Transforming Data to Make Better Predictions

Mastering JMP
Webcast
11 March 2021

Tom Donnelly, PhD, CAP
Principal Systems Engineer
JMP Defense & Aerospace Team
SAS Institute, Inc.
tom.donnelly@jmp.com

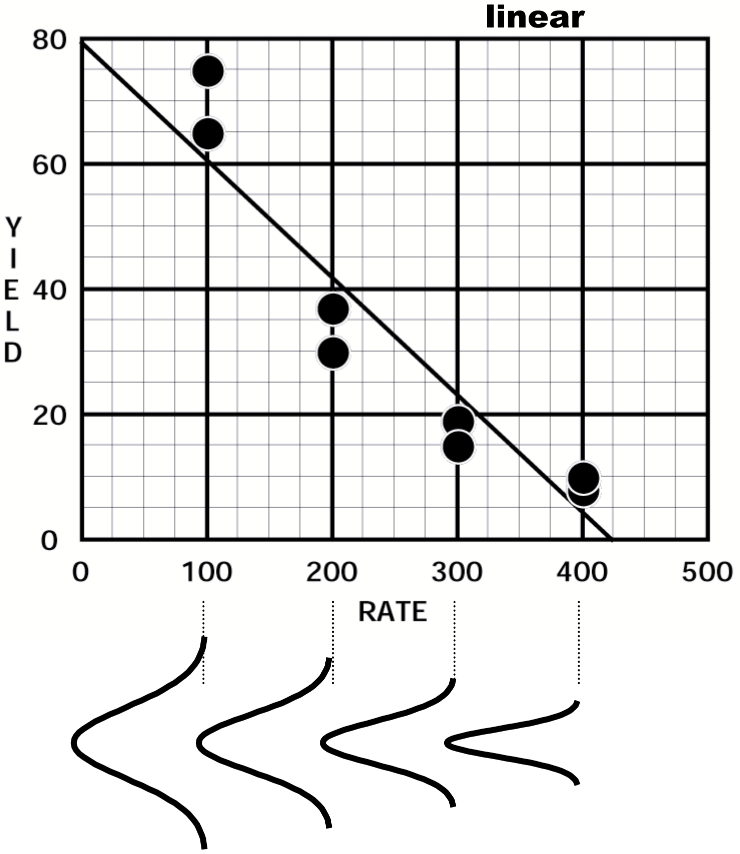
Statistical Discovery.™ From SAS.

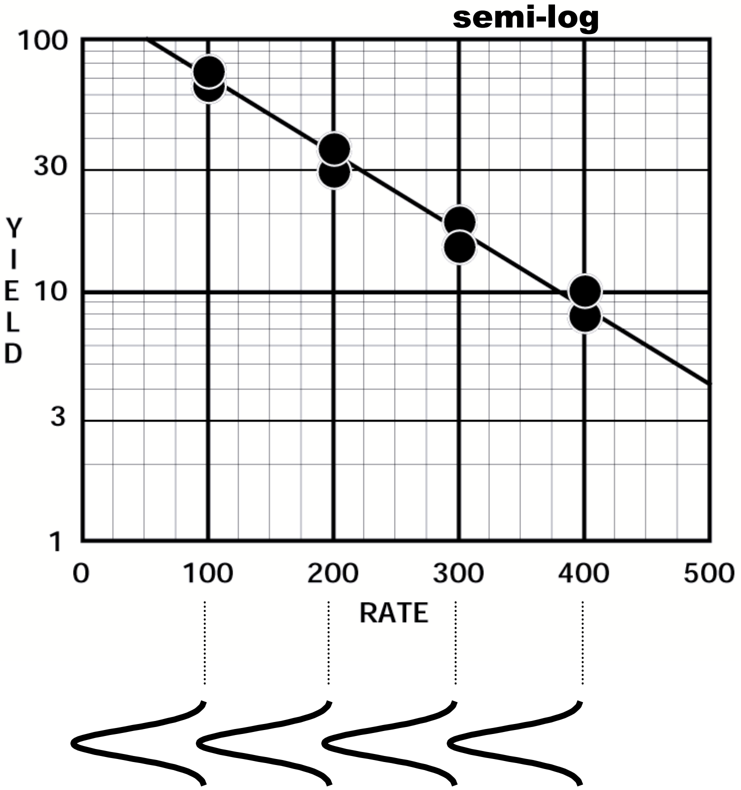
Data Transformations - Why Do Them?

- Remedy for lack of fit
- Plot predictions will not violate physical limits
  - “# of Counts” not negative;
  - “YIELD” not > 100%
- Make error more uniform across design region
  (also called “stabilizing the variance”)

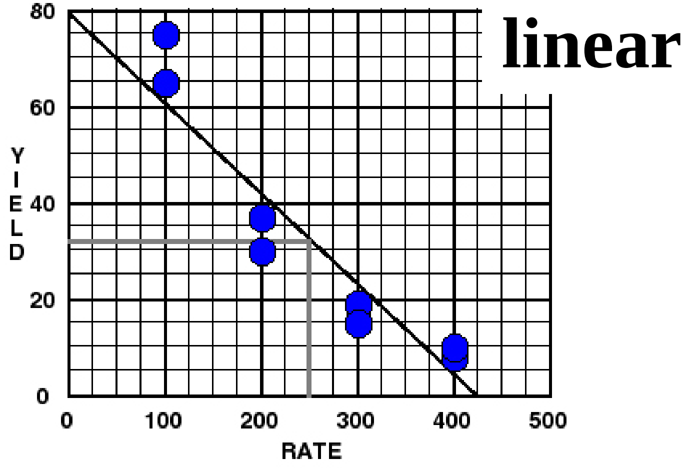
Transformations change the scale of the response to make it more nearly conform to the usual regression assumptions, the most important of which are that the data are independent and follow a **normal distribution with a constant variance.**

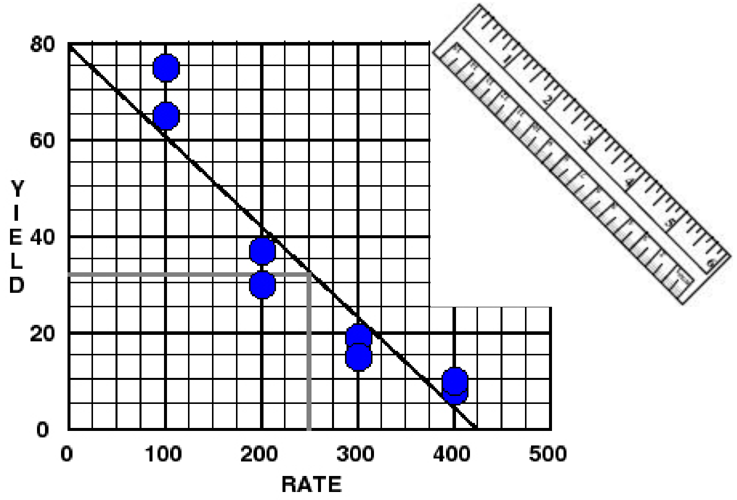
On Transformed Scale: LOF Eliminated and Error More Uniform Across Region

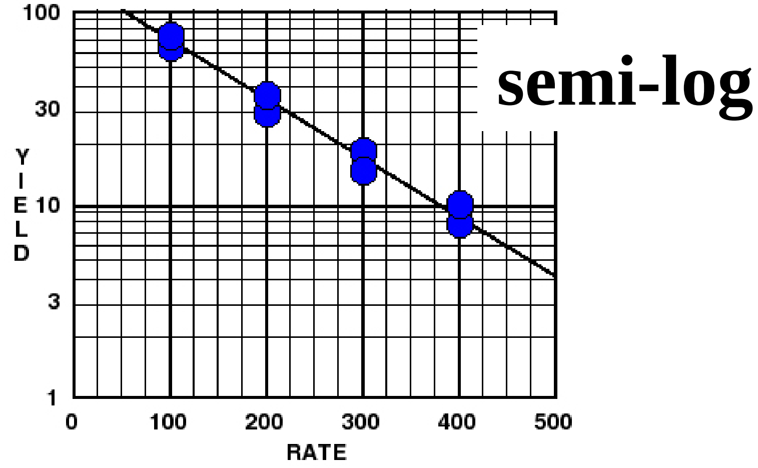


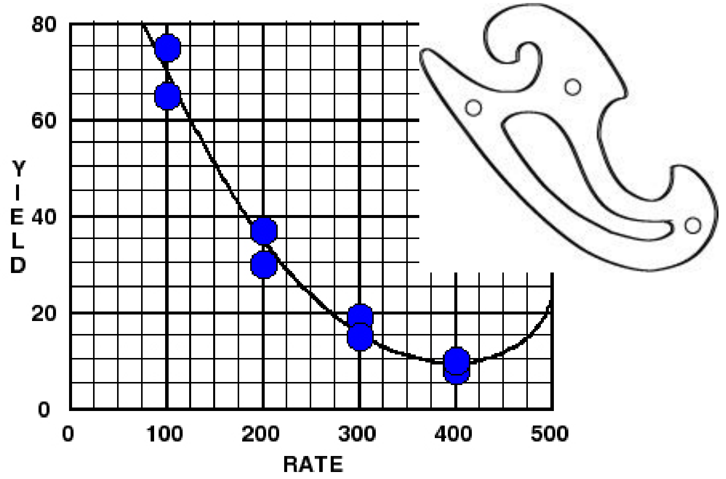


Two Remedies for Lack-of-Fit
Fancier Graph Paper or Fancier Curve





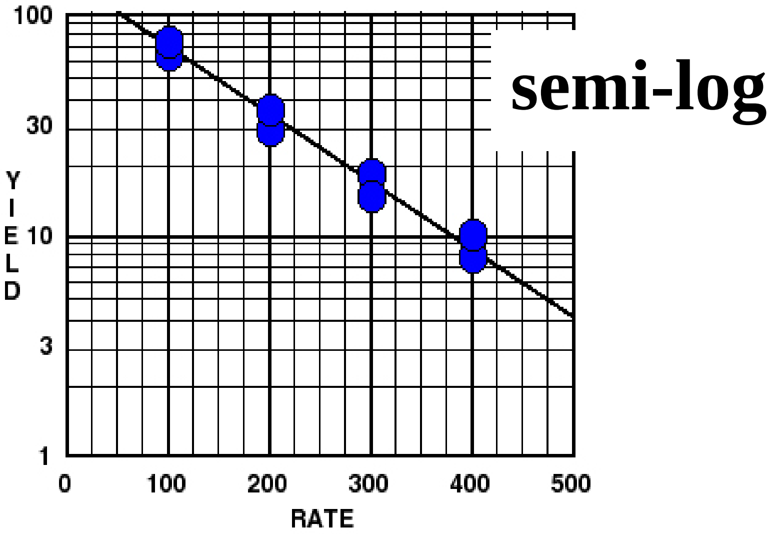




Does not require additional trials.

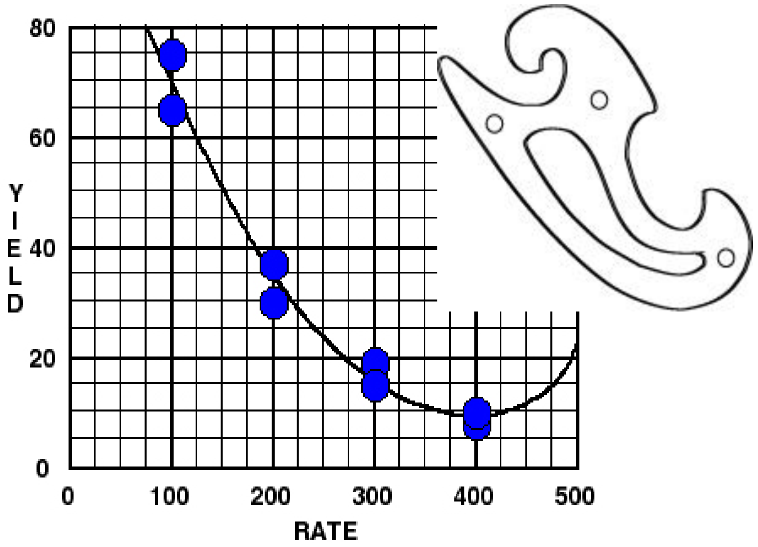
Usually requires additional trials.

Model Predictions are Virtually Same within the Range of the Factor Settings (100 to 400) but can be quite different outside the Range y



At Rate = 250
Predicted Yield is 22

At Rate = 500
Predicted Yield is 4

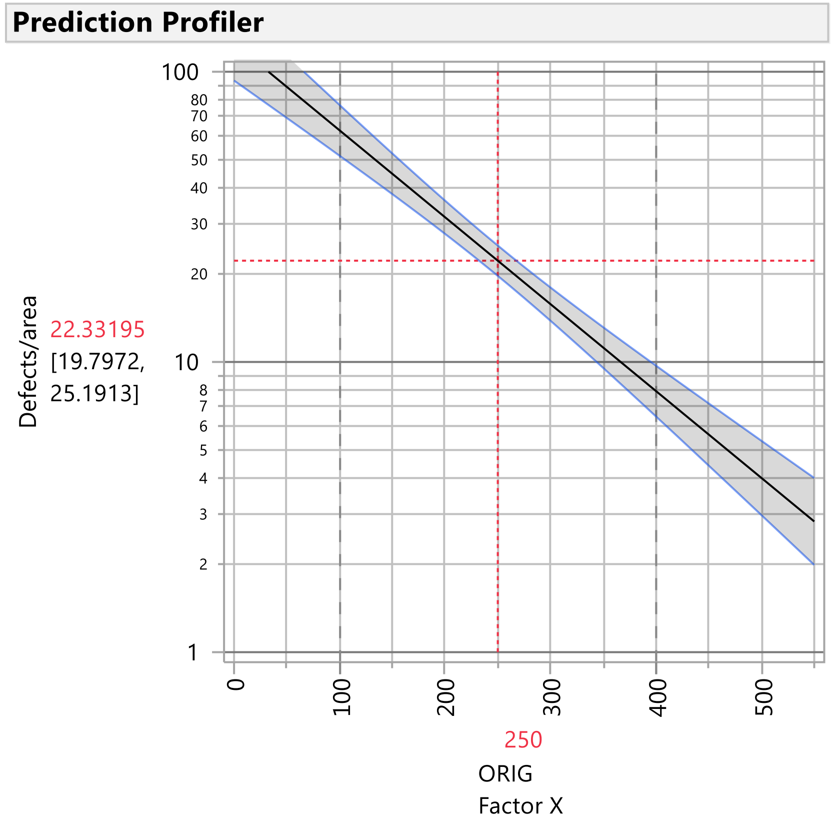


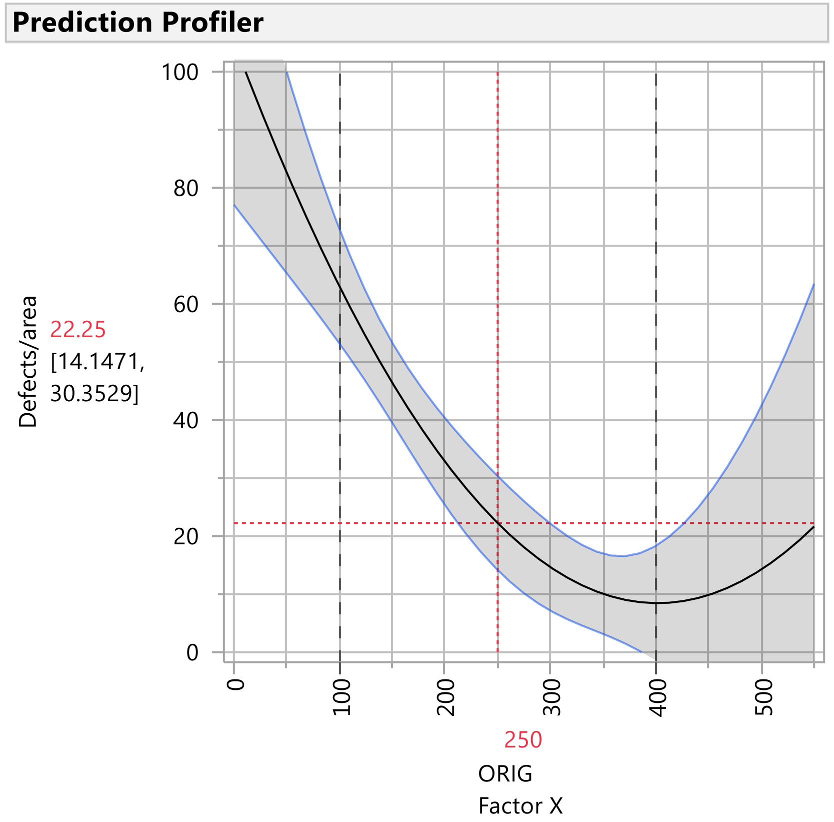
At Rate = 250
Predicted Yield is 22

At Rate = 500
Predicted Yield is 22

Which prediction at 500 is more suspect? Why?

Using Profiler we see that Predictions are Virtually Same *within the Range of the factor Settings (100 to 400)*



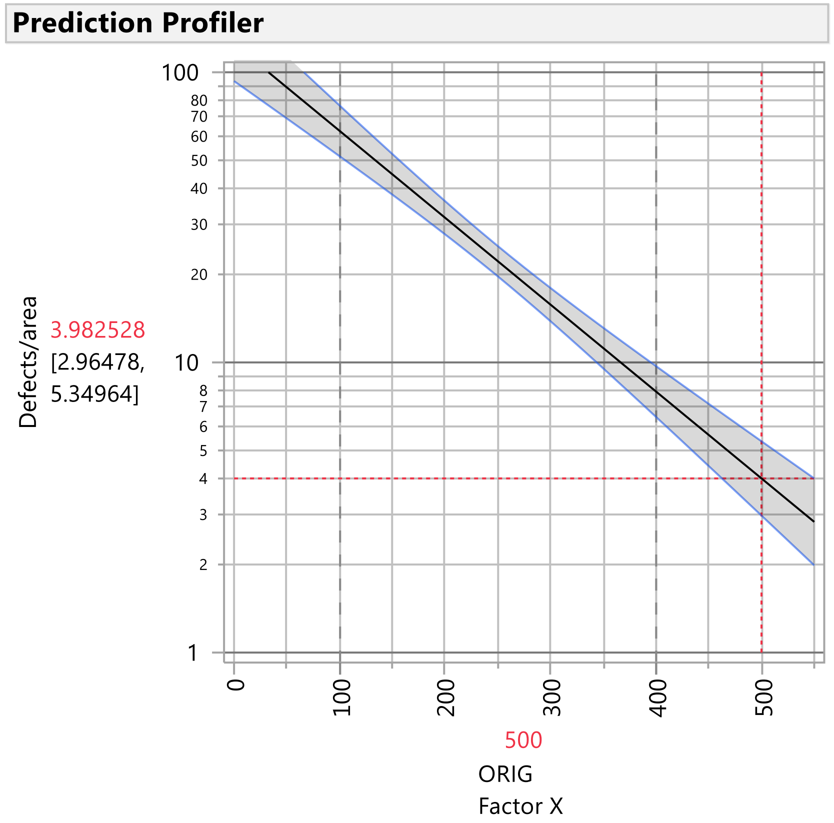


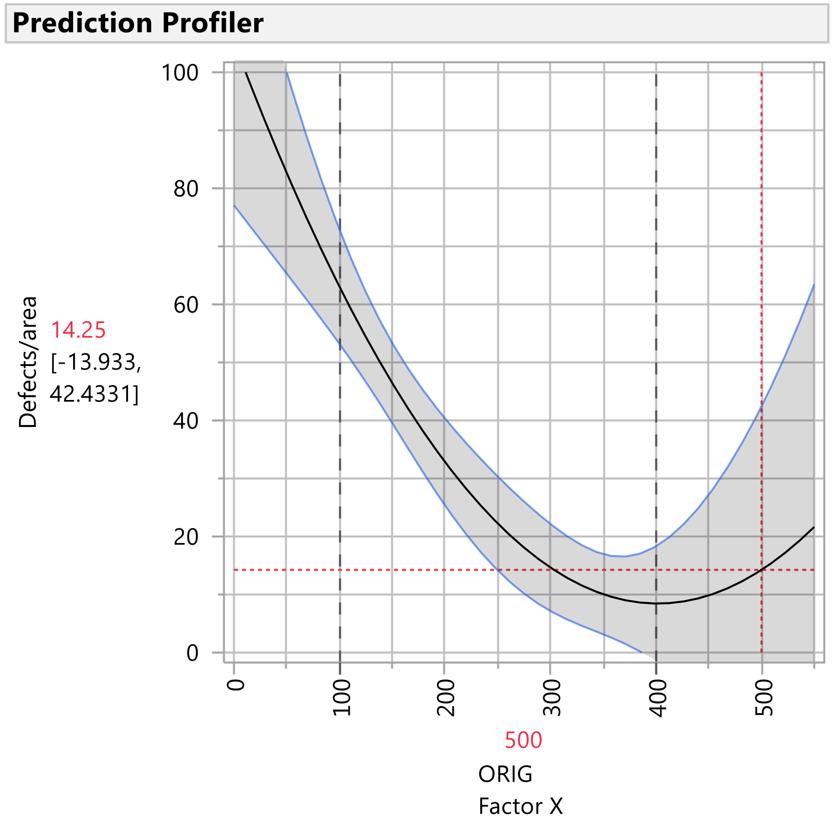
At Factor X = 250
Predicted Yield is 22.33

At Factor X = 250
Predicted Yield is 22.25

Notice the shading of the confidence interval about prediction.

Using Profiler we see that Predictions are quite different *outside the Range of the factor Settings (100 to 400)*





At Factor X = 500
Predicted Yield is 3.98

At Factor X = 500
Predicted Yield is 14.25

Notice the shading of the confidence interval about prediction.

View Extrapolated Predictions Using Profiler in Raw & Transformed Units

3 Columns of Data Used to Fit Same Quadratic Model Form For these 3 Profilers.

$y = b_0 + b_1X + b_2X^2$

Last 2 Models are *Identical*.

	Factor X	Defects/area	Defects/area 2	Log10[Defects/area]
	400 – 100	71 – 7	71 – 7	1.85 – 0.85
* 1	100	71	71	1.85
* 2	100	56	56	1.75
* 3	200	35	35	1.54
* 4	200	28	28	1.45
* 5	300	14	14	1.15
* 6	300	18	18	1.26
* 7	400	9	9	0.95
* 8	400	7	7	0.85

No transformation used
NOTE: y-axis in raw units and on a linear scale

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	67.75	5.415126	12.51	<.0001*
Factor X	-0.182	0.017612	-10.33	0.0001*
(Factor X-250)*(Factor X-250)	0.0006	0.000197	3.05	0.0285*

Log_{10} transformation used
Within Model Dialog,
NOTE: y-axis in raw units but on a Log_{10} scale

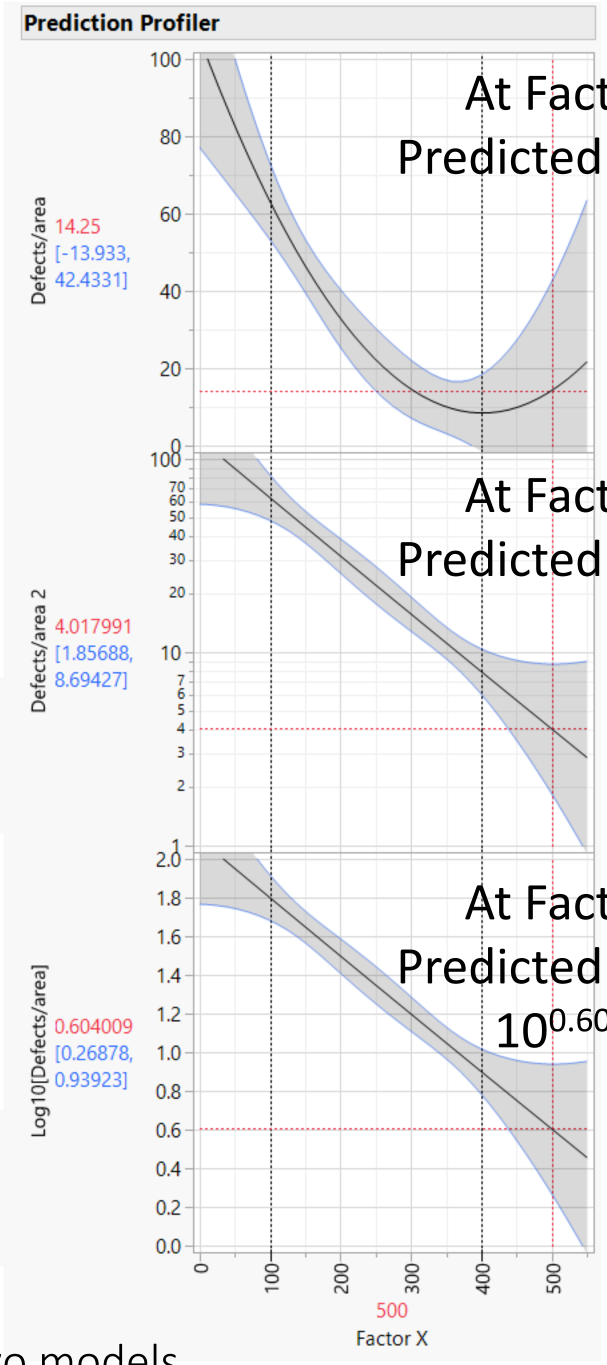
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.8279039	0.14831	32.55	<.0001*
Factor X	-0.006896	0.000482	-14.30	<.0001*
(Factor X-250)*(Factor X-250)	1.7731e-7	5.393e-6	0.03	0.9750

Log_{10} transformation used
In Data Table Column
NOTE: y-axis in Log_{10} units and on a linear scale

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.096732	0.06441	32.55	<.0001*
Factor X	-0.002995	0.000209	-14.30	<.0001*
(Factor X-250)*(Factor X-250)	7.7004e-8	2.342e-6	0.03	0.9750



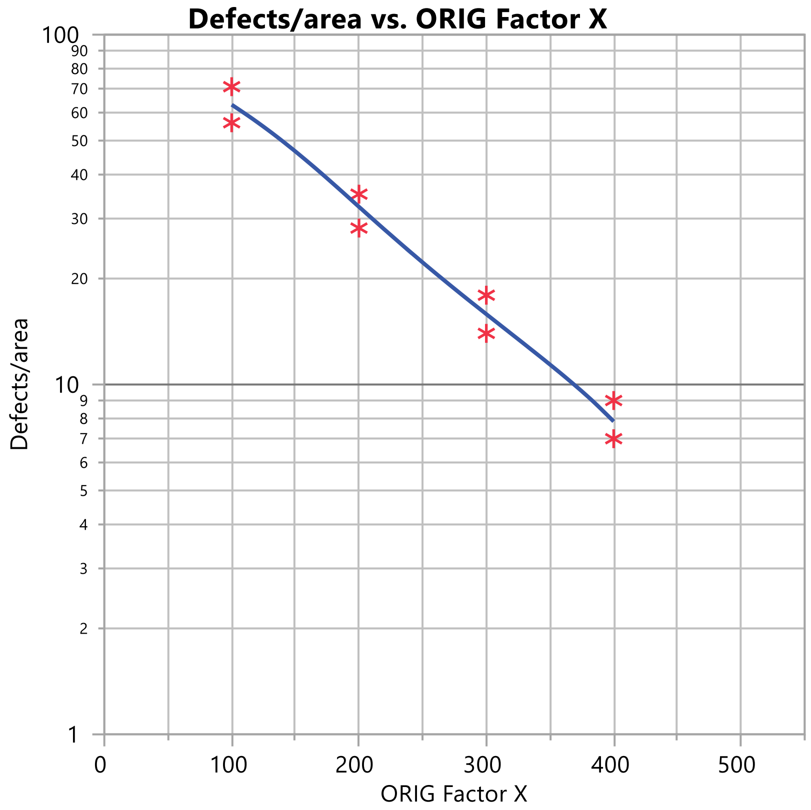
At Factor X = 500
Predicted Yield is 14.25

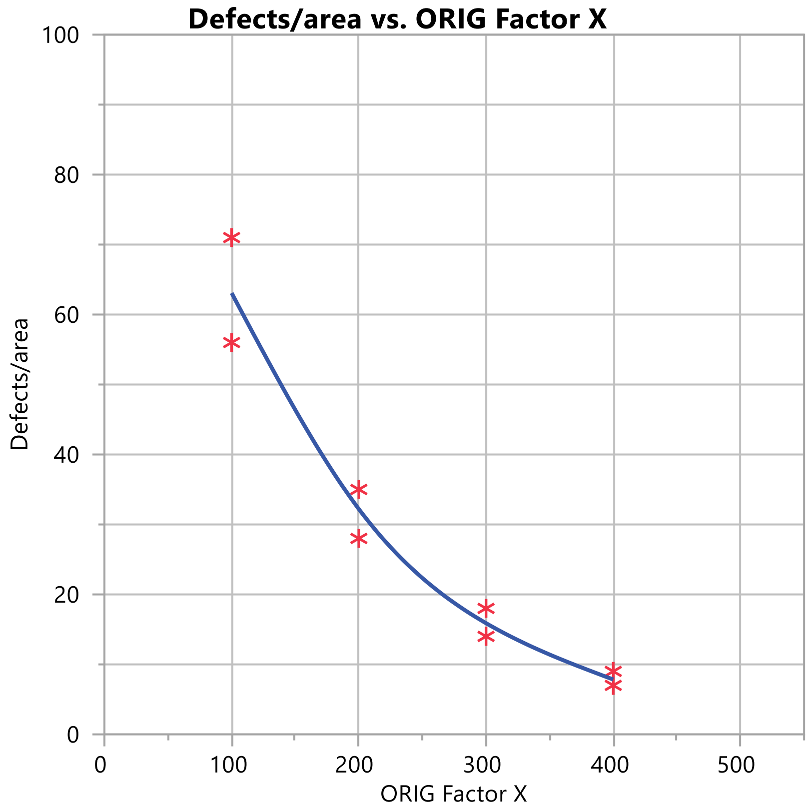
At Factor X = 500
Predicted Yield is 4.018

At Factor X = 500
Predicted Yield is 0.604
$10^{0.604} = 4.018$

NOTE: Typically we would drop the clearly NOT significant squared term in last two models.

Today with JMP we use a “SMOOTHER” in Graph Builder instead of rulers and French curves



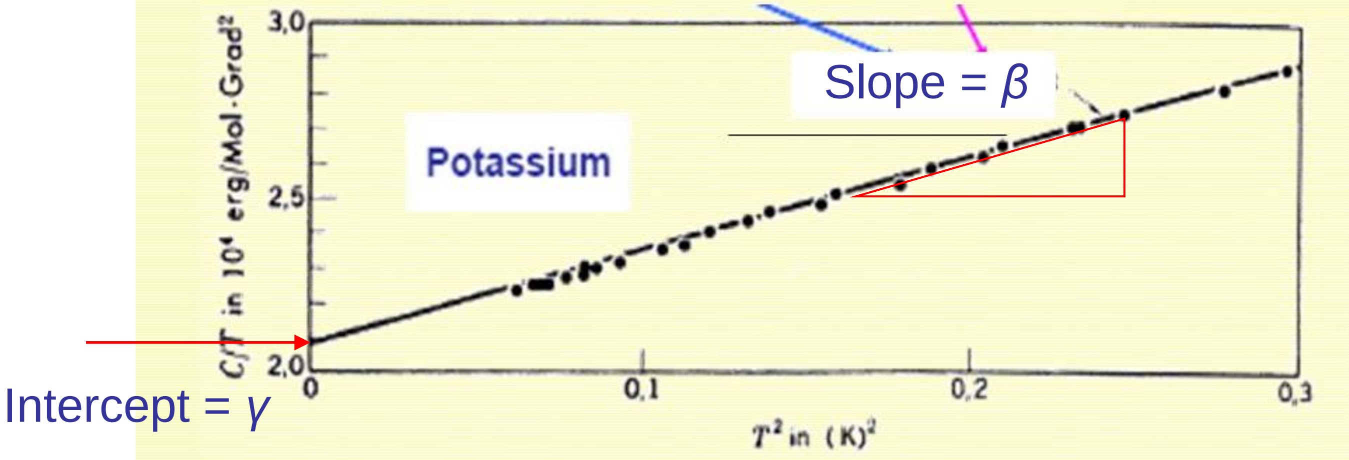


- Can also change the scale of the axes in Graph Builder
- Notice “Smoother” only visible in range of the data

Example of How Rescaling Makes the Analysis Easier

- Total metallic heat capacity at low temperatures $C = \underbrace{\gamma T}_{\text{Electronic Heat capacity}} + \underbrace{\beta T^3}_{\text{Lattice Heat Capacity}}$

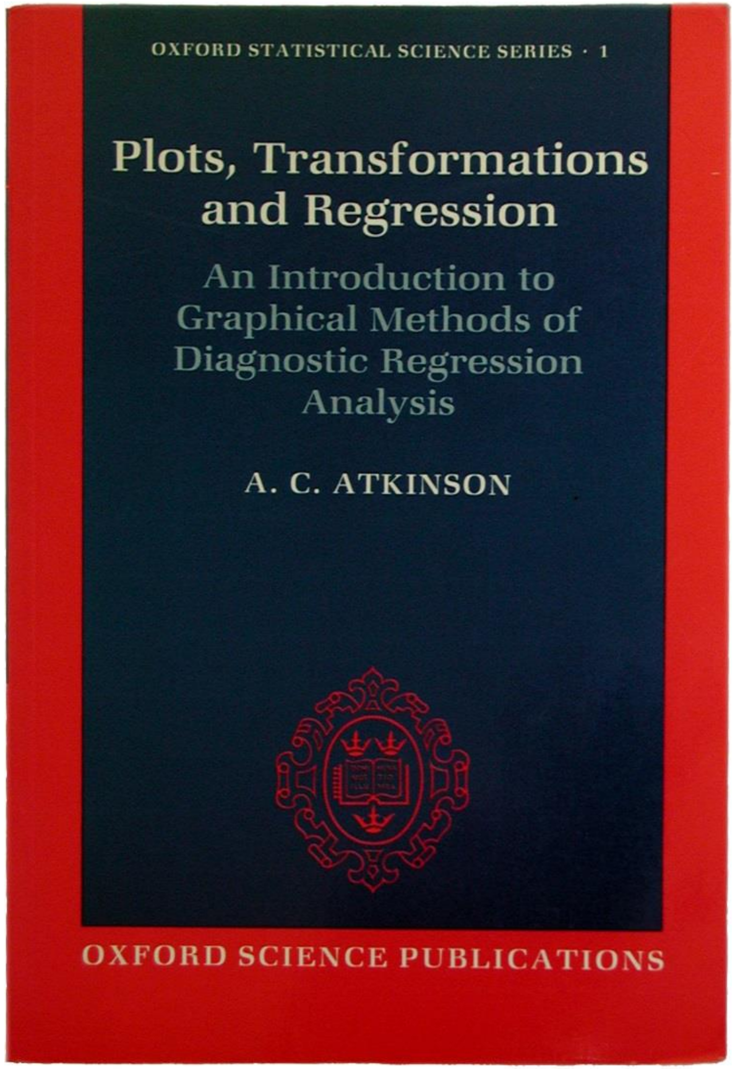
where **γ & β** are constants found plotting **C_v/T** as a function of **T^2**



jmp

Have a *Reason* to Use a Transformation - Try NOT to "Brute Force" Eliminate L-O-F

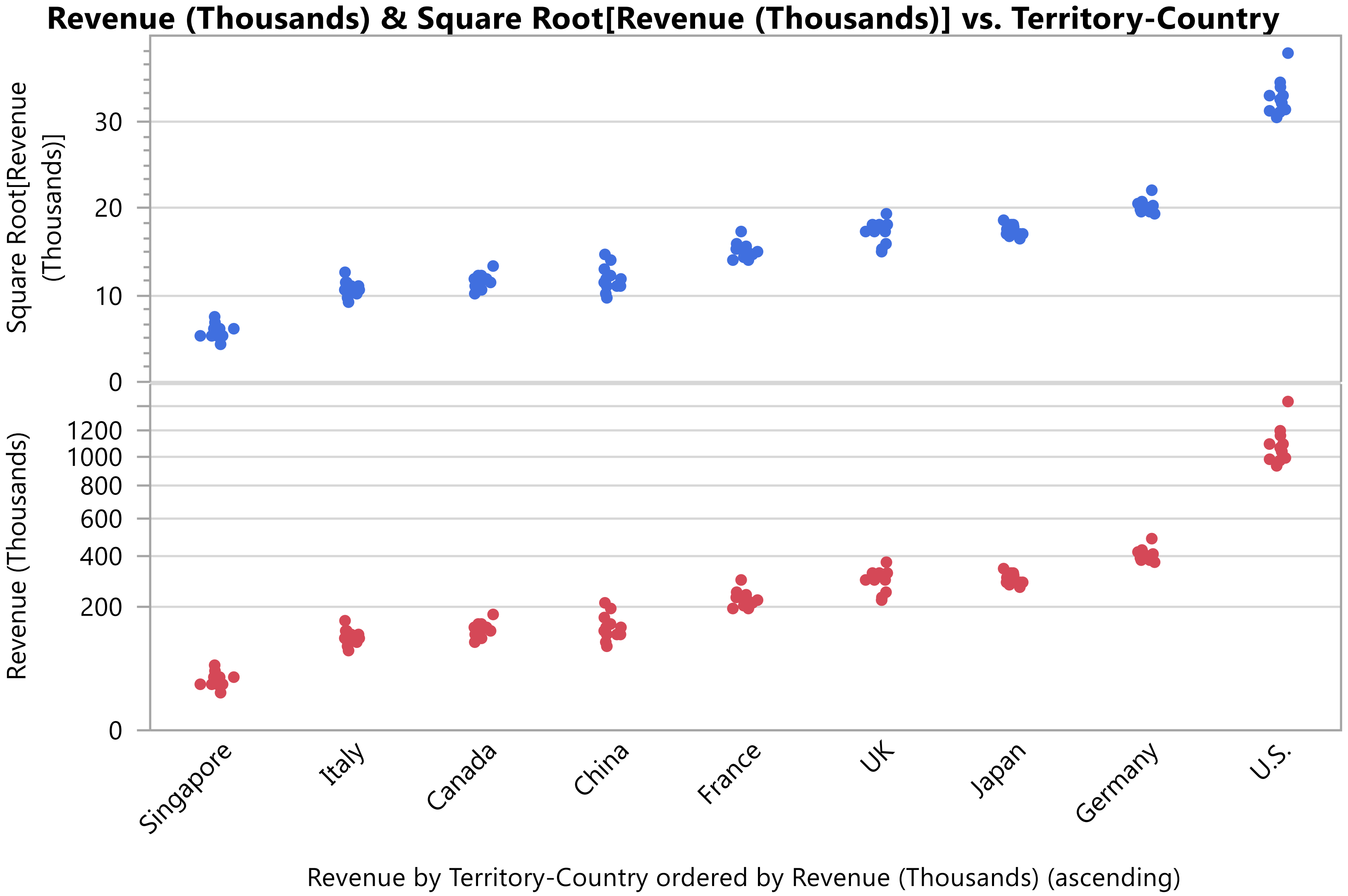
- Check publications in your field to see how others plot the same kind of data. (See previous slide)
- Consult a reference like:
- Consult your local statistical expert.
- **Remember all a transformation does is plot the data on *fancier* graph paper.**



Examples

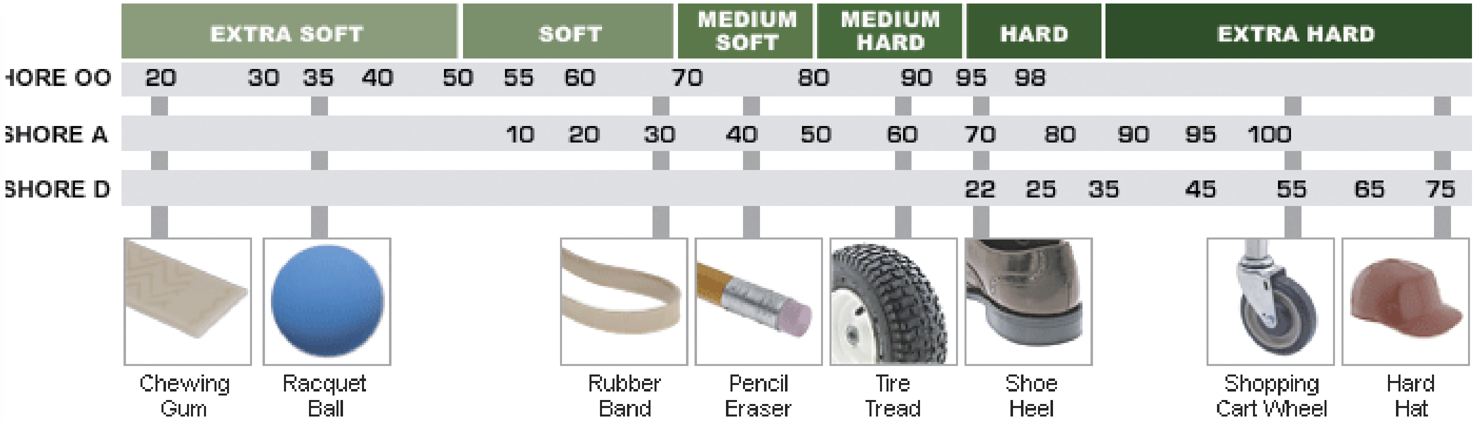
- Using Graph Builder to explore revenue by territory
  - Transform columns in graph
  - Add virtual column to table
  - Transform axes & add grid lines
- Using Box-Cox Transformation to identify appropriate transformations of the response
  - Hardness of plastic – make physical sense
  - Tensile Strength of plastic – eliminate L-O-F
  - Yield of CO_2 capture process
    - Can generate residuals to check how well transformation works
    - Not all Fit Model personalities/capabilities support transformations
      - Generalized Regression – create additional columns
      - Fit Definitive Screening - create additional columns
  - Counts of detectors
  - Briefly - Army example of transformations on the factors

SQRT Transformation on *Data* (Top) and on *Axis* (Bottom)



Want to make an informed business decisions trading off product performance and cost

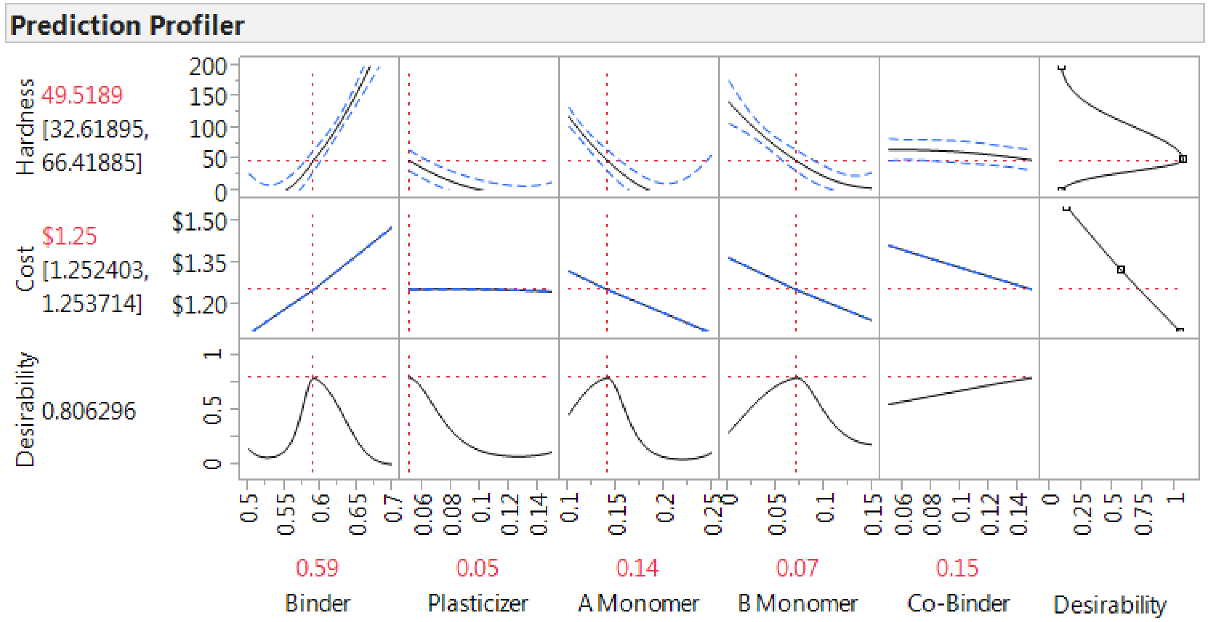
Need to predict hardness and cost of plastic



What formulation yields a Shore A hardness of 50?

What does the formulation cost?

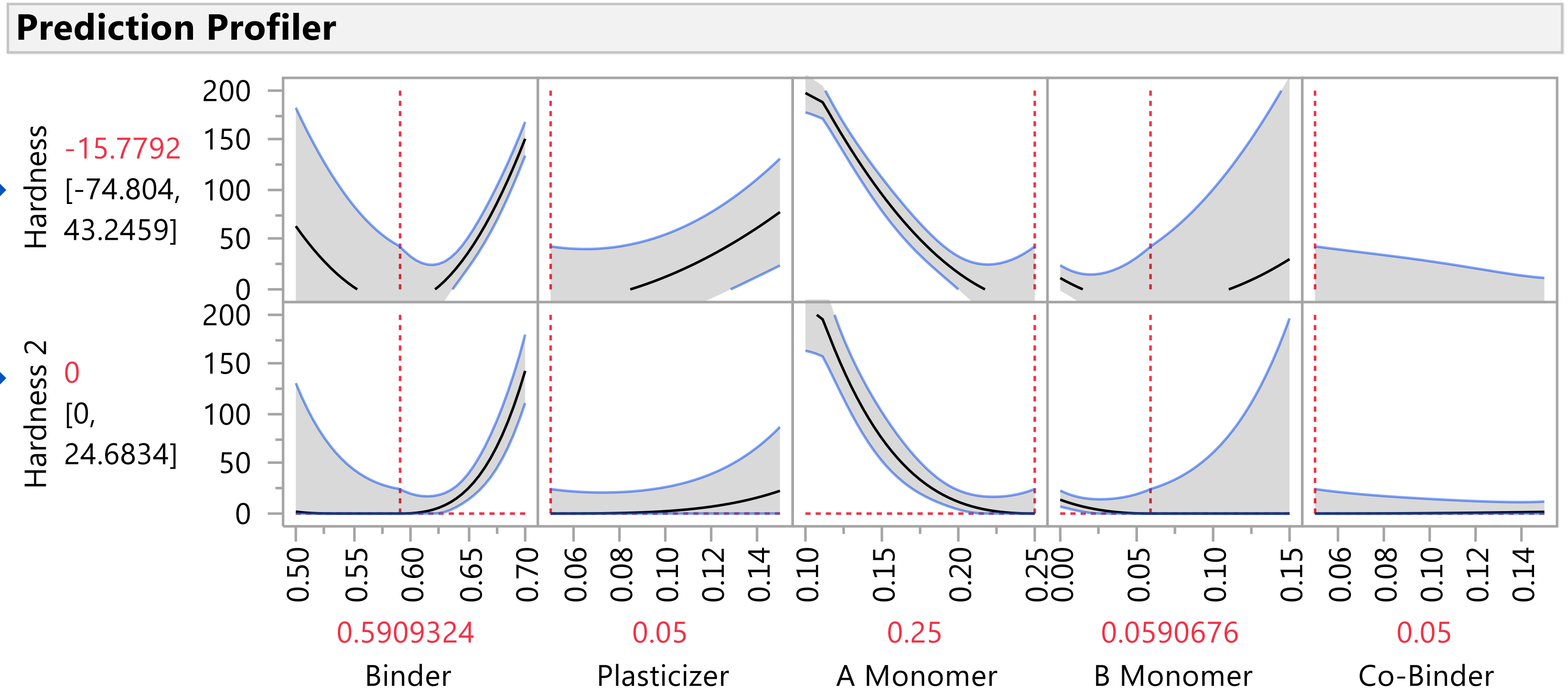
Can I trade-off hardness and cost?



Fitting Hardness of Plastic without (Top) and with SQRT Transformation (Bottom)

Potentially embarrassing predictions:

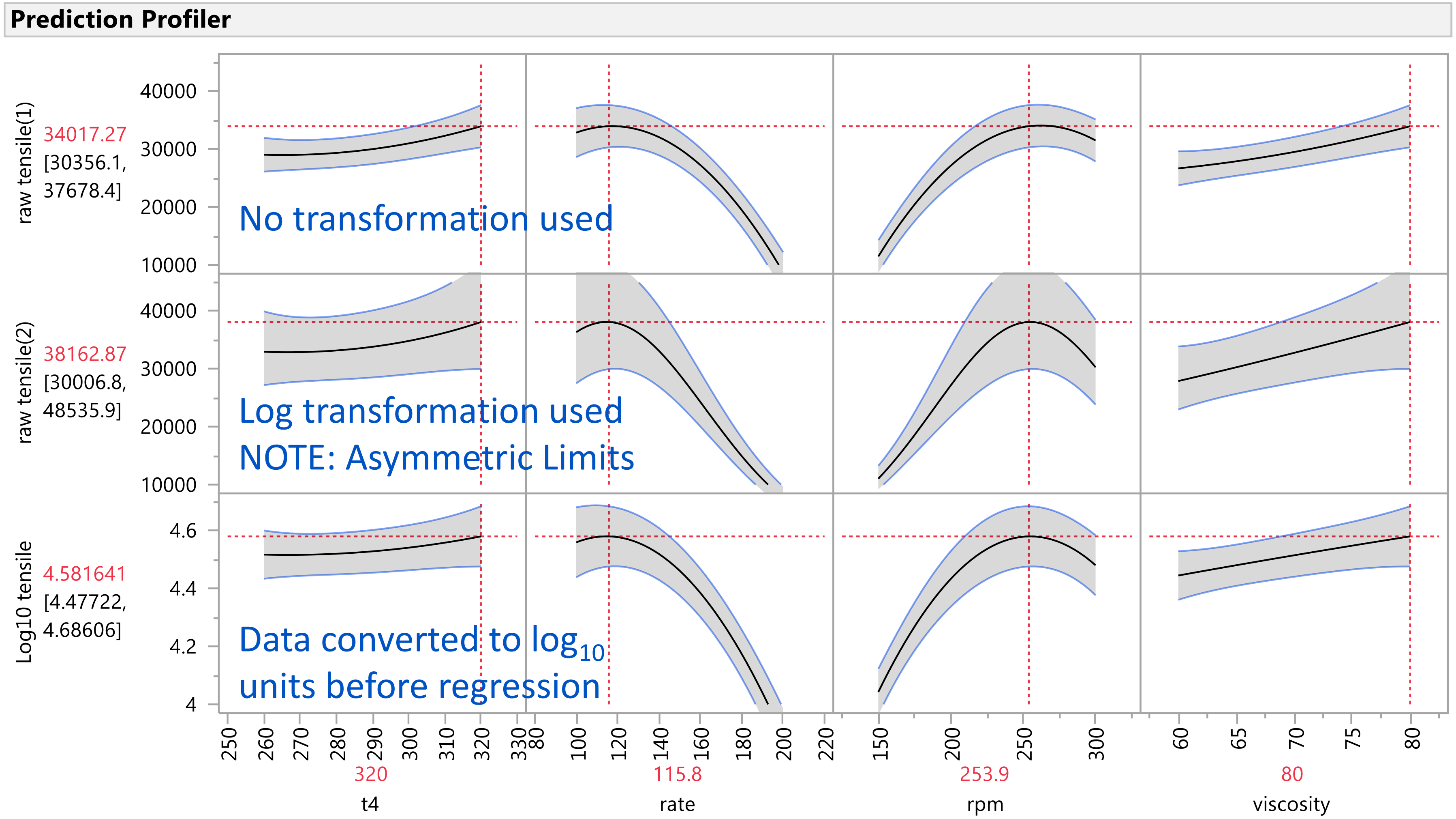
NEGATIVE Value?
NEGATIVE Low Limit?



POSITIVE Value!
ZERO Low Limit!

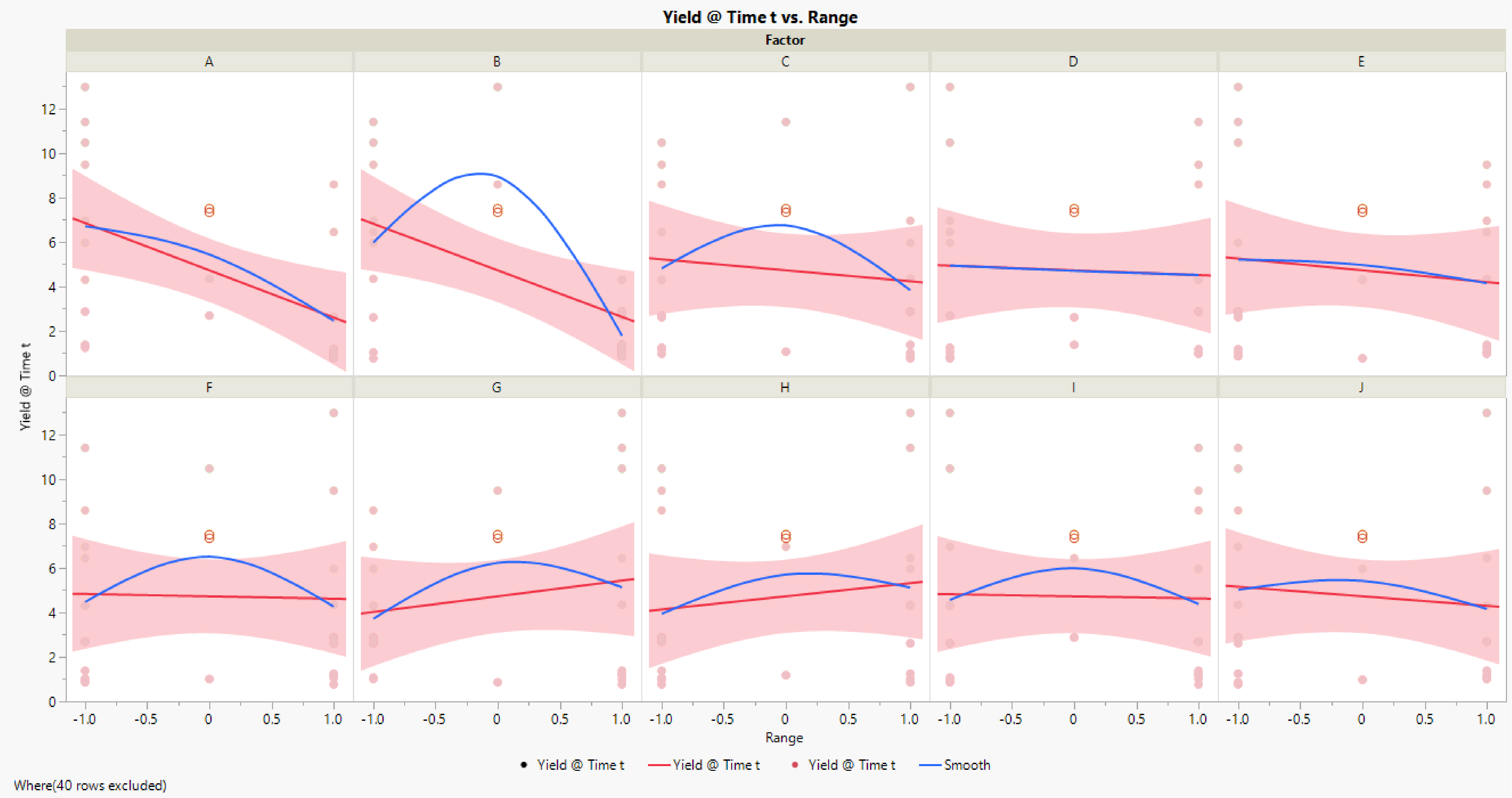
On Transformed Scale (Bottom),
Predictions Make Physical Sense

Use Profiler to View Plots in Transformed & Lab Units in JMP - Three Separate Columns of Data Used for These Plots

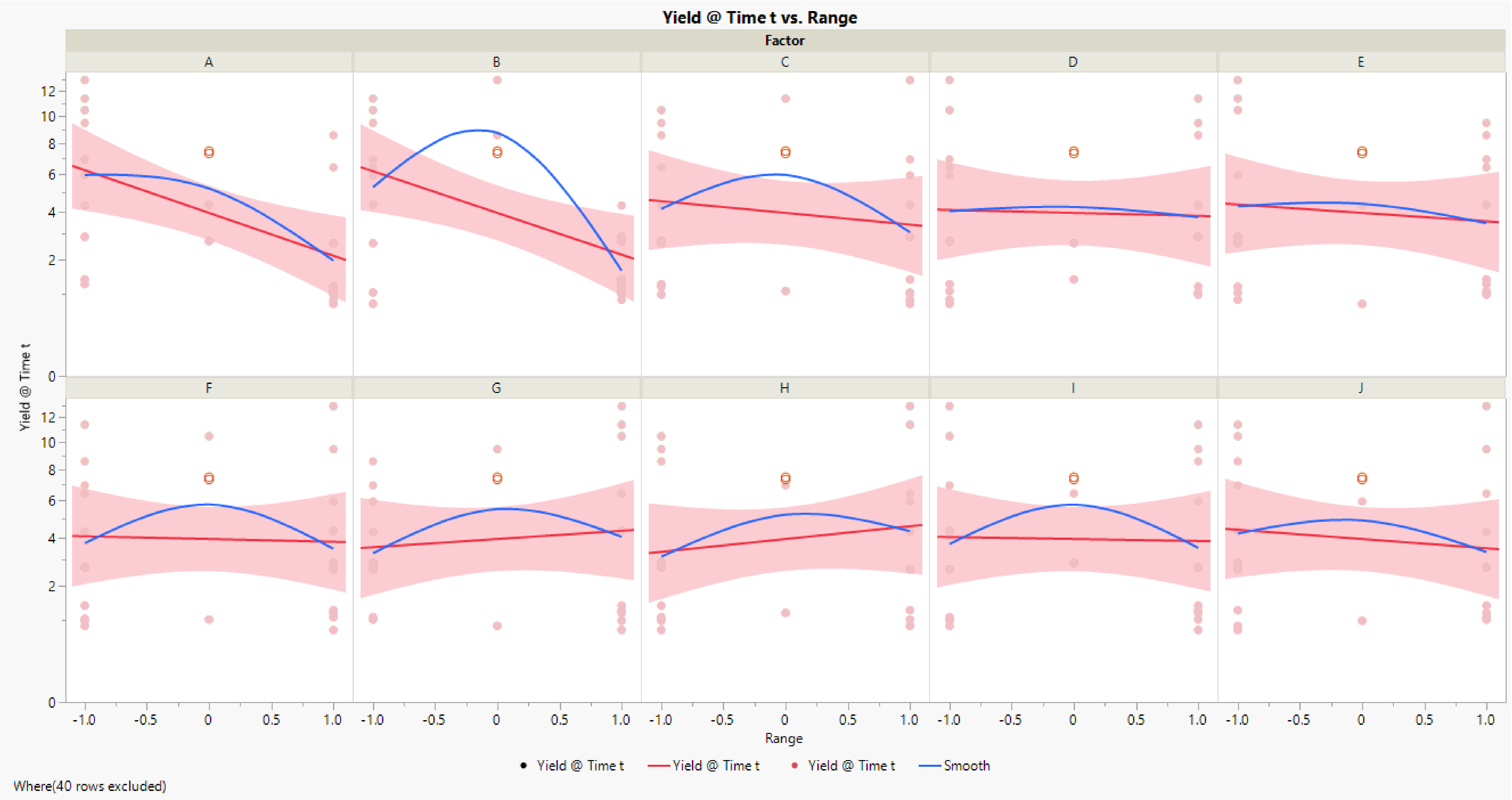


$10^{4.581641} = 38{,}162.87$

Y vs X plots of data for each X



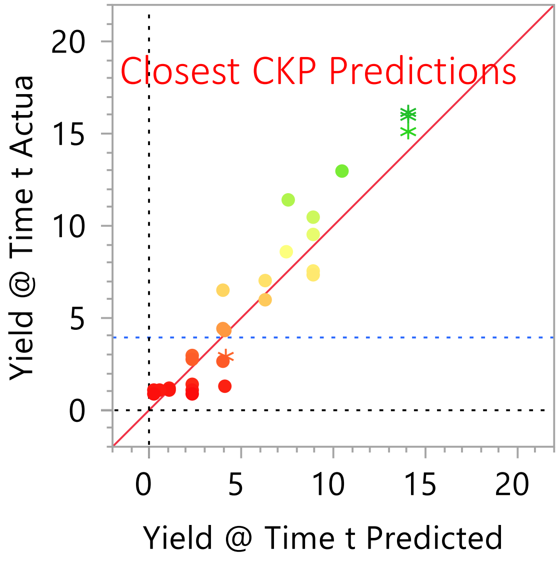
Y vs X plots of data for each X



Transformations SQRT, Log10, & NONE

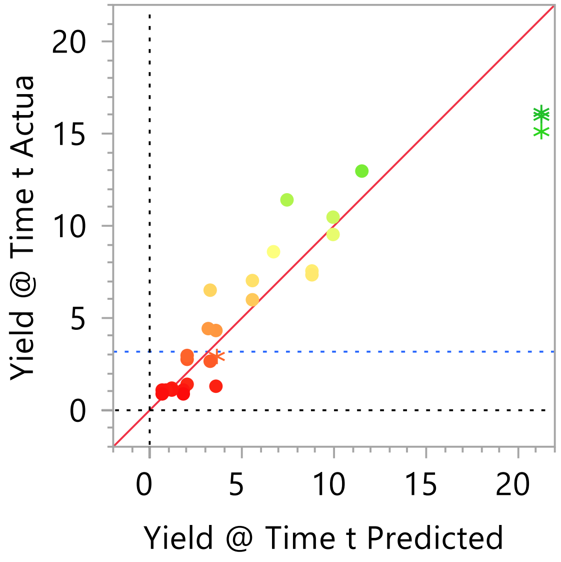
Green Asterisks* are Checkpoints NOT used in fitting data.

Actual by Predicted Plot



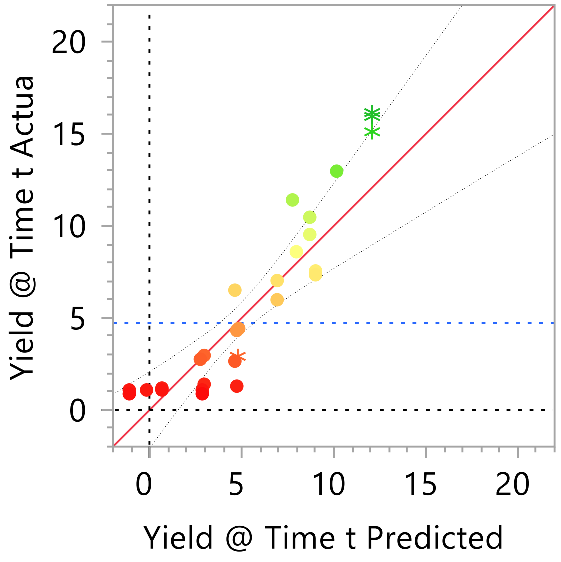
P<.0001 RSq=0.83
RMSE=0.4163

Actual by Predicted Plot



P<.0001 RSq=0.82
RMSE=0.4509

Actual by Predicted Plot



P<.0001 RSq=0.79
RMSE=1.9387

Summary of Fit

RSquare	0.825967
RSquare Adj	0.789328
Root Mean Square Error	0.416337
Mean of Response	1.983747
Observations (or Sum Wgts	24

Summary of Fit

RSquare	0.823029
RSquare Adj	0.785772
Root Mean Square Error	0.450888
Mean of Response	1.151951
Observations (or Sum Wgts	24

Summary of Fit

RSquare	0.789957
RSquare Adj	0.745738
Root Mean Square Error	1.938688
Mean of Response	4.72375
Observations (or Sum Wgts	24

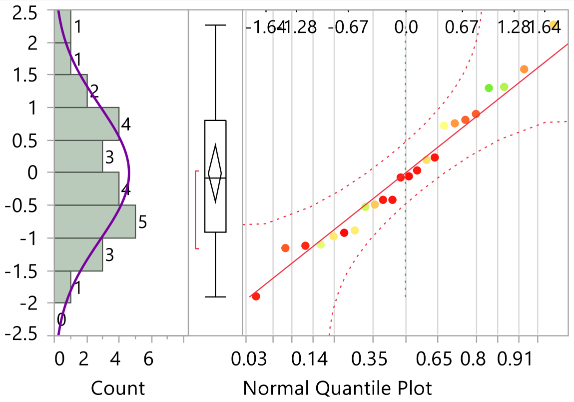
Plots of residuals for Sqrt, Log, and No Transformations

Model fit was reduced quadratic in A, B & C:

Yield = Intercept + A + B + C + B*B + A*B + B*C

Distributions

Studentized Resid Sqrt(Yield @ Time t) 2



Normal(-0.0045,1.03596)

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	-0.004478	-0.441926	0.4329688
Dispersio	σ	1.0359592	0.8051616	1.4532028

-2log(Likelihood) = 68.8047829349136

Goodness-of-Fit Test

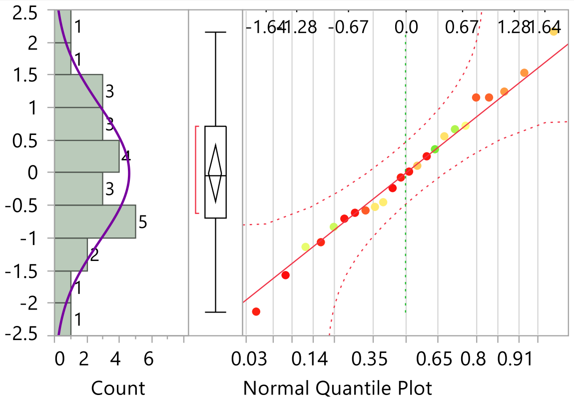
Shapiro-Wilk W Test

W	Prob<W
0.972241	0.7224

Far from LOF

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Studentized Resid Log(Yield @ Time t)



Normal(-0.008,1.03586)

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	-0.007981	-0.445387	0.4294258
Dispersio	σ	1.035863	0.8050868	1.4530679

-2log(Likelihood) = 68.8003267780461

Goodness-of-Fit Test

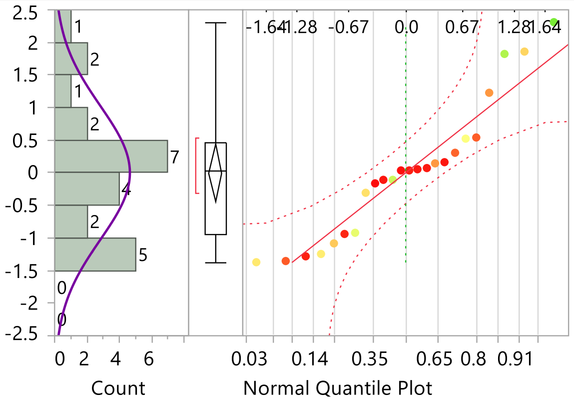
Shapiro-Wilk W Test

W	Prob<W
0.992406	0.9994

Far from LOF

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Studentized Resid Yield @ Time t



Normal(-0.0003,1.0284)

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	-0.000276	-0.434534	0.4339807
Dispersio	σ	1.0284046	0.79929	1.4426054

-2log(Likelihood) = 68.4534641248215

Goodness-of-Fit Test

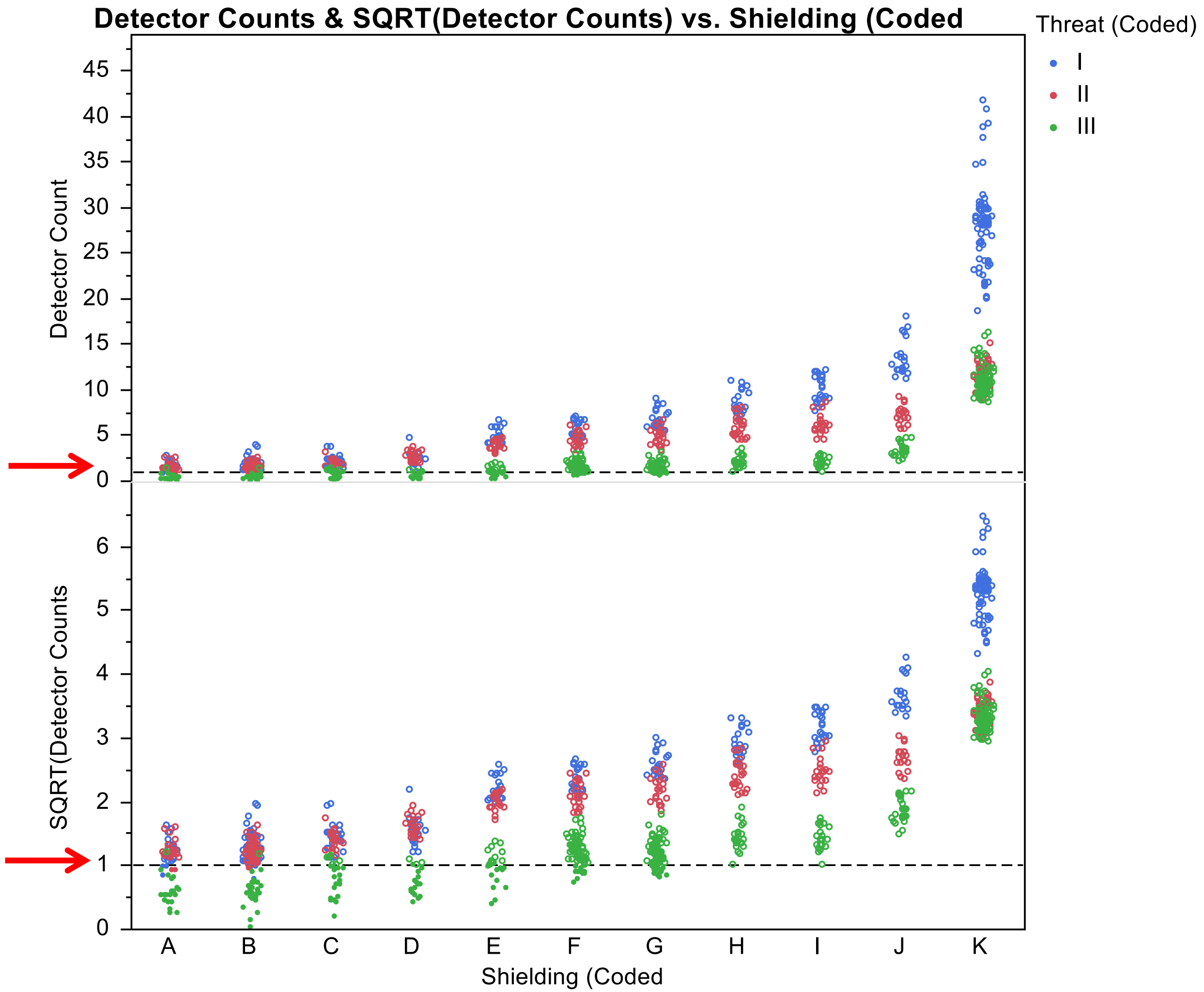
Shapiro-Wilk W Test

W	Prob<W
0.918997	0.0555

Very nearly LOF

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Detector Counts and SQRT (Detector Counts) vs. Shielding (Ordered by Attenuation) – 528 Trials

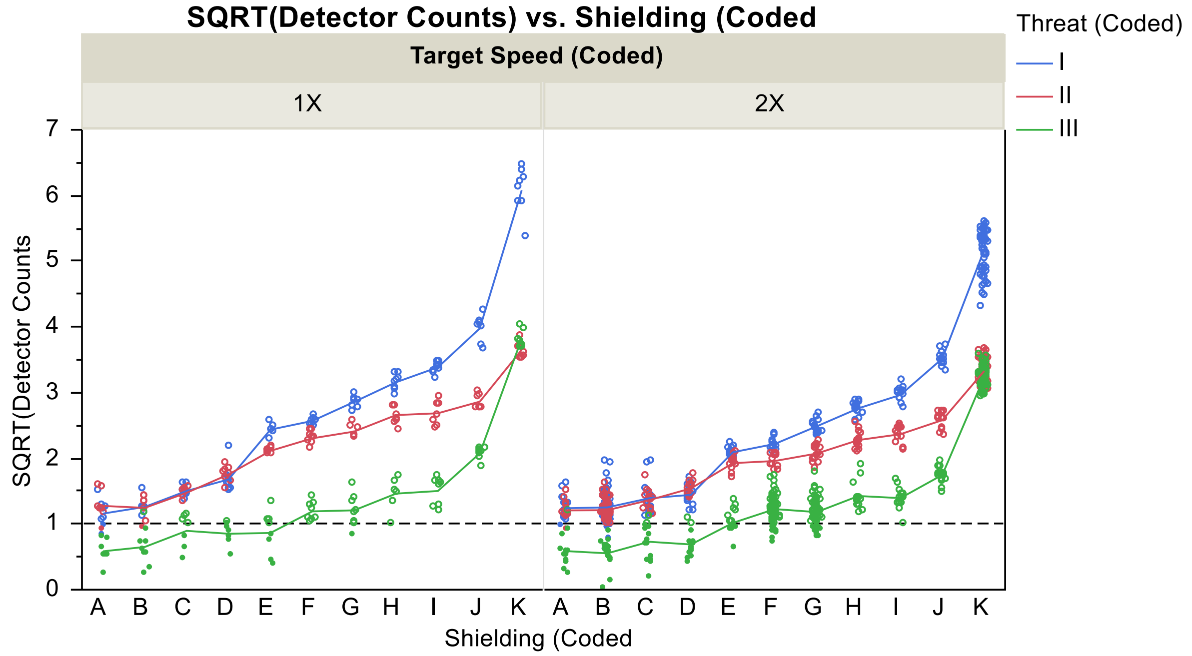


Threshold for Alarm is 1 on either scale.

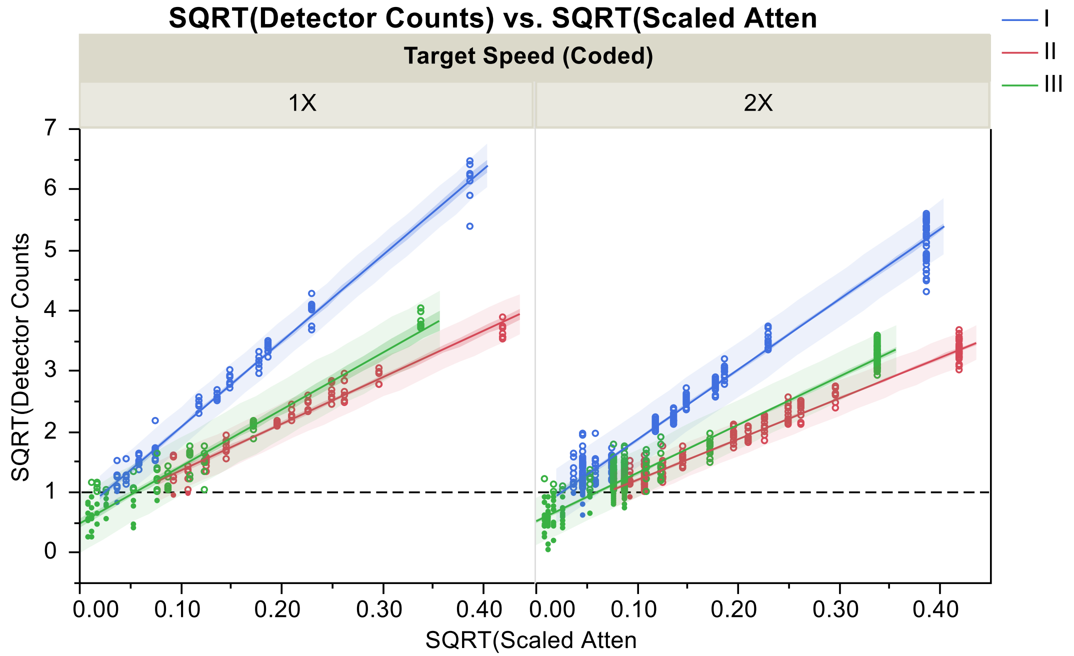
Spread of detector count data is more uniform when plotted on a square-root scale.

SQRT(Detector Counts) vs. Shielding (Ordered by Attenuation) by Target Speed

A reduction in detector counts seen at higher speed.

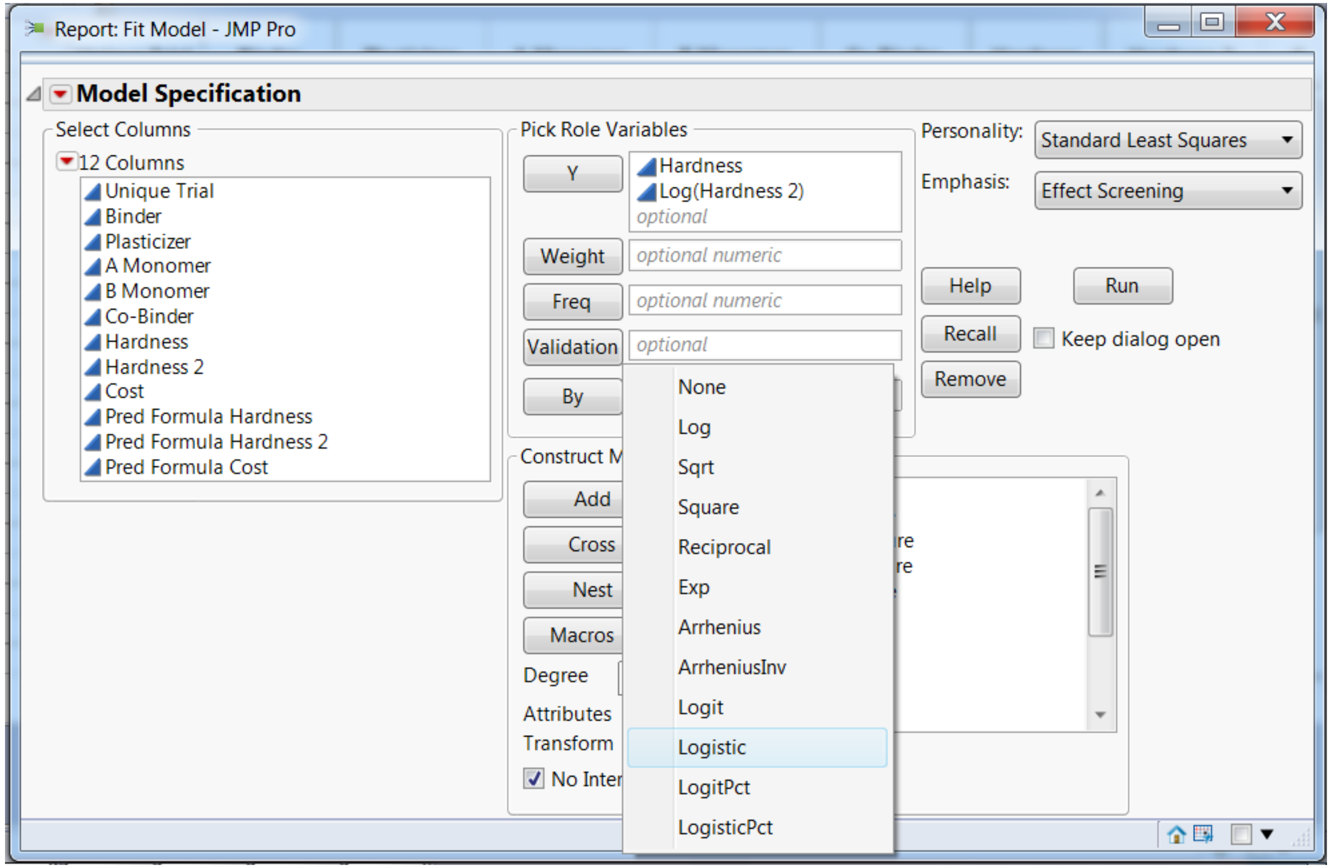


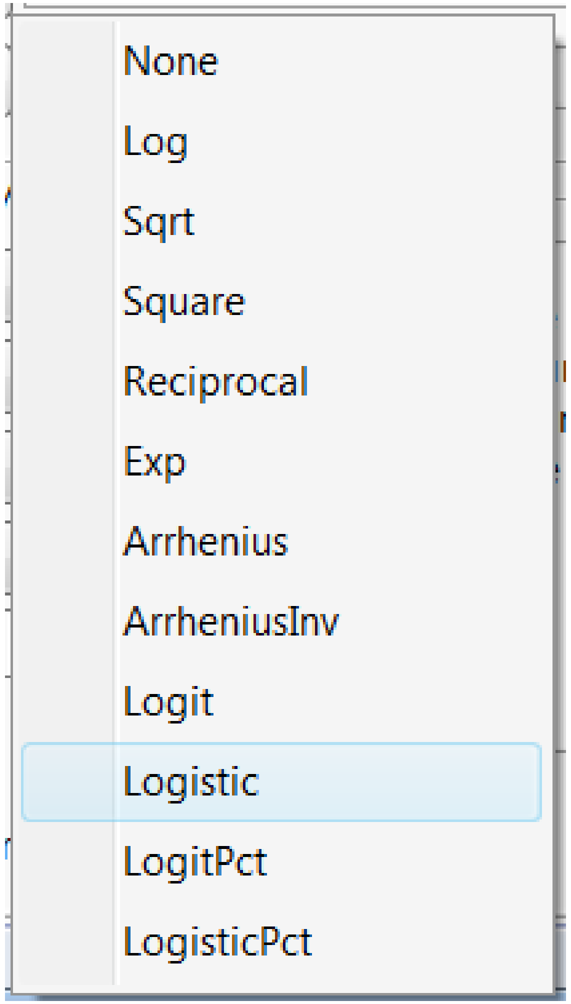
SQRT(Detector Counts) vs. SQRT (Scaled Attenuation) by Target Speed



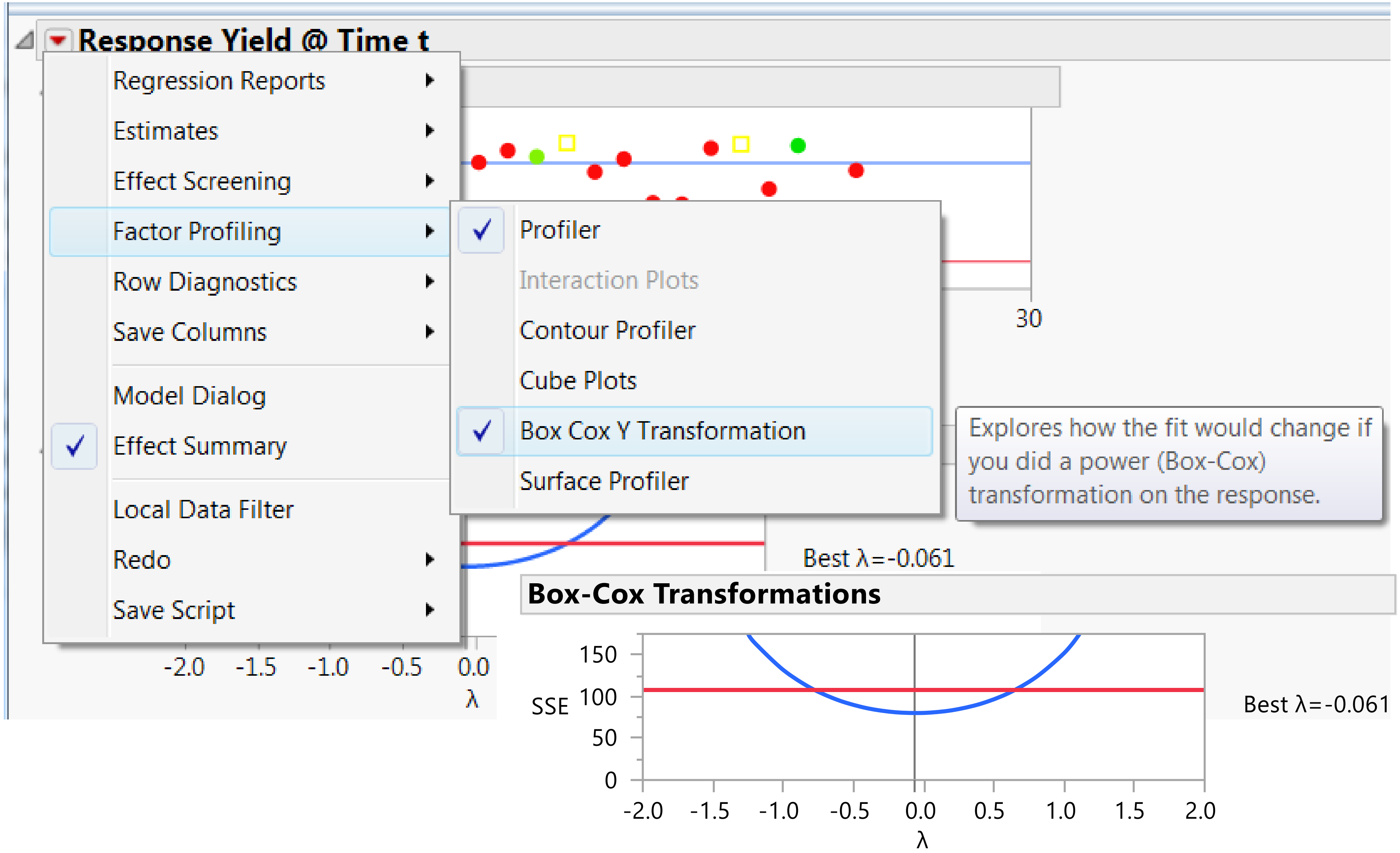
Linear relationship with uniform variance seen between SQRT(Detector Counts) and SQRT(Scaled Attenuation)

Standard Transformations in JMP are Applicable to both Response (Y) & Control (X) Variables





Box-Cox Transformation in JMP



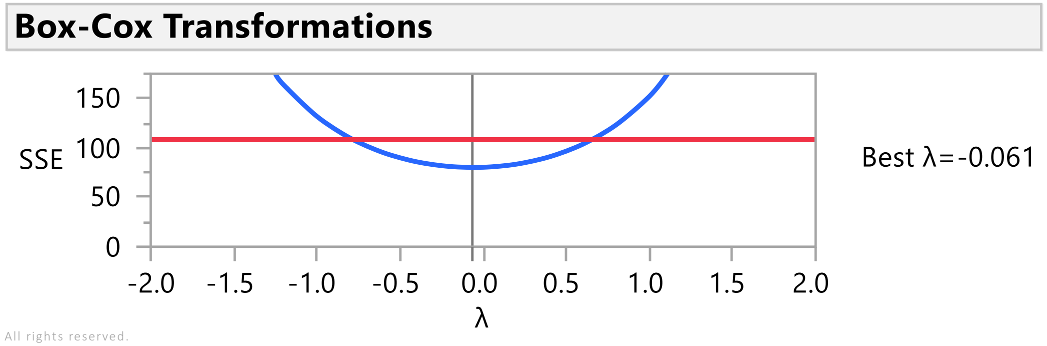
Box-Cox Transformation For Data Bounded on 1-Side

General form: $Y \propto y^{\lambda}$ (a power transformation)

λ	Trans.	λ	Trans.
2	Square	Limit ⇨ 0	Log
1	NONE	-1	Inverse
0.5	Square-Root	-2	Inverse-Square

When Box-Cox Y Transformation is selected in JMP, then a plot of λ versus sum of the squares error (SSE) is created, with the λ associated with the minimum SSE being the "best" value

Use the "best" λ value as a guide as to which "natural" power might be a good choice. If λ = -0.061, i.e. close to zero, then Log transformation is a good choice, if λ = 0.43, i.e. close to 0.5, then Square-Root transformation is a good choice.



Comparison of 10-term Quadratic and 4-term Linear Models

$$\log_{10}(y) = a_0 + a_1x_1 + a_2x_2 + a_3x_3$$
$$+ a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3$$
$$+ a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2$$

constant + linear

+ 2-way interactions

+ curvature terms

The quadratic model can support many shapes – including; mountain, valley, ridge, saddle and plane.

$$\log_{10}(y) = A_0 + A_1X_1 + A_2X_2 + A_3X_3$$

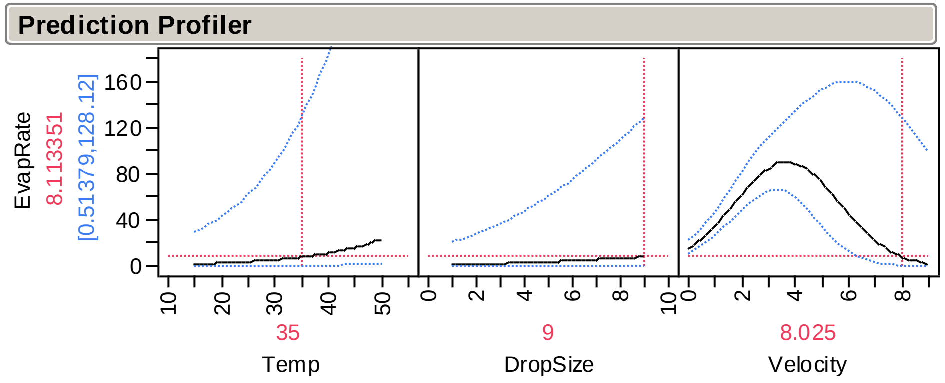
$$\text{and } X_1 = (x_1)^{-1},\ X_2 = (x_2)^{1/2},\ X_3 = (x_3)^{1/3}$$

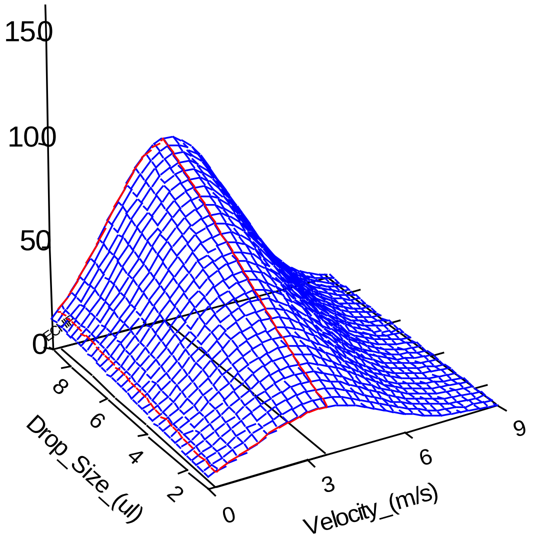
constant + linear terms

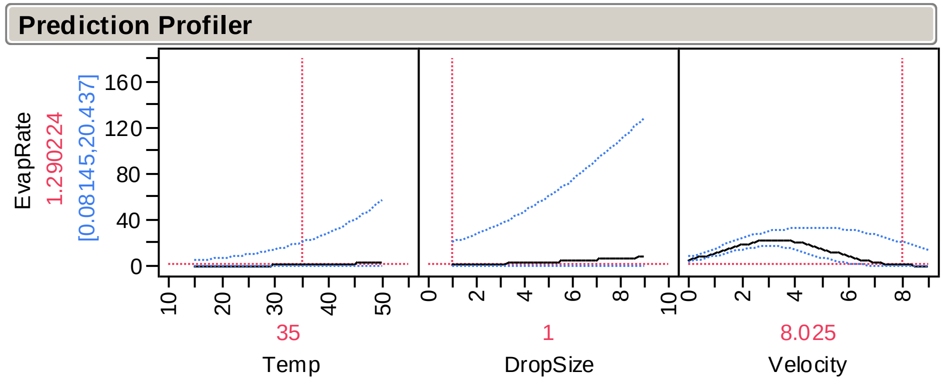
sample exponents used to "linearize" model

The linear model can only support a plane.

Extrapolation with Empirical Model Shown with JMP Prediction Profiler Plots



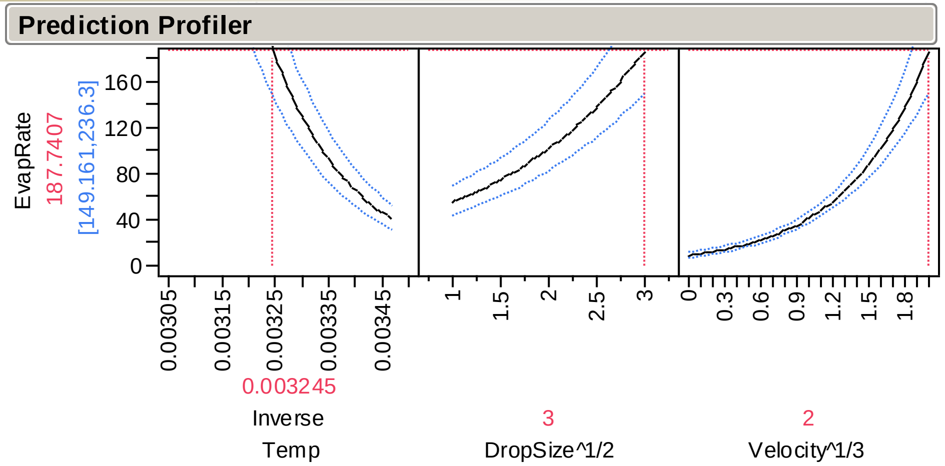


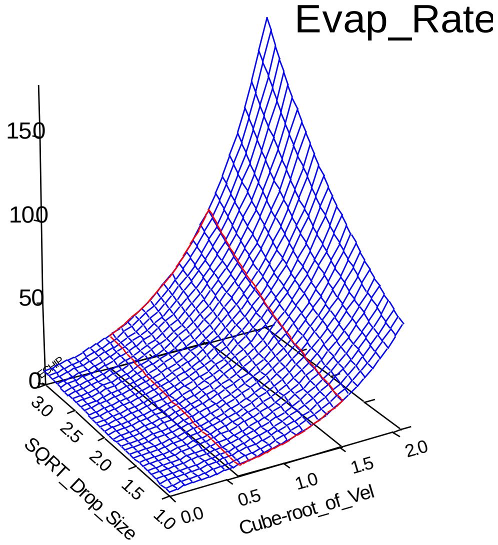


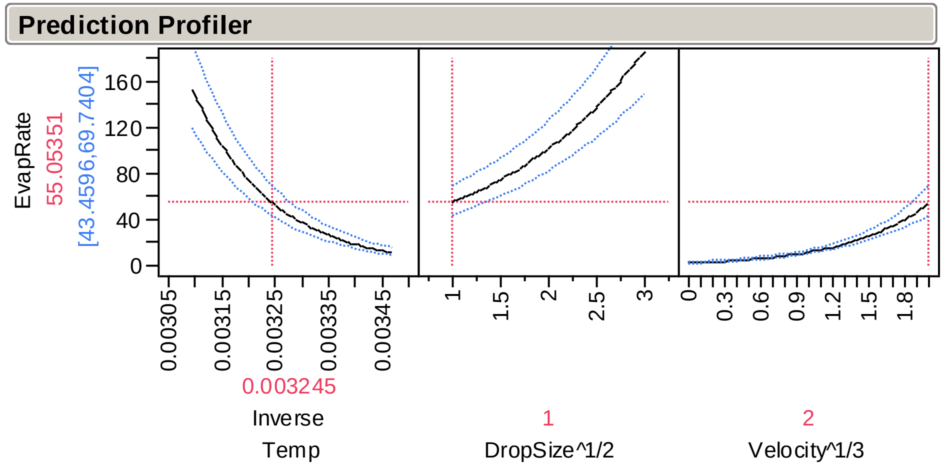
All 19 trials fit using a 10-term quadratic model

Predicted Evap_Rate
At 8 m/s = 1.3 (0.1, 17.1)

Extrapolation with Physics-Based Model Shown with JMP Prediction Profiler Plots





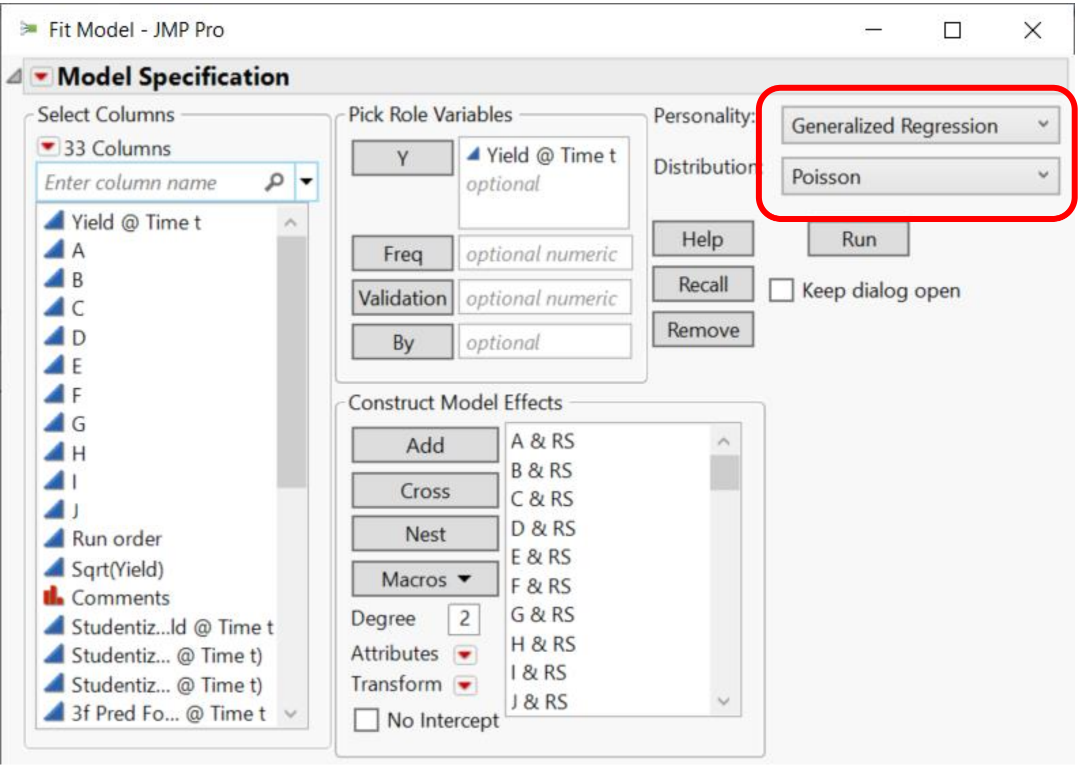


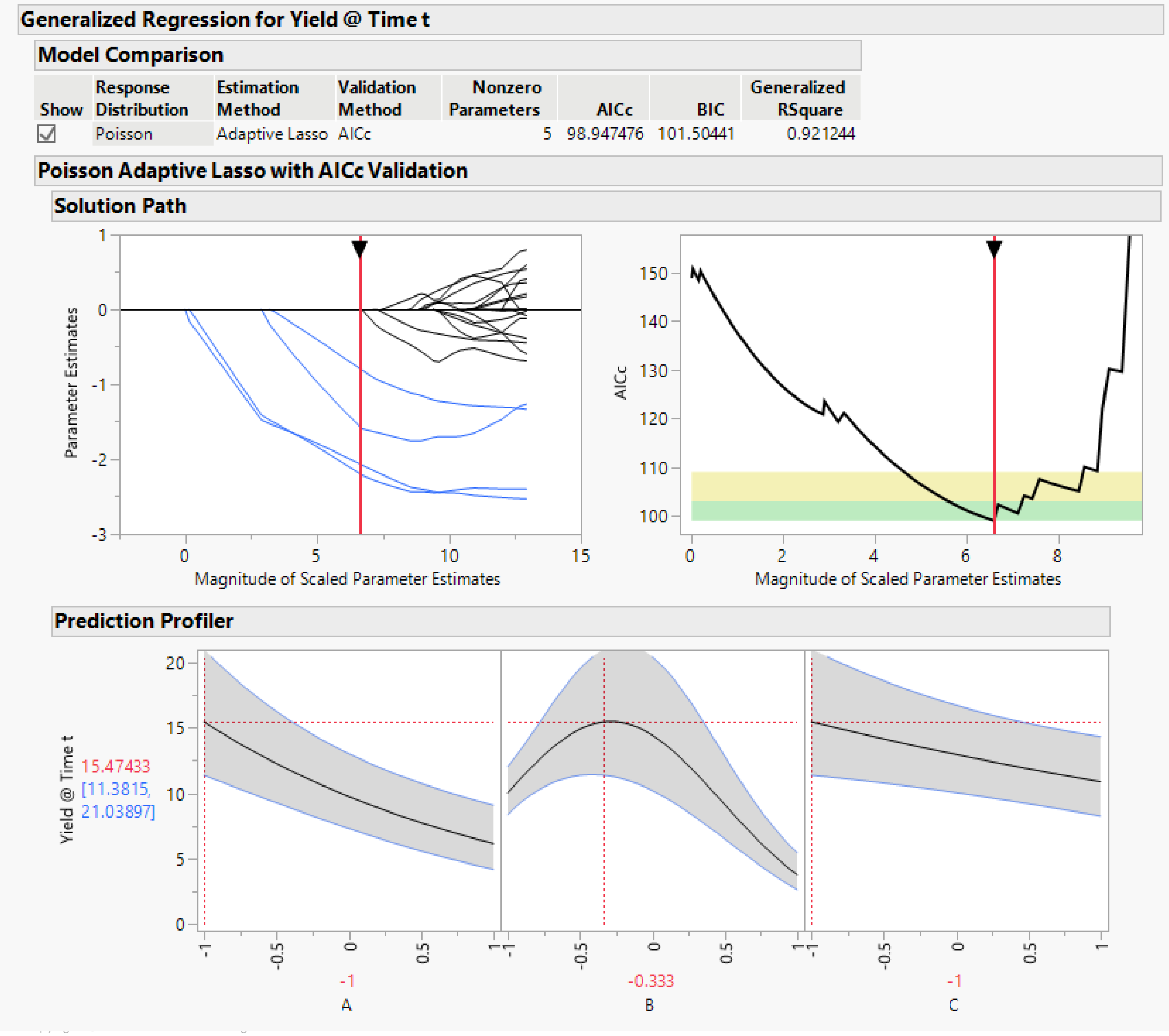
All 19 trials fit using a physics based 4-term linear model

Predicted Evap_Rate
At 8 m/s = 55.1 (36.7, 82.9)

Today with JMP Pro rather than use a transformation, one can often use the appropriate distribution of the variance for the data to fit a model.

Analysis of CO_2_Process data with Poisson Distribution instead of using SQRT Transformation to try to force the data to be normally distributed with a constant variance.





Remember All a Transformation Does is Plot the data on Fancier Graph Paper

- No new data has been taken…
- Same (or simpler) model is often used…
- Largest data point remains the largest so top of hill should be near it…
- Indicated best operating conditions without a transformation will be about the same as when the proper transformation is used.
- Take checkpoints there!

Data Transformations - Why Do Them?

- Remedy for lack of fit
- Plot predictions will not violate physical limits
  - “# of Counts” not negative;
  - “YIELD” not > 100%
- Make error more uniform across design region
  (also called “stabilizing the variance”)

Transformations change the scale of the response to make it more nearly conform to the usual regression assumptions, the most important of which are that the data are independent and follow a **normal distribution with a constant variance.**