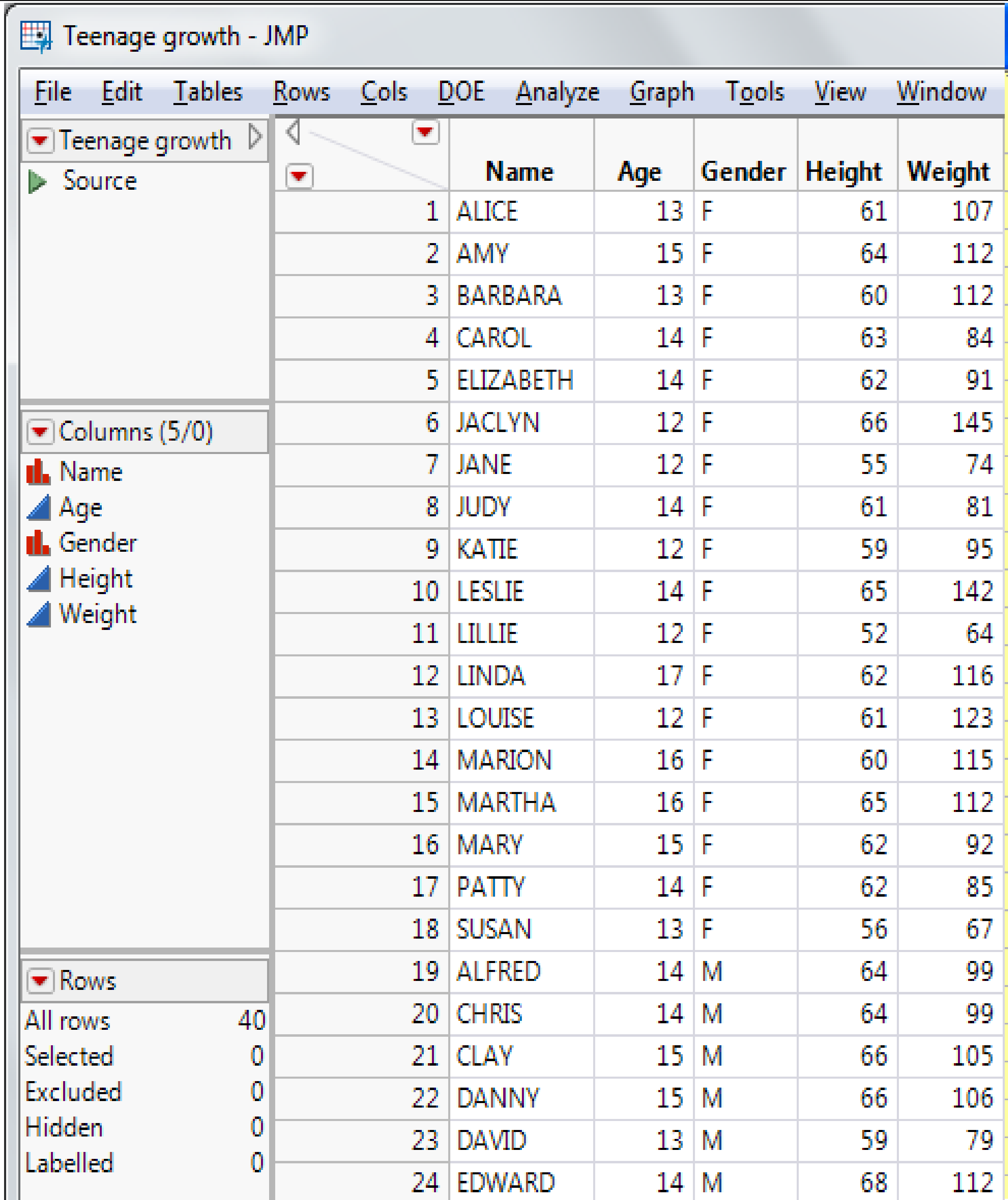
Nonlinear model	Equivalent linear model
$Y = b_0 (X_1)^{b_1} (X_2)^{b_2}$	$\log(Y) = \log(b_0) + b_1 \log(X_1) + b_2 \log(X_2)$
$Y = b_0 (b_1)^{X_1} (b_2)^{X_2}$	$\log(Y) = \log(b_0) + \log(b_1) X_1 + \log(b_2) X_2$

- In many cases, log(Y) transformations can successfully linearize nonlinear regression models
- This greatly extends the application of standard multiple regression models

Fitting regression models

Data sets \ teenage growth

Y	X_1	X_2
Height	Age	Gender
Weight	Age	Gender

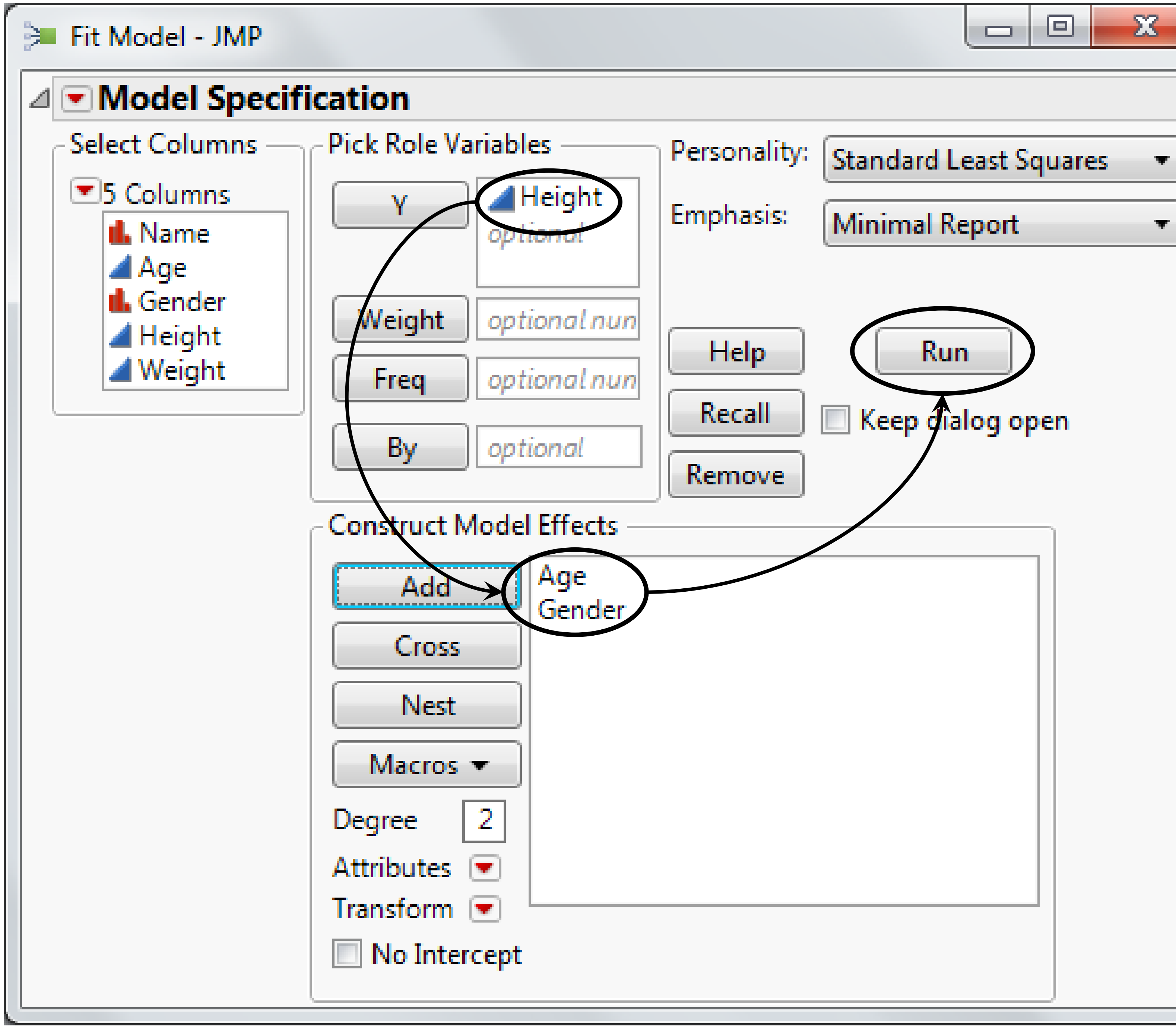


	Name	Age	Gender	Height	Weight
1	ALICE	13	F	61	107
2	AMY	15	F	64	112
3	BARBARA	13	F	60	112
4	CAROL	14	F	63	84
5	ELIZABETH	14	F	62	91
6	JACLYN	12	F	66	145
7	JANE	12	F	55	74
8	JUDY	14	F	61	81
9	KATIE	12	F	59	95
10	LESLIE	14	F	65	142
11	LILLIE	12	F	52	64
12	LINDA	17	F	62	116
13	LOUISE	12	F	61	123
14	MARION	16	F	60	115
15	MARTHA	16	F	65	112
16	MARY	15	F	62	92
17	PATTY	14	F	62	85
18	SUSAN	13	F	56	67
19	ALFRED	14	M	64	99
20	CHRIS	14	M	64	99
21	CLAY	15	M	66	105
22	DANNY	15	M	66	106
23	DAVID	13	M	59	79
24	EDWARD	14	M	68	112

Fitting models (cont'd)

Say we want to model *Height* as a function of *Age* and *Gender*

Analyze
↓
Fit Model



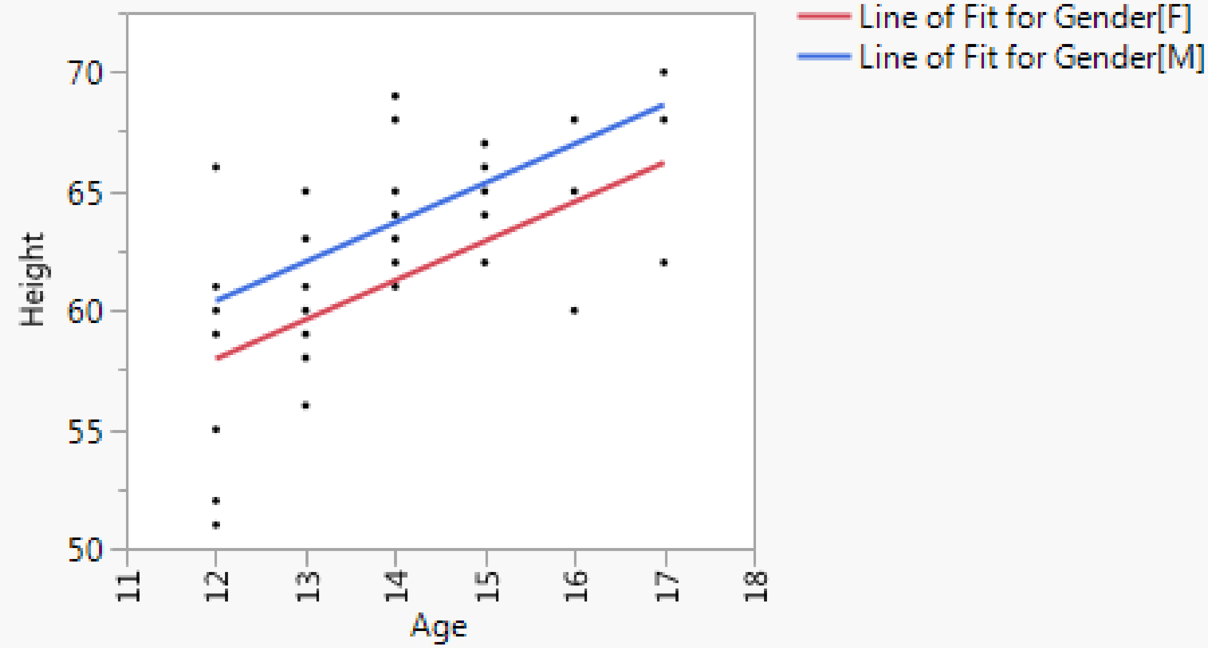
How to change options (for *Fit Model*) during analysis

- Alt-click on *Response Height* red triangle (This technique works for may JMP platforms)

- Set up as shown on next slide

Response Height

Regression Plot



Residual by Predicted Plot

Summary of Fit

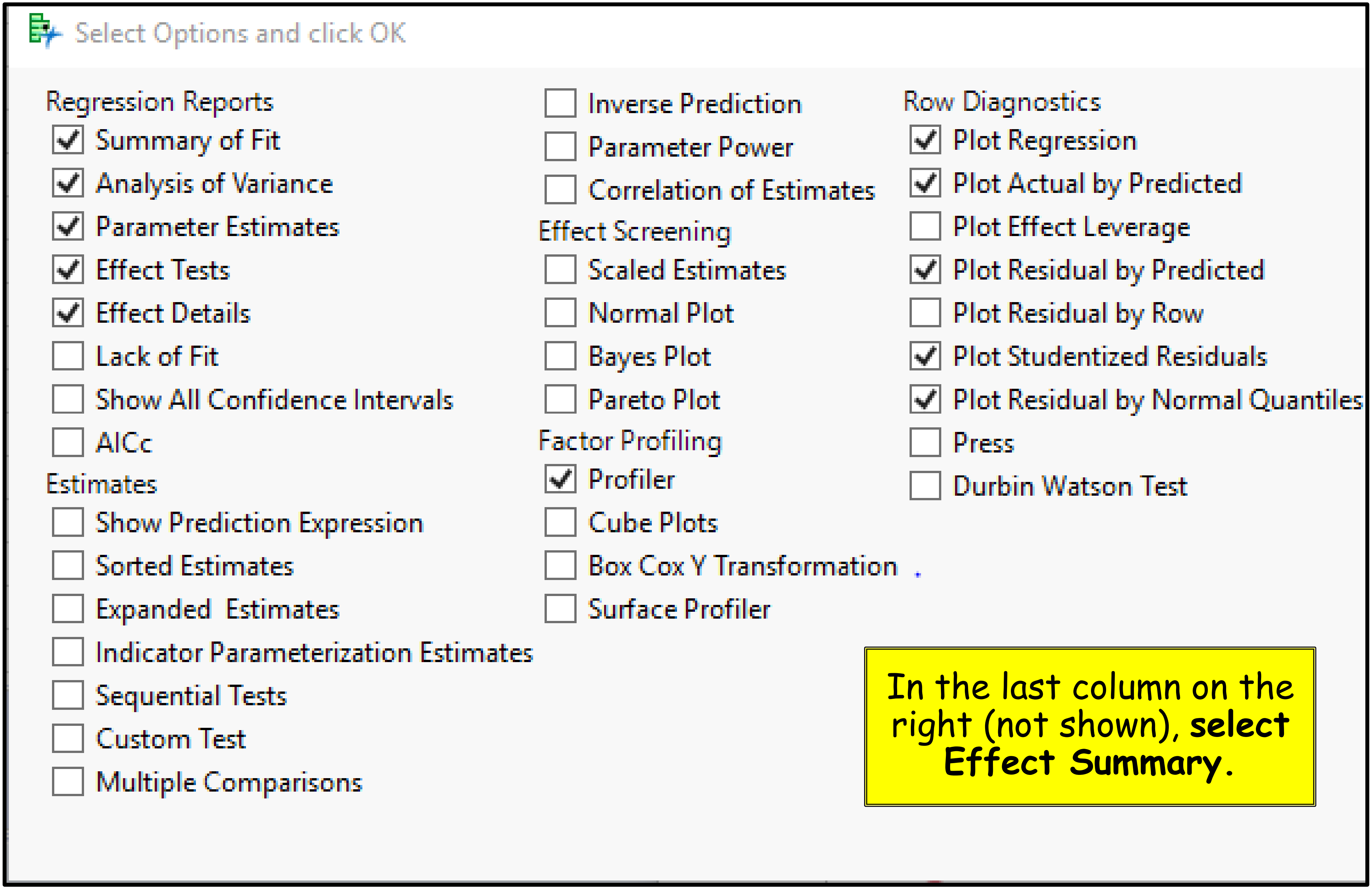
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	317.25956	158.630	15.2592
Error	37	384.64044	10.396	Prob > F
C. Total	39	701.90000		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	39.416521	4.945343	7.97	<.0001*
Age	1.6466542	0.352418	4.67	<.0001*
Gender[F]	-1.214868	0.516246	-2.35	0.0240*

Default options for *Fit Model* (cont'd)



In the last column on the right (not shown), **select Effect Summary**.

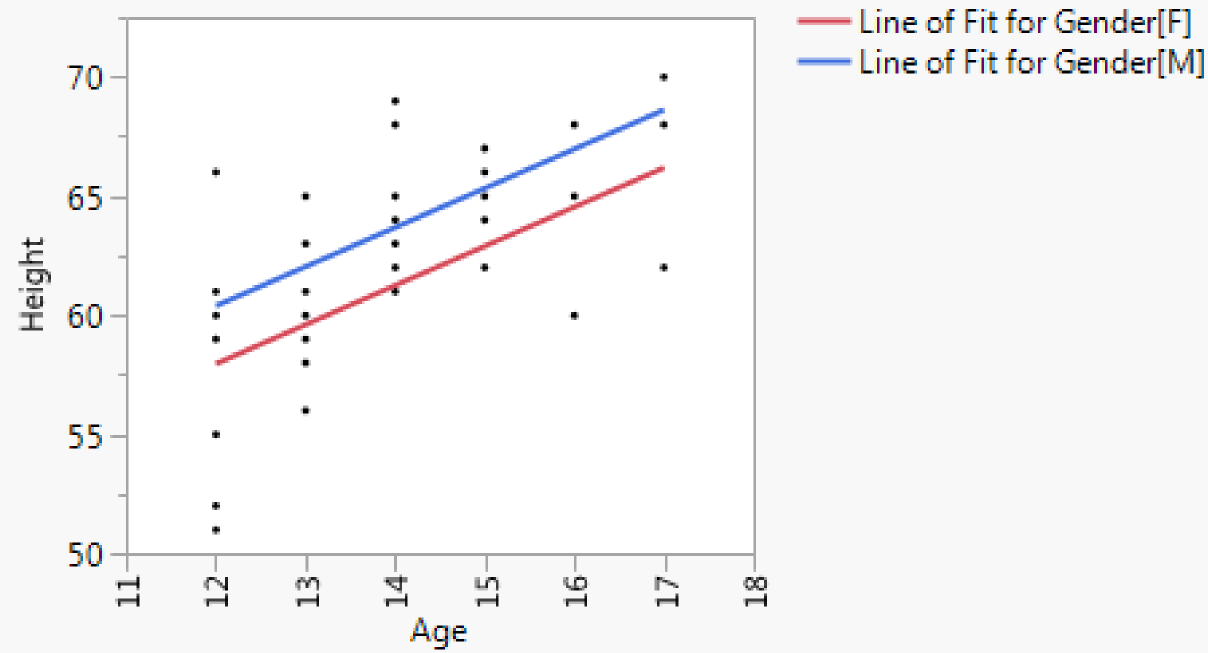
Handling categorical X variables in the model

"Indicator" or "dummy" variables are used to represent categorical variables in regression.

Indicator variable representing the effect of *Gender* in the equation

Response Height

Regression Plot



Residual by Predicted Plot

Summary of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	317.25956	158.630	15.2592
Error	37	384.64044	10.396	**Prob > F**
C. Total	39	701.90000		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	39.416521	4.945343	7.97	<.0001*
Age	1.6466542	0.352418	4.67	<.0001*
Gender[F]	-1.214868	0.516246	-2.35	0.0240*

Numeric coding for two-level categorical X

In JMP, two-level categorical factors are coded +1 and -1

$$\text{Gender[F]} = \begin{cases} +1 & \text{if Gender is F} \\ -1 & \text{if Gender is M} \end{cases}$$

$$\text{Height} = b_0 + b_1\text{Age} + b_2\text{Gender[F]}$$

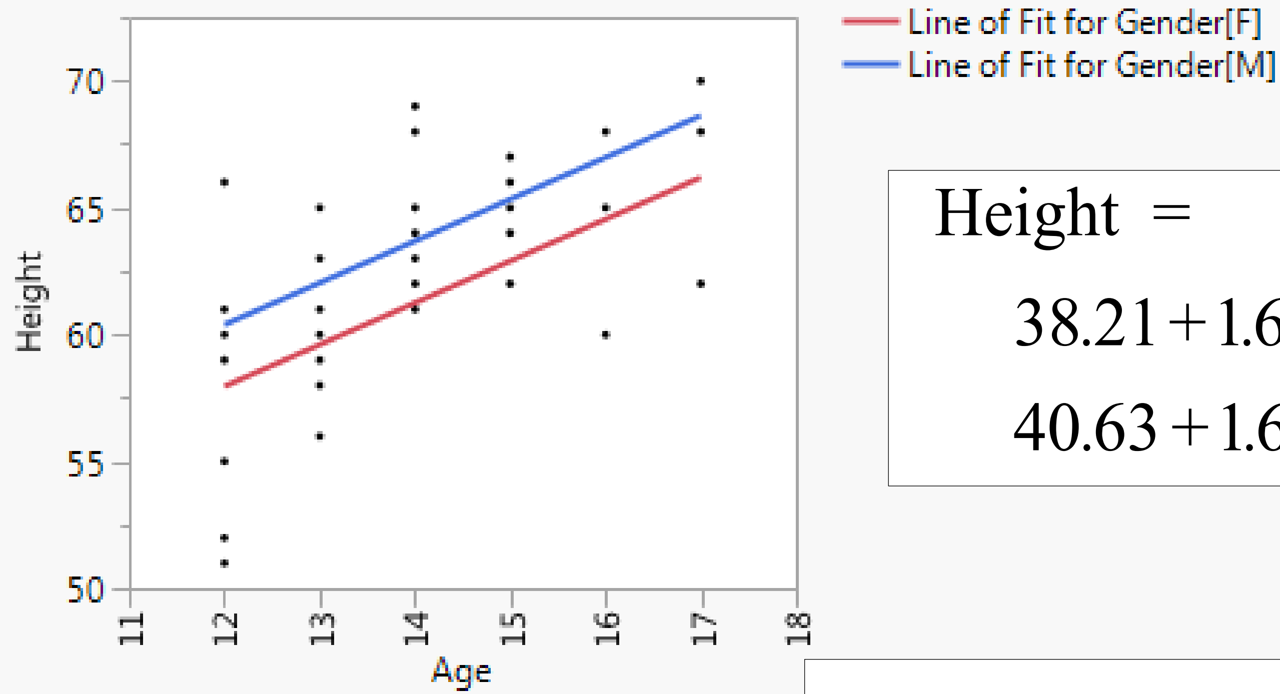
$$= \begin{cases} b_0 + b_2 + b_1\text{Age} & \text{if Gender is F} \\ b_0 - b_2 + b_1\text{Age} & \text{if Gender is M} \end{cases}$$

This results in one equation for Females and one equation for Males, with equal slopes (b_1) and different intercepts ($b_0 + b_2$ and $b_0 - b_2$).

An additional indicator variable is added for each additional level of a categorical variable.

Constructing the model equation

Regression Plot



$$\text{Height} = \begin{cases} 38.21 + 1.65\ \text{Age} & \text{if Gender} = \text{F} \\ 40.63 + 1.65\ \text{Age} & \text{if Gender} = \text{M} \end{cases}$$

$$\text{Height} = 39.42 + 1.65\ \text{Age} - 1.21\ \text{Gender[F]}$$

Residual by Predicted Plot

Summary of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	317.25956	158.630	15.2592
Error	37	384.64044	10.396	Prob > F
C. Total	39	701.90000		<.0001*

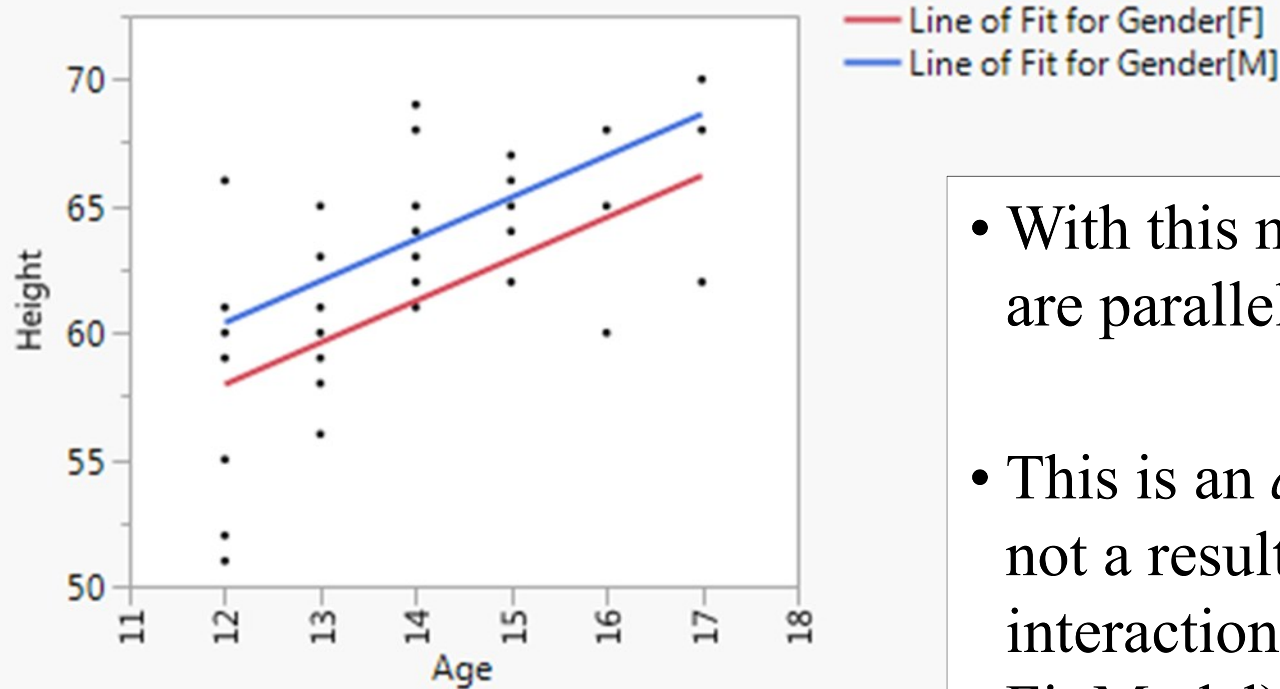
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	39.416521	4.945343	7.97	<.0001*
Age	1.6466542	0.352418	4.67	<.0001*
Gender[F]	-1.214868	0.516246	-2.35	0.0240*

If you want to verify the equation:
▼ Response Y → Estimates
→ Show Prediction Expression

The need for interaction effects

Regression Plot



- With this model, the growth curves are parallel
- This is an *assumption* of the model, not a result of the analysis (no interaction terms were included in Fit Model)
- How do we *test* for parallel curves?

Residual by Predicted Plot

Summary of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	317.25956	158.630	15.2592
Error	37	384.64044	10.396	Prob > F
C. Total	39	701.90000		<.0001*

Parameter Estimates

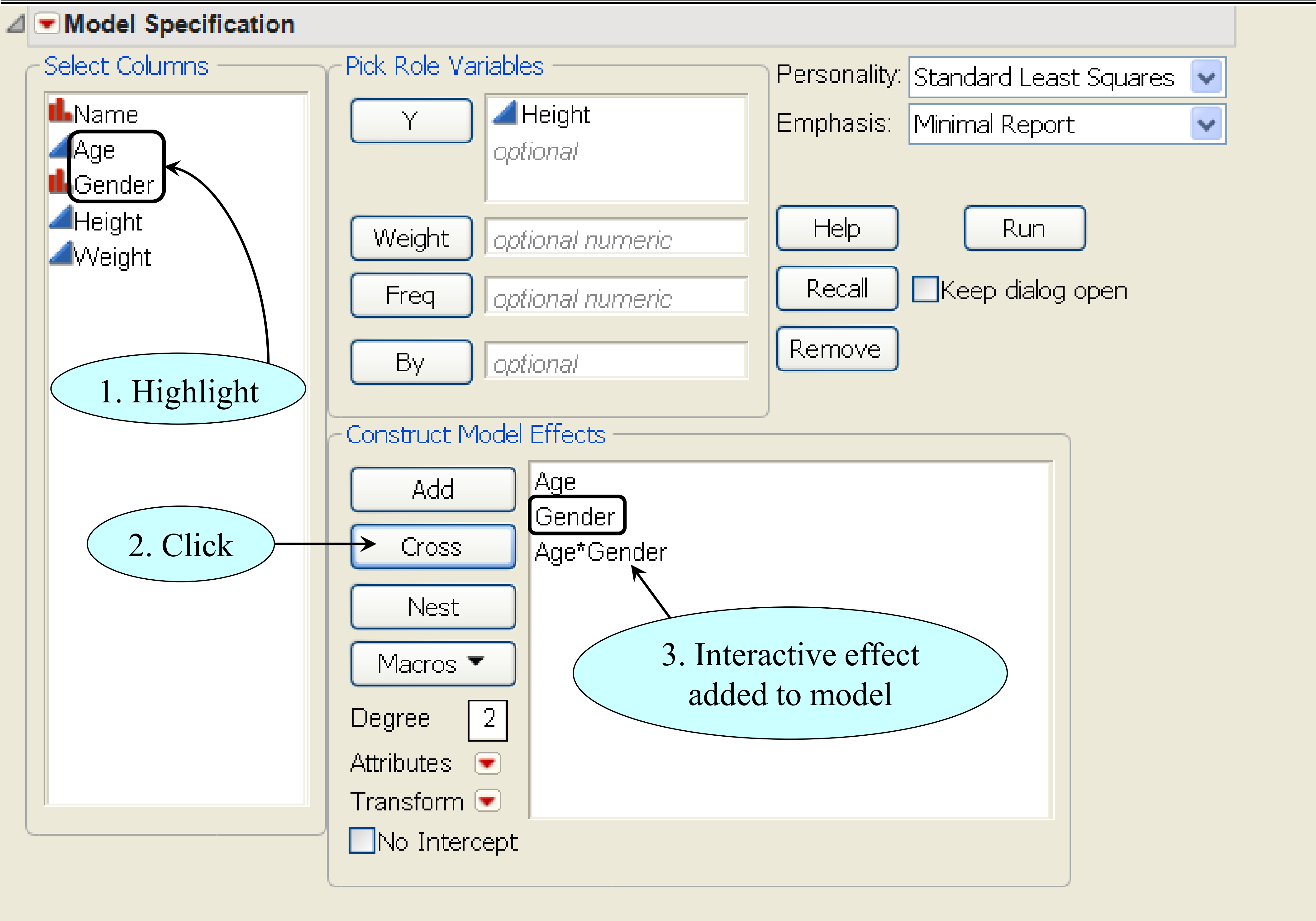
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	39.416521	4.945343	7.97	<.0001*
Age	1.6466542	0.352418	4.67	<.0001*
Gender[F]	-1.214868	0.516246	-2.35	0.0240*

Interaction effects (cont'd)

$$\text{Height} = b_0 + b_1\text{Age} + b_2\text{Gender}[F] + b_3\text{Age} * \text{Gender}[F]$$

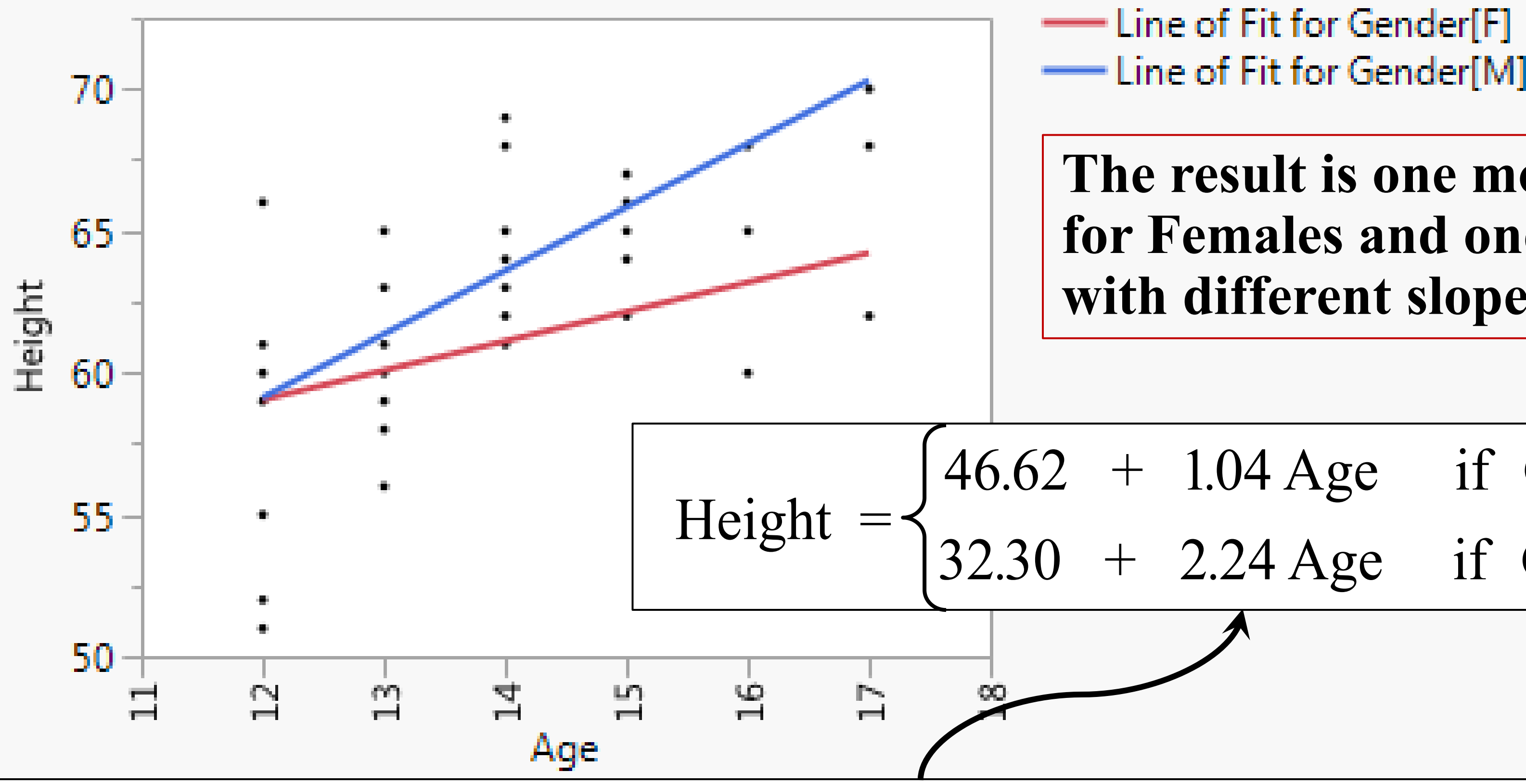
This product term allows different slopes for M and F

Adding an interaction effect



Non-parallel growth curves

Regression Plot



The result is one model equation for Females and one for Males, with different slopes and intercepts

$$\text{Height} = \begin{cases} 46.62 + 1.04\ \text{Age} & \text{if Gender} = \text{F} \\ 32.30 + 2.24\ \text{Age} & \text{if Gender} = \text{M} \end{cases}$$

$$\text{Height} = 39.46 + 1.64\ \text{Age} - 1.23\text{Gender}[\text{F}] - 0.60\text{Gender}[\text{F}] * (\text{Age} - 13.98)$$

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	39.457057	4.812681	8.20	<.0001*
Age	1.6360307	0.343014	4.77	<.0001*
Gender[F]	-1.227546	0.502444	-2.44	0.0196*
Gender[F]*(Age-13.975)	-0.600896	0.343014	-1.75	0.0883

To verify the equation:
- Response Y
  - → Estimates
  - → Show Prediction Expression

Testing the interaction effect

Response Height

Actual by Predicted Plot

Regression Plot

Effect Summary

Source	LogWorth	PValue
Age	4.518	0.00003
Gender	1.708	0.01959
Age*Gender	1.054	0.08832

Remove Add Edit ☐ FDR

Residual by Predicted Plot

Studentized Residuals

Summary of Fit

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	39.457057	4.812681	8.20	<.0001*
Gender[F]	-1.227546	0.502444	-2.44	0.0196*
Age	1.6360307	0.343014	4.77	<.0001*
(Age-13.975)*Gender[F]	-0.600896	0.343014	-1.75	0.0883

The p-value for Gender*Age indicates some evidence that growth curves for girls and boys have different slopes

- From now on we will use *Effect Summary* to find P-values. It gives the same information and allows model modification.

Summary of Fit without Interaction

RSquare	0.452001
RSquare Adj	0.42238
Root Mean Square Error	3.224234

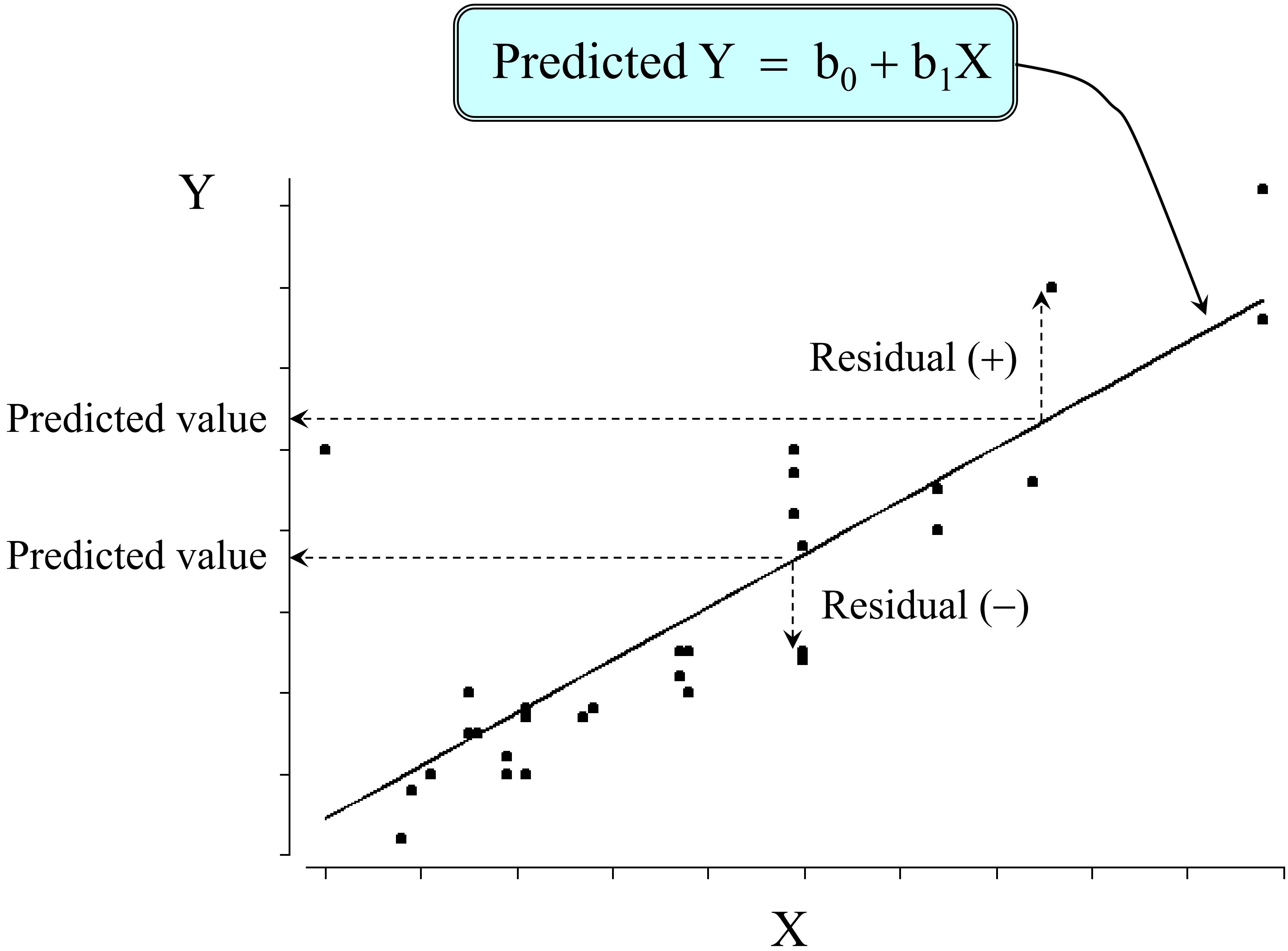
✓ Adjusted R^2 went up

✓ RMSE went down

Summary of Fit with Interaction

RSquare	0.495046
RSquare Adj	0.452967
Root Mean Square Error	3.137706

Residuals Review



A fitted model, the equation generated during regression, gives the predicted mean value of the response variable as a function of the predictor variables. These predicted mean values are also called *predicted values*, or just *predicted* for short. The *residual value* is the data (observation) value minus the predicted value. Residual values are called *residuals* for short.

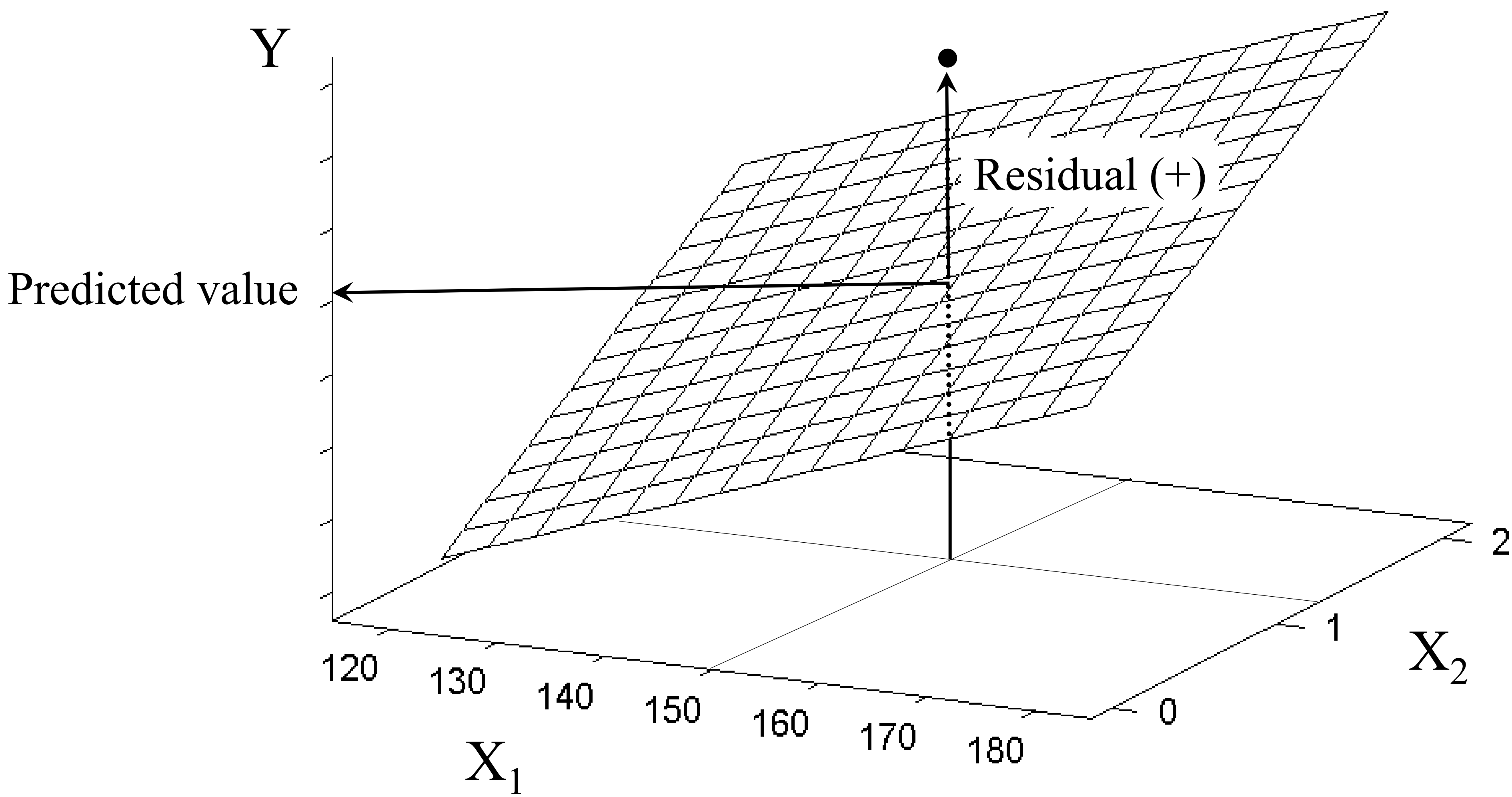
These terms are easiest to visualize in the simple linear model shown above. A predicted value is the fitted line evaluated at some X value. A residual is the difference between a measured (observed) Y value and the predicted value at the corresponding X.

Residuals contain information about the magnitude and direction of variability in the data relative to the fitted model.

- An unusually large residual might signal a measurement error, data entry error or some other type of outlier.
- A systematic trend or pattern in the residuals might signal an inadequacy in the fitted model.

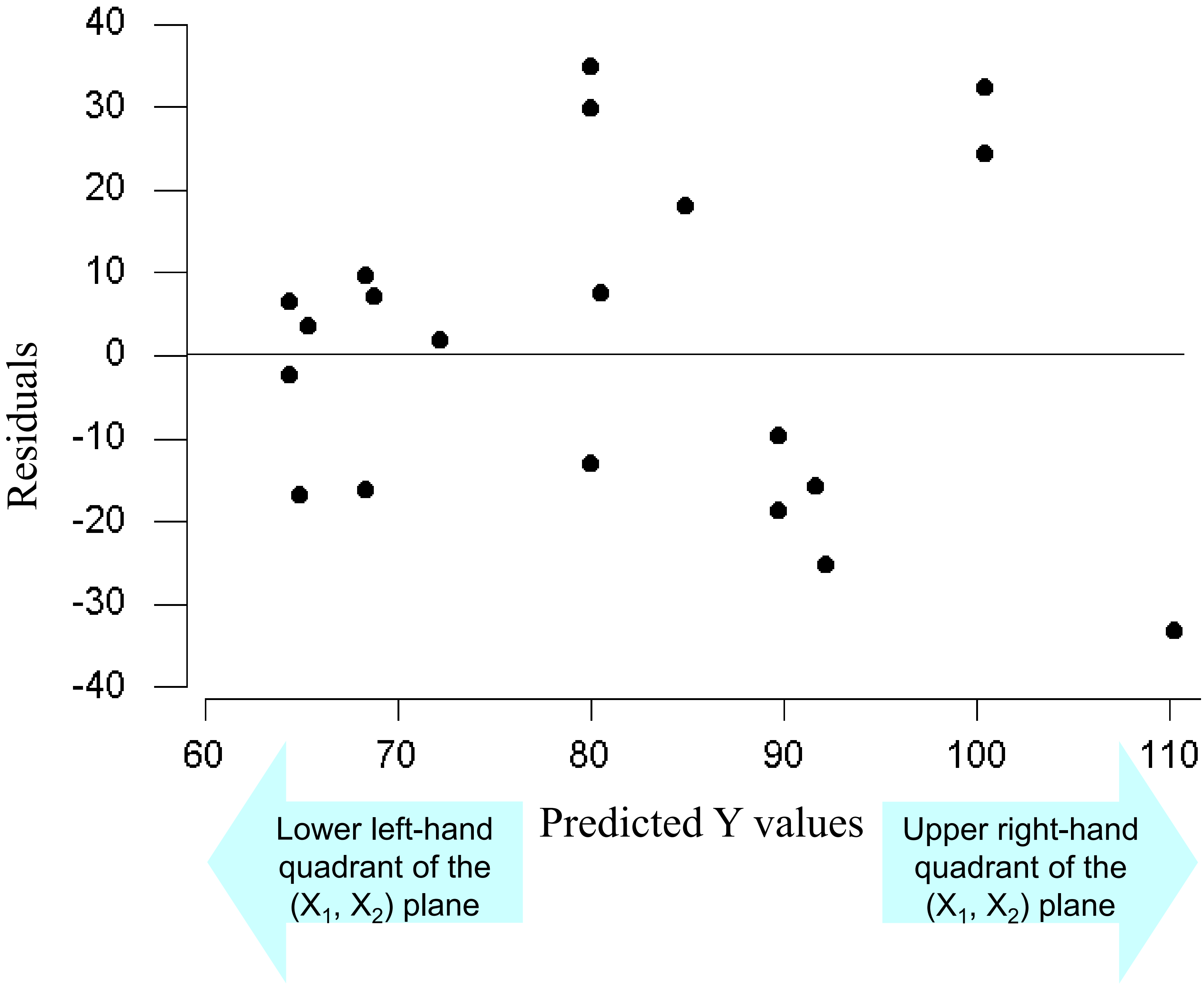
Residuals Review: Same thing for any number of X's

$$\text{Predicted Y} = b_0 + b_1X_1 + b_2X_2$$

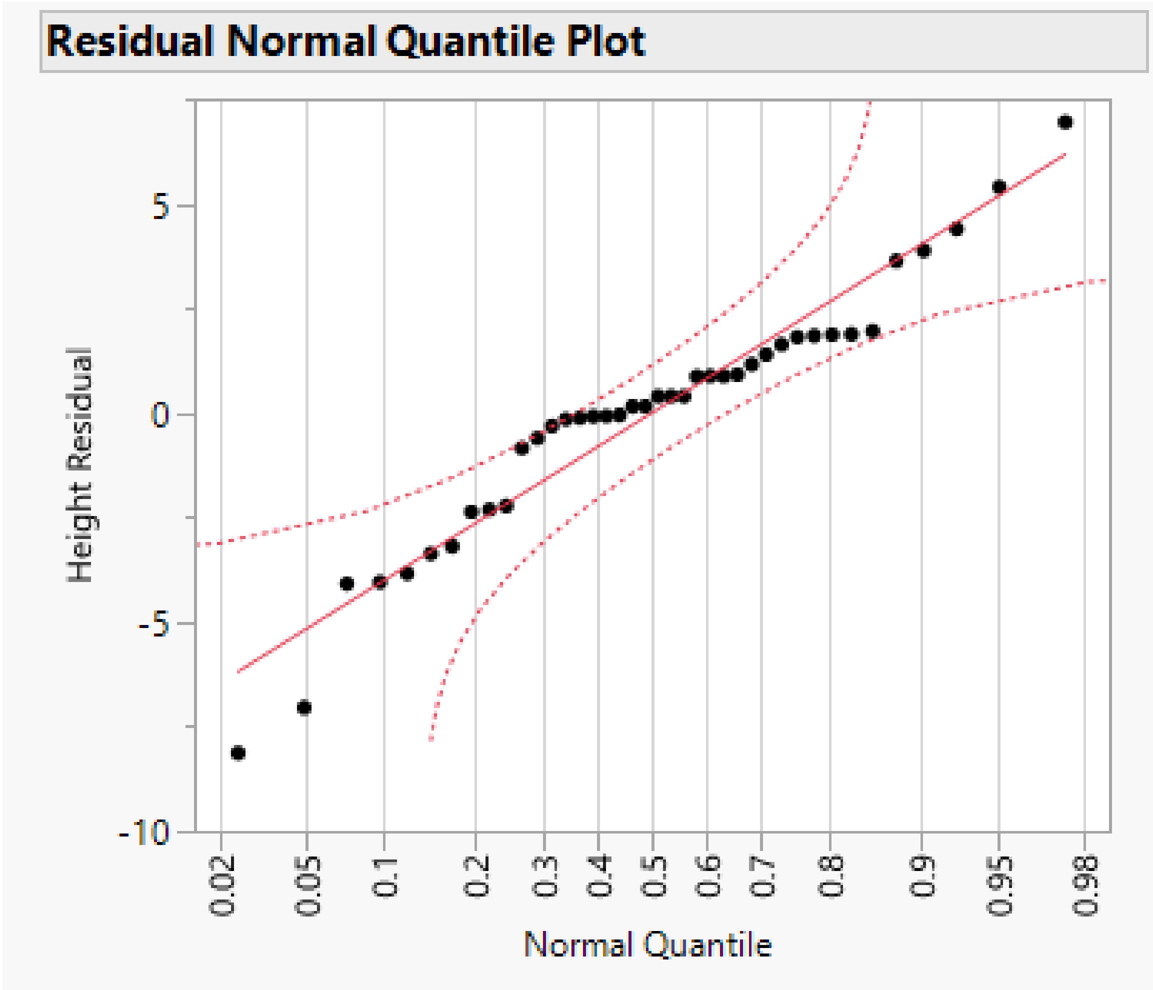


Same thing (cont'd)

Plot of residuals by predicted for any number of Xs

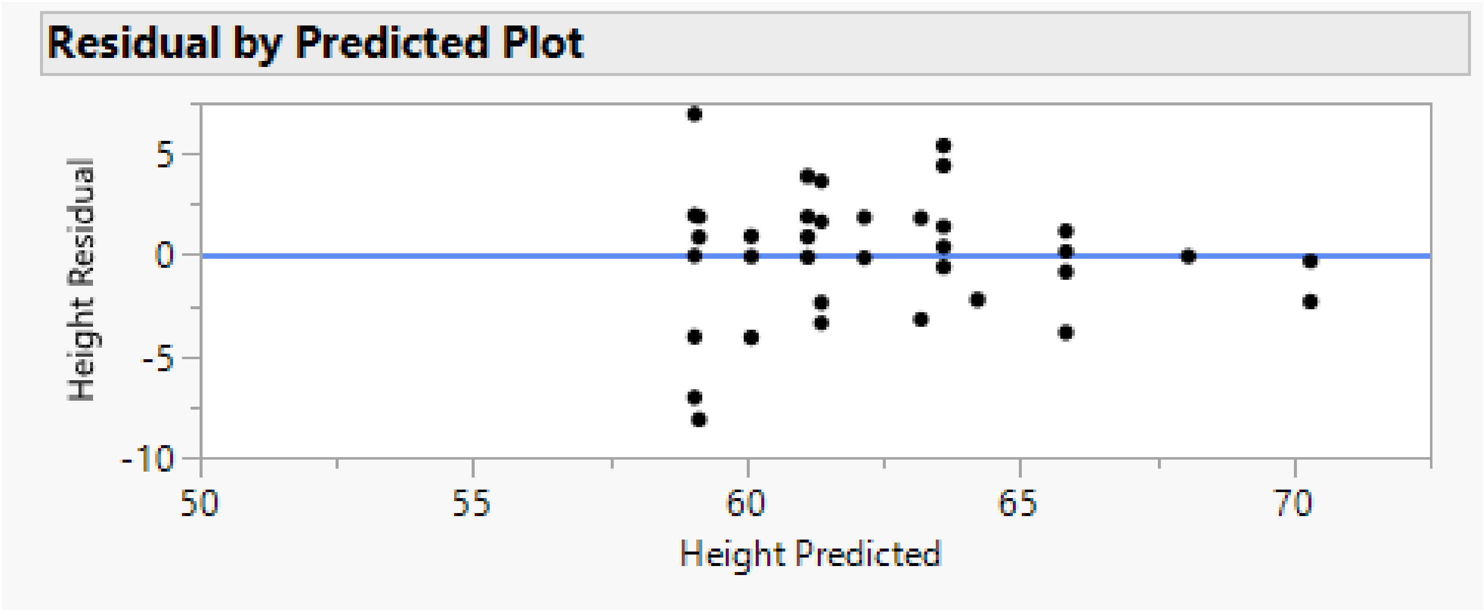


Checking model adequacy



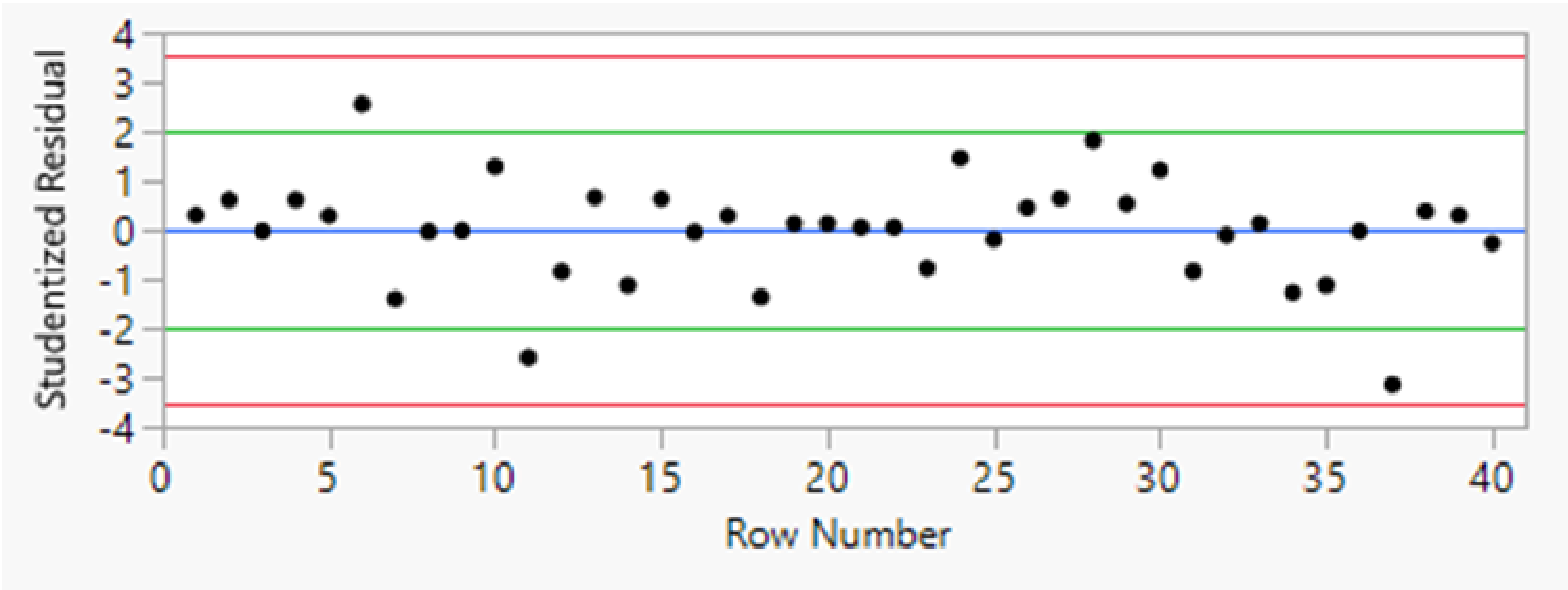
We can see points on the hyperbolic bands here, but there is not an obvious curve through the data. Given the small sample size, this is not too concerning.

Checking model adequacy (cont'd)



In this plot, we can see that the variance in the residuals is decreasing as height increases. This indicates the need for a transformation. We will see how to do this a little later in the course.

Checking model adequacy (cont'd)



There are no obvious patterns in residuals in run order, and they scatter about zero. There is no concern here.

(Points outside the red limits are considered outliers, and should be investigated. Points outside the green limits but inside the red limits are possibly outliers, but with less certainty.)

Variance Inflation Factor (VIF)

When historical or observational data is used to generate a regression model, an additional test is needed:

- The variance inflation factor (VIF) must be checked
- The VIF indicates whether the regressors (i.e. Xs or predictors) are correlated with each other
  - $VIF = 1$: regressor is independent of all other regressors
  - $1 \geq VIF \geq 5$: regressor is moderately correlated to other regressors
  - $VIF > 5$: regressor is highly correlated with other regressors
- VIFs in the final model need to be less than 5
  - When X variables are correlated (high VIFs), the analysis makes statistical determinations based on the noise between the correlated variables. This will often result in high R^2 values but insignificant p values.
  - VIFs are often lowered when insignificant terms are removed from the model, and terms should be removed one at a time. The first term removed should be the one with the highest p value unless theory implies removing a different one.
  - High VIFs are not an issue in designed experiments, as the designs prevent high correlation between terms/regressors

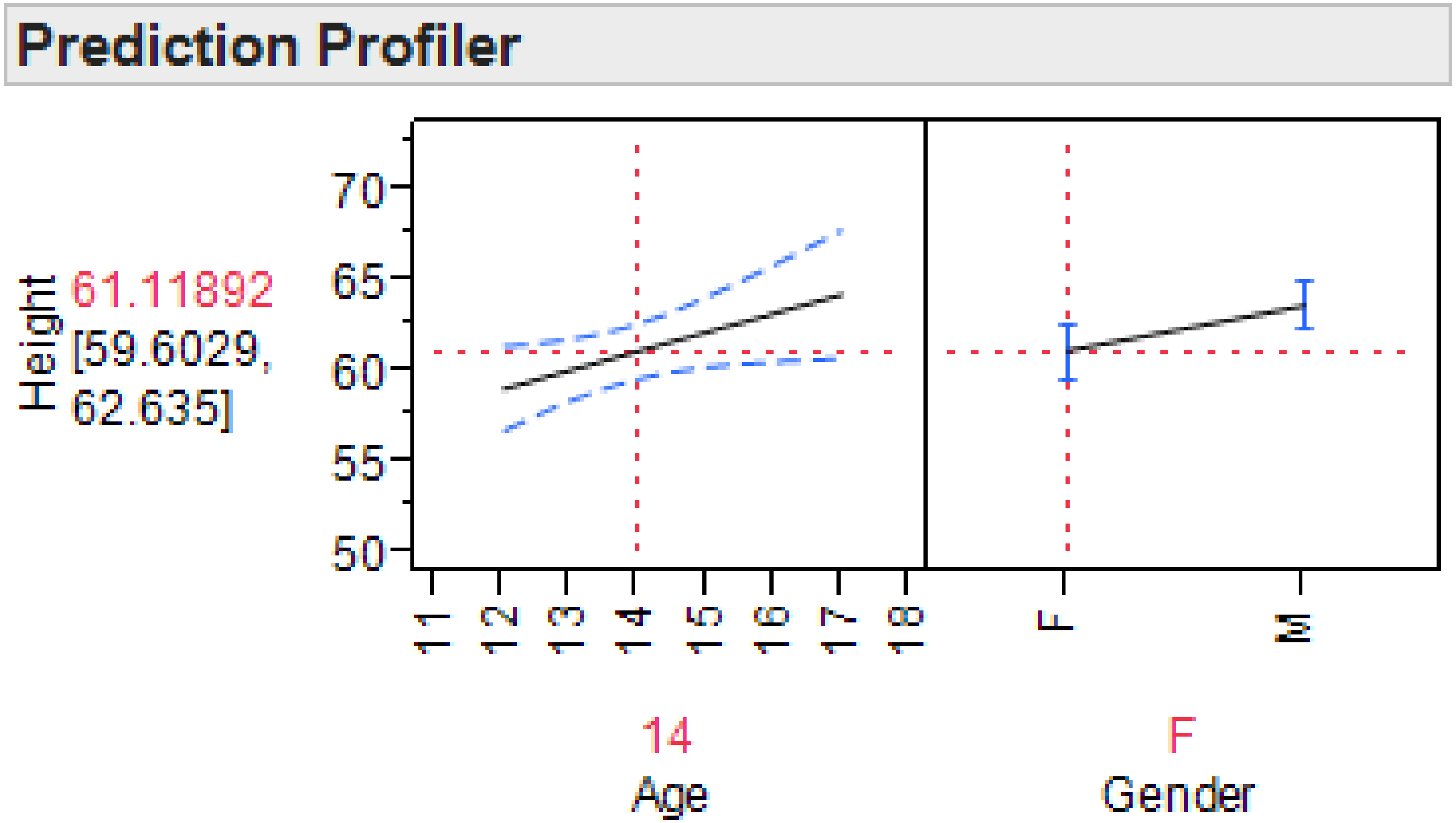
VIFs (cont'd)

Parameter Estimates					
Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**	**VIF**
Intercept	39.457057	4.812681	8.20	<.0001*	.
Gender[F]	-1.227546	0.502444	-2.44	0.0196*	1.0154192
Age	1.6360307	0.343014	4.77	<.0001*	1.0155259
(Age-13.975)*Gender[F]	-0.600896	0.343014	-1.75	0.0883	1.0004648

The variance inflation factors for all terms in the model are below 5. There is no concerning level of correlation between model terms.

To display the VIFs, right click in the Parameter Estimates section, click Columns, then VIF.

Predicted values and associated uncertainty



Predicted avg. height in the population of 14 year old girls	61.12
95% confidence interval for avg. height of 14 year old girls	[59.60, 62.64] 61.12 ± 1.52

The model without interaction gave 61.25 ± 1.55 (slightly larger margin of error).

Steps in Multiple Regression (backward elimination method)

1. Run Analyze > Fit Model in JMP to investigate the relationship between y and x's. Use the Response Surface Model (all factors, all interactions, quadratic terms for continuous variables/factors)
2. Check model adequacy by reviewing the residuals plots:
   - Residual Normal Quantile Plot
   - Residual by Predicted Plot
   - Studentized Residuals (in run order)
3. Transform the data and resolve other issues, if needed.
4. Verify all VIFs < 5. Address the issue if any are over 5.
5. Remove insignificant terms from the model, that are not needed to maintain model hierarchy (main effects must be included if a higher order term of that variable remains in the model).
6. Use $Adjusted\ R^2$ to determine the amount of variation in Y that is explained by the model.

Notes

Your instructor will go through Exercise 5.4 as an example.

Exercise 5.1

a) In the table below, record the Adjusted R^2 and RMSE from the analysis of *Height* in this section. Also, record the P-values from *Effects Tests*. Run the same analysis for *Weight* and record the corresponding results.

			P-values		
Response	**Adj. R^2**	**RMSE**	Age	Gender	Age*Gender
Height					
Weight					

b) Which variable (*Height* or *Weight*) has the greater proportion of variation explained by *Age* and *Gender*?

b) Explain why it wouldn't make sense to compare the two models in terms of RMSE.

Exercise 5.1 (cont'd)

d) Both *Age* and *Gender* were statistically significant for predicting *Height*. Is this true for *Weight*?

e) For *Height* we found evidence that the growth curves for girls and boys have different slopes. Is this true for *Weight* as well? Give the P-value that is relevant to this question and explain what it means.

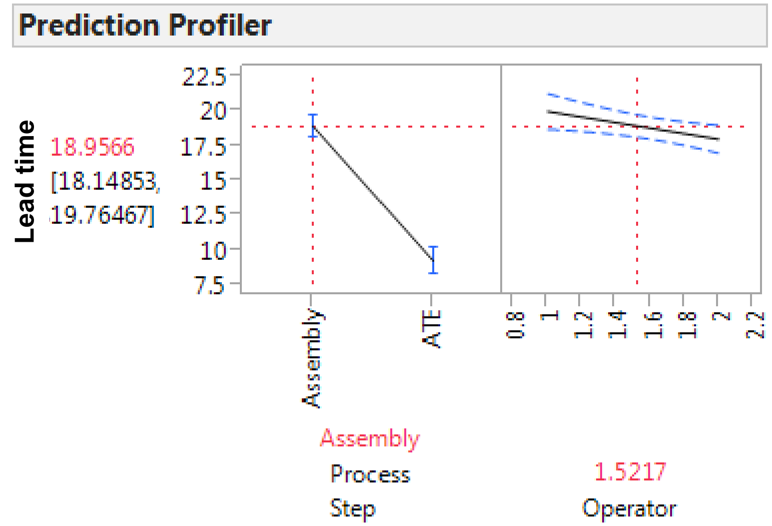
f) Give the predicted average *Weight* in the population of 15-year-old boys. Give a 95% confidence interval for this average.

g) Save your scripts, close and save the data table.

Exercise 5.2

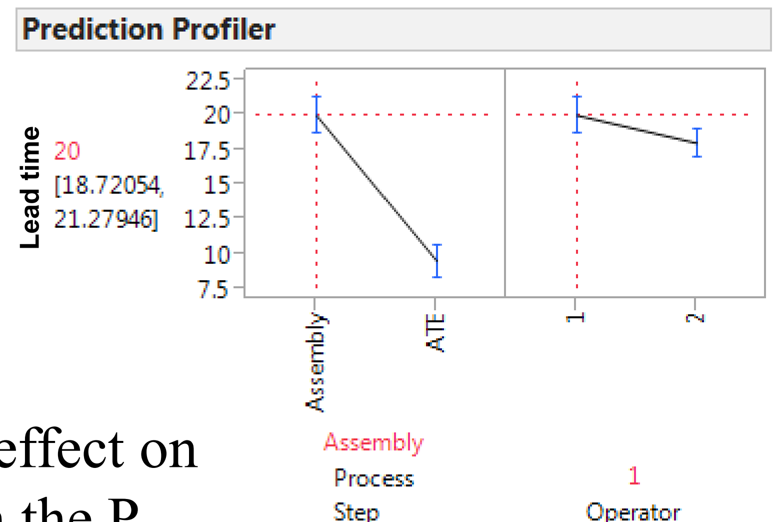
Data sets \ lead time 2.

a) Fit a model for *Lead time* including the terms *Process Step*, *Operator*, and their interactive effect. **Be sure you have the correct modeling type for *Operator*.** (If you got the upper right profiler, the modeling type for Operator is not correct. The lower right profiler is correct.)



b) Note anything concerning in the residuals plots.

c) Remove terms under *Effect Summary* with P-values exceeding 0.15 (*Remove* button). Which terms are left? Any issues with VIFs?



d) Based on the profiler, which factor has the larger effect on lead time (steeper slope)? Does this correlate with the P-values? Please explain.

e) Save your script, close and save the data table.

Exercise 5.3

Data sets \ number and size of defects.jmp.

a) Fit a model for *Max size* including the terms *Welder*, # *Defects*, their interactive effect, and the quadratic effect for # *Defects* (cross it with itself). This is the *Response Surface Model (RSM)* for one categorical factor and one continuous factor.

b) Do you see anything concerning in the residuals plots?

c) Using the *Effect Summary,* remove terms with P-values exceeding 0.15 (use the *Remove* button). Which terms are left in the model? Do all remaining terms have VIFs < 5?

d) Based on the profiler, which factor has the larger effect on *Max size*? Does this correlate with the P-values? Please explain.

e) Save your script, close and save the data table.

Exercise 5.4 [Instructor to demonstrate]

In this example you will analyze data from an optimization experiment concerning the removal of excess metal from castings by belt grinding.

The belt supplier had been recommending that belts be discarded when they are "50% used up." This rule was based on tests conducted by the supplier to define the usage point at which the total of labor and belt costs will be minimized. One of the grinders thought the supplier's rule caused grinders to discard belts too soon. Aside from being suspicious that the supplier just wanted to sell more belts, he argued that the supplier's tests did not take into account the time lost to belt changes.

This grinder developed a new standard under which belts would be discarded only after they were "75% used up." He wanted to do a comparative study to show that his method was cheaper overall. After he explains the study with his fellow grinders, 3 additional factors are added to the experiment.

Each casting in the experiment was weighed before and after the grinding operation. A technician kept track of how many belts were used and how long it took the grinder to complete each casting. From this information the total cost per unit of metal removed was calculated for each casting.

Data sets \ belt grinding.

Exercise 5.4 (cont'd) [Instructor to demonstrate]

- Y variable: *cost per unit of metal removed*
- X variables:

➢ Contact wheel land-groove ratio (LGR):	Low	or	High
➢ Contact wheel material (MATL):	Steel	or	Rubber
➢ Belt usage limit (USAGE):	"50%"	or	"75%"
➢ Belt grit size (GRIT):	30	or	50

- **Run the *Fit Model* script provided in the left panel,** by clicking on the green triangle. This is the response surface model for 4 categorical X variables.
- Check the residuals plots. Any problems?
- Using the *Effect Summary,* remove insignificant terms not needed to maintain model hierarchy, starting with the group of terms with P > 0.20, then one at a time. Which terms are left in the model?
- Use the *Prediction Profiler* to find the minimum cost factor settings.
- What do you expect the mean and standard deviation of *Cost* to be after implementing the optimal factor settings?
- Save your script, close and save the data table.

Exercise 5.5

In this example you will analyze data from an optimization experiment concerning the bond strength of potato chip bags.

Chips ‘R’ Us was receiving customer complaints about stale chips, especially from customers on airplanes. They traced the problem to the bag sealing process. The current process involved a temperature of 150°C, a pressure of 100 psi and a dwell time of 1.1 secs. The current average bond strength was about 85 psi.

Process Engineer Chip Kettle ran an experiment to increase the bond strength. Production Manager Justin Thyme reminded Chip that he would very much like to avoid an increase in the dwell time.

Justin is able to free up a bag sealer for only so much time each shift. Chip realizes he will need two shifts to complete the experiment. He decides to include *Shift* as an additional variable in the analysis just in case there is an operator and/or equipment effect.

Data sets \ heat sealing 1.

Exercise 5.5 (cont'd)

- Y variable: *bond strength*
- X variables and feasible ranges:

➢ Temperature (TEMP):	120 to 180
➢ Pressure (PRESS):	50 to 150
➢ Dwell time (DWELL):	0.2 to 2.0
➢ Shift:	1 or 2

- **Run the *Fit Model* script provided in the left panel**. This is the response surface model (RSM) for 3 continuous X's. Is anything concerning in the residuals plots?
- Remove from the model insignificant terms that are not needed to maintain model hierarchy (P > 0.15), using the *Effect Summary*. Which terms are left?
- Use the *Prediction Profiler* to maximize the average bond strength. If your solution requires a long dwell time, manually move things around in the profiler to find another solution with a short dwell time.
- What do you expect the mean and standard deviation of *bond* to be after implementing the optimal factor settings?
- Save your script, close and save the data table.

Exercise 5.6

Data sets \ outgassing process. *Current* (the Y variable) is the electrical current required to heat a filament to a specified temperature. *Resist* (one of the X variables) is the electrical resistance of the filament. *Machine* (the other X variable) identifies which of three processing units was used. We want to develop a model for *Current* as a function of *Resist* and *Machine*.

a) Fit a response surface model for *Current*. (The terms will be *Resist*, *Machine*, the interaction term *Resist***Machine*, and the quadratic term *Resist***Resist*. To get the quadratic term, highlight *Resist* both under Select Columns and under Construct Model Effects, then click Cross.)

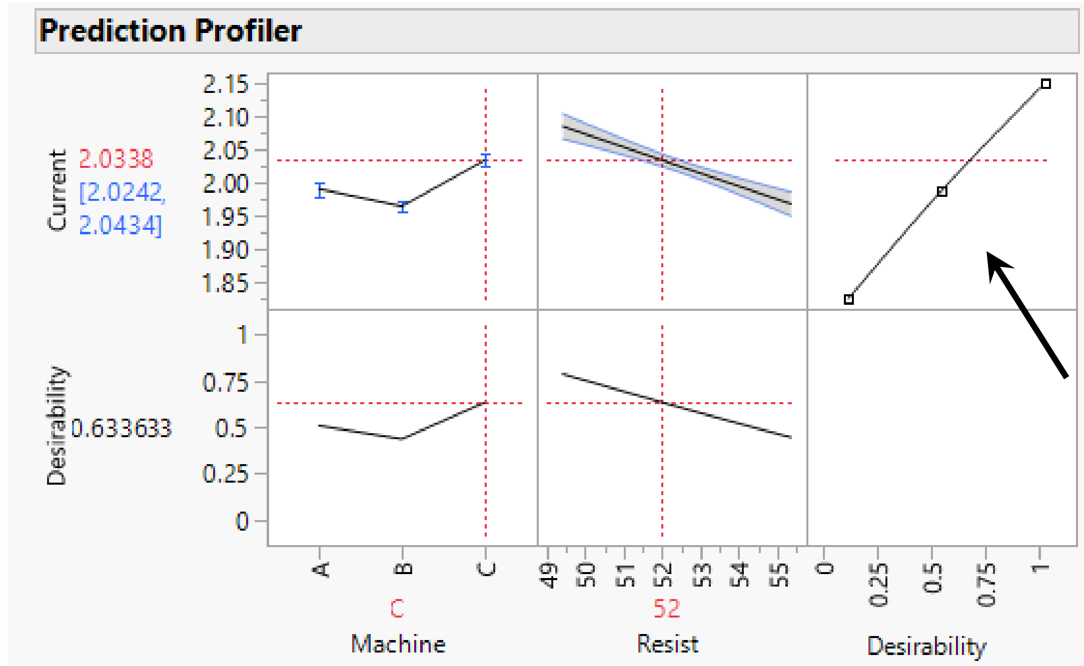
b) Do you see anything concerning in the residuals plots?

c) Remove any terms under *Effect Summary* with P value exceeding 0.15. (Use the *Remove* button.) Record the RMSE.

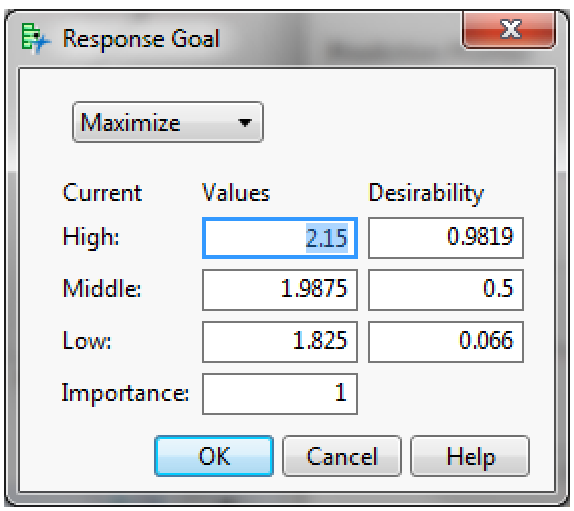
d) Use the *Prediction Profiler* to find the predicted average *Current* for each machine if we always use filaments with resistance 52.

Exercise 5.6 (cont'd)

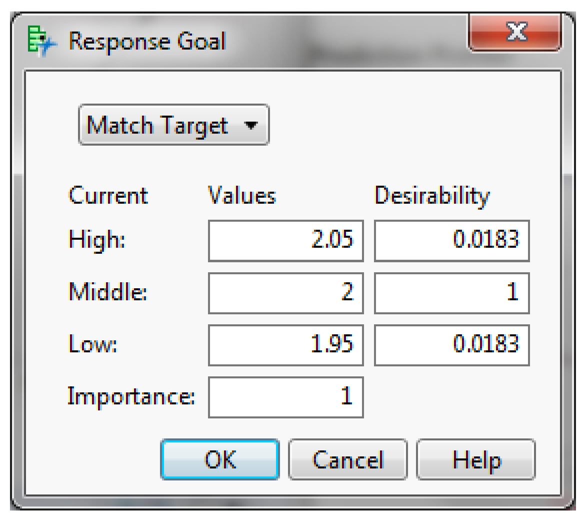
e) The target value for *Current* is 2. For each machine, we want to find the resistance for which the average current is 2. On the *Prediction Profiler* red triangle, select *Desirability Functions*. It should look like this:



f) Double click in the upper right hand panel of the profiler. (Try to avoid the plotted line.) You should get the dialog shown below.

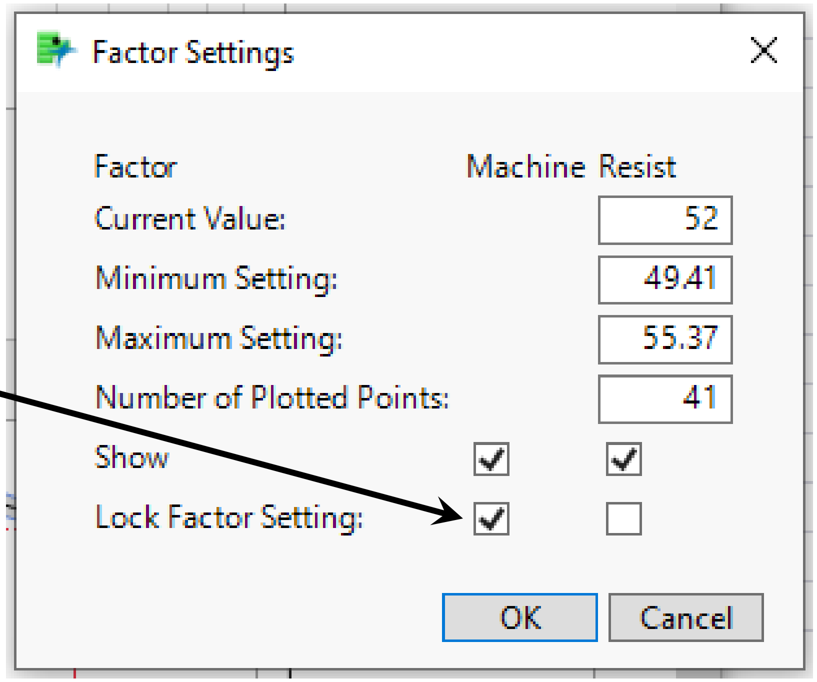


g) Modify the dialog as shown to the right, then select OK. Proceed to the next slide.

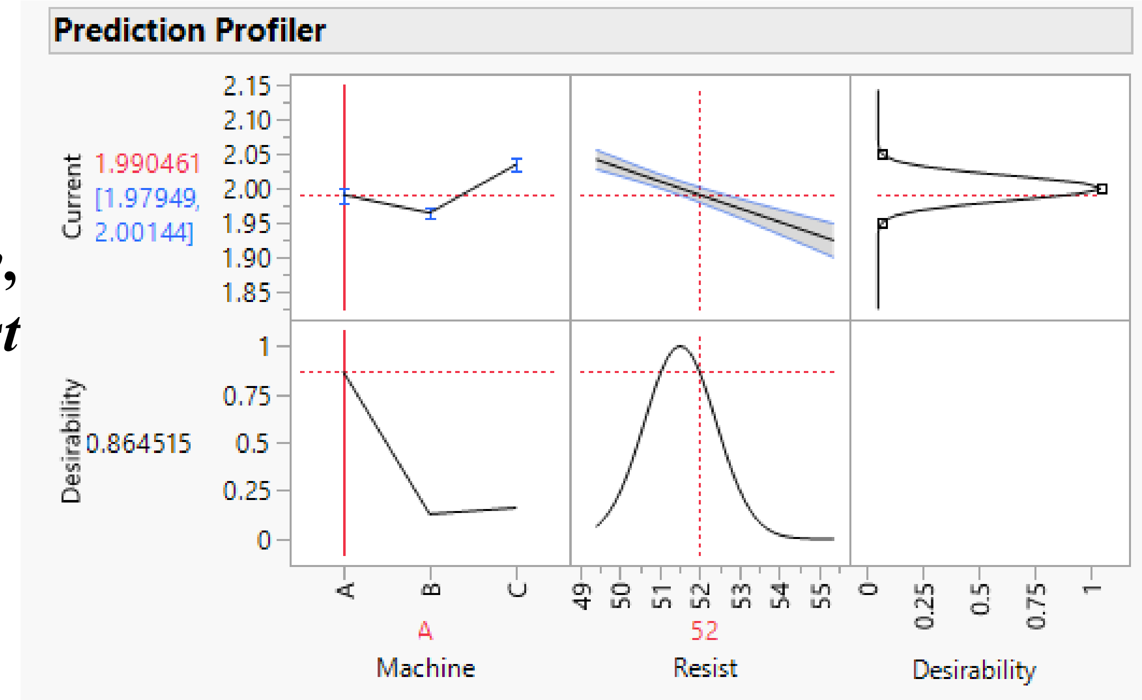


Exercise 5.6 (cont'd)

h) On the *Prediction Profiler* red triangle, select *Reset Factor Grid*. We want to lock the factor setting for *Machine*, so check the *Lock Factor Setting* box as shown here.

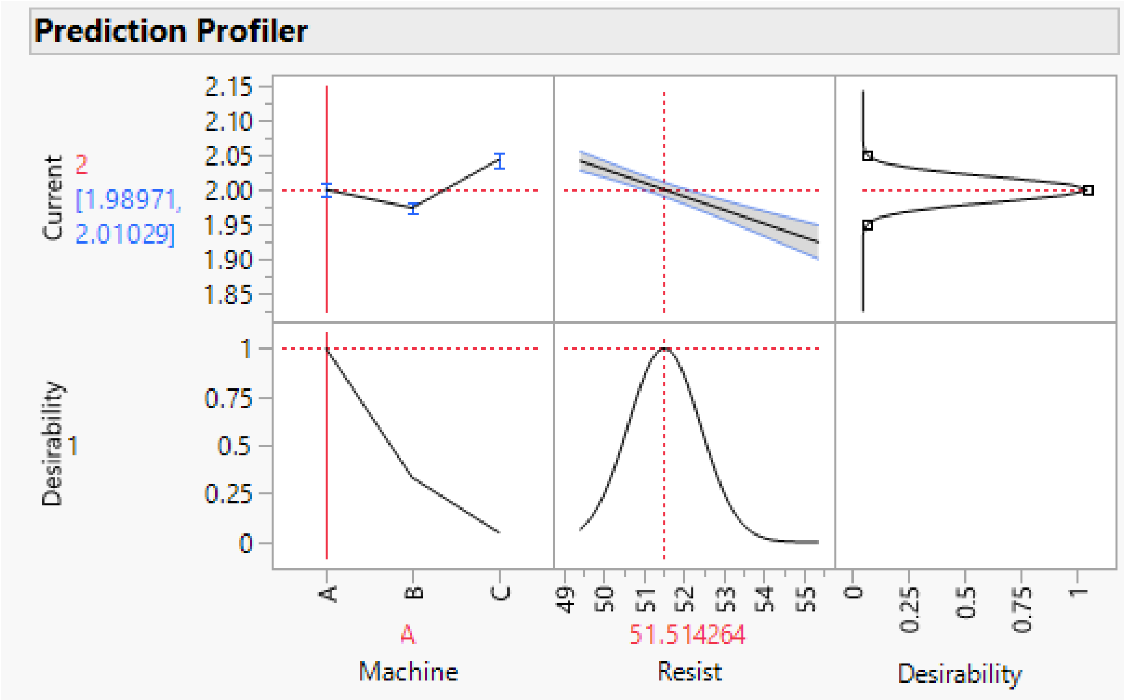


i) The vertical line for Machine should now be solid instead of dotted. **This will hold the machine setting in place during *Maximize Desirability*, which allows you to optimize *Resist* separately for each machine.** On the *Prediction Profiler* red triangle, select *Maximize Desirability*. Proceed to the next slide.



Exercise 5.6 (cont'd)

j) The optimal resistance value for Machine A is 51.5. Drag the solid vertical line across to B, then click *Maximize Desirability* to find the optimal resistance value for Machine B. Do the same for Machine C.



k) What will the average current be if we always use the optimal resistance values for each machine?

l) What will the standard deviation of current be if we always use the optimal resistance values?

m) Save your scripts, close and save the data table.

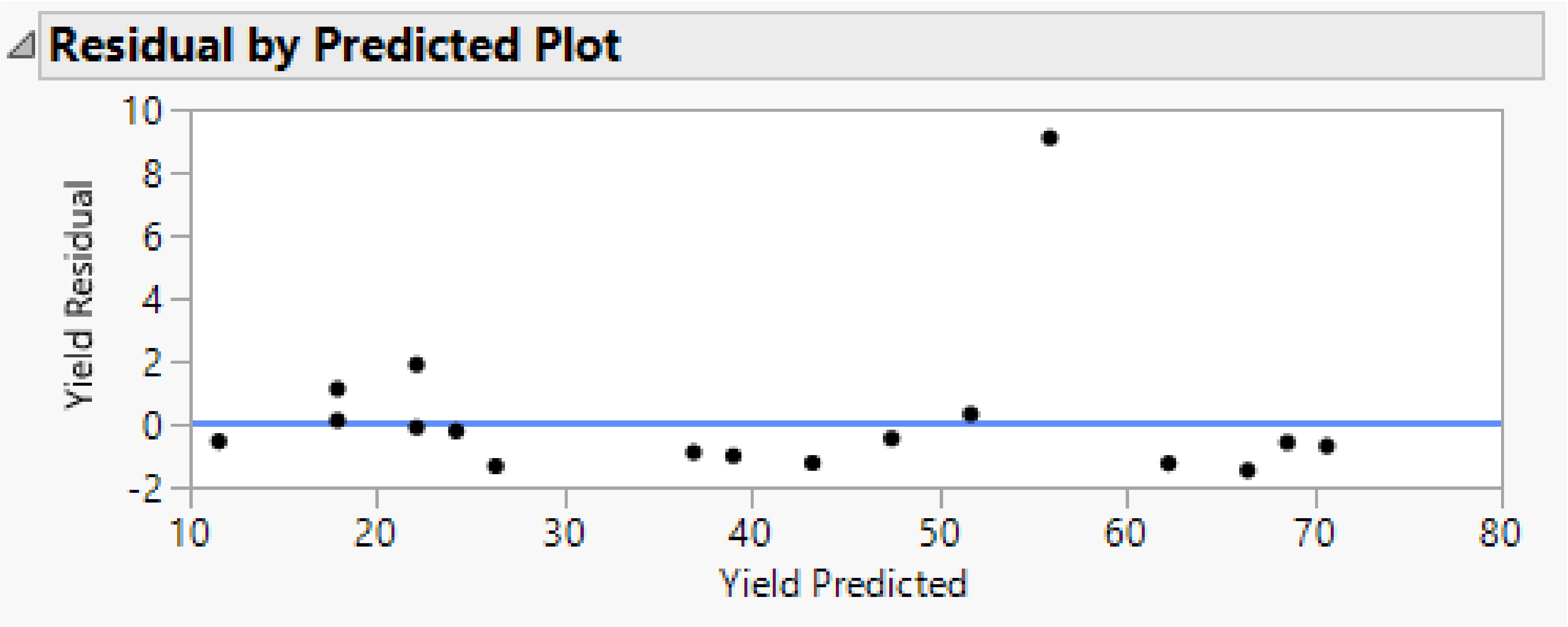
6 Dealing with Model Adequacy Issues

In this section, we will cover the most common model adequacy issues:

- Outliers
- Pattern in run order plot of residuals
- Multicollinearity (VIFs over 5)
- Unequal variance and non-normal residuals

Issue: Outliers

Outliers can easily be seen on the Residual by Predicted and Studentized Residuals (residuals by run order) plots



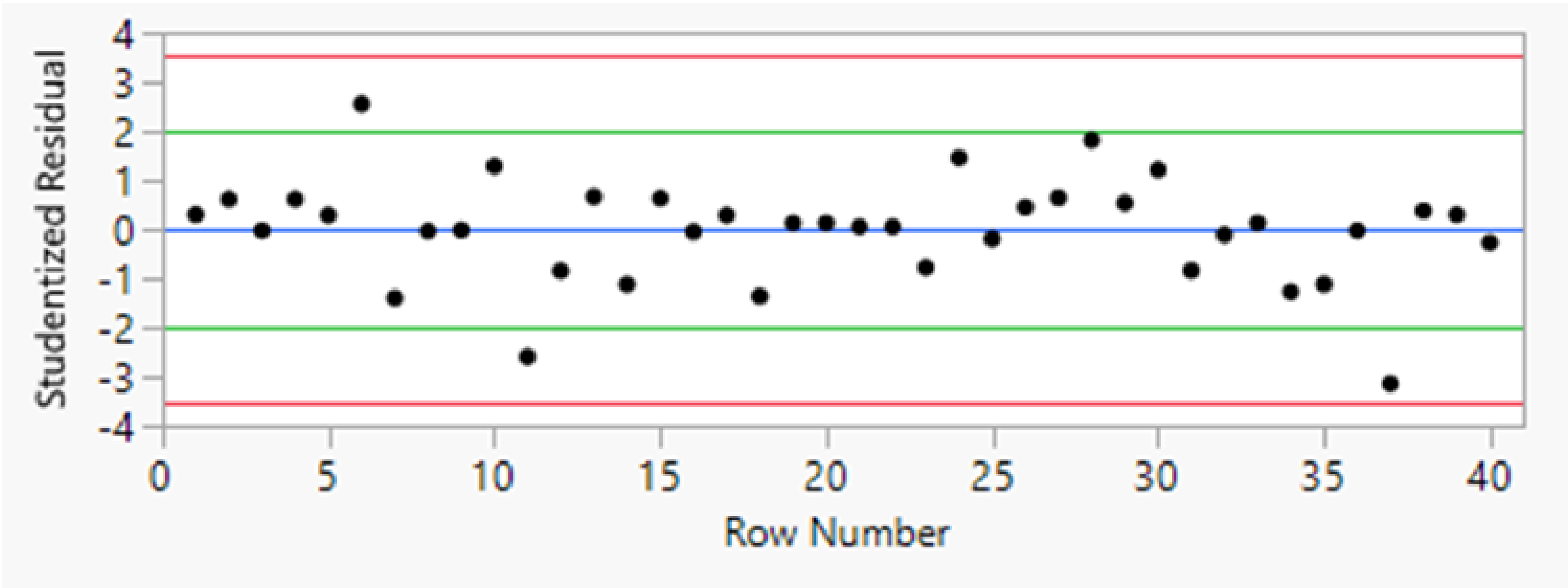
Remember, healthy residuals look like random scatter about zero.

Here, it looks like there might be a suspicious data point.

Issue: Outliers (cont'd)

- Investigate the data point.
  - If it turns out to be just a data entry error, we simply enter the correct value, then all is well. Most of the time it's not that simple.
- If you have an outlier of unknown origin:
  - Run the analysis with and without the questionable data point.
  - If you're lucky, the results will be pretty much the same both ways, hence no worries. Leave the data point in.
- If excluding the outlier does make a significant difference in the results, then you have a hard decision to make.
  - The official rule is: leave the data point in unless you can identify the cause. The idea is to throw it out only if you can demonstrate that it does not come from the population you want to study. This is the "pure" approach.
  - This should be tempered with the following practical consideration: you don't want your results to be unduly influenced by one extreme outlier, even if you can't explain it.

Issue: Pattern in run order of residuals



Remember, healthy residuals look like random scatter about zero.

There are no patterns of concern here.

Issue: Pattern in run order of residuals (cont'd)

- Runs (points in a row) of positive-negative-positive-negative residuals indicate correlation between runs in an experiment.
  - This implies that the assumption of independence has been violated.
  - **Randomization of an experiment protects against this! Do it every time!**

- This plot can show changes in variance over the time span of the experiment or data collection.
  - This could be due to increased skill as the experiment progresses, a process drift, operator fatigue, tool wear, etc.
  - This type of problem would show as an increase or decrease in spread or "scatter" of the residuals across the graph.
  - If there is x data available to support it, one remedy is to add a factor (time since tool change, number of hours of operator work, etc.)
  - Increasing or decreasing variance can also indicate the need for a transformation.

Issue: Multicollinearity (VIFs > 5)

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	4.868125	0.157585	30.89	<.0001*	.
LGR[Low]	0.616875	0.157585	3.91	0.0035*	1
Material[Rubber]	1.145625	0.157585	7.27	<.0001*	1
Usage[50%]	1.054375	0.157585	6.69	<.0001*	1
Grit[30]	-0.048125	0.157585	-0.31	0.7670	1
LGR[Low]*Grit[30]	-0.316875	0.157585	-2.01	0.0752	1
Usage[50%]*Grit[30]	0.395625	0.157585	2.51	0.0333*	1

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	14.044944	0.291958	48.11	<.0001*	.
Process Step[Assembly]	4.8792135	0.298829	16.33	<.0001*	1.0478749
Operator[1]	0.6713483	0.296556	2.26	0.0349*	1.0478749

Remember, VIF < 5 is not concerning.

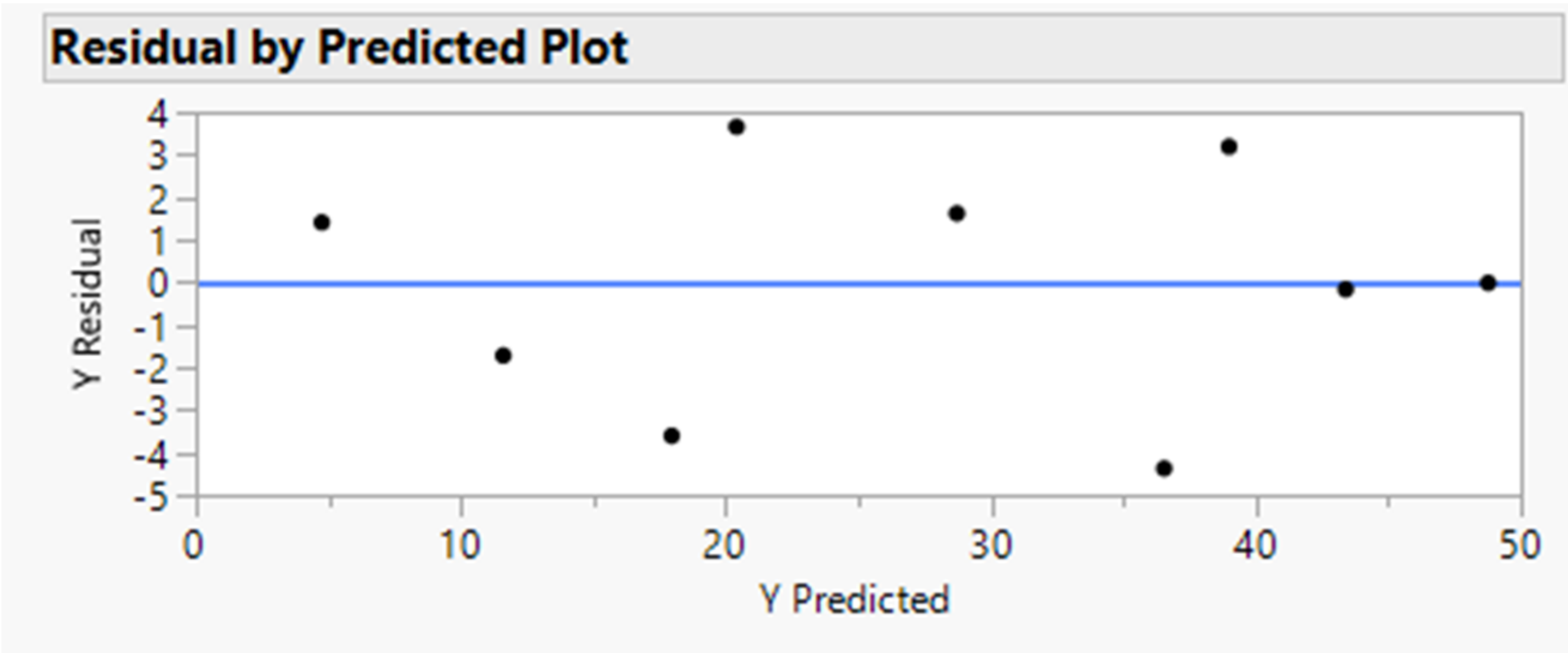
- One aspect of factorial design experiments (often called DOEs) is that they are orthogonal designs. This results in the model terms being completely uncorrelated.
- Regressors that are completely uncorrelated with others have VIF = 1.
- High correlation is only a potential issue when using historical or observational data in regression analysis.

Issue: Multicollinearity (cont'd)

Several strategies can be tried for resolving multicollinearity, but they may not be satisfactory, especially if the model will be used for prediction.

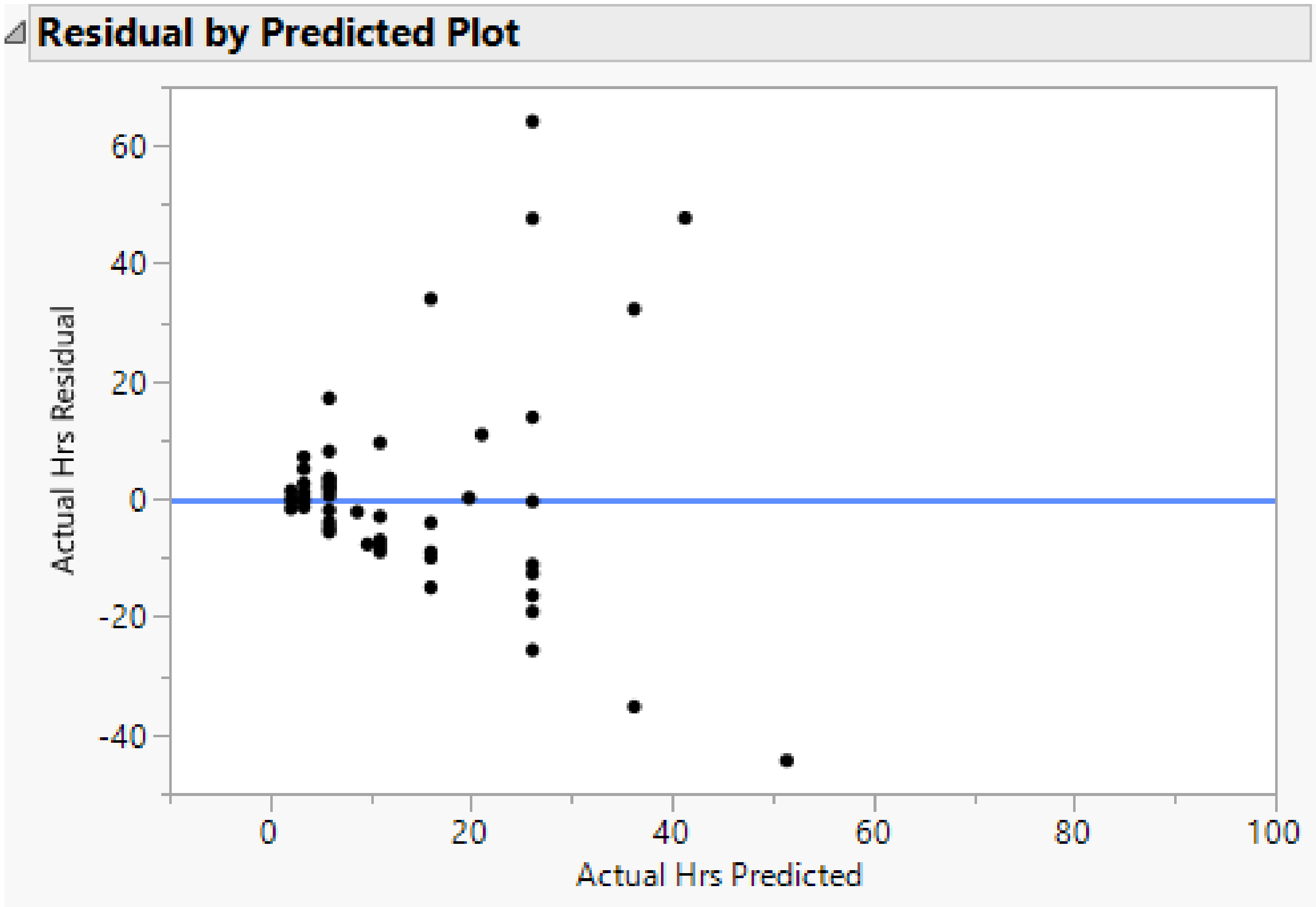
- Collect additional data in a way that breaks up the multicollinearity.
  - Historical data may contain only certain combinations of x-variables, for example, only low levels of x_1 when x_2 is at a low level and only high levels of x_1 when x_2 is at a high level
  - Note: it may not be feasible or possible to collect this additional data.
  - In some cases, the factors (x's) are inherently correlated, for example as may be the case with household income and house size.
- Respecifying the model, can help.
  - If x_1 and x_2 are nearly linearly dependent, use one term, $x = x_1 + x_2$, which preserves the information content of the original variables
  - Try removing the term with the highest p-value, and look at that model. Then, replace it and remove the term with the highest VIF. See which gives the better model.
- Use ridge or principal-component regression (way beyond the scope of this course)

Issue: Unequal variance and non-normal residuals



Remember, the variation in the residuals should be fairly constant across the Residual by Predicted Plot. There is no issue here.

Issue: Unequal variance and non-normal residuals (cont'd)

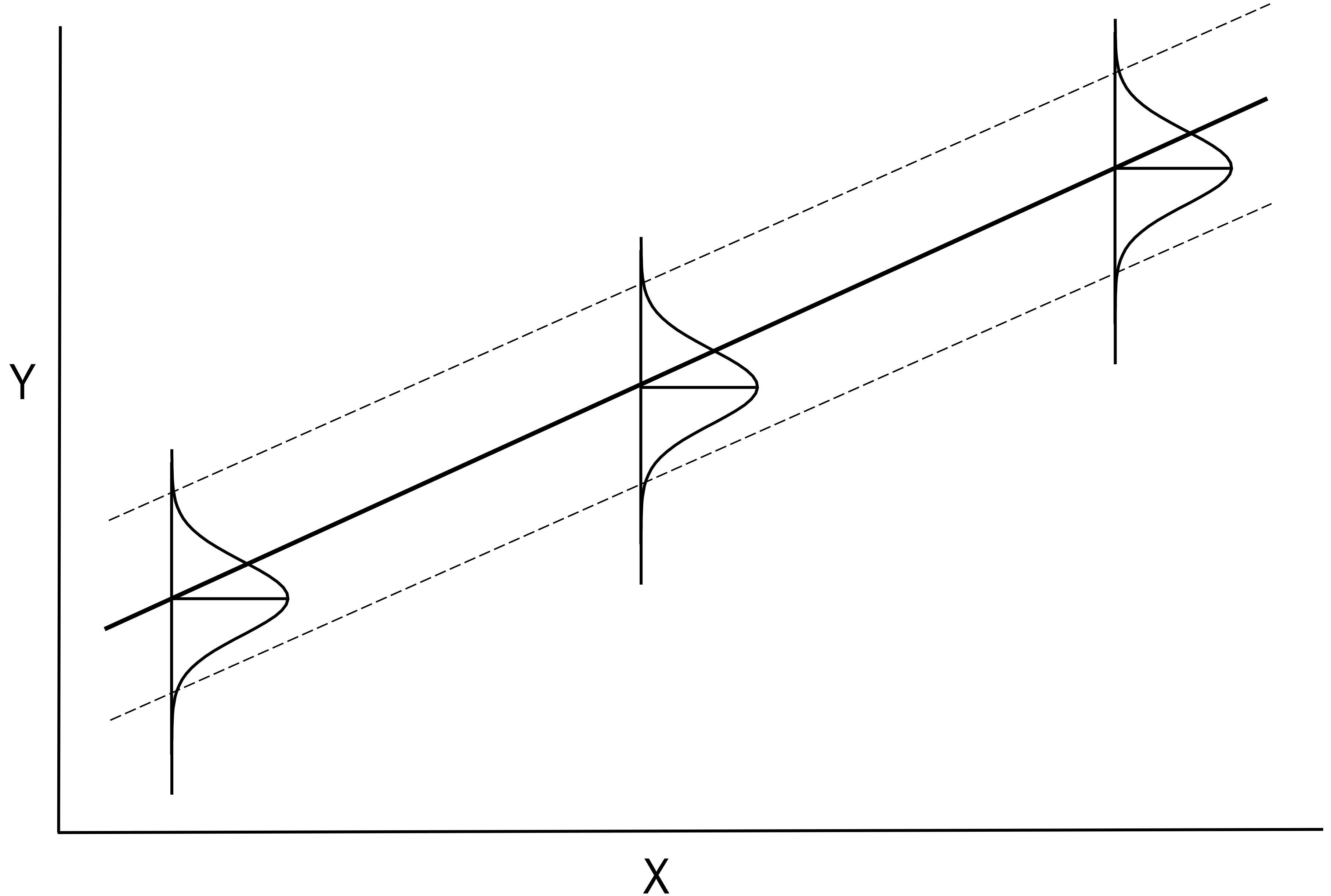


In this plot, we can see an issue.

σ_Y^2 proportional to mean Y → "sideways V"

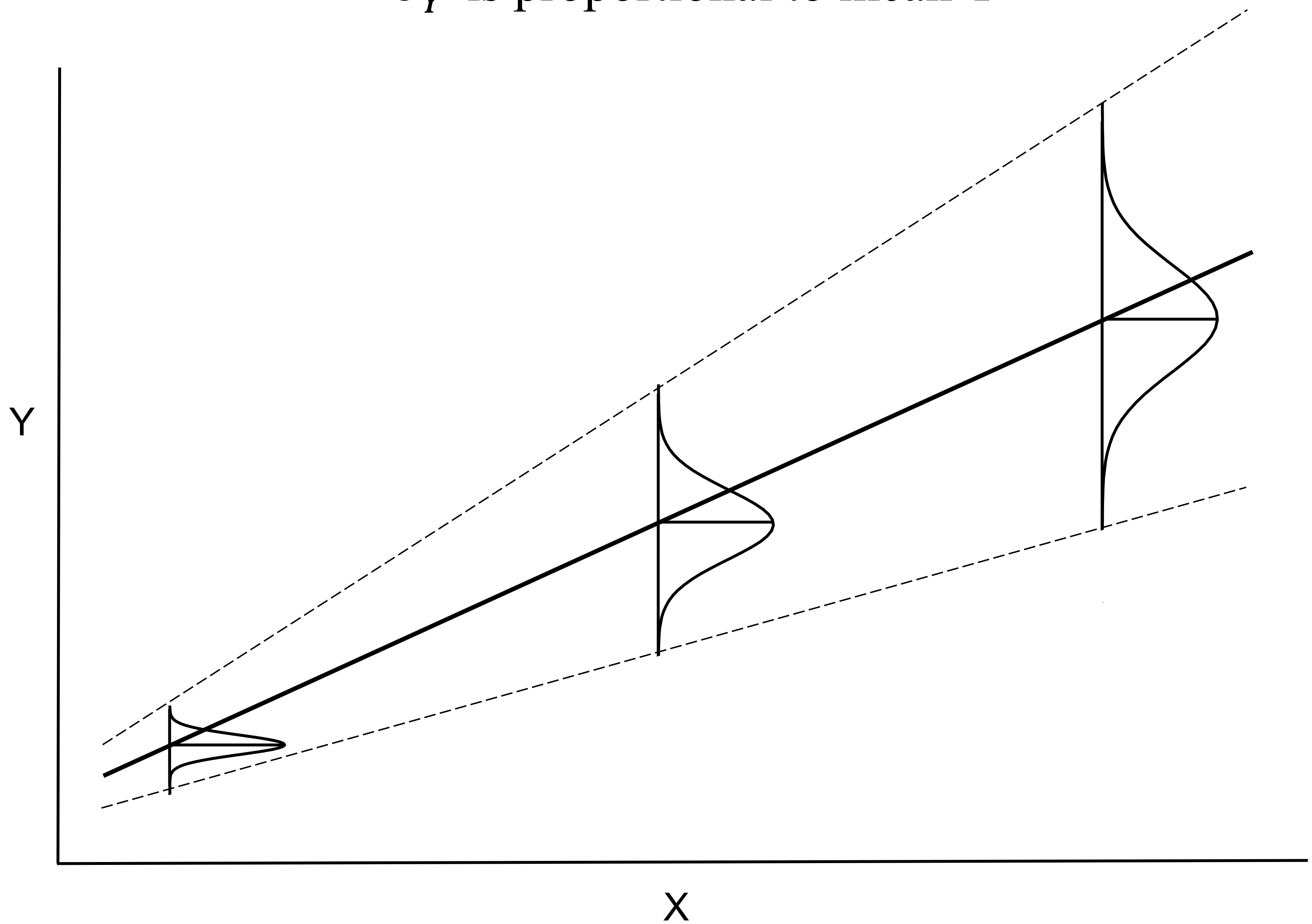
Basic model assumption: constant variance

σ_Y^2 is constant (does not depend on the X's)

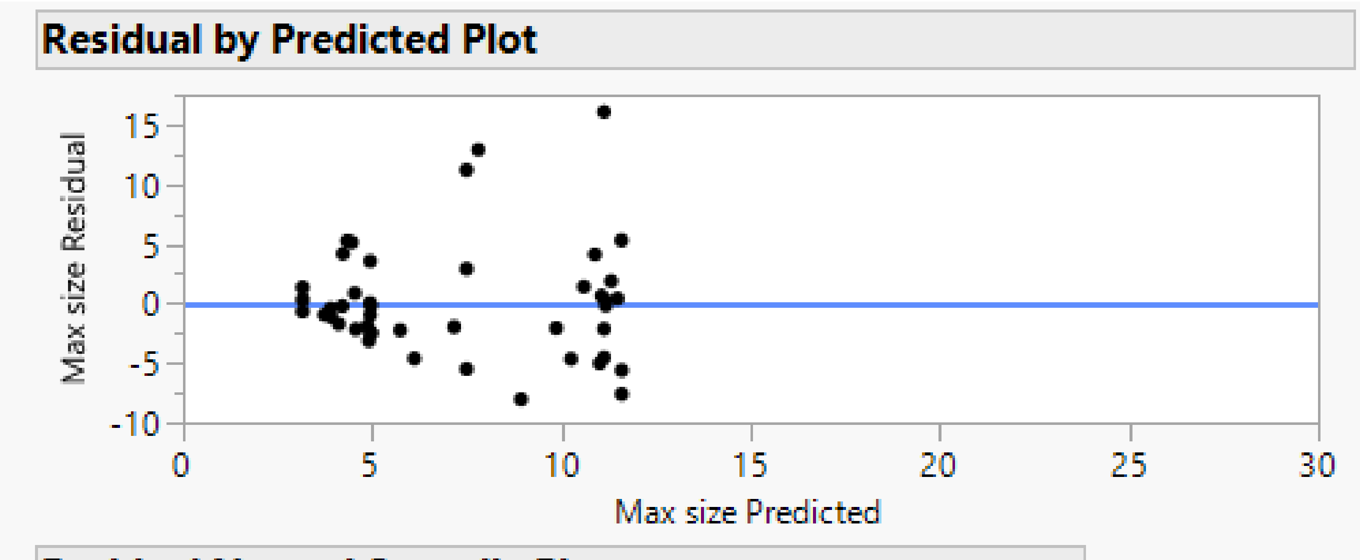


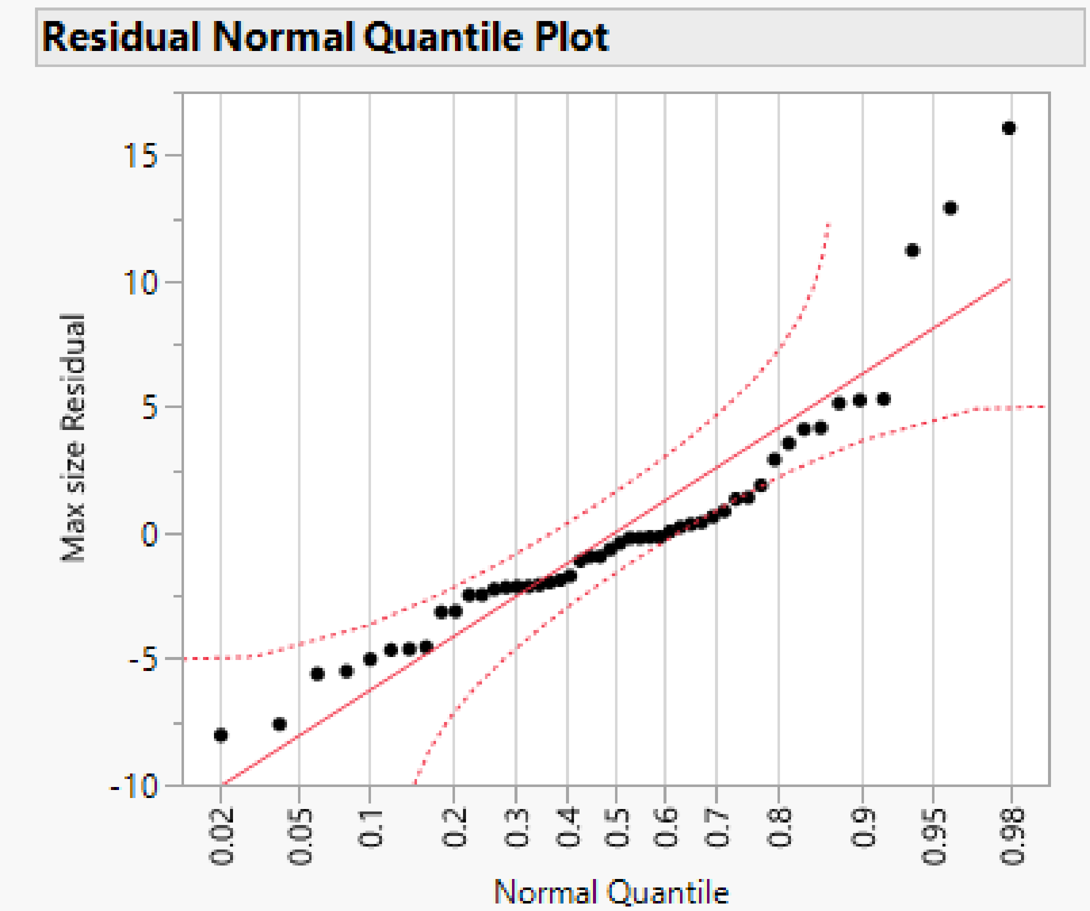
Most common violation of the basic assumption

σ_Y^2 is proportional to mean Y



Issue: Unequal variance and non-normal residuals (cont'd)





- Often, when there is an issue with constant variance, there is also the issue of non-normal residuals.
- This can be seen in these two plots
- Fortunately, they usually both resolve with the same treatment—a transformation.

The standard assumption in all comparison and correlation analyses involving a quantitative Y variable is that the noise (unexplained/error/residual) variation follows a Normal distribution with mean 0 and a standard deviation that does not depend on the X variables.

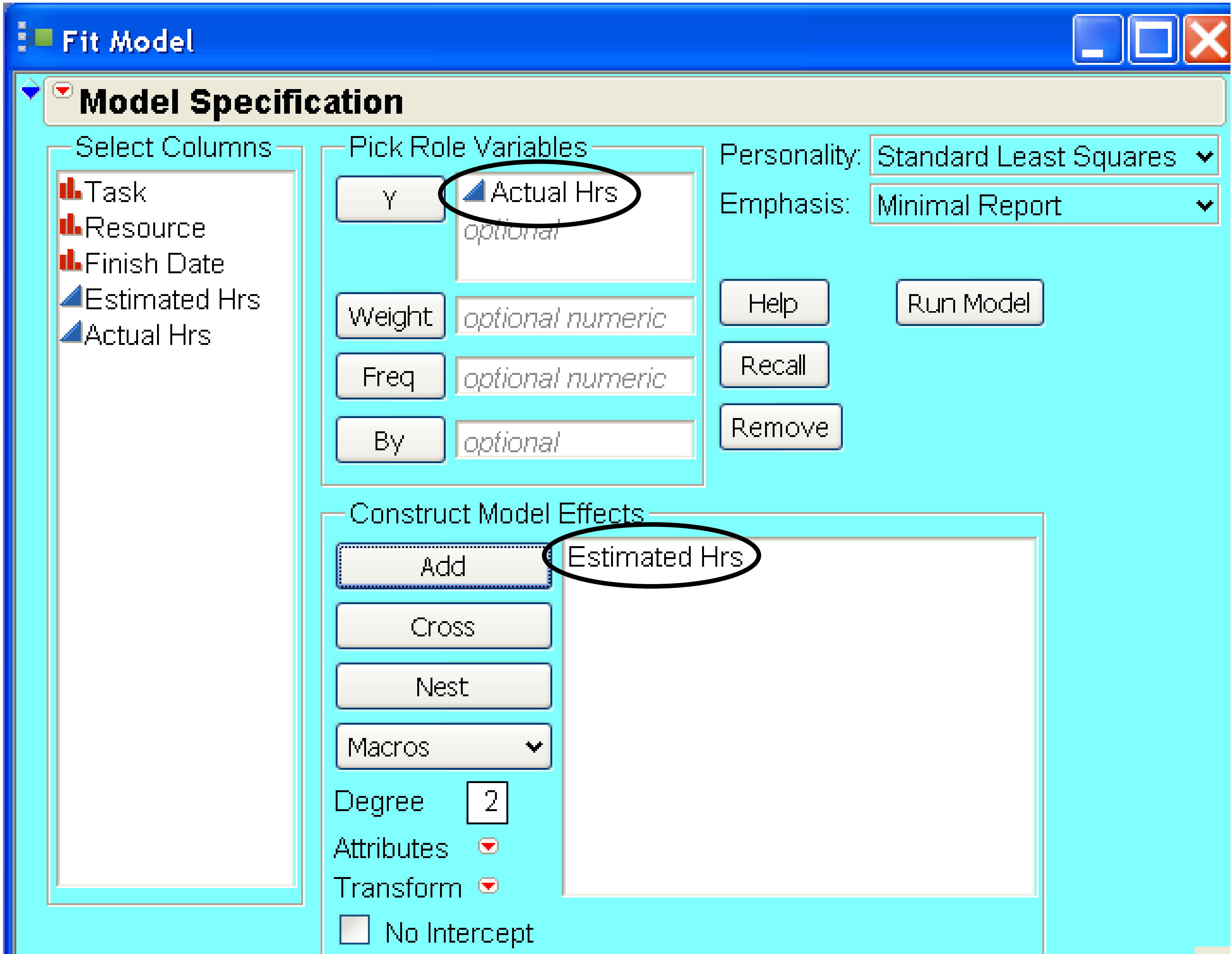
This simple model has served us well. However, when Normality or constant σ is grossly violated, something must be done. The most common remedy is to use log(Y) or sqrt(Y) as the dependent variable instead of Y. This is a transformation. This "trick of the trade" is simple and, in most cases, effective.

Transforming the Y variable

Data sets \ actual vs estimated

We want to see how accurately we can estimate the time it takes to do certain tasks

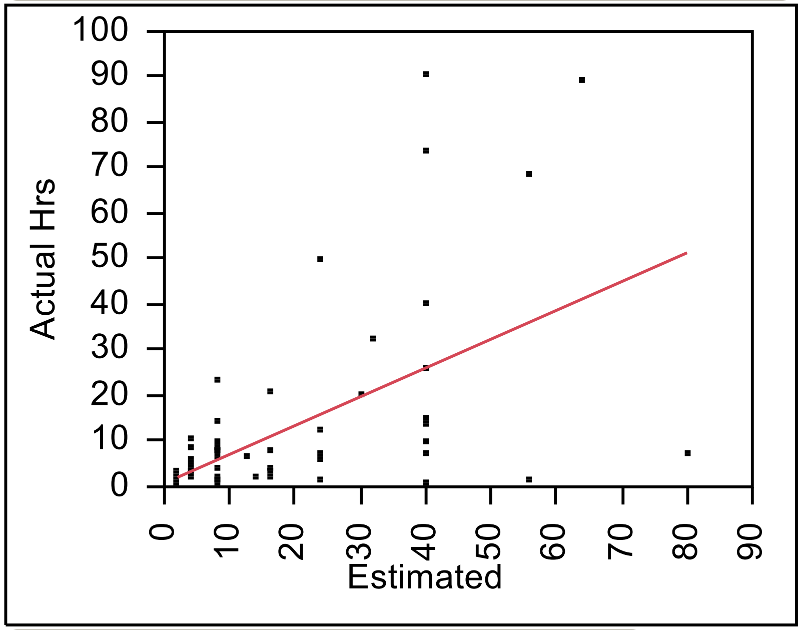
Analyze
↓
Fit Model



Transforming Y (cont'd)

Response Actual Hrs

Regression Plot



Summary of Fit

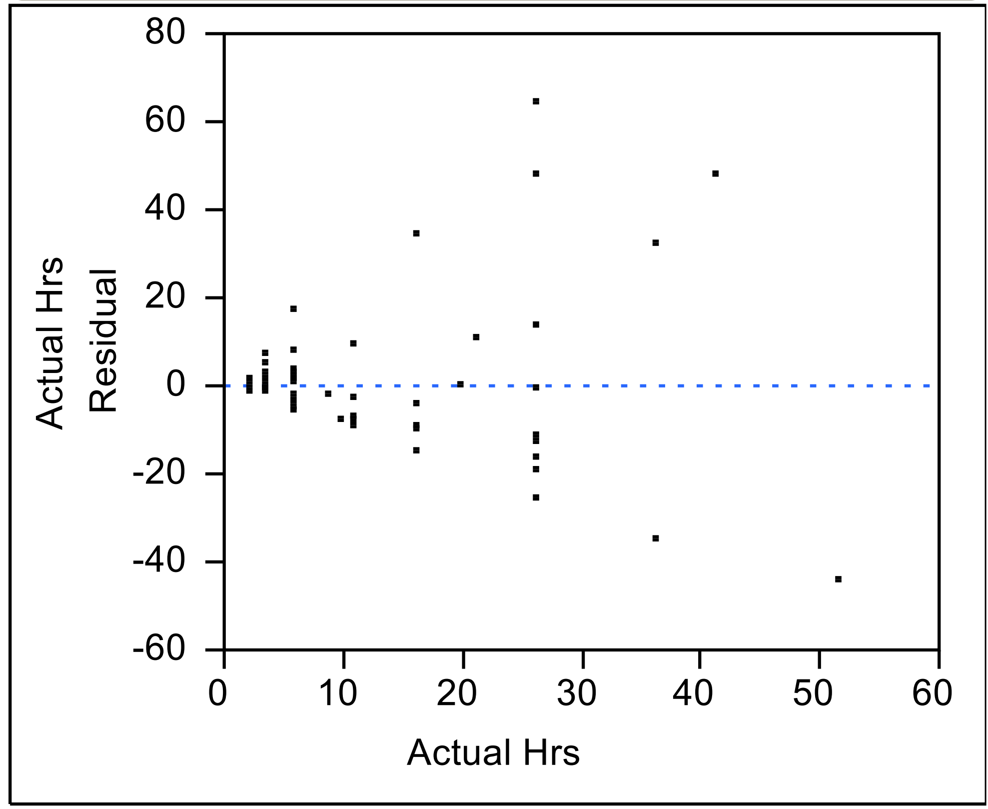
RSquare	0.307347
RSquare Adj	0.296176
Root Mean Square Error	16.95281
Mean of Response	12.23828
Observations (or Sum Wgts)	64

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.8352064	3.035964	0.28	0.7842
Estimated Hrs	0.6321871	0.120529	5.25	<.0001 *

$$Y = 0.835 + 0.632\ X$$

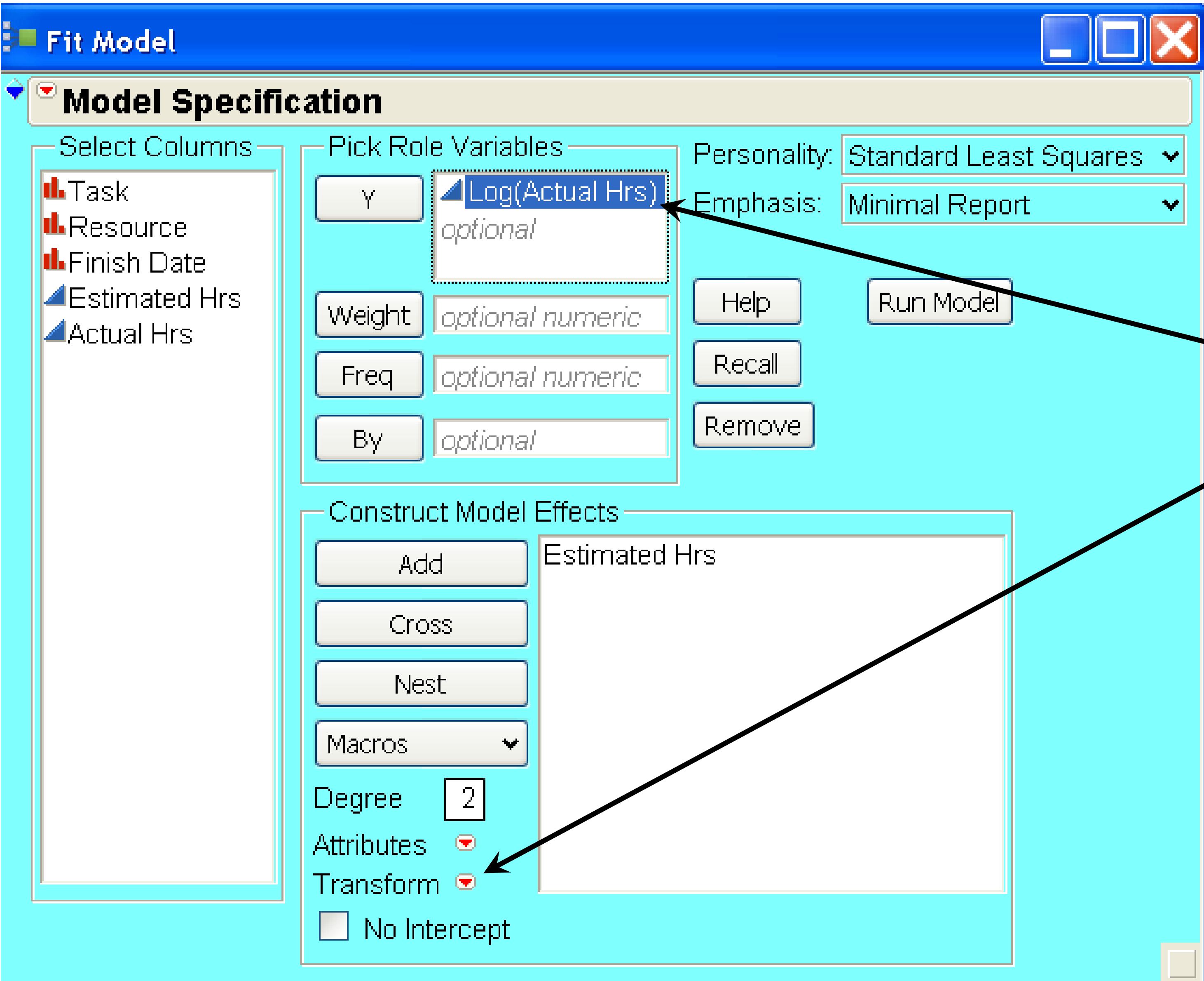
Residual by Predicted Plot



Variation increases as average *Actual Hrs* increases

Transforming Y (cont'd)

σ_Y proportional to mean Y ⟺ $\sigma_{Log(Y)}$ constant



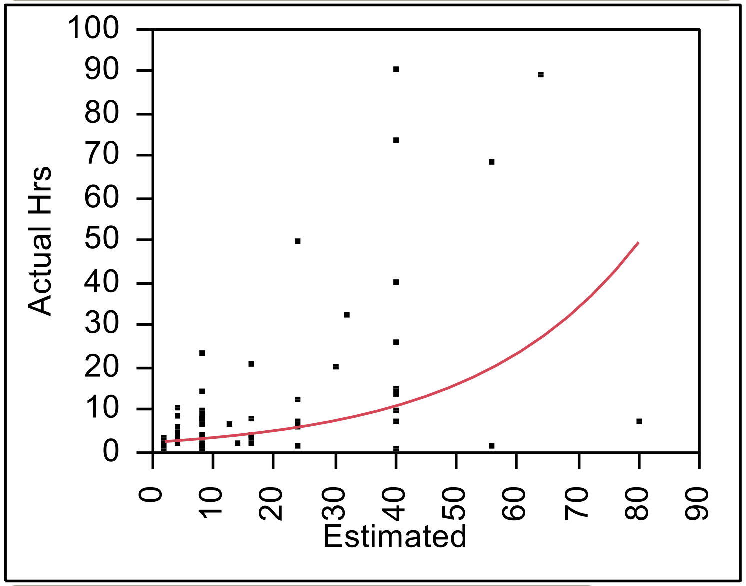
Response Actual Hrs > Model Dialog

- Click on *Actual Hrs*
- Click on *Transform* red triangle
- Select *Log*
- Run the model

Effects of log transformation

Response Log(Actual Hrs)

Regression Plot



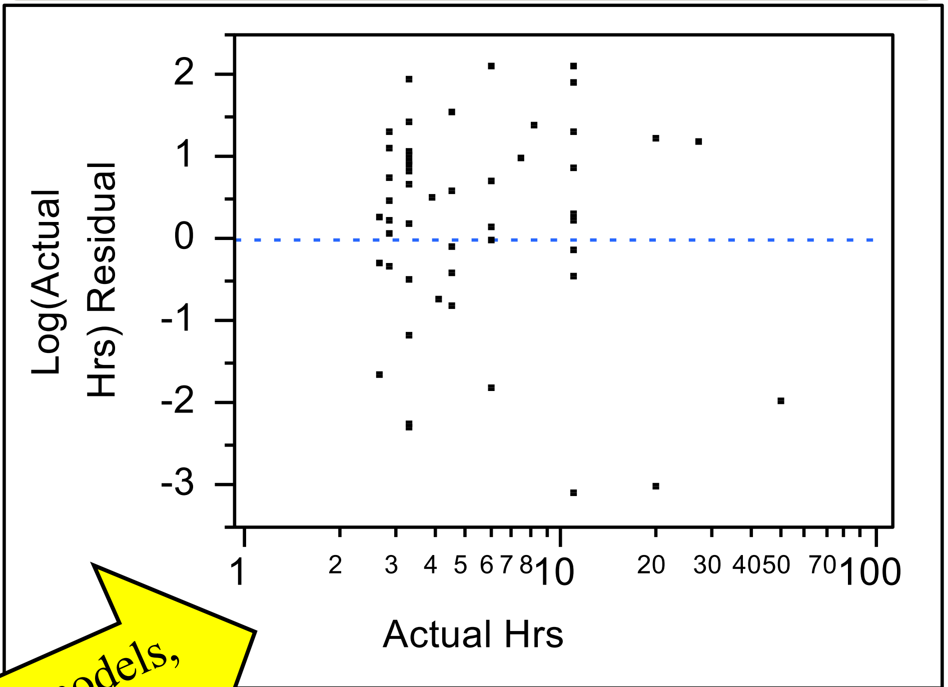
Summary of Fit

RSquare	0.233276
RSquare Adj	0.22091
Root Mean Square Error	1.217933
Mean of Response	1.576584
Observations (or Sum Wgts)	64

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.8982207	0.218111	4.12	0.0001 *
Estimated Hrs	0.0376085	0.008659	4.34	<.0001 *

Residual by Predicted Plot



For Log(Y) models, use Log scale here

Nonlinear model for Y

$$\text{Log}(Y) = 0.898 + 0.038X$$

$$Y = \exp(0.898 + 0.038X) = e^{0.898}\left(e^{0.038}\right)^X = 2.45(1.04)^X$$

Note on JMP notation, and impacts of the Log transformation

JMPs notation regarding Logs requires some clarification:

- Although JMP expresses the logarithm as "Log", it is actually base e, or the natural log, which is usually written as Ln. It is not a base 10 logarithm.
- However, the plots that use a log transformed X-axis display use base 10 log for the X-axis. This does not change the interpretation of the chart.

The impact of transformation on R^2 and p-values:

- In the previous example, a transformation was required because the residuals variance wasn't constant over the range of the predicted values.
- After the transformation, the R^2 value went down. This can lead to a belief that the non-transformed model was "better". However,
- Residuals showing this condition (heteroscedasticity) can cause p-values and R^2 to be over or under stated.
- When this condition occurs, the problem must be corrected. The resulting model, even if R^2 is lower or p-values are higher, is the more "real" model.

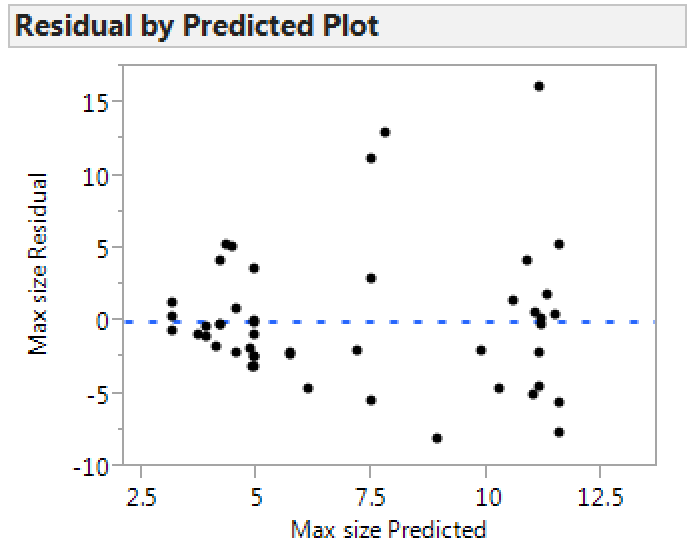
Steps in Multiple Regression (backward elimination method)

1. Run Analyze > Fit Model in JMP to investigate the relationship between y and x's. Use the Response Surface Model (all factors, all interactions, quadratic terms for continuous variables/factors)
2. Check model adequacy by reviewing the residuals plots:
   - Residual Normal Quantile Plot
   - Residual by Predicted Plot
   - Studentized Residuals (in run order)
3. Transform the data and resolve other issues, if needed.
4. Verify all VIFs < 5. Address the issue if any are over 5.
5. Remove insignificant terms from the model, that are not needed to maintain model hierarchy (main effects must be included if a higher order term of that variable remains in the model).
6. Use $Adjusted\ R^2$ to determine the amount of variation in Y that is explained by the model.

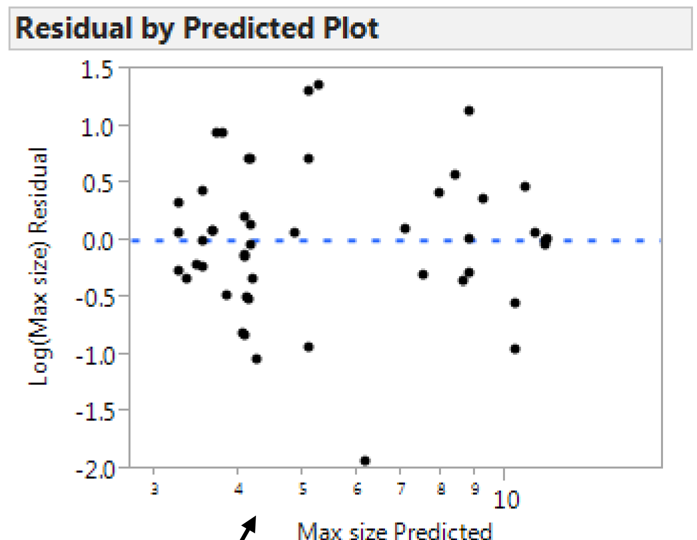
Exercise 6.1

Data sets \ number and size of defects.jmp

a) Fit a model for *Max size* including the terms *Welder*, *# Defects*, their interactive effect, and the quadratic effect for *# Defects* (*response surface model* for one continuous factor and one categorical factor). You should see a distinct sideways V. Do you see issues in any other residuals plots?



b) Select *Model Dialog* on the *Response* red triangle menu, apply a Log transformation to *Max size*, re-run the model. The sideways V isn't completely gone, but close enough. Did other residuals plots improve?



c) Use *Effect Summary* to remove terms with P > 0.15.

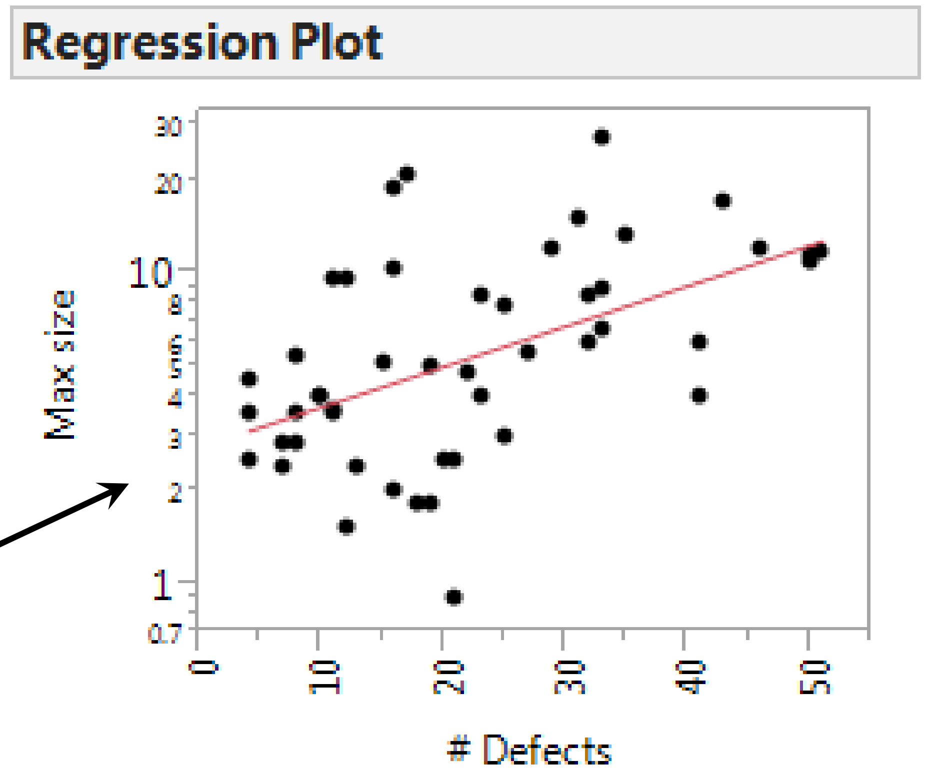
Remember to change the x-axis on the plot, as well.

Exercise 6.1 (cont'd)

d) Which terms are left in the model?

e) Now we have a log-linear simple regression.

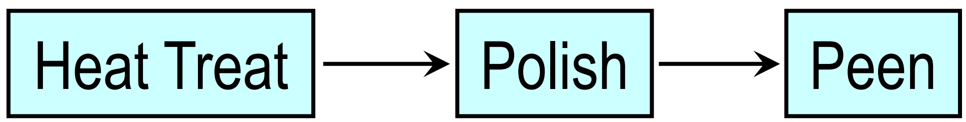
When you use a Log or square root transformation on Y, it is helpful to use same scale for the Y axes of the plots



f) Save your script, close and save the data table.

Exercise 6.2

An aerospace manufacturer uses integral castings as structural components of jet engines. Integral castings give design engineers more flexibility and simplify the assembly process. Defect-free castings are known to have long cycle fatigue life, but defects often arise in the casting process and must be weld repaired. The engine manufacturer’s metallurgical team has proposed a finishing process of the following type to ensure adequate cycle fatigue life of weld-repaired castings:



The team wants to optimize the first two steps in this process to achieve maximum cycle fatigue life. Also, though other applications of similar processes have included peening, they would like to see if it can be omitted to reduce processing time and cost.

Due to project time constraints and limited availability of test fixtures, the team can perform at most 12 cycle fatigue tests for their experiment.

Exercise 6.2 (cont'd)

- Y variable: *Cycles* (to failure)
- X variables:
  - Heat treat: Anneal or Solution/age
  - Polish: Chemical or Mechanical
  - Peen: Yes or No

- *Data sets \ weldment fatigue.jmp.*
- Run the *Model* script provided in the left panel, run the model.
- Notice the extreme sideways V on the *Residual by Predicted Plot*. Are there issues in any of the other residuals plots? If yes, what are they?
- Rerun the model using a Log transformation on *Cycles*. Did residuals plots improve?
- Remove insignificant terms from the model (P > 0.15) that are not needed to maintain model heirarchy.
- Use the *Prediction Profiler* to maximize the cycle fatigue life.

Exercise 6.3

A Black Belt wants to minimize the *leak rate* in plastic containers ultrasonically welded together. The X variables and ranges are:

- Force: 70 to 150
- Energy: 275 to 325
- Amplitude: 70 to 90

- *Data sets \ ultrasonic welding 1.jmp.*

- Run the *Model* script provided in the left panel.

- What problems do you see in the residuals plots?

Exercise 6.3 (cont'd)

- Rerun the model using the Log transformation on *leak rate*. (Be sure to change the x-scale to Log on the Residual by Predicted Plot.)
- Rerun the model using the Sqrt transformation on *leak rate*. (Be sure to change the x-scale to Sqrt on the Residual by Predicted Plot.)
- Which set of residuals plots looks better? Use whichever transformation looks like it worked better, going forward.
- Remove insignificant term(s) from the model ($P > 0.15$), while maintaining model hierarchy.
- Use the *Prediction Profiler* to minimize the leak rate.

This slide intentionally left blank

When the response variable, Y, is binary (pass/fail, yes/no, success/failure, etc.), the regression model used for a continuous Y-variable *cannot* be used.

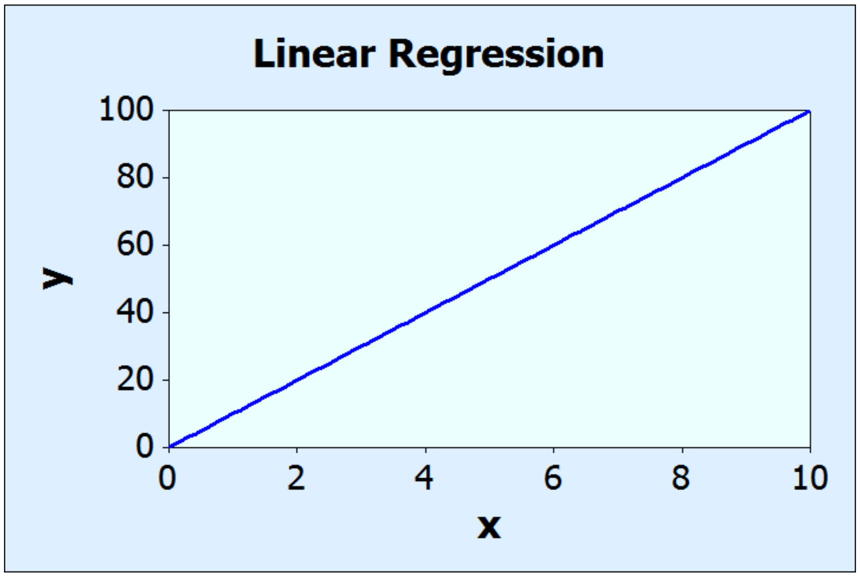
- A *logistic response function* must be used
- The resulting analysis yields an equation that allows us to calculate **event probability:**

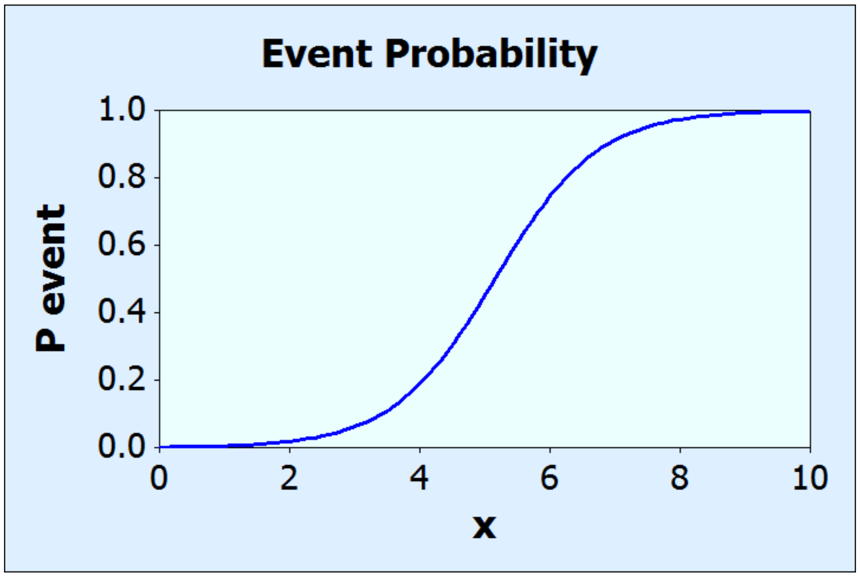
$$\boldsymbol{P_{event} = f(x_1, x_2, \ldots, x_n)}$$

- This equation is used to answer questions such as:
  - What is the probability of being in spec (at various levels of x)?
  - What is the probability of getting the contract?
  - What is the probability of a defect?

Probability Function for Pass/Fail Y

This probability function, the *logistic response function*, has a much different behavior than a linear regression function:





- The *y* values of a linear regression can have any values
- The logistic response function is an S-shaped function that can only have values between 0 and 1

To be useful in prediction, the logistic response function must be transformed into an unbounded linear function

Transforming the logistic response function (cont'd)

The *logit transformation* is used to linearize the model:

$$logit(P_{event}) = \ln\left(\frac{P_{event}}{1 - P_{event}}\right) = b_0 + b_1x_1 + \ldots + b_nx_n$$

$$\frac{P_{event}}{1-P_{event}} = e^{b_0+b_1x_1+\ldots+b_nx_n}$$

$$\boxed{P_{event} = \frac{1}{1 + e^{-(b_0+b_1x_1+\ldots+b_nx_n)}}}$$

- This is the form of the final equation in the regression analysis
- The *maximum likelihood* method is used to estimate the parameters in this probability equation . . . JMP does this work for us
- We can use this equation (model) to predict the probability of an event for various levels of $x_1, x_2, \ldots, x_n$

Using JMP for Simple Regression with Pass/Fail Y

We will see how to use JMP do the regression analysis when we have:

a) Raw data – each row represents one part or transaction

b) Tabulated data – each row represents multiple parts or transactions

Raw data

2/0 Cols / 25/0	Target speed	Result
1	200	Hit
2	205	Hit
3	210	Hit
4	215	Hit
5	220	Hit
6	225	Miss
7	230	Hit
8	235	Hit
9	240	Miss
10	245	Hit
11	250	Hit
12	255	Hit
13	260	Hit
14	265	Miss
15	270	Miss
16	275	Hit
17	280	Miss
18	285	Miss
19	290	Miss
20	295	Miss
21	300	Hit
22	305	Miss
23	310	Miss
24	315	Miss
25	320	Miss

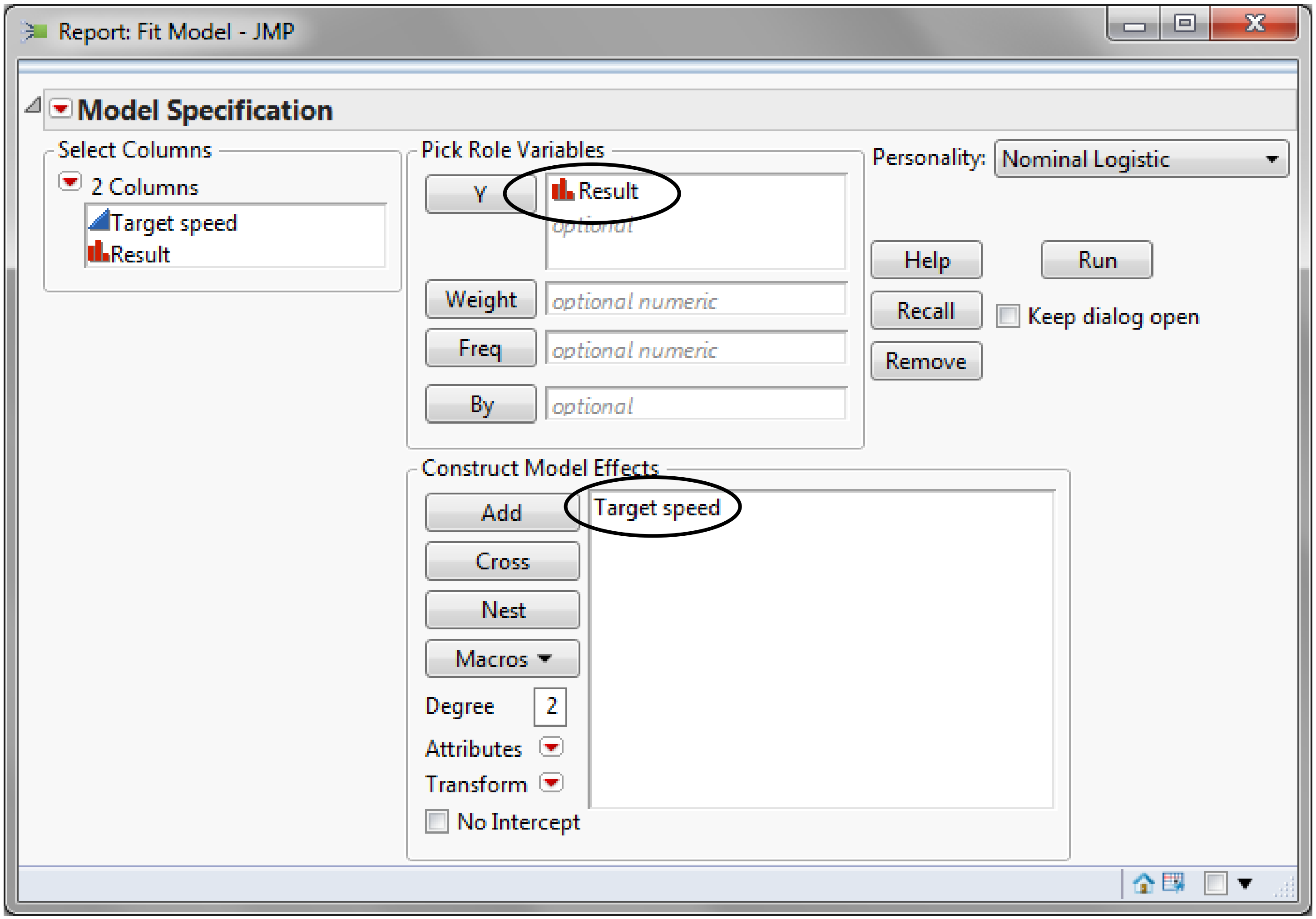
Data sets \ target practice

↓

Fit Model

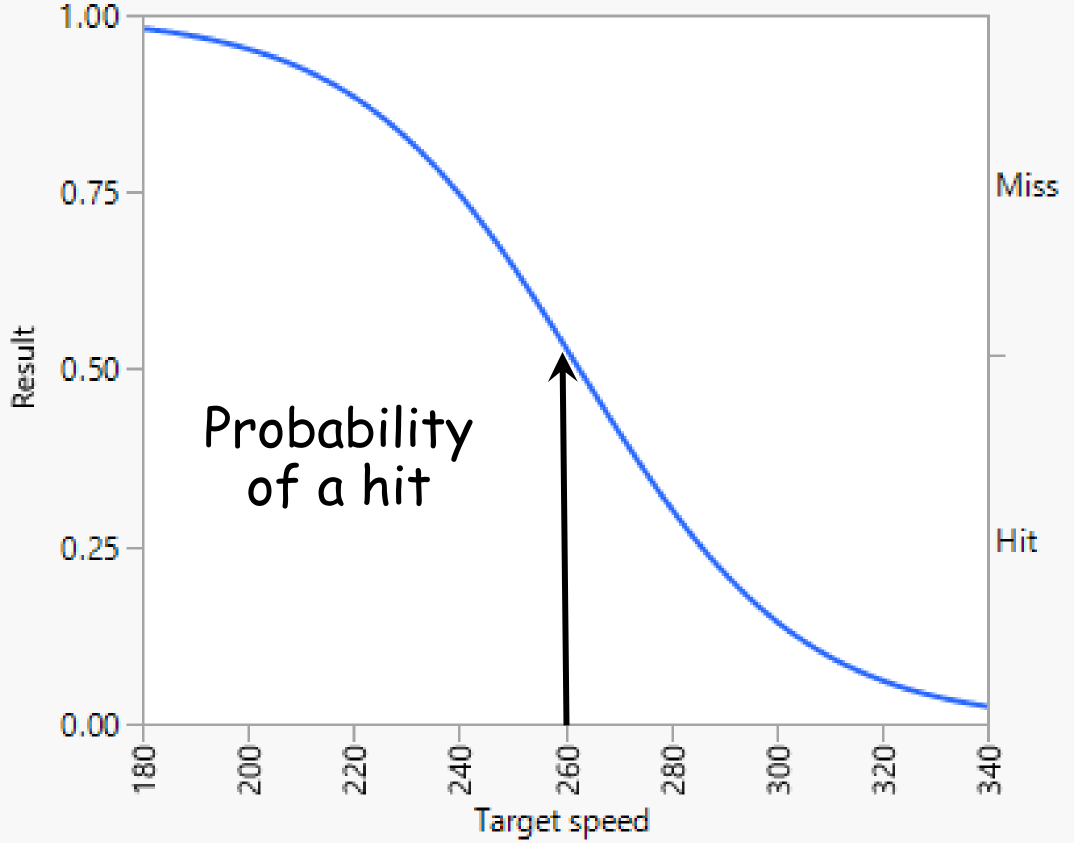
↓

Set up as shown



Analysis output

Logistic Plot



Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	12.5297022	4.7931154	6.83	0.0089*
Target speed	-0.0476836	0.0181939	6.87	0.0088*

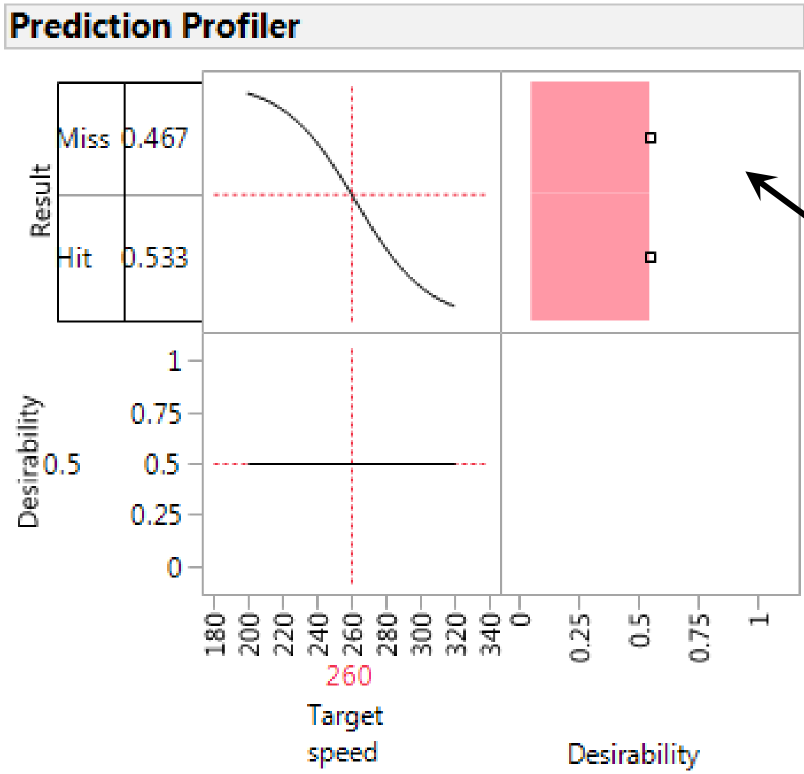
For log odds of Hit/Miss

Effect Likelihood Ratio Tests

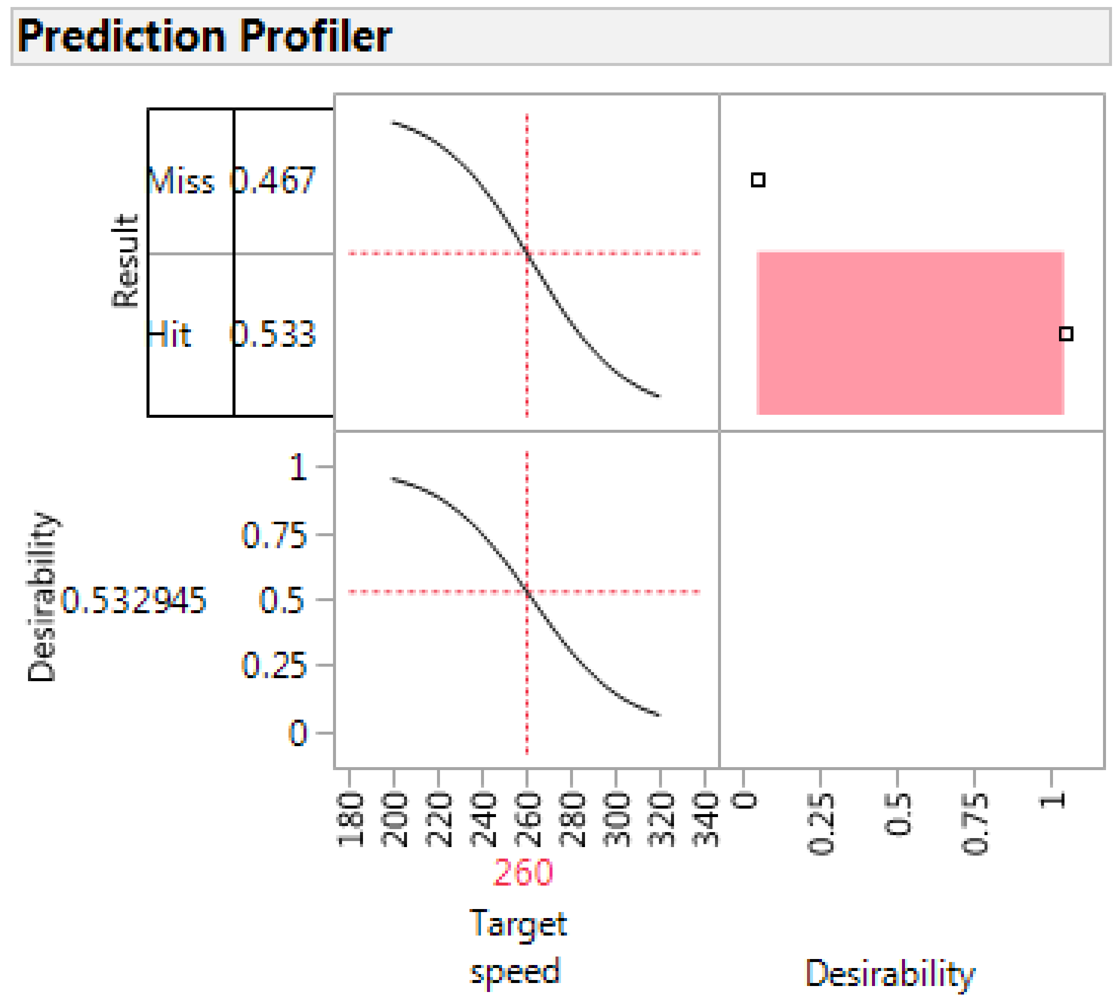
Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Target speed	1	1	11.1939322	0.0008*

- P-value for correlation (this is the one that matters)
- Very strong evidence of a negative correlation between the speed of the target and the probability of hitting it

The prediction profiler

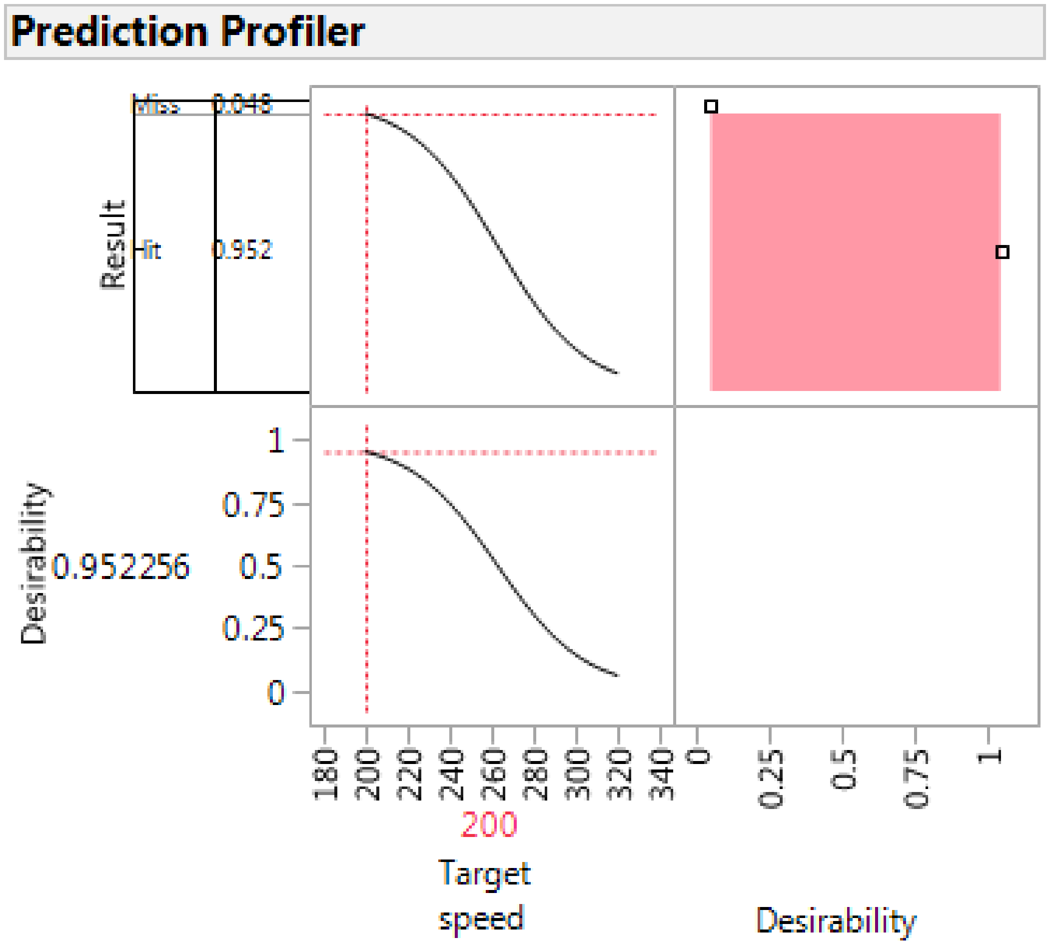


- *Red Triangle* → Profiler → *Prediction Profiler* red triangle → *Optimization and Desirability* → *Desirability Functions*
- Double-click in the blank area, enter 1 for *Hit* and 0 for *Miss* → OK → OK → next slide



Prediction profiler (cont'd)

Prediction Profiler red triangle → *Optimization and Desirability* → *Maximize Desirability*



- The target speed of 200 produces the maximum hit probability of 0.952
- The corresponding miss probability is 0.048
- The target speed of 320 produces the minimum hit probability of 0.061
- The corresponding miss probability is 0.939

Exercise 7.1

Open *Data sets \ quotation process.jmp*.

a) Fit *PO* by *TAT*. Which P-value in the output is the most reliable?

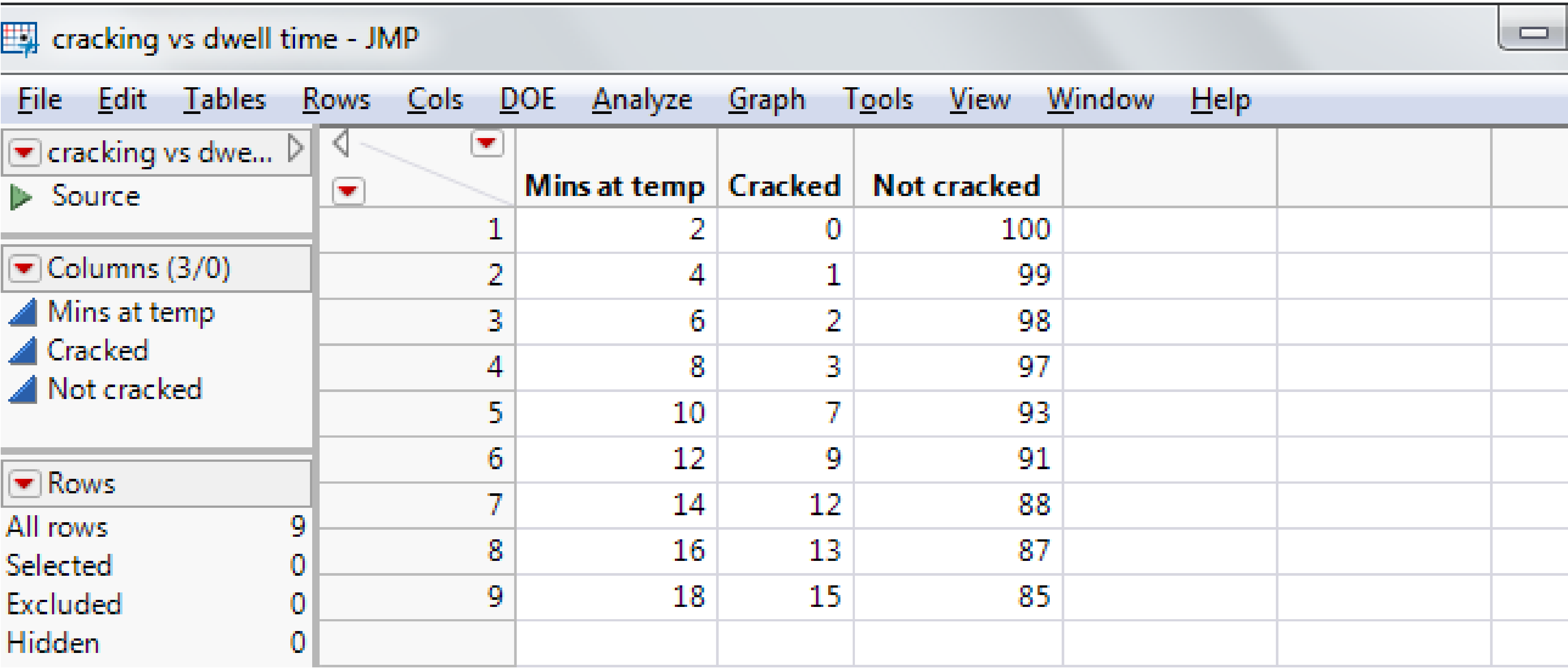
b) Does the PO hit rate increase or decrease as the TAT increases?

c) Find the PO hit rates for 3 day and 15 day turnarounds.

d) Save your script, close and save the data table.

Tabulated data

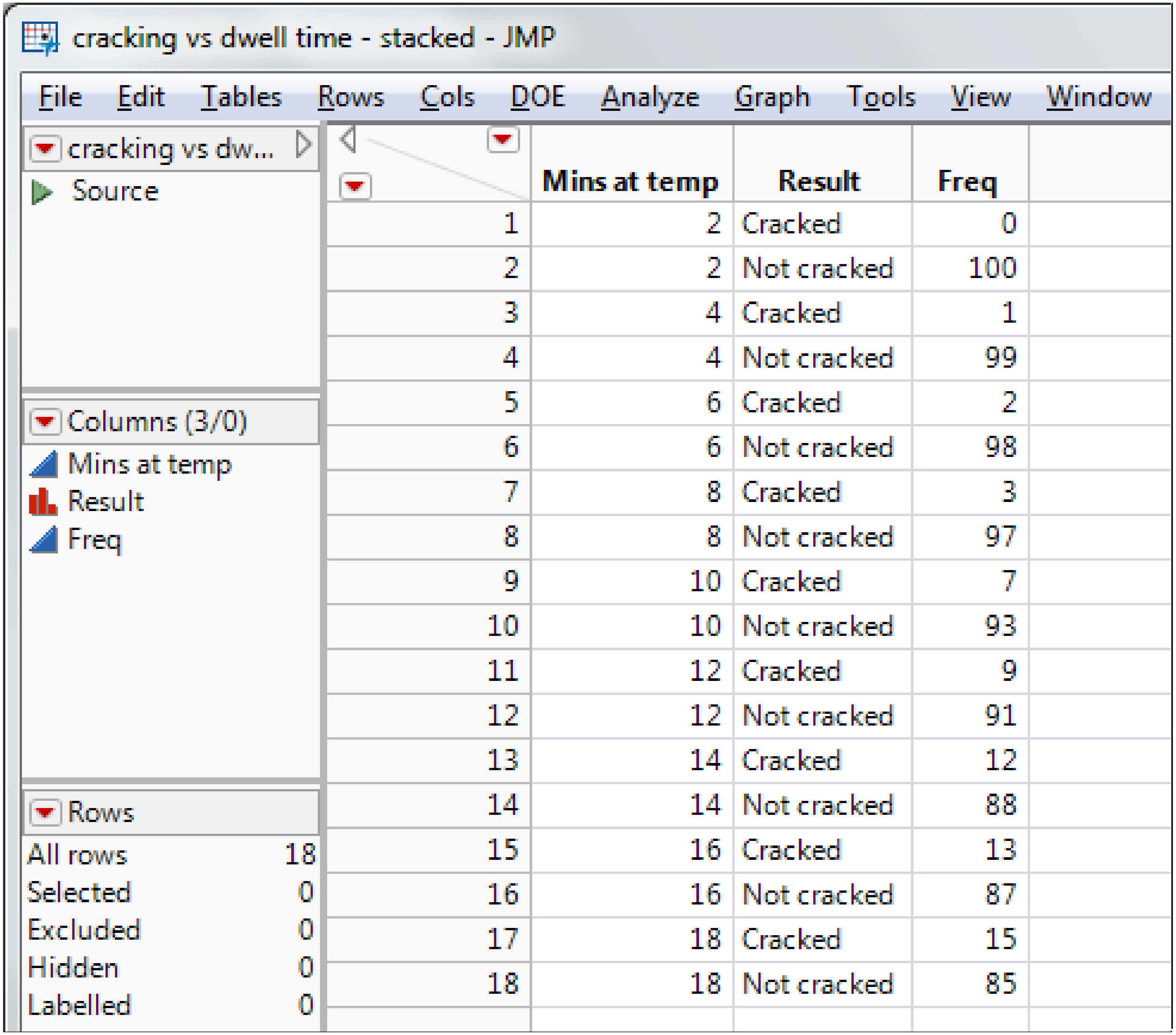
Data sets \ cracking vs dwell time



	Mins at temp	Cracked	Not cracked
1	2	0	100
2	4	1	99
3	6	2	98
4	8	3	97
5	10	7	93
6	12	9	91
7	14	12	88
8	16	13	87
9	18	15	85

1) Tables → Stack
2) Use *Cracked* and *Not cracked* as the stack columns
3) Change *Label* to *Result*, change *Data* to *Freq* → OK
4) Save as *cracking vs dwell time stacked*

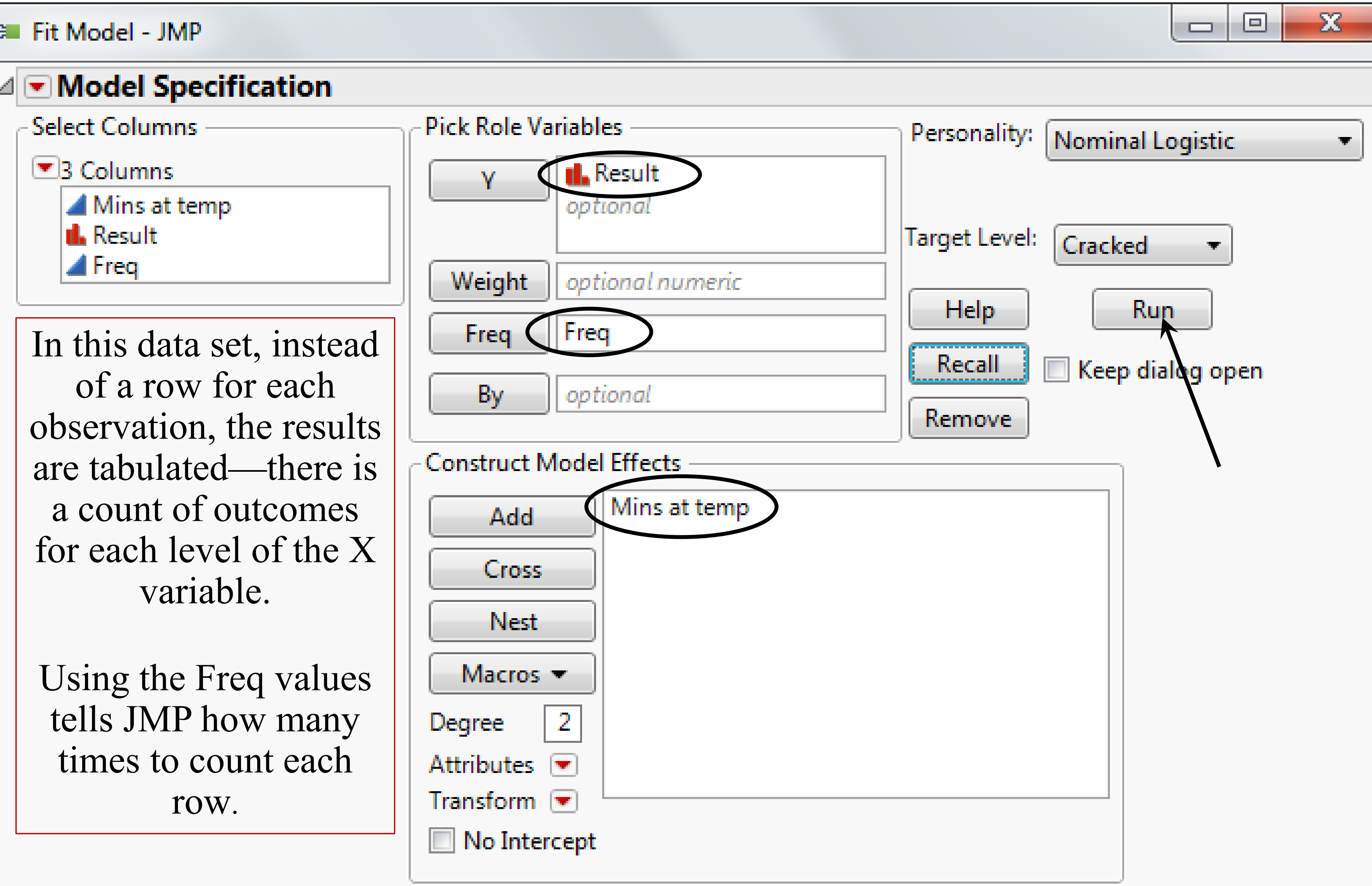
Stacked format



	Mins at temp	Result	Freq
1	2	Cracked	0
2	2	Not cracked	100
3	4	Cracked	1
4	4	Not cracked	99
5	6	Cracked	2
6	6	Not cracked	98
7	8	Cracked	3
8	8	Not cracked	97
9	10	Cracked	7
10	10	Not cracked	93
11	12	Cracked	9
12	12	Not cracked	91
13	14	Cracked	12
14	14	Not cracked	88
15	16	Cracked	13
16	16	Not cracked	87
17	18	Cracked	15
18	18	Not cracked	85

Analyze
↓
Fit Model
↓
See next slide
↓
Set up as shown

Fit Model



In this data set, instead of a row for each observation, the results are tabulated—there is a count of outcomes for each level of the X variable.

Using the Freq values tells JMP how many times to count each row.

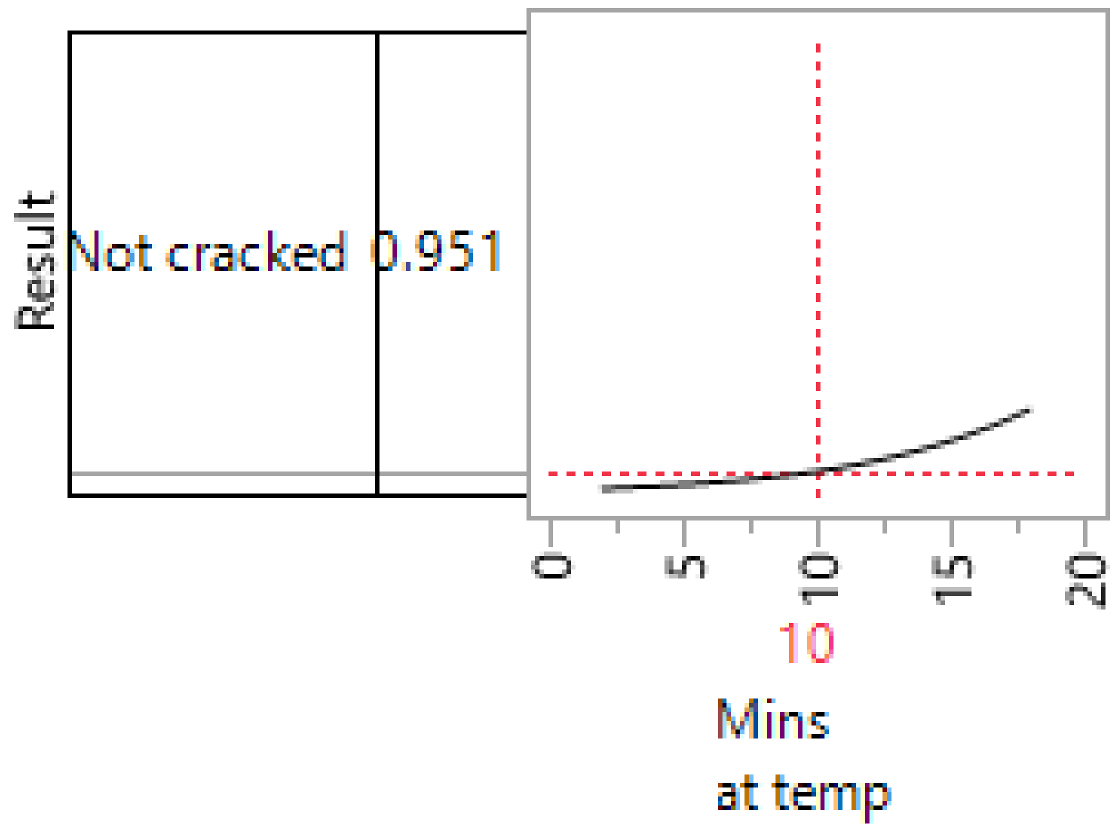
Analysis output

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Mins at temp	1	1	41.5372498	<.0001*

Very strong evidence of positive correlation between dwell time and probability of cracking

Prediction Profiler



Dwell time (mins)	Probability of cracking
5	0.020
10	0.049
15	0.114

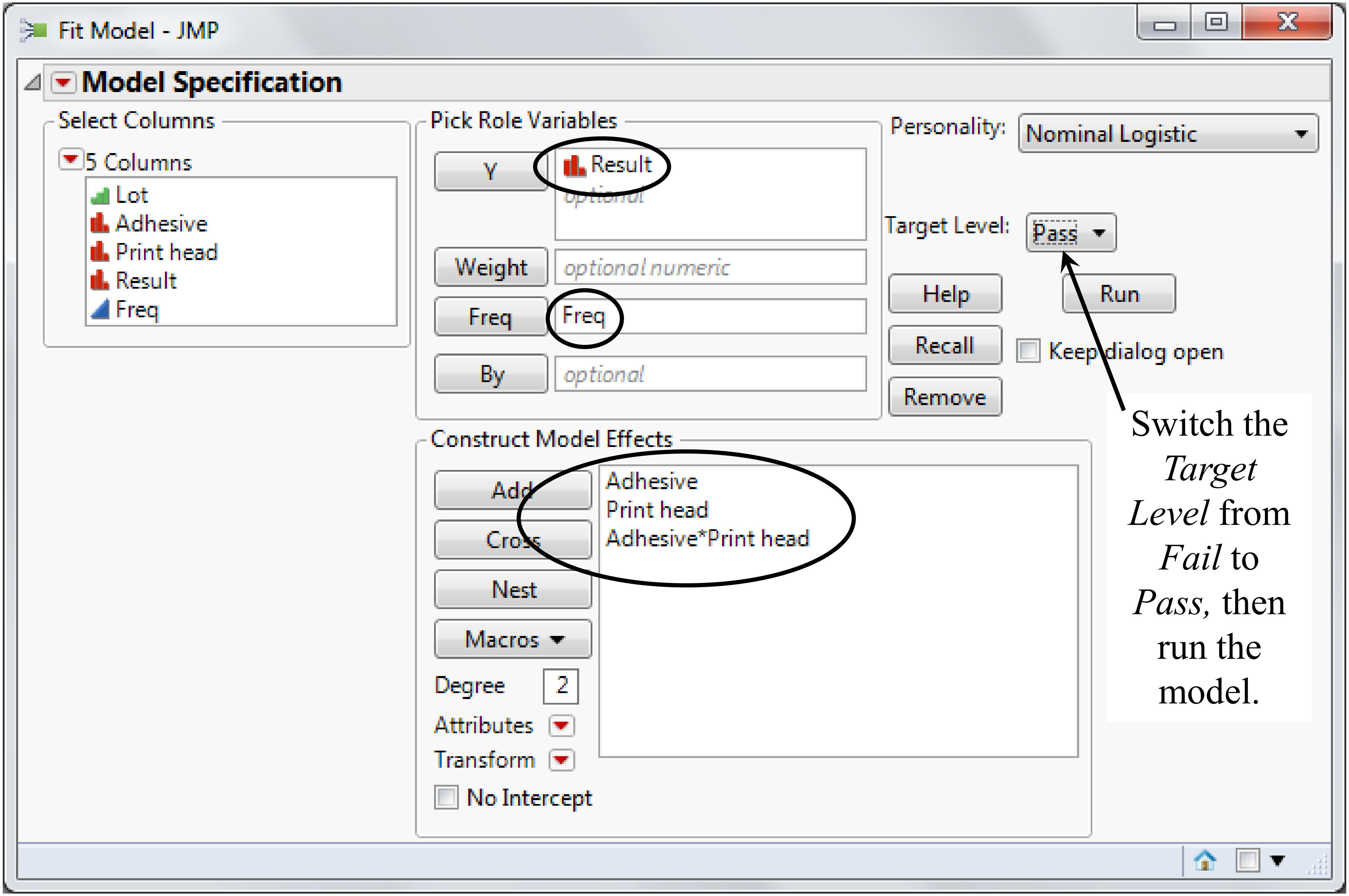
This slide intentionally left blank

8 Multiple Regression with Pass/Fail Y

- Project to reduce clogged nozzles in print heads
- Comparison of four types of adhesive and two print head designs
- Each lot = 60 print cartridges
- "Pass" = no customer detectable print defects
- *Data sets \ clogging pass-fail*
- Run the *Model* script. If necessary, bring the *Model Specification* to the front.

5/0 Cols / 32/0	Lot	Adhesive	Print head	Result	Freq
1	1	A4	D2	Fail	2
2	1	A4	D2	Pass	58
3	2	A4	D1	Fail	1
4	2	A4	D1	Pass	59
5	3	A2	D2	Fail	13
6	3	A2	D2	Pass	47
7	4	A1	D2	Fail	11
8	4	A1	D2	Pass	49
9	5	A3	D2	Fail	4
10	5	A3	D2	Pass	56
11	6	A4	D1	Fail	5
12	6	A4	D1	Pass	55
13	7	A1	D2	Fail	8
14	7	A1	D2	Pass	52
15	8	A2	D1	Fail	3
16	8	A2	D1	Pass	57
17	9	A3	D2	Fail	1
18	9	A3	D2	Pass	59
19	10	A2	D2	Fail	13
20	10	A2	D2	Pass	47
21	11	A2	D1	Fail	1
22	11	A2	D1	Pass	59
23	12	A1	D1	Fail	1
24	12	A1	D1	Pass	59
25	13	A3	D1	Fail	7

Example (cont'd)



Switch the *Target Level* from *Fail* to *Pass*, then run the model.

Example (cont'd)

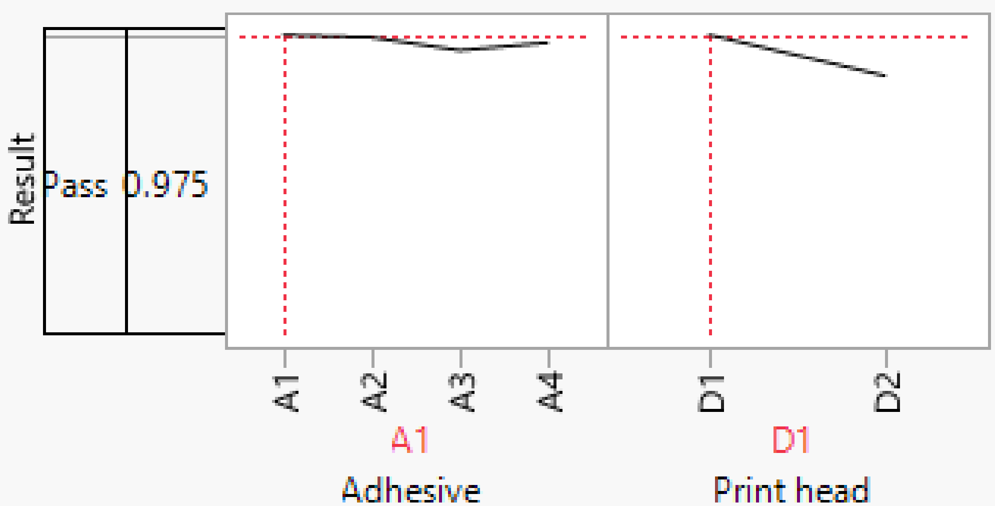
Effect Summary

Source	LogWorth		PValue	
Adhesive*Print head	3.721		0.00019	
Print head	2.254		0.00557	^
Adhesive	0.410		0.38926	^

Effect Likelihood Ratio Tests

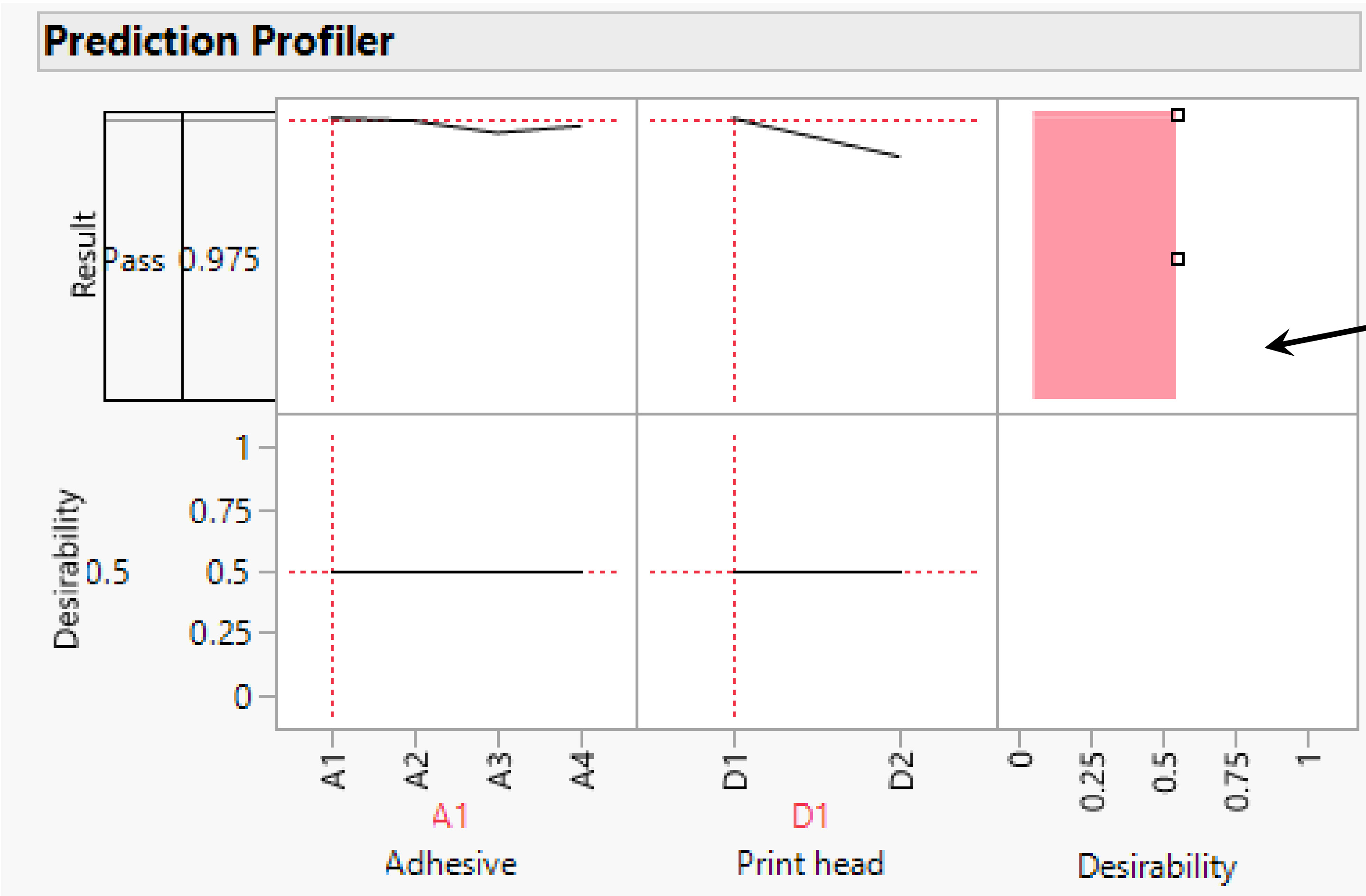
Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Adhesive	3	3	3.01536048	0.3893
Print head	1	1	7.68556658	0.0056*
Adhesive*Print head	3	3	19.7623242	0.0002*

Prediction Profiler



- The *Adhesive* factor was insignificant, but we left it in the model to preserve model hierarchy (Adhesive*Print head is significant)
- On the *Prediction Profiler* red triangle select *Optimization and Desirability* → *Desirability Functions*
- See next slide

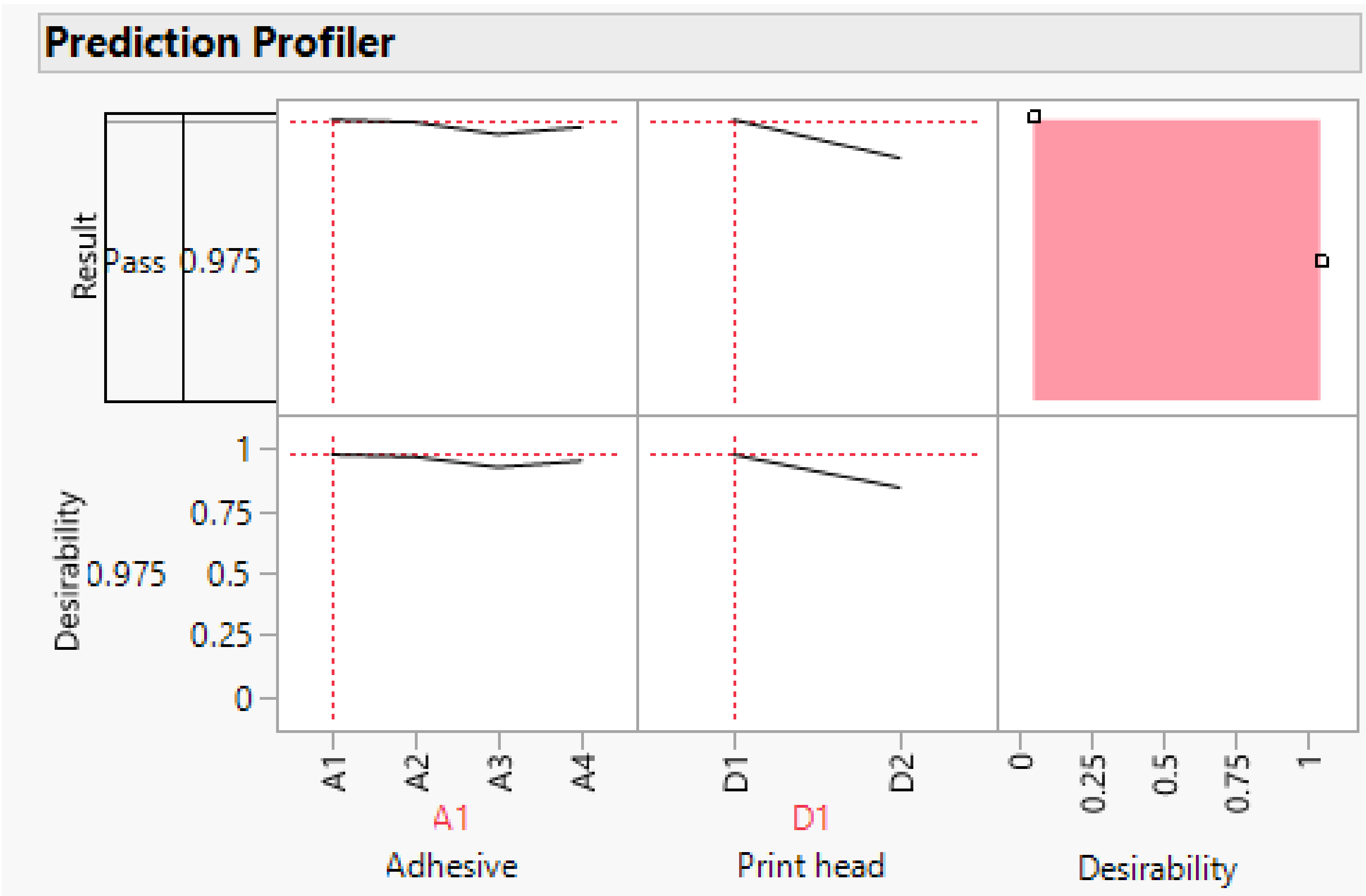
Example (cont'd)



- Double-click in the blank area
- Enter 1 for *Pass* and 0 for *Fail* → OK → OK

Example (cont'd)

- *Prediction Profiler* red triangle → *Optimization and Desirability* → *Maximize Desirability*
- The failure rate predicted from the optimization was 0.025 or 2.5% (current state failure rate was 20% or more)
- Best combination was D1 with A1



Exercise 8.1

A Black Belt wants to minimize the occurrence of bubbles and ripples in the urethane coating on truck nameplates. The X variables and ranges are:

➢ Badge temp:	20 to 40
➢ Mixing ratio:	92.6 to 94.6
➢ Curing temp:	30 to 55

- *Data sets \ urethane coating pass-fail*
- Run the *Model* script in the left panel. In the *Model Specification*, switch the *Target Level* from *Fail* to *Pass*, then run the model.
- Remove insignificant terms from the *Effect Summary* (P > 0.15).
- Use the *Prediction Profiler* to find a factor combination that maximizes the yield.
- The current state yield was about 95%. What is the predicted yield for the improved process?

Tab 3
Design of Experiments

This Page Intentionally Left Blank

1 Designed Experiments vs “File Cabinet” Data

All experiments are experiences, but not all experiences are experiments. – R. A. Fisher

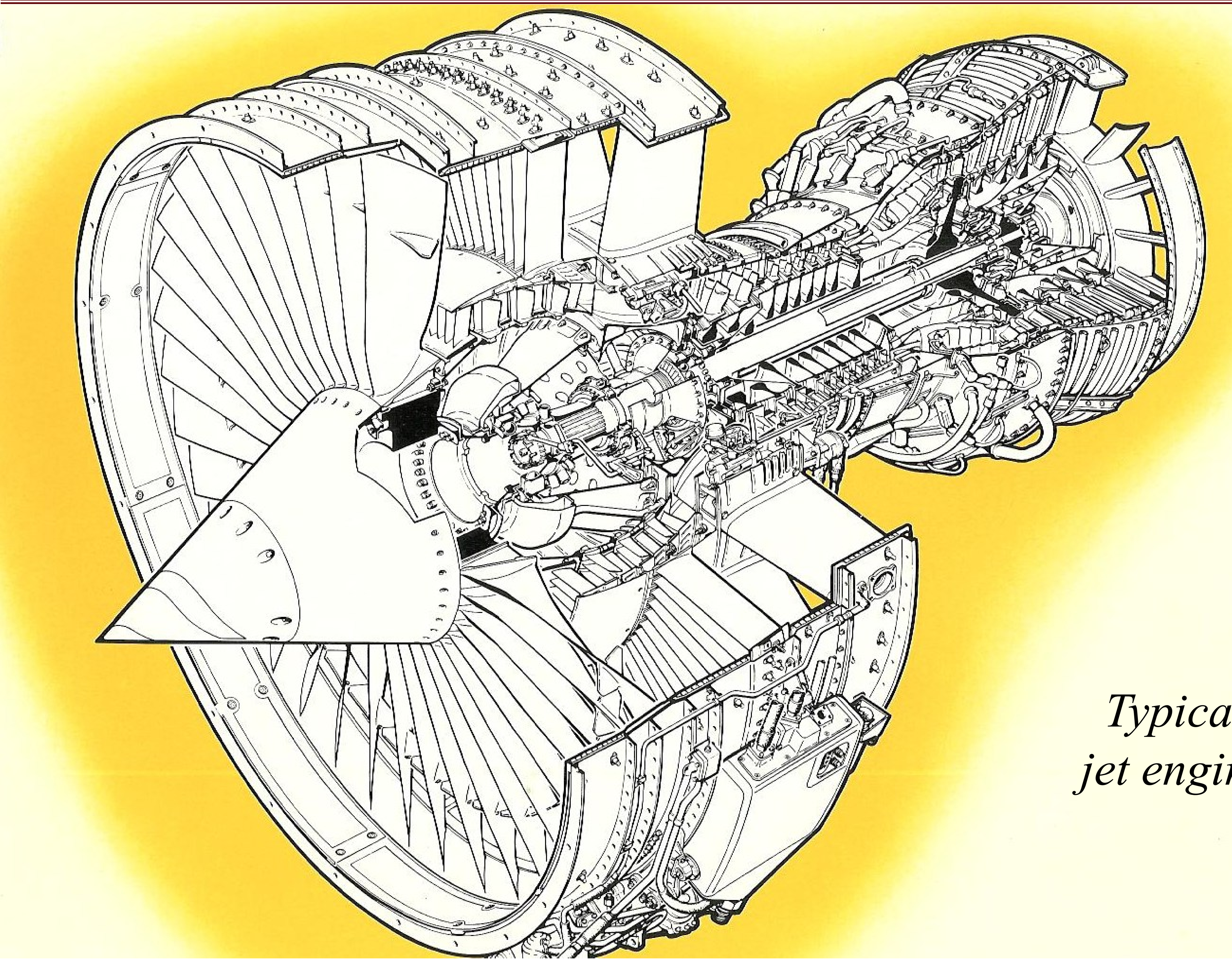
	File cabinet data	DOE
Data sets	Larger, “messy”	Smaller, “clean”
Data collection	Routine operation	Controlled conditions
Information provided	Correlations	Cause and effect
Interactive effects?	Maybe	Definitely
Time period covered	Longer	Shorter

Ronald Fisher was an English geneticist and mathematician trying to increase crop yields in the 1920s. There were limited numbers of plots available for field trials, gradients in the soil, variable proximity to water sources, differing amounts of sunlight, and long lead times. To solve these problems, Fisher developed a body of statistical methods known as Design of Experiments (DOE).

During World War II, Fisher's techniques were extended and applied to military optimization problems. After the war, they were further extended and applied to industrial problems like improving the quality and reliability of manufactured products. For his lifelong contributions to science and statistics, Dr Ronald Fisher eventually became Sir Ronald Fisher.

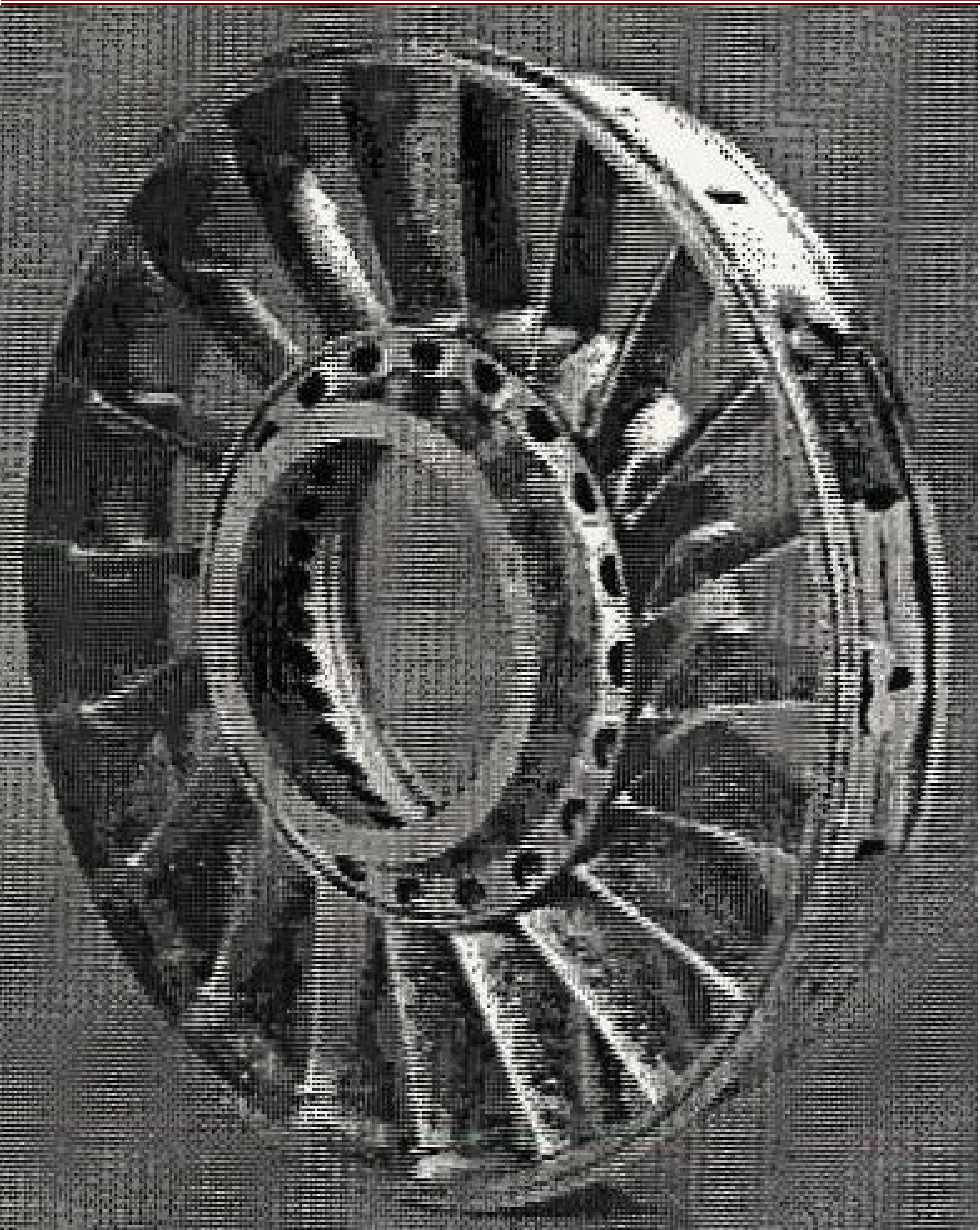
The quote above was Fisher's way of emphasizing the difference between observational studies (analysis of "file cabinet" data) and designed experiments. This distinction is as important today in Six Sigma as it was a century ago in agriculture. After all, both are concerning with increasing yields!

Case study: structural jet engine components



Typical jet engine

Case Study: Typical structural component of jet engine

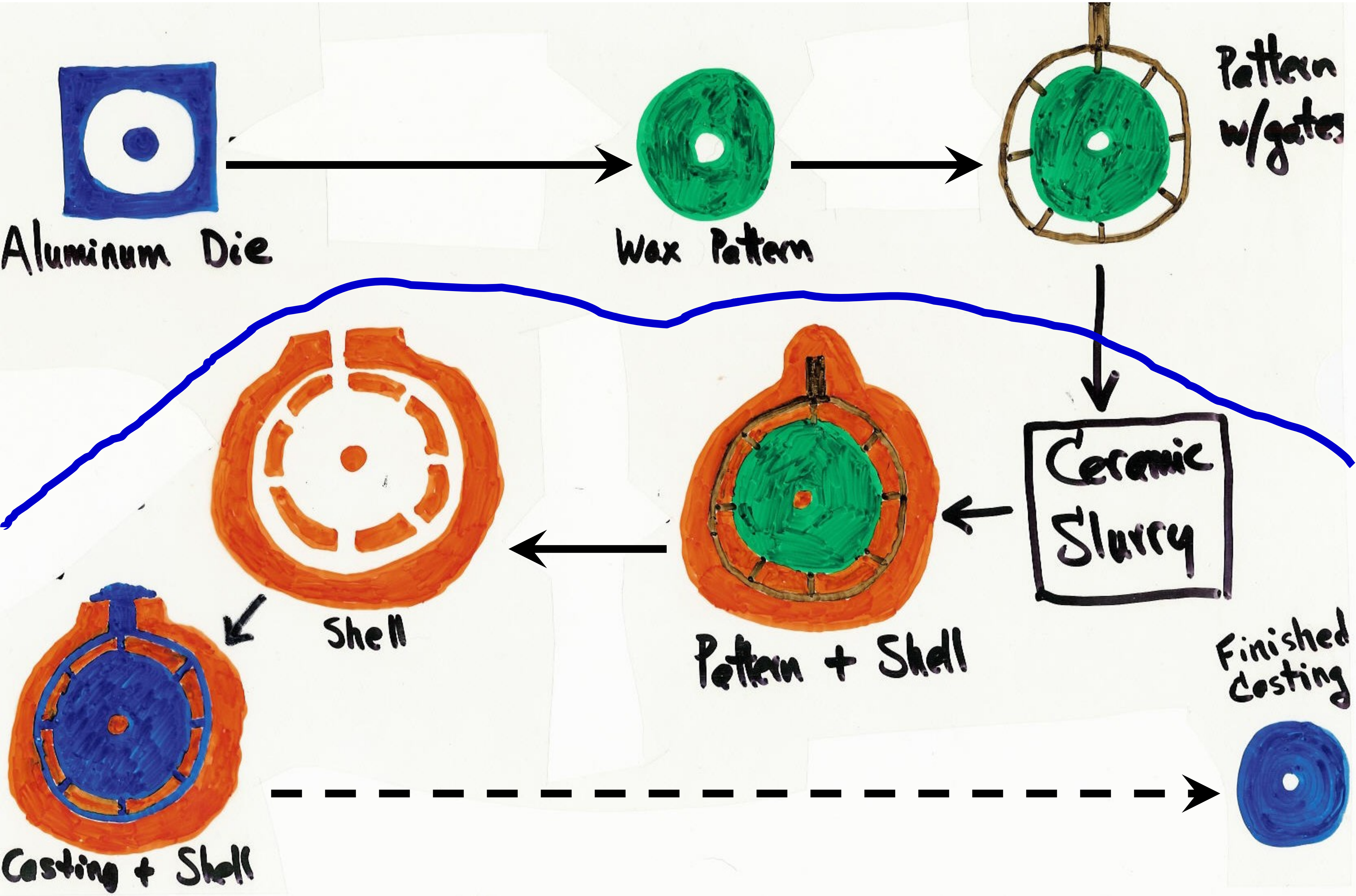


- Back in the day: many small pieces welded together
- Now: one piece casting
- 3 to 6 feet in diameter
- Stainless steel, nickel alloys, titanium alloys

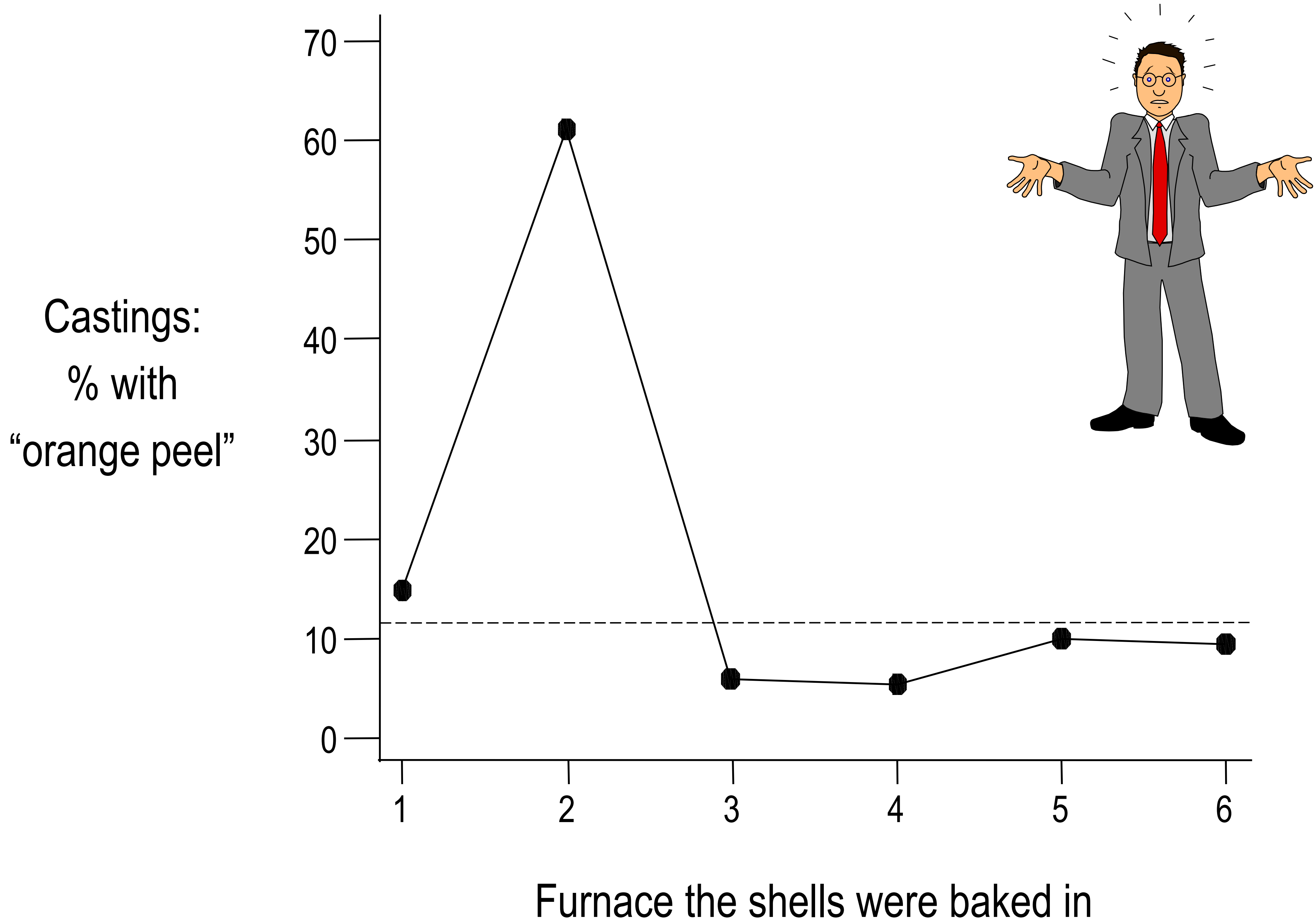
Case study (cont'd)

- Value stream: investment casting of nickel alloy structural components
- Process boundaries: shell making through backend processing
- Experiencing "orange peel" surface condition violating customer smoothness requirements
- 12% scrap rate (big parts $\rightarrow$ big \$\$)
- $Y = f(X)$: analyze existing production data

Investment casting process



A big signal

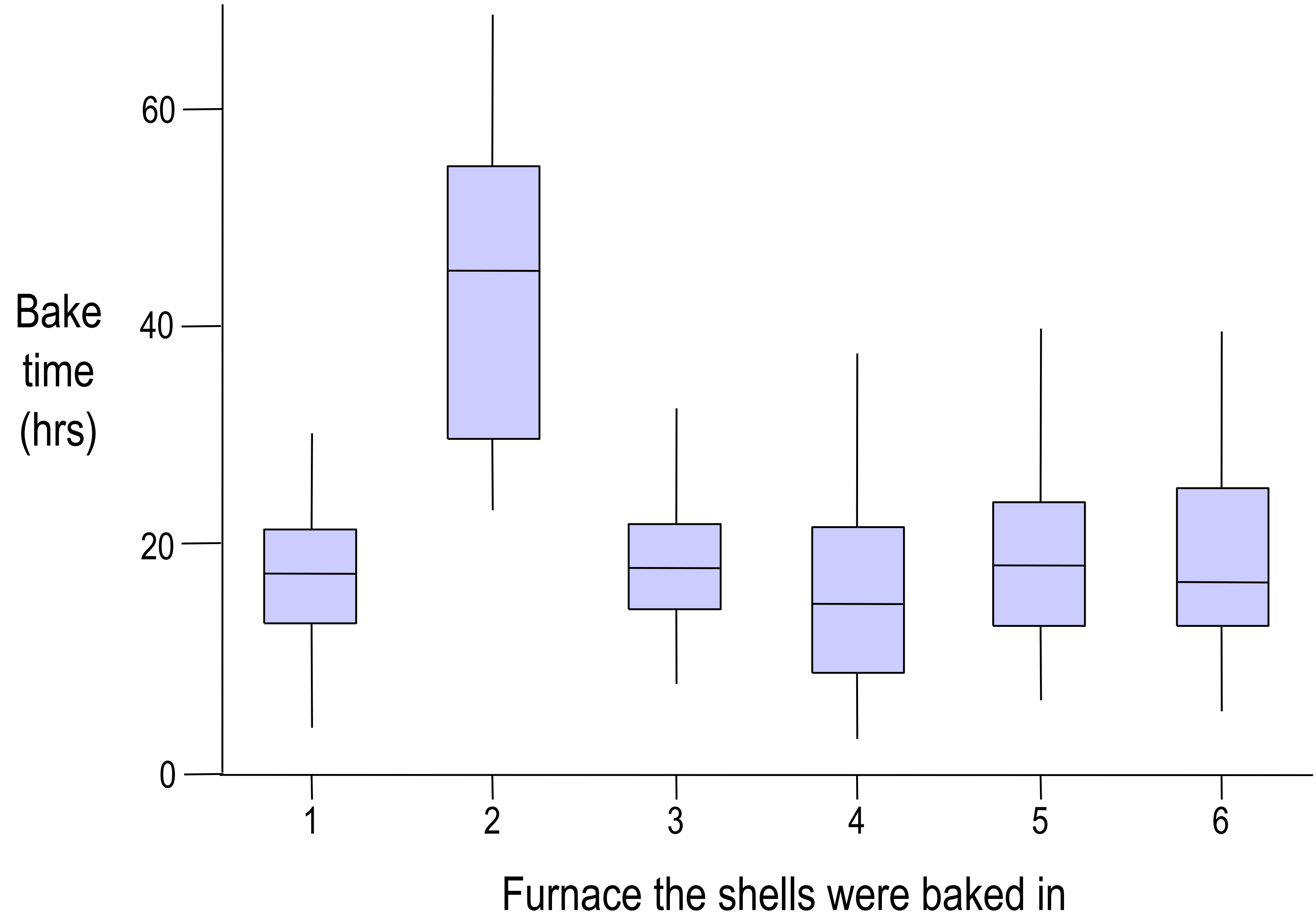

Castings:
% with
“orange peel”
70
60
50
40
30
20
10
0
1
2
3
4
5
6
Furnace the shells were baked in

The strongest correlation in the database involved one of the pre-heat furnaces used to bake the ceramic shells before transfer to the casting furnace. Furnace 2 was new and had come online just about the same time orange peel started occurring. Almost everyone agreed the new furnace was the problem.

The casting area manager refused to take Furnace #2 off-line. He needed all six pre-heats to keep the casting furnace running nonstop so he could meet his production quotas.

Process Engineer Dave (shown above) was skeptical that Furnace 2 was causing the problem. For one thing, the other pre-heats were also producing scrap castings. Also, he had spent the better part of the past three months evaluating and qualifying the new furnace.

Another big signal



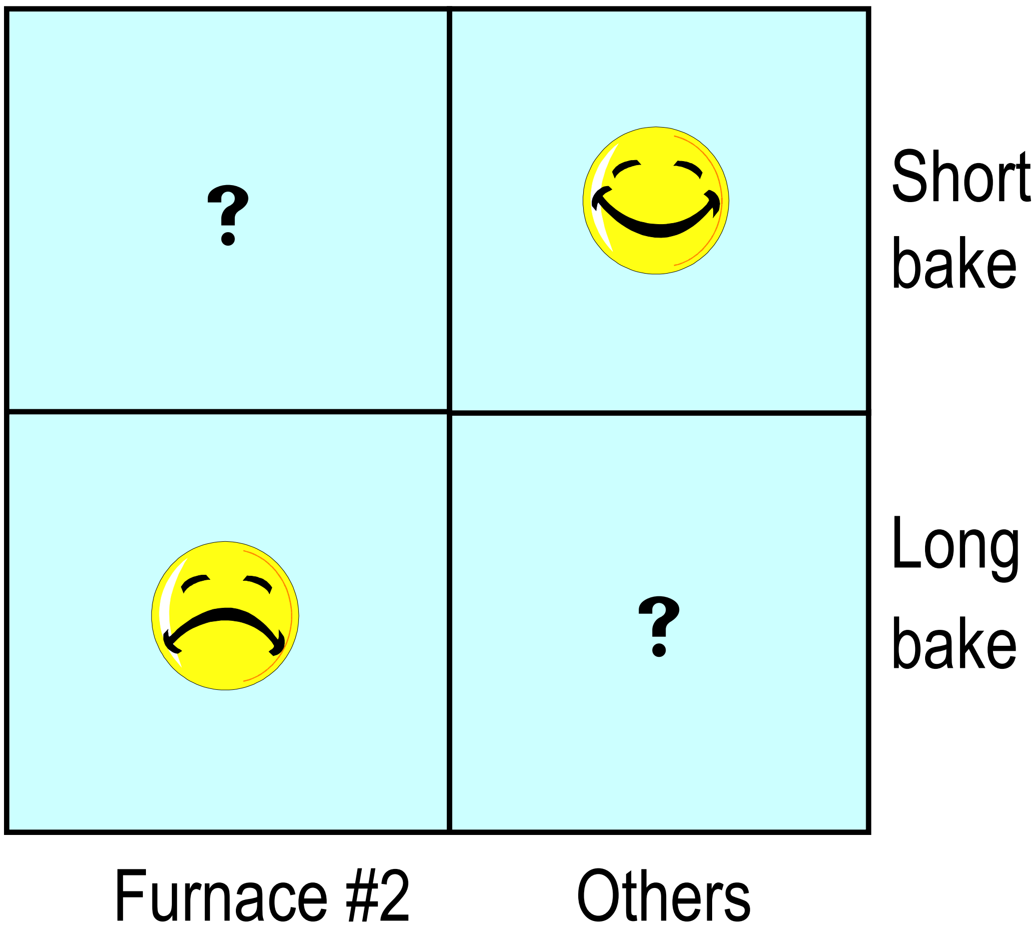
Dave pointed out that the shell bake times were much longer for Furnace 2 than for the other furnaces. There was a minimum required bake time, but no upper limit. Dave’s theory was that orange peel was caused by long bake times.

Why did shells stay longer in Furnace 2?

It turned out there wasn’t room to put the new furnace next to the original five, so it had to be located further away from the casting furnace. The fork-lift operators wouldn’t drive over there unless they had no shells ready from the closer furnaces, so shells tended to sit in Furnace 2 for a long time.

Autopsy

- The file cabinet data suggested some plausible hypotheses
- It could not establish the cause of the defect
- The *quantity* of data was not the problem
- The data lacked the *structure* required to determine cause and effect



There was lots of data in the upper right-hand and lower left-hand cells in the table above, but virtually nothing in the other two cells. Making sure that data tables like the one above are completely filled out is one of the basic principles of experimental design.

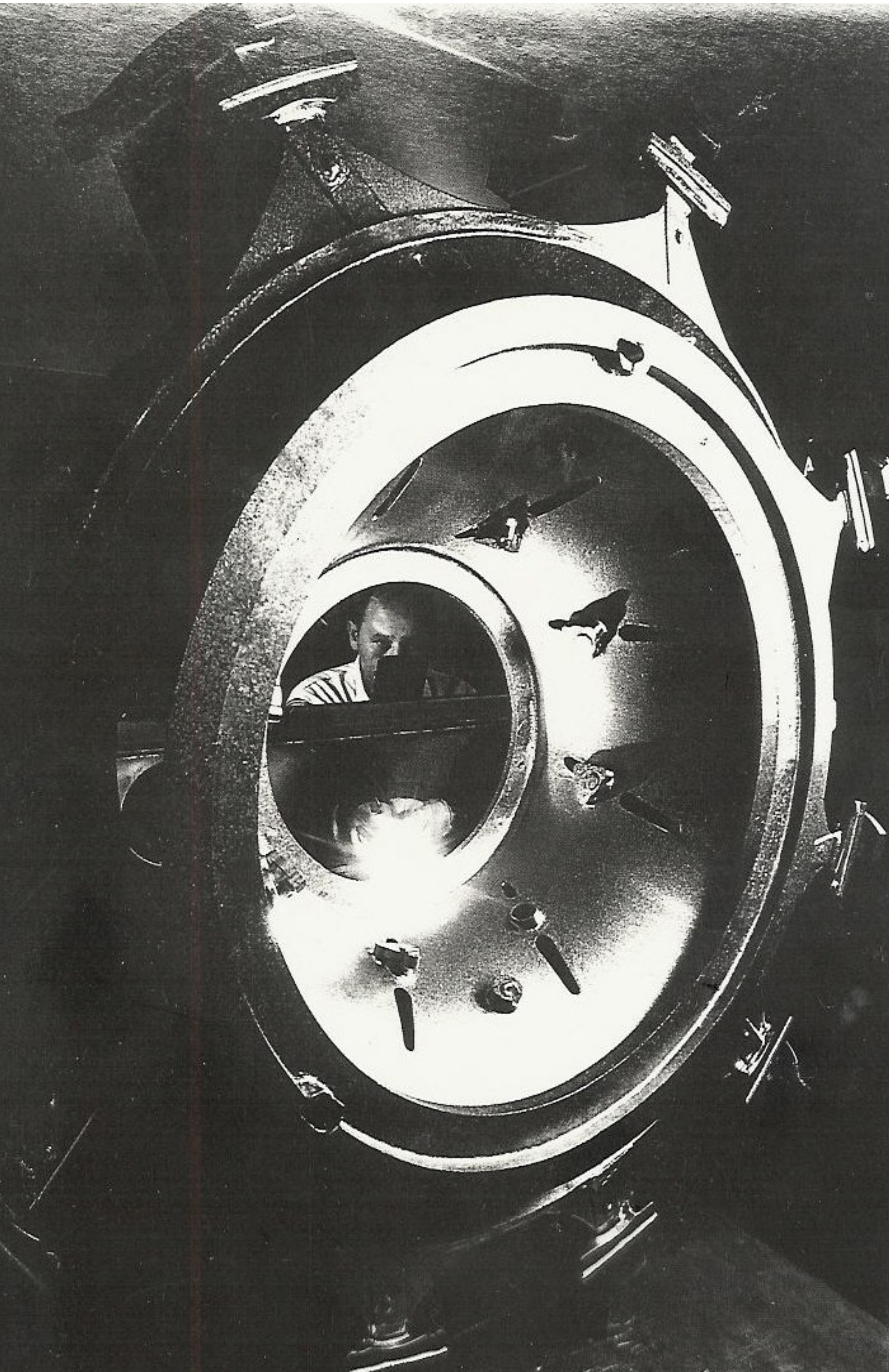
Subsequently, engineers ran enough parts in the upper left-hand corner of the table to determine that long bakes were indeed causing the problem. An upper limit on the bake time was developed and put in place. Shells that exceeded this limit were scrapped. This cost the company much less than scrapping the resulting castings.

The new procedure made the fork-lift operators' job harder, but it made the orange peel problem go away.

The Role of DOE in Process Improvement

Y = *f*(X) analysis	• DOE is an effective way to collect data for identifying critical x's, in a relatively short period of time • In a Lean Six Sigma project, data collection in the Measure phase may have produced little or no useful information.
Developing the future state	• May have multiple potential improvement ideas on the table • DOE is an effective way to evaluate these ideas prior to defining the future state

Example



- Titanium castings → strong & light
- Ti develops surface oxidation during the cooling phase
- Large Ti castings were failing the customer O_2 requirement
- Analysis of file cabinet data yielded no significant correlations
- Engineers developed a list of factors for a DOE

Example (cont'd)

Factor	Levels	Current state X variable	Possible future state solution
Slurry for shell	Batch 1 vs Batch 2	✓	
Shell thickness	14 dips vs 18 dips		✓
Shell bake time	6 hrs vs 48 hrs	✓	
Shell bake temp	1950° vs 2050°		✓
Alloy grade	Low $ vs High $		✓
Alloy status	New vs Revert	✓	
Heat shield steel	Mild vs SS		✓
Cooling fan speed	2400 vs 3200		✓

This slide intentionally left blank

2 One Factor at a Time?

- In this approach, each factor is varied with all others held constant. This way, it is felt, we can see the "pure effect" of each factor.

- This is one way to apply the scientific method, but it is not the only way, and **not the best way**!

- For any proposed one at a time experiment, there is usually a multifactor experiment providing:
  - ✓More information
  - ✓Better results
  - ✓Same (or possibly smaller) total sample size

- One at a time trials *are* useful for determining feasible ranges for factor in a DOE

Example: potato chip bags

- The current average bond strength of our potato chip bags is 86 psi

- Based on customer complaints, we need to increase the bond strength

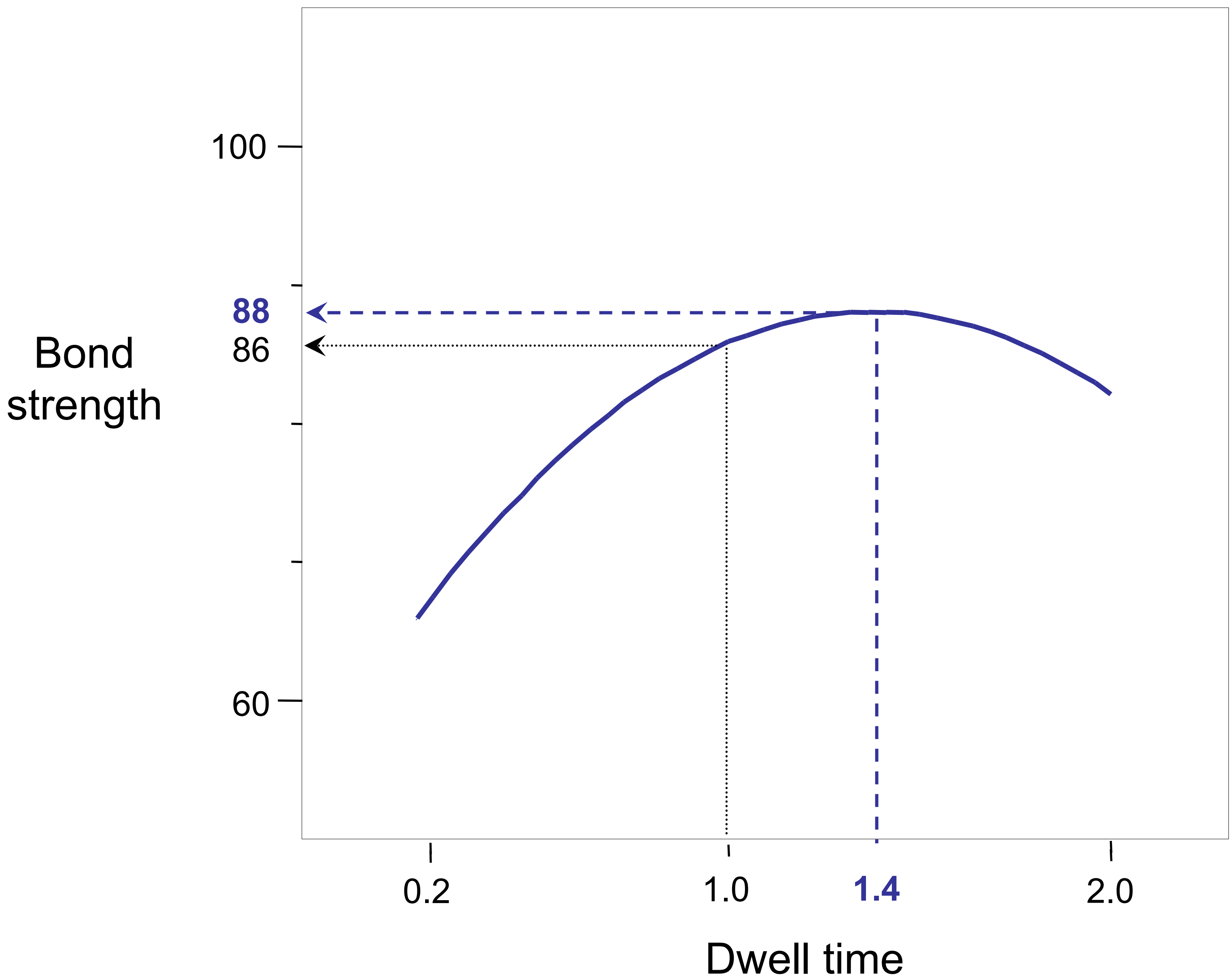
- The most important control factors in the bag sealing operation are *temperature* and *dwell time* (see below)

- Secondary objective: decrease the *dwell time* if possible

Factor	Current level	Feasible range
Temperature	150°	120 to 180
Dwell time	1.0 secs	0.2 to 2.0

One-at-a-time experiment #1

Vary *dwell time* over its feasible range while holding *temperature* at 150



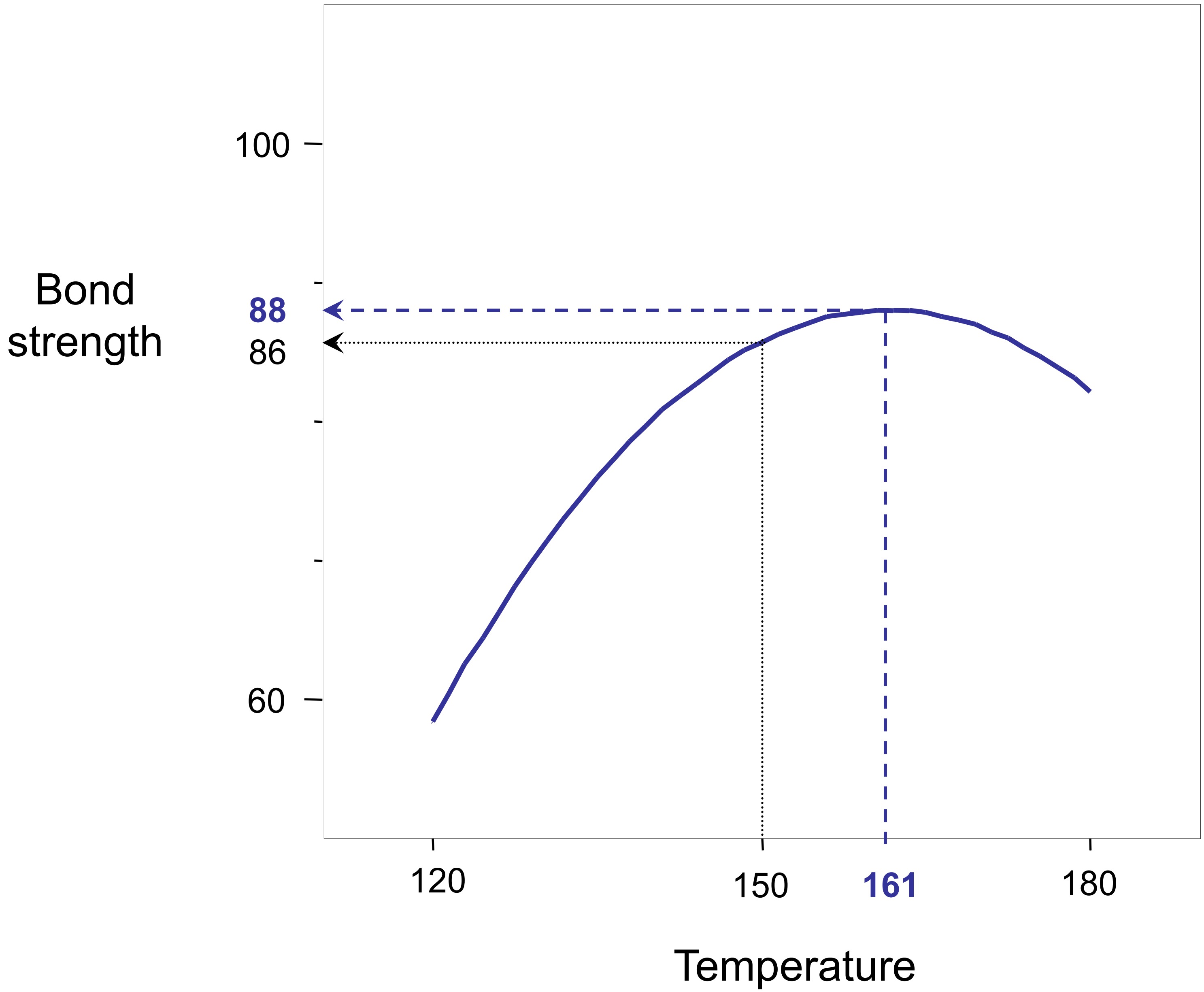
N = 9

Our process engineer Chip Kettle first studies the effect of dwell time while holding temperature constant. He seals and tests 9 bags using dwell times ranging from 0.2 to 2.0. Chip finds he can increase the bond strength by 2 psi by increasing the dwell time to 1.4.

Our production manager Justin Thyme is not pleased with the prospect of a 40% increase in dwell time.

One-at-a-time experiment #2

Vary *temperature* over its feasible range while holding *dwell time* at 1.0

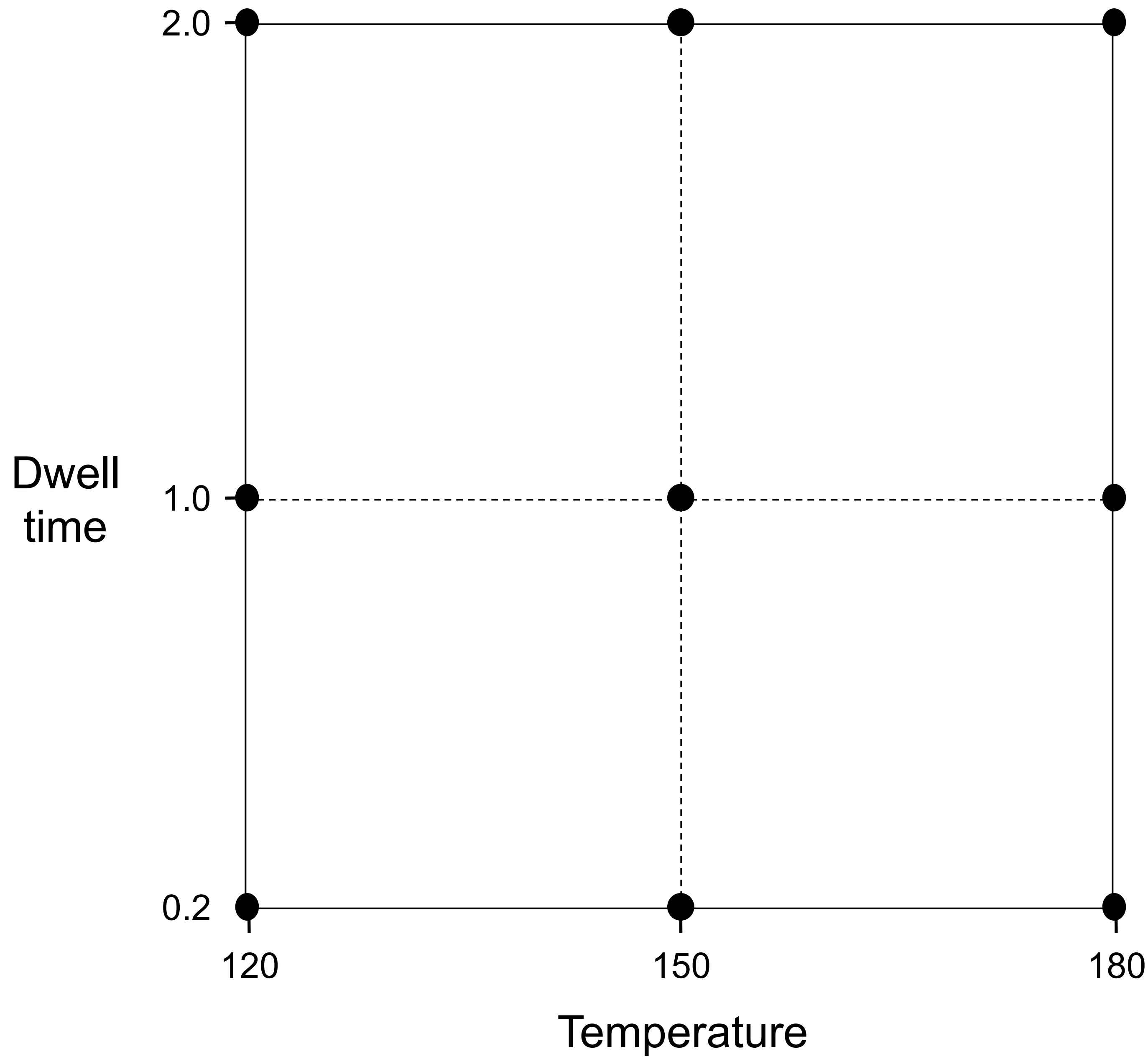


N = 9

Chip now studies the effect of temperature while holding dwell time constant. He seals and tests 9 bags using temperatures ranging from 120 to 180. Chip finds he can increase the bond strength by 2 psi by increasing the temperature to 161.

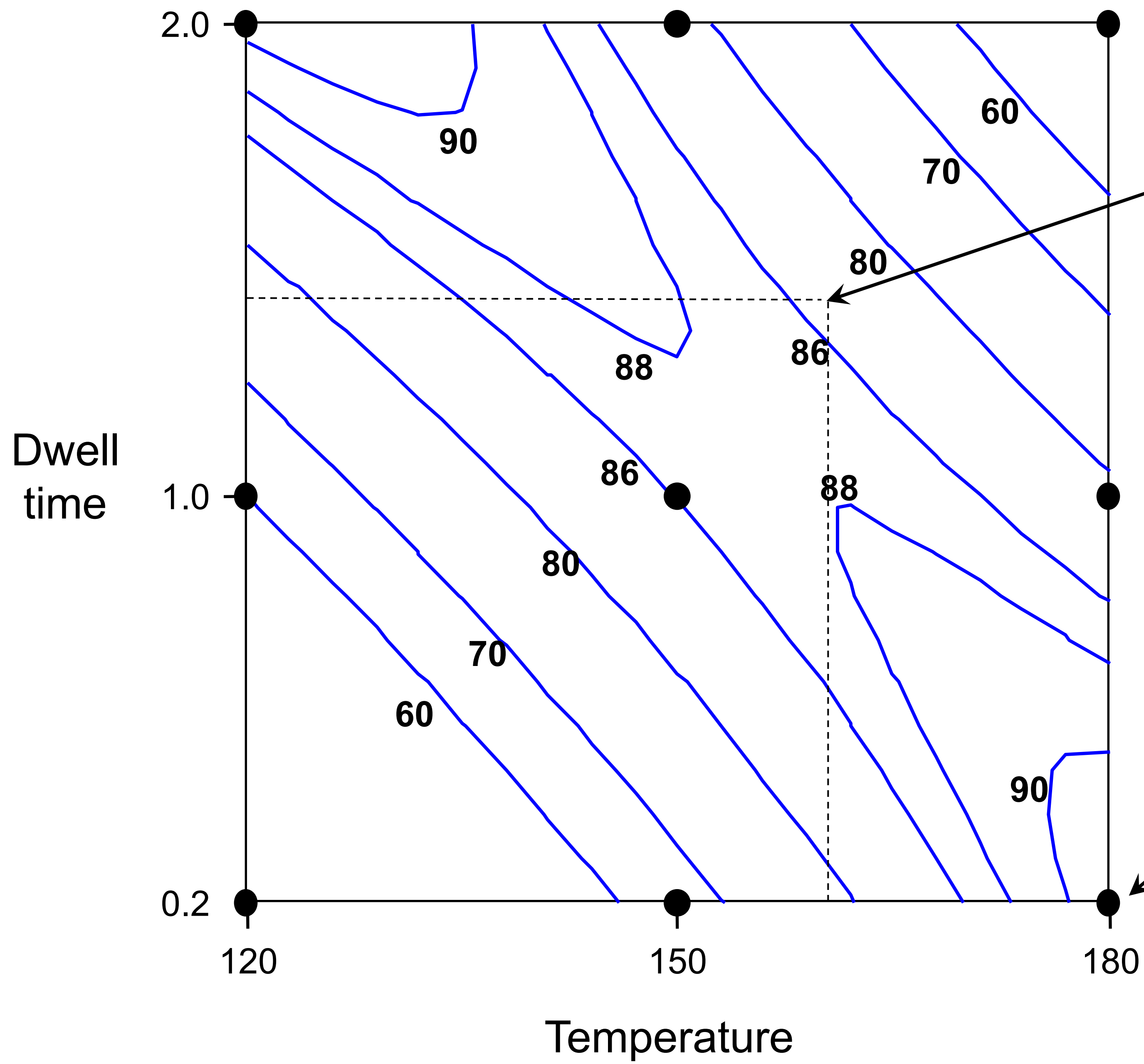
Chip predicts that changing the dwell time to 1.4 and the temperature to 161 will increase the average bond strength by 4 psi (2 + 2). However, it is highly likely that Justin will oppose the increase in dwell time, in which case the increase in average bond strength will be only 2 psi.

The multi-factor approach



- ✓ 9 design points (●)
- ✓ 2 bags sealed at each point
- ✓ Total sample size: N = 18

Contour plot of predicted average bond strength

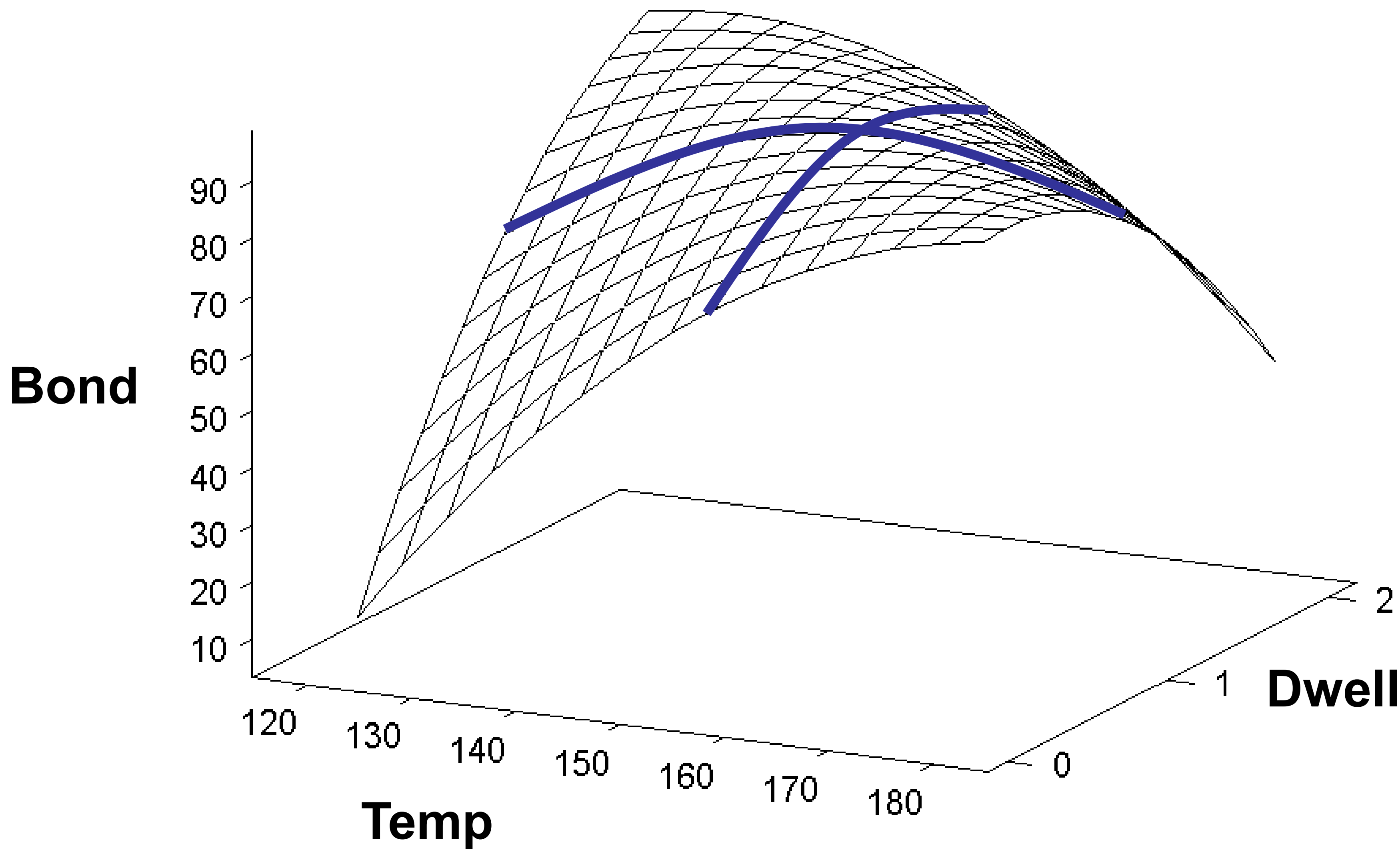


Chip's prediction of 90 psi at 161° and 1.4 secs was way off

Bond strength exceeding 90 psi at 180° and 0.2 secs

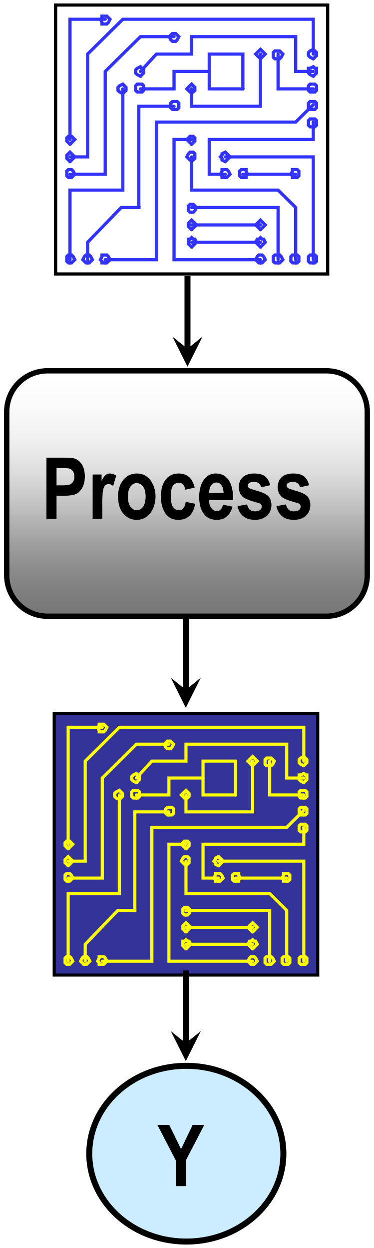
Why one-at-a-time doesn't work

The 3D perspective



When we experiment with all factors, but one held constant, we optimize sequentially over one-dimensional profiles. The sequence of solutions generated by this process is highly dependent on the starting point. It has very little chance of finding a global optimum, and often fails to move a significant distance from the starting point.

3 DOE Terminology



Experimental unit

The outcome of a single application of the process being studied

Sample size

The total number of experimental units ("number of runs")

Response variable

A Y variable measured or inspected on each experimental unit

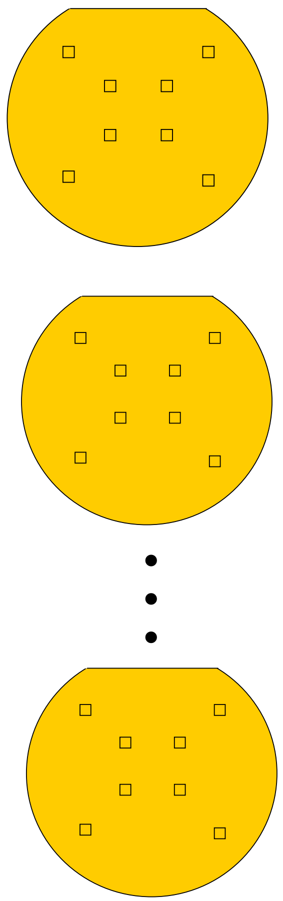
The experimental unit is often a part, lot, batch or single transaction of some kind. It may also be a test specimen or sample of material. It is important to identify the experimental unit—it provides the basis for counting sample size, and sample size is critical in determining the statistical significance of the results.

The experimental unit is determined by the process on which we are experimenting, not the measurement plan used to evaluate the results. For example, suppose we test 100 devices for product life. Suppose we measure a degradation parameter on each device every 10 hours until the end of the test at 100 hours. The sample size for the study is the number of units (100), not the number of measurements (1000).

Example

- 11 silicon wafers were subjected to vapor deposition at various temperatures, pressures, and Argon flow rates
- The thickness of the resulting layer was measured at 8 locations on each wafer
- What is the sample size?

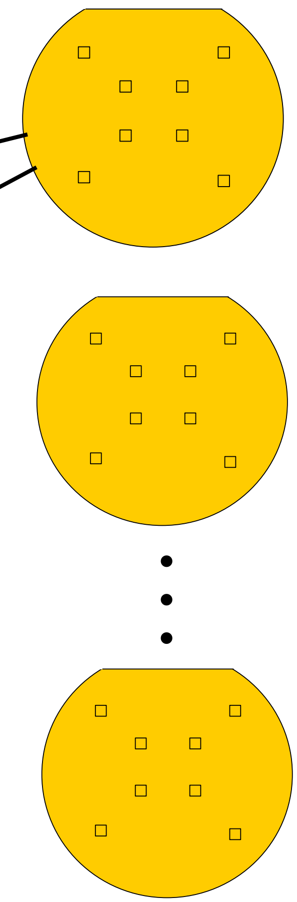
Temp	Press	Flow	Thickness
180	0.3	30	
180	0.3	30	
180	0.3	30	
160	0.4	10	
160	0.4	50	
160	0.2	50	
160	0.2	10	
200	0.4	10	
200	0.2	10	
200	0.2	50	
200	0.4	50	



Example (cont'd)

- The sample size is the number of experimental units, not the total number of measurements taken
- The response variables of interest may be statistical summaries of multiple measurements on each unit

Temp	Press	Flow	Avg.	Std. dev.
180	0.3	30		
180	0.3	30		
180	0.3	30		
160	0.4	10		
160	0.4	50		
160	0.2	50		
160	0.2	10		
200	0.4	10		
200	0.2	10		
200	0.2	50		
200	0.4	50		



DOE terminology (cont'd)

Factor

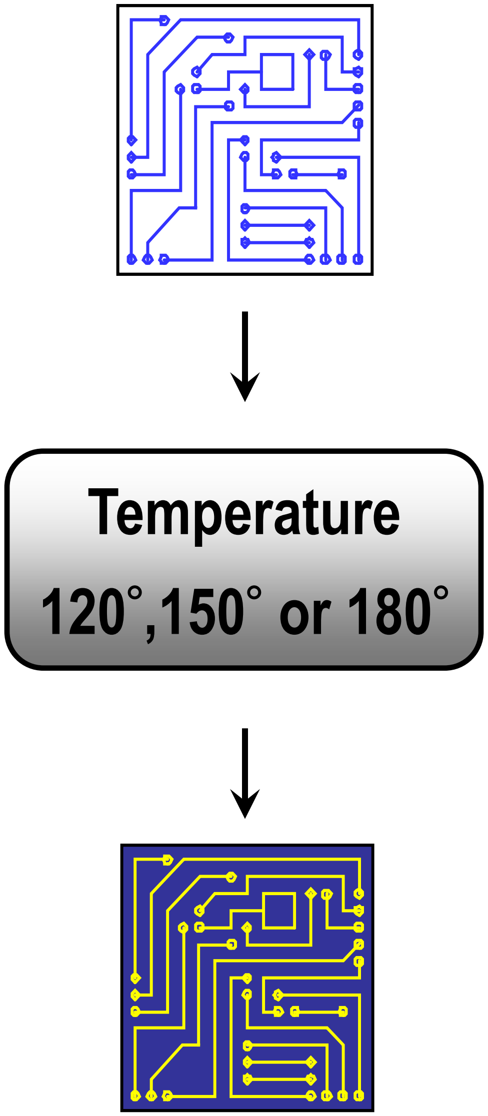
An X variable controlled in an experiment, varied on purpose to determine its effect on the responses

Level

A particular value or setting of a factor to be used in the experiment

Requirements

All levels of each factor must be logically and physically compatible with all levels of the other factors



Variables used as factors in a designed experiment may or may not be controlled in the routine process. What matters is that they can be controlled for the purpose of experimentation.

DOE terminology (cont'd)

Examples of continuous factors

Time	Volume
Temperature	Weight
Pressure	Length
Energy	Width
Voltage	Density
Resistance	Rate
Concentration	RPM
Flow	Intensity . . .

- A factor is *continuous* if it can be varied within some range on a scale of measurement
- It is generally preferable to use 3 equally-spaced levels (low, medium, and high) for continuous factors
- **Even though only two or three levels of a continuous factor will be used in an experiment, it is advantageous to identify it as continuous, rather than categorical**
- Even when some levels of a continuous factor would not be applied to the process after the experiment, it is advantageous to still treat the factor as continuous in the experimental design and analysis
  - Example: After an experiment, we find that the optimal temperature setting is 117.13º. We may choose to set the temperature to 115º or 120º. We still treat temperature as a continuous factor in our experiment.
  - Example: We know that if we determine that the optimal Introductory Time Period for an offer is 3.37 months, it wouldn't make sense to offer that to our customers. We would offer them an Introductory Time Period of 3 months. We still treat this factor as continuous in our experiment.

DOE terminology (cont'd)

Examples of categorical factors

Method	Old or New
Tool set	1, 2 or 3
Material	A, B, C or D
Supplier	X, Y or Z
Operator	Bob, Carol, Ted or Alice
Color	Cyan, Magenta or Yellow
Size	Small, Medium or Large

- A factor is *categorical* if it is not possible to have it at all values on a measurement scale

- Treating a factor as continuous implies that any value in the range can be used in the process

- If the levels used in the experiment are the only possible values, even when the categories are described by numbers, the factor should be treated as categorical
  - Example: Pizza pan sizes of 10”, 12”, 14”, 16” (10.26” doesn’t exist)
  - Example: A control parameters for certain electron microscopes has to be a power of 2.
  - Some JMP DOE platforms now have the option of *Discrete Numeric*, in addition to continuous and categorical, to better handle these cases

DOE terminology (cont'd)

Categorical factors	*Continuous factors*
Any number of levels	Usually 2 - 3 levels
Discrete set of design points	Region in factor space
Test for significant differences	Response surface modeling
Select best design point	Interpolate between design points

DOE terminology (cont'd)

Control factors	*Noise factors*
Can be controlled in the routine process ↓ Type of material Temperature Pressure Method Time ⋮	Cannot be controlled in the routine process ↓ Ambient conditions Raw materials Operators Suppliers Batches Setups Shifts Lots ⋮

Is it good practice to include noise factors in experiments?
Why or why not?

DOE terminology (cont'd)

Design point

A particular combination of levels of the factors.

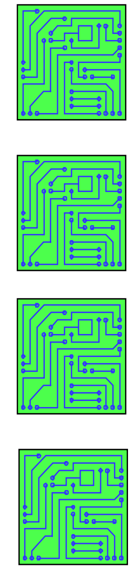
Design matrix

The set and sequence of design points to be used in the experiment.

Full factorial

The set of all possible design points for a given set of factors and levels.

Temp	Press
120	50
120	150
180	50
180	150



Experimental units

- ✓ Full factorial
- ✓ 4 design points
- ✓ No repeats (replication)
- ✓ Sample size = 4

DOE terminology (cont'd)

Replicate run

An experimental unit created independently of other units at the same design point

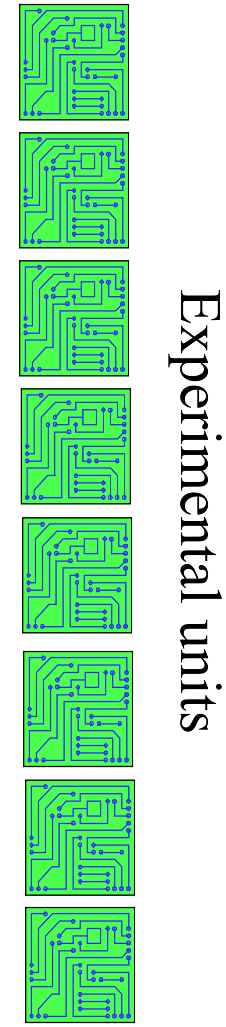
Replicate

A set of replicate runs, one for each unit in a given set (usually a replicate of a full factorial)

False repeat

- Repeated or multiple measurements on one unit
- Units in the same batch, when optimizing a batch process for which there is very little within-batch variation

Temp	Press
120	50
120	150
180	50
180	150
120	50
120	150
180	50
180	150



- ✓ Full factorial
- ✓ 4 design points
- ✓ 1 replicate
- ✓ Sample size = 8

Exercise 3.1

A bank wants to increase the yield of its credit card offers. It plans to collect VOC data by means of a DOE involving the factors in the table below. The bank plans to send out 1000 offers for each combination of the factor levels. Based on the data, they will determine the combination with the greatest % yield.

(a) What is the Y variable?

(b) What is the experimental unit? (Consider how Y will be measured)

(c) How many design points are in the full factorial?

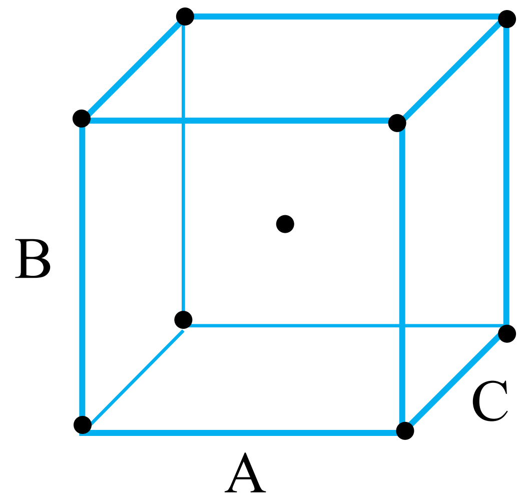
(d) What is the sample size?

(e) For each factor, decide whether you would treat it as quantitative or categorical (give your answers and reasons in the table below).

Exercise 3.1 (cont'd)

Factor	Levels	Continuous or categorical?
Introductory APR	0, 2.5 or 5%	
Introductory time period	3, 6 or 9 months	
Gift	iPhone, iPad, microwave or espresso machine	

4 The Full-Factorial Design



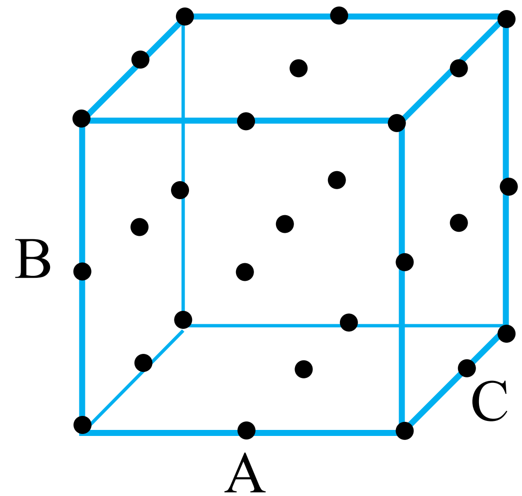
The full-factorial design contains all possible combinations of the specified factor settings

Above is an image of a 2^3 full-factorial with center points (continuous factors)

- The full-factorial requires one run at each design point (8 for this 2^3)
- 3 – 5 center points are recommended in a 2^k design
- Total runs required for this full-factorial are 11-13

A 2^k full-factorial design can estimate main effects and interactions

Full-Factorial Design (cont'd)

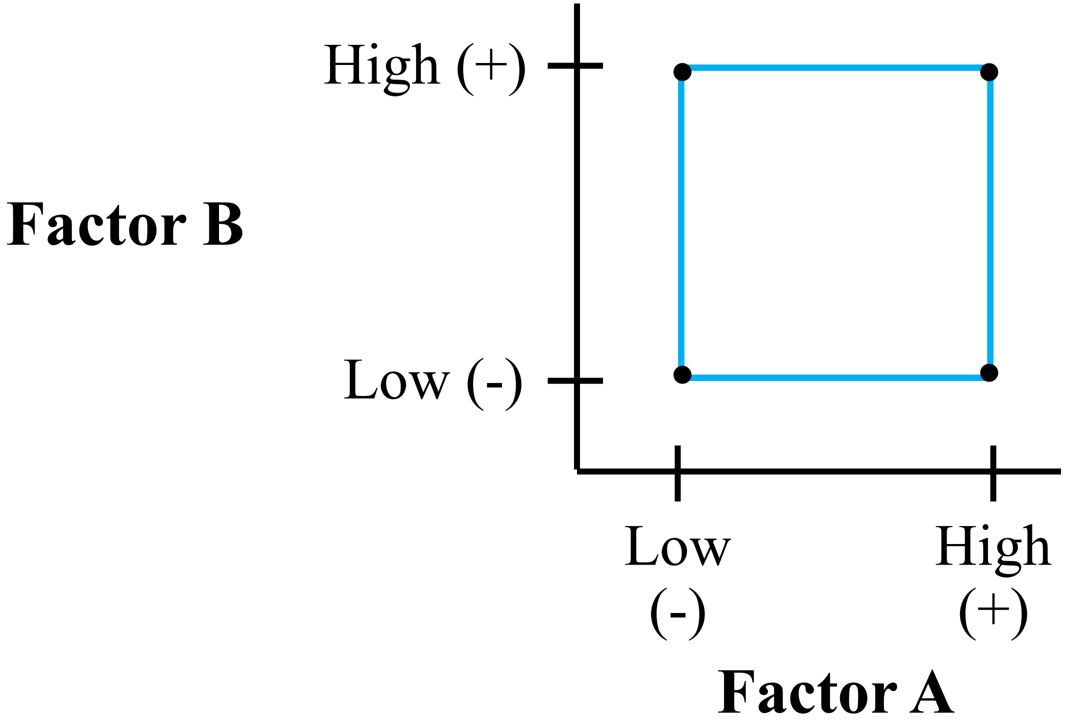


Above is an image of a 3^3 full-factorial

- The full-factorial requires one run at each design point
- "Center points" are part of the design points (the middle level of the factors)
- Total runs required for this 3^3 full-factorial is 27
- This type of design is useful when some factors are continuous, and some are categorical (there could be 3-level categorical factors in the picture above)

A three-level full-factorial (3^k) design can estimate main effects, interactions and quadratic effects, but is an inefficient design.

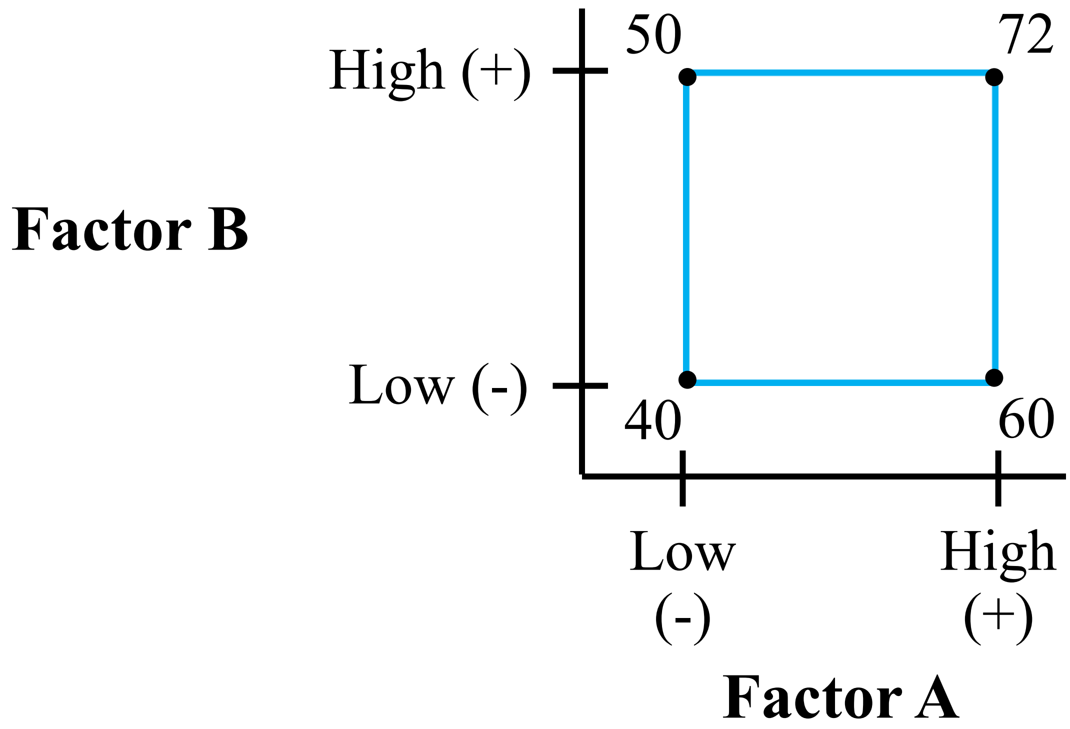
Main Effect of a Factor in a Factorial Design



$$Main\ Effect\ of\ A = Avg\ Response\ A\ (High) - Avg\ Response\ A\ (Low)$$

$$Coefficient\ A = \beta_1 = \frac{Main\ Effect\ A}{2}$$

Example: Main Effect of a Factor

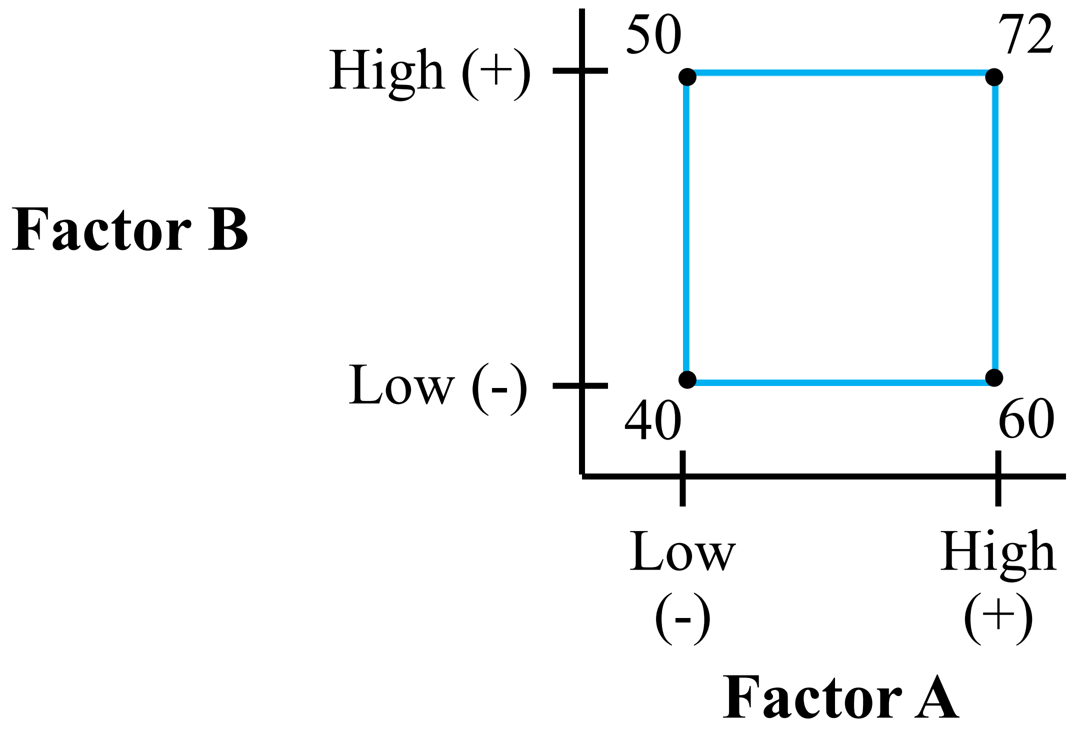


$$Main\ Effect\ of\ B = Avg\ Response\ B\ (High) - Avg\ Response\ B\ (Low)$$

$$= \frac{50+72}{2} - \frac{40+60}{2} = \frac{122}{2} - \frac{100}{2} = 11$$

What is the Main Effect of Factor A in this example?

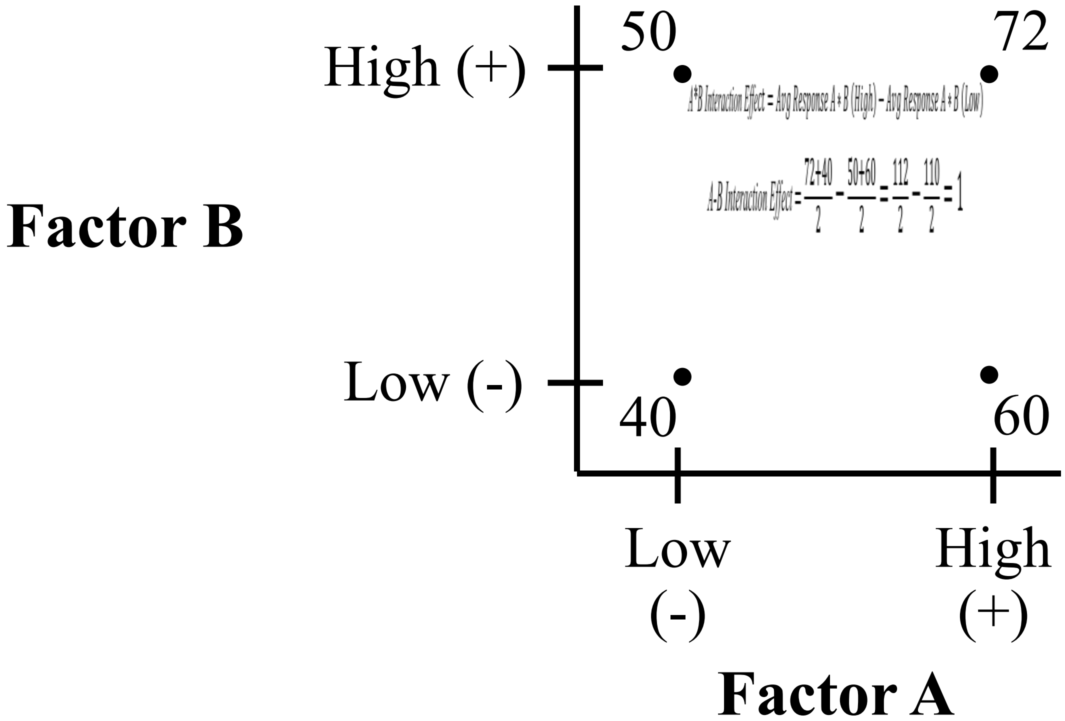
Example: Coefficient of a Factor



$$Coefficient\ B = \beta_1 = \frac{Main\ Effect\ B}{2} = \frac{11}{2} = 5.5$$

What is the coefficient for Factor A in this example?

Example: Interaction Effect



$$A * B\ Interaction\ Effect = Avg\ Response\ A * B\ (High) - Avg\ Response\ A * B\ (Low)$$

$$A\text{-}B\ Interaction\ Effect = \frac{72+40}{2} - \frac{50+60}{2} = \frac{112}{2} - \frac{110}{2} = 1$$

Example: Interaction Effect

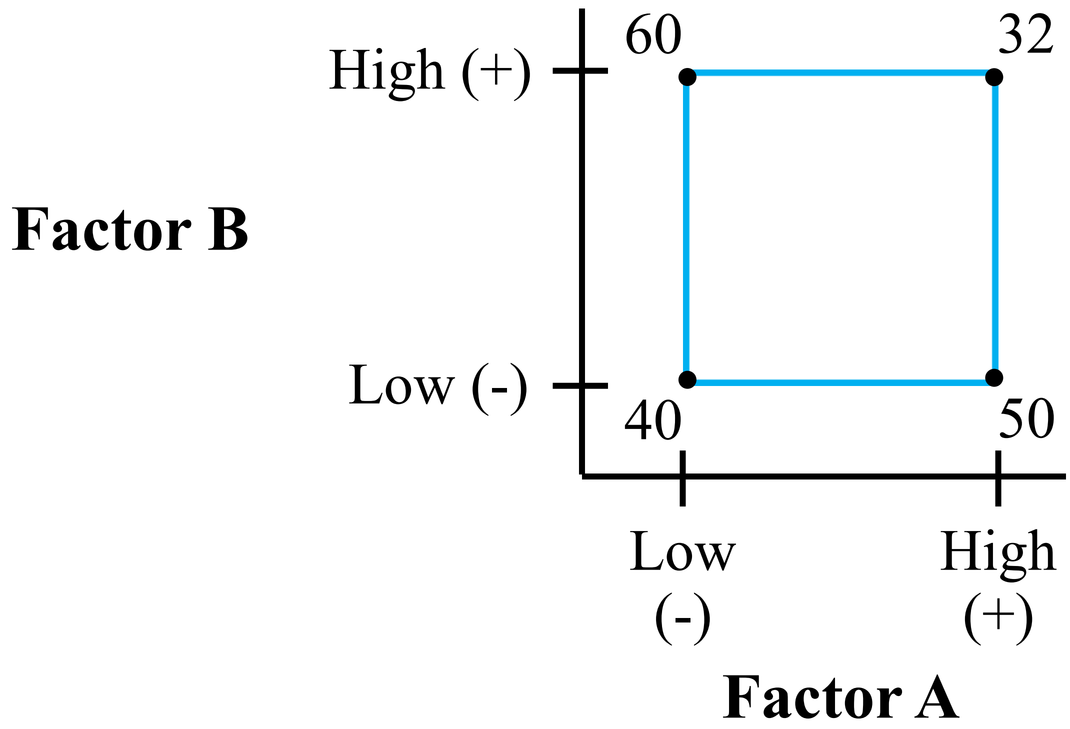
To determine which values are for A*B High and Low,

it can be helpful to refer to the experimental design matrix.

Multiply the + and – in the A and B columns in the design matrix to get the + and – for the A*B column.

	Factors			
Run	A	B	A*B	Response
1	-	-	+	40
2	-	+	-	50
3	+	-		60
4	+	+		72

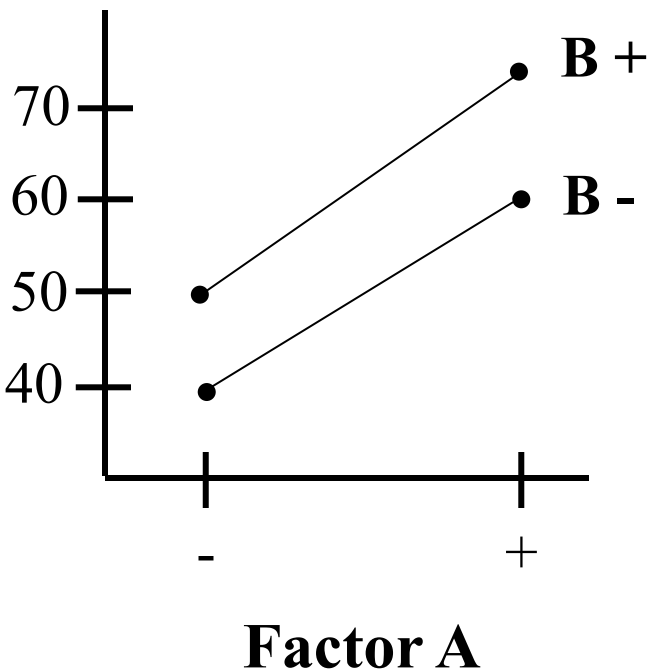
Example: Interaction Effect

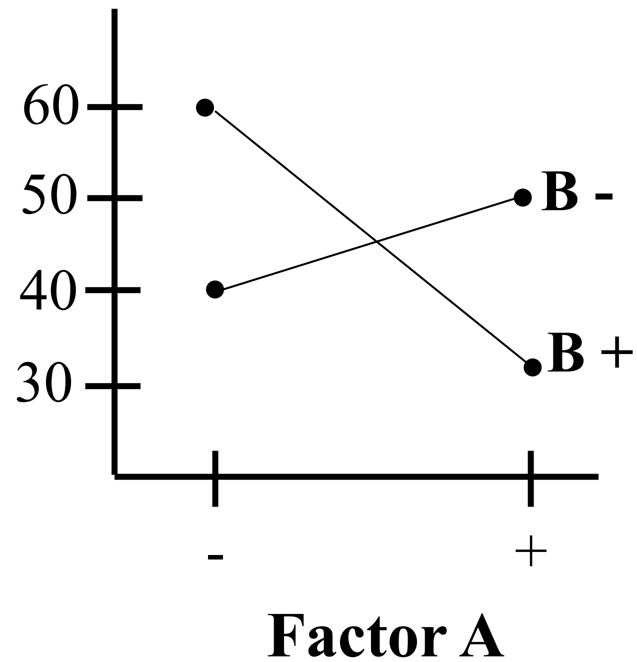


What is the A-B Interaction Effect in this example?

	Factors			
Run	A	B	A*B	Response
1	-	-		
2	-	+		
3	+	-		
4	+	+		

Interaction Plots graphically show interaction





Interaction Plot for the first example

No interaction—slopes of lines are approximately equal

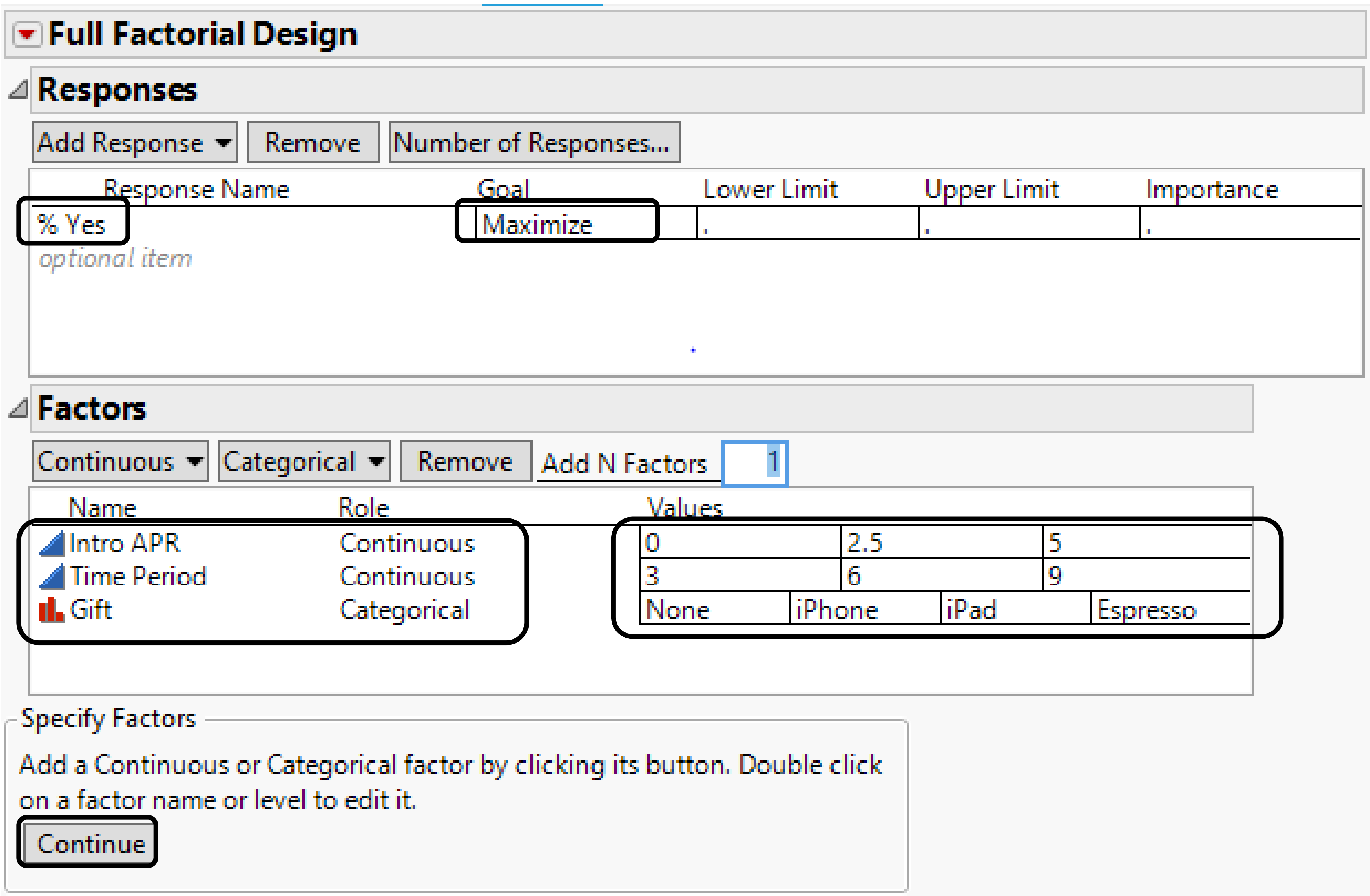
Interaction Plot for the data on the previous slide

Interaction present—lines have different slopes

Creating a Full Factorial Design

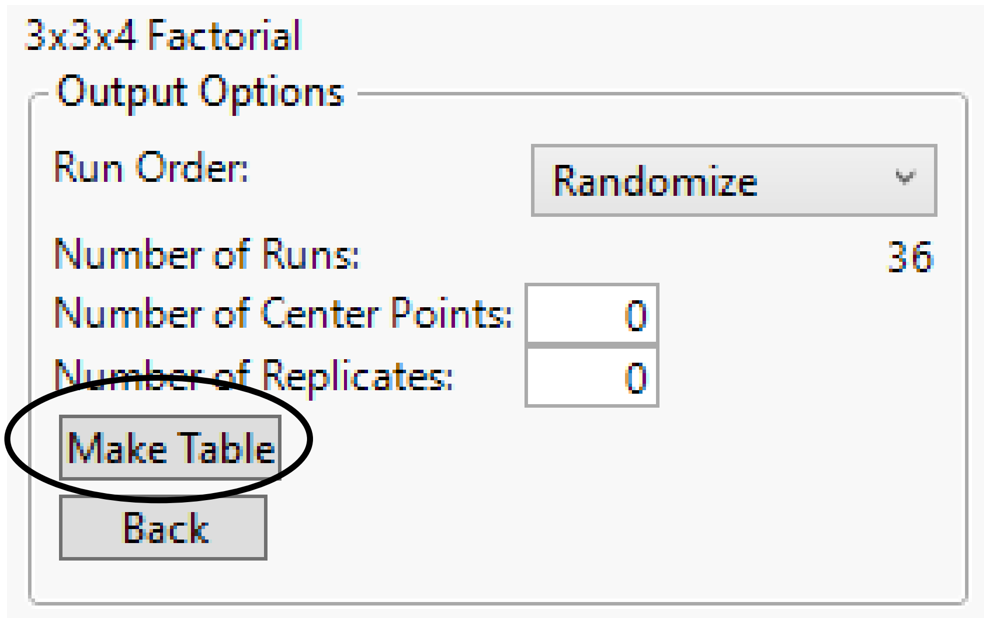
DOE → Classical → Full Factorial Design

1. Define responses, factors, numerical ranges for continuous factors, and levels for categorical factors.



Creating a full factorial (cont'd)

2. If desired, add extra center points*, request one or more replicates** and/or pre-sort the matrix. For a 2^k full-factorial, center runs are recommended. When you are ready, click *Make Table*.



	Pattern	Intro APR	Time Period	Gift	% Yes
1	312	5	3	iPhone	•
2	112	0	3	iPhone	•
3	124	0	6	Espresso	•
4	113	0	3	iPad	•
5	232	2.5	9	iPhone	•
6	231	2.5	9	None	•
7	134	0	9	Espresso	•
8	322	5	6	iPhone	•
9	332	5	9	iPhone	•
10	214	2.5	3	Espresso	•
11	331	5	9	None	•
12	121	0	6	None	•
13	223	2.5	6	iPad	•
14	314	5	3	Espresso	•
15	323	5	6	iPad	•
16	321	5	6	None	•
17	123	0	6	iPad	•
18	324	5	6	Espresso	•
19	132	0	9	iPhone	•
20	211	2.5	3	None	•
21	222	2.5	6	iPhone	•
22	311	5	3	None	•
23	213	2.5	3	iPad	•
24	333	5	9	iPad	•
25	111	0	3	None	•
26	221	2.5	6	None	•
27	212	2.5	3	iPhone	•
28	224	2.5	6	Espresso	•
29	313	5	3	iPad	•
30	334	5	9	Espresso	•
31	233	2.5	9	iPad	•
32	114	0	3	Espresso	•
33	133	0	9	iPad	•
34	234	2.5	9	Espresso	•
35	122	0	6	iPhone	•
36	131	0	9	None	•

*Each center point = one additional row (run)

**Each "replicate" = one additional set of 36 rows

Simulating response data (so we can see how analysis works)

3. Create two new columns called *Sent* and *Returned*.

4. Click on the *Sent* header → double-click on the *Sent* header *Column Properties*→ select *Formula* → enter the value 1000 in the little box → OK → OK

5. The *Returned* column is where we would enter the number of offers accepted. To simulate the data, double-click on the header and name the column → *Column Properties* → *Formula* → *Edit Formula*

6. Enter the commands shown on the next slide, then click OK.

	Pattern	Intro APR	Time Period	Gift	% Yes	Sent	Column 7
1	312	5	3	iPhone	•	1000	•
2	112	0	3	iPhone	•	1000	•
3	124	0	6	Espresso	•	1000	•
4	113	0	3	iPad	•	1000	•
5	232	2.5	9	iPhone	•	1000	•
6	231	2.5	9	None	•	1000	•
7	134	0	9	Espresso	•	1000	•
8	322	5	6	iPhone	•	1000	•
9	332	5	9	iPhone	•	1000	•
10	214	2.5	3	Espresso	•	1000	•
11	331	5	9	None	•	1000	•
12	121	0	6	None	•	1000	•
13	223	2.5	6	iPad	•	1000	•
14	314	5	3	Espresso	•	1000	•
15	323	5	6	iPad	•	1000	•
16	321	5	6	None	•	1000	•
17	123	0	6	iPad	•	1000	•
18	324	5	6	Espresso	•	1000	•
19	132	0	9	iPhone	•	1000	•
20	211	2.5	3	None	•	1000	•
21	222	2.5	6	iPhone	•	1000	•
22	311	5	3	None	•	1000	•
23	213	2.5	3	iPad	•	1000	•
24	333	5	9	iPad	•	1000	•
25	111	0	3	None	•	1000	•
26	221	2.5	6	None	•	1000	•
27	212	2.5	3	iPhone	•	1000	•
28	224	2.5	6	Espresso	•	1000	•
29	313	5	3	iPad	•	1000	•
30	334	5	9	Espresso	•	1000	•
31	233	2.5	9	iPad	•	1000	•
32	114	0	3	Espresso	•	1000	•
33	133	0	9	iPad	•	1000	•
34	234	2.5	9	Espresso	•	1000	•
35	122	0	6	iPhone	•	1000	•
36	131	0	9	None	•	1000	•

Simulating response data (cont'd)

Formula:

Random
↓
Random Integer[n1]
↓
Random Integer[50]

OK → OK

7. Define *% Yes* with the formula

$$\left[\frac{\text{Returned}}{\text{Sent}}\right]*100$$

8. Run the *Model* script provided in the left panel. (Click on the green triangle)

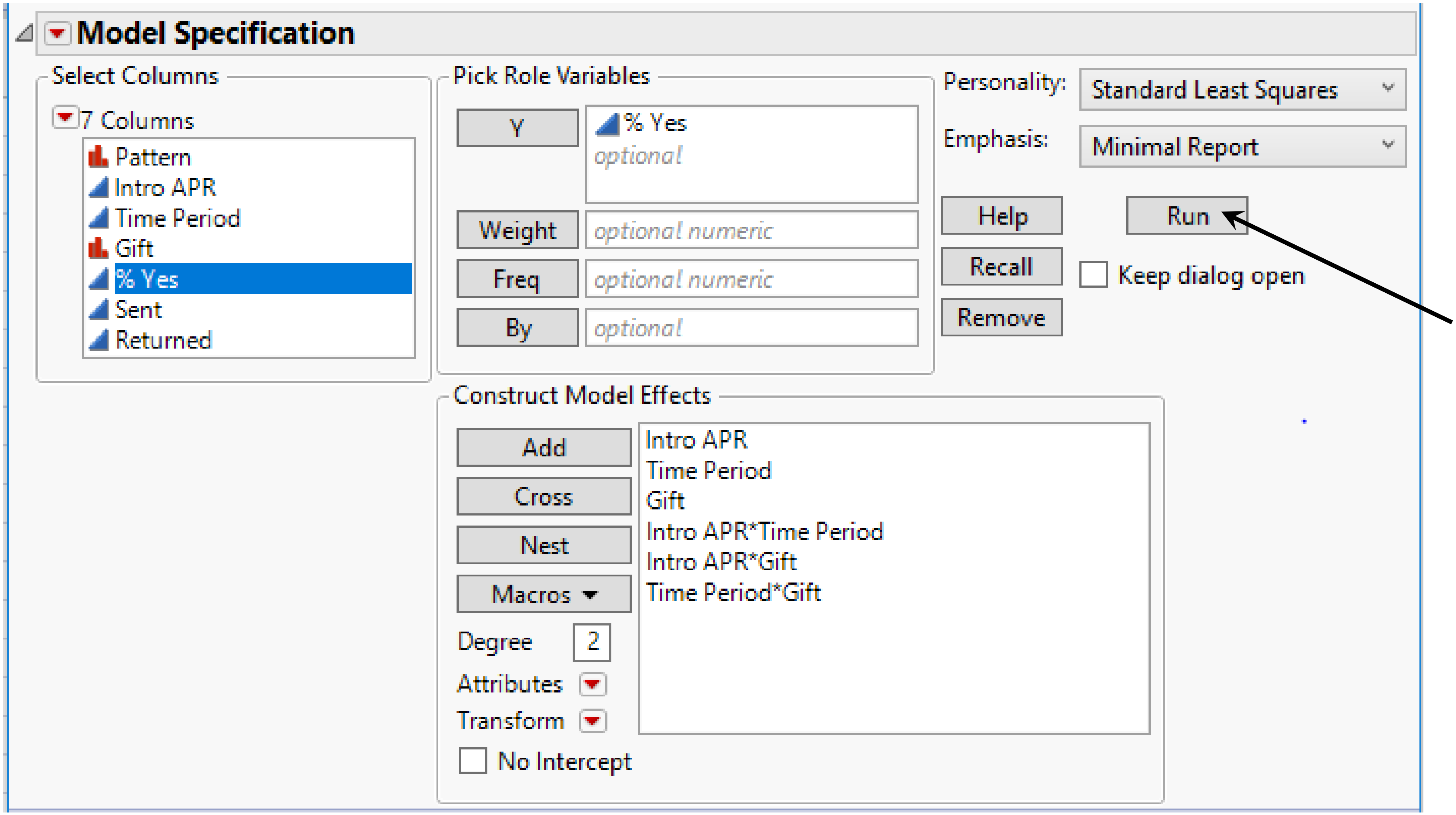
3x3x4 Factorial
Design 3x3x4 Factorial
Model
Evaluate Design
DOE Dialog

Columns (7/1)
Pattern
Intro APR
Time Period
Gift
% Yes
Sent
Returned

Rows
All rows 36
Selected 0
Excluded 0
Hidden 0
Labelled 0

	Pattern	Intro APR	Time Period	Gift	% Yes	Sent	Returned
1	312	5	3	iPhone	4.1	1000	41
2	112	0	3	iPhone	0.7	1000	7
3	124	0	6	Espresso	1.6	1000	16
4	113	0	3	iPad	3.9	1000	39
5	232	2.5	9	iPhone	2.9	1000	29
6	231	2.5	9	None	4.4	1000	44
7	134	0	9	Espresso	1.1	1000	11
8	322	5	6	iPhone	1.1	1000	11
9	332	5	9	iPhone	1.7	1000	17
10	214	2.5	3	Espresso	3.2	1000	32
11	331	5	9	None	1.7	1000	17
12	121	0	6	None	3.8	1000	38
13	223	2.5	6	iPad	5	1000	50
14	314	5	3	Espresso	1.9	1000	19
15	323	5	6	iPad	3.3	1000	33
16	321	5	6	None	4.5	1000	45
17	123	0	6	iPad	4.4	1000	44
18	324	5	6	Espresso	4.2	1000	42
19	132	0	9	iPhone	4.4	1000	44
20	211	2.5	3	None	3.5	1000	35
21	222	2.5	6	iPhone	2.5	1000	25
22	311	5	3	None	5	1000	50
23	213	2.5	3	iPad	0.3	1000	3
24	333	5	9	iPad	1.6	1000	16
25	111	0	3	None	1.7	1000	17
26	221	2.5	6	None	2.1	1000	21
27	212	2.5	3	iPhone	2.5	1000	25
28	224	2.5	6	Espresso	0.5	1000	5
29	313	5	3	iPad	0.6	1000	6
30	334	5	9	Espresso	4.6	1000	46
31	233	2.5	9	iPad	2.1	1000	21
32	114	0	3	Espresso	0.5	1000	5
33	133	0	9	iPad	4.6	1000	46
34	234	2.5	9	Espresso	5	1000	50
35	122	0	6	iPhone	1.6	1000	16
36	131	0	9	None	1	1000	10

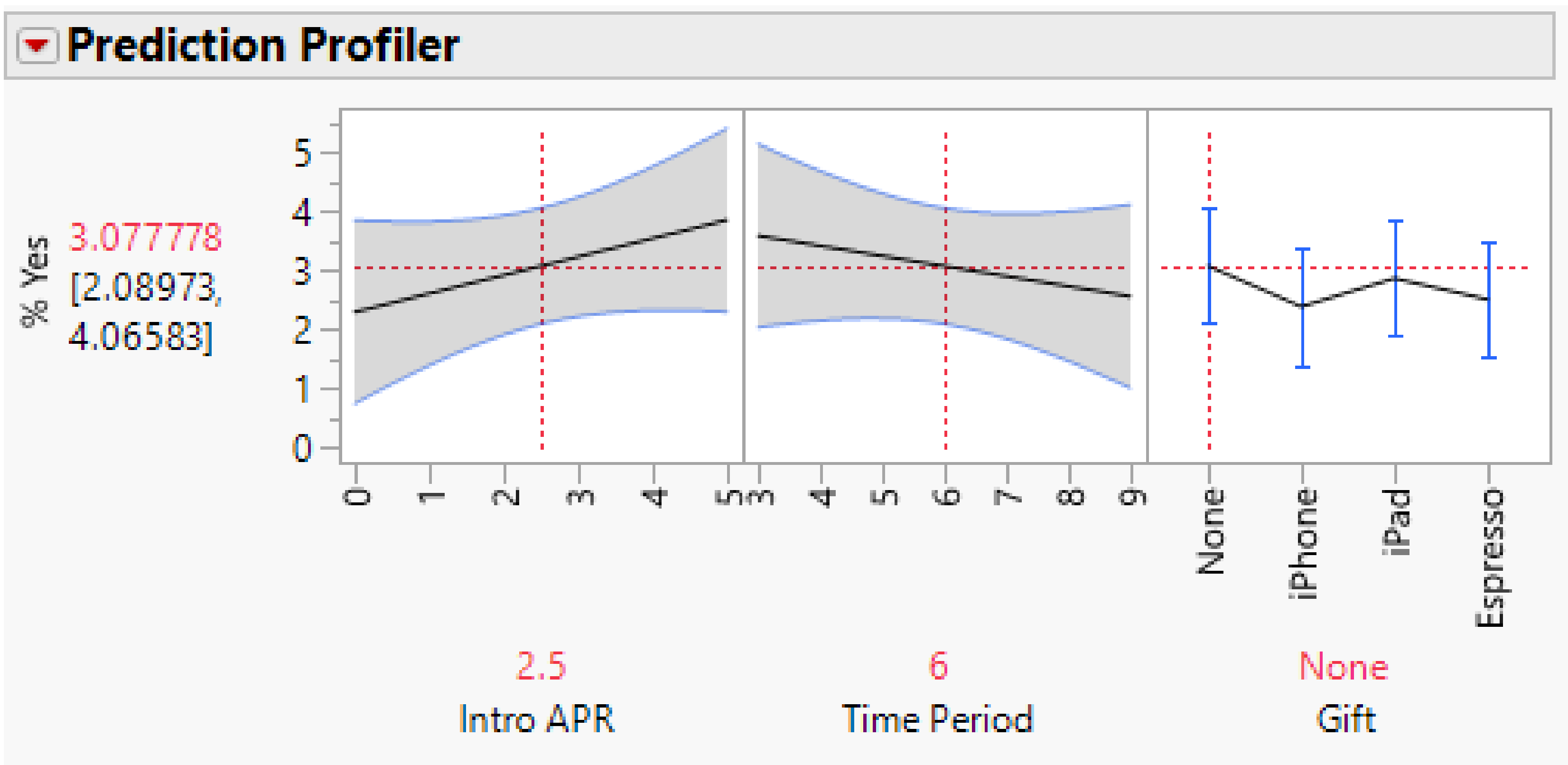
Analyzing the simulated data



When you click Run, JMP will use regression to create a "model" for the process, that includes the terms under *Construct Model Effects*.

Getting to "yes"

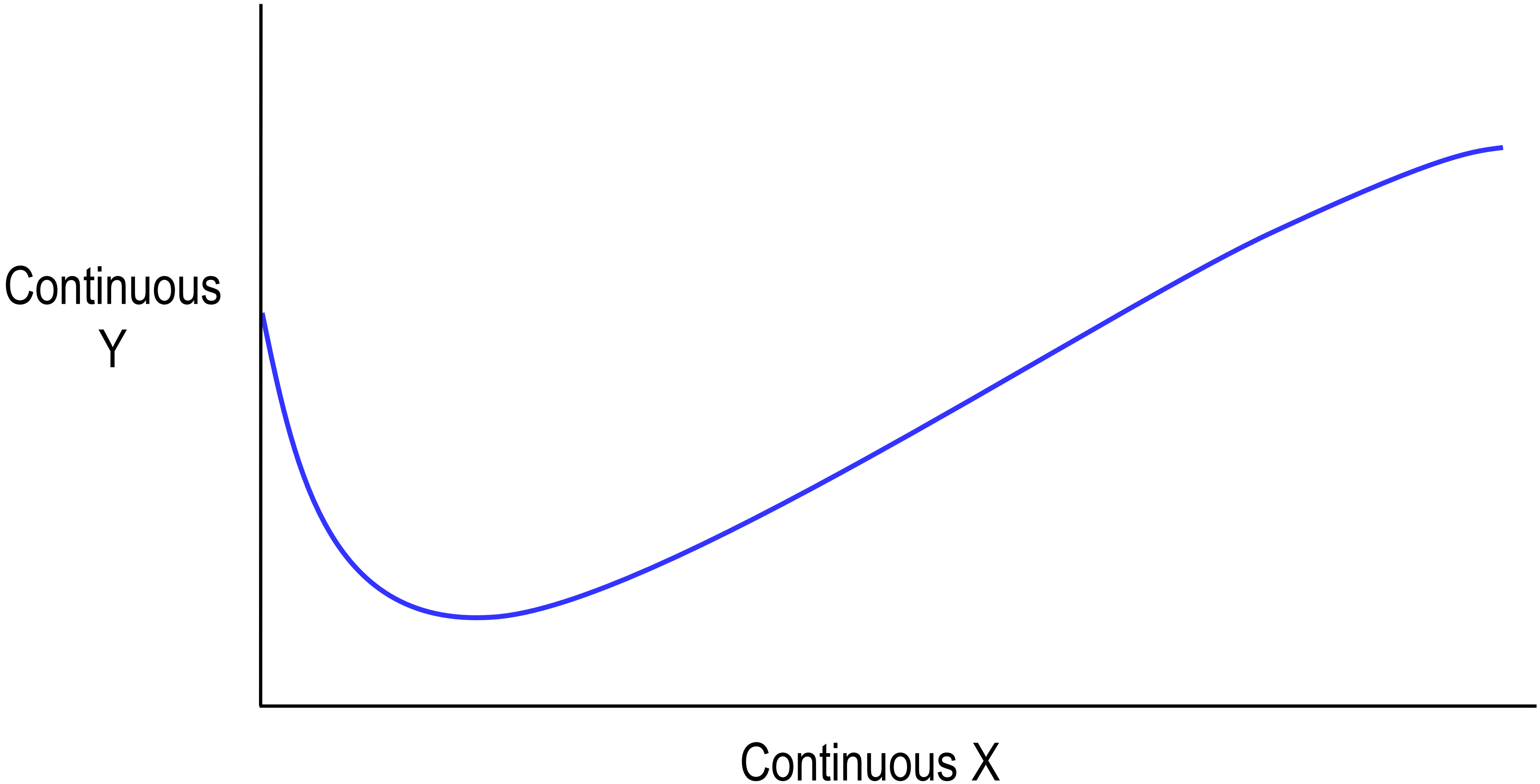
- Point and click to find the combination with the highest % Yes
- Because it is simulated data:
  - your profiler won't look exactly like this one
  - don't be alarmed if your "best" combination doesn't make sense



This slide intentionally left blank

5 Statistical Assumptions

Average Y as a function of X has no jumps or corners (assumption of *smoothness*)



A hypothetical smooth response function.

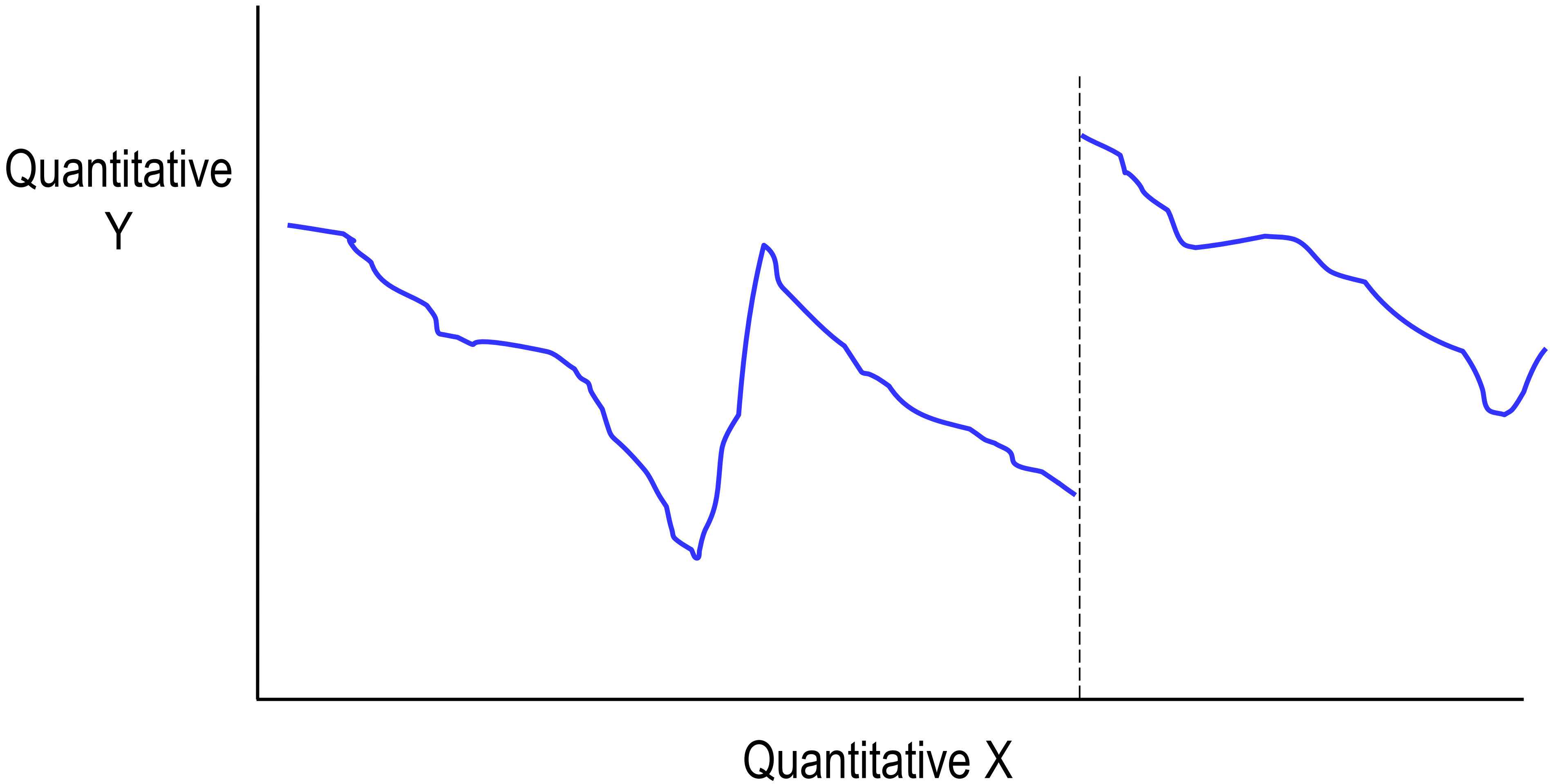
We never know the true response function, but often we have information about its general properties. For continuous X and Y, *smoothness* of the $Y = f(X)$ relationship is one such property. It means the function can be well approximated over sufficiently short intervals by a polynomial, usually linear or quadratic. This is necessary in optimization experiments where we want to *interpolate* between the experimental design points.

These experiments are designed for continuous Y response. If you have a pass-fail response, see if you can turn it into a continuous response. Here are a few ideas:

- If you measure something on a continuous scale, but only record whether it passed or failed in your normal operation, record the actual measurement during the experiment.
- If you typically use a go-no go gauge, actually measure the part during the experiment.
- Record the size of defect instead of whether there is or is not a defect.
- Other ideas?

Non-smooth response function

Average Y as a function of X has jumps and/or corners



A hypothetical non-smooth response function.

A function with jumps or sharp corners will not be well approximated by low-order polynomials in neighborhoods of the associated X values. This is a problem in optimization experiments because we want to interpolate.

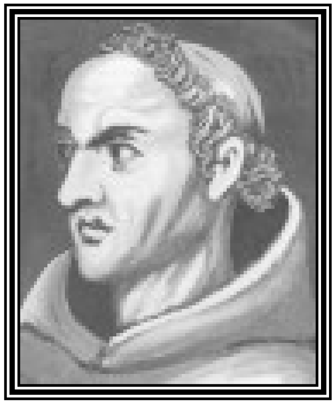
It may or may not be a problem in screening experiments, because there we are merely trying to identify factors with large first-order effects. Accurate approximation throughout the X range is not required, although we may not be able to see the impact of the factor under certain circumstances. (You can see in the picture above that the response, Y, is at nearly the same level across various X values.)

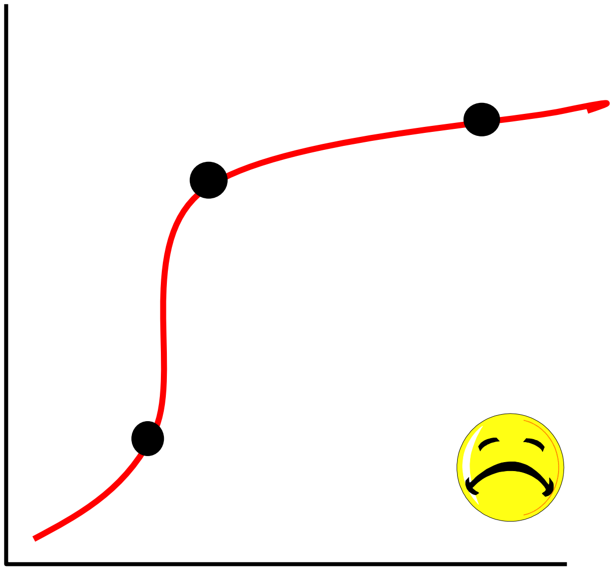
Jumps and sharp corners often occur outside the feasible operating range of the process. In fact, such discontinuities often *define* the feasible operating range. A smooth response function is usually a safe assumption as long as we are not operating too close to a "cliff."

Occam's razor

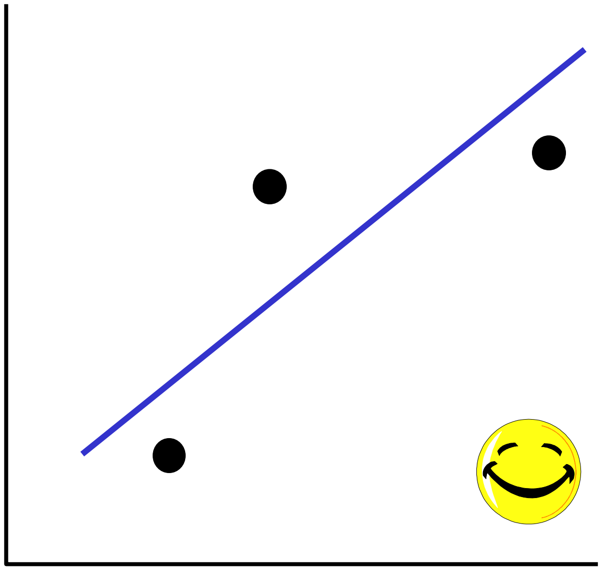
"One should not increase, beyond what is necessary, the number of entities required to explain something."

—William of Occam, medieval philosopher





Exact "French curve"



Linear plus noise

Occam’s razor represents a preference for simple explanations over complex ones. This reflects a belief that simple hypotheses are more likely to be true than complex ones. This belief is not always justified, but it is efficient in that it leads to models with just enough complexity to explain a given set of observations.

We can always find a sufficiently complex curve passing exactly through any given set of data points. **The predictive ability of this “over-fitting” method is notoriously poor.** The more successful “Occam” strategy is illustrated by random variation superimposed on a simple linear model.

Standard assumptions on the response function

- ✓ $Y = f(X_1, X_2, X_3, \ldots) + \textit{error}$
- ✓ Can't assume $f(X)$ explains everything (hence the error term)
- ✓ Can't assume $f(X)$ is linear, but quadratic model is almost always sufficient
  - $f(X)$ may include second order interactive effects
  - $f(X)$ may include quadratic effects
- ✓ Don't need cubic or higher order models
  - Don't need higher order interactive effects

Review: Least squares model fitting

For each of 18 potato chip bags, we have data on

$$T = \text{bonding temperature}$$
$$D = \text{bonding time (duration)}$$
$$Y = \text{bond strength}$$

The best fitting *response surface model* (RSM) is the one whose parameters

$$b_0, b_1, b_2, b_3, b_4, b_5$$

minimize the sum of squared residuals:

$$\sum_{\{18\text{ bags}\}} \left[Y - \left(b_0 + b_1 T + b_2 D + b_3 TD + b_4 T^2 + b_5 D^2 \right) \right]^2$$

Least squares fit of Response Surface Model (RSM)

$$\text{Avg. Y} = 87.2 + 8.3(\text{T}) + 7.7(\text{D}) - 31.8(\text{TD}) - 16.1(\text{T}^2) - 13.2(\text{D}^2)$$

	A	B	C	D	E	F	G
1	TEMP	DWELL	BOND		Prediction		Noise
2	-1	-1	11.0		10.08		0.92
3	-1	-1	8.9		10.08		-1.18
4	-1	0	63.9		62.80		1.10
5	-1	0	60.4		62.80		-2.40
6	-1	1	93.2		89.07		4.13
7	-1	1	86.5		89.07		-2.57
8	0	-1	65.7		66.30		-0.60
9	0	-1	67.7		66.30		1.40
10	0	0	88.4		87.20		1.20
11	0	0	88.0		87.20		0.80
12	0	1	82.0		81.65		0.35
13	0	1	78.5		81.65		-3.15
14	1	-1	88.1		90.37		-2.27
15	1	-1	92.1		90.37		1.73
16	1	0	77.2		79.45		-2.25
17	1	0	81.0		79.45		1.55
18	1	1	39.5		42.08		-2.58
19	1	1	45.9		42.08		3.82
20	Sum of squares (SS)		93876.58	=	93792.35	+	84.18
21	Degrees of freedom (DF)		18	=	6	+	12
22		RMSE	*Square root of noise (SS/DF)*				2.65
23							

least squares modeling.xls

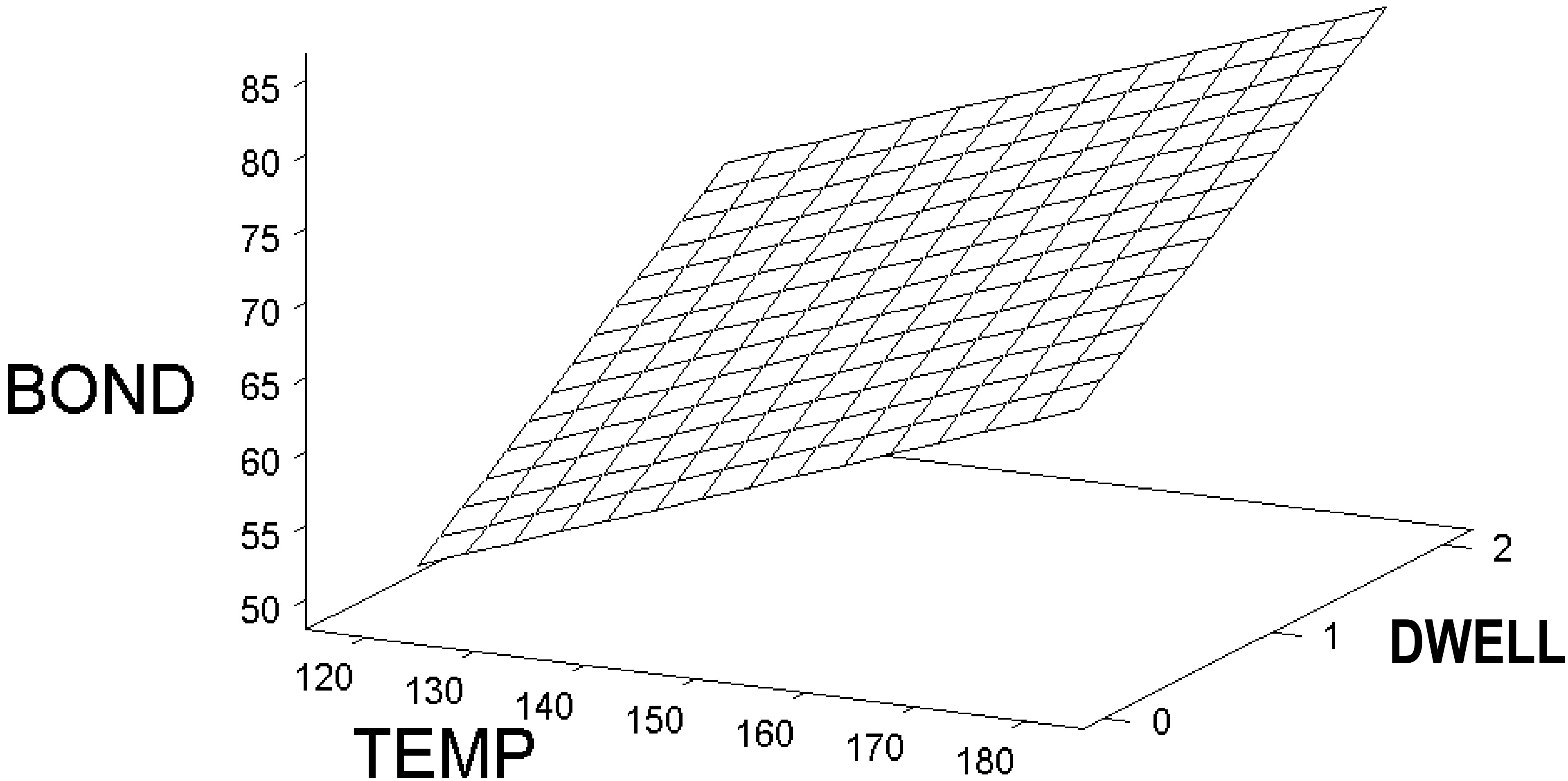
6 terms in model (equation shown above)

$2.65 = \sqrt{84.18/12}$

This slide intentionally left blank

Linear in the Xs

Average Bond = 67.2 + 8.3(TEMP) + 8.3(DWELL)



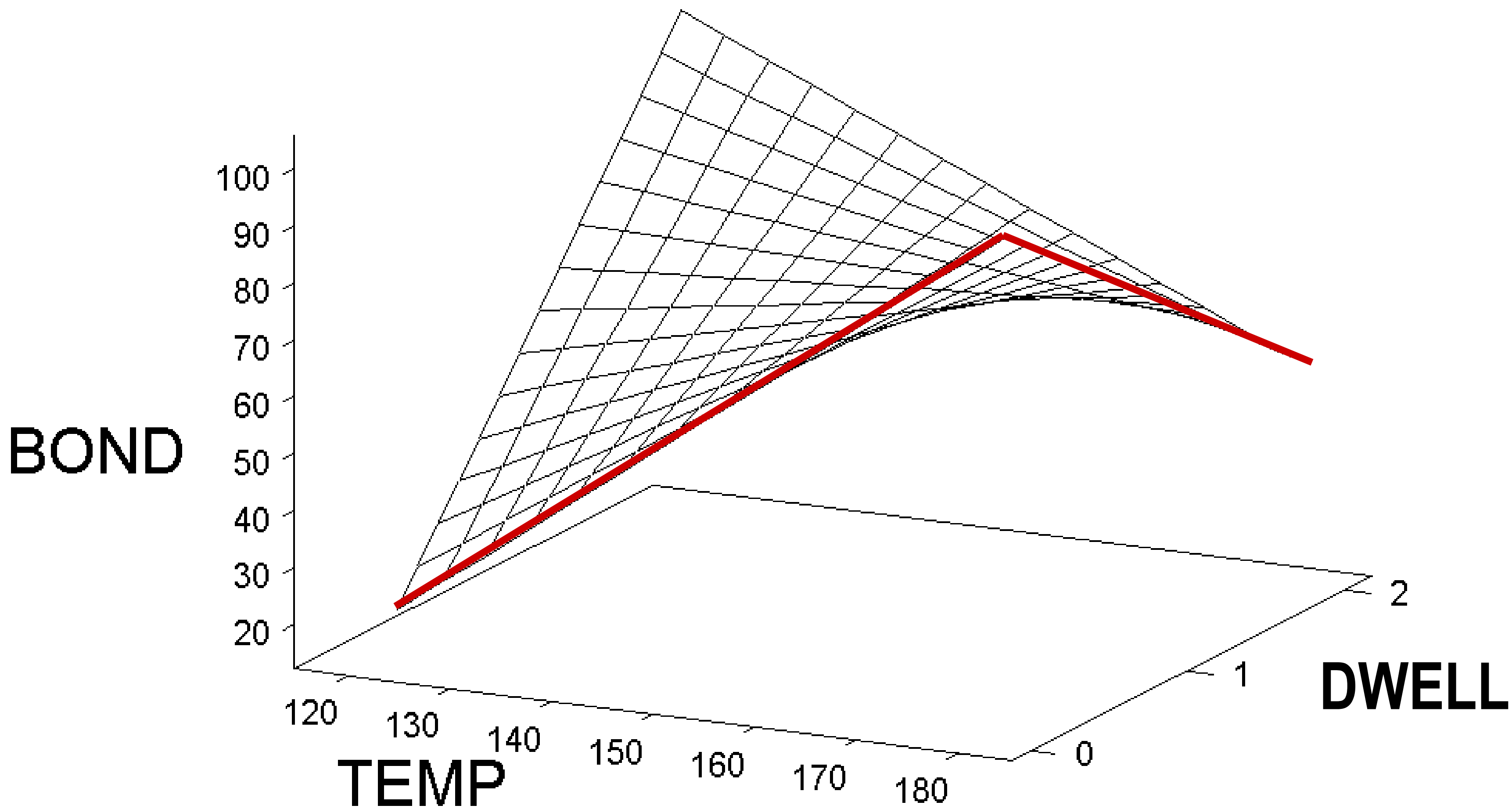
Response surface: tilted plane.

Simple linear models like the one shown above are used in screening designs. In many cases, simple linear models fit the data poorly, and do not give accurate predictions. They should not be used for optimization experiments.

Simple linear model: $Y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$

Linear interaction model

$$\text{Avg. BOND} = 67.2 + 8.3(\text{TEMP}) + 8.3(\text{DWELL}) - 31.5(\text{TEMP} \times \text{DWELL})$$



Response surface: saddle.

Linear interaction models like the one shown above usually fit the data much better than simple linear models.

They include all main effects and all interaction effects.

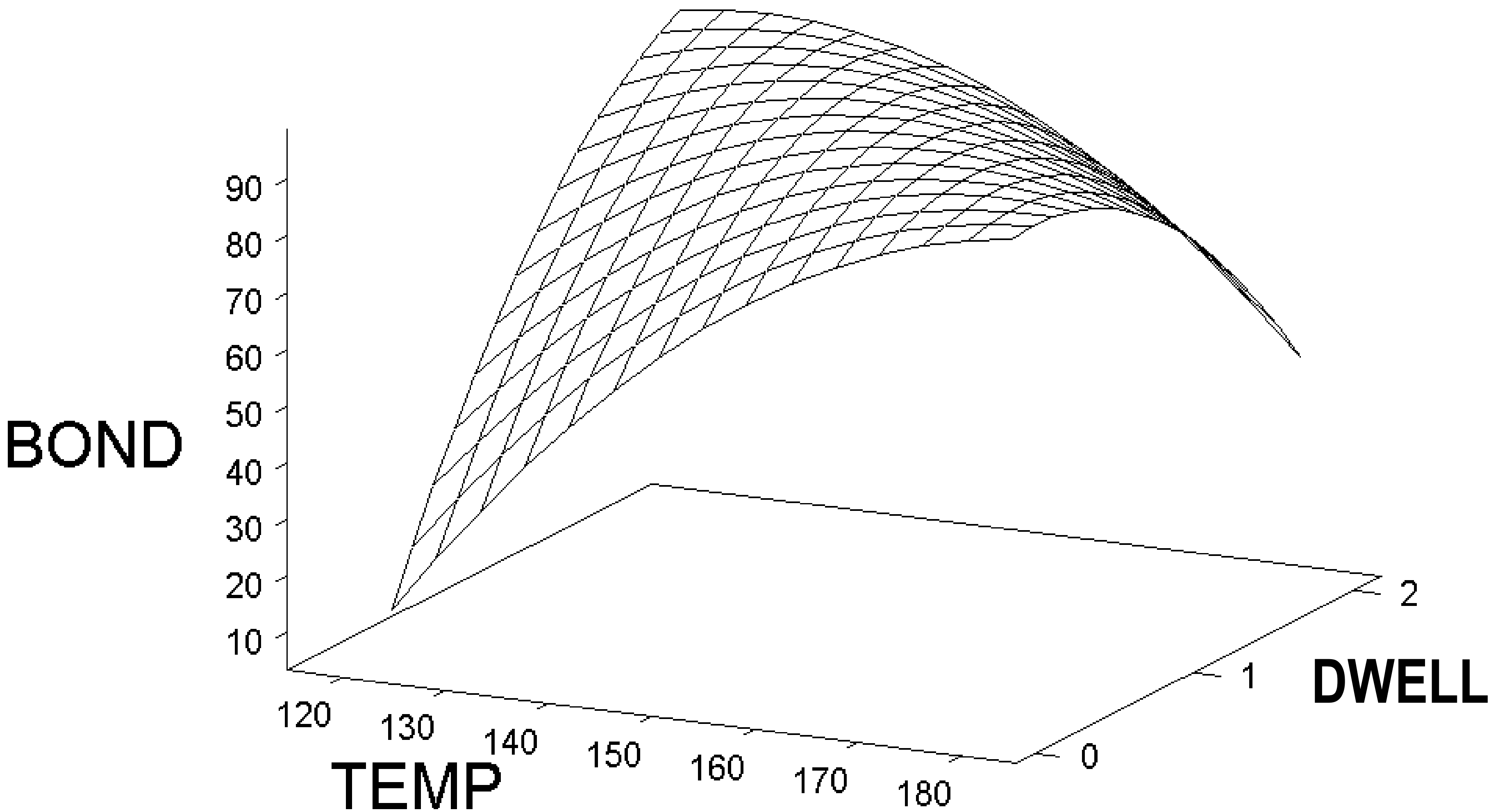
They are good for optimization experiments where all factors are categorical, but they should not be used for optimization experiments involving quantitative factors.

Linear interaction model:

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_i x_i + b_{12} x_1 x_2 + b_{13} x_1 x_3 + \cdots + b_{ij} x_i x_j$$

Response surface model (RSM)

$$\text{Avg. BOND} = 86.8 + 8.3(\text{TEMP}) + 8.1(\text{DWELL}) - 32.4(\text{TEMP}\times\text{DWELL}) - 15.5(\text{TEMP}\times\text{TEMP}) - 12.9(\text{DWELL}\times\text{DWELL})$$



Response surface: ridge.

The response surface model (RSM) shown above is the **standard model for optimization experiments**.

It differs from the linear interaction model in that it includes quadratic (squared) terms for all continuous factors, in addition to all main effects and interactions. Quadratic terms are never used with categorical factors.

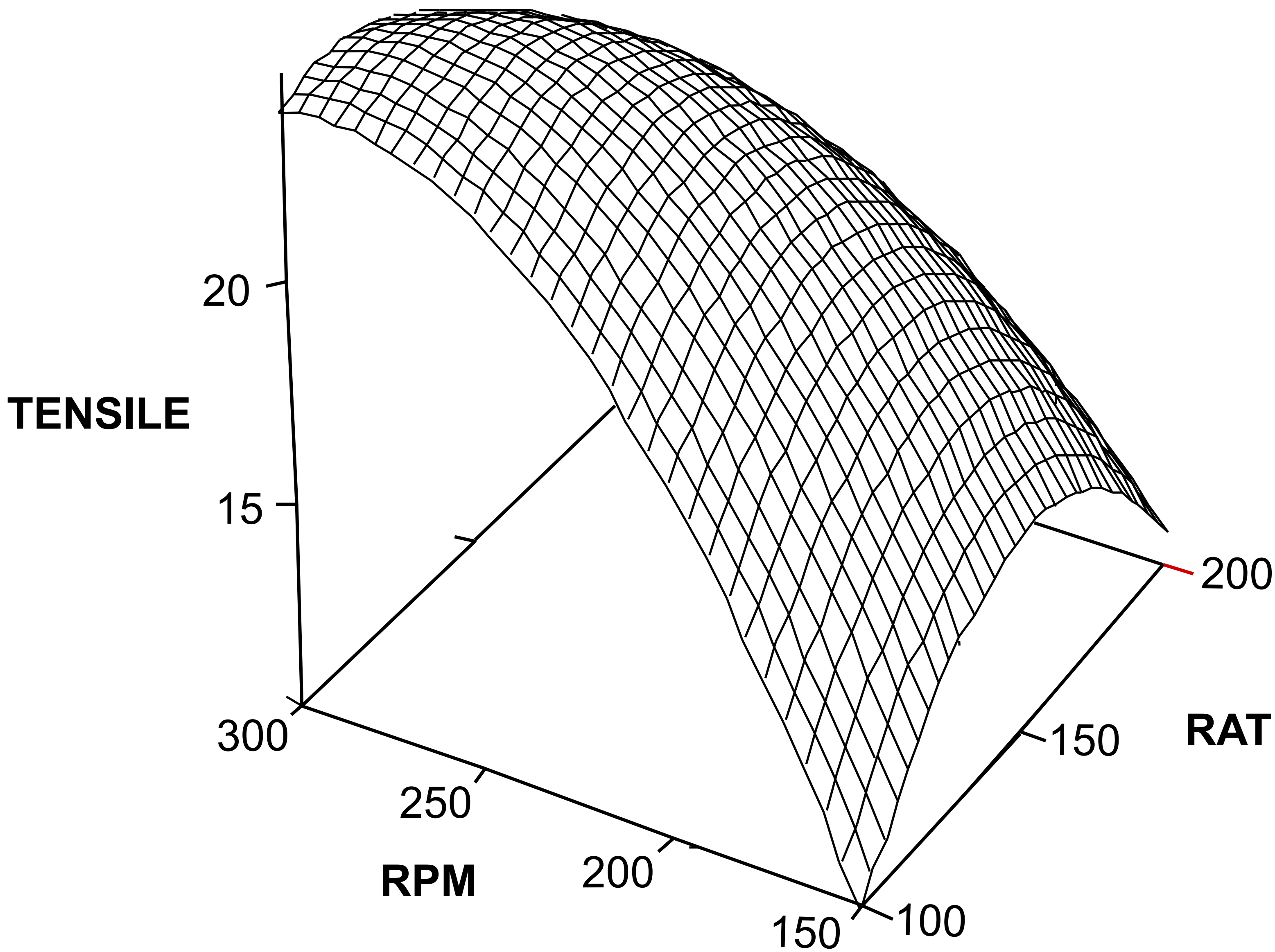
In experiments involving continuous factors, the RSM may fit the data much better than the linear interaction model. In other words, the response surface model may be a better model of the process.

Response Surface Model RSM):

$$Y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_ix_i + b_{12}x_1x_2 + b_{13}x_1x_3 + \cdots + b_{ij}x_ix_j$$
$$+b_{11}x_1^2 + b_{22}x_2^2 + \cdots + b_{ii}x_i^2$$

RSM for a different data set (process)

$$\text{Avg. TENSILE} = 22.5 - 3.3(\text{RATE}) + 3.4(\text{RPM}) - 3.6(\text{RATE} \times \text{RPM})$$
$$- 4.8(\text{RATE} \times \text{RATE}) - 5.6(\text{RPM} \times \text{RPM})$$



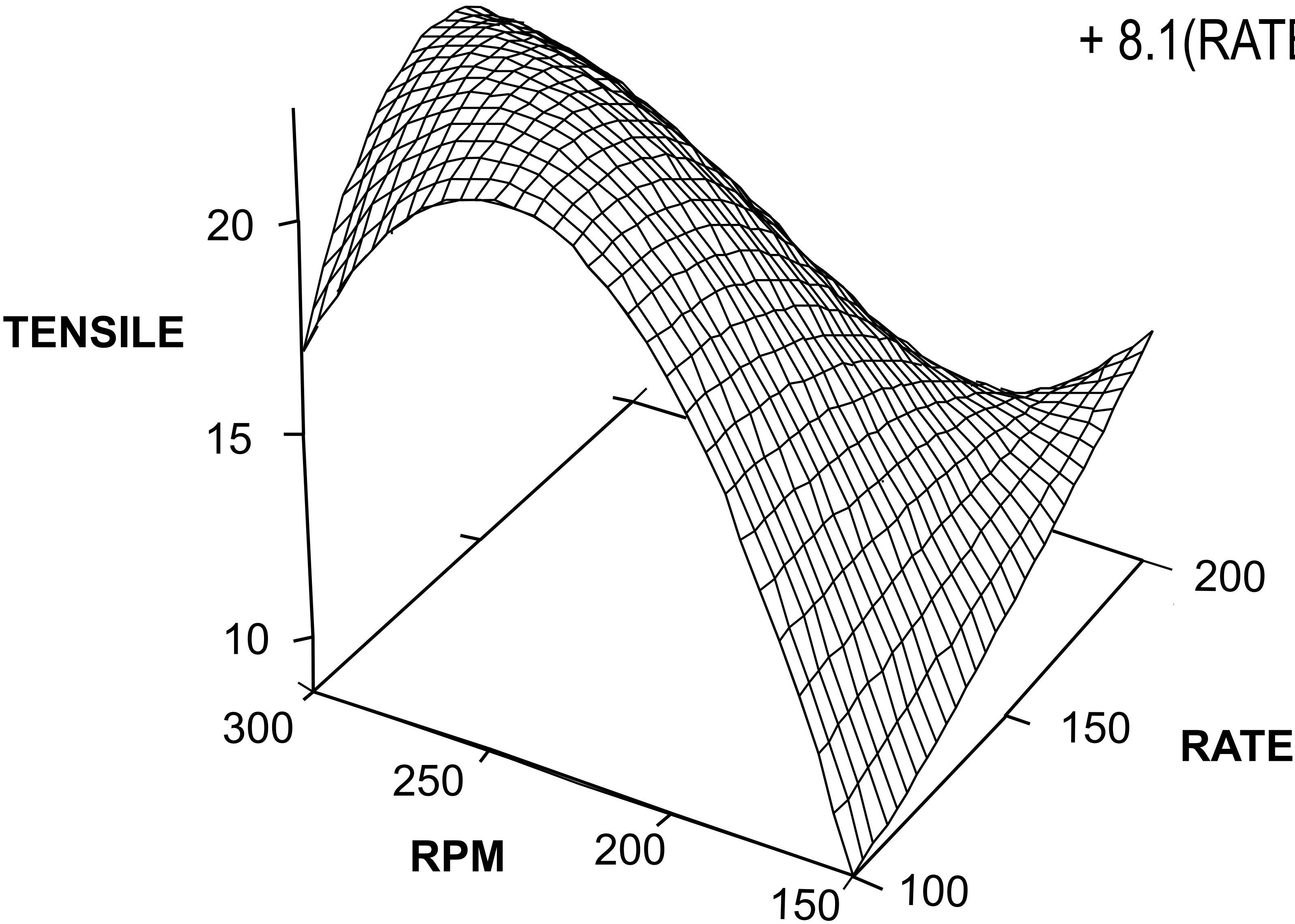
Response surface: hilltop.

Other response surface shapes include inverted saddles, inverted ridges, and bowls.

You can’t tell from the plot alone, but in this example the RSM model does not fit the data very well.

RSM plus quadratic interactions

$$\text{Avg. TENSILE} = 22.4 - 8.5(\text{RATE}) + 8.6(\text{RPM}) - 3.2(\text{RATE} \times \text{RPM})$$
$$- 6.1(\text{RATE}^2) - 4.8(\text{RPM}^2) - 7.0(\text{RATE}^2 \times \text{RPM})$$
$$+ 8.1(\text{RATE} \times \text{RPM}^2)$$



The shows a more complicated quadratic model fit to the same data as on the previous page. This model turns out to fit the data well.

Model terms like

$$\text{RATE} \times \text{RATE} \times \text{RPM}$$

$$\text{RATE} \times \text{RPM} \times \text{RPM}$$

$$\text{RATE} \times \text{RATE} \times \text{RPM} \times \text{RPM}$$

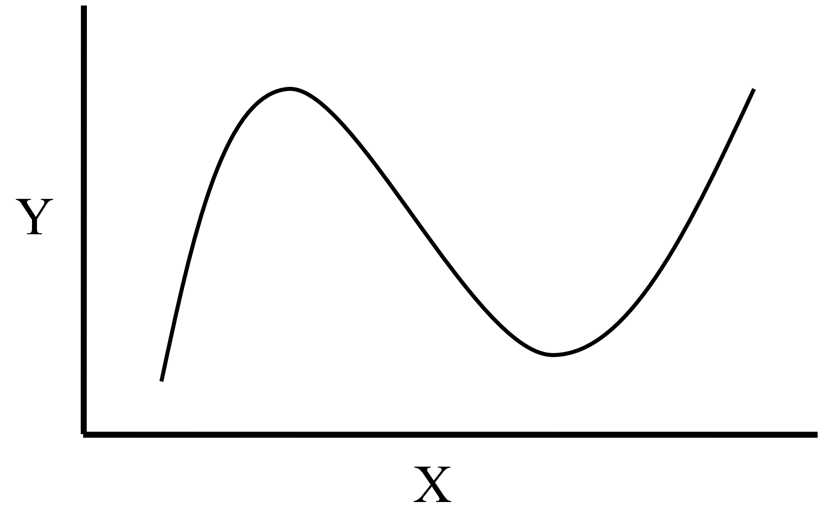
are called *quadratic interactions*. Adding one or more quadratic interactions is a good thing to try when an RSM model does not fit.

It is also possible to add other higher-level terms (cubic, three-way interactions), if the sample size is large enough to support the extra terms . . .

Higher-order polynomial models?

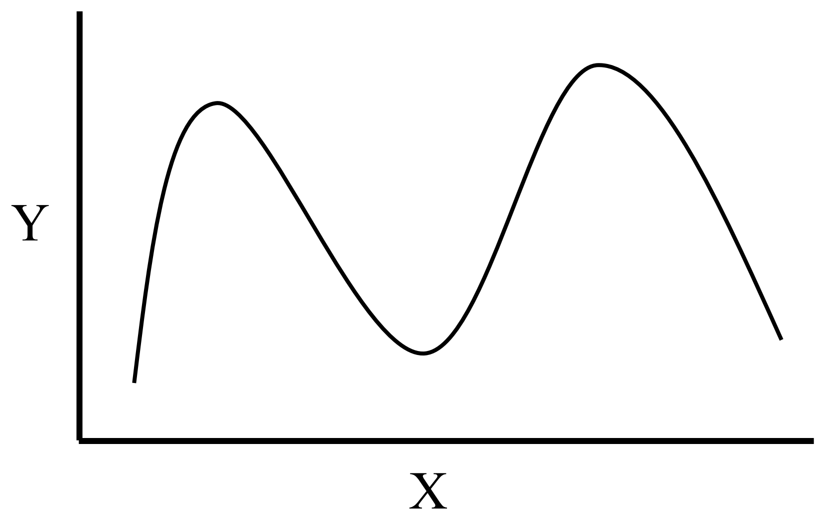
<u>3rd order polynomial (cubic)</u>

$$\text{Avg. } Y = b_0 + b_1 X + b_2 X^2 + b_3 X^3$$



<u>4th order polynomial (quartic)</u>

$$\text{Avg. } Y = b_0 + b_1 X + b_2 X^2 + b_3 X^3 + b_4 X^4$$



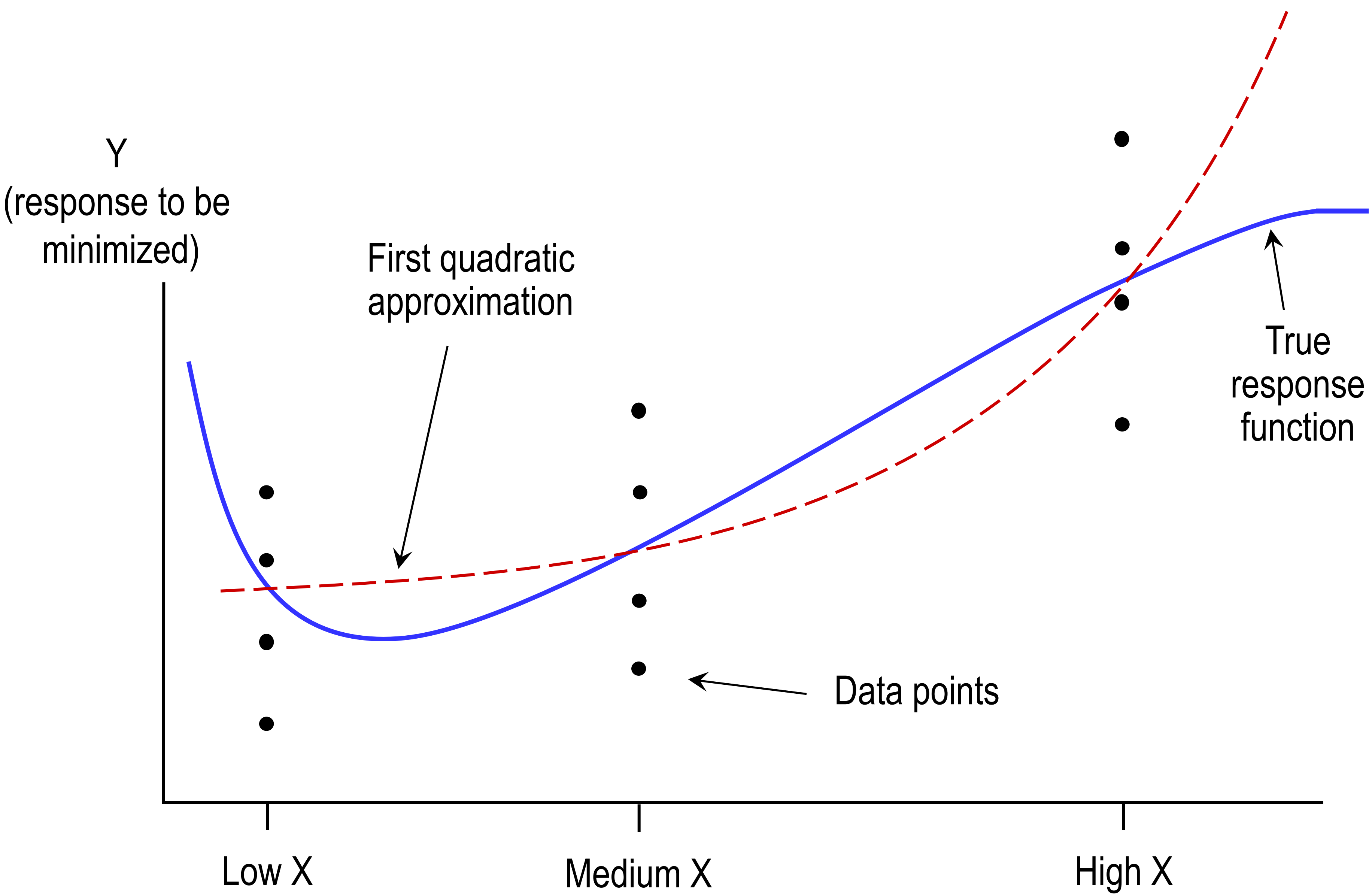
Even though third- or higher-order models may fit the data better than quadratic (second-order) models, they are rarely used in DOE. Why? They require much larger samples sizes for any given set of factors.

It is much more common to use quadratic models in an iterative fashion. A quadratic model may not fit the data well over a large initial factor space, but it almost always tells us which subset of the initial factor space is most likely to give the results we are looking for. The next step is to run another quadratric experiment in the smaller region. The smaller the factor space, the better the quadratic model will fit the data.

This concept is illustrated on the next page.

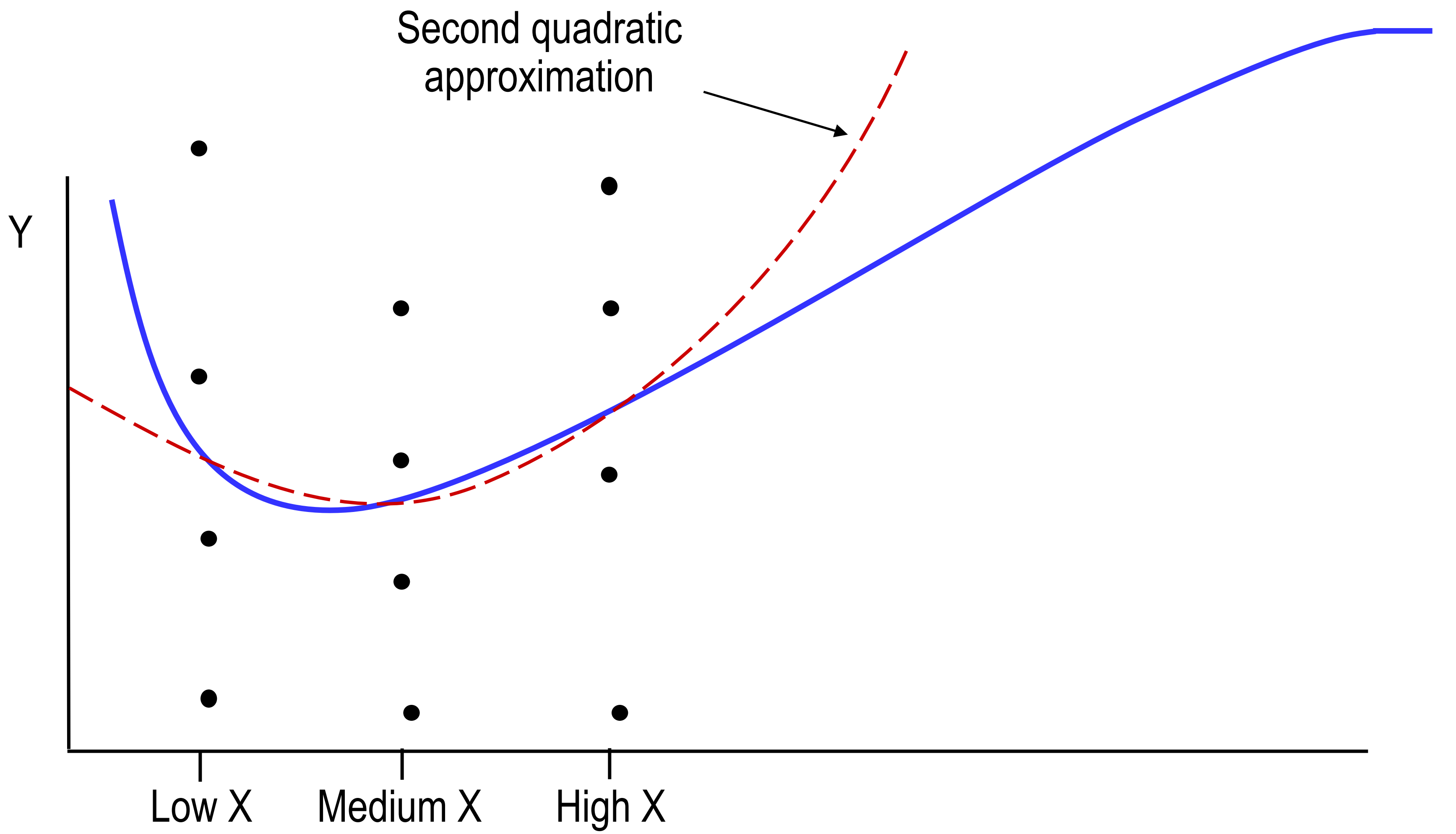
Iterated quadratic experiments

First experiment, wide ranges → "big picture"



Iterated quadratic experiments (cont'd)

Second experiment, narrow ranges → accurate modeling

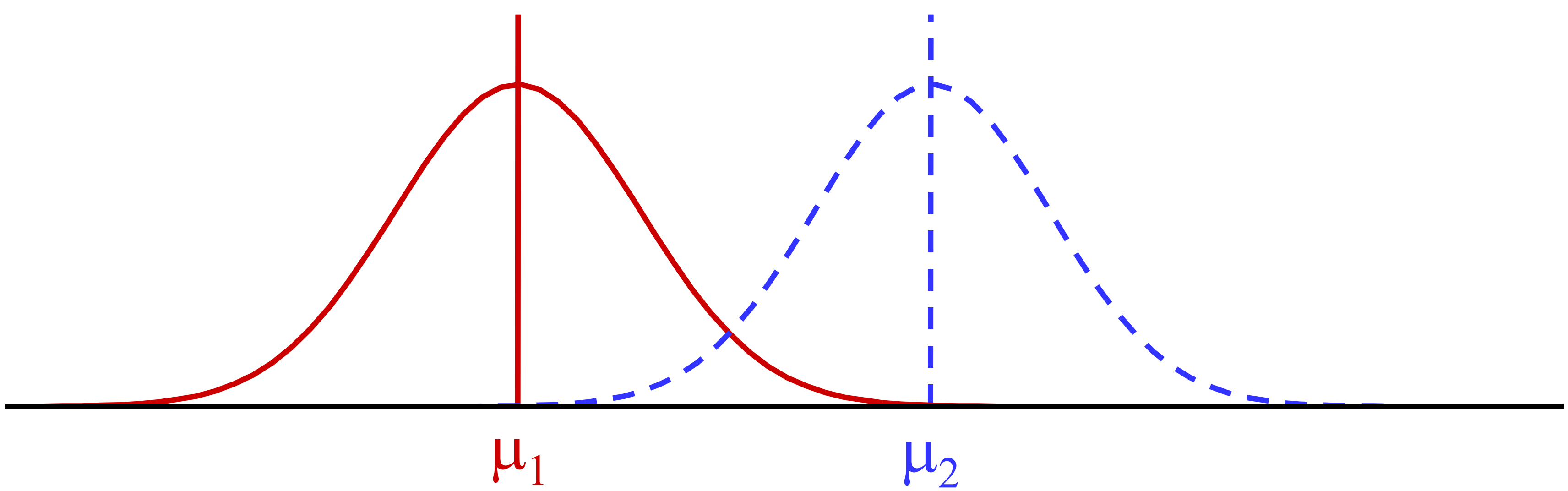


Review: Models for categorical factors

Two-level categorical factor

MATL = Steel or Rubber

$$\text{Average COST} = \begin{cases} \mu_1 & \text{if MATL} = \text{Steel} \\ \mu_2 & \text{if MATL} = \text{Rubber} \end{cases}$$



Equation form of model

Categorical factors are represented by *indicator* variables (also known as *dummy* variables)

$$\text{Average COST} = b_0 + b_1 \text{MATL[Steel]}$$

$$\text{MATL[Steel]} = \begin{cases} 1 & \text{if MATL} = \text{Steel} \\ -1 & \text{if MATL} = \text{Rubber} \end{cases}$$

Simple linear model with all factors categorical

$$\text{Avg. COST} = b_0 + b_1\,\text{LGR[Low]} + b_2\,\text{MATL[Steel]} + b_3\,\text{USAGE[50\%]} + b_4\,\text{GRIT[30]}$$

4.868125
+ Match(LGR) { "High" ⇒ -0.616875; "Low" ⇒ 0.616875; else ⇒ . }
+ Match(MATL) { "Rubber" ⇒ 1.145625; "Steel" ⇒ -1.145625; else ⇒ . }
+ Match(USAGE) { "50%" ⇒ 1.054375; "75%" ⇒ -1.054375; else ⇒ . }
+ Match(GRIT) { "30" ⇒ -0.048125; "50" ⇒ 0.048125; else ⇒ . }

- Analogy: blue book pricing of used cars
- Base price + extra for power windows
  + extra for air conditioning
  + extra for cruise control
  etc.

Categorical interaction model

$$
\begin{aligned}
\text{Avg. COST} = \; & b_0 \\
& + b_1 \, \text{LGR[Low]} \\
& + b_2 \, \text{MATL[Steel]} \\
& + b_3 \, \text{USAGE[50\%]} \\
& + b_4 \, \text{GRIT[30]} \\
& + b_5 \, \text{LGR[Low]} \times \text{MATL[Steel]} \\
& + b_6 \, \text{LGR[Low]} \times \text{USAGE[50\%]} \\
& + b_7 \, \text{LGR[Low]} \times \text{GRIT[30]} \\
& + b_8 \, \text{MATL[Steel]} \times \text{USAGE[50\%]} \\
& + b_9 \, \text{MATL[Steel]} \times \text{GRIT[30]} \\
& + b_{10} \, \text{USAGE[50\%]} \times \text{GRIT[30]}
\end{aligned}
$$

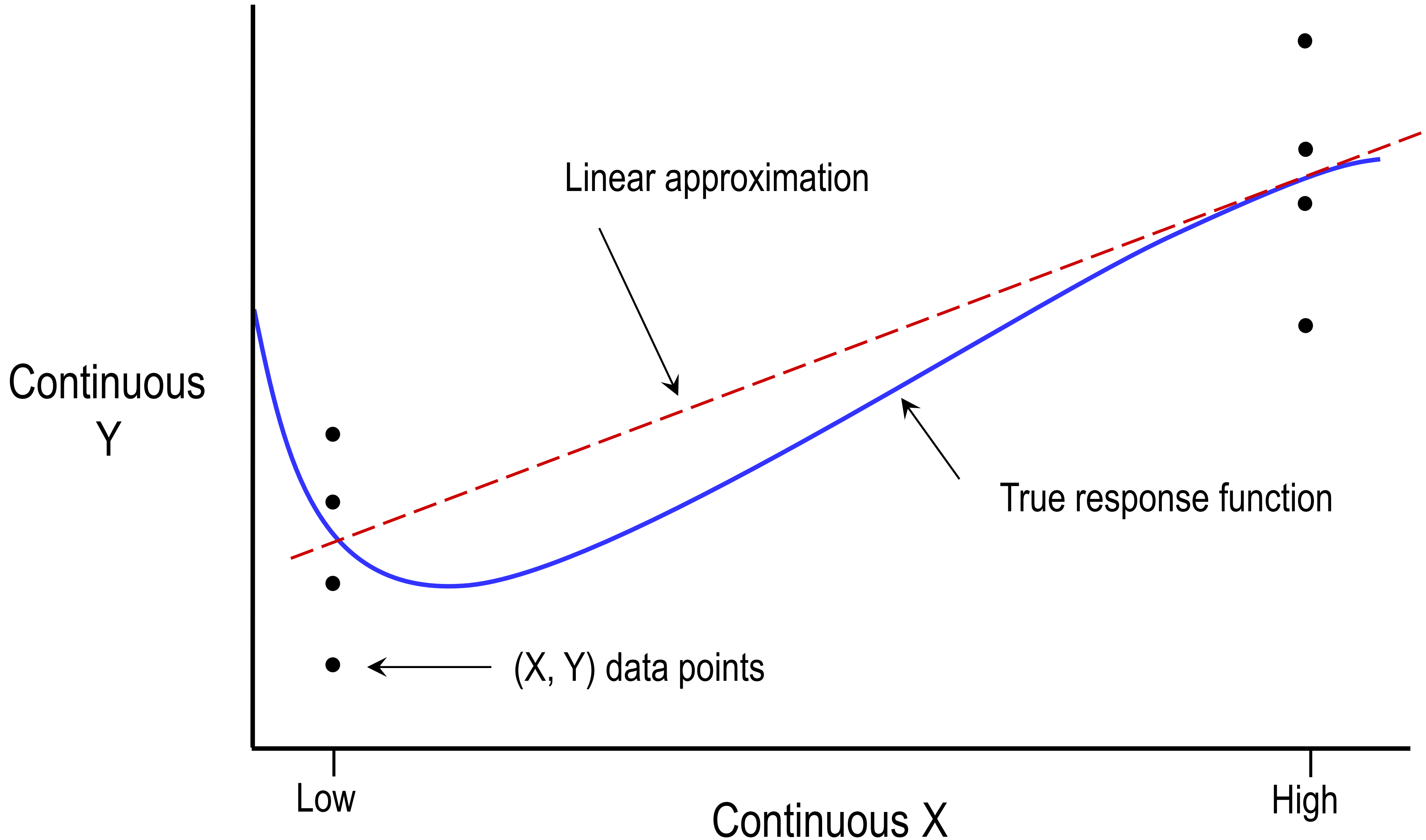
# Factors	4	5	6
Full factorial (FF)	16	32	64
Min. sample size	11	16	22
% of FF	69	50	34

7 Design Principles

- Bold strategy
- "Control group"
- Replication
- Randomization
- "Blocking"

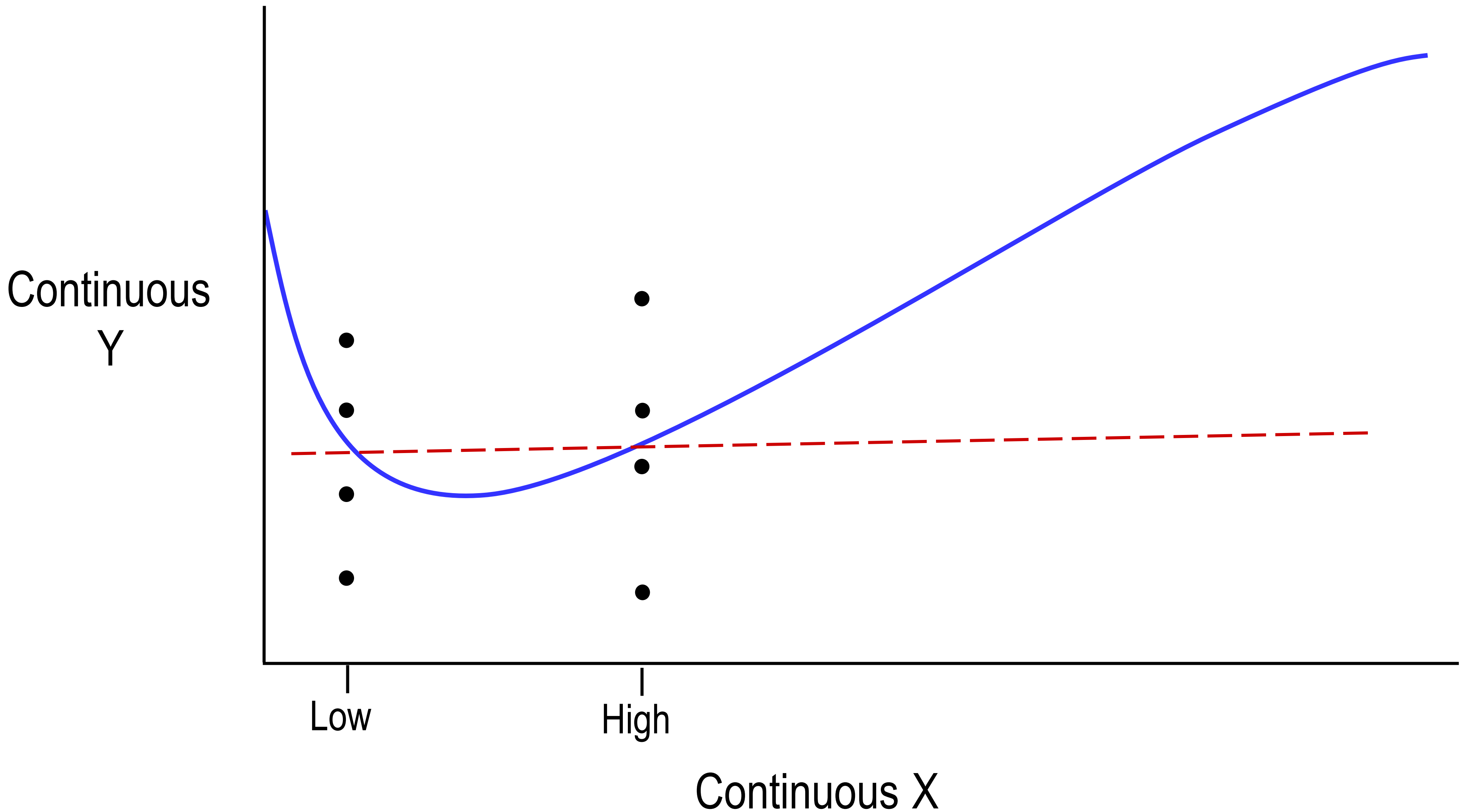
Bold strategy

Use the ***entire feasible operating range*** *in a first experiment*



Not bold enough

- Low and high levels of X are too close together
- We mistakenly conclude that X has no effect on Y



"Control group"

For each factor, one of the levels should match the current process

- Ideally, this is the middle level for continuous factors
- At least one run in the experiment should match the current process settings, for a "sanity check"
- In these types of designs, we don't usually refer to this as a "control group"

Temp	Press	Dwell	Mat'l
120	50	0.2	A
120	100	1.1	B
120	150	2.0	C
150	50	1.1	C
150	100	2.0	A
150	150	0.2	B
180	50	2.0	B
180	100	0.2	C
180	150	1.1	A

The units involved in a DOE may turn out to be uniformly different from those in current production – either better or worse. This can be due to the effects of noise variables on production units, or to special circumstances surrounding the creation and handling of experimental units.

For each factor, one of the DOE levels should match the current state value of that factor. This allows valid comparisons between current state and experimental process settings. This is especially important when non-routine measurements, tests or inspections are applied to experimental units.

Replication

Use a replicate or a replicate run to quantify the error in the experiment.

This improves estimates of coefficients and precision in determining factor significance.

Temp	Press	Experimental units
120	50	1
120	50	2
120	150	3
120	150	4
180	50	5
180	50	6
180	150	7
180	150	8

Replication forces redundancy into the experiment. This is necessary for two reasons:

- To quantify the magnitude of error in the experimental data – differences between units at the same design point are, by definition, due to error (variation in the process that is not accounted for in the factors).

- To reduce the influence of error on the experimental results by estimating "pure error." This increases the signal-to-noise ratios.

Assume that you are the person responsible for running the experiment and for the validity of the results. Is there anything about the run order shown above that makes you nervous? Please explain.

Randomization

Use a random number generator to determine the sequence in which experimental units are created and tested (JMP does this for you.)

Temp	Press	Experimental units
120	150	1
180	50	2
180	50	3
120	150	4
180	50	5
120	150	6
180	150	7
120	50	8

Randomization

Benefits

- Reduces the chance of biased results due to nuisance variables (factors not included in the experiment that may be changing while the experiment is being conducted)
- Doesn't require control of nuisance variables, which may be unknown or uncontrollable
- Results are more convincing to skeptics

What happens if you don't randomize?

- Nuisance (noise) variables may be changing during your experiment
- This increases the chance of drawing the wrong conclusions from your experiment (significant factors, best levels, etc.)
- Randomization guards against this

Drawbacks

- Impractical when some of the factors are hard to change
- We'll see what to do about this later

Blocking

Blocking allows you to account for some nuisance variables

- Nuisance variables or factors are used to divide the experiment into homogeneous "blocks"
- Effects of nuisance factors are separated from effects of other factors, for more accurate analysis of factor significance

Temp	Press	Experimental units
120	50	1
120	150	2
180	150	3
180	50	4
180	150	5
180	50	6
120	50	7
120	150	8

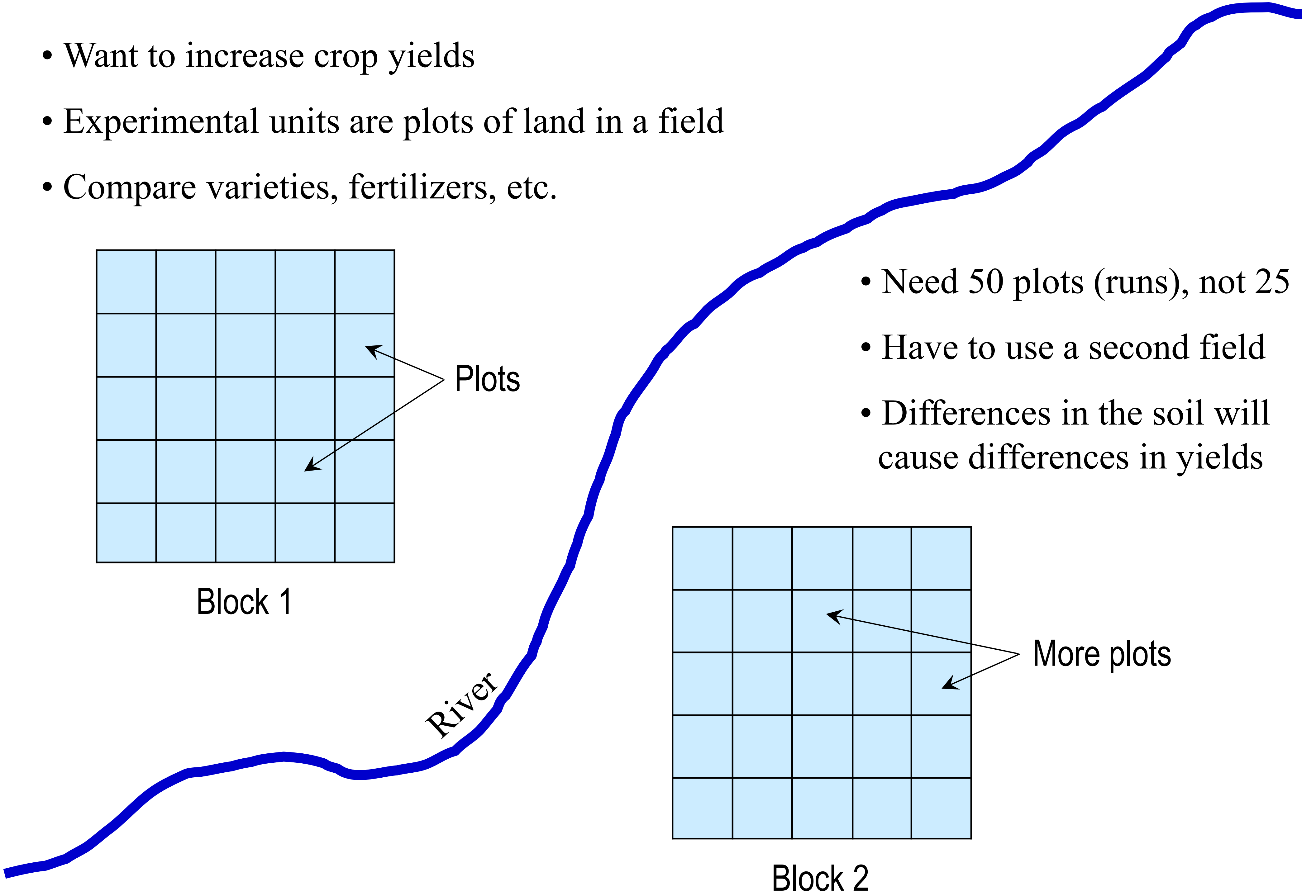
Block 1	
Operator	Bob
Shift	1
Machine	A
Material	Lot 6

Block 2	
Operator	Carol
Shift	2
Machine	B
Material	Lot 7

Agricultural origin of "blocking"

- Want to increase crop yields
- Experimental units are plots of land in a field
- Compare varieties, fertilizers, etc.

- Need 50 plots (runs), not 25
- Have to use a second field
- Differences in the soil will cause differences in yields



Why use blocking?

- Use blocking when experimental runs cannot be completed within a timeframe (shift, time allotted on a machine, etc.) or some other constraint (batch of material, space, etc.)

- Blocking systematically eliminates the effect of known, controllable nuisance (noise) factors
  - Makes predictions more reliable
  - Quantifies the effects of nuisance variables

- Improves precision with which treatment means are compared, without increasing sample size
  - Makes identification of important (significant) factors more reliable

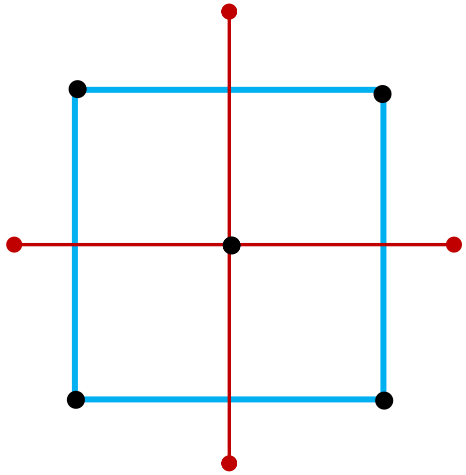
- Protects against variation due to known factors not included in the experiment

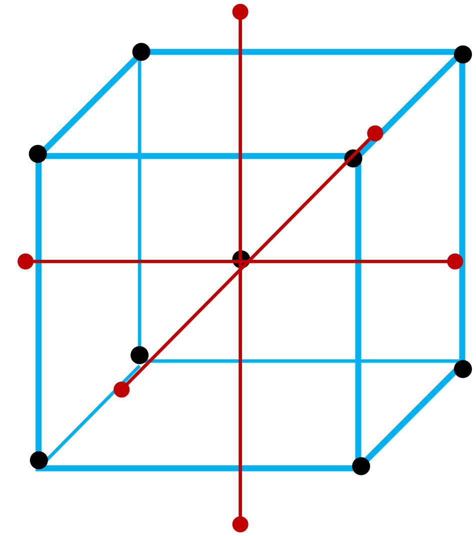
We saw the Full-Factorial Design earlier, and learned:

- A 2^k full-factorial design can estimate main effects and interactions, but cannot estimate quadratic terms
- A three level full-factorial (3^k) design can estimate main effects, interactions and quadratic effects, but is an inefficient design.

Let's look at some other designs.

Response Surface Designs





The central composite design (CCD) is a 2^k factorial with added axial or star runs.

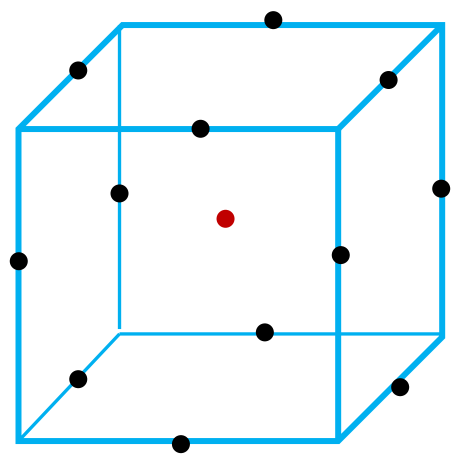
It is (was) the most used response surface design when all factors are continuous

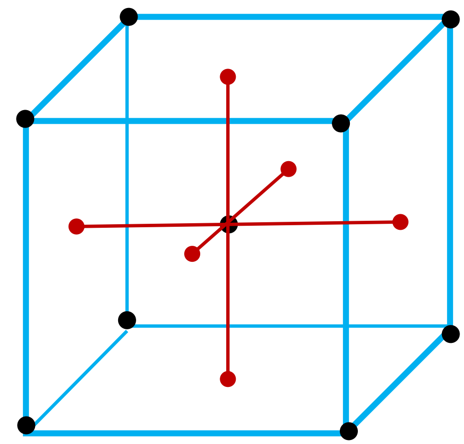
Above are images of two and three factor CCDs

- The CCD requires two axial runs for each factor, plus the 2^k design runs
- 3 – 5 center points are recommended
- Total runs required for the 3-factor CCD are 8 + 6 + center points = 17-19.

A Response Surface Design can estimate main effects, 2-factor interactions and quadratic effects, with more efficiency than the 3^k full-factorial.

Response Surface Designs (cont'd)





Box-Behnken designs (left) are spherical, and do not have any points on the corners of the "cube" contained by the limits of the factors.

The face-centered cube (right) is a variation on the Central Composite Design, with axial points on the centers of the faces of the cube (for k=3).

- 3 – 5 center points are recommended for each of these designs
- Total runs required for the 3-factor Box-Behnken design is 15-17.
- Total runs required for the face-centered cube is the same as the CCD (17-19).

As Response Surface Designs, these can estimate main effects, 2-factor interactions and quadratic effects.

Custom Designs

JMP's Custom Design platform uses modern computing power to employ a coordinate-exchange algorithm for determining the best set of points to use in a Response Surface Design, creating an "optimal design."

Often, fewer runs are required than the classical designs just presented.

When you look at the points chosen for your experiment, you may notice:

- Center points--all continuous factors at the middle level of the range given
- Points at the corners of the "cube"--all factors at high or low levels
- Points in the centers of the "cube" edges (Box-Behnken) or faces (face-centered cube)—some factors at the middle level, others at high or low levels
- You will not see axial runs extending beyond the "cube," as in the original CCD

Because fewer runs are used in these designs,
there will be some correlations and aliasing between terms.

(See Design Evaluation > Color Map on Correlations)

Steps for Creating a Custom Design

1. Specify the Responses and general goals (maximize, minimize, or match target).

2. Specify the Factors.
   - For continuous factors, specify the high and low levels.
   - For categorical factors, specify each level to be included in the experiment.

3. Specify the statistical Model (usually *RSM*).

4. Specify the blocking factor, if blocking is needed. (Click RSM again)

   - Enter the maximum number of runs that can be completed in one block (timeframe, batch of material, etc.).
   - JMP will evenly split required runs into blocks no larger than the number specified

5. Create the design matrix. (*Make Design*)

6. If desired, use *Design Evaluation* > *Power Analysis* to determine sample size.

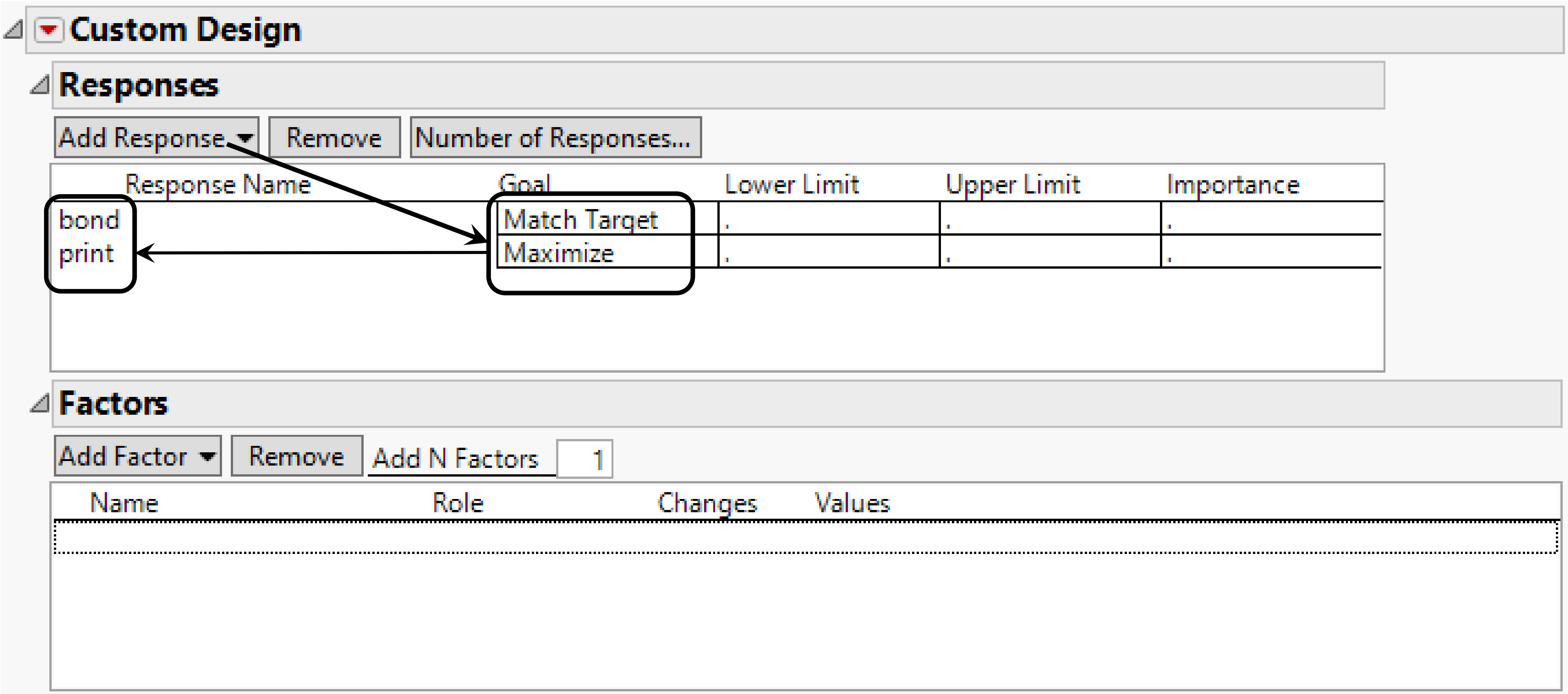
7. Back up to make changes *(Back)*, or create the data table *(Make Table)*.

8. Save the table.

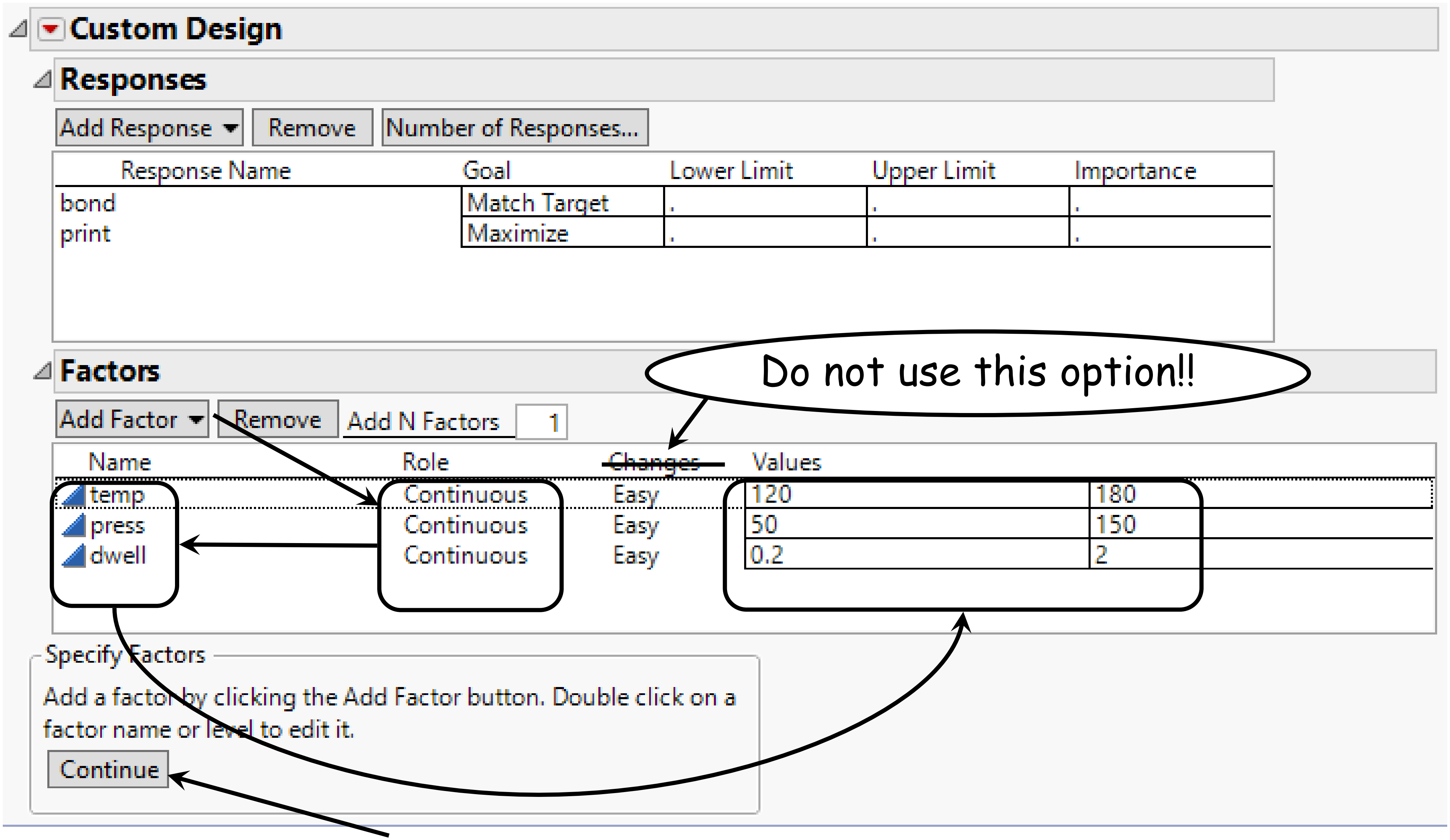
Later: Run the experiment in the order given. Enter results into table.

1. Specify the Responses and general goals

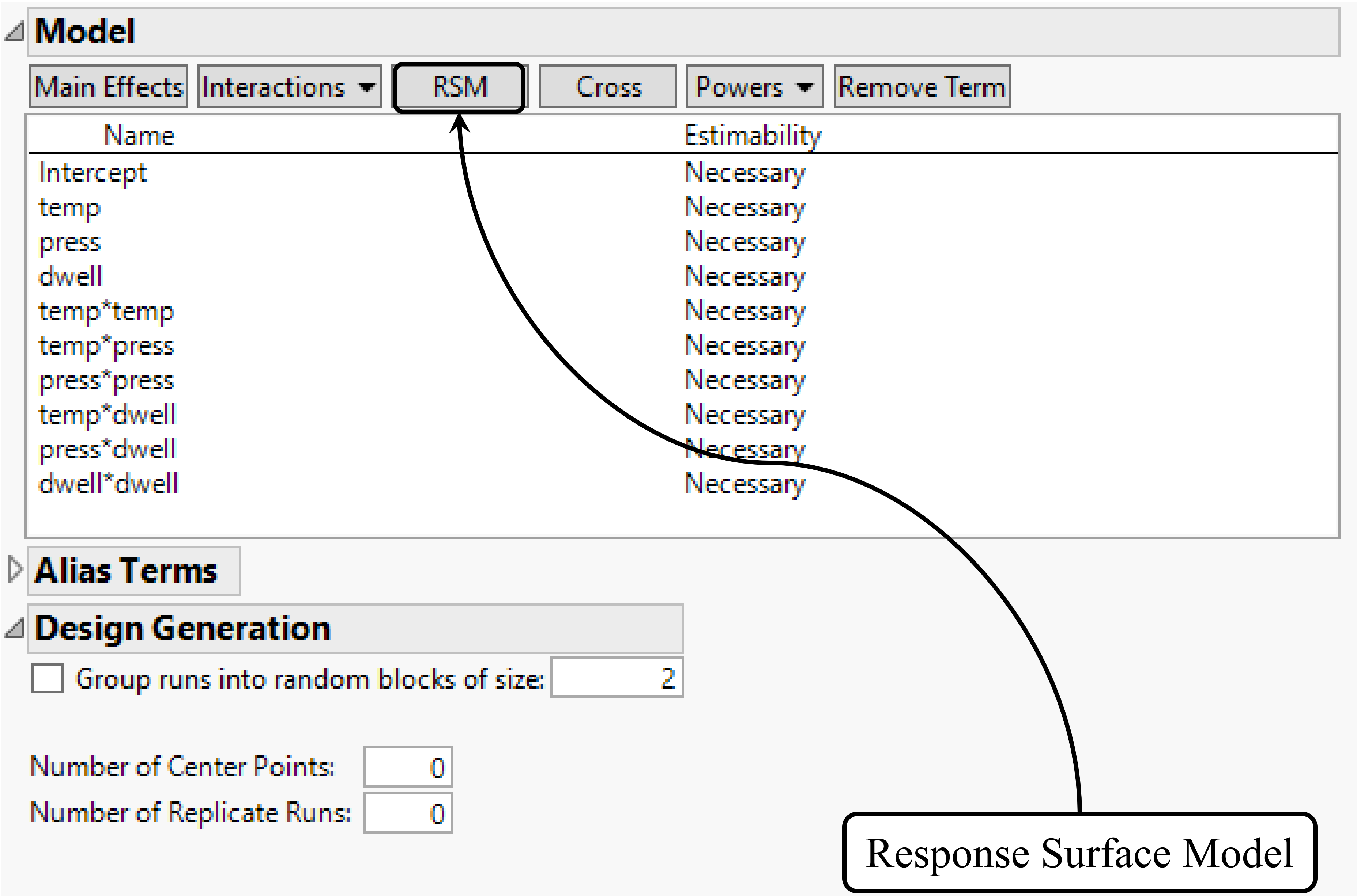
DOE → Custom Design



2. Specify the Factors



3. Specify the statistical Model (usually *RSM*)



Do not label your blocks until **after** you have done this!

4. Specify the blocking factor, if blocking is needed.

Once you specify the Model, the Default and Minimum Number of Runs are displayed.

Use this information, or User Specified Number of Runs (another sample size you've determined), to decide whether Blocking is needed.

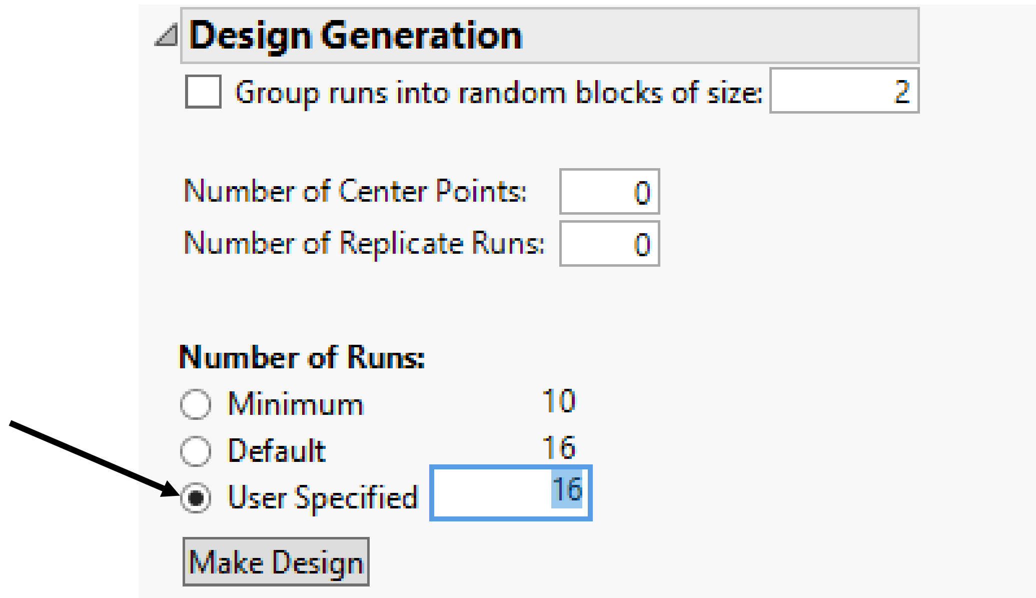
It's not a bad idea to split your experiment into blocks just in case, if it is likely to take several hours or more to complete. For example, you may have a block size equal to half of a shift, just in case there's an evacuation, or the machine goes down, or you get called away urgently, and cannot complete the experiment all at one time.

If Blocking is needed:

1. Click User Specified Number of Runs, even if you want to use the Default (this prevents JMP from increasing the sample size to a multiple of the block size),
2. Go back up to Factors to enter a Blocking factor,
3. Specify Model (click RSM) again.

4. Specify the blocking factor, if blocking is needed. (cont'd)

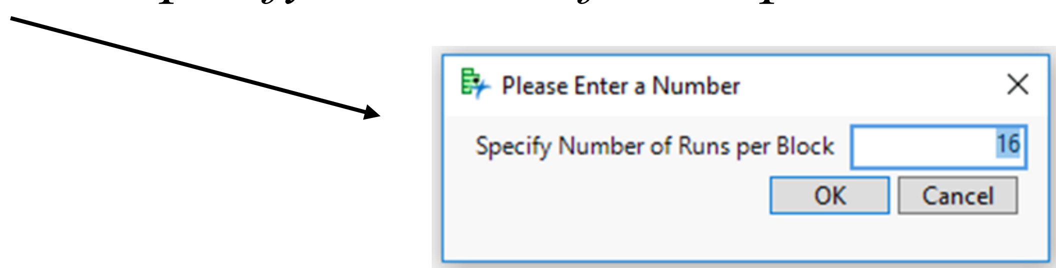
- **Select *User Specified Number of Runs*** to prevent an increase due to blocking



- Go back up to factor specification:

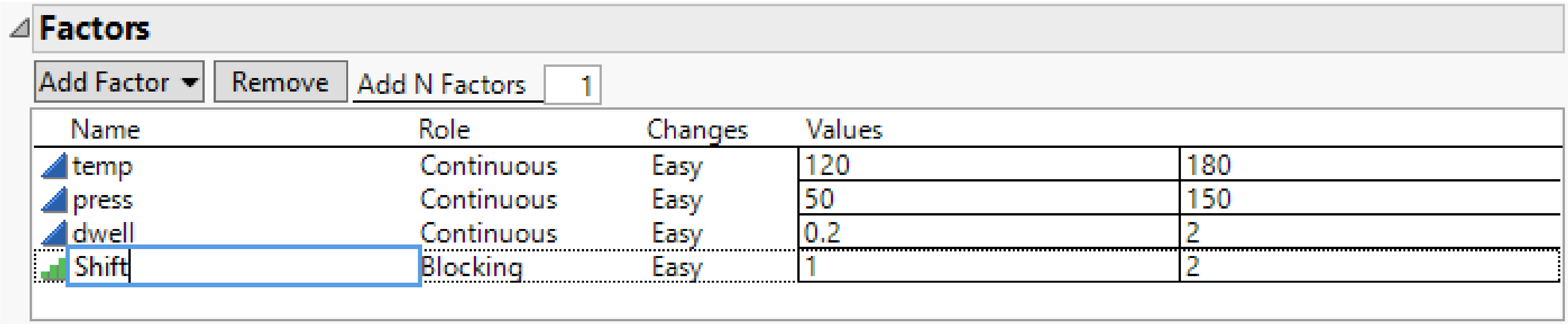
***Add Factor* > *Blocking* > Select the maximum runs possible per block**

If your maximum is not listed,
select *Other…* to *Specify Number of Runs per Block*



4. Specify the blocking factor, if blocking is needed. (cont'd)

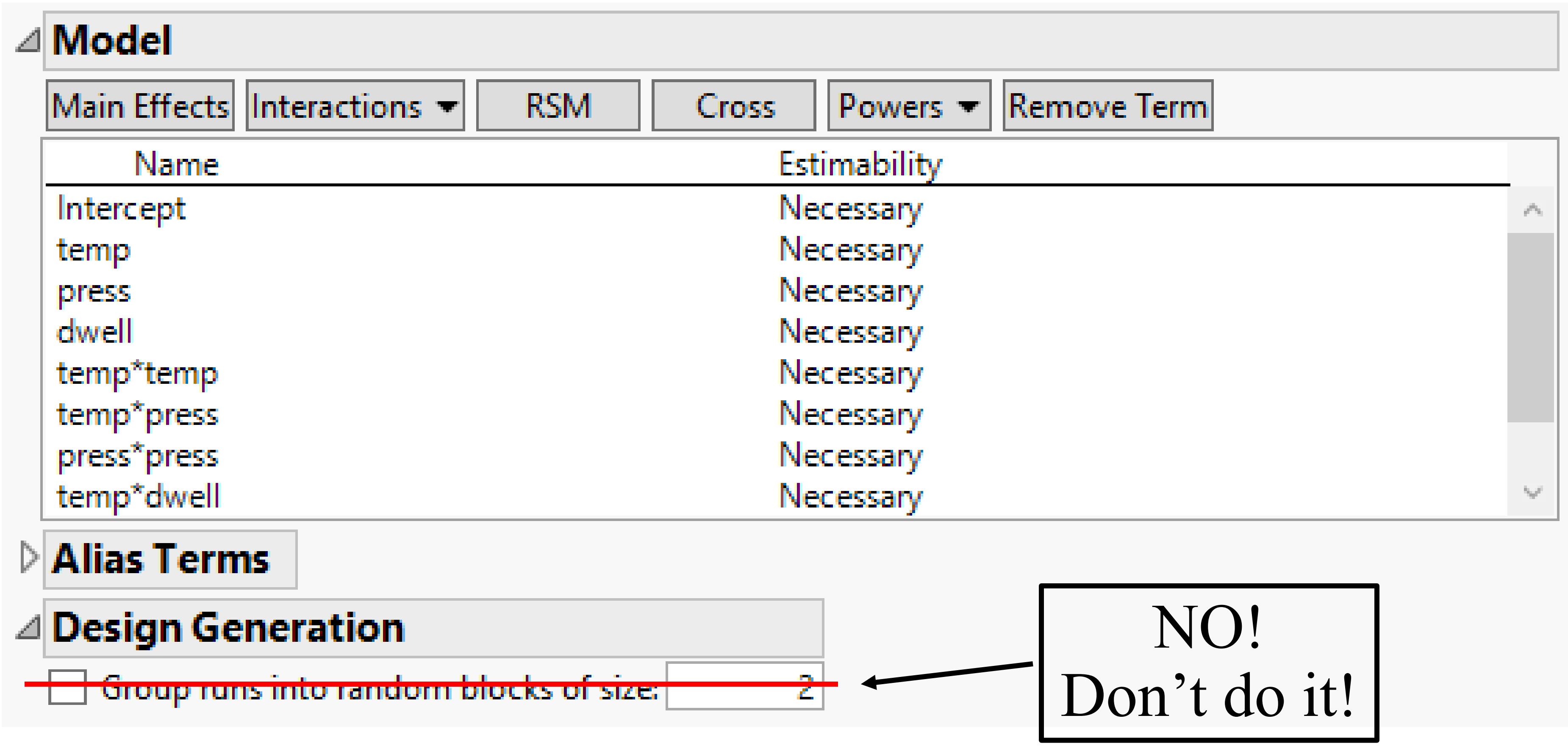
- Name the Blocking factor, so you will recognize it in the Design Matrix and Table:



- You do not need to be concerned about how many "levels" are shown under "Values." JMP will handle this when it creates the design.
- **Re-specify the Model. (Click RSM again.)** Click through JMP comments about categorical and blocking factors in RSM models.

4. Specify the blocking factor, if blocking is needed. (cont'd)

DO NOT use this option for setting up a blocking factor!



JMP will generate uneven block sizes, if this option is used.

5. Create the Design Matrix.

Model

Main Effects | Interactions ▾ | RSM | Cross | Powers ▾ | Remove Term

Name	Estimability
Intercept	Necessary
temp	Necessary
press	Necessary
dwell	Necessary
Shift	Necessary
temp*temp	Necessary
temp*press	Necessary
press*press	Necessary
temp*dwell	Necessary
press*dwell	Necessary
dwell*dwell	Necessary

Alias Terms

Design Generation

Number of Center Points: 0

Number of Replicate Runs: 0

Number of Runs:

- Minimum 10
- Default 16
- User Specified 16

Make Design

Design

Run	temp	press	dwell	Shift
1	150	100	2	1
2	120	50	0.2	1
3	180	150	0.2	1
4	180	50	2	1
5	180	50	0.2	1
6	120	50	2	2
7	180	150	2	2
8	150	100	0.2	2
9	150	100	1.1	1
10	152.7	50	1.1	2
11	150	98.5	1.1	2
12	120	100	1.1	1
13	150	150	1.1	1
14	120	150	0.2	2
15	120	150	2	2
16	180	100	1.1	2

Design Evaluation

Output Options

Data Table Options

- Save X Matrix
- Simulate Responses
- Include Run Order Column

Run Order: Randomize within Blocks

Make Table

Back

Don’t worry about the order of the blocking factor (Shift). This will be reordered when you Make Table.

6. If desired, use Power Analysis* to determine sample size.

Design Evaluation > Power Analysis

Design Evaluation

Power Analysis

Significance Level 0.05

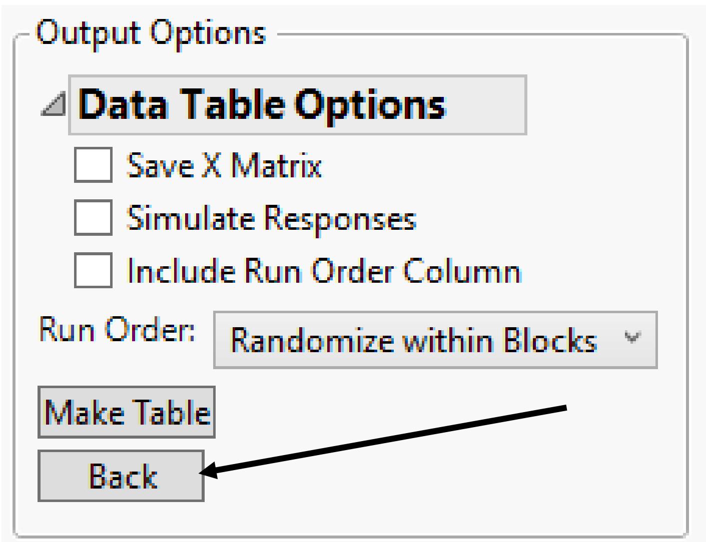
Anticipated RMSE 1

Term	Anticipated Coefficient	Power
Intercept	1	0.402
temp	1	0.706
press	1	0.706
dwell	1	0.705
Shift	1	0.865
temp*temp	1	0.262
temp*press	1	0.623
press*press	1	0.262
temp*dwell	1	0.623
press*dwell	1	0.623
dwell*dwell	1	0.263

* Details of this procedure are presented later, in the Determining Sample Size section.

7. Back up to make changes or create the data table.

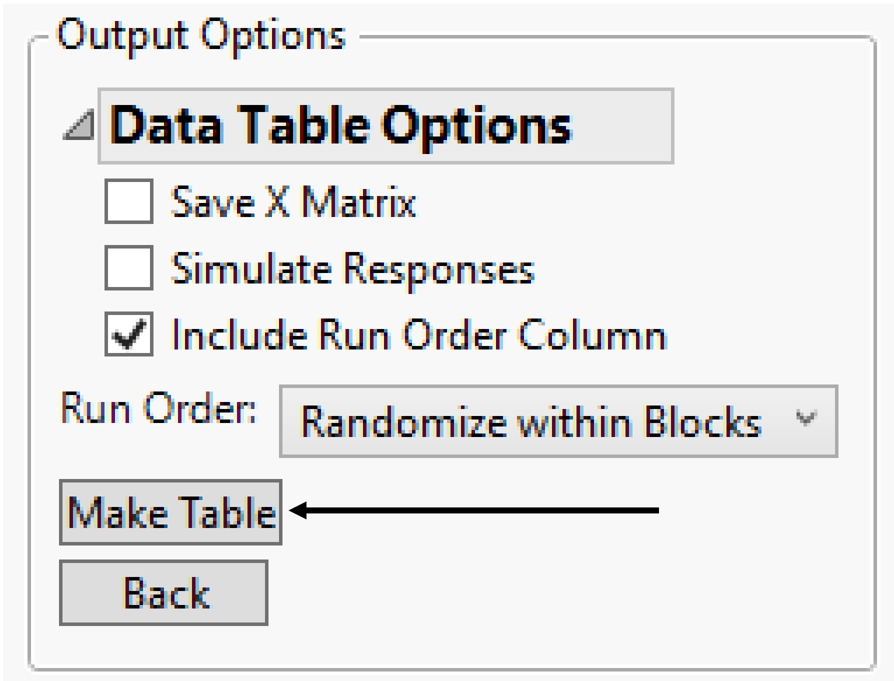
- Click *Back* to back up and adjust sample size.
- Adjust User Specified Number of Runs
- Click *Make Design*



Once the design is as needed:

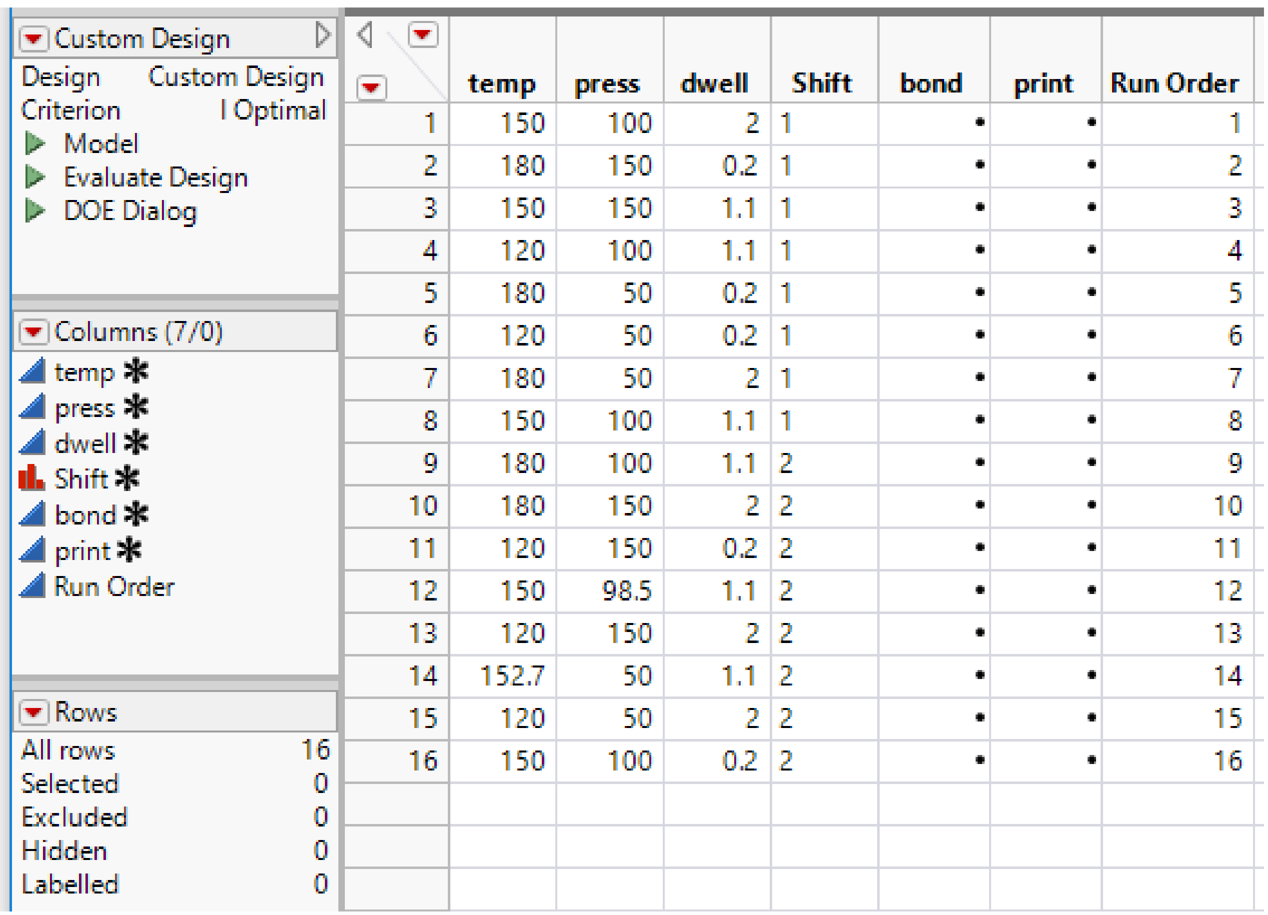
- Check *Include Run Order Column*
- click *Make Table*

JMP creates an editable table.



8. Save the table.

- You can reorder columns and adjust any odd factor levels by entering the desired value
    - Odd levels are an artifact of the procedure JMP uses to create custom designs
    - Before creating the table, you can also back up to create another design, and see if that takes care of it
    - In this example, temp of 152.7 would be changed to 150, press of 98.5 would be changed to 100



	temp	press	dwell	Shift	bond	print	Run Order
1	150	100	2	1	•	•	1
2	180	150	0.2	1	•	•	2
3	150	150	1.1	1	•	•	3
4	120	100	1.1	1	•	•	4
5	180	50	0.2	1	•	•	5
6	120	50	0.2	1	•	•	6
7	180	50	2	1	•	•	7
8	150	100	1.1	1	•	•	8
9	180	100	1.1	2	•	•	9
10	180	150	2	2	•	•	10
11	120	150	0.2	2	•	•	11
12	150	98.5	1.1	2	•	•	12
13	120	150	2	2	•	•	13
14	152.7	50	1.1	2	•	•	14
15	120	50	2	2	•	•	15
16	150	100	0.2	2	•	•	16

- Run your experiment in the order specified and enter data into this table.
- If data is entered directly into the table as the experiment is performed, it's not a bad idea to print a copy of the table and keep a hard copy also, as you go . . . just in case.

Exercises

Use the Custom Design process described on the previous slides to create Response Surface designs for the exercises on the following pages. In addition to special instructions given in each case, follow these general instructions:

- Determine whether each factor is continuous or categorical
- Use the sample size given to determine if blocking is needed.
- For each exercise, have the instructor review your matrix when you are finished.
- Make and save each design table.

Exercise 8.1

Control factors	Levels	
Heat treat	Anneal	Solution/age
Polish	Chemical	Mechanical
Peen	Yes	No

- Response variable: *Cycles to failure*
- Blocking factor: *none*
- Experimental unit: *one small test piece*
- Sample size: 12 (constraint due to availability of test fixtures)

Exercise 8.2

Control factors	Levels	
Contact wheel land-groove ratio (*LGR*)	Low	High
Contact wheel material (*Material*)	Steel	Rubber
Belt usage limit (*Usage*)	50%	80%
Belt grit size (*Grit*)	"30"	"50"

- Response variable: *Cost*
- Blocking: At most, 10 runs can be completed in a morning or an afternoon. You want to split the runs evenly between two blocks.
- Blocking factor: *Time of day* (morning vs. afternoon)
- Experimental unit: *one large casting*
- Sample size: Use the default sample size. Enter it here ________

Exercise 8.3

Control factors	Ranges
Force	70 to 150
Energy	275 to 325
Amplitude	70 to 90

- Response variable: *Leak rate*
- Blocking constraint: Due to production needs, a maximum of 20 containers can be molded in each tool cavity
- Blocking factor: *Cavity* (parts are molded from 4 tool cavities)
- Experimental unit: *one welded plastic container*
- Sample size for experiment: 68

9 Determining Sample Size for an Experiment

Sample size, N, is the total number of “runs” in the experiment.

How should sample size be determined?

- *First, you must have at least one run for each model term.*

  More factors and more complex model → more terms and more runs

- *Second, your purpose must be clear for a given experiment.*

  Process optimization with RSM require more runs for each factor than experiments for screening for important factors

  Less ambiguity in results → more runs

- *Beyond that, there are several answers to the question of how to determine sample size. Two are presented on the following slides.*

How should sample size be determined? (cont'd)

1. **The quickest answer that most statisticians experienced in experimentation give, is that the sample size depends on your budget. Run the best designed experiment you can, within your budgetary constraints**.

    - Think through your experimental strategy before running your first experiment

    - Don't use more than about 25% of your entire budget on your first experiment

    - Compare potential designs with Design Diagnostics > Compare Designs
        - Fraction of Design Space Plot, when prediction using the model, is a goal
        - Color Map on Correlations, whenever less than a full-factorial is used

How should sample size be determined? (cont'd)

2. Use JMP's Design Evaluation > Power Analysis to ensure that:

- Main Effects (e.g. Temp, Dwell, X1) have a Power of 0.9 to 0.8
- Interactions (e.g. Temp x Dwell, X1*X2) have a Power of about 0.8
- Quadratic Terms (e.g. Temp x Temp, X1*X1) have a Power of about 0.5

- Use the Power Analysis as it is when you open it, without changing Anticipated RMSE or Coefficients (this allows good detection of effects with $\beta_n \geq$ RMSE)
- Adjust Power by going Back and changing the User Specified Number of Runs

Design Evaluation

Power Analysis

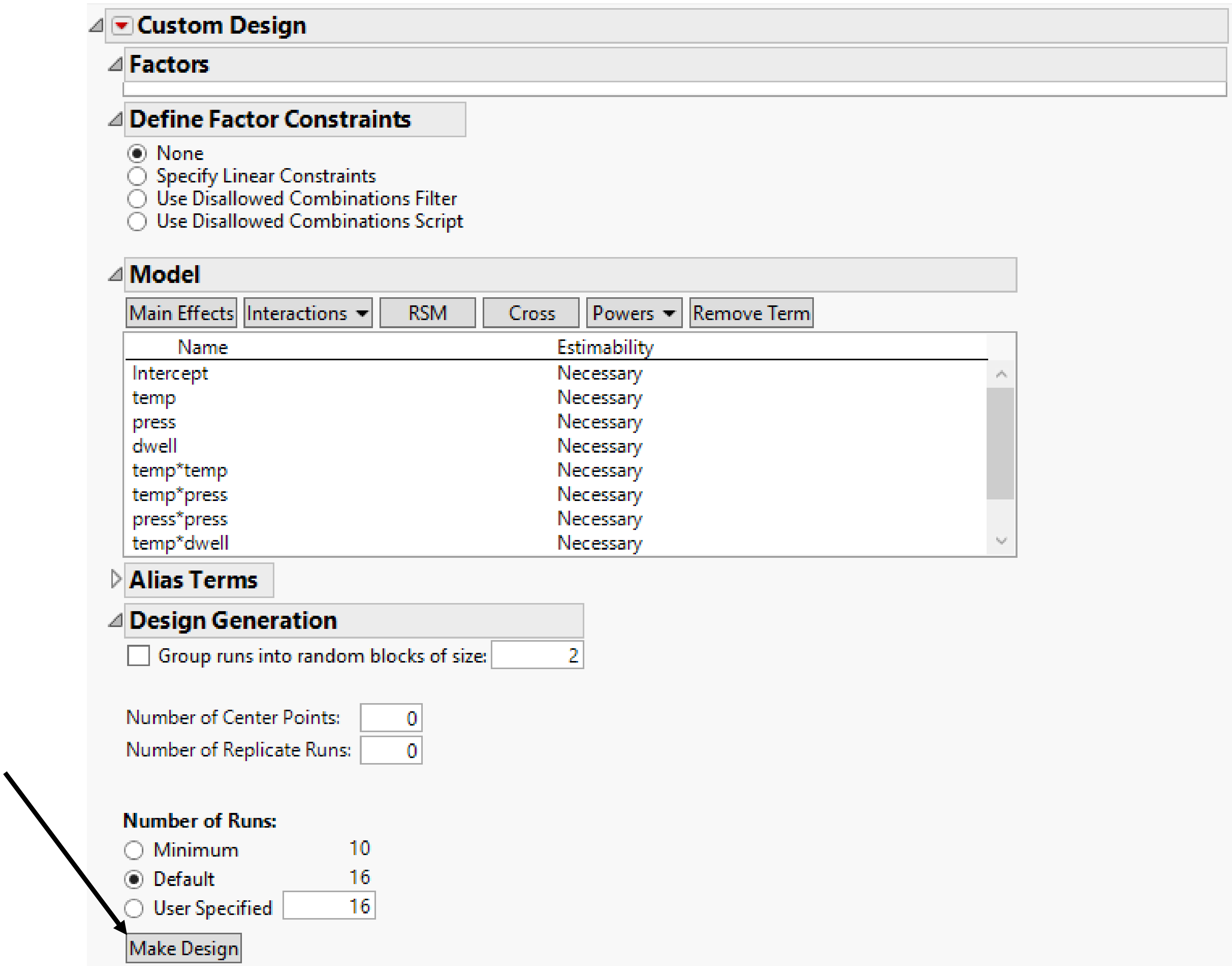
Significance Level 0.05

Anticipated RMSE 1

Term	Anticipated Coefficient	Power
Intercept	1	0.615
X1	1	0.962
X2	1	0.962
X3	1	0.962
X1*X1	1	0.547
X1*X2	1	0.899
X2*X2	1	0.547
X1*X3	1	0.899
X2*X3	1	0.899
X3*X3	1	0.547

Example: Using Power Analysis to Determine Sample Size

Set up Responses, Factors and Model, then click *Make Design*



Example: Using Power Analysis to Determine Sample Size (cont'd)

Click on the triangle next to Design Evaluation, then on the triangle next to Power Analysis to open the Power Analysis report:

Review the Power Analysis to determine if all:

- Main Effects (e.g. temp, dwell, X1) have a Power of 0.9 to 0.8
- Interactions (e.g. temp*dwell, X1*X2) have a Power of about 0.8
- Quadratic Terms (e.g. dwell*dwell, X1*X1) have a Power of about 0.5

Design Evaluation

Power Analysis

Significance Level 0.05

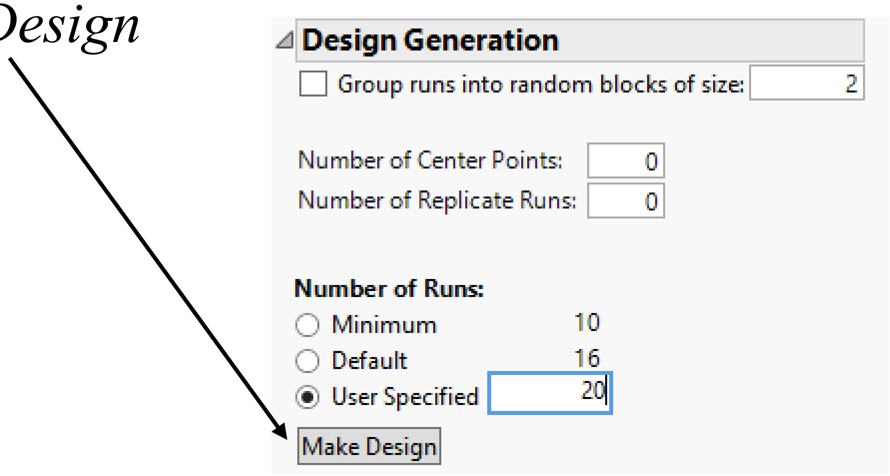
Anticipated RMSE 1

Term	Anticipated Coefficient	Power
Intercept	1	0.427
temp	1	0.75
press	1	0.75
dwell	1	0.75
temp*temp	1	0.278
temp*press	1	0.657
press*press	1	0.278
temp*dwell	1	0.657
press*dwell	1	0.657
dwell*dwell	1	0.278

In this example, all Power values are too low. The sample size needs to be increased.

Example: Using Power Analysis to Determine Sample Size (cont'd)

- Click *Back.*
- Select *User Specified* and increase the Number of Runs.
- Click *Make Design*



- Review the Power Analysis report again, to determine whether the power levels meet the requirements.
  - This may require several iterations
  - If you overshoot, go back and reduce the number of runs

Example: Using Power Analysis to Determine Sample Size (cont'd)

It took 25 runs for all model terms to exceed the desired power.

(Because every design is a little different, it's possible that a design of 24 or 26 runs could (eventually) be generated that exceed the desired power levels.)

An experimenter may choose a slightly smaller sample size, as the desired power levels are approximate ("about 0.8") and are usually conservative.

Design Evaluation

Power Analysis

Significance Level 0.05

Anticipated RMSE 1

Term	Anticipated Coefficient	Power
Intercept	1	0.615
temp	1	0.962
press	1	0.962
dwell	1	0.962
temp*temp	1	0.547
temp*press	1	0.899
press*press	1	0.547
temp*dwell	1	0.899
press*dwell	1	0.899
dwell*dwell	1	0.547

Power Analysis with Categorical Factors at more than 2 Levels

When categorical factors are at more than two levels, the Power Analysis report gets a little messy.

- Use the upper part of the Power Analysis, as before, for all continuous factor:
  - main effects
  - interactions
  - quadratic terms
- Use the little table below for all categorical factor:
  - main effects
  - interactions that include categorical factors

Design Evaluation

Power Analysis

Significance Level 0.05

Anticipated RMSE 1

Term	Anticipated Coefficient	Power
Intercept	1	0.442
Intro APR	1	0.882
Time Period	1	0.877
Gift 1	1	0.575
Gift 2	-1	0.631
Gift 3	1	0.575
Intro APR*Intro APR	1	0.297
Intro APR*Time Period	1	0.838
Time Period*Time Period	1	0.307
Intro APR*Gift 1	1	0.477
Intro APR*Gift 2	-1	0.477
Intro APR*Gift 3	1	0.477
Time Period*Gift 1	1	0.476
Time Period*Gift 2	-1	0.476
Time Period*Gift 3	1	0.476

Apply Changes to Anticipated Coefficients

Effect	Power
Gift	0.763
Intro APR*Gift	0.633
Time Period*Gift	0.629

Exercise 9.1

We are planning an experiment to optimize a monofilament extrusion process with 4 continuous factors X1 to X4. The response variable is *tensile strength*.

- Optimization experiment = Response Surface Model needed
- Use the Custom Design platform to design this experiment
- Using the Power Analysis method, determine the sample size (number of runs) required in this experiment [For consistency among class participants, find the smallest sample size that puts all factors over the recommended power levels.]

Exercise 9.2

We are planning an experiment to optimize an ultrasonic welding process with 3 continuous factors and a 4-level categorical factor. The response variable is the *weld depth*.

- Optimization experiment = Response Surface Model needed
- Use the Custom Design platform to design this experiment
- Using the Power Analysis method, determine the sample size (number of runs) required in this experiment [For consistency among class participants, find the smallest sample size that puts all factors over the recommended power levels.]

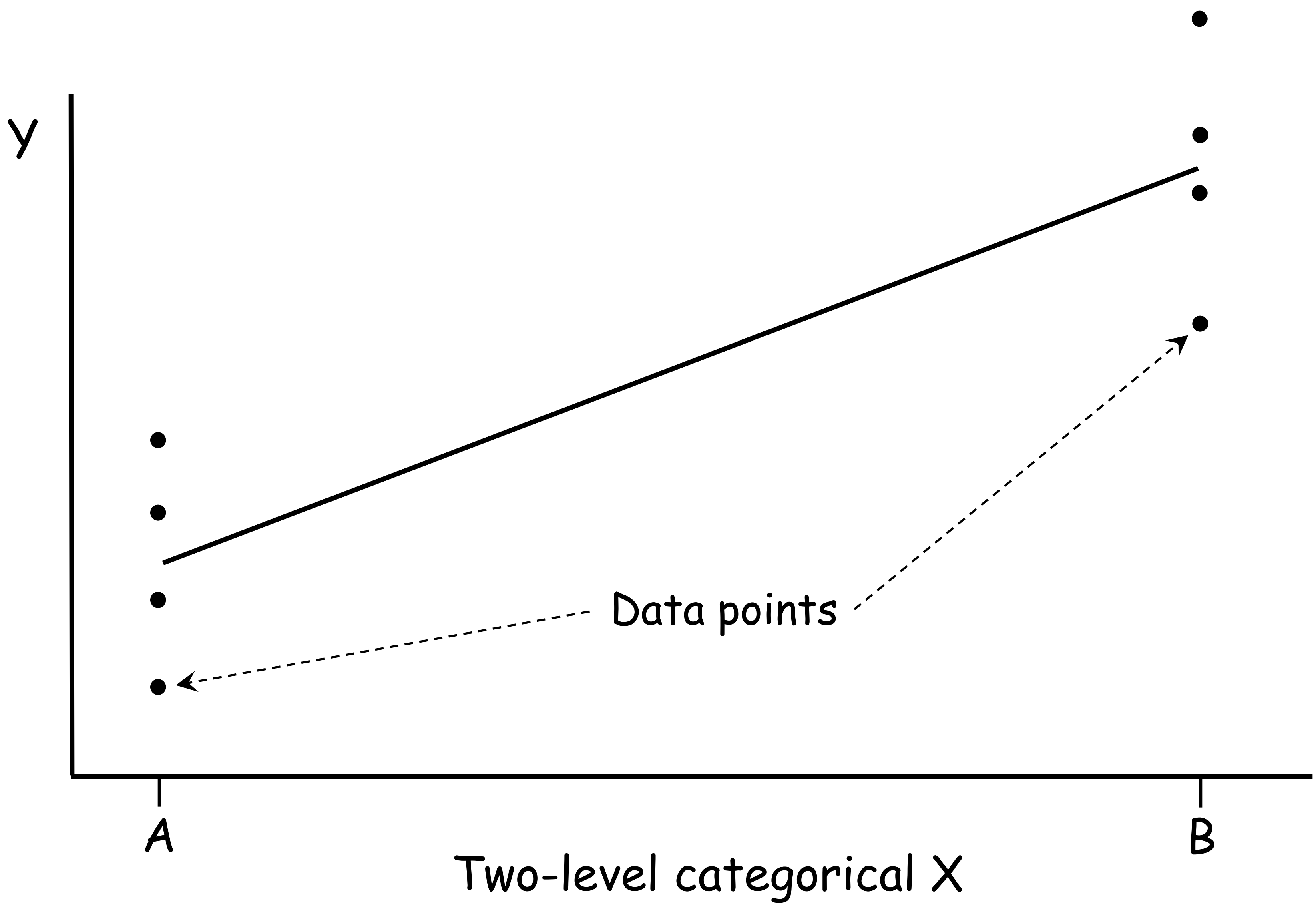
Optimization	Screening
Smaller number of factors	Larger number of factors
Main and interactive effects	Main and interactive effects if categorical factors at only 2-levels; otherwise main effects only
Quantitative factors have 3 levels	All factors have 2 levels (usually)
Identify the best factor levels	Identify the "active" factors

About screening experiments

- They are usually conducted early in the process of optimization
- They involve a relatively large number of factors
- Their objective is to identify a smaller set of influential factors for further experimentation
- It is likely that many factors considered have little or no effect on the response (sparsity-of-effects)
- They use the smallest feasible design for the given number of factors – saves time and money
- They are based on main-effect models, although with some designs, factors with interactions and quadratic effects can be identified
- They usually consist of factors at only two levels
- They rank the factors by the size of their estimated effects

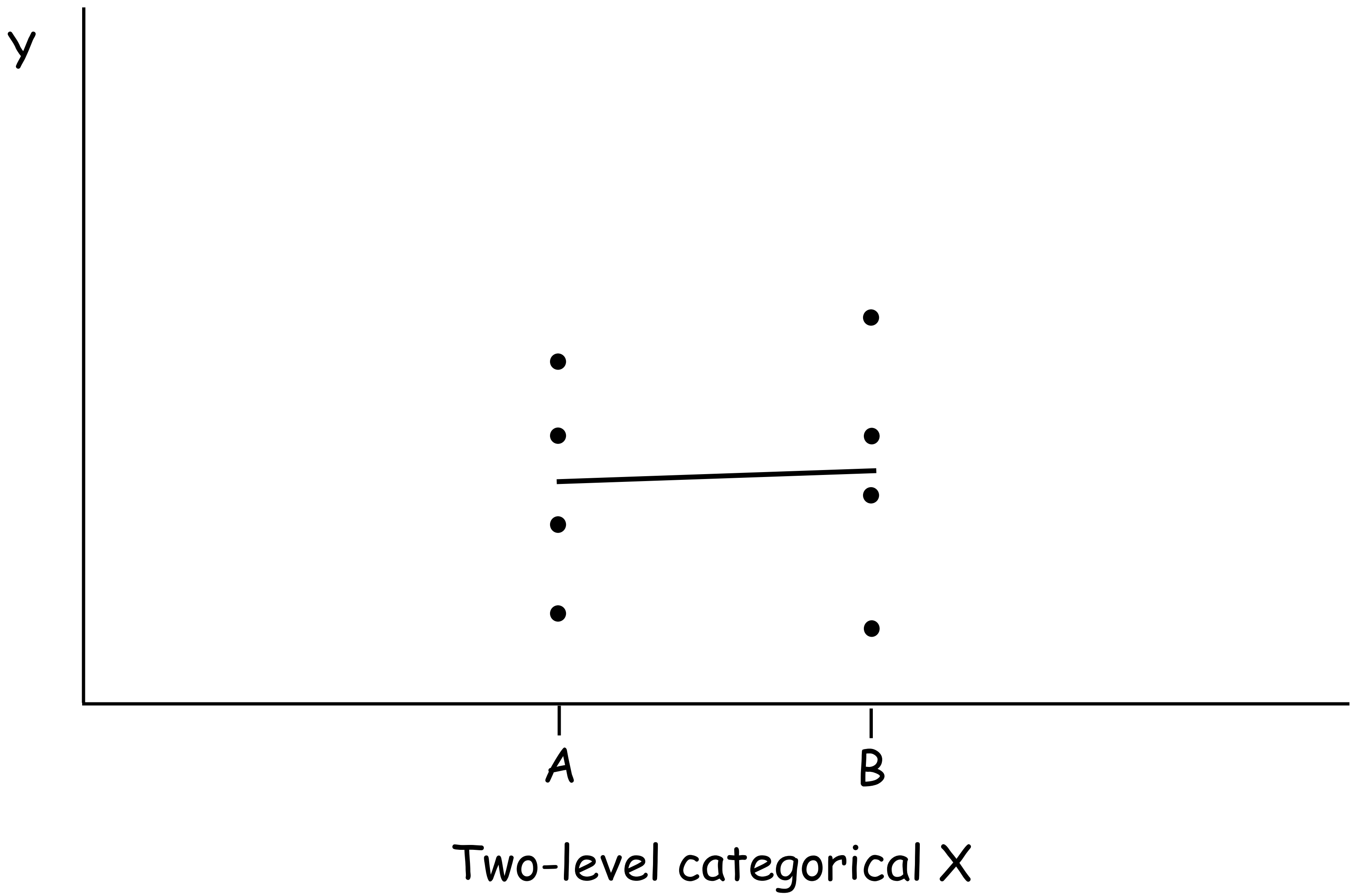
Bold strategy

Levels of X are far enough apart to quantify the effect

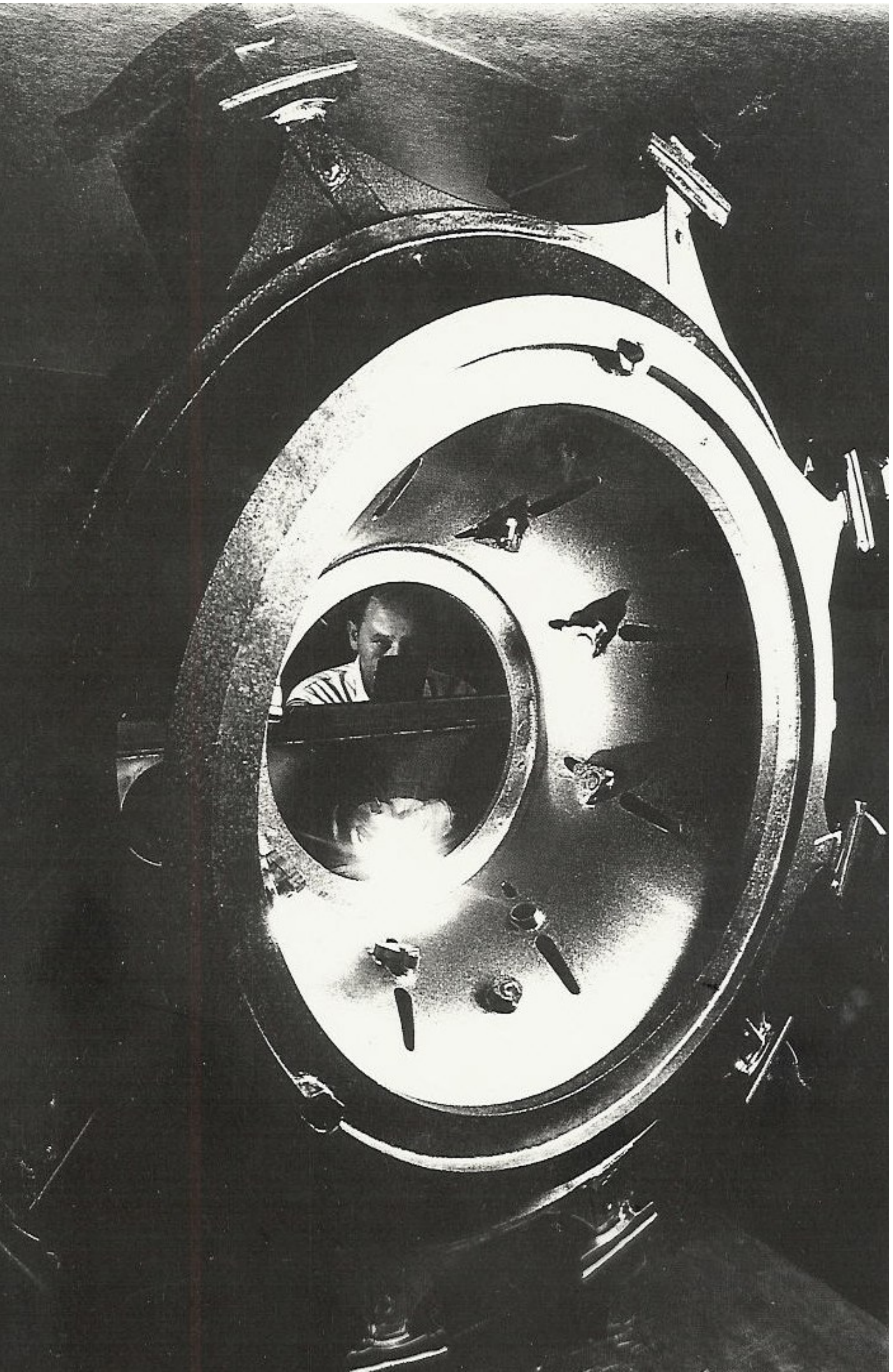


Not bold enough

Levels of X are too close to quantify the effect

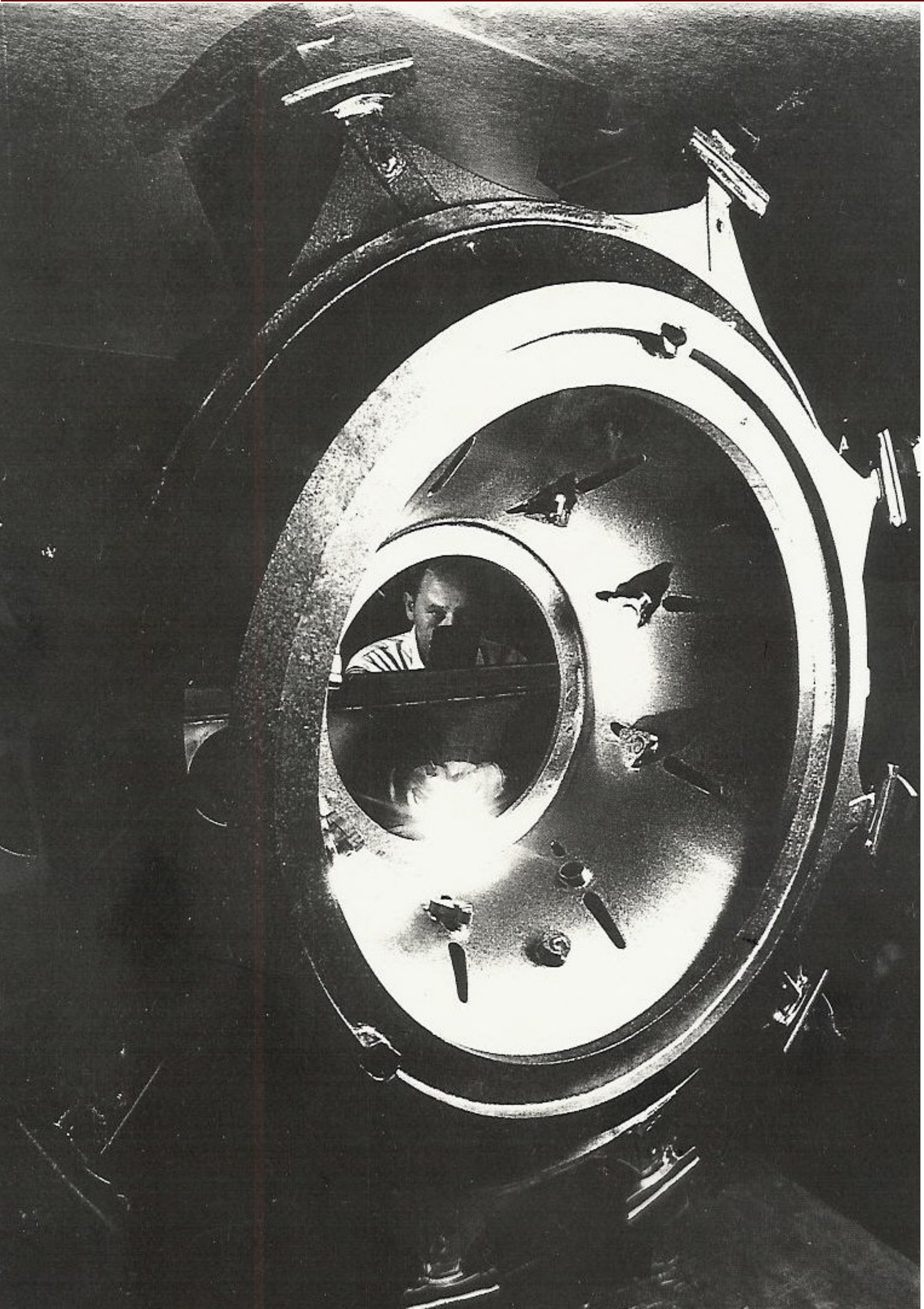


Example



- Titanium castings → strong & light
- Ti develops surface oxidation during the cooling phase
- Large Ti castings were failing the customer O_2 requirement
- Analysis of file cabinet data yielded no significant correlations

Example (cont'd)



Black Belt

"We should brainstorm factors for a DOE."

Plant manager

"We can't experiment with such an expensive part!"

Ti metallurgist

"The problem doesn't replicate on smaller parts."

Part engineer

"What have got to lose? It's been weeks since we shipped any of these!"

Example (cont'd)

Process area	Factor	Levels	Current state X variable	Possible future state solution
Shell making	Slurry	Batch 1 vs Batch 2	✓	
	# Dips	14 vs 18		✓
	Bake time	6 hrs vs 48 hrs	✓	
	Bake temp	1950° vs 2050°		✓
Casting	Alloy cost	Low vs High		✓
	Alloy status	New vs Revert	✓	
	Heat shield	Mild vs Stainless		✓
	Fan speed	2400 vs 3200		✓

Example (cont'd)

Above is the list that emerged from the brainstorming session.

- Three of the factors are variables in the current state.
- The other five are possible improvement ideas for the future state.
- Total: 8 factors
- Plant manager agreed to 16 castings
- All factors are at two levels

Steps in creating a Screening Design

1) DOE → Classical → Two Level Screening → Screening Design
2) Responses → Response Name → O2 → Goal → Minimize
3) Factors → Add all factors as in previous designs (continuous or categorical, number of levels for categorical)
4) Enter factor names and levels from the table on the previous page → Continue
5) Choose Screening Type → Construct a main effects screening design → Continue → Make Design → Make Table
6) (The matrix below has been sorted by **Slurry**, **# Dips**, **Bake time** and **Bake temp)**
7) Save as **Ti casting alpha case**

Design matrix

Ti casting alpha case - JMP

File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help

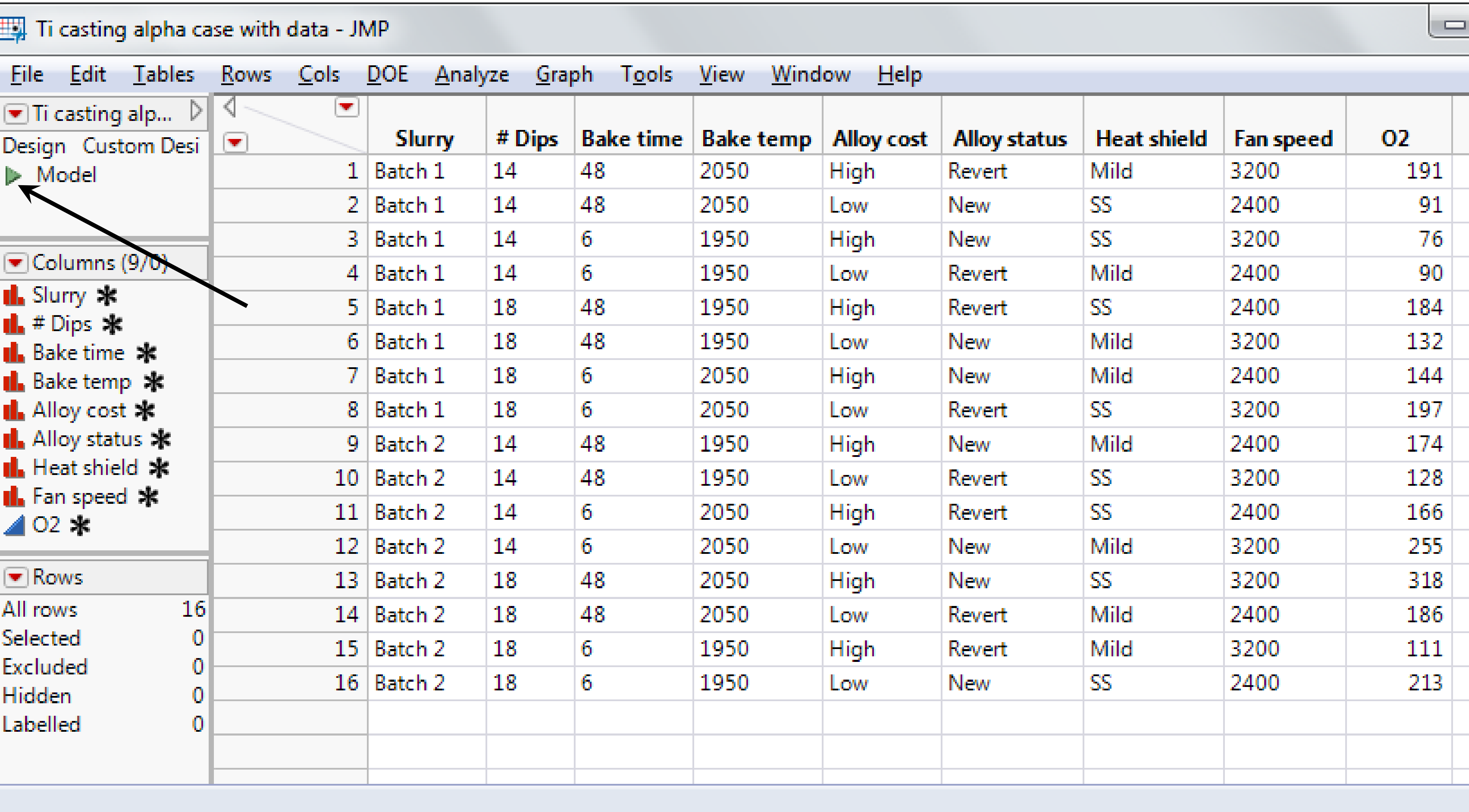
9/0 Cols
16/0

	Slurry	# Dips	Bake time	Bake temp	Alloy cost	Alloy status	Heat shield	Fan speed	O2
1	Batch 1	14	6 hrs	1950°	High	New	Mild	3200	•
2	Batch 1	14	6 hrs	1950°	High	Revert	Stainless	2400	•
3	Batch 1	14	6 hrs	2050°	Low	Revert	Mild	3200	•
4	Batch 1	14	48 hrs	2050°	High	New	Stainless	2400	•
5	Batch 1	18	6 hrs	2050°	Low	Revert	Stainless	2400	•
6	Batch 1	18	48 hrs	1950°	High	New	Mild	2400	•
7	Batch 1	18	48 hrs	1950°	Low	Revert	Stainless	3200	•
8	Batch 1	18	48 hrs	2050°	Low	New	Mild	3200	•
9	Batch 2	14	6 hrs	2050°	Low	New	Mild	2400	•
10	Batch 2	14	48 hrs	1950°	Low	Revert	Mild	2400	•
11	Batch 2	14	48 hrs	1950°	Low	New	Stainless	3200	•
12	Batch 2	14	48 hrs	2050°	High	Revert	Stainless	3200	•
13	Batch 2	18	6 hrs	1950°	High	Revert	Mild	3200	•
14	Batch 2	18	6 hrs	1950°	Low	New	Stainless	2400	•
15	Batch 2	18	6 hrs	2050°	High	New	Stainless	3200	•
16	Batch 2	18	48 hrs	2050°	High	Revert	Mild	2400	•

Analyzing the Screening Experiment . . .

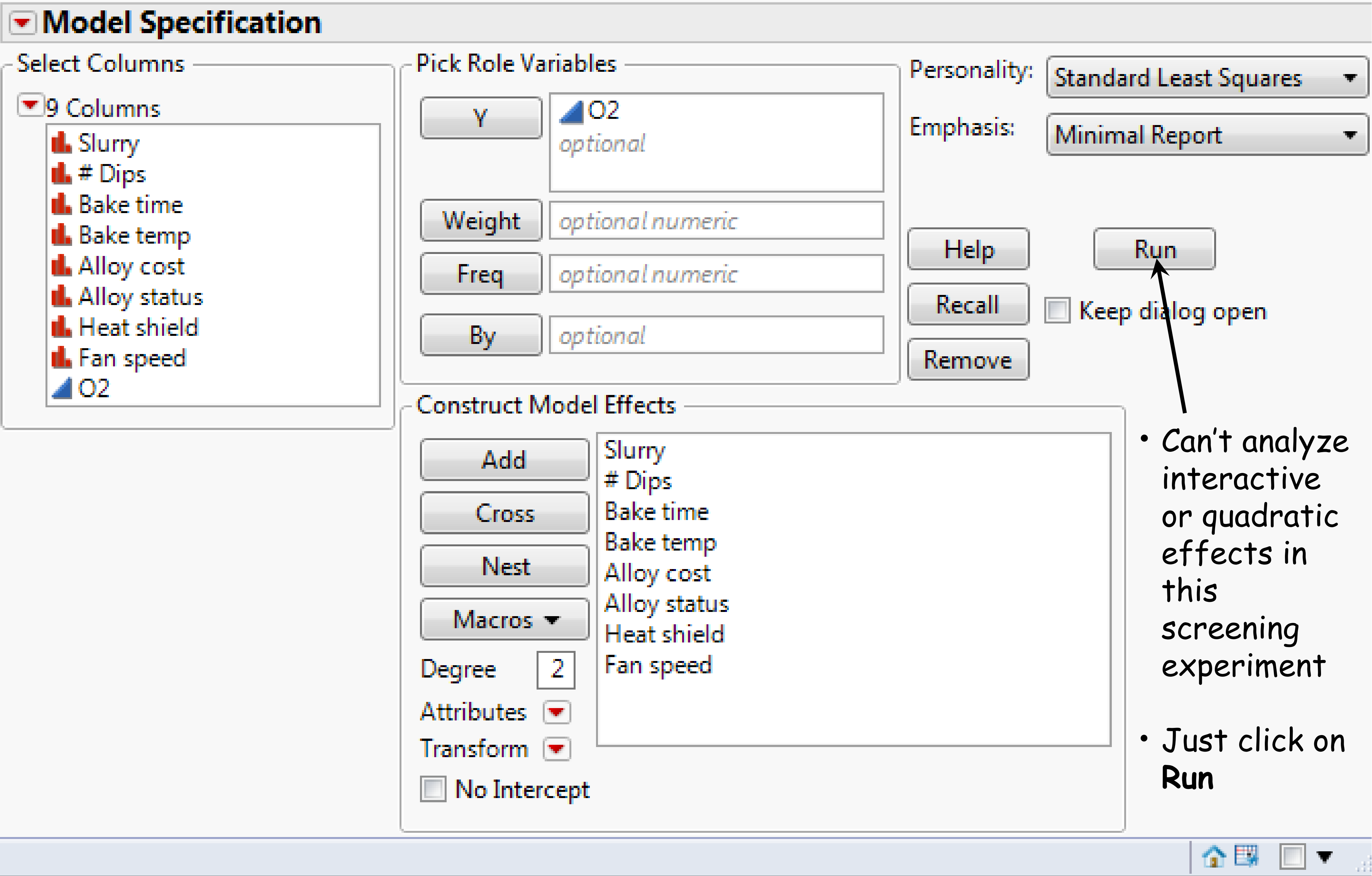
. . . two months (and many sleepless nights) later…

DOE participant files \ Ti casting alpha case with data



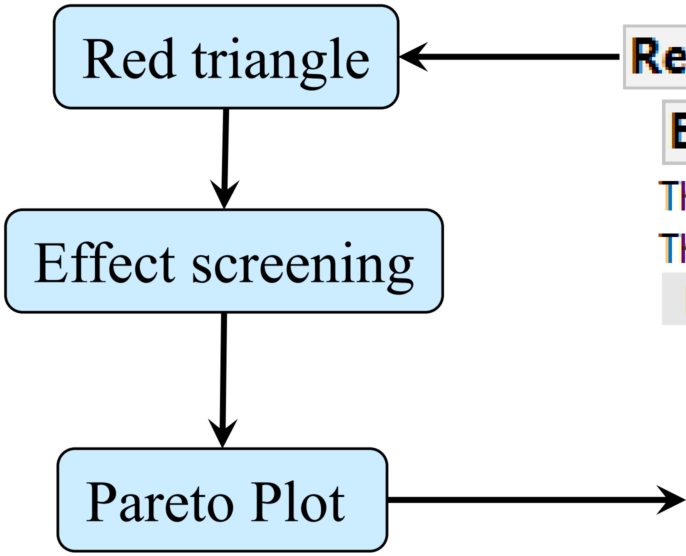
	Slurry	# Dips	Bake time	Bake temp	Alloy cost	Alloy status	Heat shield	Fan speed	O2
1	Batch 1	14	48	2050	High	Revert	Mild	3200	191
2	Batch 1	14	48	2050	Low	New	SS	2400	91
3	Batch 1	14	6	1950	High	New	SS	3200	76
4	Batch 1	14	6	1950	Low	Revert	Mild	2400	90
5	Batch 1	18	48	1950	High	Revert	SS	2400	184
6	Batch 1	18	48	1950	Low	New	Mild	3200	132
7	Batch 1	18	6	2050	High	New	Mild	2400	144
8	Batch 1	18	6	2050	Low	Revert	SS	3200	197
9	Batch 2	14	48	1950	High	New	Mild	2400	174
10	Batch 2	14	48	1950	Low	Revert	SS	3200	128
11	Batch 2	14	6	2050	High	Revert	SS	2400	166
12	Batch 2	14	6	2050	Low	New	Mild	3200	255
13	Batch 2	18	48	2050	High	New	SS	3200	318
14	Batch 2	18	48	2050	Low	Revert	Mild	2400	186
15	Batch 2	18	6	1950	High	Revert	Mild	3200	111
16	Batch 2	18	6	1950	Low	New	SS	2400	213

The model dialog



- Can't analyze interactive or quadratic effects in this screening experiment
- Just click on **Run**

Analysis



Response O2

Effect Screening

The parameter estimates have equal variances.

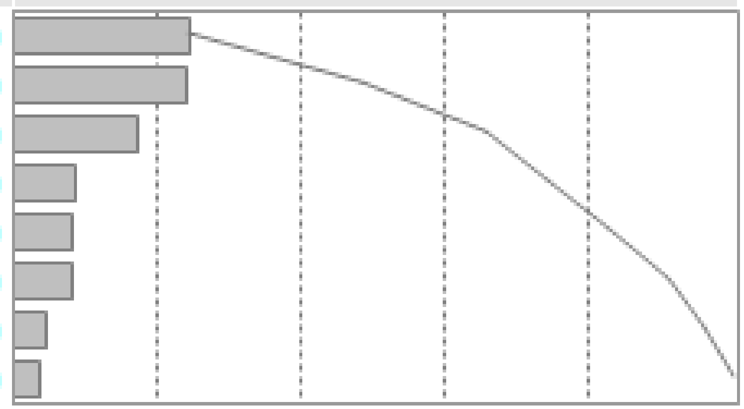
The parameter estimates are not correlated.

Lenth PSE

14.625

Pareto Plot of Estimates

Term	Estimate
Slurry[Batch 1]	-27.87500
Bake temp[1950]	-27.50000
# Dips[14]	-19.62500
Fan speed[2400]	-10.00000
Bake time[48]	9.50000
Alloy status[New]	9.37500
Heat shield[Mild]	-5.62500
Alloy cost[Low]	-4.50000



"Big hitters"

Slurry

Bake temp

Dips

- *Slurry* is a variable in the current state
- The O_2 values for castings made from Batch 1 shells were much lower than those from Batch 2
- The operators did not report any differences in the make-up of the two batches

To interpret screening experiments, use the *Effects Screening* analysis element as shown above. It shows showing the relative magnitude of the factor effects. The idea is to use the factors with the largest effects in a subsequent optimization experiment.

The interactive and quadratic effects are left out of the model. This biases the signal-to-noise ratios downward. The P-values are not to be trusted, so factors appear less significant than they really are.

Ideal follow-up plan

- Do a screening experiment in the shell-making area
- Include *Bake temp*, # *Dips* and the important shell-making variables in an optimization experiment

What actually happened

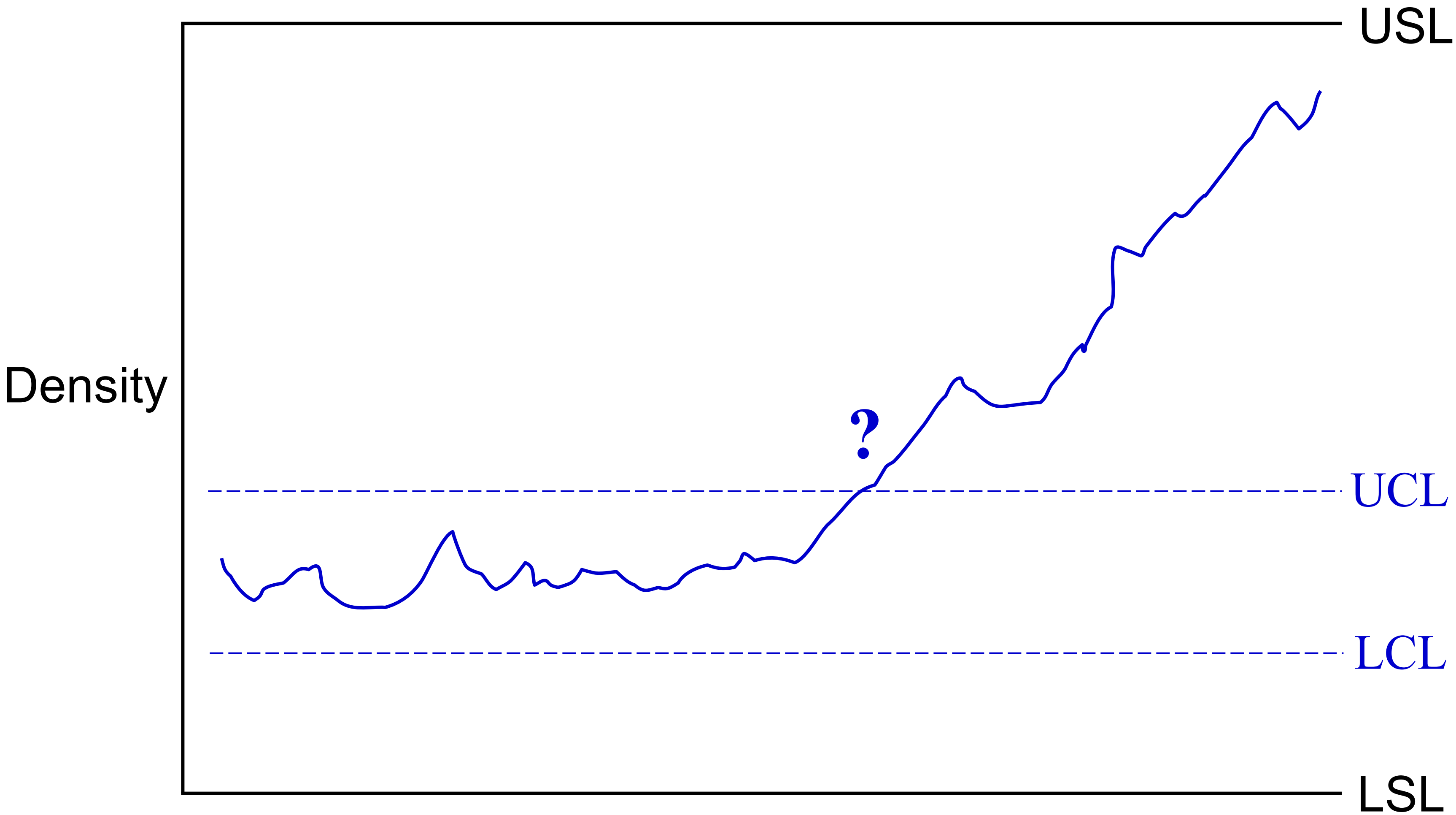
- They changed *Bake temp* to 1950 and # *Dips* to 14 (easy)
- The problem immediately went away
- 13 of the 16 DOE castings were good to ship as is
- Only 1 eventually scrapped
- Worst-case annual cost avoidance: $20.8M
- No immediate follow-up

Root causes

- Investigation of the slurry effect eventually lead to the root cause of the problem
  - → The density of the ceramic powder used to make the shell had increased over time, resulting in heavier shells
  - → The increase had been noted, but no action was taken because the densities were still within spec limits
  - → At the time, shell weights were not monitored

- Why no significant correlations in the “file cabinet” data?
  - → The O_2 data in the engineering database was post rework rather than first pass

Control limits vs. spec limits



- The data was trying to tell us something
- Disaster could have been averted

Exercise 10.1

a) Create a standard screening design matrix for the 10 factors shown below. Note: A sample size of 16 would have been adequate, but the project team decided to use a sample size of 24.

b) Save the table of factors for use in the next exercise:

Click the red triangle next to Screening Design > *Save Factors* (table opens)

File > Save as… > extrusion design factors

c) Save your design matrix as *extrusion design 1.jmp*.

d) *DOE Participant Files \ extrusion 0.jmp*. Analyze the data as shown for standard screening designs.

e) Based on the results for *Strength* and *Ductility*, find the best set of 4 factors for a subsequent optimization experiment.

Exercise 10.1 (cont'd)

Factors	*Feasible ranges*
Polymer variables	
Smoother	0.0 to 0.5
Filler	2.0 to 4.0
Viscosity	60 to 80
Moisture	0.1 to 0.25
Process variables	
Zone 1 temp	260 to 320
Zone 2 temp	260 to 320
Zone 3 temp	260 to 320
Zone 4 temp	260 to 320
Rate	100 to 200
RPM	150 to 300

Responses are *Strength* and *Ductility* of the extrusions

Another way to analyze

The experiment in the previous example was conducted years ago.

JMP can now analyze this experiment differently, giving more information!

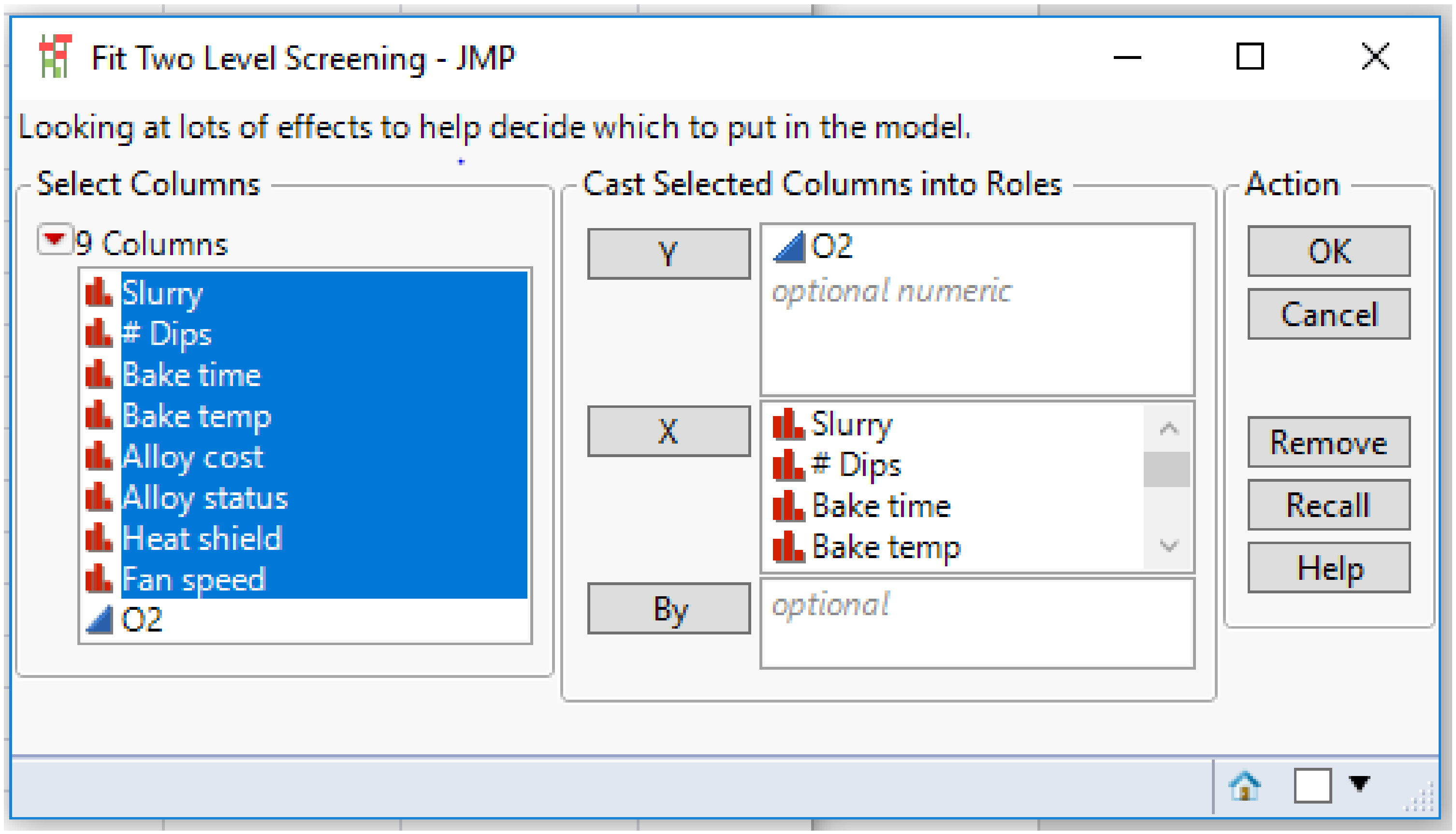
The O2 experiment can be analyzed using JMP's Fit Two Level Screening

- Requirement for this type of analysis: All factors are at 2 levels
- Reports and interpretation are very different
- Based on the assertion that relatively few of the effects are active
- Most are inactive (insignificant), meaning their effects are negligible
- Often, in screening experiments, there are no degrees of freedom for error
- Estimates of inactive effects are used to estimate random error in this analysis
- Some information can be gained about 2-factor interactions
- 2-Factor interactions are aliased with each other

Fit Two Level Screening

DOE participant files \ Ti casting alpha case with data

- DOE > Classical > Two Level Screening > Fit Two Level Screening
- Set up as shown (all factors are cast into X)
- Click OK



Fit Two Level Screening (cont'd)

Below is the Contrasts report:

- Contrast column shows the regression parameter estimate
  - An asterisk shows estimate is not the same as the regression estimate
  - An asterisk would indicate that we need to use the Fit Model platform
  - There are no asterisks in this report
- Individual p-Values indicate significant effects
- Bar Chart shows terms significant at the 0.10 level
- Analysis may not be exactly the same if re-run, due to the analysis process
- Note that there is an interaction that is significant!
  - We cannot tell if the significant interaction is Bake temp*Fan speed
  - It could be any of the interactions under Aliases
  - The estimate of the effect (Contrast) is actually the sum of all of the aliased interactions
  - This is because this is a screening design
  - Additional experimentation is needed determine the active interaction

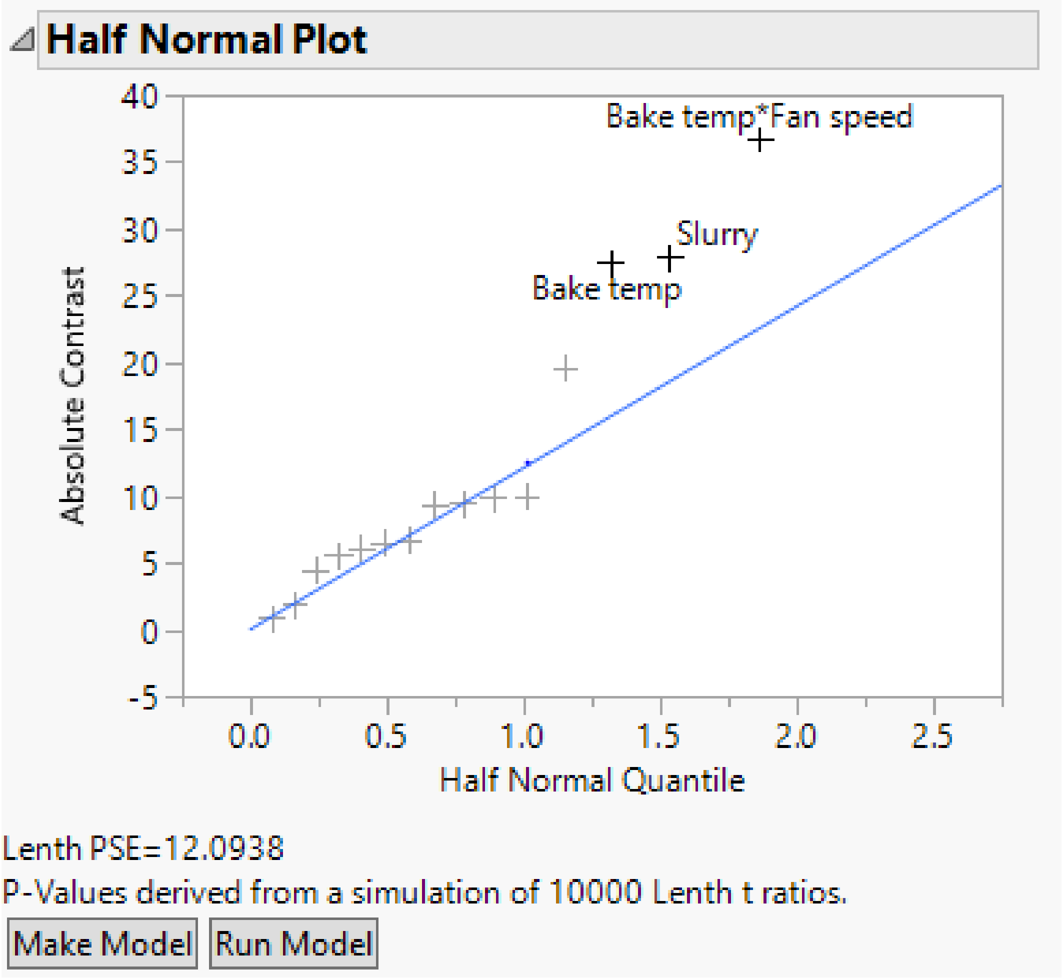
Contrasts report

Contrasts

Term	Contrast	Lenth t-Ratio	Individual p-Value	Simultaneous p-Value	Aliases
Slurry	27.8750	2.30	0.0415*	0.3224	
Bake temp	27.5000	2.27	0.0434*	0.3370	
# Dips	19.6250	1.62	0.1126	0.7222	
Fan speed	10.0000	0.83	0.3818	1.0000	
Bake time	9.5000	0.79	0.4068	1.0000	
Alloy status	-9.3750	-0.78	0.4133	1.0000	
Heat shield	5.6250	0.47	0.6697	1.0000	
Alloy cost	4.5000	0.37	0.7320	1.0000	
Slurry*Bake temp	9.8750	0.82	0.3882	1.0000	# Dips*Bake time, Fan speed*Alloy status, Heat shield*Alloy cost
Slurry*# Dips	-6.5000	-0.54	0.6237	1.0000	Bake temp*Bake time, Alloy status*Heat shield, Fan speed*Alloy cost
Bake temp*# Dips	-1.8750	-0.16	0.8871	1.0000	Slurry*Bake time, Fan speed*Heat shield, Alloy status*Alloy cost
Slurry*Fan speed	-0.8750	-0.07	0.9474	1.0000	Bake temp*Alloy status, Bake time*Heat shield, # Dips*Alloy cost
Bake temp*Fan speed	36.7500	3.04	0.0190*	0.1617	Slurry*Alloy status, # Dips*Heat shield, Bake time*Alloy cost
# Dips*Fan speed	-6.1250	-0.51	0.6434	1.0000	Bake time*Alloy status, Bake temp*Heat shield, Slurry*Alloy cost
Fan speed*Bake time	6.7500	0.56	0.6115	1.0000	# Dips*Alloy status, Slurry*Heat shield, Bake temp*Alloy cost

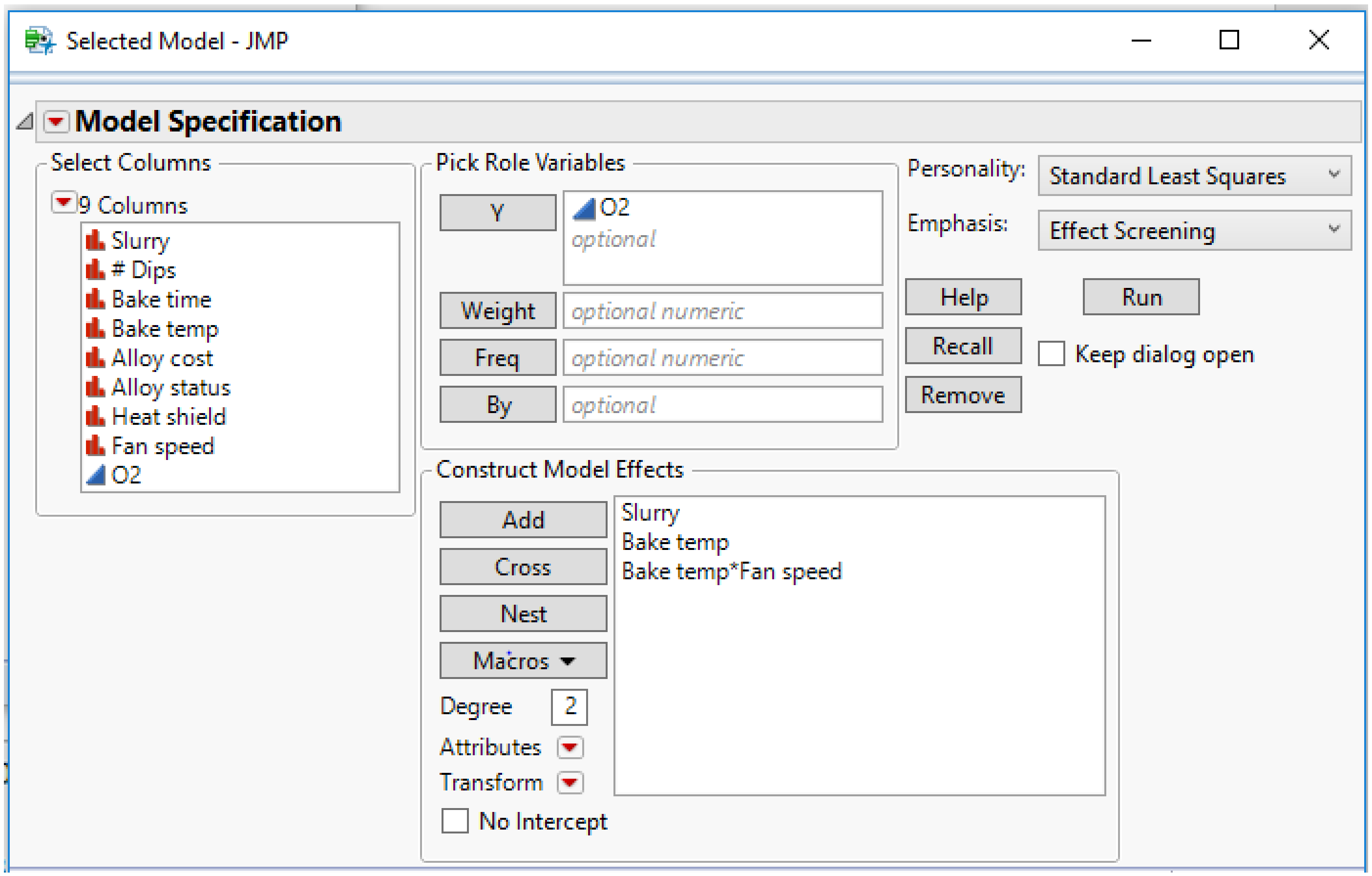
The Half Normal Plot graphically identifies significant effects

- Significant effects or terms fall off (away from) the blue line
  - The additional point off the line is # Dips, which was near the cut-off
  - Here, it appears to be significant
  - One could choose to carry this term forward



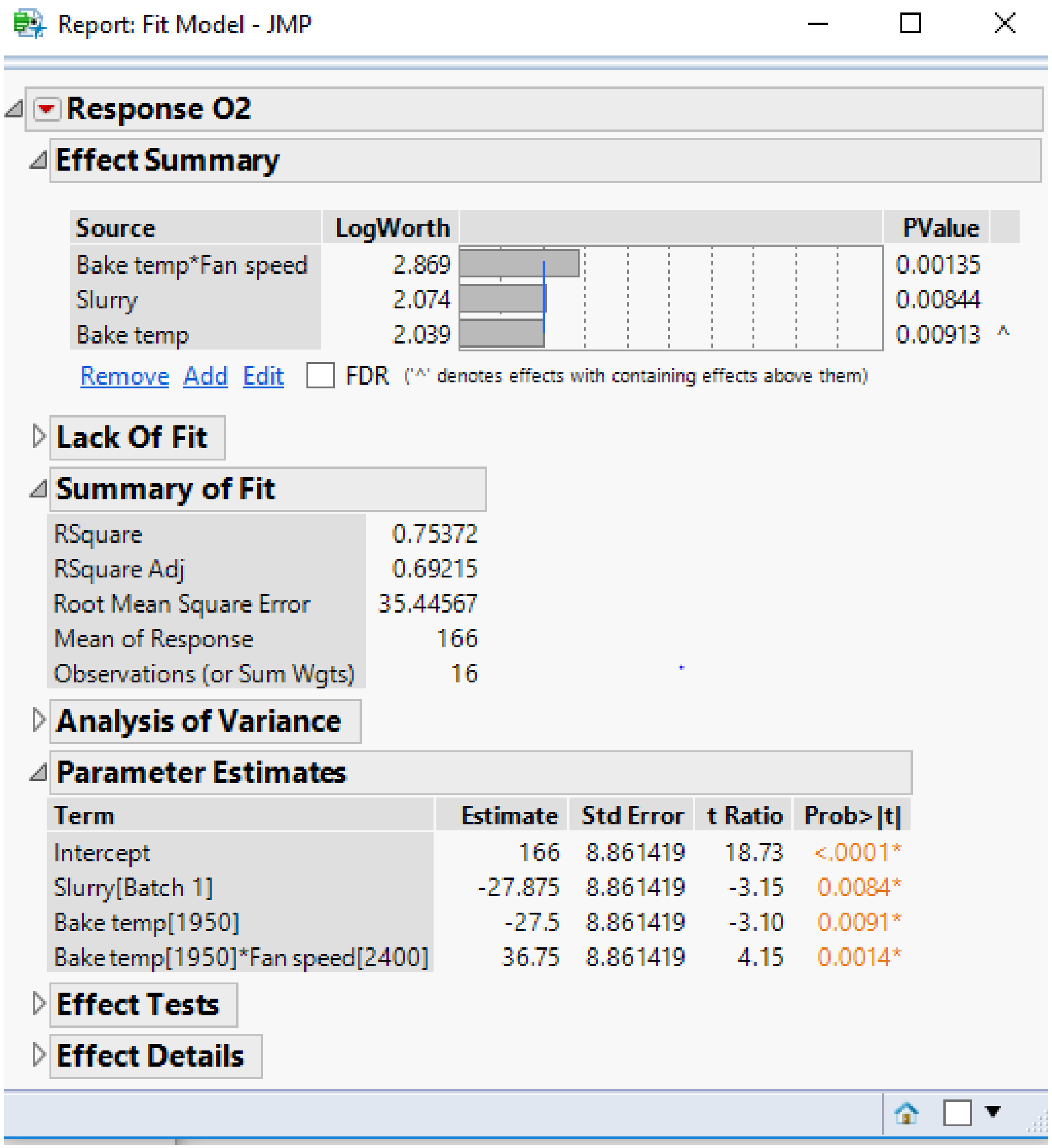
Fit Two Level Screening (cont'd)

- Click Make Model
- Fit Model window will come up
  - Significant terms have been carried forward
  - Terms can be added to the model
  - # Dips could be added (probably should be, based on Half Normal Plot)



Fit Two Level Screening (cont'd)

- Click Run
- This familiar report comes up
- This analysis got us further
  - Presence of interaction
  - Need higher level terms
- Additional experimentation to:
  - Determine interaction
  - Optimize



Report: Fit Model - JMP

Response O2

Effect Summary

Source	LogWorth		PValue
Bake temp*Fan speed	2.869		0.00135
Slurry	2.074		0.00844
Bake temp	2.039		0.00913 ^

Remove Add Edit FDR ('^' denotes effects with containing effects above them)

Lack Of Fit

Summary of Fit

RSquare	0.75372
RSquare Adj	0.69215
Root Mean Square Error	35.44567
Mean of Response	166
Observations (or Sum Wgts)	16

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	166	8.861419	18.73	<.0001*
Slurry[Batch 1]	-27.875	8.861419	-3.15	0.0084*
Bake temp[1950]	-27.5	8.861419	-3.10	0.0091*
Bake temp[1950]*Fan speed[2400]	36.75	8.861419	4.15	0.0014*

Effect Tests

Effect Details

Definitive Screening Design

A Definitive Screening Design is a very effective screening design

- Factors must be either continuous or two-level categorical
- It can be a good alternative to a Custom Design when six or more factors

"A minimum run-size DSD is capable of correctly identifying active terms with high probability if the number of active effects is less than about half the number of runs and if the effects sizes exceed twice the standard deviation. However, by augmenting a minimum run-size DSD with four or more properly selected runs, you can identify substantially more effects with high probability. . . . Extra Runs substantially increase the design's ability to detect second-order effects."

--From JMP's Overview of the Fit Definitive Screening Platform

"Effect sizes exceed twice the standard deviation" → $\frac{b_n}{\sigma} \geq 1$,

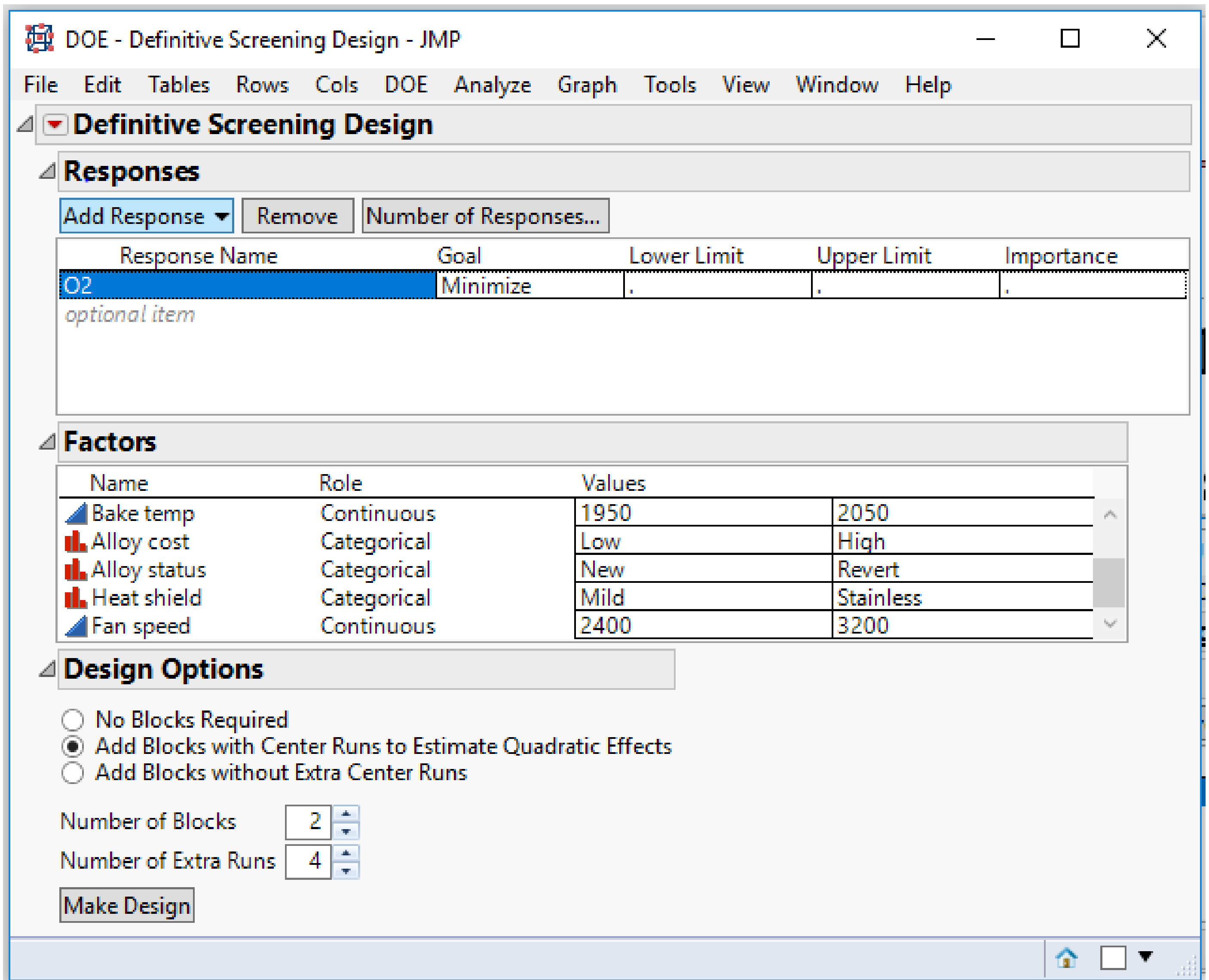
which means that the difference between the average response at the high level and at the low level is 2σ, or 2 * std dev. (Remember, the coefficient is the effect/2.)

"Second order effects" include 2-level interactions and quadratic terms.

Example

Using the same situation as in the previous example:

- Enter response and factors, as usual
- Set up Design Options, as shown. (4 Extra Runs are recommended!!!)



Example (cont'd)

This Definitive Screening Design requires 22 runs

- In the previous example, only 16 runs were required
- However, a follow-on optimization experiment was needed

The Definitive Screening can be run, then augmented, if needed

- This requires many fewer runs (and other resources) overall

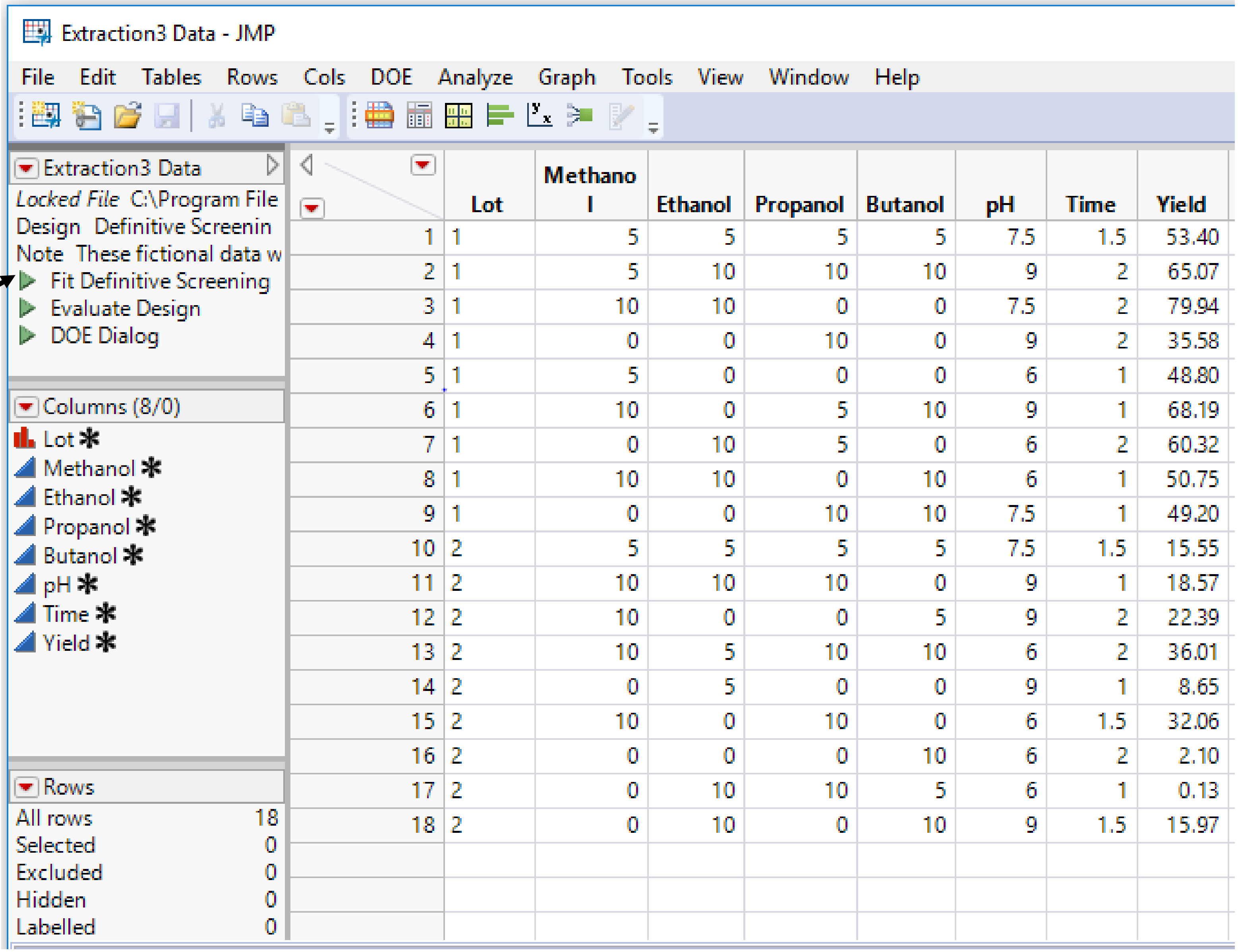
	Block	Slurry	# Dips	Bake time	Bake temp	Alloy cost	Alloy status	Heat shield	Fan speed	O2
1	1	Batch 2	18	27	2000	High	Revert	Stainless	2800	•
2	1	Batch 2	14	6	2050	Low	Revert	Stainless	2800	•
3	1	Batch 1	18	6	1950	Low	New	Stainless	3200	•
4	1	Batch 2	18	48	1950	Low	New	Stainless	3200	•
5	1	Batch 1	14	6	2050	High	Revert	Mild	2400	•
6	1	Batch 1	14	48	2050	High	New	Stainless	3200	•
7	1	Batch 1	18	48	1950	High	New	Mild	2800	•
8	1	Batch 2	18	27	2050	High	Revert	Stainless	3200	•
9	1	Batch 1	14	27	2000	Low	New	Mild	2800	•
10	1	Batch 2	14	48	2050	High	Revert	Mild	2400	•
11	1	Batch 2	18	6	1950	Low	Revert	Mild	2400	•
12	1	Batch 1	14	27	1950	Low	New	Mild	2400	•
13	2	Batch 2	18	48	2050	Low	New	Stainless	2400	•
14	2	Batch 1	18	6	2050	Low	New	Stainless	2400	•
15	2	Batch 2	18	48	1950	High	Revert	Stainless	2400	•
16	2	Batch 1	18	48	2050	Low	Revert	Mild	3200	•
17	2	Batch 1	14	6	1950	High	Revert	Mild	3200	•
18	2	Batch 1	14	6	2050	Low	New	Mild	3200	•
19	2	Batch 2	14	48	1950	High	Revert	Mild	3200	•
20	2	Batch 2	14	6	1950	High	New	Stainless	2400	•
21	2	Batch 1	14	48	2000	Low	Revert	Stainless	2400	•
22	2	Batch 2	18	6	2000	High	New	Mild	3200	•

Analyzing the Definitive Screening Design

When you create a Definitive Screening Design in JMP, the Table will contain a script for analysis

Help > Sample Data Library   Design Experiment / Extraction 3 Data

- Run the experiment
- Enter data into the table
- Click on the green triangle to analyze the data (run the script)
- **You must use Fit Definitive Screening for the analysis, to take advantage of the design structure**



	Lot	Methanol	Ethanol	Propanol	Butanol	pH	Time	Yield
1	1	5	5	5	5	7.5	1.5	53.40
2	1	5	10	10	10	9	2	65.07
3	1	10	10	0	0	7.5	2	79.94
4	1	0	0	10	0	9	2	35.58
5	1	5	0	0	0	6	1	48.80
6	1	10	0	5	10	9	1	68.19
7	1	0	10	5	0	6	2	60.32
8	1	10	10	0	10	6	1	50.75
9	1	0	0	10	10	7.5	1	49.20
10	2	5	5	5	5	7.5	1.5	15.55
11	2	10	10	10	0	9	1	18.57
12	2	10	0	0	5	9	2	22.39
13	2	10	5	10	10	6	2	36.01
14	2	0	5	0	0	9	1	8.65
15	2	10	0	10	0	6	1.5	32.06
16	2	0	0	0	10	6	2	2.10
17	2	0	10	10	5	6	1	0.13
18	2	0	10	0	10	9	1.5	15.97

Analyzing the Definitive Screening Design (cont'd)

- JMP does all the work:
  - Stage 1 tests Main Effects
  - Stage 2 tests interactions and quadratic terms of significant Main Effects
  - Combined Model includes both

Stage 1 - Main Effect Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Methanol	9.7133	0.3674	26.438	<.0001*
Ethanol	2.3166	0.3674	6.3055	0.0015*
Time	4.0798	0.3674	11.104	0.0001*

Statistic	Value
RMSE	1.3747
DF	5

Stage 2 - Even Order Effect Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	34.568	1.3459	25.683	0.0015*
Lot[1]	17.197	0.7757	22.171	0.0020*
Methanol*Ethanol	-0.367	0.7127	-0.515	0.6581
Methanol*Time	0.5266	0.7127	0.7389	0.5369
Ethanol*Time	9.8258	0.8534	11.514	0.0075*
Methanol*Methanol	7.637	1.4914	5.1208	0.0361*
Ethanol*Ethanol	-1.449	1.477	-0.981	0.4299
Time*Time	-3.297	1.477	-2.232	0.1552

Statistic	Value
RMSE	2.0626
DF	2

Combined Model Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	34.568	1.0452	33.074	<.0001*
Lot[1]	17.197	0.6023	28.552	<.0001*
Methanol	9.7133	0.4281	22.691	<.0001*
Ethanol	2.3166	0.4281	5.4118	0.0010*
Time	4.0798	0.4281	9.5307	<.0001*
Methanol*Ethanol	-0.367	0.5534	-0.663	0.5287
Methanol*Time	0.5266	0.5534	0.9516	0.3730
Ethanol*Time	9.8258	0.6627	14.828	<.0001*
Methanol*Methanol	7.637	1.1581	6.5945	0.0003*
Ethanol*Ethanol	-1.449	1.1469	-1.264	0.2468
Time*Time	-3.297	1.1469	-2.875	0.0238*

Statistic	Value
RMSE	1.6017
DF	7

Make Model | Run Model

Click Run Model

Analyzing the Definitive Screening Design (cont'd)

A familiar report comes up

- Proceed as before: Check residuals and remove insignificant terms
- **Note that interactions and quadratic terms are estimated!**
- This is what is meant by Definitive Screening
- In this case, an additional optimization experiment is not necessary!

Effect Summary

Source	LogWorth	PValue	
Lot	7.779	0.00000	
Methanol(0,10)	7.087	0.00000	
Ethanol*Time	5.818	0.00000	
Time(1,2)	4.532	0.00003	^
Methanol*Methanol	3.514	0.00031	
Ethanol(0,10)	3.002	0.00100	^
Time*Time	1.623	0.02382	
Ethanol*Ethanol	0.608	0.24682	
Methanol*Time	0.428	0.37300	
Methanol*Ethanol	0.277	0.52873	

Remove Add Edit ☐ FDR ('^' denotes effects with containing effects above them)

Full Factorial vs. Definitive Screening Design (not randomized)

Full Factorial Design with 4 Center Runs:

X1	X2	X3	X4	Y
-1	-1	-1	-1	•
-1	-1	-1	1	•
-1	-1	1	-1	•
-1	-1	1	1	•
-1	1	-1	-1	•
-1	1	-1	1	•
-1	1	1	-1	•
-1	1	1	1	•
1	-1	-1	-1	•
1	-1	-1	1	•
1	-1	1	-1	•
1	-1	1	1	•
1	1	-1	-1	•
1	1	-1	1	•
1	1	1	-1	•
1	1	1	1	•
0	0	0	0	•
0	0	0	0	•
0	0	0	0	•
0	0	0	0	•

Definitive Screening Design with 4 Extra Runs and 2 Center Runs:

X1	X2	X3	X4	X5	X6	Y
0	1	1	1	1	1	•
0	-1	-1	-1	-1	-1	•
1	0	1	1	-1	1	•
-1	0	-1	-1	1	-1	•
1	-1	0	1	1	-1	•
-1	1	0	-1	-1	1	•
1	-1	-1	0	1	1	•
-1	1	1	0	-1	-1	•
1	1	-1	-1	0	1	•
-1	-1	1	1	0	-1	•
1	-1	1	-1	-1	0	•
-1	1	-1	1	1	0	•
1	1	-1	1	-1	-1	•
-1	-1	1	-1	1	1	•
1	1	1	-1	1	-1	•
-1	-1	-1	1	-1	1	•
0	0	0	0	0	0	•
0	0	0	0	0	0	•

Note the structural differences in these two classes of designs.

Exercise 10.2

Using the same factors and levels as Exercise 10.1, create a Definitive Screening Design.

- When you are ready to enter the factors:
  - Click the red triangle next to Definitive Screening Design > *Load Factors* (select the file *extrusion design factors* saved during Exercise 10.1)
- **Be sure to add the recommended 4 runs!**
- The previous experiment required 16 runs, but they used 24 runs. Further experimentation would be needed with that screening design.
- How many runs does this Definitive Screening Design require?

Factors from Exercise 10.1

Factors	*Feasible ranges*
Polymer variables	
Smoother	0.0 to 0.5
Filler	2.0 to 4.0
Viscosity	60 to 80
Moisture	0.1 to 0.25
Process variables	
Zone 1 temp	260 to 320
Zone 2 temp	260 to 320
Zone 3 temp	260 to 320
Zone 4 temp	260 to 320
Rate	100 to 200
RPM	150 to 300

Responses are *Strength* and *Ductility* of the extrusions

This slide intentionally left blank

12. Experiments with Hard-to-Change Factors

Sometimes it's not feasible to completely randomize, because a factor is hard-to-change

There are many situations when this is the case. Here are a few examples:

- Temperature in a furnace takes a very long time (hours) to stabilize after changing
- Special material needed (a factor) are made in large batches and cannot be stored, or it is run in a continuous flow through the process
- Material or part used in a machine is difficult to change, requiring a complete breakdown and cleaning
- Type of irrigation on a plot of land is very difficult and costly to change (an example of the origin of split-plot designs)

What are examples in your workplace?

Experiments with Hard-to-change factors (cont'd)

When you have hard-to-change factors that cannot be randomized, you need to create and analyze your experiment as a "split-plot" design

If you don't do this (if you design and analyze as usual), you are more likely to:

- Conclude that unimportant factors are important among the hard-to-change factors
  - You think that a factor (X) is impacting your response (Y), when it is not
  - This is a Type I error
  - Hard-to-change factors are those in the "Whole Plots" or main treatments, that <u>were not</u> randomized
- Fail to recognize factors that are significant among the easy-to-change factors
  - You think that a factor (X) is NOT impacting your response (Y), when it is
  - This is a Type II error
  - Easy-to-change factors are those in the "Subplots" or split-plots, that <u>were</u> randomized

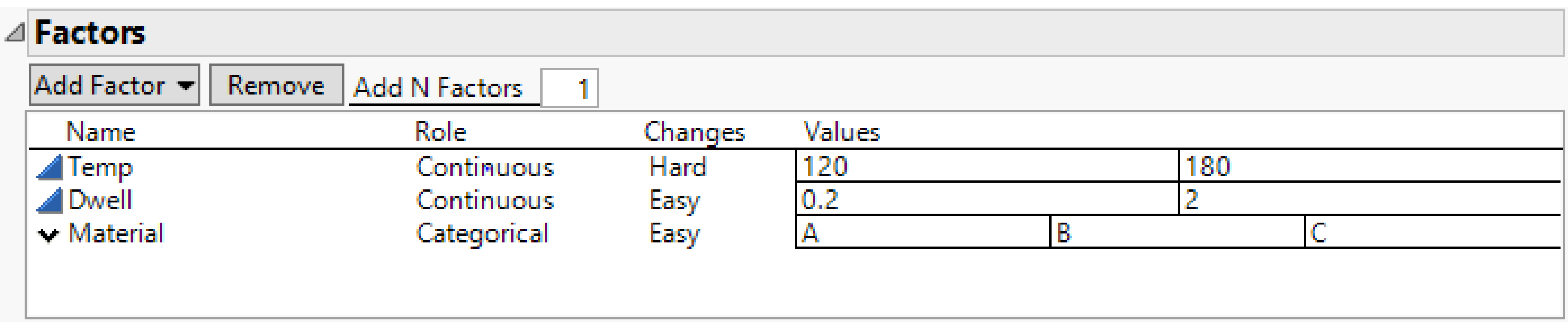
Experiments with Hard-to-change factors (cont'd)

The decision to consider a factor as "hard-to-change" should not be taken lightly

- Subplot (easy-to-change) factors are compared with higher precision
    - Usually, subplot error is smaller than whole-plot error
    - Whenever possible, the treatment(s) or factors we are most interested in should be assigned to the subplots
- To increase the precision of the test on whole-plot (hard-to-change) factors, additional replicates of the experiment or additional whole-plots are needed
    - Clearly, this takes more time and resources
    - Several (3-6) replicates could be needed to gain an adequate level of precision
    - So, you could be back to changing that hard-to-change factor many times

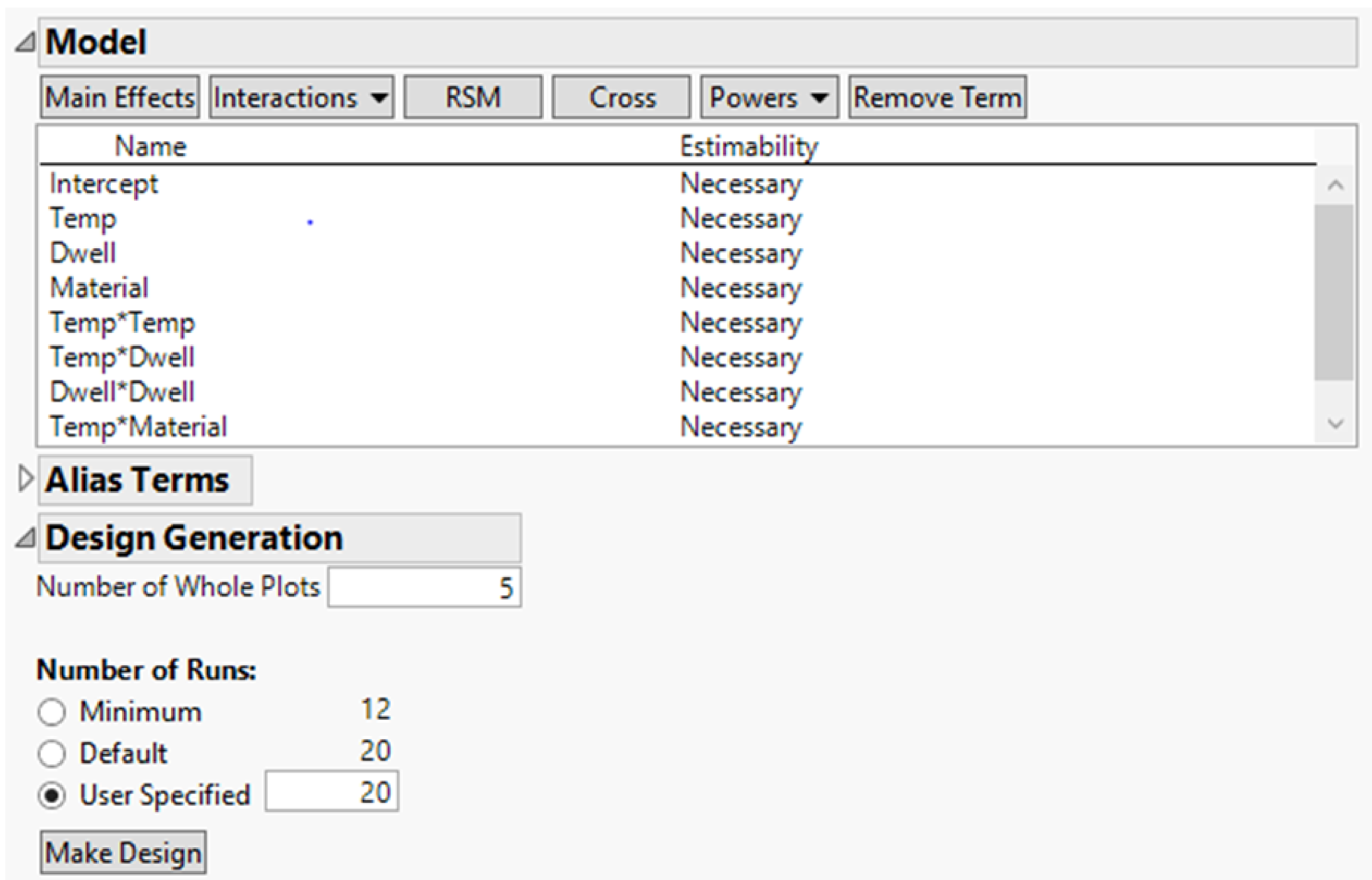
Creating a Split-Plot Design

- DOE > Custom Design
- Enter the factors as usual, except double-click on “Changes” and change to Hard for the hard to change factor
- Click Continue



Creating a Split-Plot Design (cont'd)

- Click on RSM.
- JMP will suggest a reasonable number of Whole Plots for the number of factors and levels entered
- The number of Whole Plots shows the number of times the hard-to-change factor will need to be changed in the experiment
- Click Make Design



Creating a Split-Plot Design (cont'd)

- The design is presented.
- As before, click Back to make adjustments. Click Make Table.
- Run the experiment in the order shown in the table.

Design

Run	Whole Plots	Temp	Dwell	Material
1	1	150	1.1	A
2	1	150	0.2	C
3	1	150	0.2	B
4	1	150	2	B
5	2	180	0.2	A
6	2	180	1.1	C
7	2	180	1.1	B
8	2	180	2	A
9	3	120	0.2	A
10	3	120	1.1	C
11	3	120	2	A
12	3	120	1.1	B
13	4	150	1.1	B
14	4	150	0.2	C
15	4	150	2	C
16	4	150	1.1	A
17	5	150	0.2	B
18	5	150	1.1	A
19	5	150	2	B
20	5	150	2	C

Table:

Whole Plots	Temp	Dwell	Material	Y1
1	150	1.1	A	•
1	150	0.2	C	•
1	150	0.2	B	•
1	150	2	B	•
2	180	0.2	A	•
2	180	1.1	C	•
2	180	1.1	B	•
2	180	2	A	•
3	120	0.2	A	•
3	120	1.1	C	•
3	120	2	A	•
3	120	1.1	B	•
4	150	1.1	B	•
4	150	0.2	C	•
4	150	2	C	•
4	150	1.1	A	•
5	150	0.2	B	•
5	150	1.1	A	•
5	150	2	B	•
5	150	2	C	•

Blocking in a Split-Plot Design

What if there are too many runs to complete in one day (or lot of material, or by one tester, etc.)?

- Once you see that there are too many runs, click Back (before making the table)
- Add a Categorical Factor with the number of levels as the number of batches or days or shifts, etc. needed for the experiment (In this example, two days will be needed to run the experiment, so a 2-Level Categorical Factor was added.)
- Name the factor something that you can easily pick out of the lists of terms (Here it is named REMOVE.)
- Set Changes for this factor to Very Hard
- Click Continue

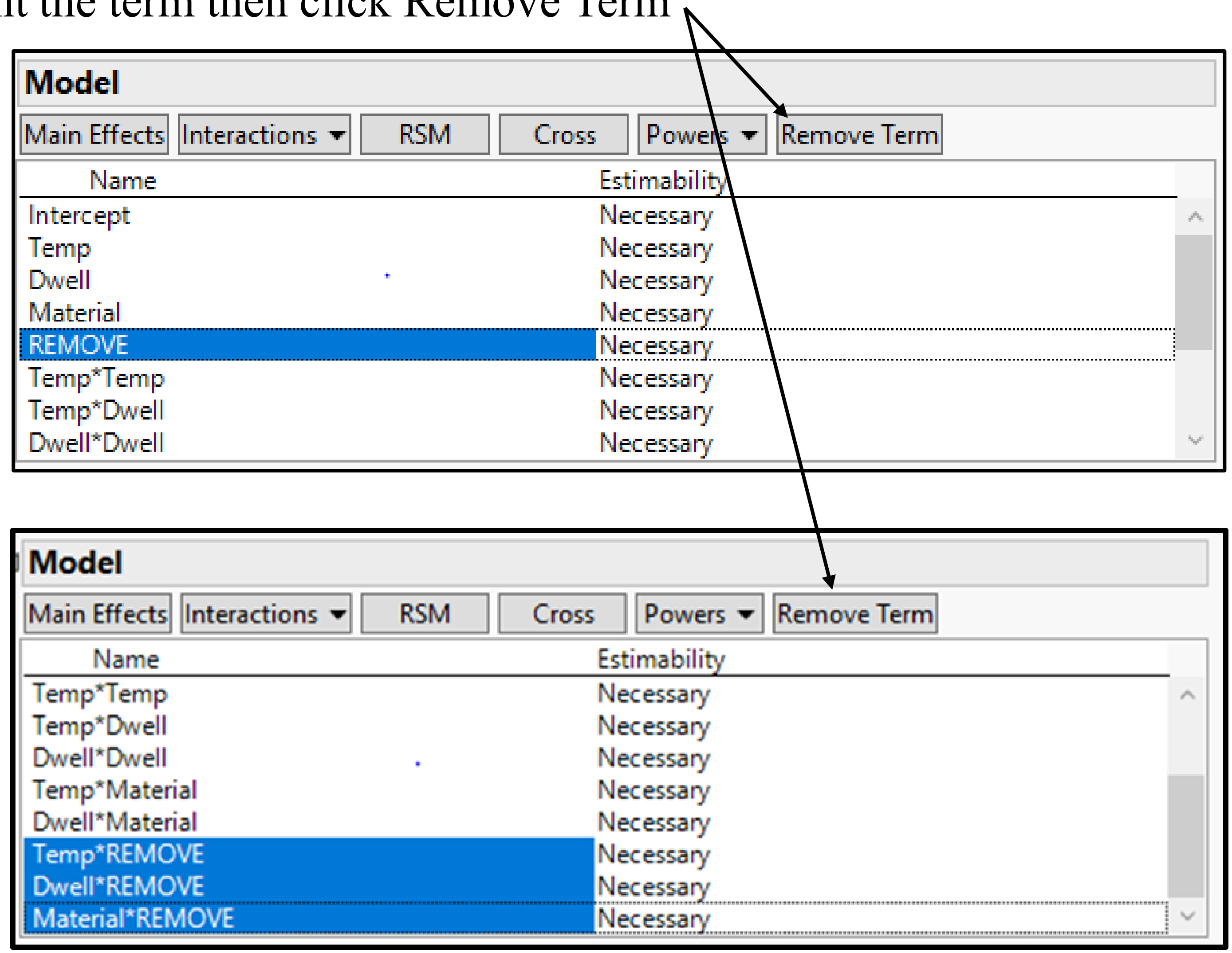
Factors

Add Factor | Remove | Add N Factors 1

Name	Role	Changes	Values		
Temp	Continuous	Hard	120	180	
Dwell	Continuous	Easy	0.2	2	
Material	Categorical	Easy	A	B	C
REMOVE	Categorical	Very Hard	L1	L2	

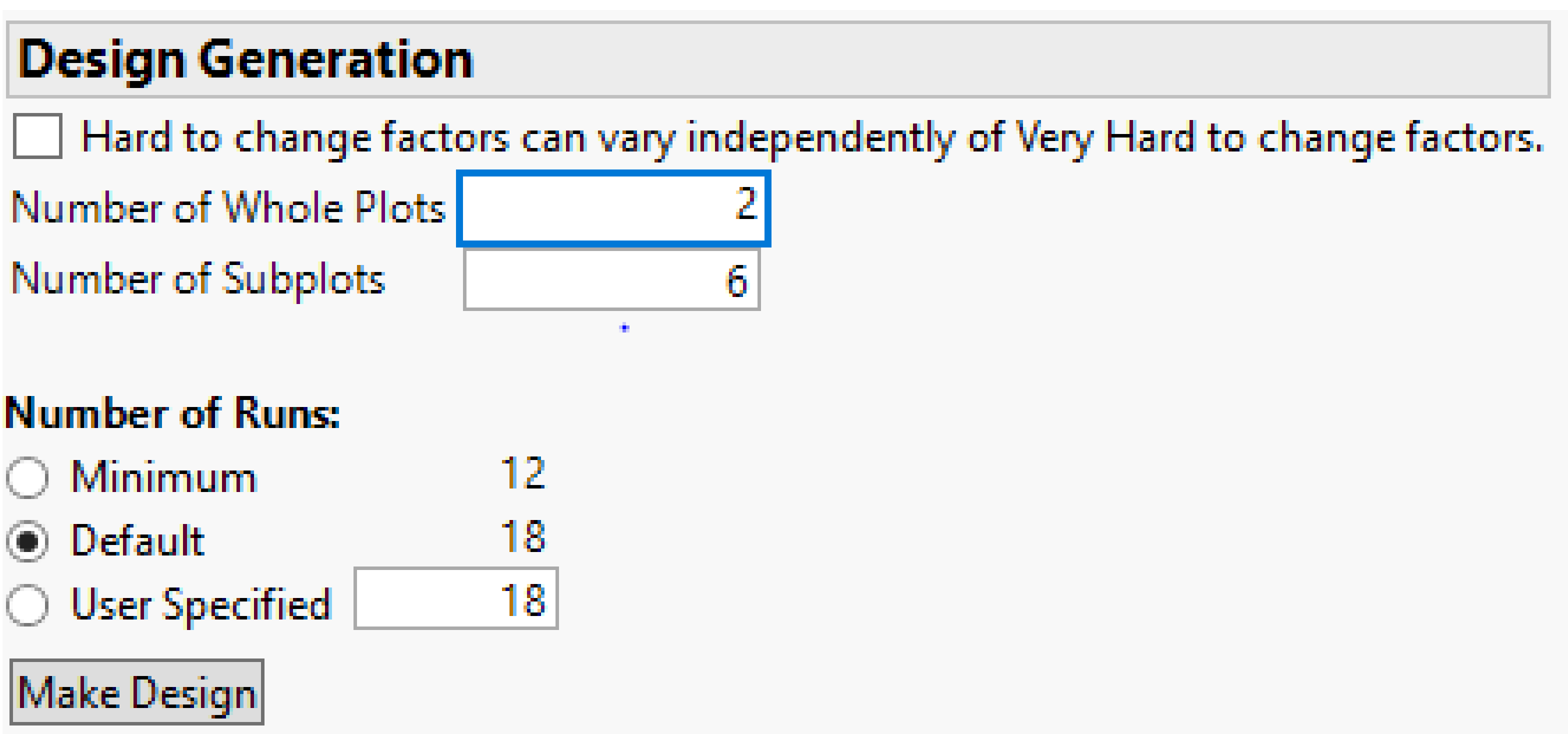
Blocking in a Split-Plot Design (cont'd)

- Click RSM
- Remove from the Model every term that contains the Categorical factor that you added
  - Highlight the term then click Remove Term



Blocking in a Split-Plot Design (cont'd)

- Change the number of Whole Plots to the number of levels of the Categorical Factor
  - In this example, two days were needed
  - So, a 2-Level Categorical Factor called REMOVE was added
  - Now, the Number of Whole Plots is changed to 2
- Click make Design



Blocking in a Split-Plot Design (cont'd)

- The Design is developed
- Whole Plots show the number of days required
- REMOVE is still in the table, as it was entered as a factor
- Click Make Table

Design

Run	Whole Plots	Subplots	Temp	Dwell	Material	REMOVE
1	1	1	120	0.2	C	L1
2	1	1	120	2	A	L1
3	1	1	120	1.1	B	L1
4	1	2	180	2	C	L1
5	1	2	180	0.2	A	L1
6	1	2	180	1.1	B	L1
7	1	3	150	1.1	C	L1
8	1	3	150	1.1	A	L1
9	1	3	150	2	B	L1
10	2	4	120	1.1	B	L1
11	2	4	120	0.2	A	L1
12	2	4	120	2	C	L1
13	2	5	150	0.2	B	L1
14	2	5	150	1.1	C	L1
15	2	5	150	1.1	A	L1
16	2	6	180	1.1	B	L1
17	2	6	180	2	A	L1
18	2	6	180	0.2	C	L1

If you get this warning, it's okay to ignore it, IN THIS CASE, because you are not trying to estimate effects of the whole plot

At least one more whole plot is strongly recommended. This design does not have enough whole plots to estimate the whole plot variance. The whole plot effects are not testable.

Blocking in a Split-Plot Design (cont'd)

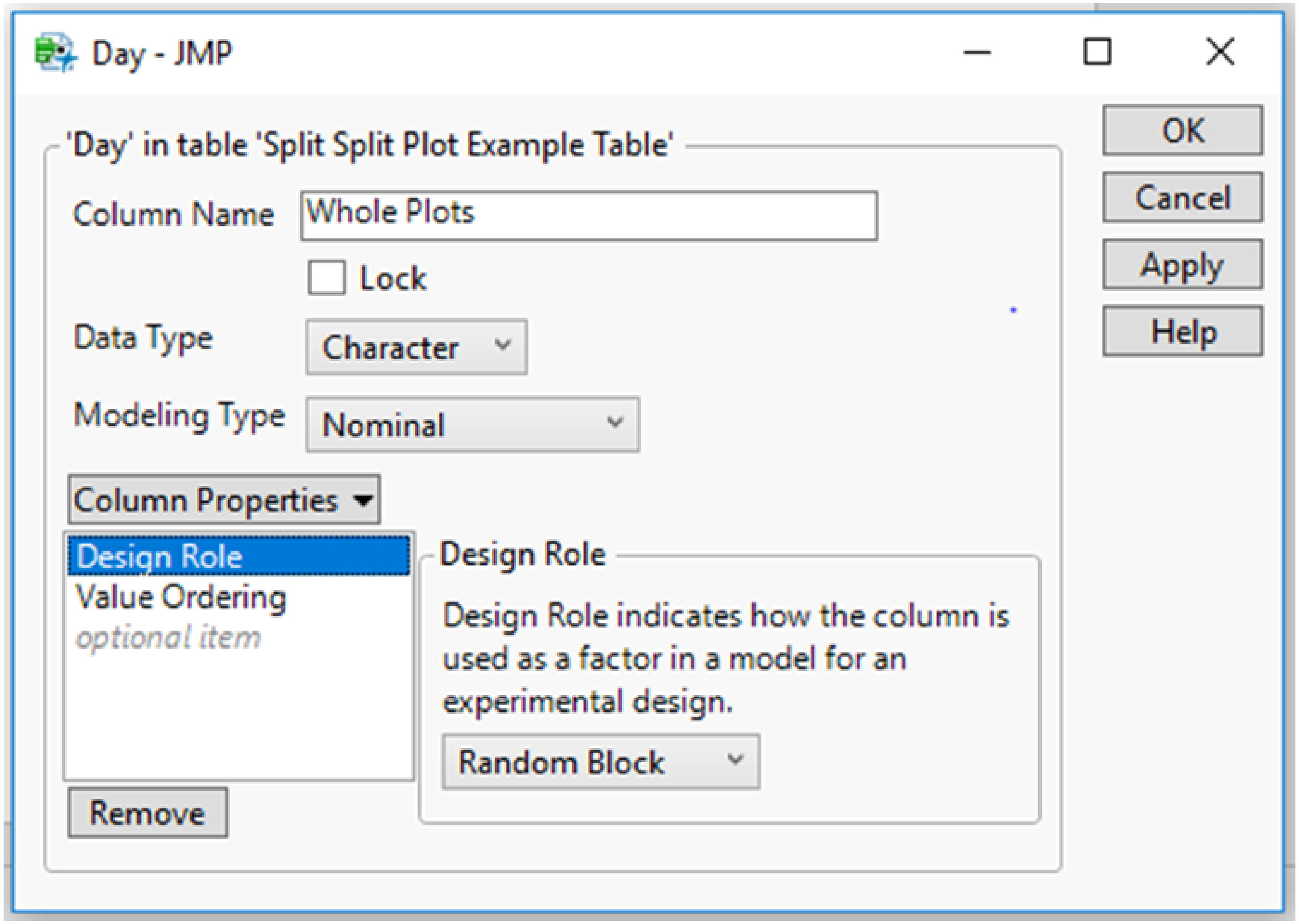
- The table is generated
- Click on the column of the Categorical Factor ("REMOVE" in this example).
- Cols > Delete Columns to delete the column from the table

Whole Plots	Subplots	Temp	Dwell	Material	REMOVE	Y1
1	1	120	0.2	C	L1	•
1	1	120	2	A	L1	•
1	1	120	1.1	B	L1	•
1	2	180	2	C	L1	•
1	2	180	0.2	A	L1	•
1	2	180	1.1	B	L1	•
1	3	150	1.1	C	L1	•
1	3	150	1.1	A	L1	•
1	3	150	2	B	L1	•
2	4	120	1.1	B	L1	•
2	4	120	0.2	A	L1	•
2	4	120	2	C	L1	•
2	5	150	0.2	B	L1	•
2	5	150	1.1	C	L1	•
2	5	150	1.1	A	L1	•
2	6	180	1.1	B	L1	•
2	6	180	2	A	L1	•
2	6	180	0.2	C	L1	•

Whole Plots	Subplots	Temp	Dwell	Material	REMOVE	Y1
1	1	120	0.2	C	L1	•
1	1	120	2	A	L1	•
1	1	120	1.1	B	L1	•
1	2	180	2	C	L1	•
1	2	180	0.2	A	L1	•
1	2	180	1.1	B	L1	•
1	3	150	1.1	C	L1	•
1	3	150	1.1	A	L1	•
1	3	150	2	B	L1	•
2	4	120	1.1	B	L1	•
2	4	120	0.2	A	L1	•
2	4	120	2	C	L1	•
2	5	150	0.2	B	L1	•
2	5	150	1.1	C	L1	•
2	5	150	1.1	A	L1	•
2	6	180	1.1	B	L1	•
2	6	180	2	A	L1	•
2	6	180	0.2	C	L1	•

Blocking in a Split-Plot Design (cont'd)

- If you open the Column Info for Whole Plots, you'll see that the Design Role is Random Block (JMP is pretty smart!)
- Rename the Whole Plots column with the name of your block



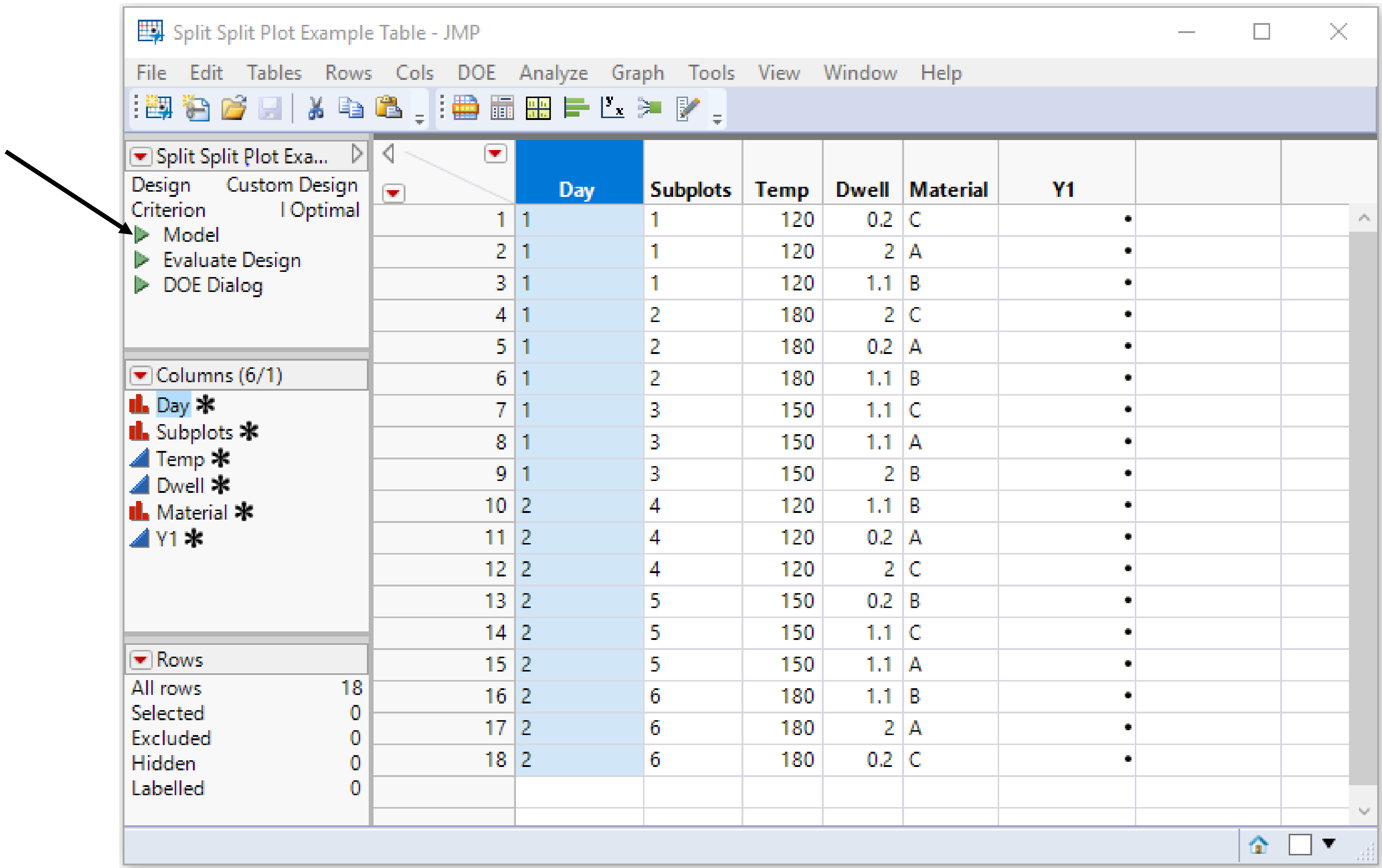
Blocking in a Split-Plot Design (cont'd)

- This shows the final table, with Whole Plots renamed to Day
- This experiment is designed to be run in two days
- What you actually have now is a split-split-plot design

Day	Subplots	Temp	Dwell	Material	Y1
1	1	120	0.2	C	•
1	1	120	2	A	•
1	1	120	1.1	B	•
1	2	180	2	C	•
1	2	180	0.2	A	•
1	2	180	1.1	B	•
1	3	150	1.1	C	•
1	3	150	1.1	A	•
1	3	150	2	B	•
2	4	120	1.1	B	•
2	4	120	0.2	A	•
2	4	120	2	C	•
2	5	150	0.2	B	•
2	5	150	1.1	C	•
2	5	150	1.1	A	•
2	6	180	1.1	B	•
2	6	180	2	A	•
2	6	180	0.2	C	•

Analyzing the Split-Plot Design

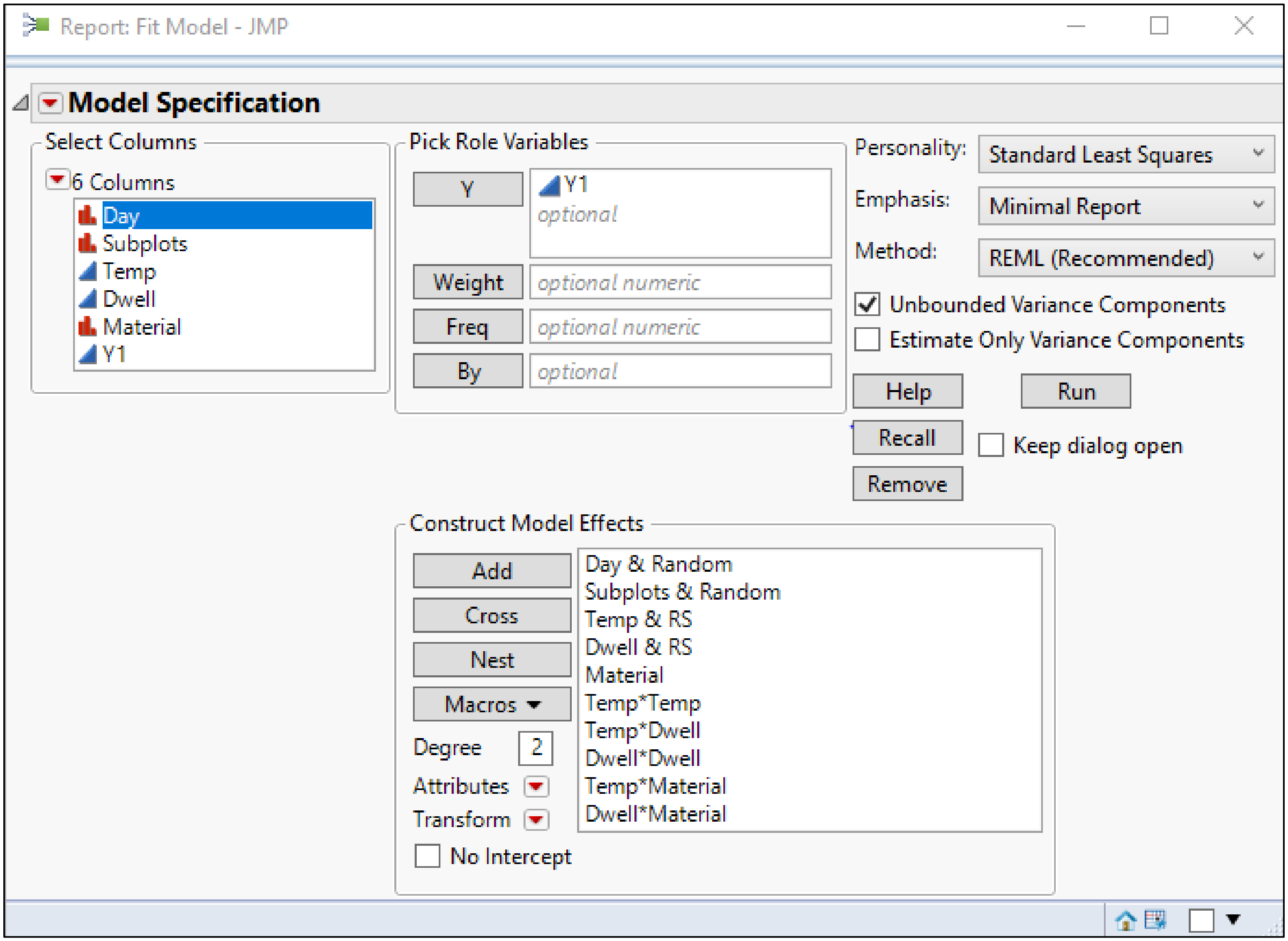
- For the Split-Plot or the Split-Split Plot design, click on the green triangle next to Model after entering data into the table.



	Day	Subplots	Temp	Dwell	Material	Y1
1	1	1	120	0.2	C	•
2	1	1	120	2	A	•
3	1	1	120	1.1	B	•
4	1	2	180	2	C	•
5	1	2	180	0.2	A	•
6	1	2	180	1.1	B	•
7	1	3	150	1.1	C	•
8	1	3	150	1.1	A	•
9	1	3	150	2	B	•
10	2	4	120	1.1	B	•
11	2	4	120	0.2	A	•
12	2	4	120	2	C	•
13	2	5	150	0.2	B	•
14	2	5	150	1.1	C	•
15	2	5	150	1.1	A	•
16	2	6	180	1.1	B	•
17	2	6	180	2	A	•
18	2	6	180	0.2	C	•

Analyzing the Split-Plot Design

- The Fit Model window will look a little different. Leave as is!
- Click Run
- Analyze the residuals and remove terms as with other experiments



This slide intentionally left blank

13 Multiple Response Optimization

- Experiments may have more than one response variable
- You can optimize each response separately . . .
- . . . but you will get different answers for each response!

It is not uncommon to have multiple response variables in a DOE. If you think you have just one, you might want to solicit feedback from one or more knowledgeable colleagues.

In this section we introduce and illustrate the most widely used technique for joint optimization of multiple responses.

Example 1: heat sealing process

- *DOE Participant Files \ heat sealing 2.jmp*
- Run the *Model* script
- Response variables:
  - ✓ *Bond* (bond strength)
  - ✓ *Print* (higher-is-better cosmetic quality rating)
- *Shift* is the only factor we can eliminate
- All other factors are significant for at least one response

Response Bond

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Shift	1	1	3.578	0.8499	0.3671
Temp(120,180)	1	1	1540.835	366.0070	<.0001*
Press(50,150)	1	1	8.439	2.0046	0.1715
Dwell(0.2,2)	1	1	1606.813	381.6793	<.0001*
Temp*Temp	1	1	1363.630	323.9142	<.0001*
Temp*Press	1	1	14.607	3.4697	0.0766
Press*Press	1	1	1.385	0.3290	0.5724
Temp*Dwell	1	1	20235.249	4806.642	<.0001*
Press*Dwell	1	1	0.759	0.1804	0.6754
Dwell*Dwell	1	1	715.715	170.0096	<.0001*

Response Print

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Shift	1	1	0.137812	1.7253	0.2032
Temp(120,180)	1	1	6.821113	85.3929	<.0001*
Press(50,150)	1	1	25.625986	320.8095	<.0001*
Dwell(0.2,2)	1	1	2.121674	26.5611	<.0001*
Temp*Temp	1	1	2.148242	26.8937	<.0001*
Temp*Press	1	1	0.300304	3.7595	0.0661
Press*Press	1	1	0.257674	3.2258	0.0869
Temp*Dwell	1	1	1.613751	20.2024	0.0002*
Press*Dwell	1	1	1.065140	13.3344	0.0015*
Dwell*Dwell	1	1	1.372401	17.1810	0.0005*

Example 1 (cont'd)

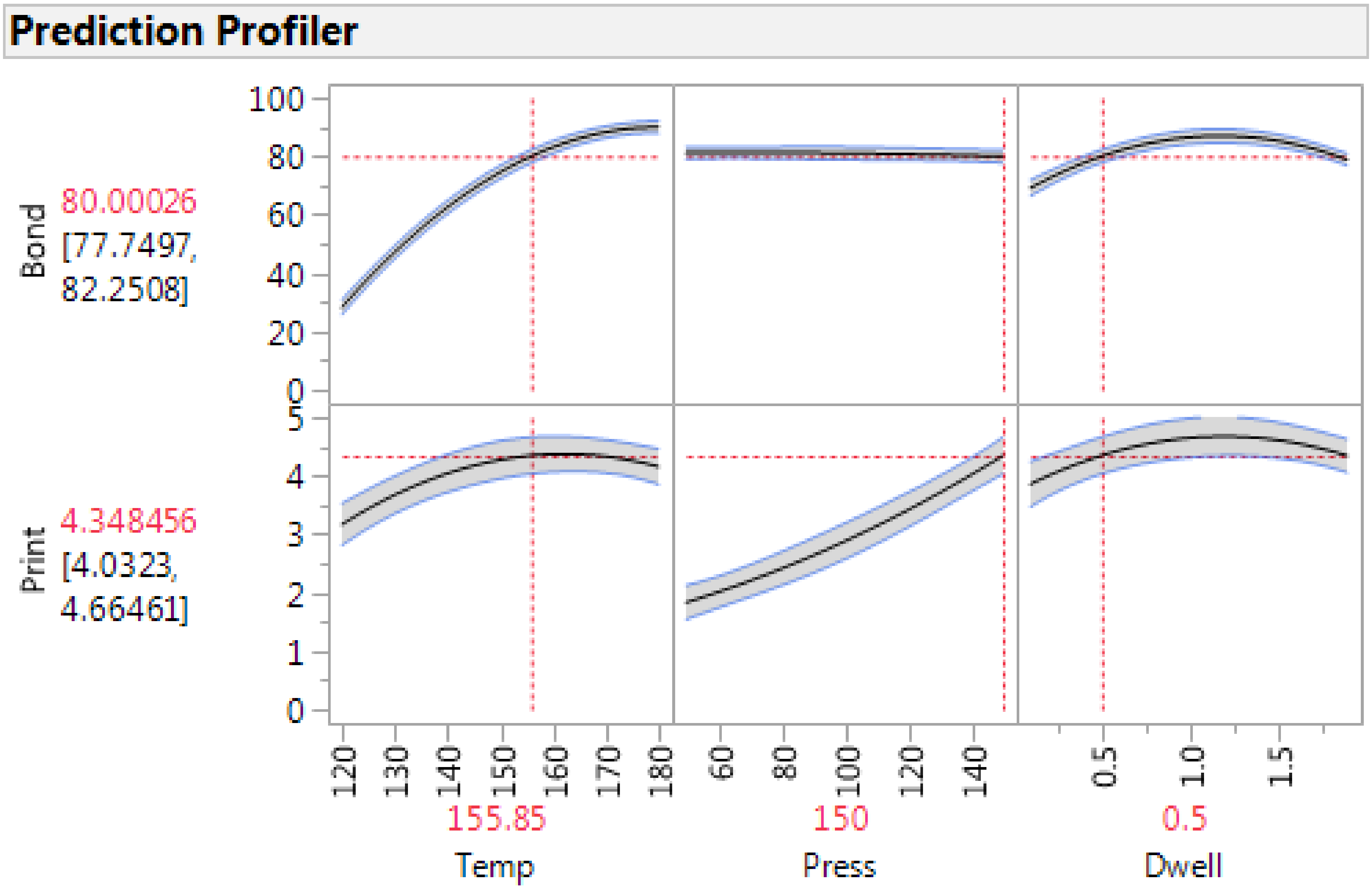
- The Effect Summary displays the lowest p-value from each of the response's Effects Tests
- This makes it easy to find terms to remove from the model
- Remove insignificant terms, as before, using the Effect Summary

Effect Summary

Source	LogWorth	PValue	
Temp*Dwell	25.559	0.00000	
Dwell(0.2,2)	14.223	0.00000	^
Temp(120,180)	14.041	0.00000	^
Temp*Temp	13.515	0.00000	
Press(50,150)	13.473	0.00000	
Dwell*Dwell	10.809	0.00000	
Press*Dwell	2.827	0.00149	
Temp*Press	1.180	0.06606	
Press*Press	1.061	0.08689	
Shift	0.692	0.20319	

Example 1 (cont'd)

We want *Bond* = 80 and *Print* as large as possible.
Here is a solution based on manually exploring the *Prediction Profiler.*



In this example is it easy to find solutions by manually exploring the *Prediction Profiler*.

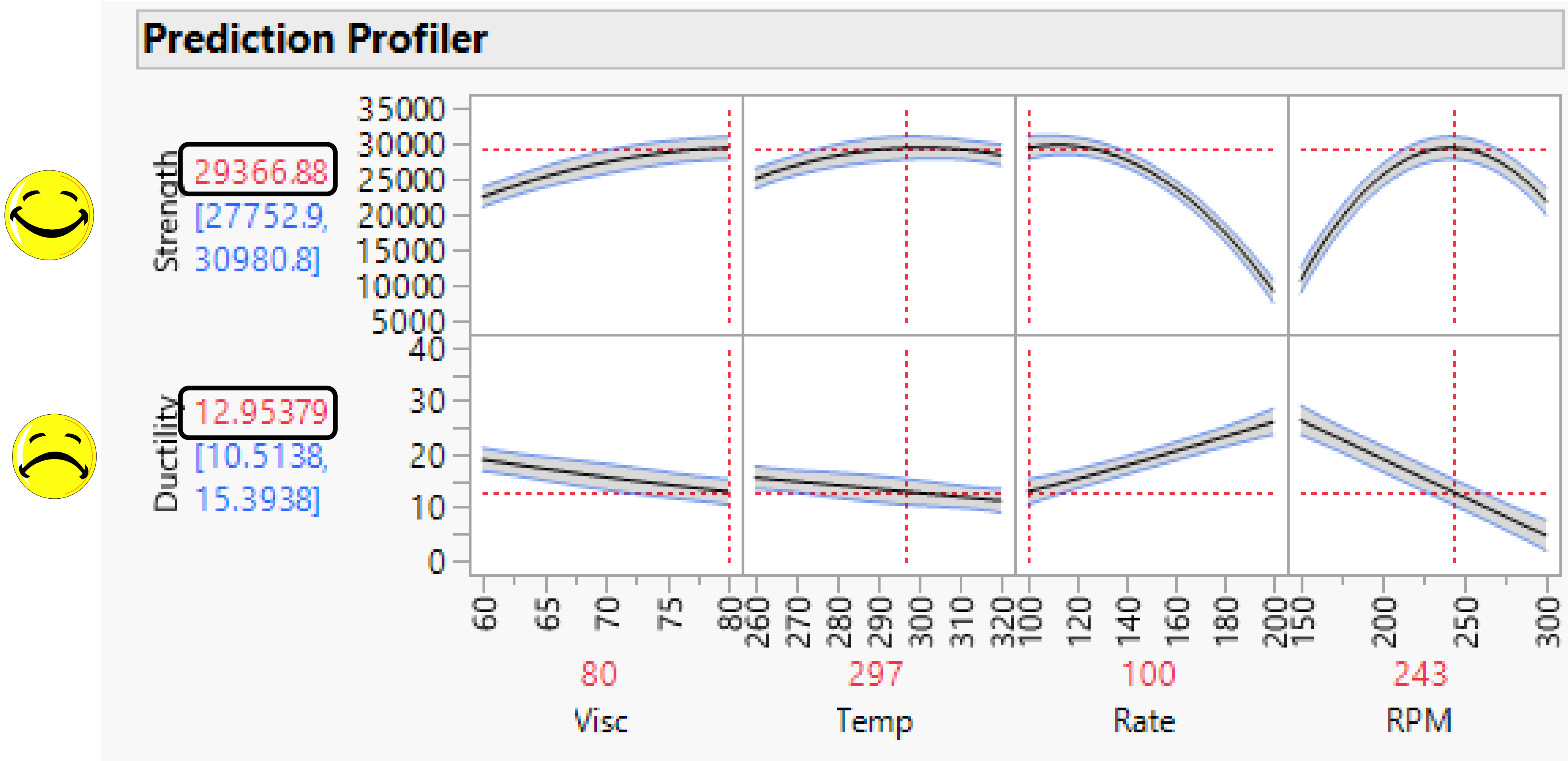
- ✓ *Press* should be set to 150, because this increases *Print* without significantly affecting *Bond*.
- ✓ The baseline value for *Dwell* was 1.0. Reducing this to 0.5 increases throughput while staying above the lowest feasible dwell time (0.2)
- ✓ Once these settings are in place, we can manipulate *Temp* to achieve something very close to 80 psi for *Bond*.

Joint optimization of response variables was not needed in this example. In most applications, however, manual optimization will not achieve the desired results. Extreme versions of this are illustrated in the next two examples.

Close the analysis window and the data table without saving.

Example 2: extrusion process

$$(Visc, Temp, Rate, RPM) \approx (80, 297, 100, 243)$$
$$Ductility \approx 13$$



Data sets \ extrusion 2

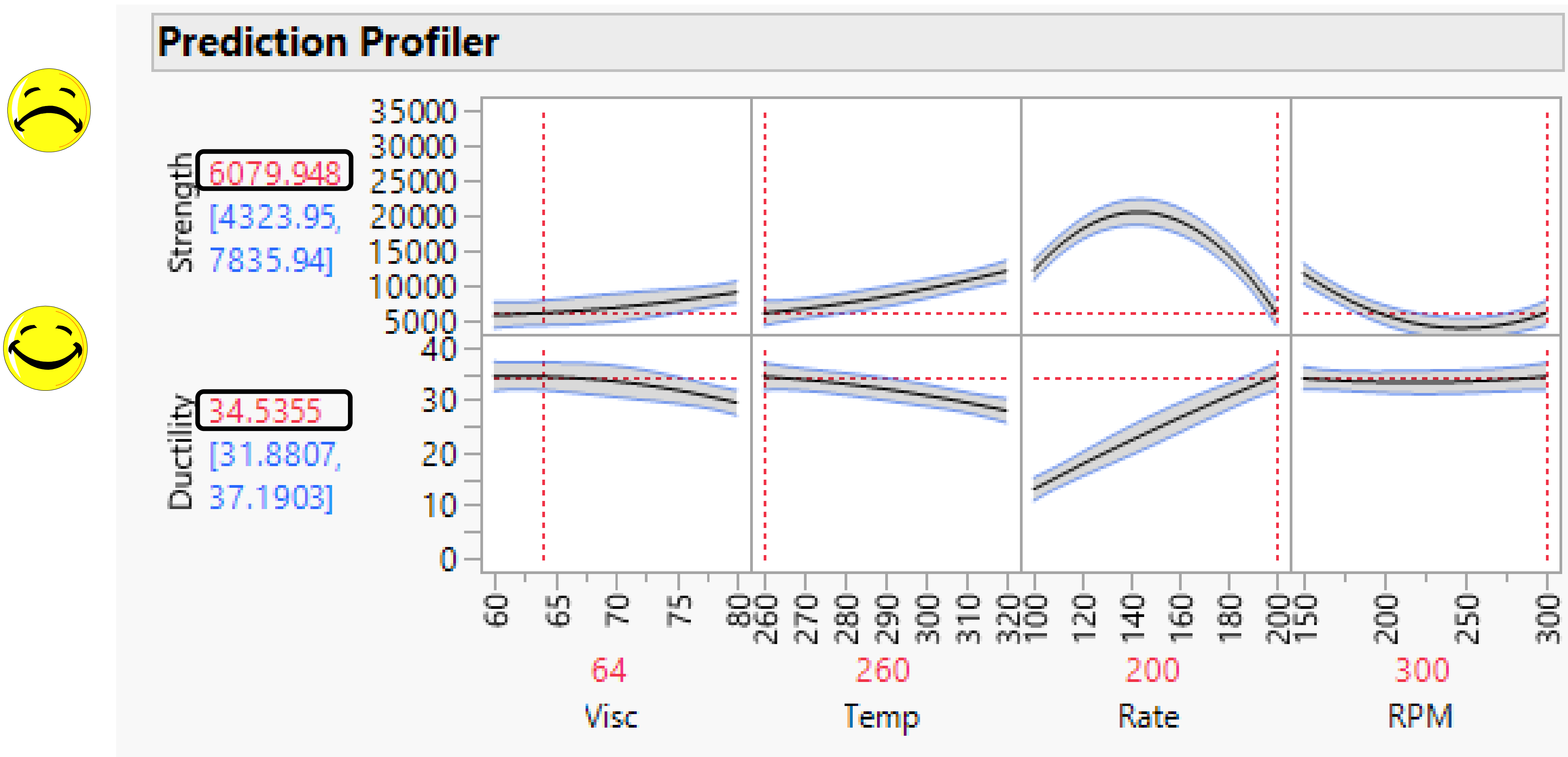
This example is based on data from an experiment to optimize the mechanical properties of an extruded plastic material. We want *Strength* to be as high as possible while maintaining a lower bound of 20 for *Ductility*.

The solution for *Strength* (29367) shown above was found by visually exploring the *Prediction Profiler*. However, this approach resulted in an unacceptably low *Ductility* (13).

Example 2 (cont'd)

$$(Visc, Temp, Rate, RPM) \approx (64, 260, 200, 300)$$

$$Strength \approx 6080$$



The solution for *Ductility* (35) shown above was found by visually exploring the *Prediction Profiler*. However, this approach resulted in an unacceptably low *Strength* (6080).

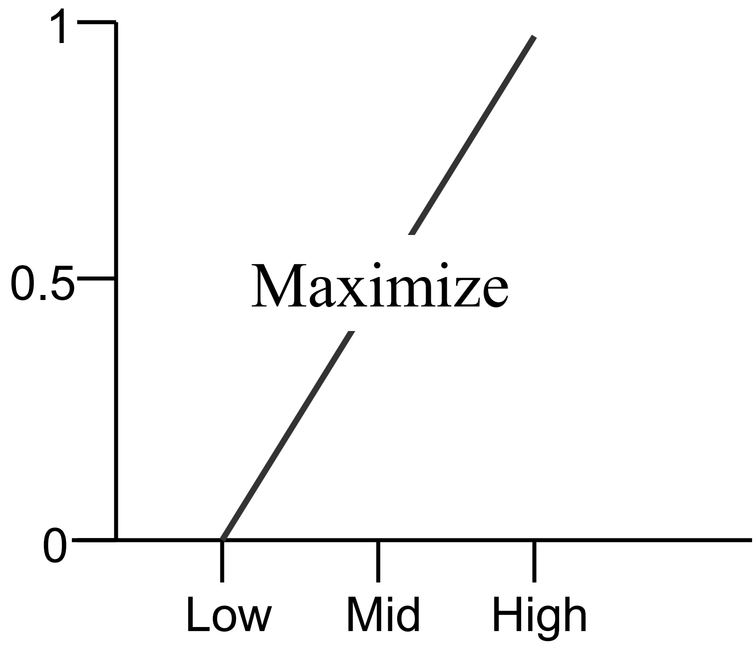
Joint optimization of responses

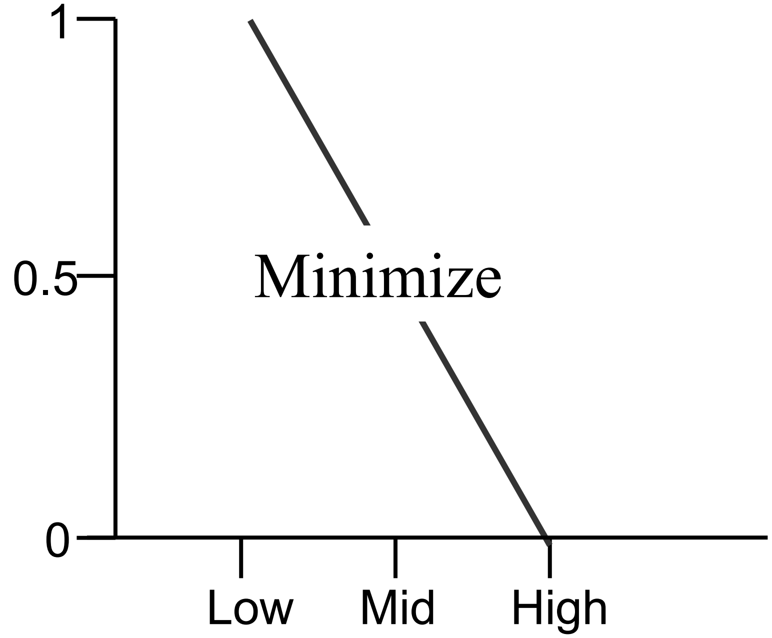
- Each response has a goal (minimize, maximize or target)
- Define a "desirability function" for each response
- Combine the individual desirabilities into a single overall desirability function
- Maximize the overall desirability to jointly optimize all responses

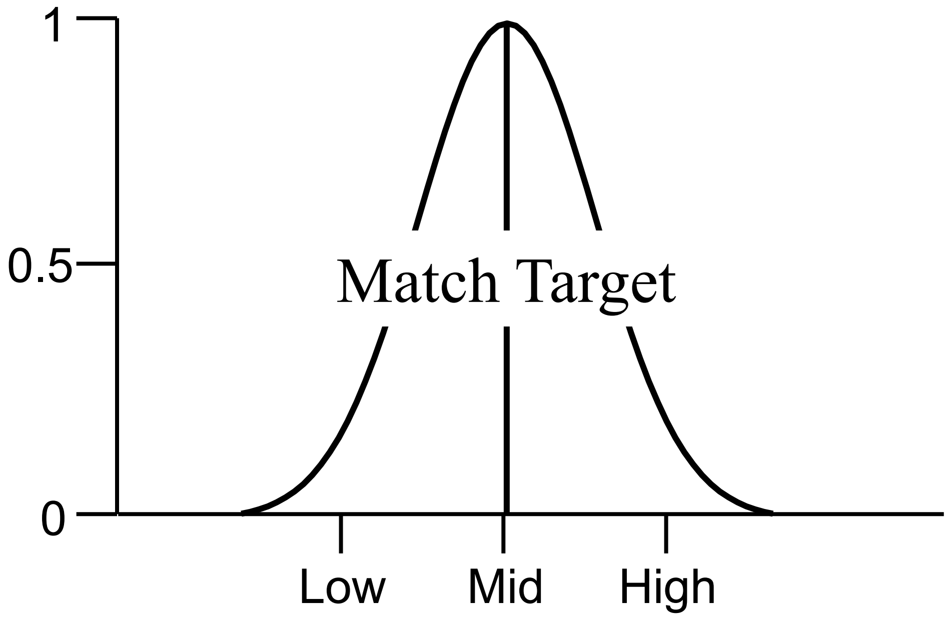
Desirability is a unitless quantity between 0 and 1, defined so that higher is better. JMP supplies default desirability functions based on the experimental data for your response variables. You must redefine the desirability functions so that they represent your objectives for each response variable.

You start by setting the general goal for each response: *Maximize*, *Minimize* or *Match Target*. Then you specify low, middle, and high data values to fine tune the shape of the desirability functions.

Default desirability functions







The desirability function is increasing for *Maximize* responses and decreasing for *Minimize* responses. It is bell-shaped for *Match Target* responses.

For *Minimize* responses with a lower bound of 0, it is a good idea to make the *Low* value equal to 0. Examples are number of defects, fraction defective, cycle time, standard deviation, cost of waste, etc.

The low and high values for a *Match Target* response are used to define the allowable deviation from the target value.

Overall desirability

- The overall desirability function for the response variables (Y_1, Y_2, · · ·) is

$$\sqrt{\left(Y_1 \text{ desirability}\right) \times \left(Y_2 \text{ desirability}\right) \times \cdots}$$

- It is the geometric mean of the desirability functions for all the individual response variables

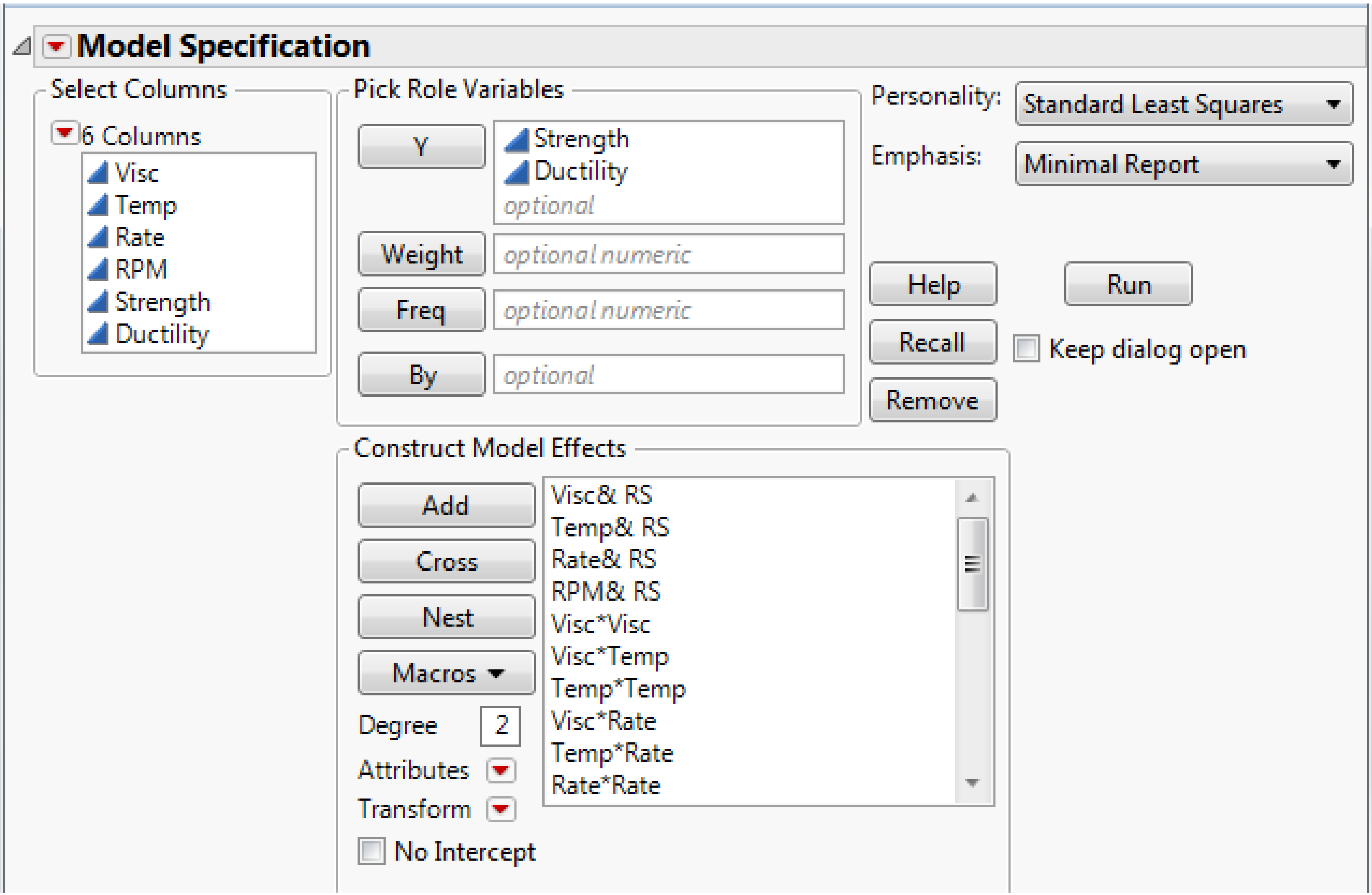
- With a geometric mean, the overall desirability will be zero whenever any individual response desirability is zero

A *weighted* geometric mean can be used. The weights (called *importance* in JMP) allow users to specify relative priorities for the responses. The higher the importance, the greater the influence the response has in determining the overall solution found by the optimization algorithm.

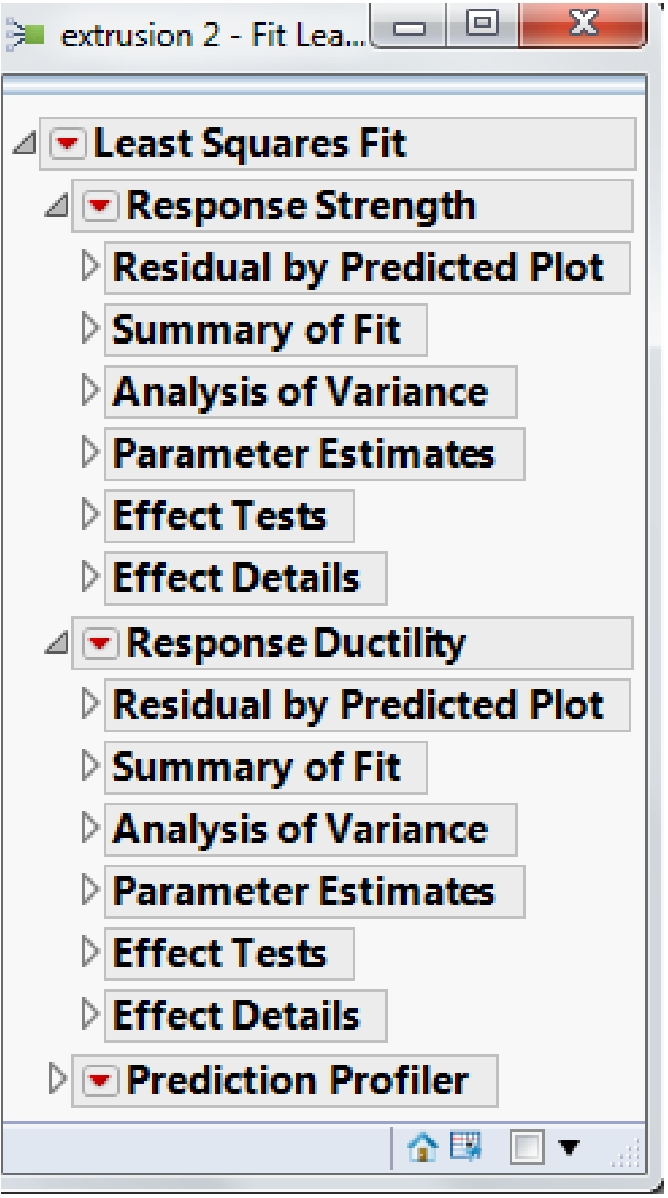
The vast majority of users do not go into this level of detail.

Example 2 (revisited)

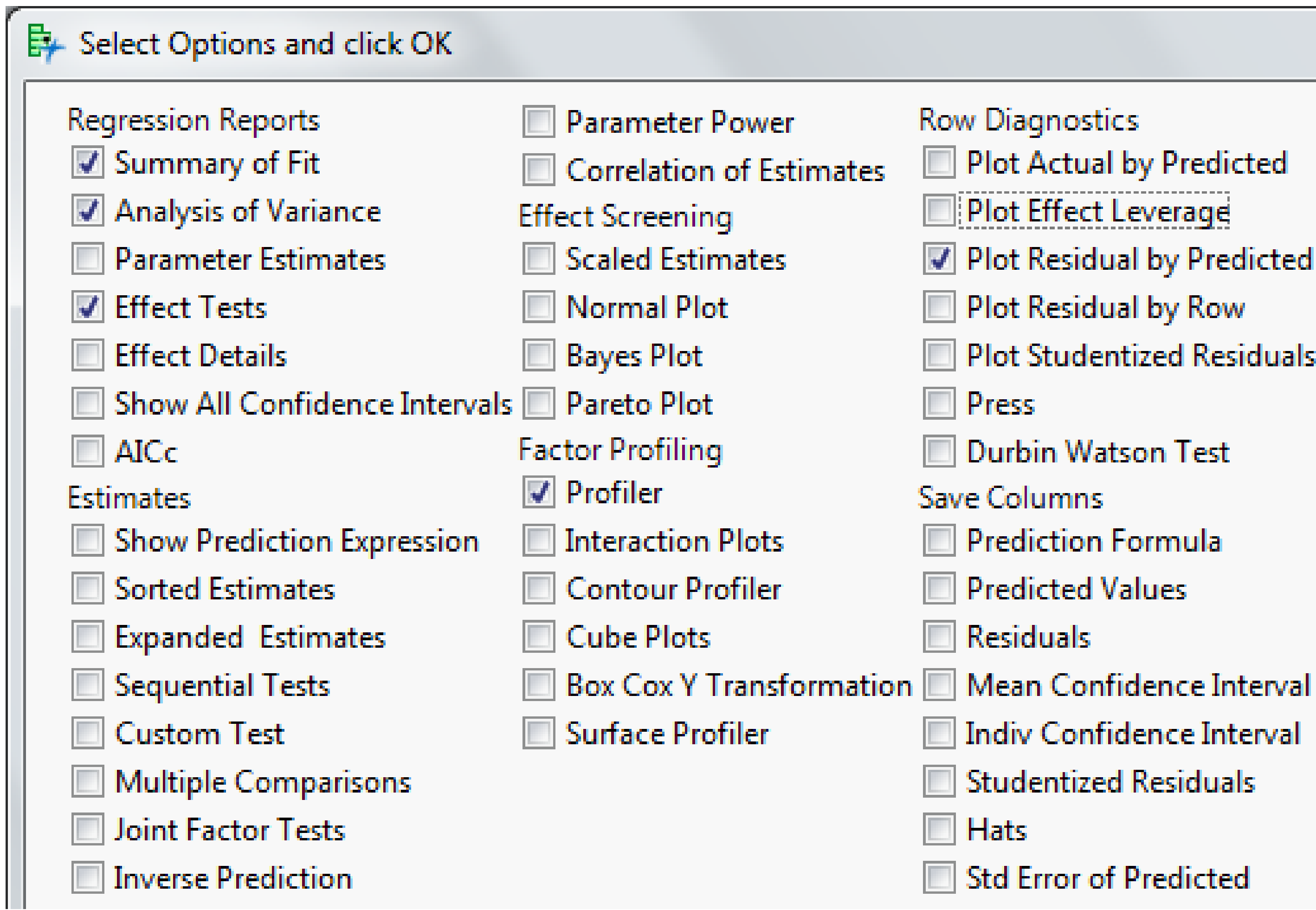
DOE Participant Files \ extrusion 2.jmp → *Model* script → *Model Specification* → Run



Example 2 (cont'd)



- Alt-click on *Response Strength* red triangle → uncheck *Parameter Estimates*, *Effect Details*, *Plot Effect Leverage* → OK
- Repeat for *Response Ductility*



“Pruning” the models

- The *Effect Summary* combines the P-values for all responses
- Removing terms here applies to the *Effects Tests* for one or more responses
- The usual threshold is P > 0.15

Effect Summary

Source	LogWorth	PValue	
Rate*RPM*RPM	7.277	0.00000	
Rate*RPM	7.181	0.00000	^
Rate(100,200)	6.383	0.00000	^
Rate*Rate*RPM	6.300	0.00000	
RPM(150,300)	6.165	0.00000	^
Rate*Rate	5.394	0.00000	^
RPM*RPM	4.809	0.00002	^
Temp(260,320)	3.121	0.00076	
Visc(60,80)	2.913	0.00122	
Temp*RPM	1.805	0.01565	
Visc*Visc*Temp	1.568	0.02705	
Visc*Rate	1.559	0.02763	
Visc*RPM	1.549	0.02822	
Visc*Temp*Temp	1.526	0.02980	
Visc*Visc*RPM	1.429	0.03723	
Temp*Temp*Rate	1.212	0.06143	
Visc*Visc*Rate	0.844	0.14308	
Temp*Temp*RPM	0.826	0.14926	
Temp*Rate*Rate	0.808	0.15550	
Temp*RPM*RPM	0.792	0.16134	
Temp*Temp	0.668	0.21486	^
Visc*Rate*Rate	0.571	0.26852	
Visc*Temp	0.470	0.33863	^
Temp*Rate	0.358	0.43877	^
Visc*Visc	0.299	0.50227	^
Visc*RPM*RPM	0.215	0.60885	

Remove Add Edit FDR ('^' denotes effects with containing effects above them)

“Pruning” the models (cont’d)

Effect Summary

Source	LogWorth	PValue
Rate*RPM*RPM	11.346	0.00000
Rate*RPM	10.339	0.00000
Rate*Rate*RPM	10.023	0.00000
Rate(100,200)	9.762	0.00000
RPM(150,300)	9.741	0.00000
Rate*Rate	8.273	0.00000
RPM*RPM	7.572	0.00000
Visc(60,80)	6.093	0.00000
Temp(260,320)	4.727	0.00002
Temp*RPM	2.347	0.00449
Visc*RPM	2.138	0.00727
Visc*Visc*RPM	1.935	0.01163
Visc*Temp*Temp	1.853	0.01404
Visc*Rate	1.815	0.01531
Visc*Visc*Temp	1.499	0.03171
Temp*Temp*RPM	1.238	0.05774
Temp*Temp*Rate	1.197	0.06350
Temp*Rate*Rate	1.074	0.08435
Visc*Visc*Rate	1.000	0.10006

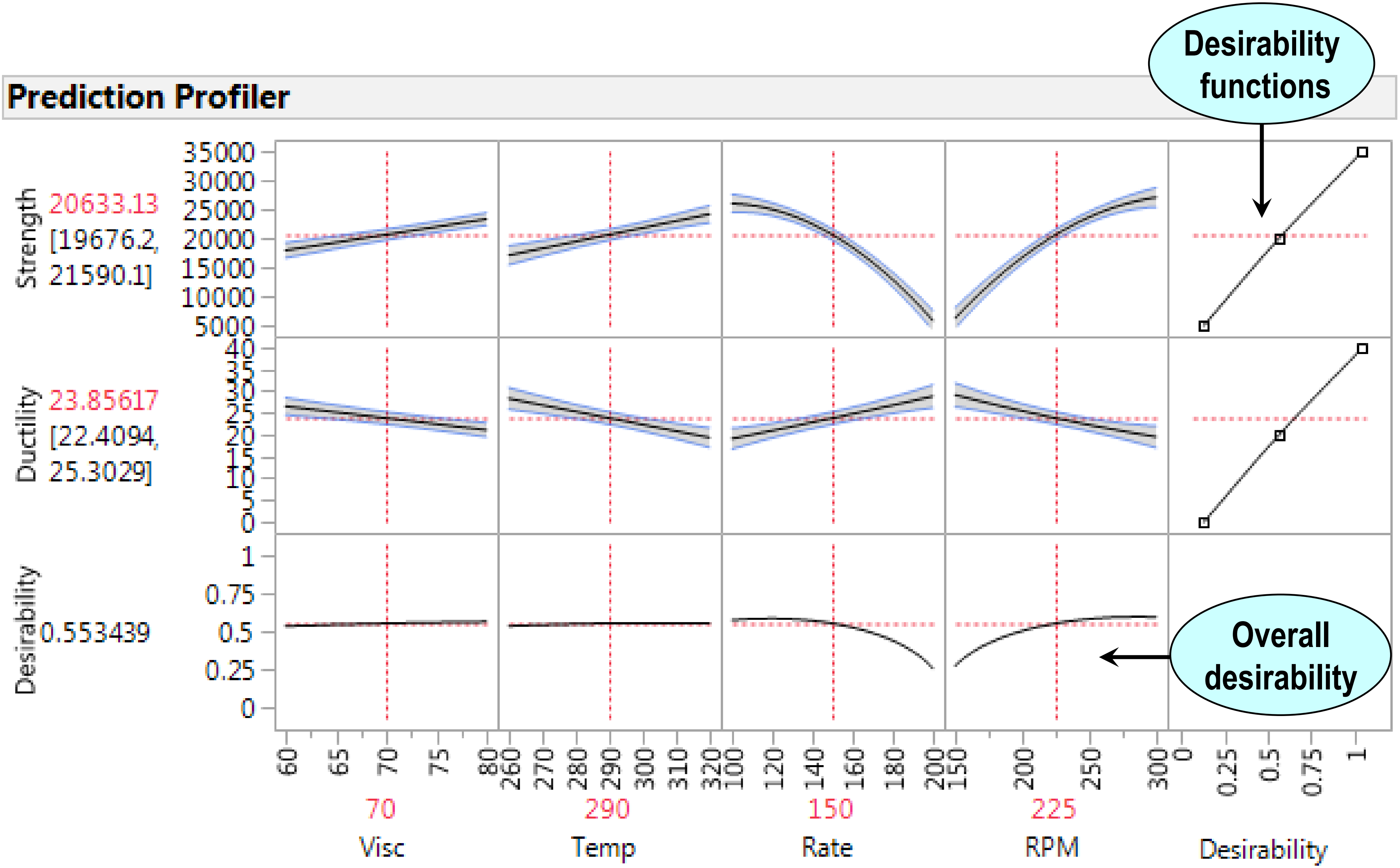
Effect Tests for Strength

Source	Prob > F
Visc(60,80)	<.0001*
Temp(260,320)	<.0001*
Rate(100,200)	<.0001*
RPM(150,300)	<.0001*
Visc*Rate	0.0153*
Rate*Rate	<.0001*
Visc*RPM	0.0073*
Temp*RPM	0.0045*
Rate*RPM	<.0001*
RPM*RPM	<.0001*
Visc*Visc*Temp	0.0317*
Visc*Visc*Rate	0.1001
Visc*Visc*RPM	0.0116*
Visc*Temp*Temp	0.0140*
Temp*Temp*Rate	0.0635
Temp*Temp*RPM	0.0577
Temp*Rate*Rate	0.0844
Rate*Rate*RPM	<.0001*
Rate*RPM*RPM	<.0001*

Effect Tests for Ductility

Source	Prob > F
Visc(60,80)	<.0001*
Temp(260,320)	0.0001*
Rate(100,200)	0.0003*
RPM(150,300)	0.0005*
Visc*Rate	0.4624
Rate*Rate	0.8364
Visc*RPM	0.5440
Temp*RPM	0.7358
Rate*RPM	<.0001*
RPM*RPM	0.4084
Visc*Visc*Temp	0.0527
Visc*Visc*Rate	0.8994
Visc*Visc*RPM	0.8700
Visc*Temp*Temp	0.8114
Temp*Temp*Rate	0.9857
Temp*Temp*RPM	0.3483
Temp*Rate*Rate	0.3080
Rate*Rate*RPM	0.9424
Rate*RPM*RPM	0.5257

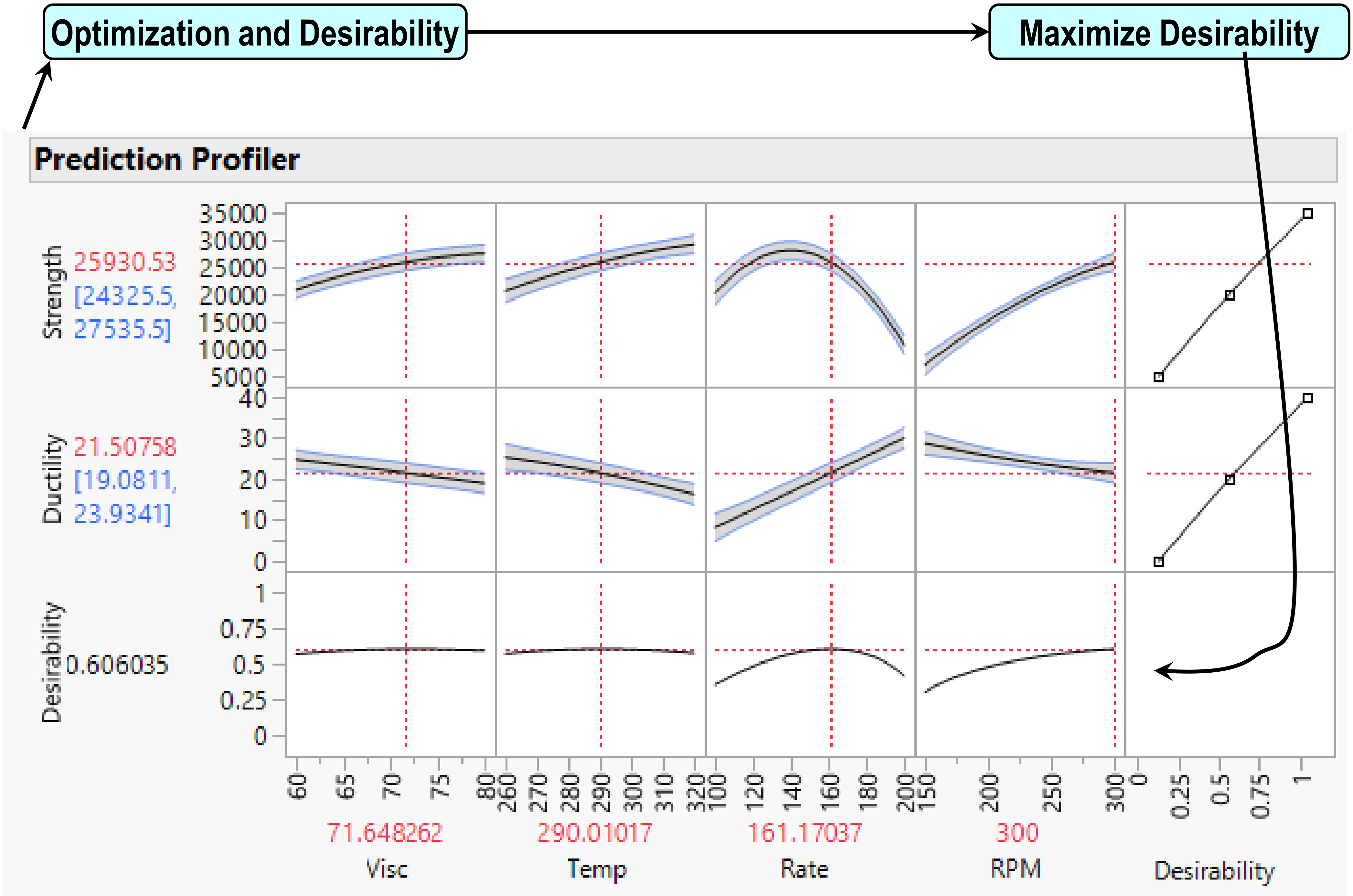
Example 2 (cont'd)



Here is the default *Prediction Profiler* for the four-factor extrusion experiment. The individual desirability functions are shown in the right-most column. In this case they are both increasing functions because our general objective for both responses is *Maximize*.

The overall desirability is a function of the experimental factors, and is shown in the bottom row. By default, it is the unweighted geometric mean of the individual desirability functions.

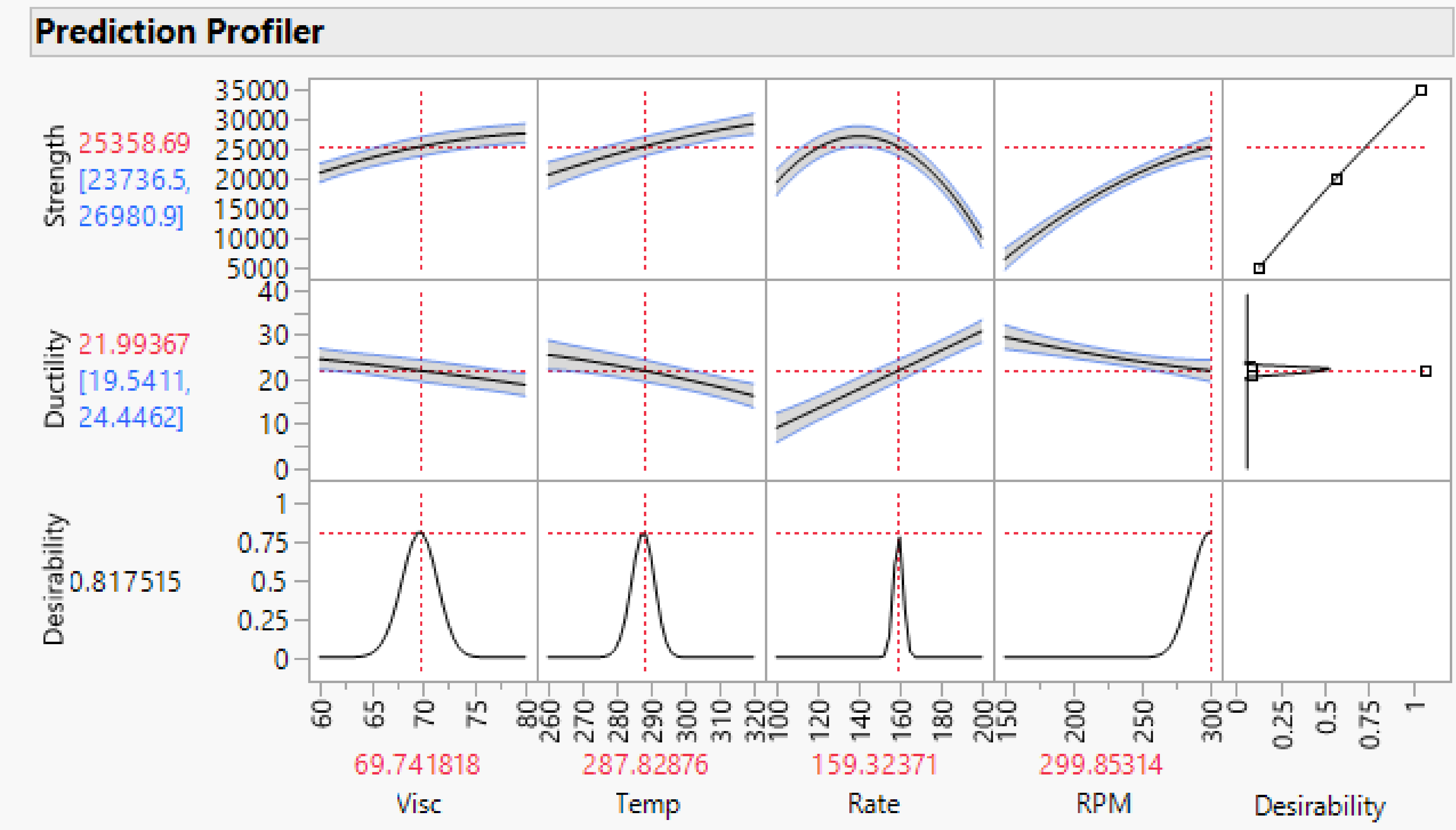
Example 2 (cont'd)



Shown above is the *Prediction Profiler* after selecting *Maximize Desirability* from the red triangle menu. We have increased average *Strength* to 25930, and decreased average *Ductility* to 21.5.

Example 2 (cont'd)

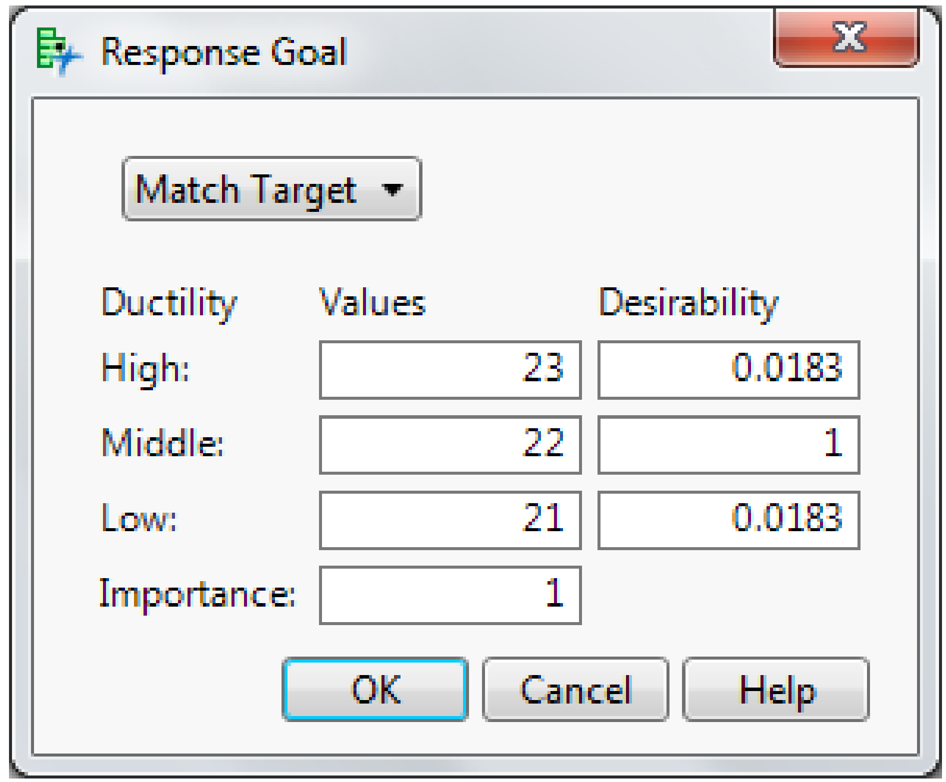
Using a *Match Target* objective (see next slide)



To obtain the results shown above, double-click in the individual *Desirability* pane (on the right) for *Ductility*. Change the specifications as shown below, click OK, run *Maximize Desirability* again.

Predicted average *Strength* is now 25359, predicted average *Ductility* is 22.

The 95% confidence interval is (19.5, 24.4). This is an improvement over the previous confidence interval (19.0, 24.0), which would have allowed *Ductility* to vary a little further below 20.



Note: Due to the iterative process used in the prediction profiler, results may vary slightly from what's shown in the above slide.

Least Squares Fit red triangle → Save Script → To Data Table → Save Script As → Name: *Fit Least Squares* → OK.

Exercise 12.1

(a) *DOE Participant Files \ heat sealing 2*. Run the model script. Use the *Effect Summary* to remove model terms with P > 0.15.

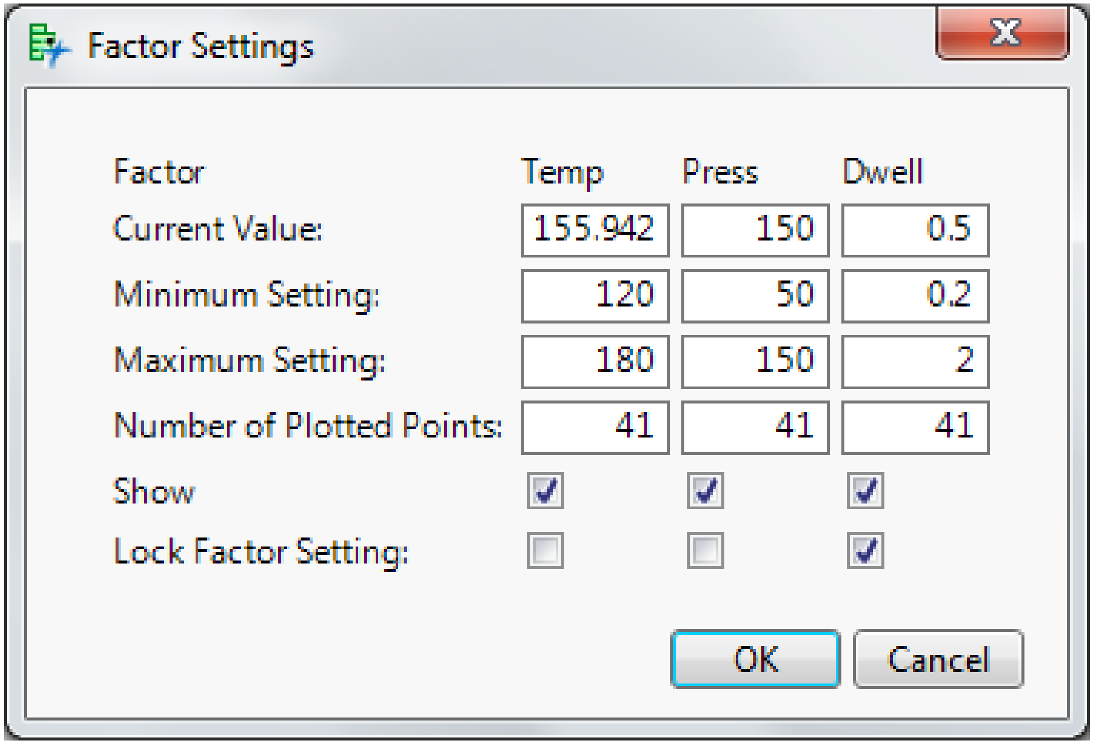
(b) Go to the *Prediction Profiler*. Our target for average *Bond* is 80, with a tolerance of ±5. The highest possible value for average *Print* is 5. Average *Print* must exceed 4. Modify the desirability functions for *Bond* and *Print* accordingly. Click *Prediction Profiler* red triangle → *Optimization and Desirability* → *Save Desirabilities*.

(c) Click *Prediction Profiler* red triangle → *Optimization and Desirability* → *Maximize Desirability*.

d) The Production Manager is unhappy with our solution. It achieves excellent bond strength (80) and print quality (4.8), but the proposed increase in dwell time would reduce throughput from 300 to 50 bags per minute!

To look for a compromise, select *Reset Factor Grid* on the *Prediction Profiler* red triangle. We want to hold *Dwell* at a low value, say 0.5. Type 0.5 for *Current Value,* check the *Lock Factor Setting* box, then click OK. The vertical line on the *Dwell* profile should now be solid.

Exercise 12.1 (cont'd)



Factor Settings

Factor	Temp	Press	Dwell
Current Value:	155.942	150	0.5
Minimum Setting:	120	50	0.2
Maximum Setting:	180	150	2
Number of Plotted Points:	41	41	41
Show	☑	☑	☑
Lock Factor Setting:	☐	☐	☑

OK Cancel

e) Run *Maximize Desirability* again. The optimal factor settings are shown in the *Current Value* row. The response averages are 80.08 for *Bond* and 4.35 for *Print*.

f) Save your script, close and save the data table.

Exercise 12.2

a) Assembly of inkjet print cartridges includes an ultrasonic welding operation with X variables *Force*, *Energy, Amplitude,* and *Cavity* (identifies the tool cavity in which each plastic cartridge was molded). The response variables are *Weld depth* and *Leak rate*.

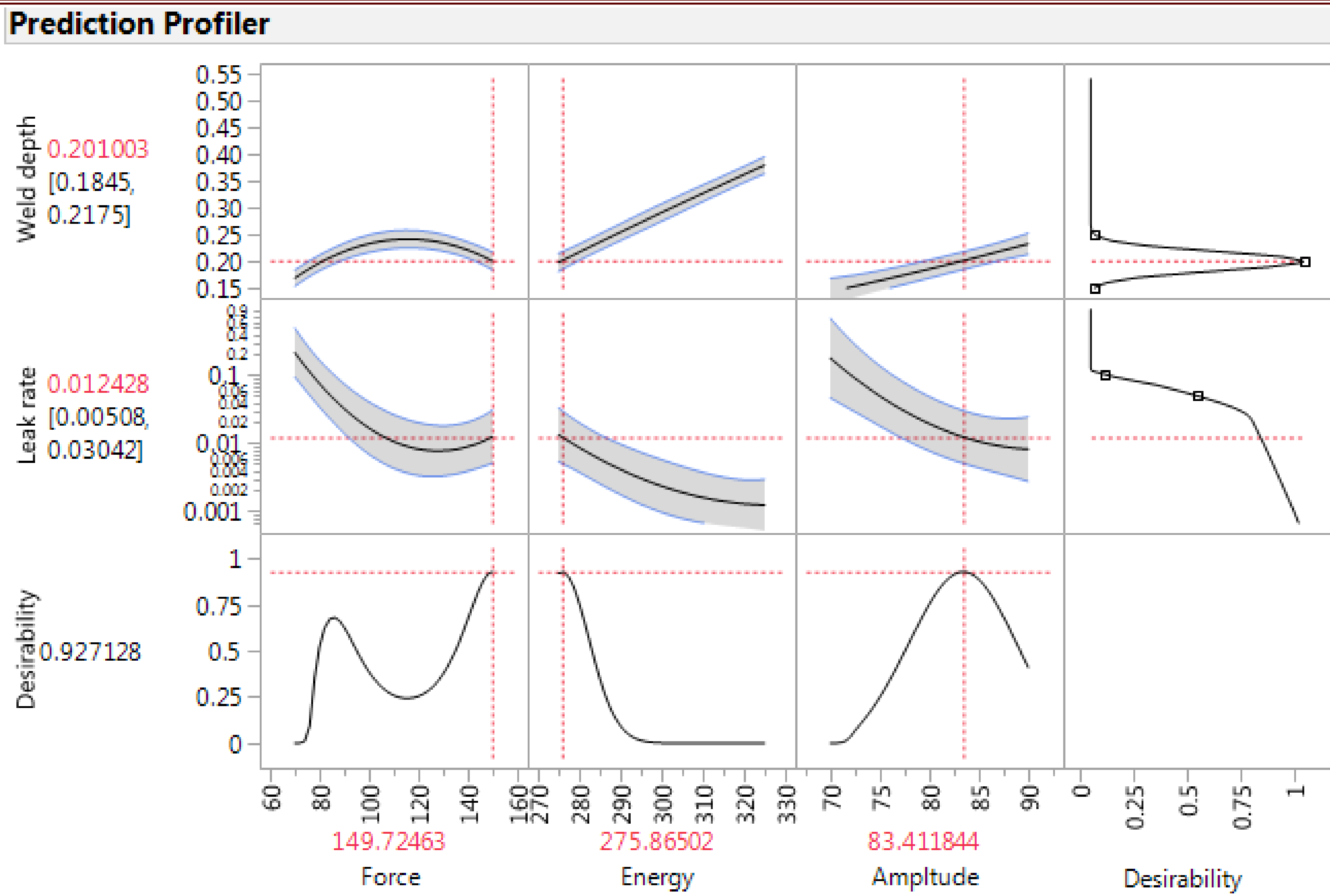
b) *DOE Participant Files \ ultrasonic welding 2.* Run the model script. Use a Log transformation for *Leak rate*. Use *Effect Summary* to prune the models.

c) Go to the *Prediction Profiler*. The target for average *Weld depth* is 0.20, with a tolerance of ± 0.05. The lowest possible value for average *Leak rate* is 0. We require mean *Leak Rate* to be no larger than 0.10.

d) Modify the desirability functions for *Weld depth* and *Leak rate* accordingly. Click *Prediction Profiler* red triangle → *Optimization and Desirability* → *Save Desirabilities*.

e) Click *Prediction Profiler* red triangle → *Optimization and Desirability* → *Maximize Desirability*. See next slide.

Exercise 12.2 (cont'd)



f) Least Squares Fit → *Save Script* → *To Data Table* → Name: *Fit Least Squares* → OK

g) Save data table.

Exercise 12.3 (Homework)

a) *DOE Participant Files \ electron microscope.* Run the *Model* script. In this case, it will take you directly to the *Model Dialog*. Apply Log transformations to all 4 response variables, then run the model.

b) Click *Least Squares Fit* red triangle → *Effect Summary* → prune the models. See slide below.

c) Go to the *Prediction Profiler*. We want to minimize all 4 responses. Use the same desirability functions for all 4 responses: High = 2, Middle = 1, Low = 0. Click *Prediction Profiler* red triangle → *Optimization and Desirability* → *Save Desirabilities*.

d) Click *Prediction Profiler* red triangle → *Reset Factor Grid* → *Factor Settings* → click the *Lock Factor Setting* box under *Tool* → OK. See next page.

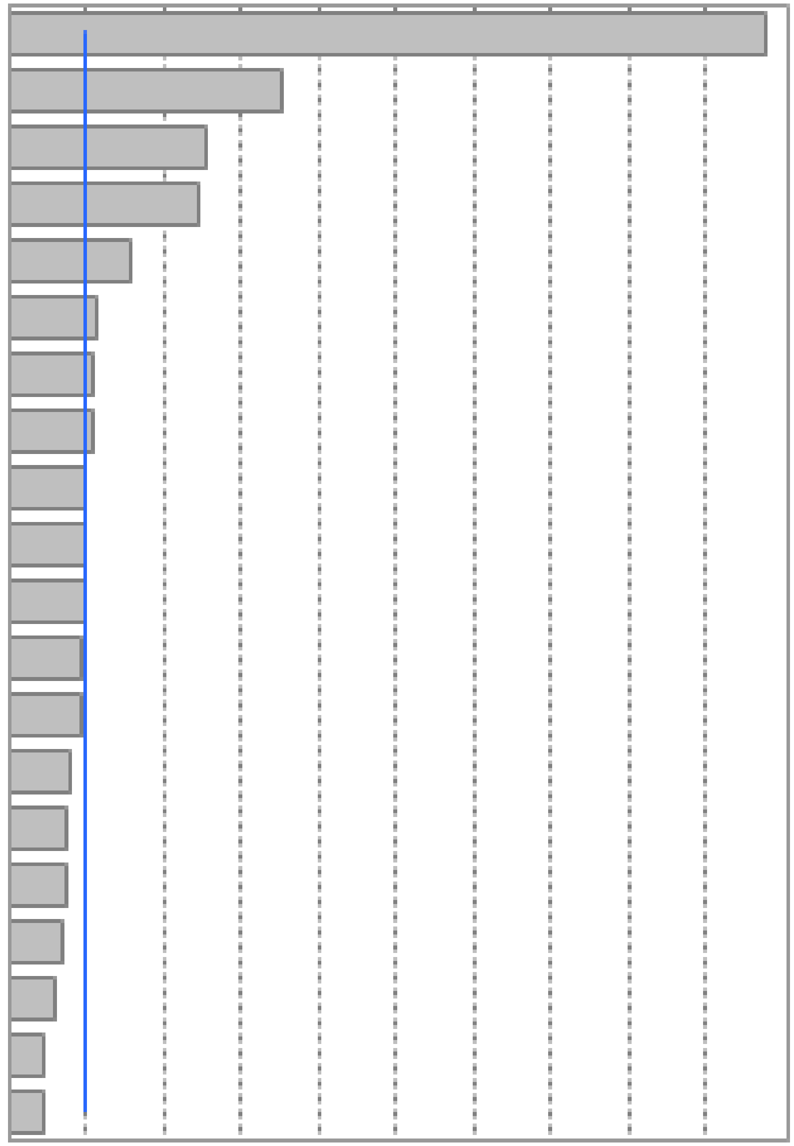
e) Run *Maximize Desirability* separately for each *Tool* (A, B, C). Give the average values of the 4 responses for each tool. See next page.

f) Save your script, close and save the data table.

Exercise 12.3 (cont'd)

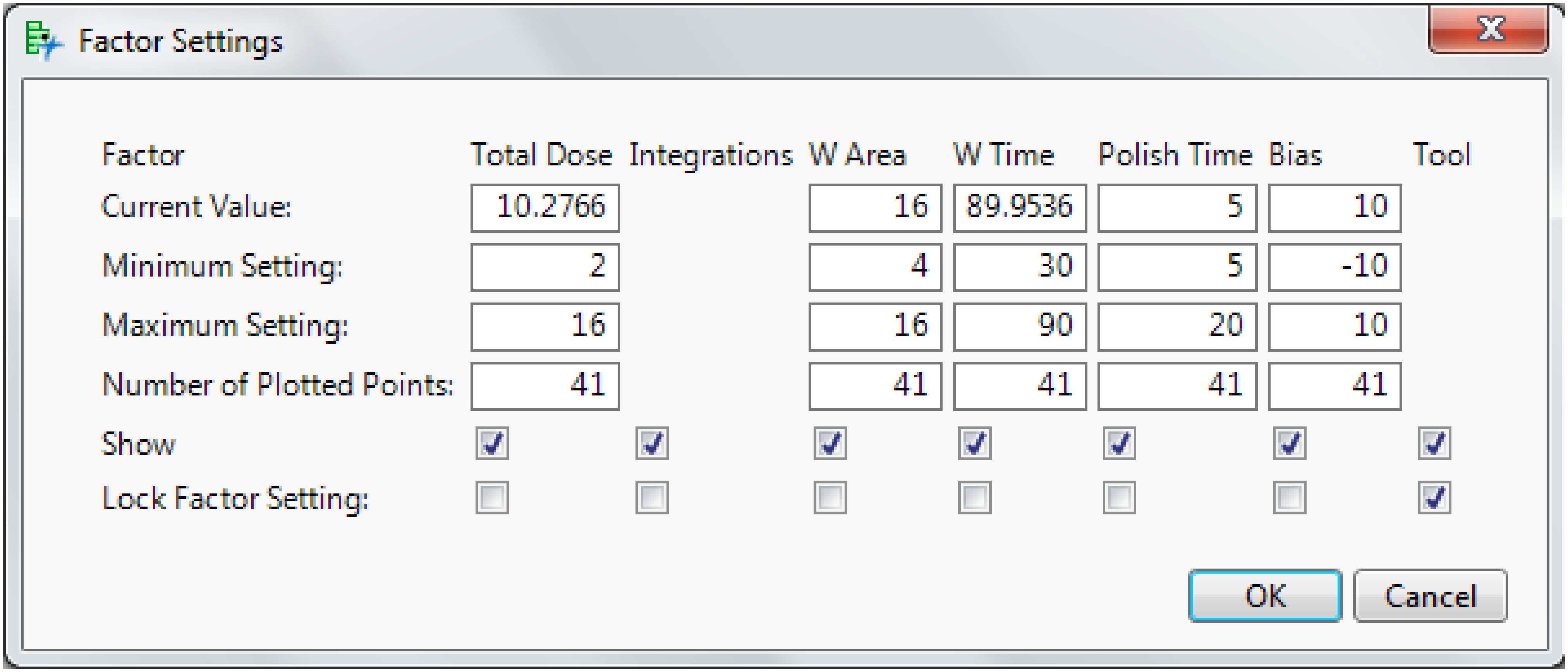
(b) Effect Summary

Source	LogWorth	PValue
Tool	19.514	0.00000
Total Dose(2,16)	7.057	0.00000
Bias*Tool	5.140	0.00001
Bias(-10,10)	4.892	0.00001
Total Dose*Tool	3.203	0.00063
W Time*Bias	2.294	0.00509
W Time(30,90)	2.232	0.00586
Total Dose*Total Dose	2.229	0.00590
Bias*Bias	2.003	0.00994
Integrations	1.961	0.01094
W Time*Tool	1.957	0.01103
Total Dose*W Time	1.915	0.01216
W Area(4,16)	1.858	0.01388
W Area*Tool	1.596	0.02536
Integrations*W Time	1.499	0.03172
W Time*W Time	1.483	0.03288
Integrations*W Area	1.371	0.04255
Polish Time(5,20)	1.247	0.05662
W Area*W Time	0.950	0.11211
W Area*Bias	0.941	0.11449



Exercise 12.3 (cont'd)

(d) Reset Factor Grid



Factor	Total Dose	Integrations	W Area	W Time	Polish Time	Bias	Tool
Current Value:	10.2766		16	89.9536	5	10	
Minimum Setting:	2		4	30	5	-10	
Maximum Setting:	16		16	90	20	10	
Number of Plotted Points:	41		41	41	41	41	
Show	☑	☑	☑	☑	☑	☑	☑
Lock Factor Setting:	☐	☐	☐	☐	☐	☐	☑

Exercise 12.3 (cont'd)

(e) Average responses by tool

Tool	S-Height	S-Width	D-Height	D-Width
A	1.33	1.13	1.10	0.95
B	1.41	0.76	1.36	1.08
C	1.48	1.32	1.94	1.57