S-Matrix®

FUSION AE

Verification of Automated Regression Modeling in the Fusion AE Software Platform

Prepared by: S-Matrix Corporation
WWW.SMATRIX.COM

January, 2013

Table of Contents

Executive Summary

Design of Experiments (DOE) methodology, which incorporates multiple linear regression (MLR) equation building (modeling) techniques, is an integral component of the Quality by Design (QbD) paradigm now mandated by the government agencies which regulate the pharmaceutical industry. QbD is now being aggressively adopted by the industry globally in response to this mandate. To assist the working pharmaceutical scientists and engineers required to implement QbD, S-Matrix has developed the only fully automated MLR routine of its kind, and implemented this capability within its Fusion AE software platform. This revolutionary capability, which represents an extraordinary accomplishment, enables scientists and engineers to obtain statistically defensible MLR equations (models) from their DOE experimental data with one-click analysis.

This validation exercise demonstrates that the Automated MLR capability developed by S-Matrix Corporation and implemented in the Fusion AE software contains all the relevant component analysis operations required of a fully automated MLR routine, and that these routines are implemented in the correct operational sequence. This exercise further demonstrates that the calculations executed by these component routines are correct, and that the routines have the required sensitivity to boundary conditions in response data sets. In addition to deriving the correct model form from MLR analysis of the response data, identifying an outlier, and determining and implementing the appropriate data transformation, Fusion AE's summary report correctly identifies MLR results which require investigation into the source data used in the analysis.

It is especially noteworthy that about a decade ago a major international corporation spent over $1,000.000.00 attempting to develop an automated MLR analysis capability. They were unsuccessful. This validation exercise therefore demonstrates that Fusion AE's automated MLR software is at once both correctly implemented and a major advance in the field of MLR analysis.

Douglas C. Montgomery, Ph.D.

Regents' Professor of Industrial Engineering & Statistics
ASU Foundation Professor of Engineering
School of Computing, Informatics and Decision Systems Engineering
Arizona State University
Tempe, AZ 85287-8809
480.965.3836
doug.montgomery@asu.edu

Overview

Design of Experiments (DOE) methodology is an integral component of the Quality by Design (QbD) paradigm instituted in government regulatory guidances for the pharmaceutical industry. QbD is being aggressively adopted by the industry globally. However, the formal education path of most working pharmaceutical scientists does not include either the statistical experimental design or the multiple linear regression (MLR) equation-building (modeling) techniques which are integral elements of DOE. To support these working pharmaceutical scientists and engineers, S-Matrix has developed the only fully automated MLR routine of its kind, and implemented this capability within its Fusion AE software platform. This revolutionary capability, which represents an extraordinary accomplishment, enables scientists and engineers to obtain statistically defensible MLR equations (models) from their experimental data with one-click analysis.

It is extremely important that an automated MLR routine integrated within a DOE-based software program designed for working scientists correctly sequences and executes the capabilities listed below.

1. Component routines which perform the following operations:
    a. “Combined Mode” of MLR (Forward Stepwise + Backward Elimination).
    b. Augmented Box-Cox Transformation Analysis – additional transformations to address “nonlinear” response functions (e.g. Lognormal – Lower Limit, Lognormal – Upper Limit, and Tangent).
    c. Experimental Error Analysis using internal replicate run data.
    d. Outlier diagnostics.
2. Correct sequencing of the component routines within the automated MLR analysis.
3. Internal diagnostic routines and operational loops to assure proper parameterization of subsequent routines in the automated analysis chain (e.g. routines which diagnose and self correct to minimize the possibility of over-fitting or under-fitting the data).
4. A reporting capability which communicates the adequacy of the model fit to the scientist.

This report presents the results of an independent verification conducted by Dr. Douglas C. Montgomery of the automated MLR analysis capability in the S-Matrix Fusion AE software program. The capability was evaluated by comparing the analysis results for several specially-designed data sets obtained from Fusion AE and the JMP Software Package (SAS Institute). These data sets were specifically designed and constructed by Dr. Montgomery to challenge Fusion AE’s automated regression analysis capabilities in four critical areas:

1. Statistical sufficiency of the final model.
2. Outlier analysis and data handling.
3. Data transformation analysis and handling.
4. Pareto analysis and ranking of model coefficients.

This report contains four sections: Automated MLR Test Strategy, Test Bed Creation, Automated MLR Test Results, and Conclusions. The first section, Automated MLR Test Strategy, describes test data set design approach. The second section, Test Bed Creation, defines the statistical experimental design used in each test bed and the details of each data set creation. The third section, Automated MLR Test Results, defines the comparative results between Fusion AE and JMP, and Fusion AE performance in the four critical areas listed above. The last section, Conclusions, summarizes the results and lists further evaluation requirements.

Automated MLR Test Strategy

Each data set used in this exercise consists of a statistical experimental design (the X data) and a response data set (the Y data). In each case the experimental design was selected based on the equation (model) underlying the design (the design model) – in most cases this is a full quadratic equation of the general form:

$$\hat{Y} = \beta_0 + \beta_i(x_i) + \beta_{ii}(x_i)^2 + \beta_{ij}(x_i * x_j)$$

The response data set was generated using a specific model (the data model) with pre-defined model term coefficients (β_i values). A random error was then added to the individual response values. The magnitude of the random error was deliberately set to be equal in magnitude to many of the coefficients in the model used to generate the response data. This was done to determine if the automated regression analysis had the sensitivity required to make the following determinations when analyzing data sets with large inherent error:

- Identify an outlier which was deliberately included in the response data set, and automatically eliminate the outlier as part of the automated analysis.
- Identify the need for a data transformation, select the correct transformation function, and correctly transform the response data when the response data are constructed using a specific transformation.
- Derive the correct sign and rank of the prediction model coefficients for all terms in the model used to generate the response data set.

Model Sufficiency Metrics

Model Sufficiency is defined as the ability of the final model to re-predict the individual response data values with an acceptable prediction error across the response data range. This means that the model prediction error is of the same magnitude as overall experimental error (i.e., the model is not under-fitting or over-fitting the response data). Model sufficiency is determined using three statistical metrics:

- Error %

  The Error% is the proportional amount of the total variation in the response data across all experiment runs which can be attributed to overall experimental error, expressed as a percent. The magnitude of the overall experimental error variation is estimated from the replicate run groups in the experimental data set. When the estimated Error % is greater than 25% the software should warn the user that the model should be used with caution.

- MSR/MSE F-ratio

  The MSR is the amount of the total variation in the response data across all experiment runs accounted for by the model. The MSE is the difference between the total variation in the response data across all experiment runs and that amount of variation accounted for by the model. The MSR/MSE ratio is compared to a *threshold F-ratio value* with an associated probability P-value equal to 0.0500. An F-ratio value greater than the threshold value indicates that the model MSR is statistically larger than the MSE. This indicates that the experiment variables have effects, i.e., the response data are not purely random, and that the model is able to account for at least some of these effects.

- MS-LoF/MS-PE F-ratio

  The MS-PE is the “Pure Error” variation estimated from the replicate runs in the experimental design, i.e., the model independent residual error variation. The MS-LoF is equal to the residual error variation – the total prediction error in the model – minus the MS-PE. It is therefore the model dependent residual error variation. In other words, it is the amount of prediction error in the model which cannot be attributed to experimental error. The MS-LoF/MS-PE F-ratio is used to determine if the difference in magnitude between the MS-LoF and the MS-PE is statistically significant. A statistically significant F-ratio means that the current model “does not explain” all of the response data variation above the amount of variation which can be attributed to overall experimental error alone.

Outliers

Outliers can best be determined by hypothesis testing of the R-student residuals, designated as t_i, using a t-test where each t_i follows a t_{n-p-1} t-distribution. An outlier is defined as a response data value for which the magnitude of the corresponding R-student residual value is ≥ the ±4 standard deviation value of the t-distribution. When more than one response data value is determined to be an outlier, it is critical that the automated regression analysis operation eliminate only the one outlier with the largest R-student residual value.

Data Transformations

Regression analysis by the method of least squares operates under the assumption that the overall experimental error associated with the response being analyzed is "iid", i.e., independent and identically distributed. In other words, the error is of uniform magnitude and with the same distribution across the response data range. Therefore, when the response data have a sufficiently large *relative* error, the data normally require a transformation (pre-treatment) to "normalize" the error – attempt to make it approximately iid – prior to regression analysis. A relative error is one in which the magnitude of the error is dependent on the magnitude of the response. The error can be either directly proportional or inversely proportional to the magnitude of the response across the data range.

The Box-Cox transformation method provides a diagnostic to first determine the need for a response data transformation prior to regression analysis, and second to identify the most appropriate transformation for the given response data set from among several alternative transformation functions.

Sign and Rank Order of Model Term Coefficients

For any model term selection strategy, the sign and the rank of the terms should be similar to the sign and rank of the terms in the model that generated the data set. The sign and rank of the terms should hold true even when a large error, compared to the model terms, has been imparted to the prediction values within a data set. When error is imparted to the data from an experiment design that can support a quadratic model, it is not uncommon that terms that were not terms in the prediction model are significant. This is especially true when the magnitude of the error is of similar size as the magnitude of the prediction model's coefficients.

Test Data Creation

All designs used for this exercise were created by software other than Fusion AE. Each response data set for a given design was created in the following four steps.

1. Create a predicted response value for each run in the statistical experimental design using the required prediction equation. This is done using Microsoft Excel. For example, Design 1is a three-factor design which supports a full quadratic model. The prediction equation used for the “Y1” response data set is:

   $$\hat{Y} = 45 + 2(A) + 3(B) + 3(C) + 2(A)^2 + (B)^2 + 2(A*B) + 3(A*C)$$

   which is represented in Excel spreadsheet format as presented below, where A, B, and C represent experiment variables A, B, and C in the experiment design.

Coded Variable	Coefficient Value
Constant	45
A	2
B	3
C	3
(A)^2	2
(B)^2	1
A*B	2
A*C	3

2. Using Excel’s random error generator, generate a random error value for each experiment run using the Normal Distribution function, then add the error to the response data value generated for the run by the model in Step 1 above. As an example, a random normal error with Mean = 0 and Standard Deviation (Std. Dev.) = ±2.0 can be generated in Cell A1 of an Excel spreadsheet by entering the function presented below into the cell:

   Cell A1=NORMINV(RAND(),0,2)

   The final result value for data analysis is obtained by adding the error value generated in this step to the predicted response value generated in Step 1 above.

3. When the test data set requires an outlier, it is generated for a randomly selected experiment design run in the Y data set by first multiplying the absolute value of the Std. Dev. associated with the data set by five (5*|Std. Dev.|), and then increasing the magnitude of the designated outlier run value by the resulting value. For example, given a Std. Dev. of ±2.0, the magnitude of the designated outlier value would be increased by 10 (5* 2.0).

4. Testing the software’s ability to correctly determine the need for the response data to be transformed prior to analysis requires pre-treating the response data set with the inverse transformation of the transformation function selected for the test. For example, if the test should determine that the Natural Log function is the correct data transformation, then the response data values must be transformed prior to analysis using the Exponential function. The Exponential function can be applied to a response data value in an Excel spreadsheet, in Cell A1 for example, by entering the Exponential function presented below into Cell B1:

   Cell B1=EXP(A1)

Experimental Design 1 – Three Study Factors

The 3-factor experimental design is a rotatable full factorial central composite design with six (6) center points. The design, which contains 20 runs, supports a full quadratic model.

The first response data set for the 3-factor design, designated "Y1", was generated within Excel in the following steps:

1. A predicted response (Y data) value is generated for each experimental design run using the model shown below, where A, B, and C represent the three experimental design factors.

$$\hat{Y}_1 = 45 + 2(A) + 3(B) + 3(C) + 2(A)^2 + 1(B)^2 + 2(A * B) + 3(A * C)$$

Coded Variable	Coefficient Value
Constant	45
A	2
B	3
C	3
(A)^2	2
(B)^2	1
A*B	2
A*C	3

2. A random error value was generated for each of the 20 runs in the experimental design. A standard normal distribution was used to create the random error values. The distribution parameters for the standard normal error generator within Excel were set to Mean = 0, Std. Dev. = ±2.0.

3. The final Y1 data values for the for the 3-factor design were generated by adding the random error value for each run to the associated model-predicted response value for the run.

A second response data set for the 3-factor design, designated "Y2", was generated within Excel in the following steps:

1. Twenty predictions are created from the 20 run experimental design using the model shown below, where A, B, and C represent the three experiment design factors.

$$\hat{Y}_2 = 1.0 + 0.5(A) + 0.2(B) - 0.1(C) + 1.0(A)^2 + 0.2(C)^2 - 0.5(A * B) + 0.3(A * C)$$

Coded Variable	Coefficient Value
Constant	1.0
A	0.5
B	0.2
C	-0.1
(A)^2	1.0
(C)^2	0.2
A*B	-0.5
A*C	0.3

2. The 20 predicted Y2 values were transformed using the exponent transformation to generate a data set that required a Natural Log transformation.

NOTE – a random error was not incorporated into this Y2 response data set.

Experimental Design 2 – Five Study Factors

The 5-factor experimental design is a rotatable full factorial central composite design with eight (8) center points. The design, which contains 50 runs, supports a full quadratic model.

The first response data set for the 5-factor design, designated "Y1", was generated within Excel in the following steps:

1. A predicted response value was generated for each of the 50 experimental design runs using the model shown below, where A, B, C, D, and E represent the five experiment design factors.

$$\hat{Y}_1 = 45.0 + 2.0(A) + 3.0(B) + 4.0(C) + 3.0(D) + 2.0(E) + 1.0(A)^2 + 2.0(B)^2 + 1.5(C)^2 + 1.5(D)^2 + 2.0(A*B) + 2.0(A*C) + 1.5(A*D) + 1.8(A*E) + 1.2(B*C) + 1.0(D*E) + 1.0(A^2*C)$$

Coded Variable	Coefficient Value
Constant	45.0
A	2.0
B	3.0
C	4.0
D	3.0
E	2.0
(A)^2	1.0
(B)^2	2.0
(C)^2	1.5
(D)^2	1.5
A*B	2.0
A*C	2.0
A*D	1.5
A*E	1.8
B*C	1.2
D*E	1.0
(A)^2*C	1.0

2. A random error value was generated for each of the 50 runs in the experimental design. A standard normal distribution was used to create the random error values. The distribution parameters for the standard normal error generator within Excel were set to Mean = 0, Std. Dev. = ±1.0.

3. The final Y1 data values for the 50 runs were generated by adding the random error value for each run to the associated model-predicted response value for the run.

A second response data set for the 5-factor design, designated "Y2", was generated within Excel in the following steps:

1. A predicted response value was generated for each of the 50 experimental design runs, where A, B, C, D, and E represent the five experiment design factors.

$$\hat{Y}_2 = -0.15 + 0.60(A) - 0.90(B) + 1.20(C) - 0.90(D) + 0.60(E) - 0.30(A)^2 + 0.60(B)^2 - 0.50(C)^2 + 0.40(D)^2 + 0.70(A*B) + 0.70(A*C) + 0.50(A*D) + 0.60(A*E) + 0.40(B*C) + 0.30(D*E)$$

Coded Variable	Coefficient Value
Constant	-0.15
A	0.60
B	-0.90
C	1.20
D	-0.90
E	0.60
(A)^2	-0.30
(B)^2	0.60
(C)^2	-0.50
(D)^2	0.40
A*B	0.70
A*C	0.70
A*D	0.50
A*E	0.60
B*C	0.40
D*E	0.30

2. A random error value was generated for each of the 50 runs in the experimental design. A standard normal distribution was used to create the random error values. The distribution parameters for the standard normal error generator within Excel were set to Mean = 0, Std. Dev. = ±0.2.

3. The Y2 data values were updated by adding the random error value for each run to the associated model-predicted response value for the run.

4. The updated Y2 data values were then transformed using the arctangent transformation to create a data set that required a Tangent transformation.

Experimental Design 3 – Eight Study Factors

The 8-factor experimental design is a D-optimal design. The design included five (5) replicated non-center point runs. The design, which contains 55 runs, supports a full quadratic model.

The first response data set for the 8-factor design, designated "Y1", was generated within Excel in the following steps:

1. A predicted response value was generated for each of the 55 experimental design runs, where A, B, C, D, E, F, G, and H represent the eight experimental design factors.

$$\hat{Y}_1 = 100.0 + 6.0(A) - 4.0(B) + 3.0(C) + 8.0(D) - 2.0(A)^2 + 1.8(B)^2 - 4.0(A*B) - 3.0(A*C) + 2.0(A*D) - 2.5(B*C) + 2.4(A*B*D)$$

Coded Variable	Coefficient Value
Constant	100.0
A	6.0
B	-4.0
C	3.0
D	8.0
(A)^2	-2.0
(B)^2	1.8
A*B	-4.0
A*C	-3.0
A*D	2.0
B*C	-2.5
A*B*D	2.4

2. A random error value was generated for each of the 55 runs in the experimental design. A standard normal distribution was used to create the random error values. The distribution parameters for the standard normal error generator within Excel were set to Mean = 0, Std. Dev. = ±2.0.

3. The final Y1 data values for the 55 runs were generated by adding the random error value for each run to the associated model-predicted response value for the run.

A second response data set for the 8-factor design, designated "Y2", was generated within Excel in the following steps:

1. A predicted response value was generated for each of the 55 experimental design runs, where A, B, C, D, E, F, G, and H represent the eight experiment design factors.

$$\hat{Y}_2 = 100.0 + 12.0(A) - 4.0(B) + 3.0(C) + 8.0(D) + 2.0(A)^2 + 5.0(B)^2 - 4.0(A*B) - 3.0(A*C) + 2.0(A*D) - 2.5(B*C) + 2.4(A*B*D)$$

Coded Variable	Coefficient Value
Constant	100.0
A	12.0
B	-4.0
C	3.0
D	8.0
(A)^2	2.0
(B)^2	5.0
A*B	-4.0
A*C	-3.0
A*D	2.0
B*C	-2.5
A*B*D	2.4

2. A random error value was generated for each of the 55 runs in the experimental design. A standard normal distribution was used to create the random error values. The distribution parameters for the standard normal error generator within Excel were set to Mean = 0, Std. Dev. = ±1.0.

3. The Y2 data values for the 55 runs were generated by adding the random error value for each run to the associated model-predicted response value for the run.

4. An outlier is generated for a randomly selected experiment design run in the Y2 data set (Run 14 in this data set) by first multiplying the absolute value of the Std. Dev. associated with the data set by five (5*|Std. Dev.|), and then increasing the magnitude of the designated outlier run value by the resulting value. For Run 14, the original value was 121.20, and the final outlier value is 126.20.

A third response data set for the 8-factor design, designated "Y3", was generated within Excel in the following steps:

1. A predicted response value was generated for each of the 55 experimental design runs, where A, B, C, D, E, F, G, and H represent the eight experiment design factors.

$$\hat{Y}_3 = 100.0 + 6.0(A) - 4.0(B) + 3.0(C) + 8.0(D) - 2.0(A)^2 + 1.8(B)^2 - 4.0(A*B) - 3.0(A*C) + 2.0(A*D) - 2.5(B*C) + 2.4(A*B*D)$$

Coded Variable	Coefficient Value
Constant	100.0
A	6.0
B	-4.0
C	3.0
D	8.0
(A)^2	-2.0
(B)^2	1.8
A*B	-4.0
A*C	-3.0
A*D	2.0
B*C	-2.5
A*B*D	2.4

2. A random error value was generated for each of the 55 runs in the experimental design. A standard normal distribution was used to create the random error values. The distribution parameters for the standard normal error generator within Excel were set to Mean = 0, Std. Dev. = ±4.0.

3. The Y3 data values for the 55 runs were generated by adding the random error value for each run to the associated model-predicted response value for the run.

Experimental Design 4 – Ten Study Factors

The 10-factor experimental design is a rotatable Response Surface Reduced Cubic Model Central Composite design with six (6) center points. The design, which contains 82 runs, supports a full quadratic model.

The first response data set for the 10-factor design, designated "Y1", was generated within Excel in the following steps:

1. A predicted response value was generated for each of the 55 experimental design runs, where A, B, C, D, E, F, G, H, I, and J represent the 10 experimental design factors.

   $$\hat{Y}_1 = 45 + 4(A) + 5(B) + 3(C) - 8(D) + 3(A)^2 + 4(B)^2 - 2(C)^2 + 6(A*B) - 3(A*C) + 2(A^2*B)$$

Coded Variable	**Coefficient Value**
Constant	45
A	4
B	5
C	3
D	-8
(A)^2	3
(B)^2	4
(C)^2	-2
A*B	6
A*C	-3
(A)^2*B	2

2. A random error value was generated for each of the 82 runs in the experimental design. A standard normal distribution was used to create the random error values. The distribution parameters for the standard normal error generator within Excel were set to Mean = 0, Std. Dev. = ±2.0.

3. The Y1 data values for the 82 runs were generated by adding the random error value for each run to the associated model-predicted response value for the run.

4. An outlier is generated for a randomly selected experiment design run in the Y1 data set (Run 8 in this data set) by first multiplying the absolute value of the Std. Dev. associated with the data set by five (5*|Std. Dev.|), and then increasing the magnitude of the designated outlier run value by the resulting value. For Run 8, the original value was 37.66, and the final outlier value is 47.66.

A second response data set for the 8-factor design, designated "Y2", was generated within Excel in the following steps:

1. A predicted response value was generated for each of the 82 experimental design runs, where A, B, C, D, E, F, G, H, I, and J represent the 10 experiment design factors.

$$\hat{Y}_2 = 100.0 - 4.0(A) + 3.0(B) + 1.6(C) - 2.0(D) + 2.0(A)^2 + 1.5(B)^2 + 3.0(B)^2 + 2.0(D)^2 \\ - 1.0(A * B) + 1.7(A * C) + 1.5(A * D) + 1.0(B * C) + 2.0(A^2 * B)$$

Coded Variable	Coefficient Value
Constant	100.0
A	-4.0
B	3.0
C	1.6
D	-2.0
(A)^2	2.0
(B)^2	1.5
(C)^2	3.0
(D)^2	2.0
A*B	-1.0
A*C	1.7
A*D	1.5
B*C	1.0
(A)^2*B	2.0

2. A random error value was generated for each of the 82 runs in the experimental design. A standard normal distribution was used to create the random error values. The distribution parameters for the standard normal error generator within Excel were set to Mean = 0, Std. Dev. = ±1.0.

3. The Y2 data values for the 82 runs were generated by adding the random error value for each run to the associated model-predicted response value for the run.

Automated MLR Test Results

A table of analysis results was compiled for each of the analyzed test data sets. Each table contains five key MLR metrics: Model Sufficiency, Model Standard Error, Outlier Analysis, Transformation Analysis, and Model Coefficient Rank and Sign.

Model Sufficiency has two components: the MSR/MSE f-ratio and the MS-LoF/MS-PE f-ratio. The MSR result is considered sufficient if the MSR/MSE f-ratio is statistically significant and if the MS-LoF/MS-PE f-ratio is not statistically significant.

The Model Standard Error assessment evaluates the magnitude of the Model Standard Error relative to the magnitude of the standard normal error added to the test data set. The Model Standard Error result is considered sufficient if it is not significantly different from the standard normal error.

The Outlier Analysis result is correct if (1) it correctly identifies the outlier run as an outlier when one is present in the test data set, and (2) does not identify an outlier when one is not present in the test data set.

The Transformation Analysis result is correct if (1) it identifies the specific transformation required for MLR when the corresponding inverse transformation has been applied to the test data set, and (2) it does not identify the need for a transformation when one is not required for MLR analysis of the test data set.

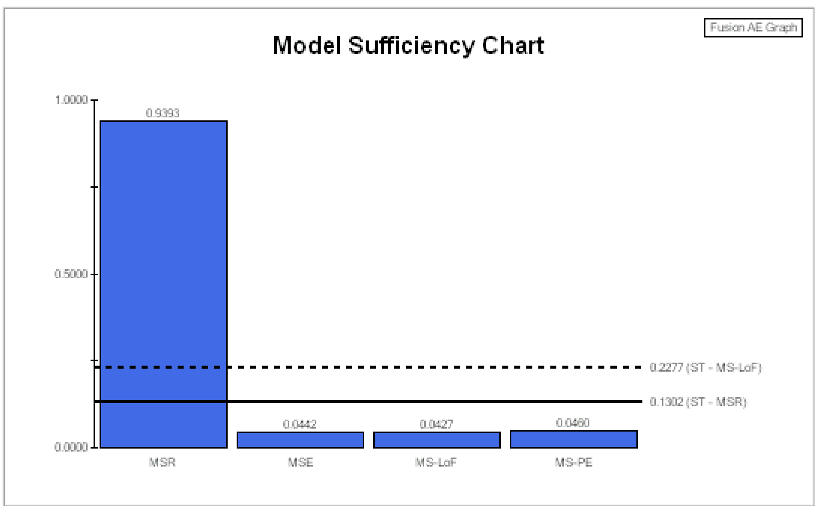
In each of the test data sets, the standard normal error added to the data was of the same magnitude as that of several of the coefficients in the model used to generate the data set (the data model). In such cases it is expected that the final model obtained from the automated MLR analysis will have a greater number of statistically significant terms than the number of terms in the model used to generate the data set. Model Coefficient Rank and Sign result is therefore considered correct if (1) the terms in the model used to generate the data are a perfect subset of the terms in the final MLR model, (2) the rank order and sign of the terms in the model used to generate the data match the rank order and sign of the terms in the final MLR model, and (3) the complement of terms in the final MLR model is a perfect subset of the complement of terms obtained from a matching MLR analysis executed within the JMP statistical analysis software package (SAS Institute). It should be noted that Fusion AE is the only software product with a fully automated and complete MLR analysis. Therefore, a manual multi-step analysis procedure was required in JMP to obtain a comparable analysis result.

Analysis Results – Three Study Factors – Y1 Response

Analysis Results Table - Summary

Suitability Metric	Expected Result	Actual Result	Pass/Fail
Model Sufficiency	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	Pass
Model Standard Error	±2.0	±1.9	Pass
Outlier Analysis	Outlier is not present	Outlier is not present	Pass
Transformation Analysis	Transformation = NONE	Transformation = NONE	Pass
Model Coefficient Rank and Sign	Rank and sign of terms are similar to the rank and sign of terms from the data model. Selected terms are a perfect subset of the terms selected by JMP.	Rank and sign of terms matches. Selected terms are a perfect subset of the terms selected by JMP.	Pass

1. Model Sufficiency



Regression Statistic	Computed Value	Scaled Value
R Square	0.9393	---
Adj. R Square	0.8951	---
Error %	10.92	---
MSR	73.7642	0.9393
MSE	3.4696	0.0442
MSR/MSE F-ratio	21.2600	---
MSR Significance Threshold	10.2284	0.1302
MS-LoF	3.3507	0.0427
MS-PE	3.6124	0.0460
MS-LoF/MS-PE F-ratio	0.9275	---
MS-LoF Significance Threshold	17.8823	0.2277

2. Model Error Analysis

Regression Statistic	Computed Value
R Square	0.9393
Adj. R Square	0.8951
Standard Error (+/-)	**1.8627**
Observations	20

3. Outlier Analysis

Analysis did not find an outlier.

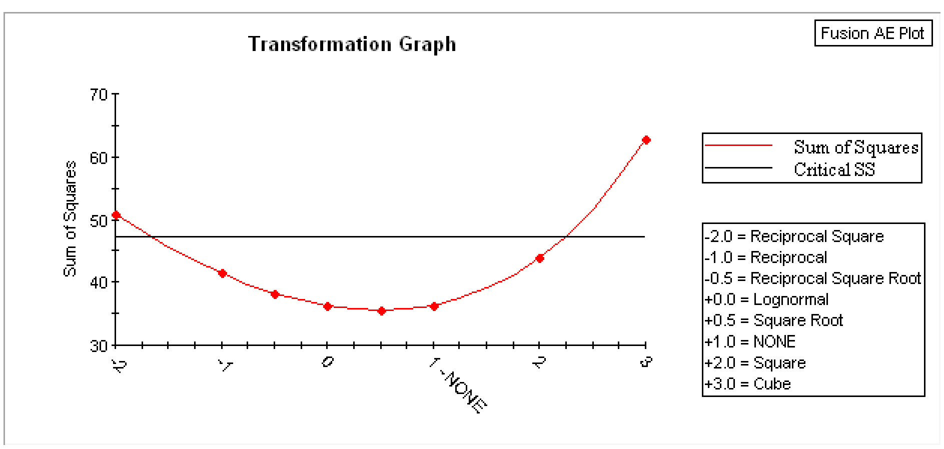
Largest R-student Residual value from Residuals Table.

Run No.	R-student Residuals
4	-2.1853

4. Transformation Analysis

Response Measurement Limit Settings	
Absolute Upper Limit	NONE
Absolute Lower Limit	NONE

Data Transformations	Lambda	Sum of Squares
Equivalent:		
NONE	1	36.3428
Square Root	0.5	35.5244
Natural Log	0	36.2340
Reciprocal Square Root	-0.5	38.2502
Reciprocal	-1	41.4200
Square	2	43.9015
Critical SS	---	**47.1942**
Not Recommended:		
Reciprocal Square	-2	50.8818
Cube	3	62.8504
Recommended Transformation		NONE
Selected Transformation		NONE



5. Model Coefficient Rank and Sign

Model Term	Data model Coefficient	Fusion AE Coefficient Value	JMP Coefficient Value
Constant	45	45.3184	45.3184
A	2	1.9224	1.9224
B	3	3.3191	3.3191
C	3	3.3524	3.3524
(A)^2	2	2.3914	2.3914
(B)^2	1	1.2795	1.2795
A*B	2	1.5638	1.5638
A*C	3	3.5063	3.5063
B*C	---	1.4988	1.4988

Expectation – Model Terms in Final Model

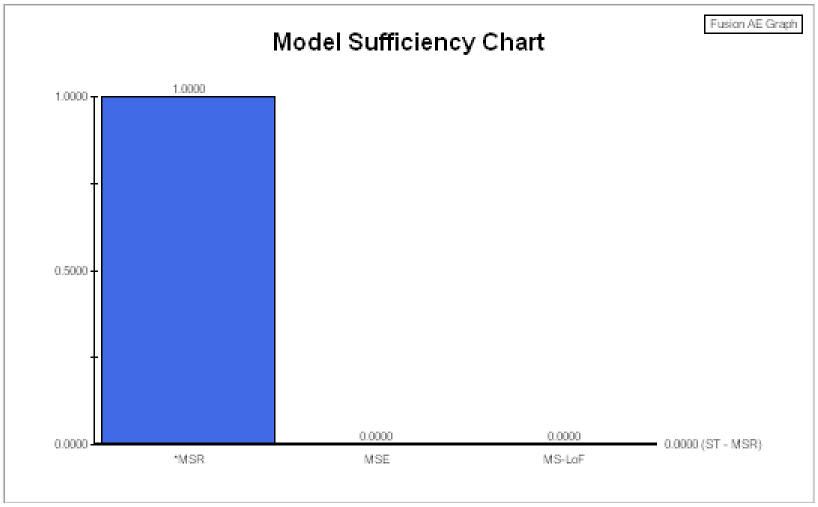
As expected, when error of a sufficient magnitude is added to simulated response data, some terms that were not in the data model such as the B*C term in the above example, will be included into the final regression models obtained by both JMP and Fusion AE.

Analysis Results – Three Study Factors – Y2 Response

Analysis Results Table - Summary

Suitability Metric	Expected Result	Actual Result	Pass/Fail
Model Sufficiency	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	Pass
Model Standard Error	±0.0	±0.0	Pass
Outlier Analysis	Outlier is not present	Outlier is not present	Pass
Transformation Analysis	Transformation = Natural Log	Transformation = Natural Log	Pass
Model Coefficient Rank and Sign	Rank and sign of terms are similar to the rank and sign of terms from the data model. Selected terms are a perfect subset of the terms selected by JMP.	Rank and sign of terms matches. Selected terms are a perfect subset of the terms selected by JMP.	Pass

1. Model Sufficiency



Regression Statistic	Computed Value	Scaled Value
R Square	1.0000	---
Adj. R Square	1.0000	---
Error %	< 0.01	---
MSR	1.6249	1.0000
MSE	0.0000	0.0000
MSR/MSE F-ratio	2.35×10^{19}	---
MSR Significance Threshold	0.0000	0.0000
MS-LoF*	0.0000	0.0000
MS-PE	0.0000	0.0000
MS-LoF/MS-PE F-ratio	---	---
MS-LoF Significance Threshold	0.0000	0.0000

2. Model Error Analysis

Regression Statistic	Computed Value
R Square	1.0000
Adj. R Square	1.0000
Standard Error (+/-)	**2.6E-10**
Observations	19

3. Outlier Analysis

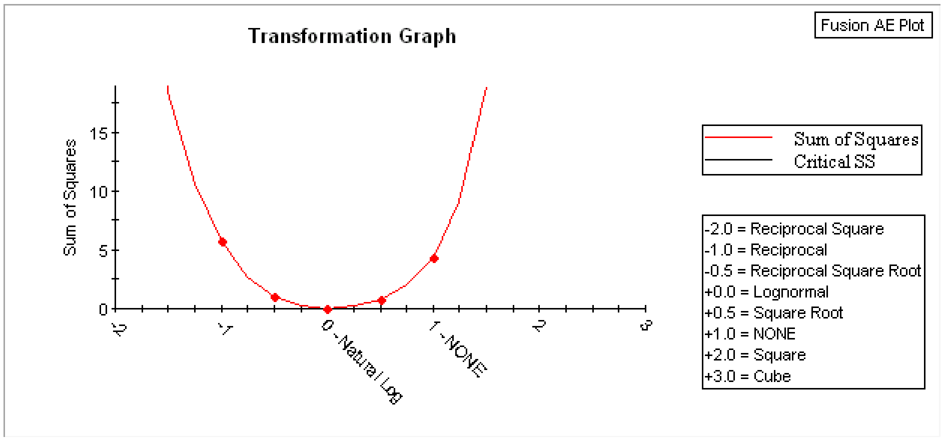
Analysis did not find an outlier.
Largest R-student Residual value from Residuals Table

Run No.	R-student Residuals
12	3.1213

4. Transformation Analysis

Response Measurement Limit Settings	
Absolute Upper Limit	NONE
Absolute Lower Limit	NONE

Data Transformations	Lambda	Sum of Squares
Equivalent:		
Natural Log	0	<±0.0001
Critical SS	---	**<±0.0001**
Not Recommended:		
Square Root	0.5	0.8309
Reciprocal Square Root	-0.5	1.0618
NONE	1	4.3389
Reciprocal	-1	5.7075
Reciprocal Square	-2	49.8996
Square	2	89.6263
Cube	3	1,959.2988
Recommended Transformation		Natural Log
Selected Transformation		Natural Log



5. Model Coefficient Rank and Sign

Model Term	Data model Coefficient	Fusion AE Coefficient Value	JMP Coefficient Value
Constant	1.0	1.0000	1.0000
A	0.5	0.5000	0.5000
B	0.2	0.2000	0.2000
C	-0.1	-0.1000	-0.1000
(A)^2	1.0	1.0000	1.0000
(B)^2	---	---	2.213E-10
(C)^2	0.2	0.2000	0.2000
A*B	-0.5	-0.5000	-0.5000
A*C	0.3	0.3000	0.3000
B*C	---	---	7.044E-11
(A)^2*B	---	---	5.547E-11
(A)^2*C	---	---	3.452E-11
A*(B)^2	---	---	1.772E-10
A*B*C	---	---	-1.17E-11

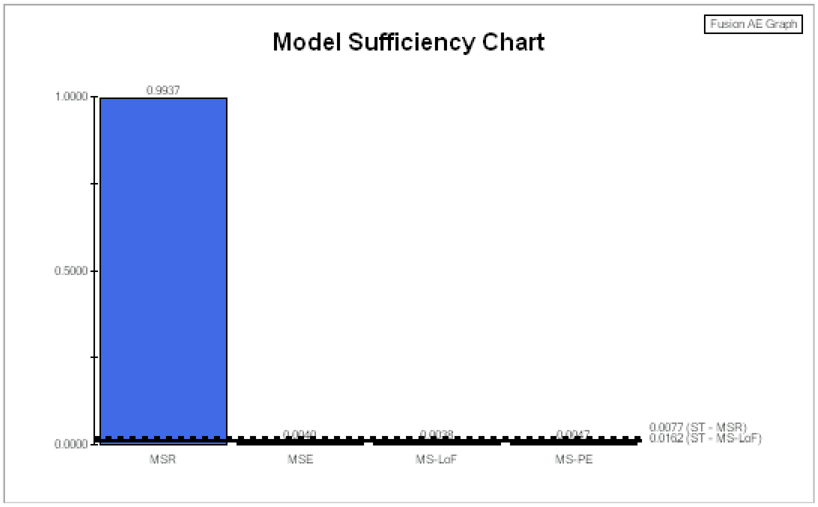
Note that in this analysis the Fusion AE result has fewer extraneous terms resulting from the high induced error than the corresponding JMP result.

Analysis Results – Five Study Factors – Y1 Response

Analysis Results Table - Summary

Suitability Metric	Expected Result	Actual Result	Pass/Fail
Model Sufficiency	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	Pass
Model Standard Error	±2.0	±1.8	Pass
Outlier Analysis	Outlier is not present	Outlier is not present	Pass
Transformation Analysis	Transformation = NONE	Transformation = NONE	Pass
Model Coefficient Rank and Sign	Rank and sign of terms are similar to the rank and sign of terms from the data model. Selected terms are a perfect subset of the terms selected by JMP.	Rank and sign of terms matches. Selected terms are a perfect subset of the terms selected by JMP.	Pass

1. Model Sufficiency



Regression Statistic	Computed Value	Scaled Value
R Square	0.9937	---
Adj. R Square	0.9897	---
Error %	1.22	---
MSR	161.6825	0.9937
MSE	0.6477	0.0040
MSR/MSE F-ratio	249.6314	---
MSR Significance Threshold	1.2599	0.0077
MS-LoF	0.6103	0.0038
MS-PE	0.7705	0.0047
MS-LoF/MS-PE F-ratio	0.7921	---
MS-LoF Significance Threshold	2.6335	0.0162

2. Model Error Analysis

Regression Statistic	Computed Value
R Square	0.9937
Adj. R Square	0.9897
Standard Error (+/-)	**0.8048**
Observations	50

3. Outlier Analysis

Analysis did not find an outlier.

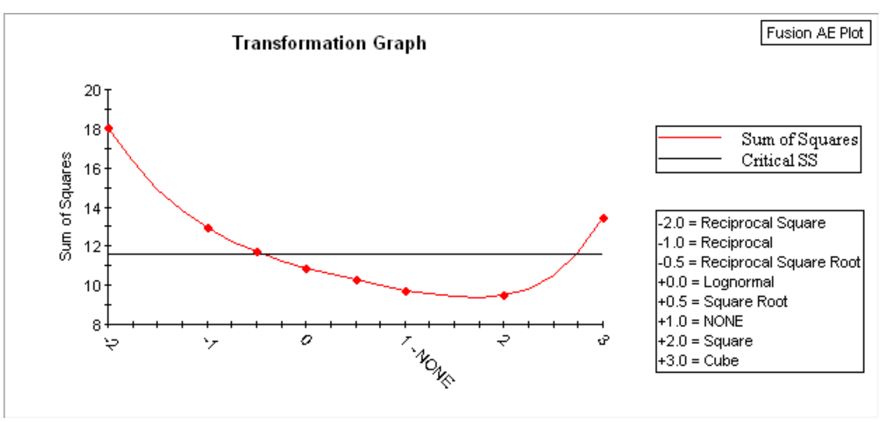
Largest R-student Residual value from Residuals Table.

Run No.	R-student Residuals
24	-2.6284

4. Transformation Analysis

Response Measurement Limit Settings	
Absolute Upper Limit	NONE
Absolute Lower Limit	NONE

Data Transformations	Lambda	Sum of Squares
Equivalent:		
NONE	1	9.7521
Square	2	9.4787
Square Root	0.5	10.2657
Natural Log	0	10.8920
Critical SS	---	**11.5791**
Not Recommended:		
Reciprocal Square Root	-0.5	11.7168
Reciprocal	-1	12.9468
Cube	3	13.4292
Reciprocal Square	-2	18.0942
Recommended Transformation		NONE
Selected Transformation		NONE



5. Model Coefficient Rank and Sign

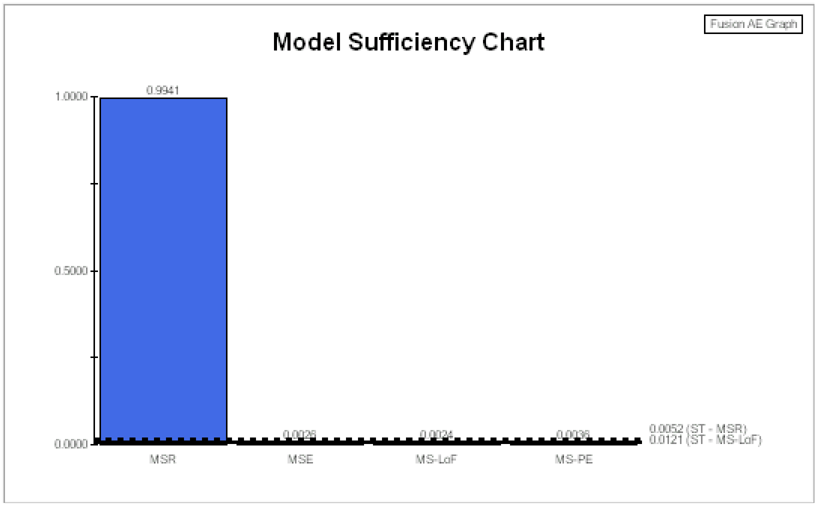
Model Term	Data model Coefficient	Fusion AE Coefficient Value	JMP Coefficient Value
Constant	45	45.0947	45.0947
A	2	2.0984	2.0984
B	3	3.4107	3.4107
C	4	3.4591	3.4591
D	3	2.3314	2.3314
E	2	1.9736	1.9736
(A)^2	1	1.1212	1.1212
(B)^2	2	2.0092	2.0092
(C)^2	1.5	1.4118	1.4118
(D)^2	1.5	1.4406	1.4406
A*B	2	2.1477	2.1477
A*C	2	1.8133	1.8133
A*D	1.5	1.5113	1.5113
A*E	1.8	1.6190	1.6190
B*C	1.2	1.1430	1.1430
D*E	1	0.8122	0.8122
(A)^2*C	1	1.5000	1.5000
(A)^2*D	---	0.9512	0.9512
A*C*D	---	0.2921	0.2921
C*D*E	---	0.3088	0.3088

Analysis Results – Five Study Factors – Y2 Response

Analysis Results Table - Summary

Suitability Metric	Expected Result	Actual Result	Pass/Fail
Model Sufficiency	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	Pass
Model Standard Error	±0.20	±0.22	Pass
Outlier Analysis	Outlier is not present	Outlier is not present	Pass
Transformation Analysis	Transformation = Tangent	Transformation = Tangent	Pass
Model Coefficient Rank and Sign	Rank and sign of terms are similar to the rank and sign of terms from the data model. Selected terms are a perfect subset of the terms selected by JMP.	Rank and sign of terms matches. Selected terms are a perfect subset of the terms selected by JMP.	Pass

1. Model Sufficiency



Regression Statistic	Computed Value	Scaled Value
R Square	0.9941	---
Adj. R Square	0.9914	---
Error %	1.16	---
MSR	18.5069	0.9941
MSE	0.0487	0.0026
MSR/MSE F-ratio	379.6508	---
MSR Significance Threshold	0.0961	0.0052
MS-LoF	0.0442	0.0024
MS-PE	0.0663	0.0036
MS-LoF/MS-PE F-ratio	0.6662	---
MS-LoF Significance Threshold	0.2250	0.0121

2. Model Error Analysis

Regression Statistic	Computed Value
R Square	0.9941
Adj. R Square	0.9914
Standard Error (+/-)	**0.2208**
Observations	50

3. Outlier Analysis

Analysis did not find an outlier.

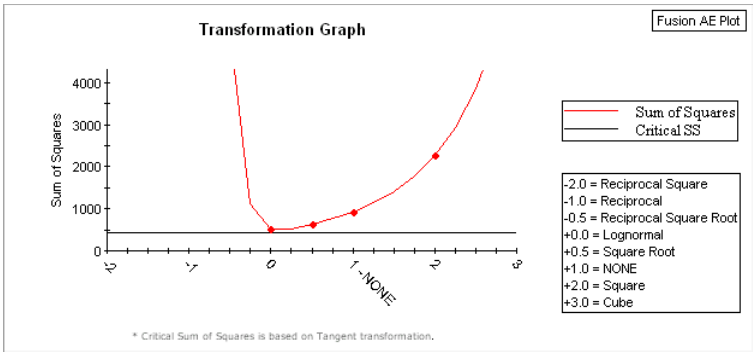
Largest R-student Residual value from Residuals Table.

Run No.	R-student Residuals
48	2.2593

4. Transformation Analysis

Response Measurement Limit Settings	
Absolute Upper Limit	100.0000
Absolute Lower Limit	0.0000

Data Transformations	Lambda	Sum of Squares
Equivalent:		
Tangent	---	353.0342
Critical SS	---	**431.2619**
Not Recommended:		
Natural Log - LL	---	488.7544
Natural Log	0	514.1314
Square Root	0.5	625.4904
Natural Log - UL	---	741.0863
Linear - Inverse	---	879.5104
NONE	1	933.4742
Square	2	2,273.0402
Reciprocal Square Root	-0.5	5,050.9057
Cube	3	6,975.8444
Reciprocal	-1	142,425.5182
Reciprocal Square	-2	139,618,952.4308
Recommended Transformation		Tangent
Selected Transformation		Tangent



5. Model Coefficient Rank and Sign

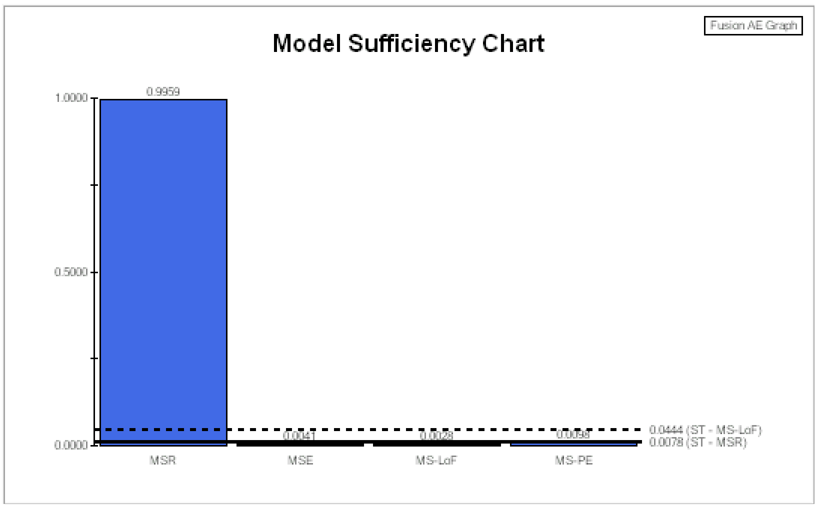
Model Term	Data model Coefficient	Fusion AE Coefficient Value	JMP Coefficient Value
Constant	-0.15	-0.1272	-0.1272
A	0.60	0.6321	0.6321
B	-0.90	-0.9225	-0.9225
C	1.20	1.1802	1.1802
D	-0.90	-0.9199	-0.9199
E	0.60	0.6005	0.6005
(A)^2	-0.30	-0.2777	-0.2777
(B)^2	0.60	0.6004	0.6004
(C)^2	-0.50	-0.5428	-0.5428
(D)^2	0.40	0.3416	0.3416
A*B	0.70	0.7561	0.7561
A*C	0.70	0.6820	0.6820
A*D	0.50	0.4759	0.4759
A*E	0.60	0.5174	0.5174
B*C	0.40	0.4659	0.4659
D*E	0.30	0.2440	0.2440

Analysis Results – Eight Study Factors – Y1 Response

Analysis Results Table - Summary

Suitability Metric	Expected Result	Actual Result	Pass/Fail
Model Sufficiency	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	Pass
Model Standard Error	±2.0	±0.9	Pass
Outlier Analysis	Outlier is not present	Outlier is not present	Pass
Transformation Analysis	Transformation = NONE	Transformation = NONE	Pass
Model Coefficient Rank and Sign	Rank and sign of terms are similar to the rank and sign of terms from the data model. Selected terms are a perfect subset of the terms selected by JMP.	Rank and sign of terms matches. Selected terms are a perfect subset of the terms selected by JMP.	Pass

1. Model Sufficiency



Regression Statistic	Computed Value	Scaled Value
R Square	0.9959	---
Adj. R Square	0.9918	---
Error %	1.96	---
MSR	215.0844	0.9959
MSE	0.8868	0.0041
MSR/MSE F-ratio	242.5351	---
MSR Significance Threshold	1.6892	0.0078
MS-LoF	0.6085	0.0028
MS-PE	2.1112	0.0098
MS-LoF/MS-PE F-ratio	0.2882	---
MS-LoF Significance Threshold	9.5877	0.0444

2. Model Error Analysis

Regression Statistic	Computed Value
R Square	0.9959
Adj. R Square	0.9918
Standard Error (+/-)	**0.9417**
Observations	55

3. Outlier Analysis

Analysis did not find an outlier.

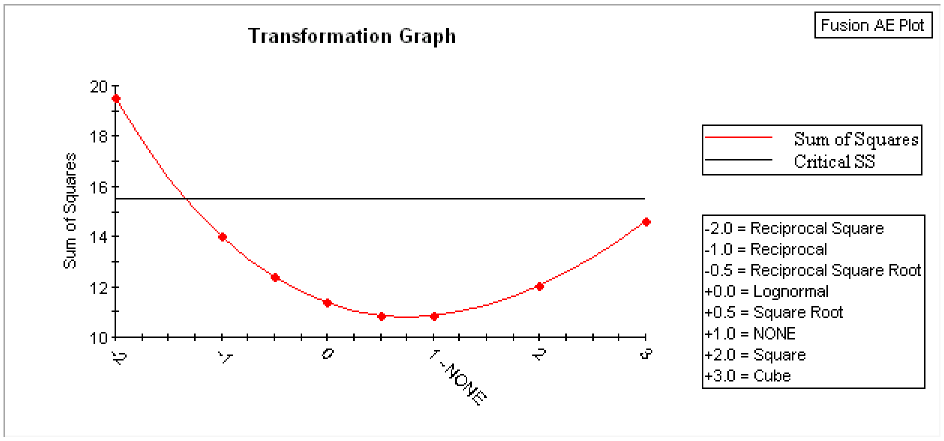
Largest R-student Residual value from Residuals Table.

Run No.	R-student Residuals
26	-3.4863

4. Transformation Analysis

Response Measurement Limit Settings	
Absolute Upper Limit	NONE
Absolute Lower Limit	NONE

Data Transformations	Lambda	Sum of Squares
Equivalent:		
NONE	1	10.8428
Square Root	0.5	10.8629
Natural Log	0	11.3557
Square	2	12.0297
Reciprocal Square Root	-0.5	12.3831
Reciprocal	-1	14.0201
Cube	3	14.6300
Critical SS	---	**15.5295**
Not Recommended:		
Reciprocal Square	-2	19.4994
Recommended Transformation		NONE
Selected Transformation		NONE



5. Model Coefficient Rank and Sign

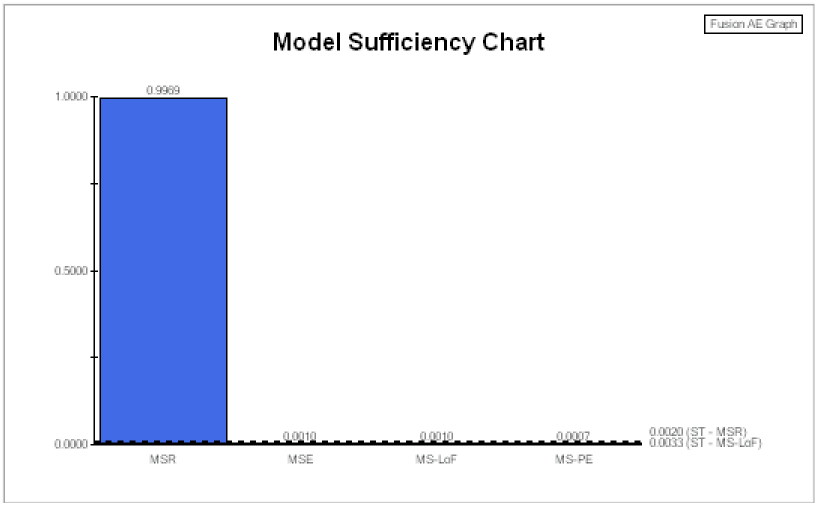
Model Term	Data model Coefficient	Fusion AE Coefficient Value	JMP Coefficient Value
Constant	100.0	100.8290	100.8290
A	6.0	6.2576	6.2576
B	-4.0	-3.6750	-3.6750
C	3.0	3.1031	3.1031
D	8.0	8.4346	8.4346
F	---	0.5373	0.5373
G	---	-0.5052	-0.5052
(A)^2	-2.0	-1.9654	-1.9654
(B)^2	1.8	-1.3472	-1.3472
(H)^2	---	1.3489	1.3489
A*B	-4.0	-3.6353	-3.6353
A*C	-3.0	-2.8876	-2.8876
A*D	2.0	2.5350	2.5350
A*E	---	-1.1373	-1.1373
A*G	---	-0.8971	-0.8971
A*H	---	0.6478	0.6478
B*C	-2.5	-2.7342	-2.7342
B*G	---	0.6832	0.6832
C*D	---	0.5897	0.5897
C*H	---	-0.5723	-0.5723
D*E	---	-0.9011	-0.9011
D*H	---	-0.5169	-0.5169
E*G	---	-1.3583	-1.3583
E*H	---	-0.6099	-0.6099
G*H	---	0.8646	0.8646
A*B*D	2.4	2.2845	2.2845
A*F*H	---	-1.9746	-1.9746
B*D*H	---	-1.0681	-1.0681

Analysis Results – Eight Study Factors – Y2 Response

Analysis Results Table - Summary

Suitability Metric	Expected Result	Actual Result	Pass/Fail
Model Sufficiency	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	Pass
Model Standard Error	±1.0	±0.9	Pass
Outlier Analysis	Outlier is present (Run 14; 126.20)	Outlier is present (Run 14; 126.20)	Pass
Transformation Analysis	Transformation = NONE	Transformation = NONE	Pass
Model Coefficient Rank and Sign	Rank and sign of terms are similar to the rank and sign of terms from the data model. Selected terms are a perfect subset of the terms selected by JMP.	Rank and sign of terms matches. Selected terms are a perfect subset of the terms selected by JMP.	Pass

1. Model Sufficiency



Regression Statistic	Computed Value	Scaled Value
R Square	0.9969	---
Adj. R Square	0.9960	---
Error %	0.30	---
MSR	747.5934	0.9969
MSE	0.7446	0.0010
MSR/MSE F-ratio	1,004.0102	---
MSR Significance Threshold	1.4697	0.0020
MS-LoF	0.7656	0.0010
MS-PE	0.5556	0.0007
MS-LoF/MS-PE F-ratio	1.3781	---
MS-LoF Significance Threshold	2.4859	0.0033

2. Model Error Analysis

Regression Statistic	Computed Value
R Square	0.9969
Adj. R Square	0.9960
Standard Error (+/-)	**0.8629**
Observations	54

3. Outlier Analysis

Analysis found and removed an outlier.

Outlier Run No.	14
Outlier P-Value	<±0.001
Dropped From Analysis	Yes

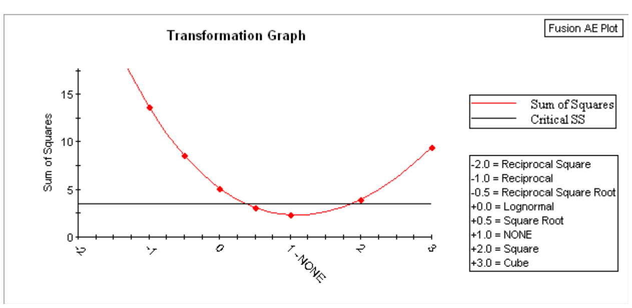
Largest R-student Residual value from Residuals Table.

Run No.	R-student Residuals
14	4.2073

4. Transformation Analysis

Response Measurement Limit Settings	
Absolute Upper Limit	NONE
Absolute Lower Limit	NONE

Data Transformations	Lambda	Sum of Squares
Equivalent:		
NONE	1	2.3357
Square Root	0.5	3.1108
Critical SS	---	**3.5334**
Not Recommended:		
Square	2	3.9495
Natural Log	0	5.1115
Reciprocal Square Root	-0.5	8.5405
Cube	3	9.4189
Reciprocal	-1	13.6747
Reciprocal Square	-2	30.6354
Recommended Transformation		NONE
Selected Transformation		NONE



5. Model Coefficient Rank and Sign

Model Term	Data model Coefficient	Fusion AE Coefficient Value	JMP Coefficient Value
Constant	100.0	99.6231	99.2457
A	12.0	11.9334	11.8878
B	-4.0	-3.6645	-3.6021
C	3.0	2.8210	2.8161
D	8.0	7.9916	8.0740
G	---	0.3266	0.3188
(A)^2	2.0	2.2885	2.3266
(B)^2	5.0	5.1171	5.0429
(D)^2	---	---	0.5934
A*B	-4.0	-3.6741	-3.6620
A*C	-3.0	-3.1202	-3.1585
A*D	2.0	1.8831	1.8434
B*C	-2.5	-2.6010	-2.5693
D*E	---	---	-0.3229
D*H	---	---	0.3260
F*H	---	-0.4409	-0.4210
A*B*D	2.4	2.6230	2.5099

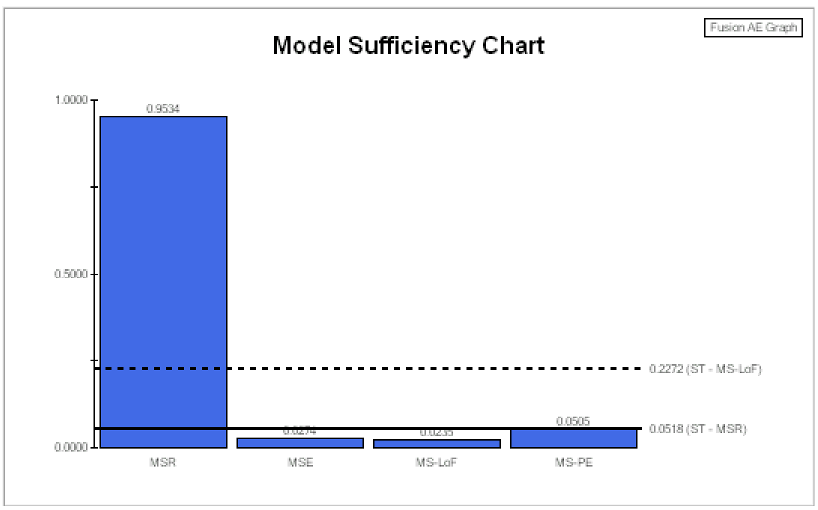
Note that (1) the terms in the data model are a perfect subset of the terms in the final MLR model, (2) the rank order and sign of the terms in the data model are consistent with the rank order and sign of the terms in the final MLR model, and (3) the complement of terms in the final MLR model is a perfect subset of the complement of terms obtained from a matching MLR analysis executed within JMP. Also note that the Fusion AE result has fewer extraneous terms resulting from the high induced error than the corresponding JMP result.

Analysis Results – Eight Study Factors – Y3 Response

Analysis Results Table - Summary

Suitability Metric	Expected Result	Actual Result	Pass/Fail
Model Sufficiency	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	Pass
Model Standard Error	±4.0	±3.0	Pass
Outlier Analysis	Outlier is not present.	Outlier is not present.	Pass
Transformation Analysis	Transformation = NONE	Transformation = NONE	Pass
Model Coefficient Rank and Sign	Rank and sign of terms are similar to the rank and sign of terms from the data model. Selected terms are a perfect subset of the terms selected by JMP.	Rank and sign of terms matches. Selected terms are a perfect subset of the terms selected by JMP.	Pass

1. Model Sufficiency



Regression Statistic	Computed Value	Scaled Value
R Square	0.9534	---
Adj. R Square	0.9259	---
Error %	13.63	---
MSR	324.5652	0.9534
MSE	9.3375	0.0274
MSR/MSE F-ratio	34.7593	---
MSR Significance Threshold	17.6266	0.0518
MS-LoF	7.9834	0.0235
MS-PE	17.1911	0.0505
MS-LoF/MS-PE F-ratio	0.4644	---
MS-LoF Significance Threshold	77.3613	0.2272

2. Model Error Analysis

Regression Statistic	Computed Value
R Square	0.9534
Adj. R Square	0.9259
Standard Error (+/-)	3.0557
Observations	55

3. Outlier Analysis

Analysis did not find an outlier.

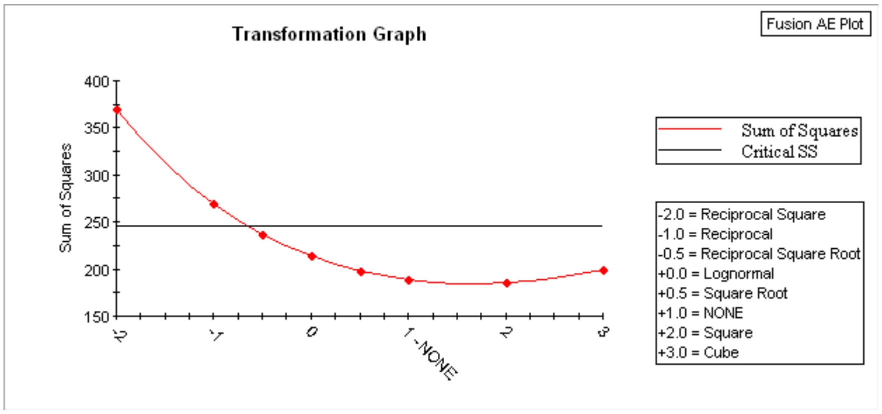
Largest R-student Residual value from Residuals Table.

Run No.	R-student Residuals
46	-2.8684

4. Transformation Analysis

Response Measurement Limit Settings	
Absolute Upper Limit	NONE
Absolute Lower Limit	NONE

Data Transformations	Lambda	Sum of Squares
Equivalent:		
NONE	1	188.8516
Square	2	185.4571
Square Root	0.5	198.3839
Cube	3	199.9200
Natural Log	0	214.1932
Reciprocal Square Root	-0.5	237.4370
Critical SS	---	**246.3800**
Not Recommended:		
Reciprocal	-1	269.6538
Reciprocal Square	-2	369.6882
Recommended Transformation		NONE
Selected Transformation		NONE



5. Model Coefficient Rank and Sign

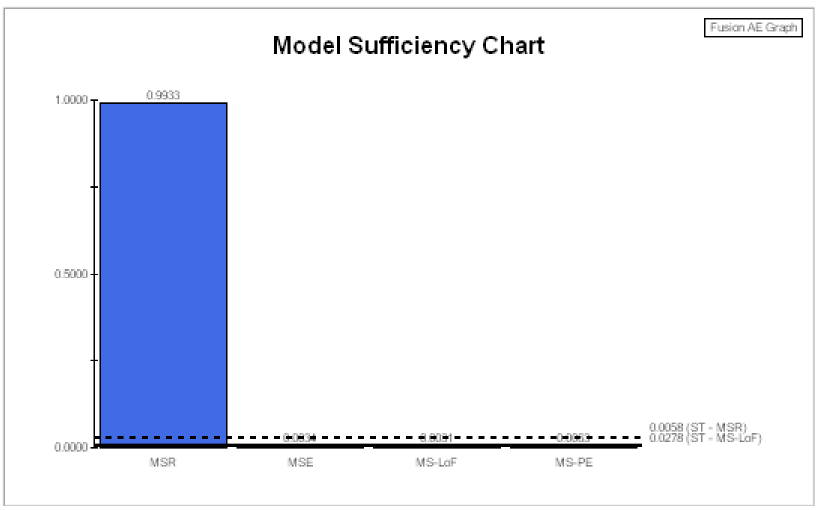
Model Term	Data model Coefficient	Fusion AE Coefficient Value	JMP Coefficient Value
Constant	100.0	98.1766	98.1766
A	6.0	6.5174	6.5174
B	-4.0	-4.7486	-4.7486
C	3.0	2.7772	2.7772
D	8.0	8.0559	8.0559
E	---	-1.4877	-1.4877
(A)^2	-2.0	-2.8876	-2.8876
(B)^2	1.8	---	---
(F)^2	---	3.6166	3.6166
(G)^2	---	3.1329	3.1329
A*B	-4.0	-4.7262	-4.7262
A*C	-3.0	-3.9899	-3.9899
A*D	2.0	1.1951	1.1951
A*E	---	-1.1388	-1.1388
A*H	---	1.2974	1.2974
B*C	-2.5	-1.2940	-1.2940
C*D	---	1.1384	1.1384
C*H	---	1.9070	1.9070
D*G	---	1.5177	1.5177
E*H	---	-1.2062	-1.2062
F*G	---	1.2468	1.2468
G*H	---	-1.6280	-1.6280
A*B*D	2.4	---	---

Analysis Results – Ten Study Factors – Y1 Response

Analysis Results Table - Summary

Suitability Metric	Expected Result	Actual Result	Pass/Fail
Model Sufficiency	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	Pass
Model Standard Error	±2.0	±1.3	Pass
Outlier Analysis	Outlier is present (Run 8; 47.66)	Outlier is present (Run 8; 47.66)	Pass
Transformation Analysis	Transformation = NONE	Transformation = NONE	Pass
Model Coefficient Rank and Sign	Rank and sign of terms are similar to the rank and sign of terms from the data model. Selected terms are a perfect subset of the terms selected by JMP.	Rank and sign of terms matches. Selected terms are a perfect subset of the terms selected by JMP.	Pass

1. Model Sufficiency



Regression Statistic	Computed Value	Scaled Value
R Square	0.9933	---
Adj. R Square	0.9898	---
Error %	1.85	---
MSR	483.6481	0.9933
MSE	1.6695	0.0034
MSR/MSE F-ratio	289.6967	---
MSR Significance Threshold	2.8331	0.0058
MS-LoF	1.5262	0.0031
MS-PE	3.0450	0.0063
MS-LoF/MS-PE F-ratio	0.5012	---
MS-LoF Significance Threshold	13.5431	0.0278

2. Model Error Analysis

Regression Statistic	Computed Value
R Square	0.9933
Adj. R Square	0.9898
Standard Error (+/-)	**1.2921**
Observations	81

3. Outlier Analysis

Analysis found and removed an outlier.

Outlier Run No.	8
Outlier P-Value	<±0.001
Dropped From Analysis	Yes

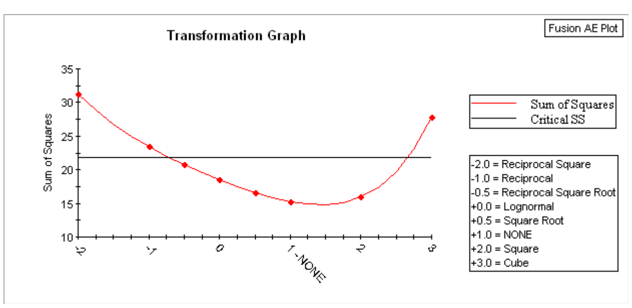
Largest R-student Residual value from Residuals Table.

Run No.	R-student Residuals
8	5.4705

4. Transformation Analysis

Response Measurement Limit Settings	
Absolute Upper Limit	NONE
Absolute Lower Limit	NONE

Data Transformations	Lambda	Sum of Squares
Equivalent:		
NONE	1	15.2639
Square	2	15.9431
Square Root	0.5	16.5955
Natural Log	0	18.4704
Reciprocal Square Root	-0.5	20.7478
Critical SS	---	**21.8616**
Not Recommended:		
Reciprocal	-1	23.4611
Cube	3	27.8095
Reciprocal Square	-2	31.2215
Recommended Transformation		NONE
Selected Transformation		NONE



5. Model Coefficient Rank and Sign

Model Term	Data model Coefficient	Fusion AE Coefficient Value	JMP Coefficient Value
Constant	45	44.7441	44.7981
A	4	4.5161	4.4765
B	5	5.0461	5.0461
C	3	2.4813	2.5781
D	-8	-9.2761	-9.2761
E	---	-1.1988	-1.1988
(A)^2	3	3.1158	3.1287
(B)^2	4	4.3009	4.3138
(C)^2	-2	-1.6524	-1.6395
(E)^2	---	---	-0.5874
(I)^2	---	---	0.4219
A*B	6	6.3762	6.4427
A*C	-3	-3.0278	-3.1674
B*E	---	0.6042	0.6443
B*F	---	-0.5032	-0.3667
B*G	---	-0.3911	-0.3743
B*H	---	---	-0.5172
B*I	---	---	-0.4355
B*J	---	0.4929	0.4873
C*E	---	---	-0.3899
C*F	---	0.5329	0.6826
C*H	---	0.6961	0.4610
C*J	---	---	0.4188
D*F	---	0.4538	0.4606
D*G	---	-0.4484	-0.3145
D*H	---	---	0.5203
D*J	---	0.4027	---
E*F	---	-0.4707	-0.5515
E*G	---	-0.7508	-0.9548
F*H	---	---	0.3439
G*I	---	---	0.3348
G*J	---	-0.6718	-0.4595
(A)^2*B	2	1.5511	1.6087
(A)^2*D	---	1.8059	1.5173
(A)^2*E	---	1.5841	1.6529
(A)^2*F	---	0.9686	0.8094
(A)^2*I	---	-0.5826	-0.6246

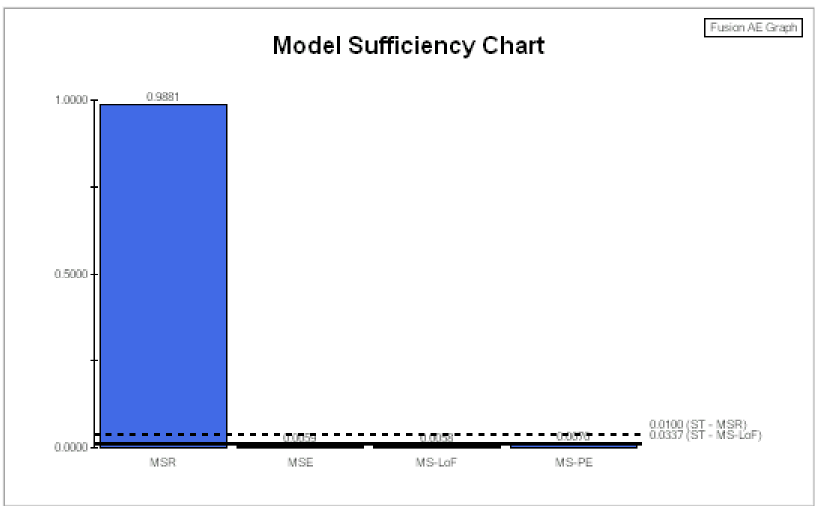
Note in this analysis that (1) the terms in the data model are a perfect subset of the terms in the final MLR model, (2) the rank order and sign of the terms in the data model are consistent with the rank order and sign of the terms in the final MLR model, and (3) the complement of terms in the final MLR model is a perfect subset of the complement of terms obtained from a matching MLR analysis executed within JMP.

Analysis Results – Ten Study Factors – Y2 Response

Analysis Results Table - Summary

Suitability Metric	Expected Result	Actual Result	Pass/Fail
Model Sufficiency	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	MSR/MSE F-ratio is significant. MS-LoF/MS-PE F-ratio is not significant.	Pass
Model Standard Error	±1.0	±1.0	Pass
Outlier Analysis	Outlier is not present.	Outlier is not present.	Pass
Transformation Analysis	Transformation = NONE	Transformation = NONE	Pass
Model Coefficient Rank and Sign	Rank and sign of terms are similar to the rank and sign of terms from the data model. Selected terms are a perfect subset of the terms selected by JMP.	Rank and sign of terms matches. Selected terms are a perfect subset of the terms selected by JMP.	Pass

1. Model Sufficiency



Regression Statistic	Computed Value	Scaled Value
R Square	0.9881	---
Adj. R Square	0.9822	---
Error %	2.27	---
MSR	156.1570	0.9881
MSE	0.9369	0.0059
MSR/MSE F-ratio	166.6690	---
MSR Significance Threshold	1.5863	0.0100
MS-LoF	0.9104	0.0058
MS-PE	1.1969	0.0076
MS-LoF/MS-PE F-ratio	0.7606	---
MS-LoF Significance Threshold	5.3214	0.0337

2. Model Error Analysis

Regression Statistic	Computed Value
R Square	0.9881
Adj. R Square	0.9822
Standard Error (+/-)	**0.9680**
Observations	82

3. Outlier Analysis

Analysis did not find an outlier.

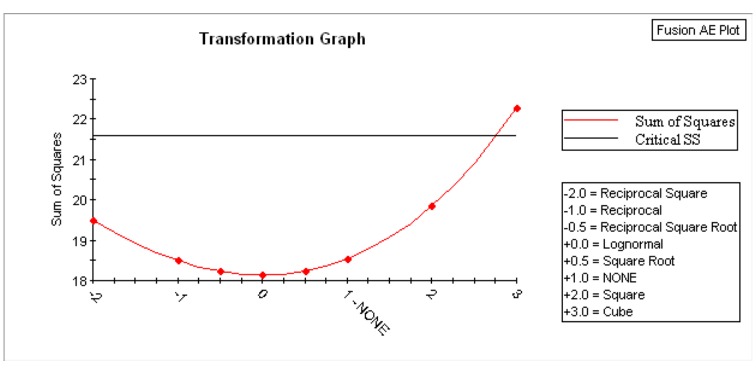
Largest R-student Residual value from Residuals Table.

Run No.	R-student Residuals
51	2.4724

4. Transformation Analysis

Response Measurement Limit Settings	
Absolute Upper Limit	NONE
Absolute Lower Limit	NONE

Data Transformations	Lambda	Sum of Squares
Equivalent:		
NONE	1	18.5326
Natural Log	0	18.1358
Reciprocal Square Root	-0.5	18.2299
Square Root	0.5	18.2306
Reciprocal	-1	18.4970
Reciprocal Square	-2	19.4999
Square	2	19.8472
Critical SS	---	**21.5908**
Not Recommended:		
Cube	3	22.2937
Recommended Transformation		NONE
Selected Transformation		NONE



5. Model Coefficient Rank and Sign

Coded Name	Generating Model Coefficient Value	Fusion AE Coefficient Value[1]	JMP Coefficient Value[2]
Constant	100.0	100.0536	100.0536
A	-4.0	-4.3130	-4.3130
B	3.0	3.6501	3.6501
C	1.6	1.9449	1.9449
D	-2.0	-1.8630	-1.8630
F	---	1.5884	1.5884
J	---	-1.0530	-1.0530
(A)^2	2.0	1.8212	1.8212
(B)^2	1.5	1.2648	1.2648
(C)^2	3.0	3.1365	3.1365
(D)^2	2.0	1.9456	1.9456
A*B	-1.0	-0.7601	-0.7601
A*C	1.7	1.8629	1.8629
A*D	1.5	1.6161	1.6161
A*F	---	0.5016	0.5016
B*C	1.0	1.1446	1.1446
B*D	---	-0.3085	-0.3085
B*F	---	0.3057	0.3057
C*G	---	0.2973	0.2973
C*H	---	-0.3149	-0.3149
C*J	---	0.3447	0.3447
D*E	---	0.3488	0.3488
D*J	---	0.3341	0.3341
E*H	---	0.4313	0.4313
F*J	---	-0.4274	-0.4274
(A)^2*B	2	1.1033	1.1033
(A)^2*F	---	-1.3741	-1.3741
(A)^2*J	---	1.1093	1.1093

Note that for this analysis the terms in the data model are a perfect subset of the terms in the final MLR model, and the rank order and sign of the terms in the data model are consistent with the rank order and sign of the terms in the final MLR model.

Conclusions

This validation exercise demonstrates that the Automated MLR capability developed by S-Matrix Corporation and implemented in the Fusion AE software contains all the relevant component analysis operations required of a fully automated MLR routine, and that these routines are implemented in the correct operational sequence. This exercise further demonstrates that the calculations executed by these component routines are correct, and that the routines have the required sensitivity to boundary conditions in response data sets. In addition to deriving the correct model form from MLR analysis of the response data, identifying an outlier, and determining and implementing the appropriate data transformation, Fusion AE's summary report correctly identifies MLR results which require investigation into the source data used in the analysis.

It is especially noteworthy that about a decade ago a major international corporation spent over $1,000.000.00 attempting to develop an automated MLR analysis capability. They were unsuccessful. This validation exercise therefore demonstrates that Fusion AE's automated MLR software is at once both correctly implemented and a major advance in the field of MLR analysis.

Douglas C. Montgomery, Ph.D.

Regents' Professor of Industrial Engineering & Statistics
ASU Foundation Professor of Engineering
School of Computing, Informatics and Decision Systems Engineering
Arizona State University
Tempe, AZ 85287-8809
480.965.3836
doug.montgomery@asu.edu