These hands-on activities allow you to practice predictive modeling using JMP. Extract the data in **Predictive Modeling Hands-on Activity.zip** for use in the following activities. The solutions follow the instructions for the activities.

Continuous response, continuous predictors

1. Use the data in Practice 1.jmp to build predictive models. These data were collected over time.
   a. Visualize the data using Graph Builder and Multivariate. Are there any data problems?
   b. Build a response surface model using Fit Model and the Response Surface macro. What is the R^2 of the full model? Are there any problems with the model fit seen in the Residual by Predicted plot? Which predictor variables are most important for predicting Y? Save the prediction formula to the data table.
   c. Build a neural network model using the default settings of the Neural platform. What is the R^2 of the model on the validation set? Are there any problems with the model fit seen in the Residual by Predicted plot? Fit another model with 50 nodes. Does this model have appreciably better predictive capability? Save the prediction formula to the data table.
   d. Build a decision tree model using the Partition platform with a 25% validation set. What is the R^2 of the model on the validation set? How many splits are in the model? Are there any problems with the model fit seen in the Actual by Predicted plot? Which variables are most important for predicting Y? Save the prediction formula to the data table.
   e. Build an Actual by Predicted graph for the three models using Graph Builder. Which model do you prefer?

Categorical response, continuous and categorical predictors

2. Use the data in Practice 2.jmp to build predictive models.
   a. Visualize the data using Graph Builder. Are there any data problems?
   b. Build an ordinal regression model using Fit Model and the Response Surface macro. What is the misclassification rate of the full model? Hint: open the Fit Details report. What is the misclassification rate for the Acceptable group? Hint: open the Confusion Matrix report.
   c. Build a neural network model using the default settings of the Neural platform. What is the misclassification rate of the model on the validation set? In particular, what is the misclassification rate of the Acceptable group? Fit another model with 50 nodes. What is the misclassification rate of the model on the validation set? What is the misclassification rate of the Acceptable group?
   d. Build a decision tree model using the Partition platform with a 25% validation set. What is the misclassification rate of the model on the validation set? What is the misclassification rate of the Acceptable group?

Solutions

Continuous response, continuous predictors

1. Use the data in PM 1.jmp to build predictive models. These data were collected over time.

a. Visualize the data using Graph Builder and Multivariate. Are there any data problems?

1) Open **PM 1.jmp**.

PM 1

Notes: Simulated data

Columns (11/0)

- Y
- X1
- X2
- X3
- X4
- X5
- X6
- X7
- X8
- X9
- X10

Rows

All rows	7,875
Selected	0
Excluded	0
Hidden	0
Labeled	0

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
1	54.574	12.04	4.07	6.96	21.79	32.22	31.74	10.04	28.81	29.1	9.86
2	53.993	13.25	8.21	-2.77	34.42	39.13	26.7	18.03	24.76	26.73	9.89
3	54.126	18.7	11.87	-3.65	22.36	38.7	14.08	32.03	22.35	24.72	20.45
4	53.838	19.53	13.02	-6.18	20.79	38.71	18.56	25.5	24.26	21.48	12.53
5	53.758	14.01	10.14	-0.67	24.87	46.62	26.21	21.32	30.51	29.84	5.55
6	54.773	9.94	11.88	10.98	29.64	34.78	25.1	7.84	19.97	21.92	10.84
7	57.167	15.28	14.98	26.74	39.46	13.07	16.53	4.39	8.6	6.22	19.14
8	59.063	27.99	14.96	32.69	35.39	0.14	5.95	11.22	5.14	5.18	29.65
9	58.673	21.73	9.14	21.81	35.64	7.95	10.74	4.59	11.59	9.42	20.32
10	59.308	25.18	4.58	29.88	30.45	17.08	12.51	16.89	17.03	20.23	24.72
11	60.572	17.3	2.93	40.27	36.32	16.64	16.49	19.04	17.94	20.75	18.26
12	61.386	17.81	1.34	38.78	26.83	21.81	21.89	25.82	24.31	30.78	21.65
13	62.137	26.46	3.11	33.75	24.17	25.85	19.14	30.05	25.57	34.42	22.05
14	61.299	22.29	5.98	26.55	26.64	23.76	23.03	30.88	24.43	26.67	15.44
15	59.525	17.23	7.67	18.92	33.2	16.78	30.07	12.12	16.43	13.97	8.45
16	57.247	16.71	3.87	9.17	31.87	16.42	29.76	1.97	16.24	12.46	10.73
17	56.526	26.19	3.11	6.99	15.28	13.62	21.68	20.03	22.25	29.38	24.37
18	55.55	21.02	3.13	-0.46	26.75	22.55	23.76	32.02	28.03	35.25	19.12

Build a time plot

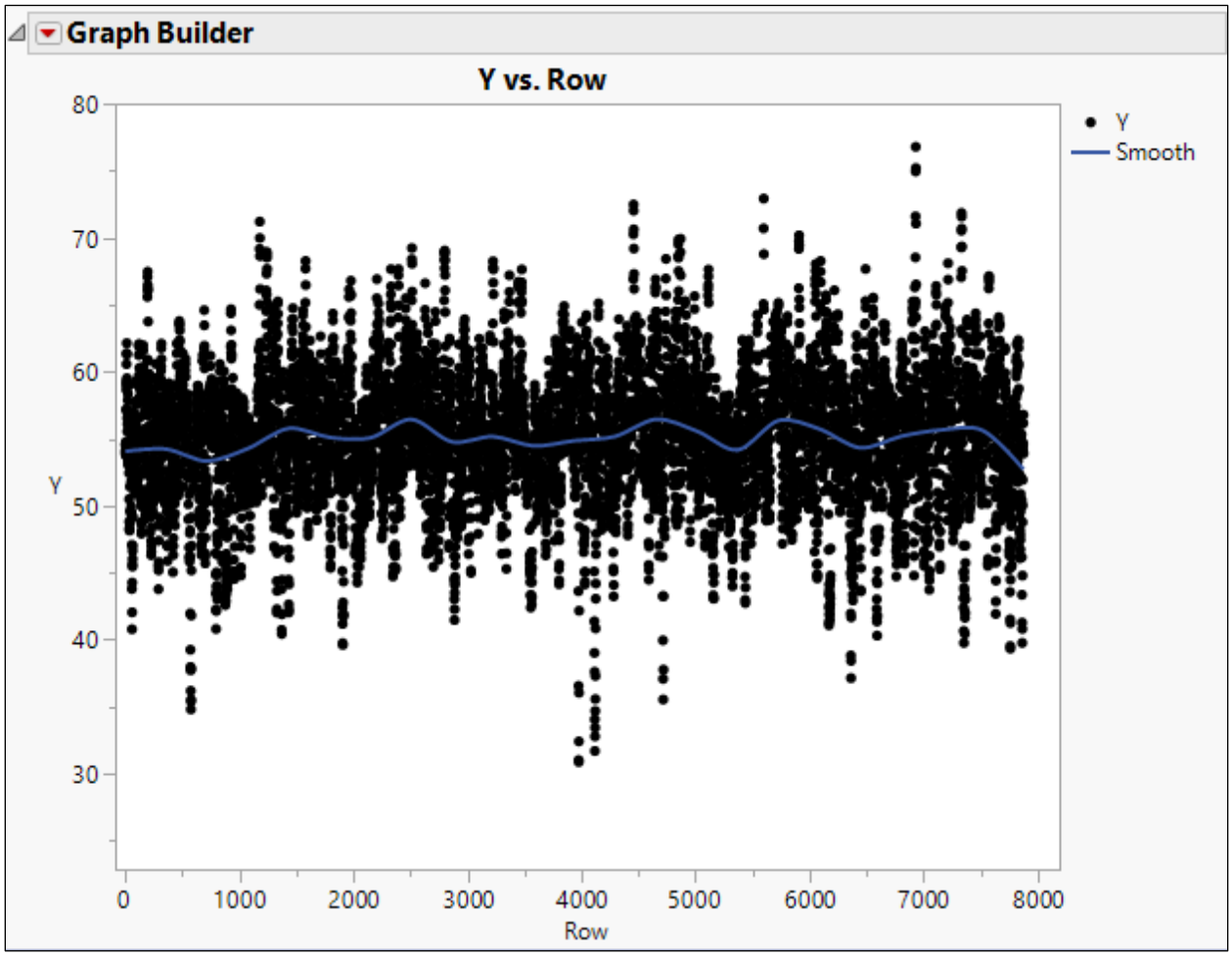
2) Select **Graph** > **Graph Builder**.

3) Drag **Y** to the Y zone.

4) Right-click **Y** and select **Row** > **Row**.

5) Drag **Row** to the X zone.

6) Click **Done**.

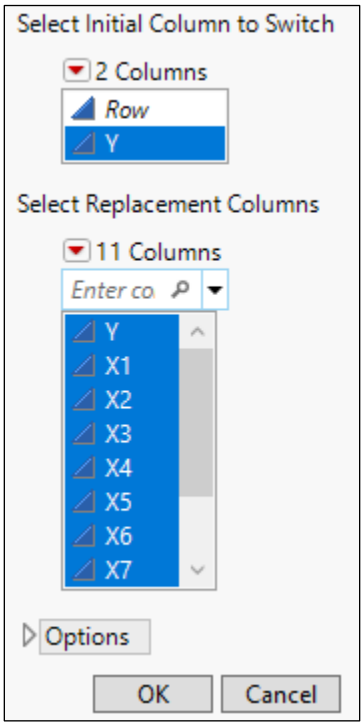


The data appear to be correlated over the rows. In this case, the rows correspond to time order. The mean and variability of the response appear to be constant over time. Add a column switcher to reproduce this graph for the predictor variables.

7) Click the red triangle next to **Graph Builder** and select **Redo** > **Column Switcher**.

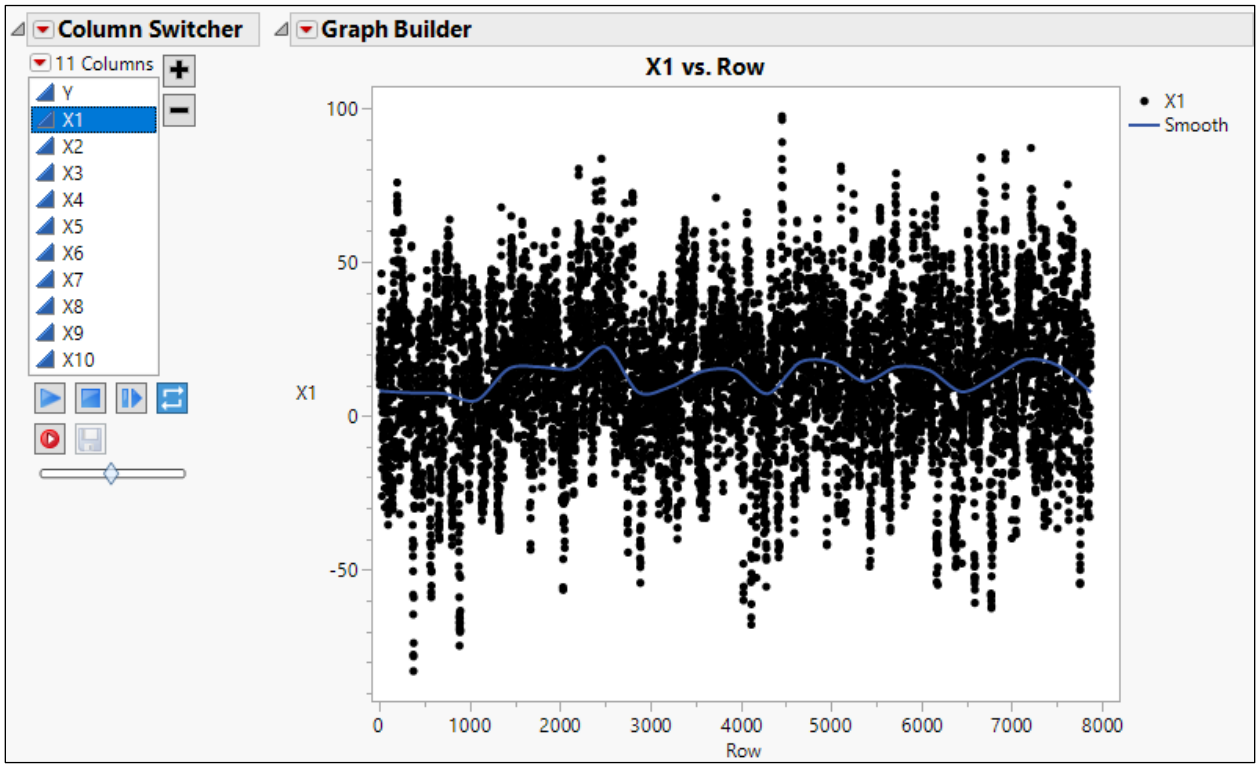
8) Select **Y** from the Select Initial Column to Switch box.

9) Select all columns from the Select Replacement Columns box.



10) Click **OK**.

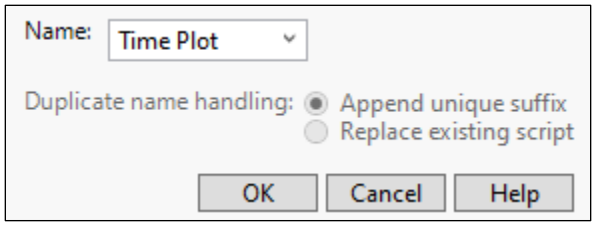
11) Click on each variable in the Column Switcher in turn.



Many variables appear autocorrelated. No other data problems are evident. Save the script to the data table.

12) Click the red triangle next to **Graph Builder** and select **Save Script** > **To Data Table**.

13) Enter **Time Plot** for the name.

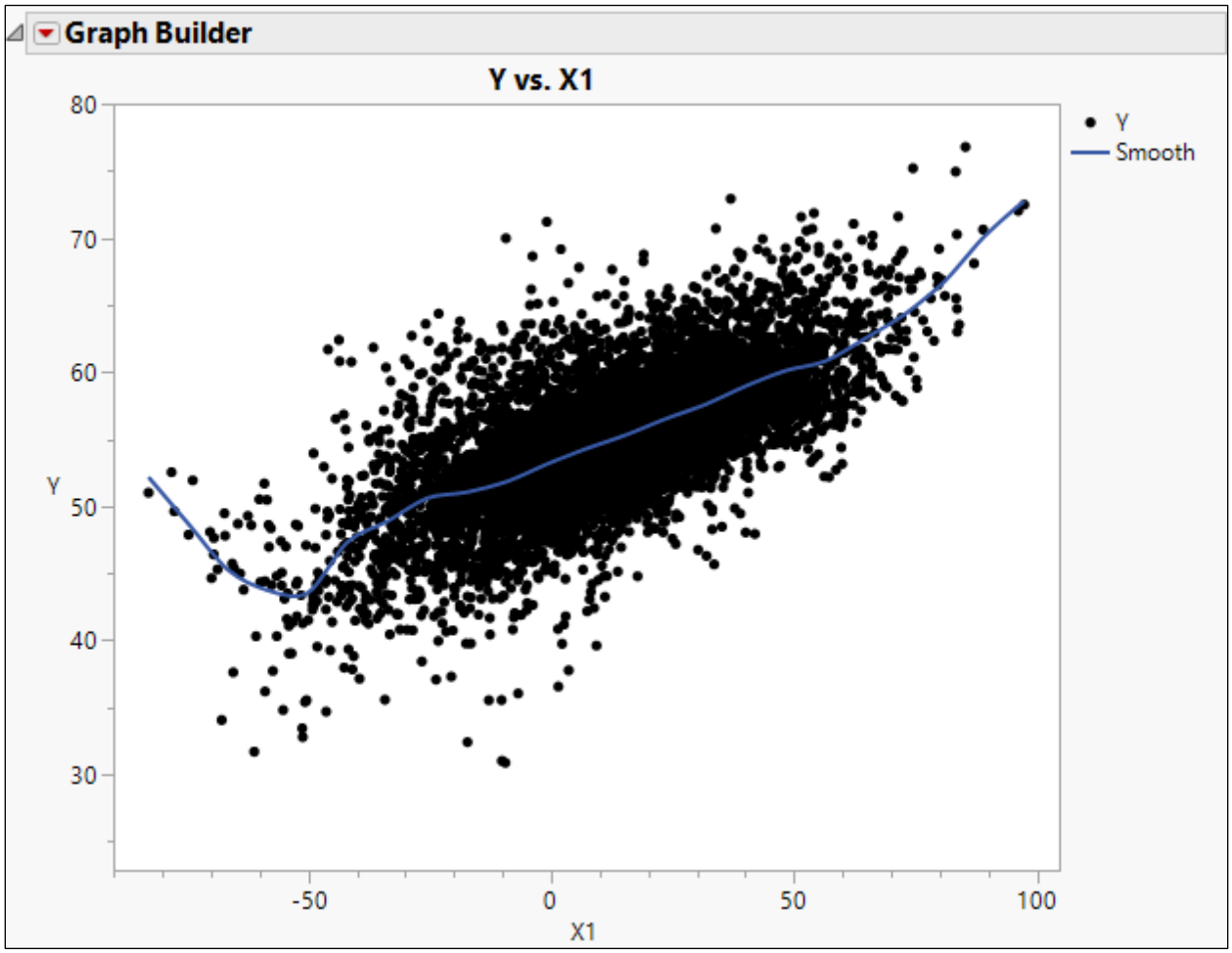


14) Click **OK**.

Build a scatter plot

1) Select **Graph** > **Graph Builder**.
2) Drag **Y** to the Y zone.
3) Drag **X1** to the X zone.

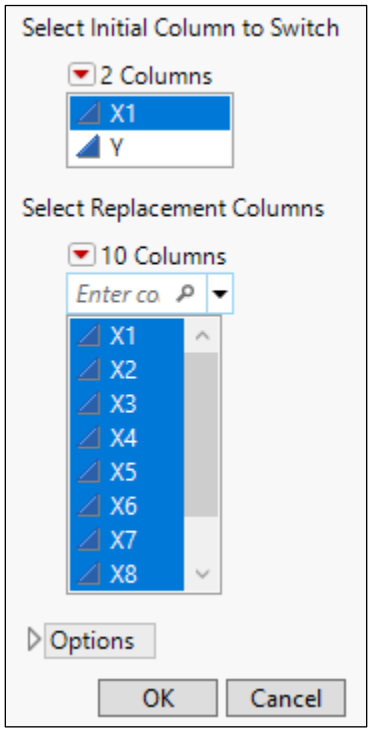
4) Click **Done**.



The relationship between Y and X1 appears strong, positive, and linear. Add a column switcher to reproduce this graph for other predictors.

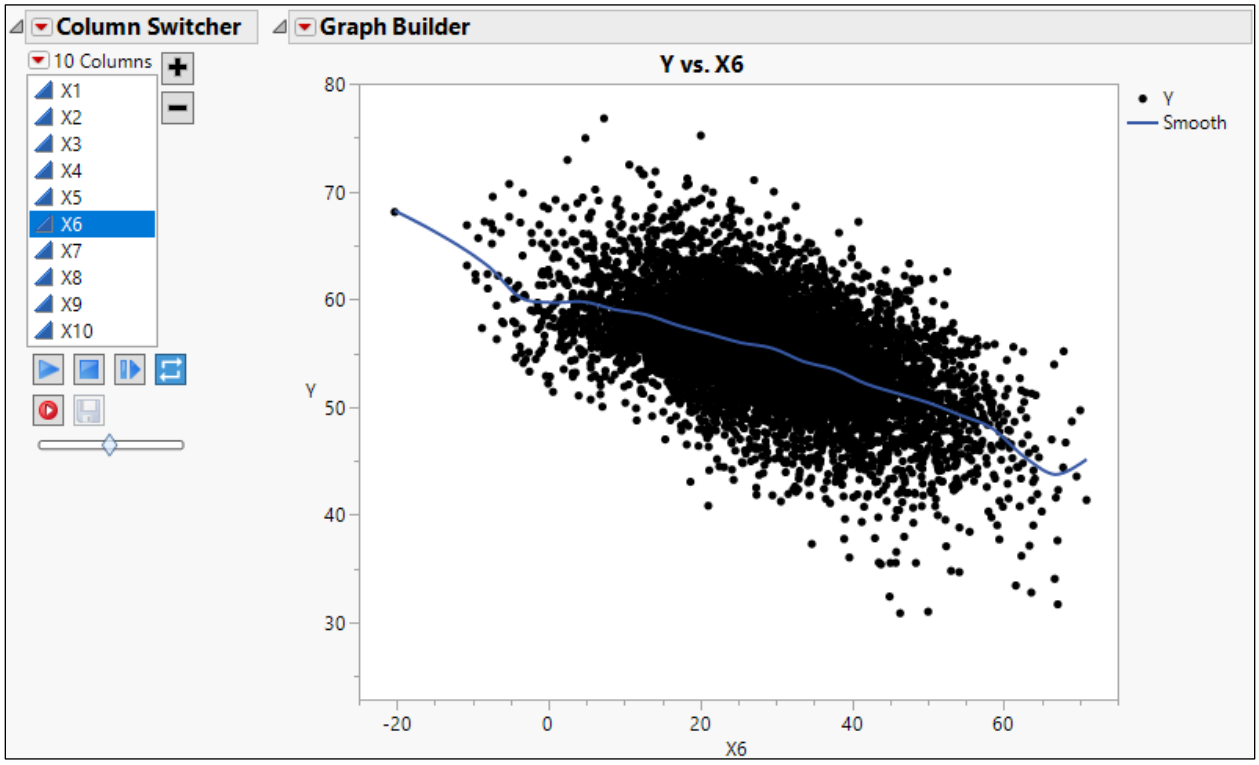
5) Click the red triangle next to **Graph Builder** and select **Redo** > **Column Switcher**.

6) Select all columns in the Select Replacement Columns box.



7) Click **OK**.

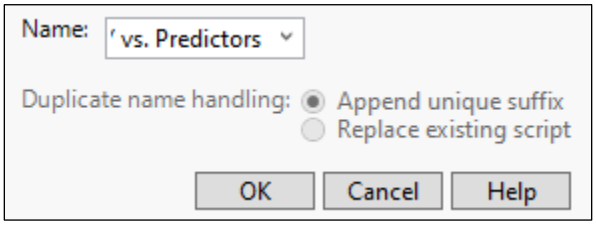
8) Click on each variable in the Column Switcher in turn.



Other relationships are also strong. Save the script to the data table.

9) Click the red triangle next to **Graph Builder** and select **Save Script** > **To Data Table**.

10) Enter **Y vs. Predictors** for the name.



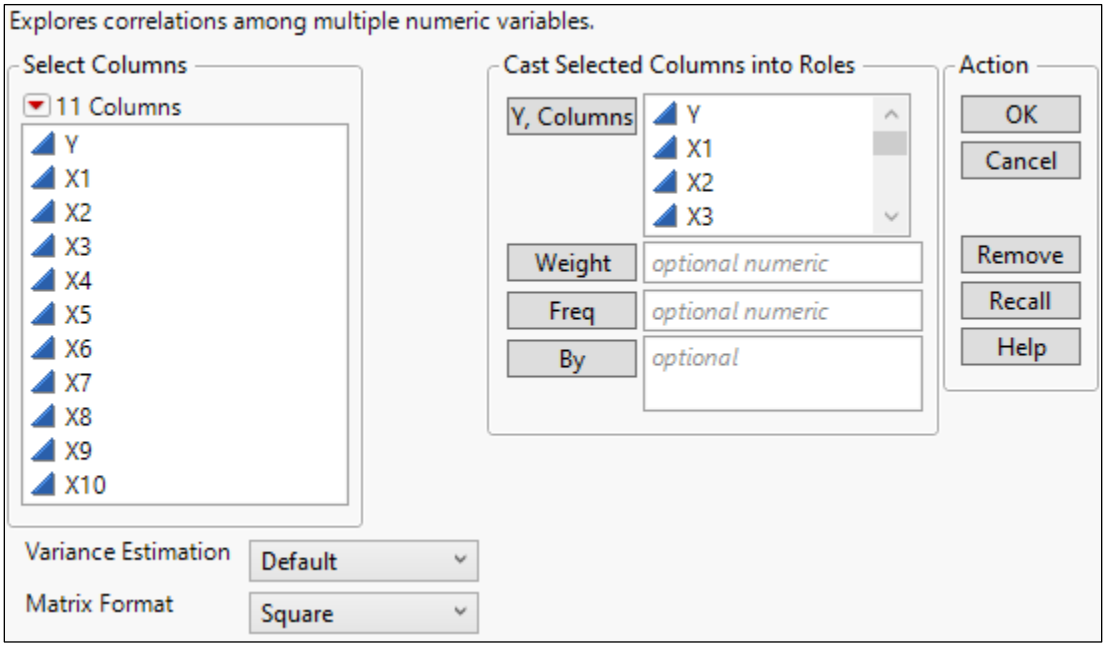
11) Click **OK**.

Another way to view the relationships among variables is with the Multivariate platform.

Multivariate

1) Select **Analyze** > **Multivariate Methods** > **Multivariate**.

2) Select all columns, then click **Y, Columns**.

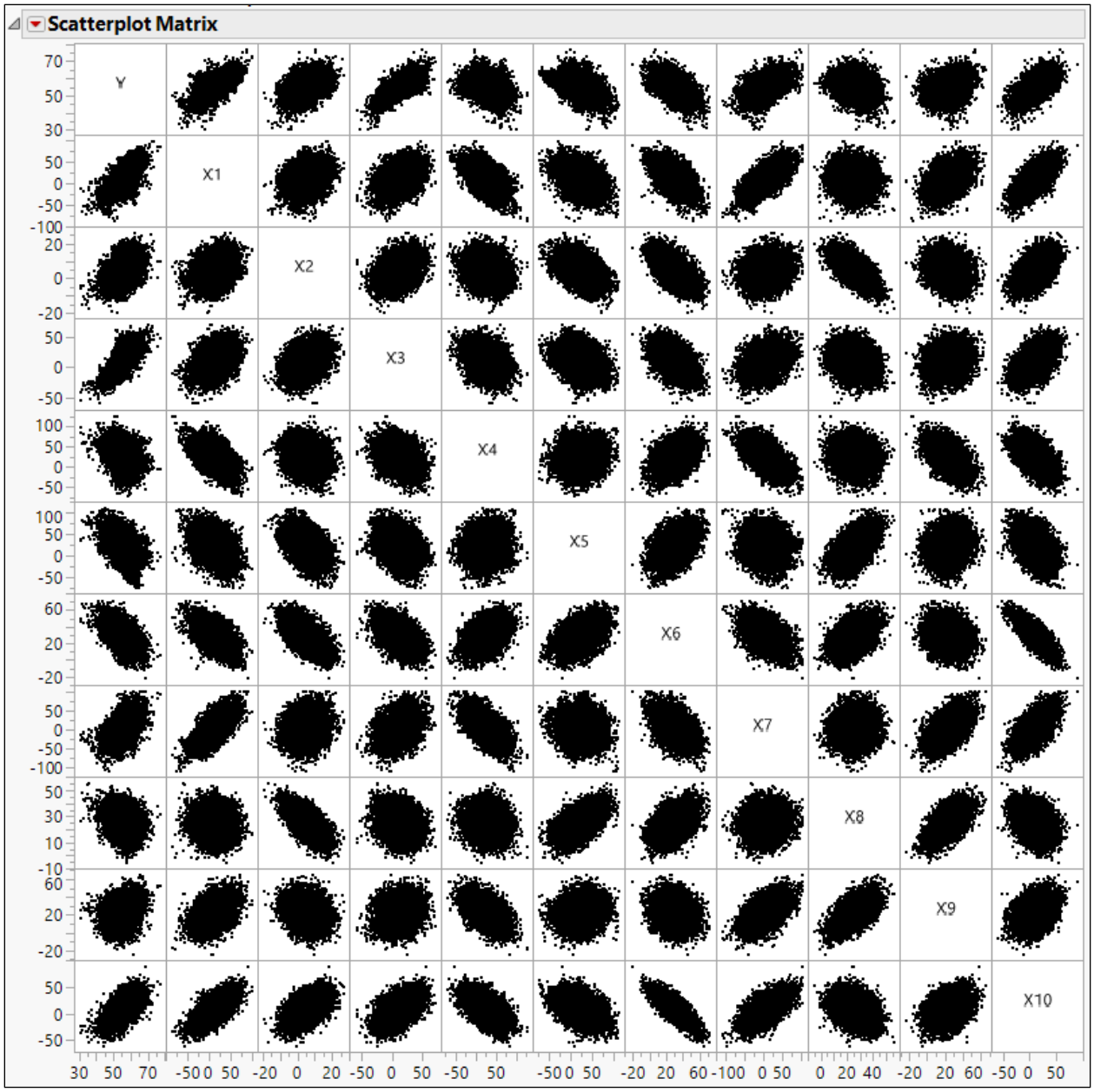


3) Click **OK**.

Correlations

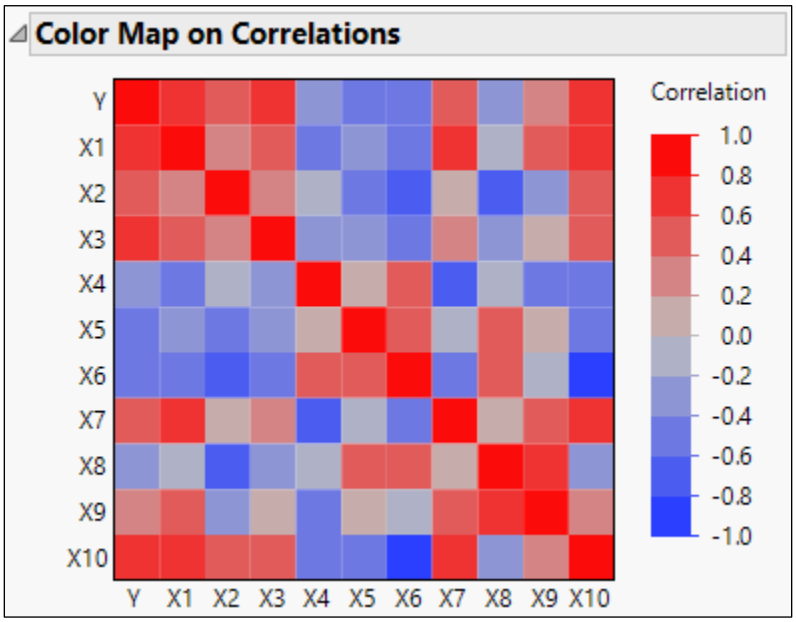
	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Y	1.0000	0.6719	0.4120	0.7486	-0.2675	-0.4234	-0.5353	0.5048	-0.2367	0.2372	0.6006
X1	0.6719	1.0000	0.3036	0.4164	-0.5867	-0.3466	-0.5931	0.7058	-0.0762	0.4266	0.6786
X2	0.4120	0.3036	1.0000	0.3444	-0.1384	-0.4566	-0.6195	0.1828	-0.6868	-0.2185	0.5397
X3	0.7486	0.4164	0.3444	1.0000	-0.3218	-0.3054	-0.4937	0.3547	-0.2412	0.1735	0.5240
X4	-0.2675	-0.5867	-0.1384	-0.3218	1.0000	0.1029	0.4253	-0.6105	-0.1421	-0.5274	-0.5574
X5	-0.4234	-0.3466	-0.4566	-0.3054	0.1029	1.0000	0.4806	-0.1802	0.5904	0.1900	-0.4268
X6	-0.5353	-0.5931	-0.6195	-0.4937	0.4253	0.4806	1.0000	-0.5046	0.4988	-0.1932	-0.8238
X7	0.5048	0.7058	0.1828	0.3547	-0.6105	-0.1802	-0.5046	1.0000	0.1086	0.5619	0.6371
X8	-0.2367	-0.0762	-0.6868	-0.2412	-0.1421	0.5904	0.4988	0.1086	1.0000	0.6370	-0.3387
X9	0.2372	0.4266	-0.2185	0.1735	-0.5274	0.1900	-0.1932	0.5619	0.6370	1.0000	0.3705
X10	0.6006	0.6786	0.5397	0.5240	-0.5574	-0.4268	-0.8238	0.6371	-0.3387	0.3705	1.0000

There are many large positive (blue) and negative (red) correlations among the variables.



The same scatter plots that you created in Graph Builder can be seen all at once in the Scatterplot Matrix.

4) Click the red triangle next to **Multivariate** and select **Color Maps** > **Color Map on Correlations**.



This color map duplicates the information in the correlation matrix but might be easier to read.

No data problems other than the autocorrelation, are detected. Save the script to the data table.

5) Click the red triangle next to **Multivariate** and select **Save Script** > **To Data Table**.

6) Click **OK**.

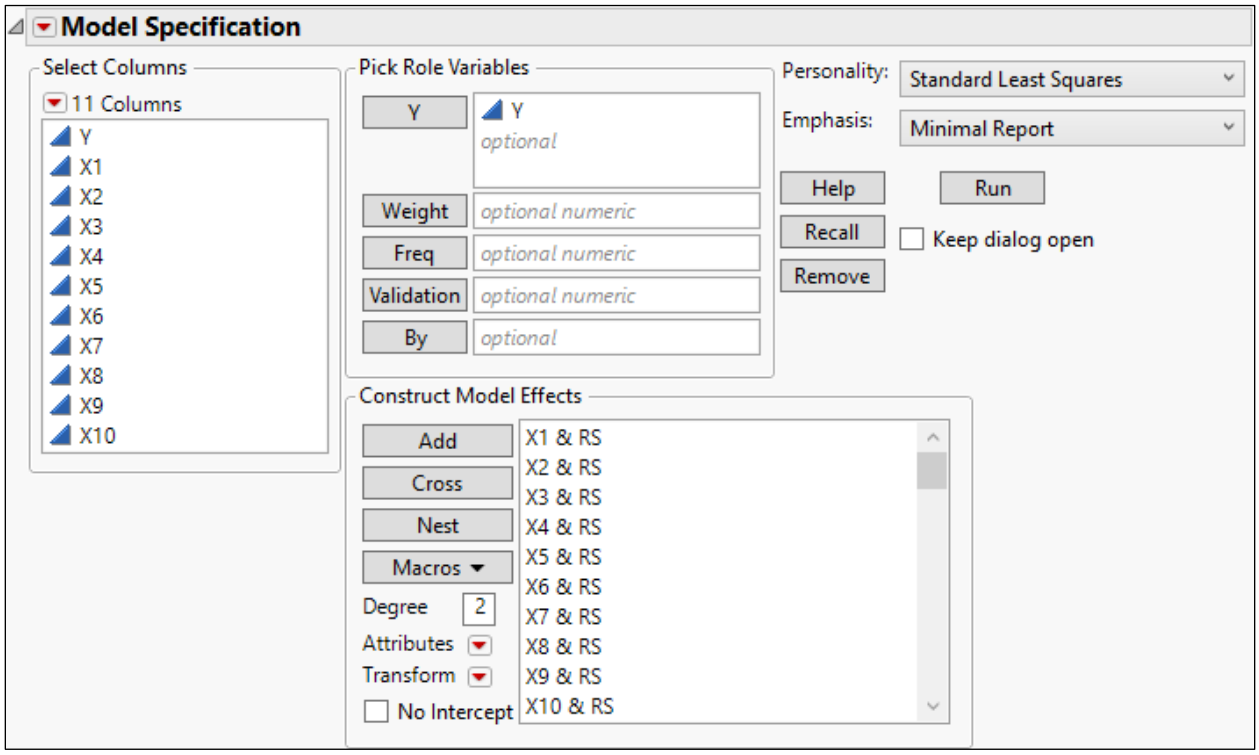
7) Save the data table.

b. Build a response surface model using Fit Model and the Response Surface macro. What is the R^2 of the full model? Are there any problems with the model fit seen in the Residual by Predicted plot? Which predictor variables are most important for predicting Y? Save the prediction formula to the data table.

1) Select **Analyze** > **Fit Model**.

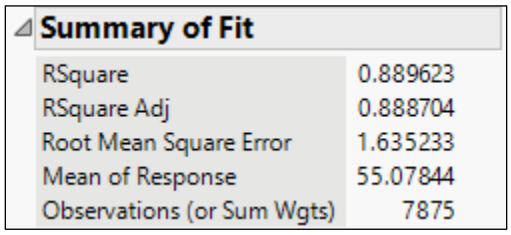
2) Select **Y**, then click **Y**.

3) Select **X1** through **X10**, then click **Macros** > **Response Surface**.



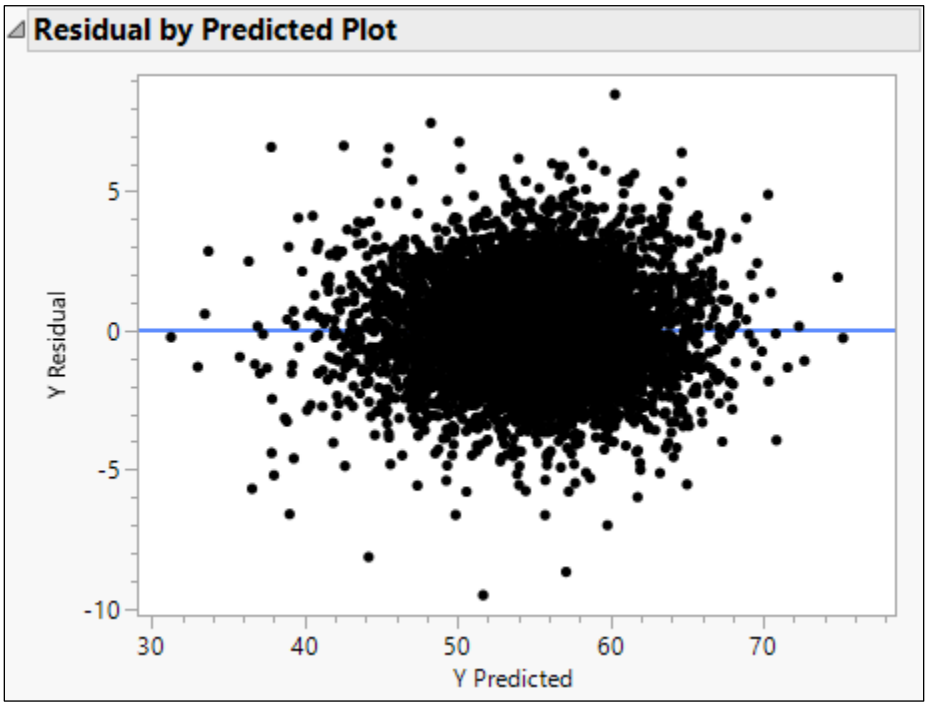
4) Click **Run**.

5) Close the Effect Summary report.



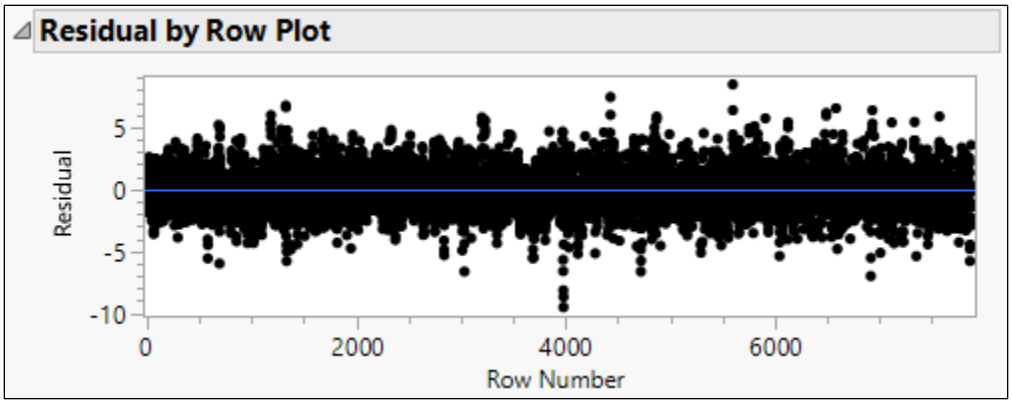
R^2 is about 89%, meaning that 89% of the variability in Y is explained by this model. One standard deviation of the unexplained variability is about 1.6. Check model assumptions.

6) Click the red triangle next to **Response Y** and select **Row Diagnostics** > **Plot Residual by Predicted**.



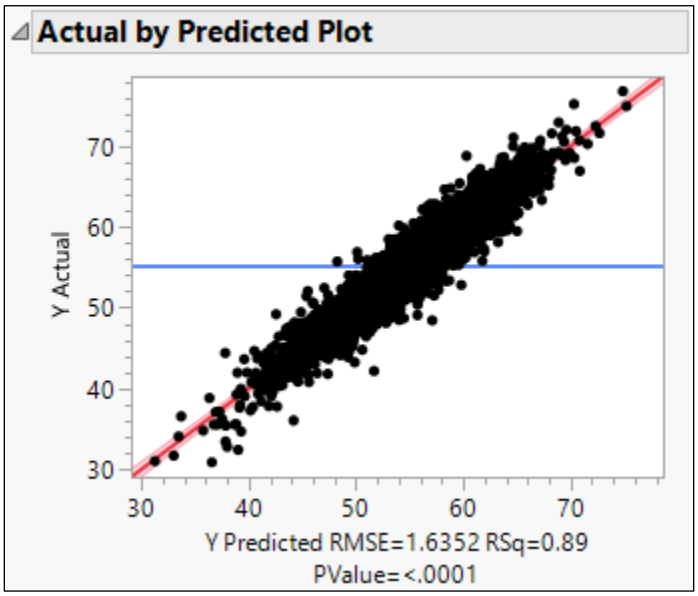
No problems are seen in the residual plot. Plot the residuals in time order (row order).

7) Click the red triangle next to **Response Y** and select **Row Diagnostics** > **Plot Residual by Row**.



Autocorrelation in the residuals is also seen. Examine the Actual by Predicted plot to see how much the autocorrelation affects the predictions.

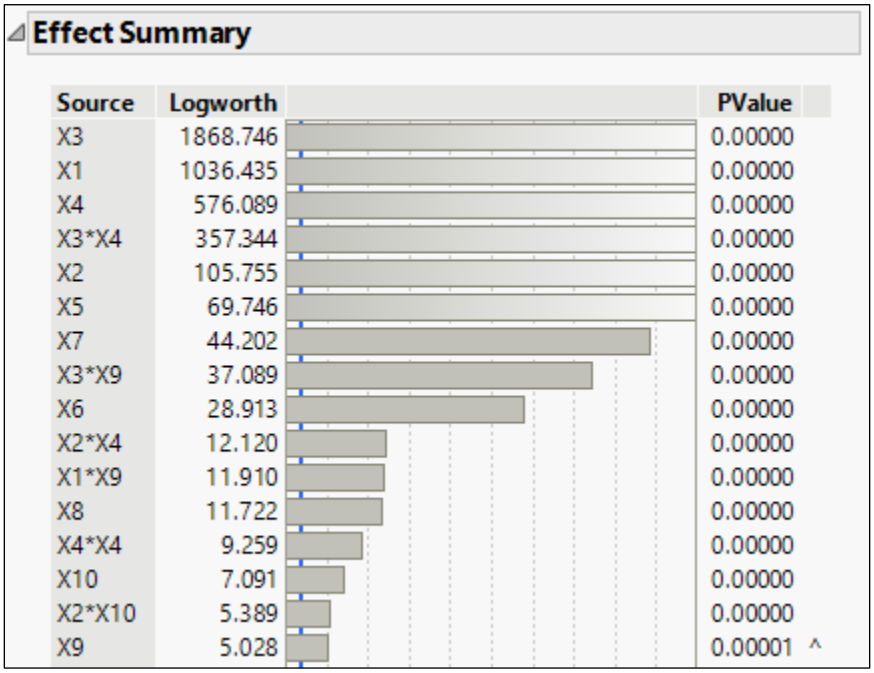
8) Click the red triangle next to **Response Y** and select **Row Diagnostics** > **Plot Actual by Predicted**.



No bias is seen in the predictions.

Because there are so many observations, looking at *p*-values to determine variable importance indicates that all variables are important. See the Parameter Estimates report. To find the most important variables, examine the Effect Summary report.

9) Open the Effect Summary outline.



Source	Logworth	PValue
X3	1868.746	0.00000
X1	1036.435	0.00000
X4	576.089	0.00000
X3*X4	357.344	0.00000
X2	105.755	0.00000
X5	69.746	0.00000
X7	44.202	0.00000
X3*X9	37.089	0.00000
X6	28.913	0.00000
X2*X4	12.120	0.00000
X1*X9	11.910	0.00000
X8	11.722	0.00000
X4*X4	9.259	0.00000
X10	7.091	0.00000
X2*X10	5.389	0.00000
X9	5.028	0.00001

The most important variables for predicting Y are X3, X1, X4, X2, and X5. All main effects and many two-factor interactions are important.

Save the prediction formula.

10) Click the red triangle next to **Response Y** and select **Save Columns** > **Prediction Formula**.

11) Return to the data table and rename the last column as **Pred Y RS**.

12) Save the data table.

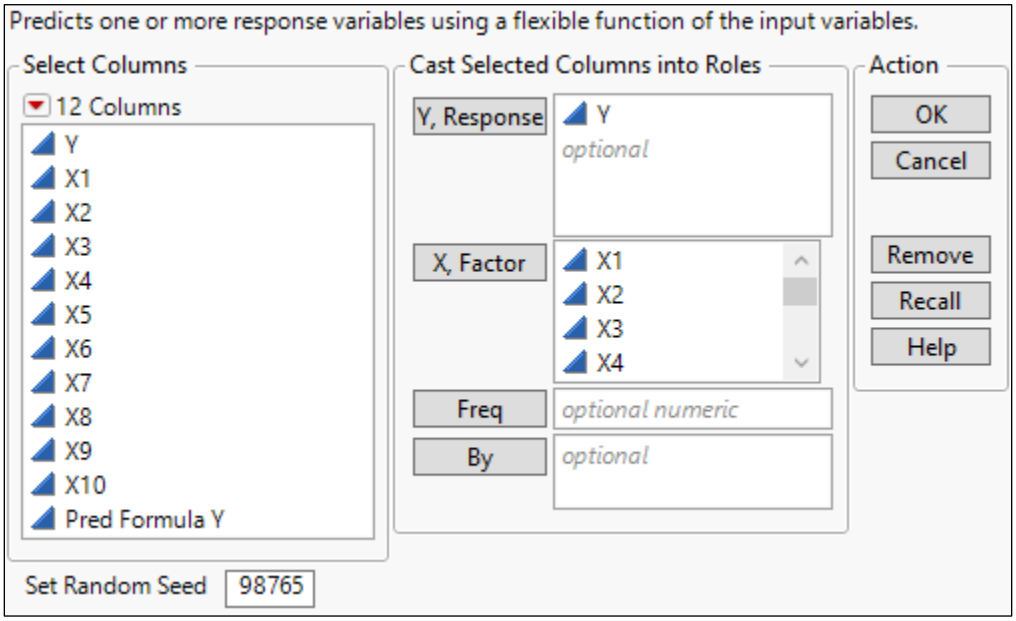
c. Build a neural network model using the default settings of the Neural platform. What is the R^2 of the model on the validation set? Are there any problems with the model fit seen in the Residual by Predicted plot? Fit another model with 50 nodes. Does this model have appreciably better predictive capability? Save the prediction formula to the data table.

1) Select **Analyze** > **Predictive Modeling** > **Neural**.

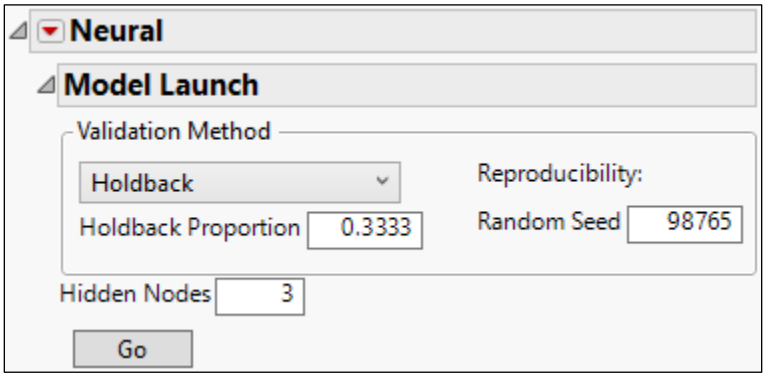
2) Select **Y**, then click **Y, Response**.

3) Select **X1** through **X10**, then click **X, Factor**.

4) If you want your results to match this solution, enter 98765 as the random seed.



5) Click **OK**.



6) Click **Go**.

Model NTanH(3)

Training

Y

Measures	Value
RSquare	0.8816073
RASE	1.697623
Mean Abs Dev	1.3222562
-LogLikelihood	10227.88
SSE	15130.1
Sum Freq	5250

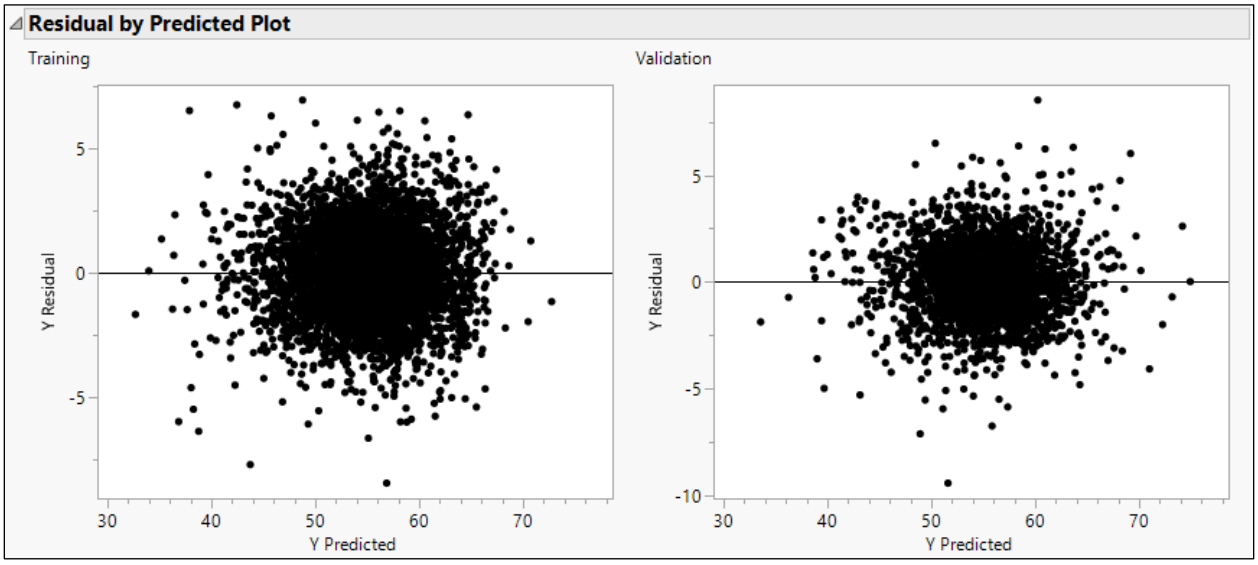
Validation

Y

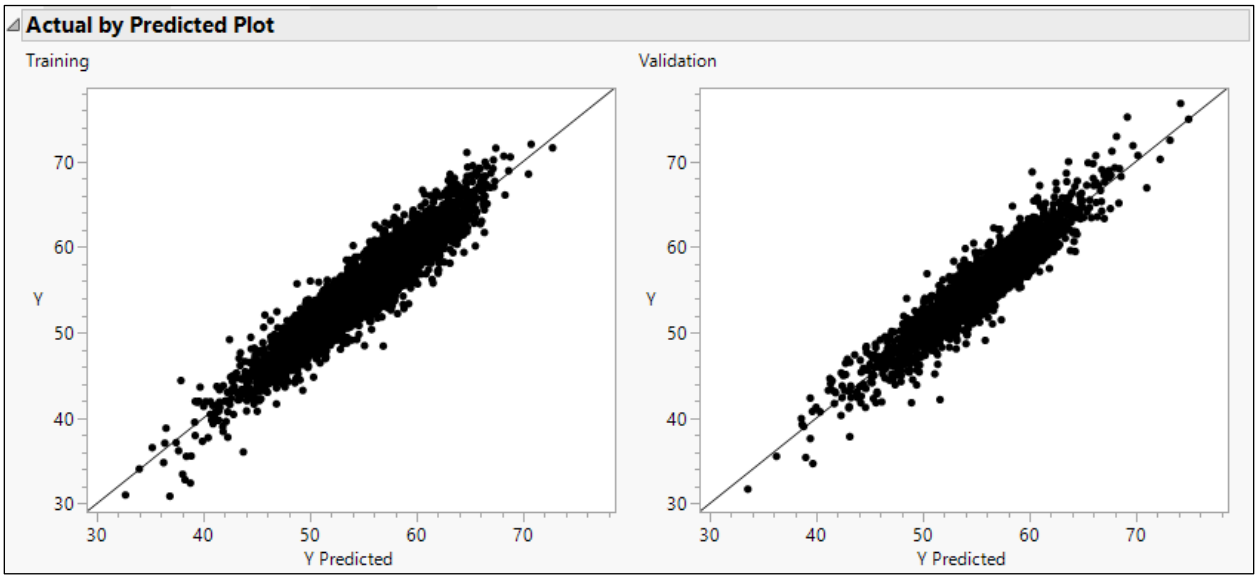
Measures	Value
RSquare	0.8809392
RASE	1.6680197
Mean Abs Dev	1.282874
-LogLikelihood	5067.7611
SSE	7303.5107
Sum Freq	2625

R^2 on the validation set is about 88%. Examine the model diagnostics.

7) Click the red triangle next to **Model NTanH(3)** and select **Plot Residual by Predicted**.



8) Click the red triangle next to **Model NTanH(3)** and select **Plot Actual by Predicted.**



No model problems are evident. Fit a model with more nodes to see if it improves the fit.

9) Open the **Model Launch** outline.

10) Enter 50 in the Hidden Nodes box.

11) Click **Go**.

Model NTanH(50)

Training

Y

Measures	Value
RSquare	0.8998253
RASE	1.5615568
Mean Abs Dev	1.2157681
-LogLikelihood	9789.2644
SSE	12801.913
Sum Freq	5250

Validation

Y

Measures	Value
RSquare	0.8895211
RASE	1.6067805
Mean Abs Dev	1.2409576
-LogLikelihood	4969.5739
SSE	6777.0766
Sum Freq	2625

R^2 on the validation set has increased by less than one percent. RASE and Mean Abs Dev are lower in the second decimal place. The model is not appreciably better.

Save the prediction formula from the model with three nodes.

Note: JMP Pro has the capability of adding nodes of other activation functions. If you have nonstationary data collected over time, that is, if the mean is drifting over time, it is recommended to add one linear node as well. For this data set, because the autocorrelation is stationary, adding one (or more) nodes with linear activation functions does not improve the predictive power of the model appreciably.

Save the prediction formula for the simpler model, then save the analysis.

12) Click the red triangle next to **Model NTanH(3)** and select **Save Profile Formulas**.

13) Return to the data table and rename the last column as **Pred Y NN**.

14) Return to the Neural report, then click the red triangle next to **Neural** and select **Save Script** > **To Data Table**.

15) Click **OK**.

16) Save the data table.

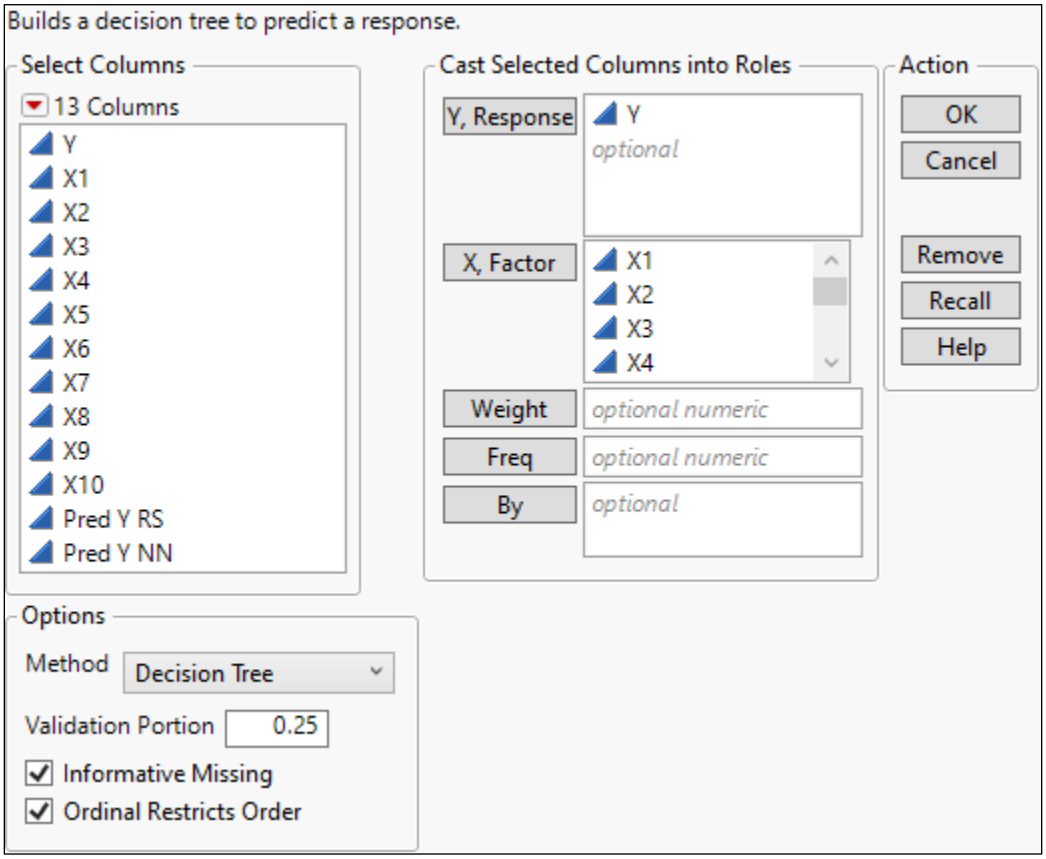
d. Build a decision tree model using the Partition platform with a 25% validation set. What is the R^2 of the model on the validation set? How many splits are in the model? Are there any problems with the model fit seen in the Actual by Predicted plot? Which variables are most important for predicting Y? Save the prediction formula to the data table.

1) Select **Analyze** > **Predictive Modeling** > **Partition**.

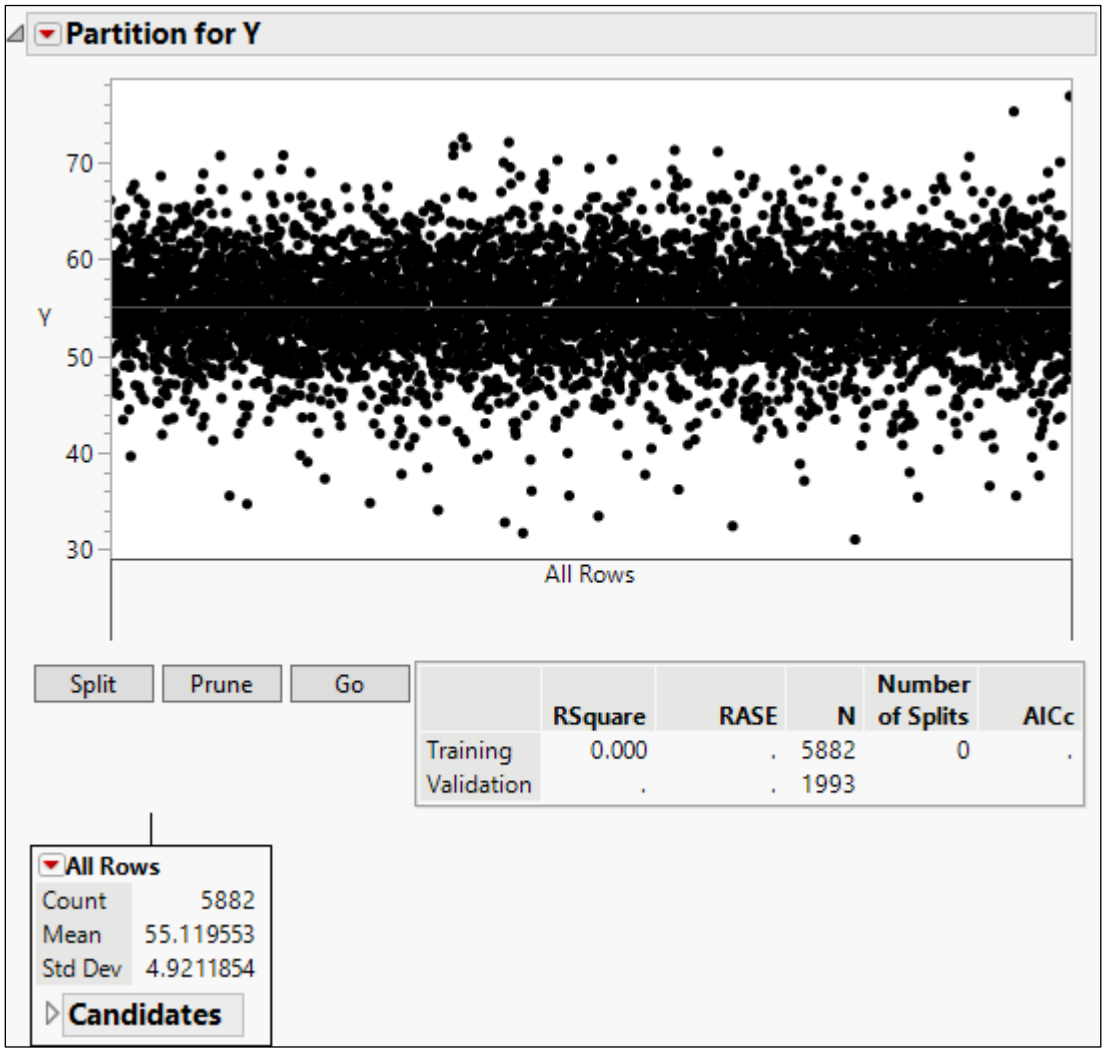
2) Select **Y**, then click **Y, Response**.

3) Select **X1** through **X10**, then select **X, Factor**.

4) For Validation Portion, enter 0.25.

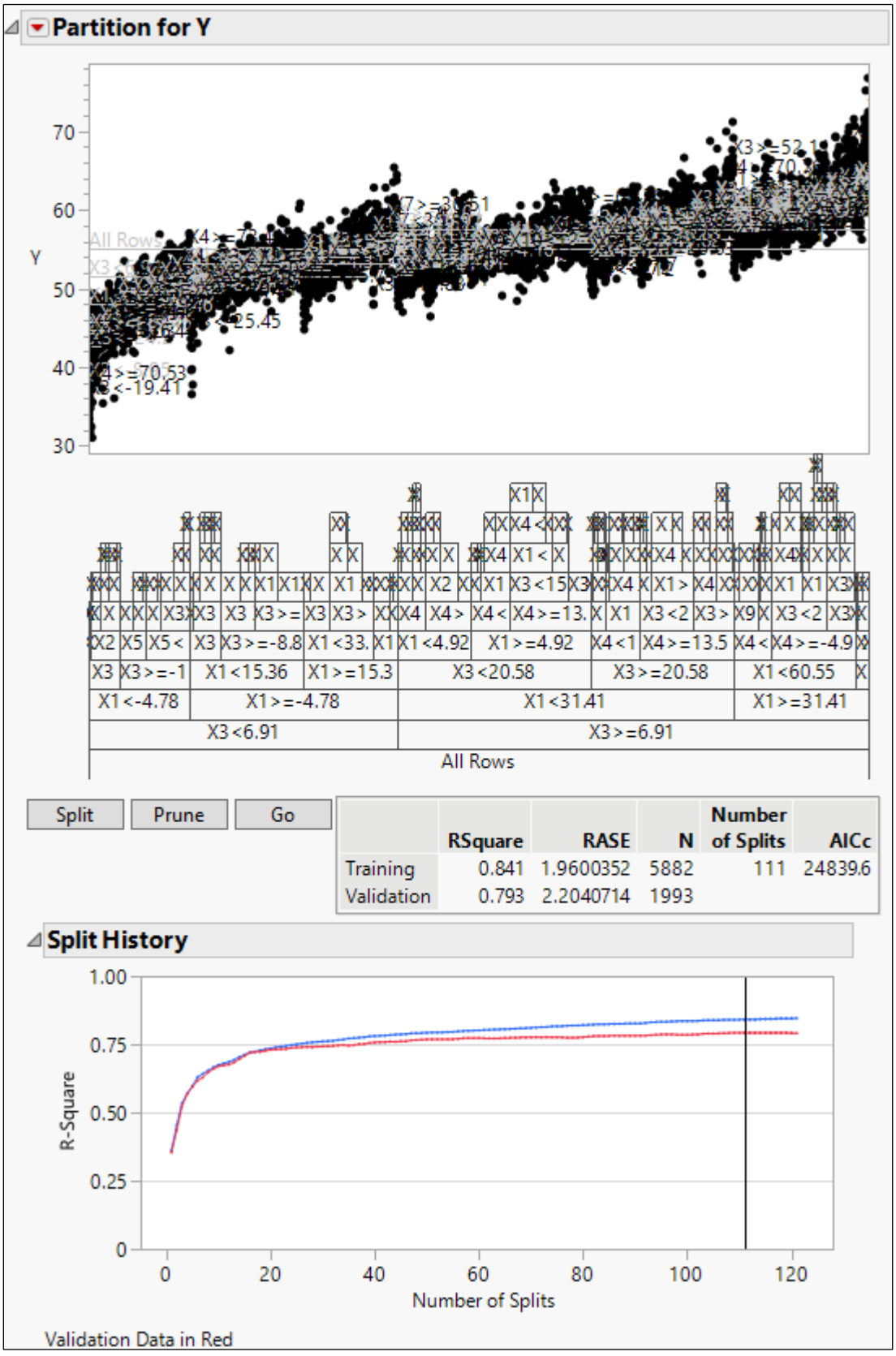


5) Click **OK**.



Your results will be slightly different because your validation data are different.

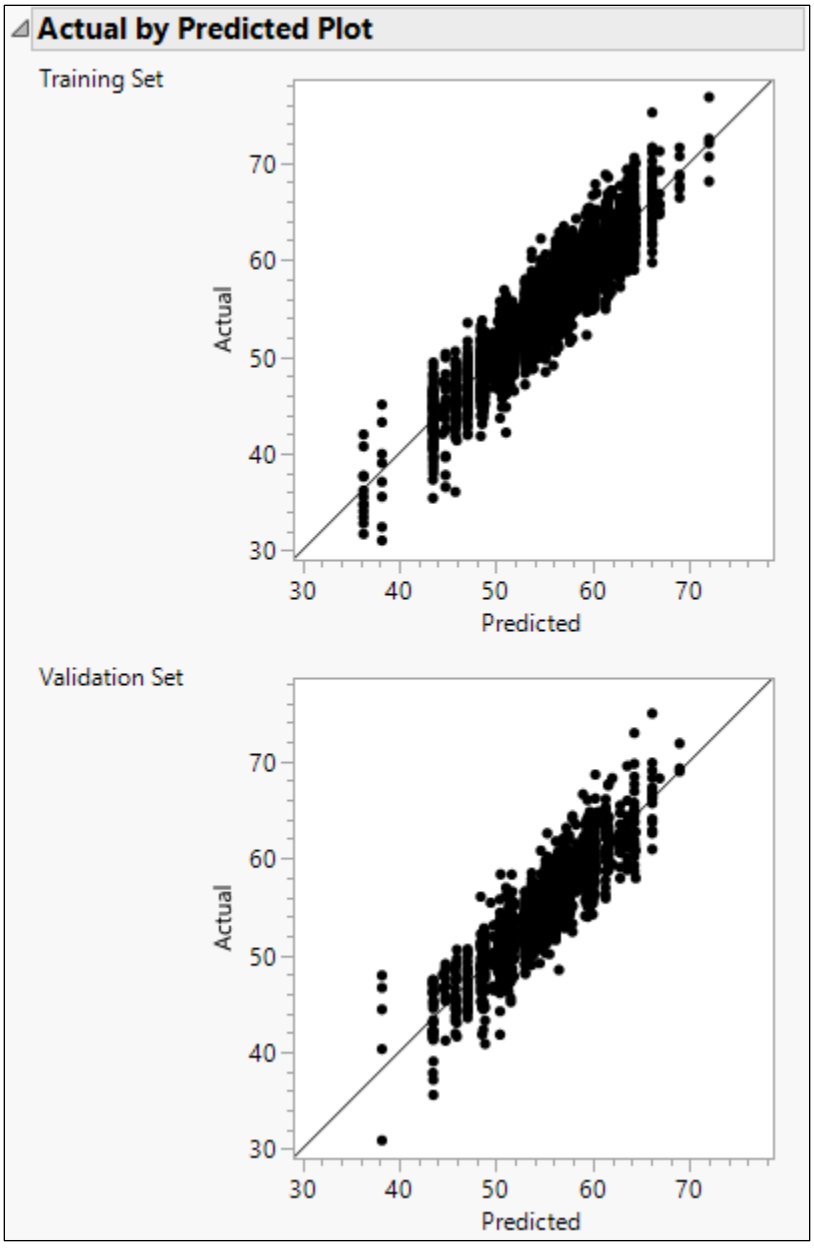
6) Click **Go**.



The data are split until the next 10 splits do not improve R^2 on the validation set, then the model is pruned back. R^2 on the validation set is about 79%, a decrease from the linear model or the neural network model. There were 111 splits.

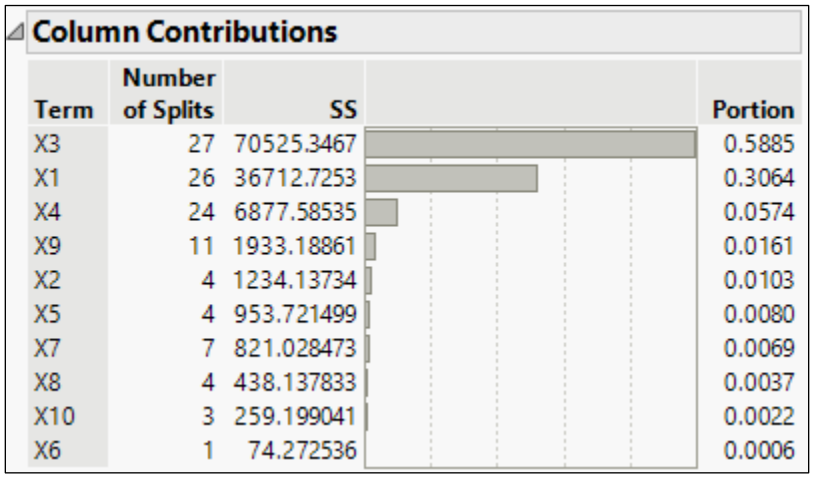
Assess the model.

7) Click the red triangle next to **Partition for Y**, then select **Plot Actual by Predicted**.



There are no problems evident in the model fit.

8) Click the red triangle next to **Partition for Y**, then select **Column Contributions**.



Column Contributions

Term	Number of Splits	SS	Portion
X3	27	70525.3467	0.5885
X1	26	36712.7253	0.3064
X4	24	6877.58535	0.0574
X9	11	1933.18861	0.0161
X2	4	1234.13734	0.0103
X5	4	953.721499	0.0080
X7	7	821.028473	0.0069
X8	4	438.137833	0.0037
X10	3	259.199041	0.0022
X6	1	74.272536	0.0006

The report is ordered by the portion of variability in Y explained by the predictors. X3, X1, and X4 are the top three predictors.

Save the prediction formula.

9) Click the red triangle next to **Partition for Y**, then select **Save Columns** > **Save Prediction Formula**.

10) Click the red triangle next to **Partition for Y**, then select **Save Script** > **To Data Table**.

11) Click **OK**.

12) Return to the data table and rename the last column to **Pred Y DT**.

13) Save the data table.

e. Build an Actual by Predicted graph for the three models using Graph Builder. Which model do you prefer?

1) Select **Graph** > **Graph Builder**.

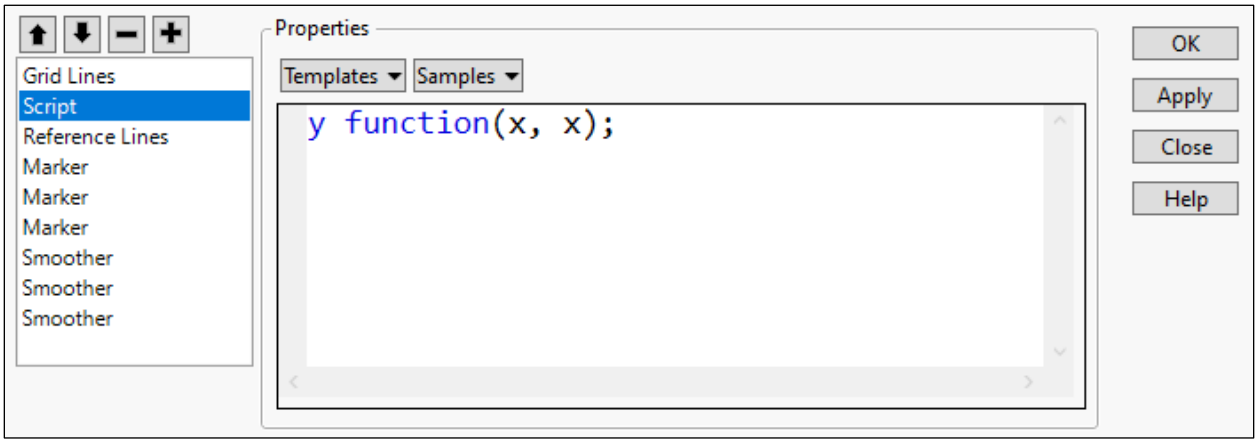
2) Drag **Y** to the Y zone.

3) Drag **Pred Y RS**, **Pred Y NN**, and **Pred Y DT** to the X zone together.

4) Right-click in the graph and select **Customize**.

5) Click the plus sign icon  to add a new graphics script.

6) Enter `y function(x, x);` in the script editor box.

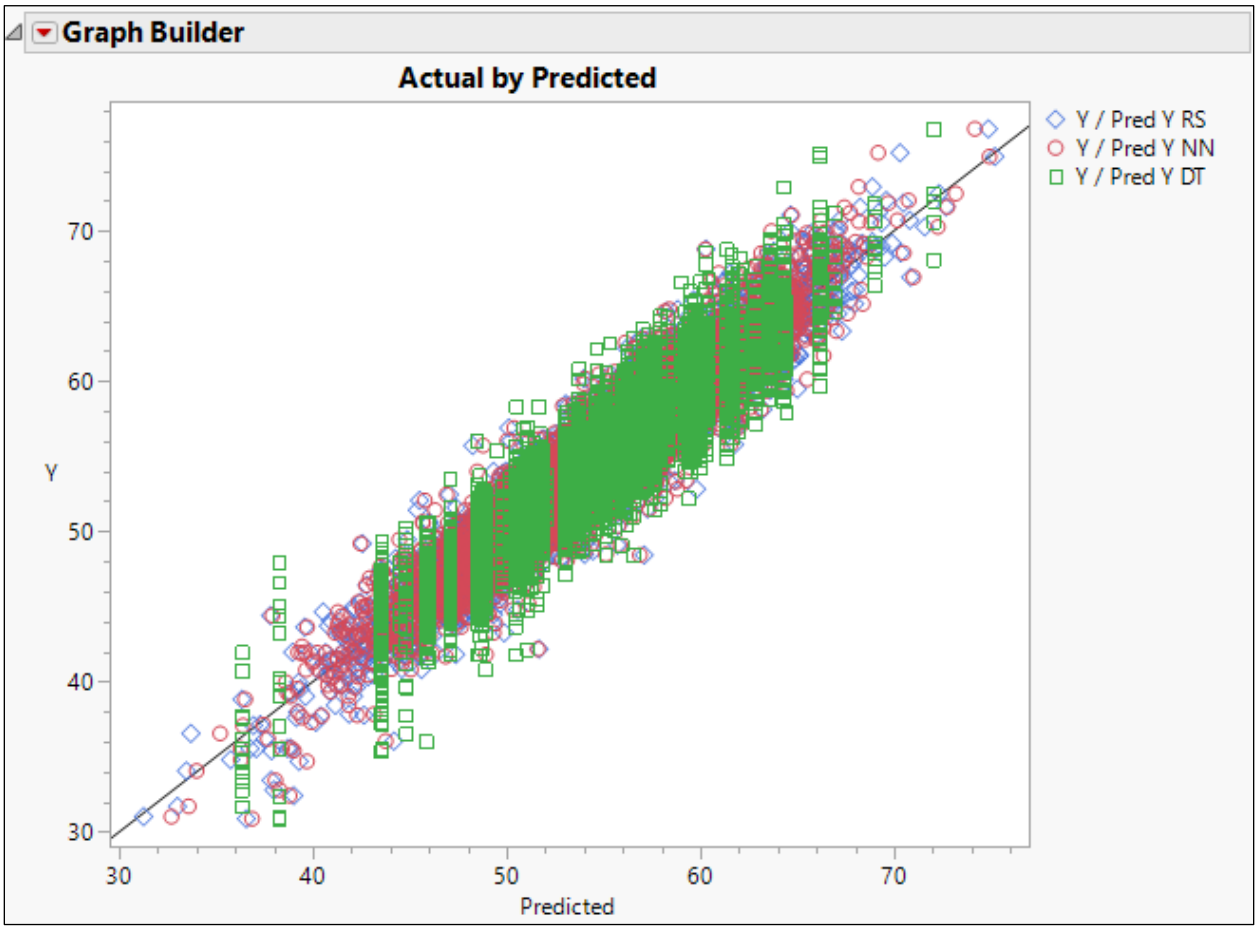


7) Click **OK**.

8) Remove the smooth curve.

9) Click **Done**.

10) If you are going to publish this graph, change the graph title to Actual by Predicted, change the X axis title to Predicted, and change the markers for each X variable.



The fits are fairly similar. The decision tree model gives chunky predictions because it is using discrete leaves to predict a continuous response.

Categorical response, continuous and categorical predictors

2. Use the data in PM 2.jmp to build predictive models.

a. Visualize the data using Graph Builder. Are there any data problems?

1) Open **PM 2.jmp**.

PM 2
Notes Simulated data
Columns (25/0)
Y *
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12
X13
X14
X15
X16
X17
X18
X19
X20
X21
X22
X23
X24
Rows
All rows 2,751
Selected 0
Excluded 0
Hidden 0
Labeled 0

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24
1	Very Go...	18.89	28.78	5.61	21.17	28	4.84	8.65	20.14	-9.18	7.05	13.81	-7.55	Red	15.72	11.91	27.14	25.2	B	28.22	-7.72	26.49	-0.84	7.85	19.79
2	Perfect	31.73	24.92	11.45	21.38	20.56	-0.52	20.43	15.03	-2.26	10.94	2.52	-2.07	White	30.21	20.87	15.85	22.56	A	36.83	-19.94	25.69	-12.71	-4.28	29.56
3	Good	13.18	31.32	8.19	19.8	22.99	4.48	18.42	16.71	1.36	10.61	14.55	0.89	White	21.99	31.39	18.87	25.13	B	15.1	8.76	23.81	20.6	-0.99	23.6
4	Good	24.57	19.63	-0.18	17.72	26.89	-7.64	15.46	12.74	-0.65	8.65	7.36	-10.02	White	21.45	12.42	25.61	19.56	C	43.76	-12.88	28.99	-5.92	0.58	33.93
5	Very Go...	24.23	26.39	2.73	25.24	27.08	5.81	16.45	20.82	-9.03	9.52	3.81	-6.39	White	21.01	14.05	20.22	26.25	A	28.53	-1.93	26.73	3.24	4.01	21.38
6	Good	9.78	30.65	3.84	17.08	27.43	-0.39	15.17	23.23	4.95	11.91	19.7	-3.25	White	17.76	19.99	17.72	24.45	B	20.73	9.47	23.46	10.87	0.78	25.83
7	Very Go...	26.74	33.69	9.11	21.29	25.39	9.81	25.62	20.52	0.07	11.51	17.4	2.82	White	17.42	13.47	17.74	30.72	A	12.92	-0.35	23	22.96	5.93	1.28
8	Good	10.33	26.36	1.51	2.16	28.45	2.12	7.27	6.64	5.38	13.29	15.84	-10.17	White	18.53	22.66	20.95	22.58	C	20.79	-0.85	22.08	13.13	0.2	30.09
9	Perfect	28.31	18.62	3.51	24.49	20.72	-7.69	13.13	-4.24	2.33	13.92	4.25	-4.42	White	27.72	21.25	20.82	18.94	A	39.96	-19.77	22.59	-5.88	-8.14	21.9
10	Very Go...	25.95	21.77	5.06	21.84	22.41	8.24	19.33	10.8	1.63	11.94	2.41	0.11	White	24.93	23.67	17.84	26.78	A	21.65	-9	24.6	13.63	-0.81	15.68
11	Good	21.6	26.75	4.46	11.76	24.61	8.44	29.08	28.39	-5.49	4.74	5.98	-0.61	Red	15.8	18.85	18.96	26.56	B	17.82	3.01	28.47	16.28	10.22	29.37
12	Good	29.63	25.14	5.43	16.21	24.73	-1.78	28.88	17.33	-4.85	9.88	0.82	-6.67	White	21.22	17.71	19.34	22.54	B	32.01	-9.11	25.65	6.93	1.19	30.27
13	Good	6.79	32.53	4.56	5.46	30.75	2.81	3.46	17.92	0.84	13.48	21.88	-8.04	White	17.7	17.17	22.17	25.42	C	26.33	5.31	22.23	15.94	2.73	22.82
14	Very Go...	16.13	34.89	4	18.66	22.73	4.43	29.44	12.94	6.44	14.36	16.28	-3.53	Blue	24.97	20.47	19.1	25.48	A	30.14	-7.15	19.9	6.48	-2.96	8.56
15	Very Go...	13.65	37.54	3.04	21.1	26.88	5.09	26.56	21.64	-4.95	9.54	17.86	-3.38	White	17.17	17.53	21.39	31.45	A	19.22	3.02	23.13	21.2	9.71	11.6
16	Perfect	32.08	27.1	16	33.29	19.89	8.1	37.4	28.77	0.28	8.86	-0.29	5.73	White	23.7	27.06	16.23	26.51	A	20.13	-5.74	25.46	2.64	-0.27	22.8
17	Very Go...	30.7	27.3	8.03	26.84	19.09	6.85	43.92	11.32	2.15	14.22	7.59	0.86	Blue	25.19	23.09	16.63	26.24	A	25.55	-17.37	22	7.57	-3.16	16.04
18	Very Go...	30.72	24.87	9.82	27.62	18.39	-6.47	27.28	20.48	0.54	8.07	-9.48	3.62	White	27.88	22.2	15.88	20.57	B	26.35	-12.58	26.94	2.56	-5.58	14.15
19	Good	18.07	29.7	5.99	20.32	22.14	9.81	31.28	26.73	-2.4	9.49	7.86	2.02	White	22.07	23.86	16.4	27.22	A	25.64	4.81	24.54	18.43	2.25	16.76
20	Very Go...	22.09	30.32	8.36	22.66	16.81	10.28	26.31	19.91	3.56	12.63	8.35	5.67	White	27.68	35.93	12.68	24.97	A	17.8	-9.04	20.1	4.05	-3.42	18.6
21	Good	25.33	36.93	18.32	16.43	15.47	9.77	18.73	40.02	-0.3	8.66	2.74	15.82	White	23.75	47.25	-7.37	29.51	A	-8.49	14.44	19.86	22.7	-1.42	18.89
22	Very Go...	16.23	28.08	-0.74	8.45	29.18	-1.67	17.68	21.07	-7.71	10.35	15.8	-9.62	White	17.8	7.58	30.19	28.13	B	29.31	-5.97	25.02	5.72	8.72	13.01
23	Good	9.61	27.25	-2.62	6.19	27.98	-0.53	14.23	2.63	-0.13	13.7	16.21	-11.02	White	19.67	15.07	20.29	23.12	B	25.59	-12.58	21.51	12.82	-0.04	15.06
24	Good	6.95	38.54	5.76	9.58	23.36	12.79	4.59	8.82	8.43	18.23	25.12	1.56	White	23.24	32.65	4.36	28.68	B	9.75	14.13	18.2	30.24	-4.28	12.84
25	Very Go...	14.02	27	3.52	24.77	27.01	2.21	30.84	5.09	-1.1	14.22	14.55	-4.22	White	23.03	10.83	18.97	26.12	A	34.32	-6.61	20.75	11.18	1.91	5.4
26	Very Go...	36.91	33.98	15.37	27.22	18.84	7.33	35.82	37.54	-5.83	5.24	6.5	9.2	White	16.22	27.94	9.39	27.83	A	20.99	-3.94	28.91	8.86	6.29	19.58
27	Very Go...	18.13	21.43	-3	10.48	30.01	-12.09	6.28	7.04	-8.33	9.21	5.35	-14.76	White	20.2	8.21	30.52	20.57	C	49.94	-16.59	27.11	-6.66	0.74	28.41
28	Very Go...	9.26	32.01	-5.75	15.35	30.27	7.26	27.66	18.03	1.81	12.61	24.84	-3.02	White	13.69	13.63	23.91	27.91	A	23.06	12.05	21.43	17.58	6.52	-1.54
29	Good	12.94	32.12	0.83	17.24	30.18	6.48	7.21	20.7	-4.27	11.66	24.02	-6.3	White	14.08	15.24	25.12	28.22	B	18.5	6.65	23.9	24.51	6.93	8.54

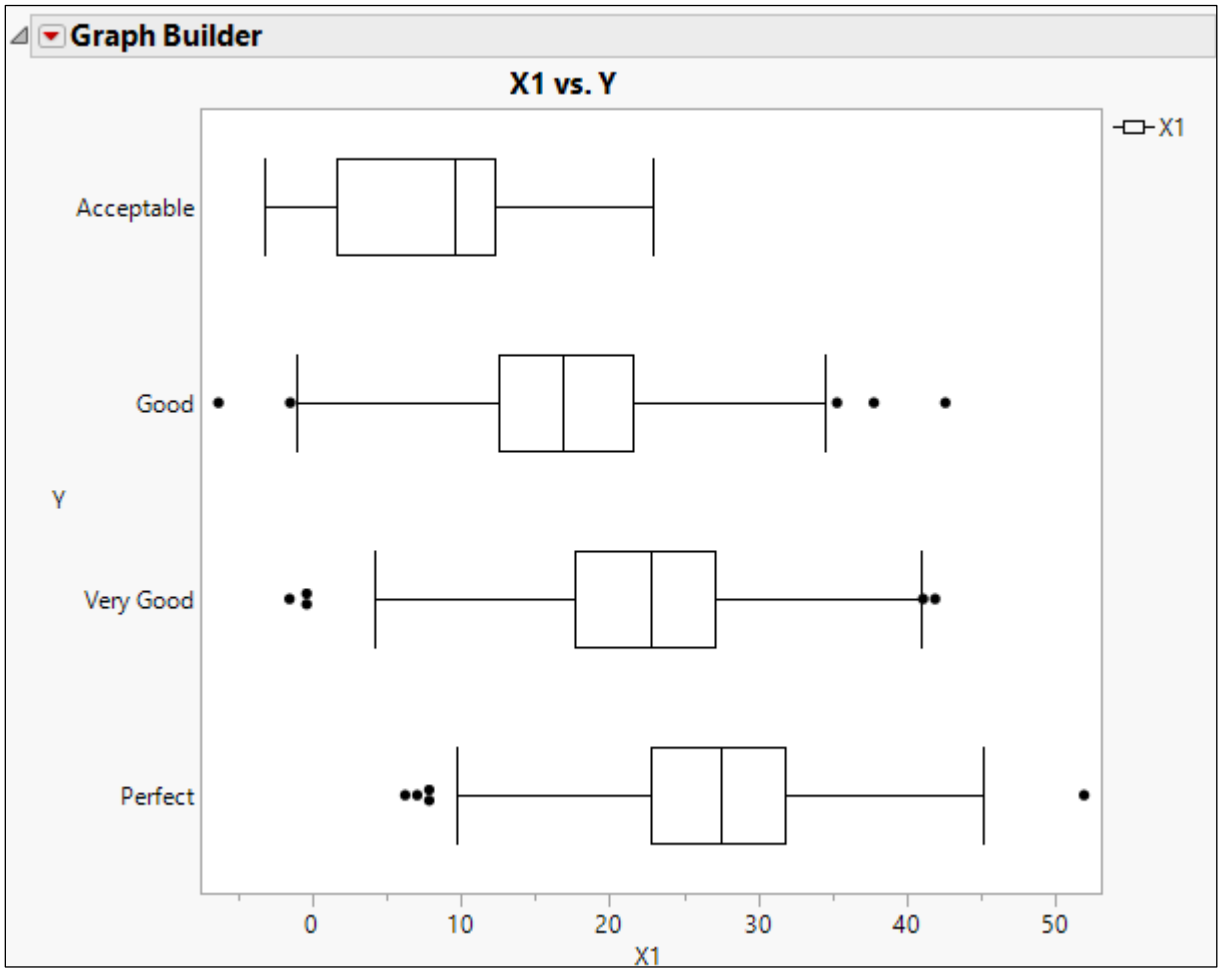
2) Select **Graph** > **Graph Builder**.

3) Drag **Y** to the Y zone.

4) Drag **X1** to the X zone.

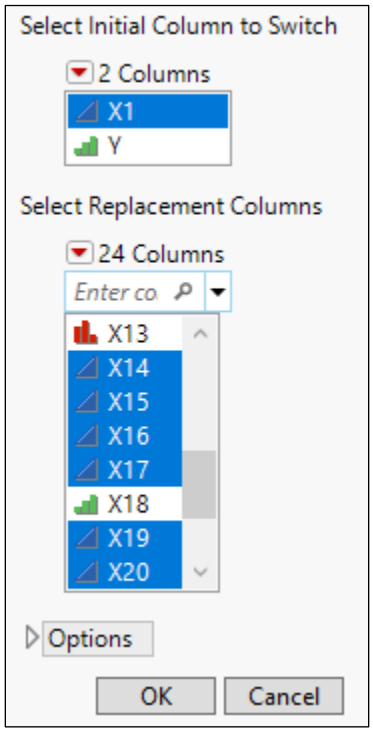
5) Select the Boxplot element on the Elements bar.

6) Click **Done**.

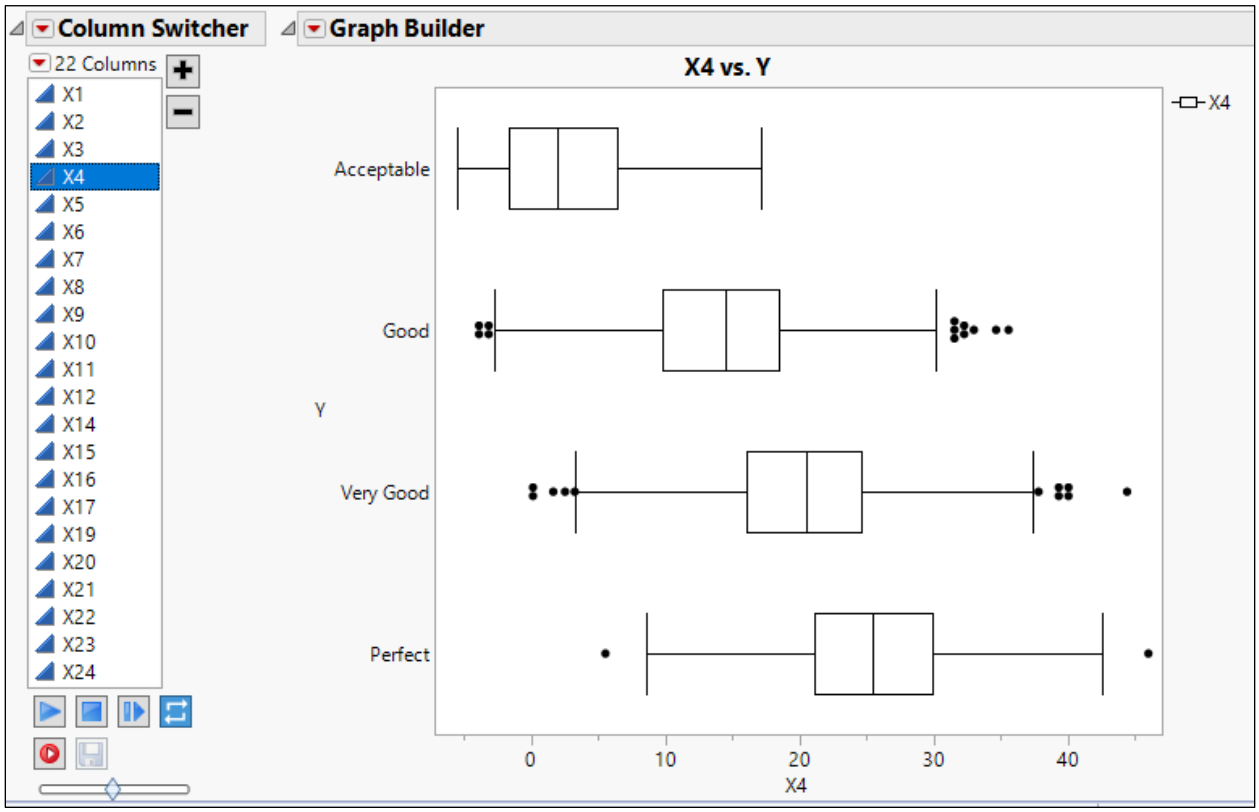


7) Click the red triangle next to **Graph Builder** and select **Redo** > **Column Switcher**.

8) In the Select Replacement Columns box, select all continuous columns.



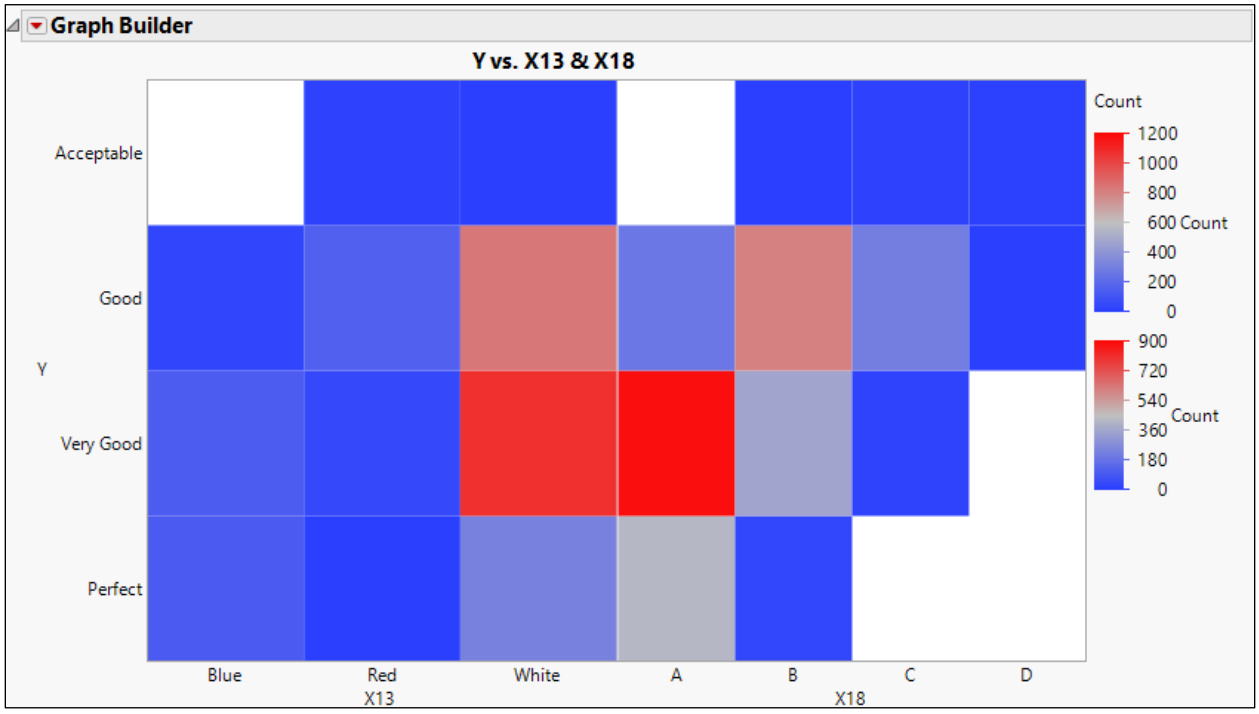
9) Click **OK**.



Several strong relationships between Y and the predictors can be seen.

Build visualizations for the categorical predictors.

10) Select **Graph** > **Graph Builder**.
11) Drag **Y** to the Y zone.
12) Drag **X13** to the X zone.
13) Drag **X18** to the X axis to the right of **X13**.
14) Select the Heatmap element from the Elements bar.
15) Click **Done**.



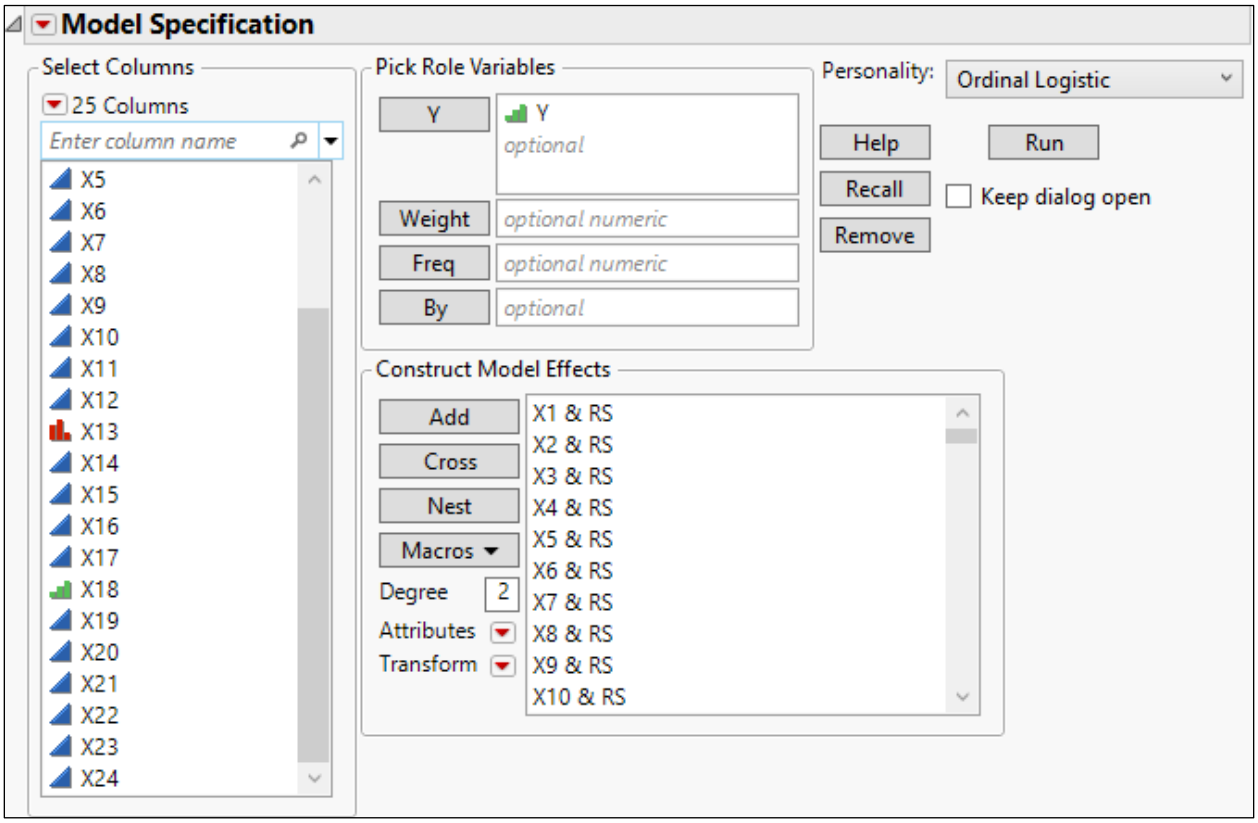
There appears to be an association between Y and both categorical predictors.

No data problems are evident in either graph.

b. Build an ordinal regression model using Fit Model and the Response Surface macro. What is the misclassification rate of the full model? Hint: open the Fit Details report. What is the misclassification rate for the Acceptable group? Hint: open the Confusion Matrix report.

1) Select **Analyze** > **Fit Model**.
2) Select **Y**, then click **Y**.

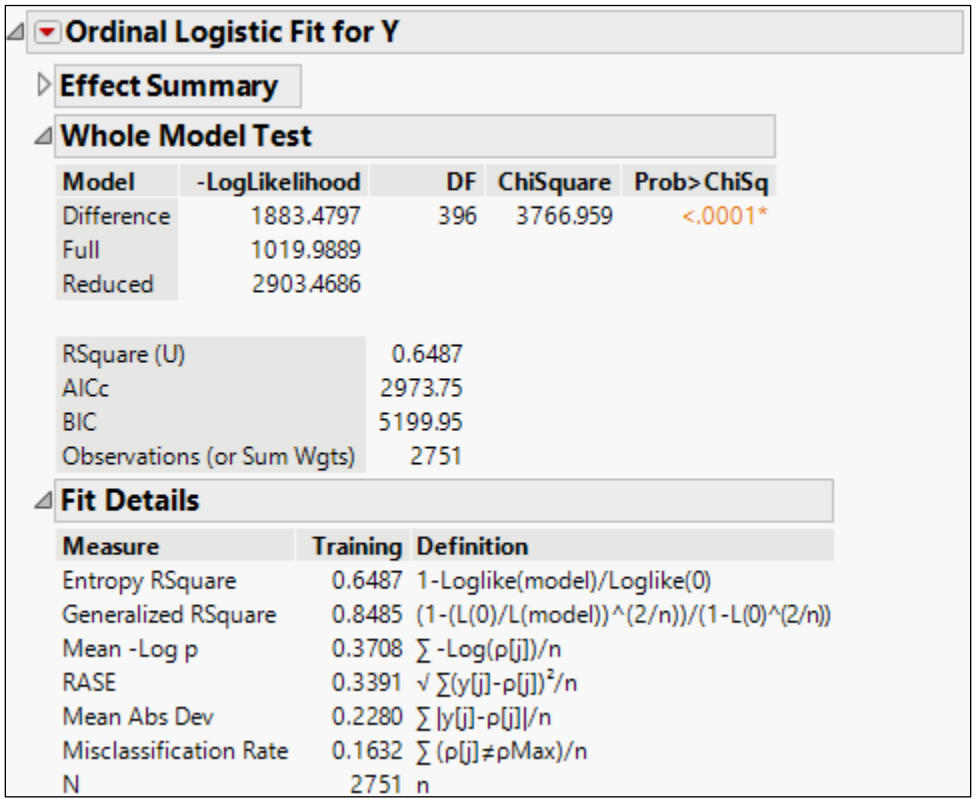
3) Select **X1** through **X24**, then click **Macros** > **Response Surface**.



4) Click **Run**.

5) Close the Effect Summary outline.

6) Open the Fit Details outline.



Ordinal Logistic Fit for Y

Effect Summary

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	1883.4797	396	3766.959	<.0001*
Full	1019.9889			
Reduced	2903.4686			

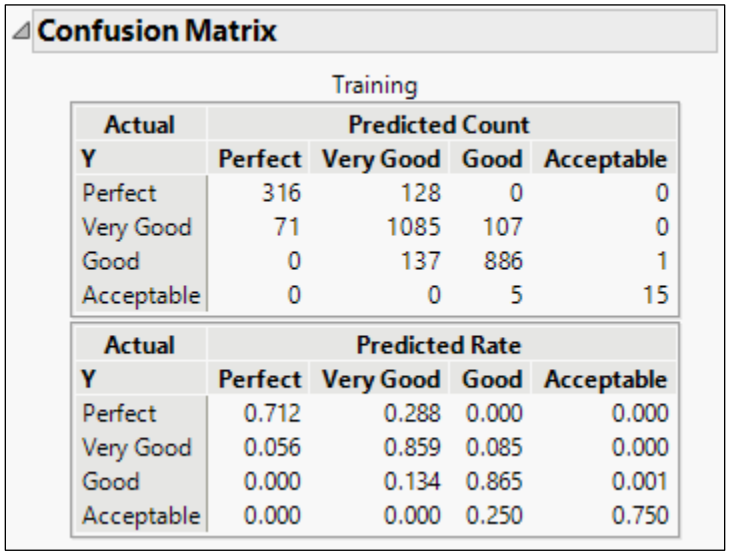
RSquare (U)	0.6487
AICc	2973.75
BIC	5199.95
Observations (or Sum Wgts)	2751

Fit Details

Measure	Training	Definition
Entropy RSquare	0.6487	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.8485	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.3708	∑ -Log(ρ[j])/n
RASE	0.3391	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.2280	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.1632	∑ (ρ[j]≠ρMax)/n
N	2751	n

The misclassification rate is about 16%.

7) Click the red triangle next to **Ordinal Logistic Fit for Y** and select **Confusion Matrix**.



Confusion Matrix

Training

Actual Y	Predicted Count Perfect	Very Good	Good	Acceptable
Perfect	316	128	0	0
Very Good	71	1085	107	0
Good	0	137	886	1
Acceptable	0	0	5	15

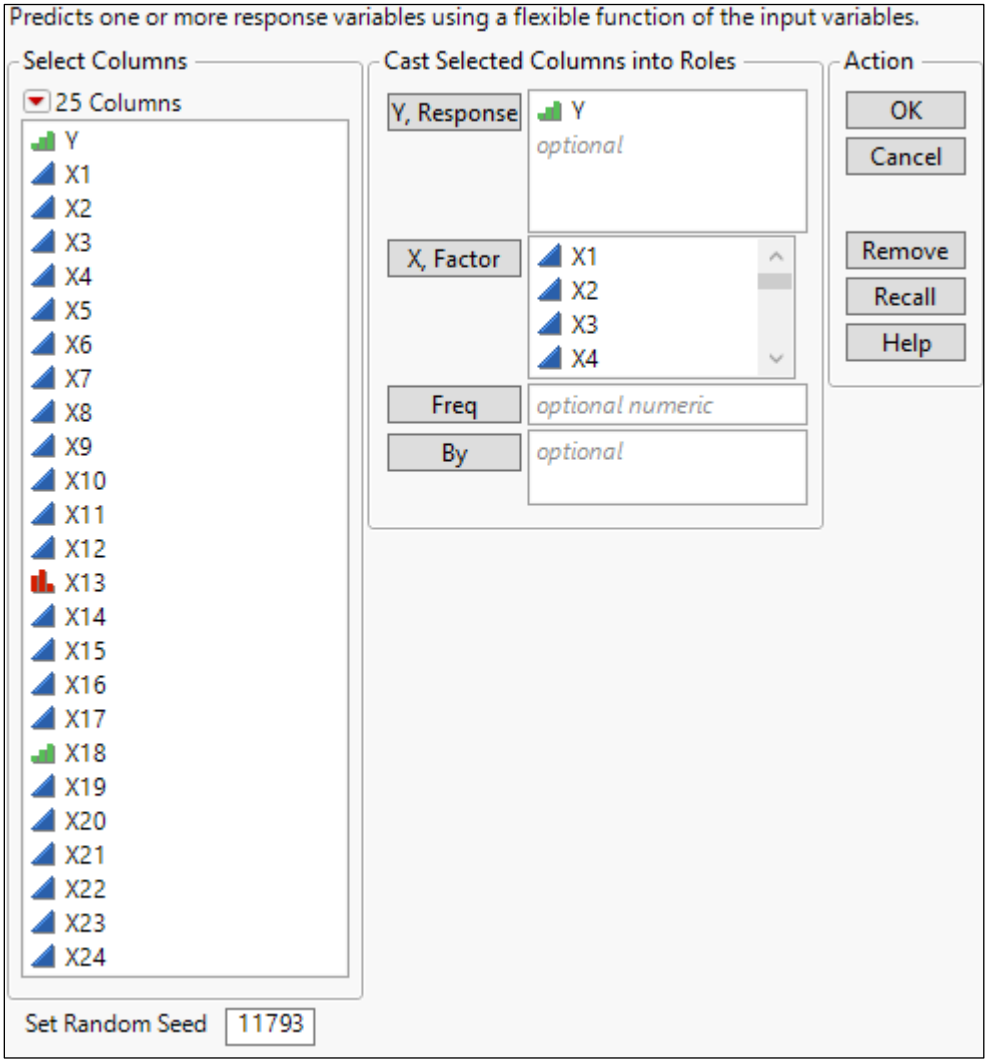
Actual Y	Predicted Rate Perfect	Very Good	Good	Acceptable
Perfect	0.712	0.288	0.000	0.000
Very Good	0.056	0.859	0.085	0.000
Good	0.000	0.134	0.865	0.001
Acceptable	0.000	0.000	0.250	0.750

The misclassification rate for the Acceptable group is 1 – 0.75 = 0.25. Five Acceptable observations were classified as Good.

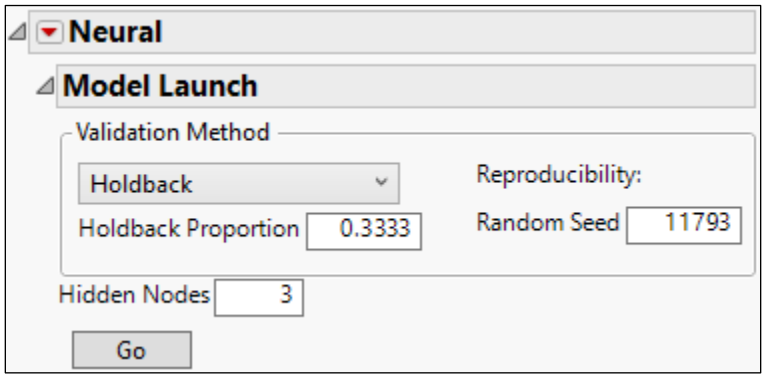
c. Build a neural network model using the default settings of the Neural platform. What is the misclassification rate of the model on the validation set? In particular, what is the misclassification rate of the Acceptable group? Fit another model with 50 nodes. What is the

misclassification rate of the model on the validation set? What is the misclassification rate of the Acceptable group?

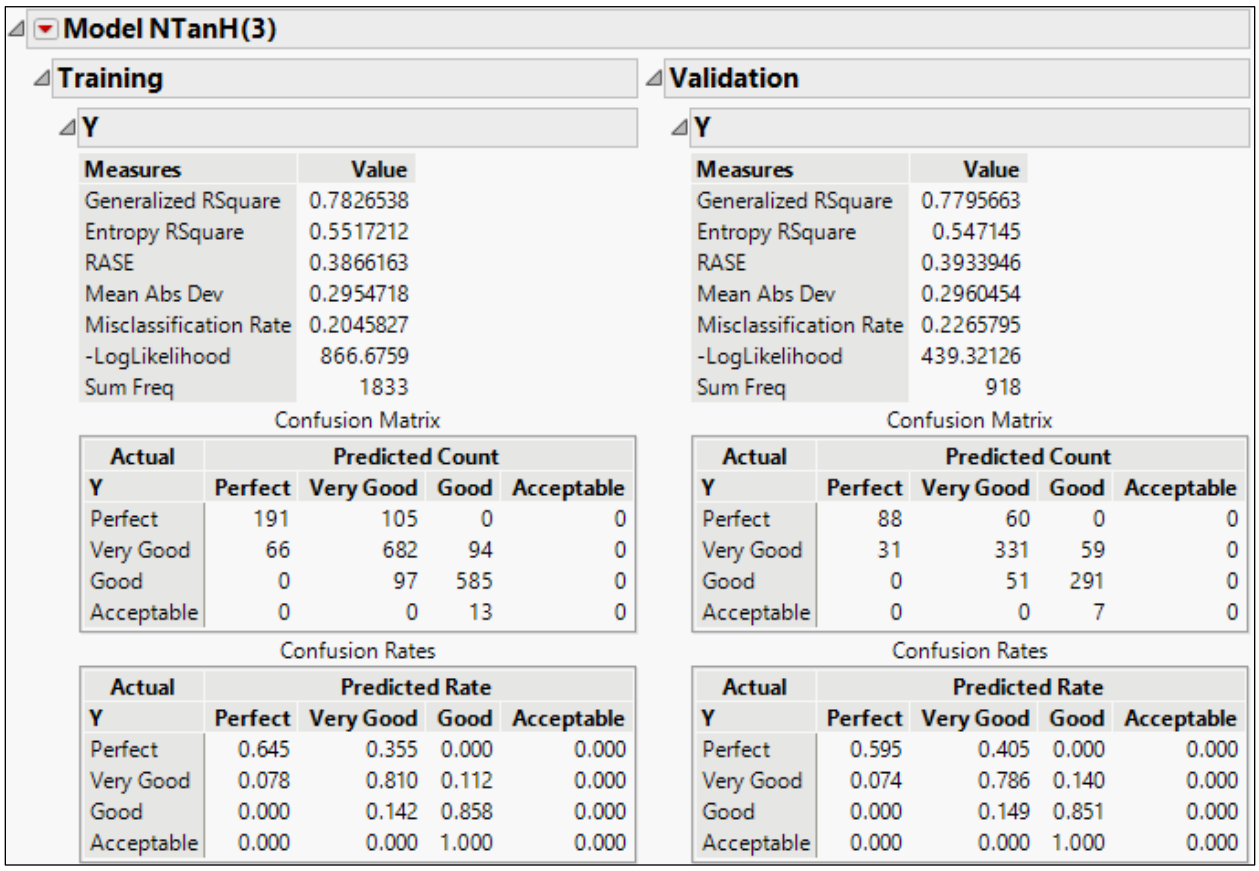
1) Select **Analyze** > **Predictive Modeling** > **Neural**.
2) Select **Y**, then click **Y, Response**.
3) Select **X1** through **X24**, then click **X, Factor**.
4) If you want your results to match this solution, enter 11793 as the random seed.



5) Click **OK**.



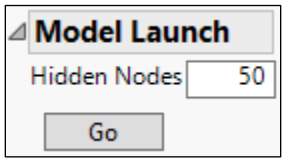
6) Click **Go**.



The misclassification rate on the validation set is about 23%. The misclassification rate of the Acceptable group is 100%. All seven observations in the Acceptable group were misclassified as Good. All thirteen observations in Acceptable group in the Training data were also misclassified as Good.

7) Open the Model Launch outline.

8) Change the number of Hidden Nodes to 50.



9) Click **Go**.

Model NTanH(50)

Training

Y

Measures	Value
Generalized RSquare	0.8760636
Entropy RSquare	0.6962804
RASE	0.3150258
Mean Abs Dev	0.2159405
Misclassification Rate	0.1352973
-LogLikelihood	587.19368
Sum Freq	1833

Confusion Matrix

Actual	Predicted Count			
Y	Perfect	Very Good	Good	Acceptable
Perfect	221	75	0	0
Very Good	40	749	53	0
Good	1	72	607	2
Acceptable	0	0	5	8

Confusion Rates

Actual	Predicted Rate			
Y	Perfect	Very Good	Good	Acceptable
Perfect	0.747	0.253	0.000	0.000
Very Good	0.048	0.890	0.063	0.000
Good	0.001	0.106	0.890	0.003
Acceptable	0.000	0.000	0.385	0.615

Validation

Y

Measures	Value
Generalized RSquare	0.7325104
Entropy RSquare	0.4886919
RASE	0.4114429
Mean Abs Dev	0.2833961
Misclassification Rate	0.2352941
-LogLikelihood	496.02749
Sum Freq	918

Confusion Matrix

Actual	Predicted Count			
Y	Perfect	Very Good	Good	Acceptable
Perfect	96	52	0	0
Very Good	41	315	65	0
Good	0	55	286	1
Acceptable	0	0	2	5

Confusion Rates

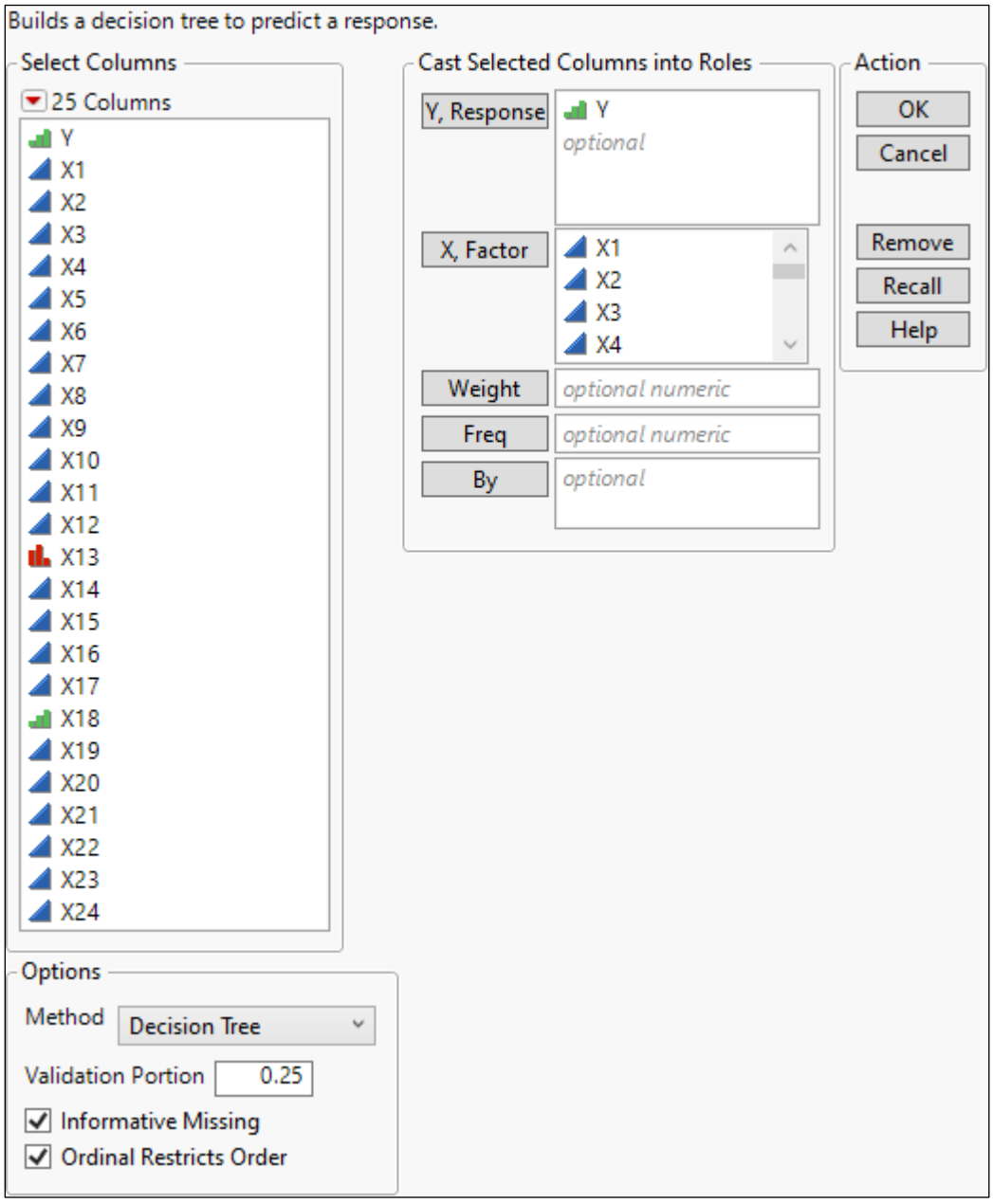
Actual	Predicted Rate			
Y	Perfect	Very Good	Good	Acceptable
Perfect	0.649	0.351	0.000	0.000
Very Good	0.097	0.748	0.154	0.000
Good	0.000	0.161	0.836	0.003
Acceptable	0.000	0.000	0.286	0.714

The misclassification rate on the validation set is again about 23%. The misclassification rate of the Acceptable group is about 29%.

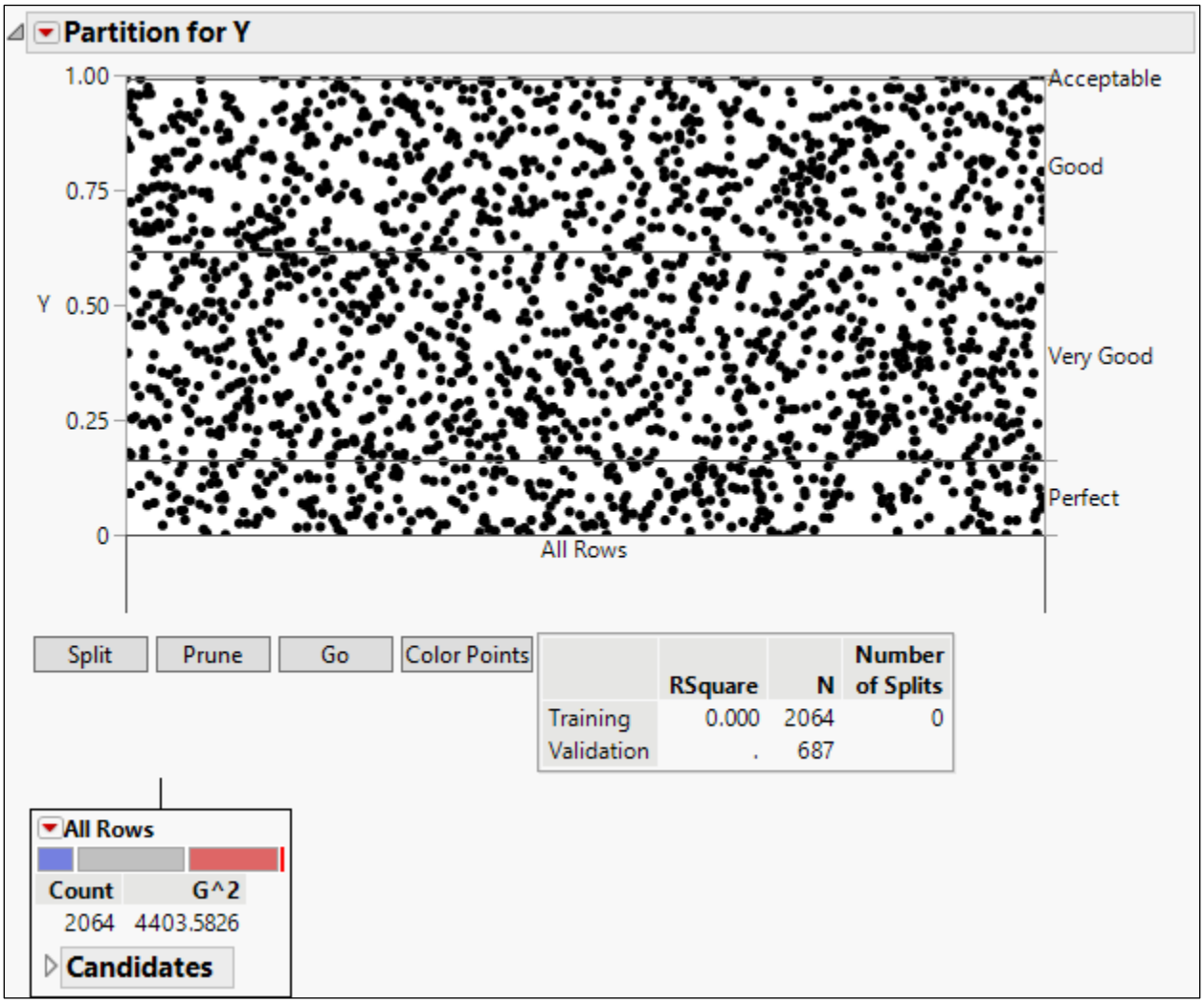
d. Build a decision tree model using the Partition platform with a 25% validation set. What is the misclassification rate of the model on the validation set? What is the misclassification rate of the Acceptable group?

1) Select **Analyze** > **Predictive Modeling** > **Partition**.
2) Select **Y**, then click **Y, Response**.
3) Select **X1** through **X24**, then click **X, Factor**.

4) For the Validation Portion, enter 0.25.

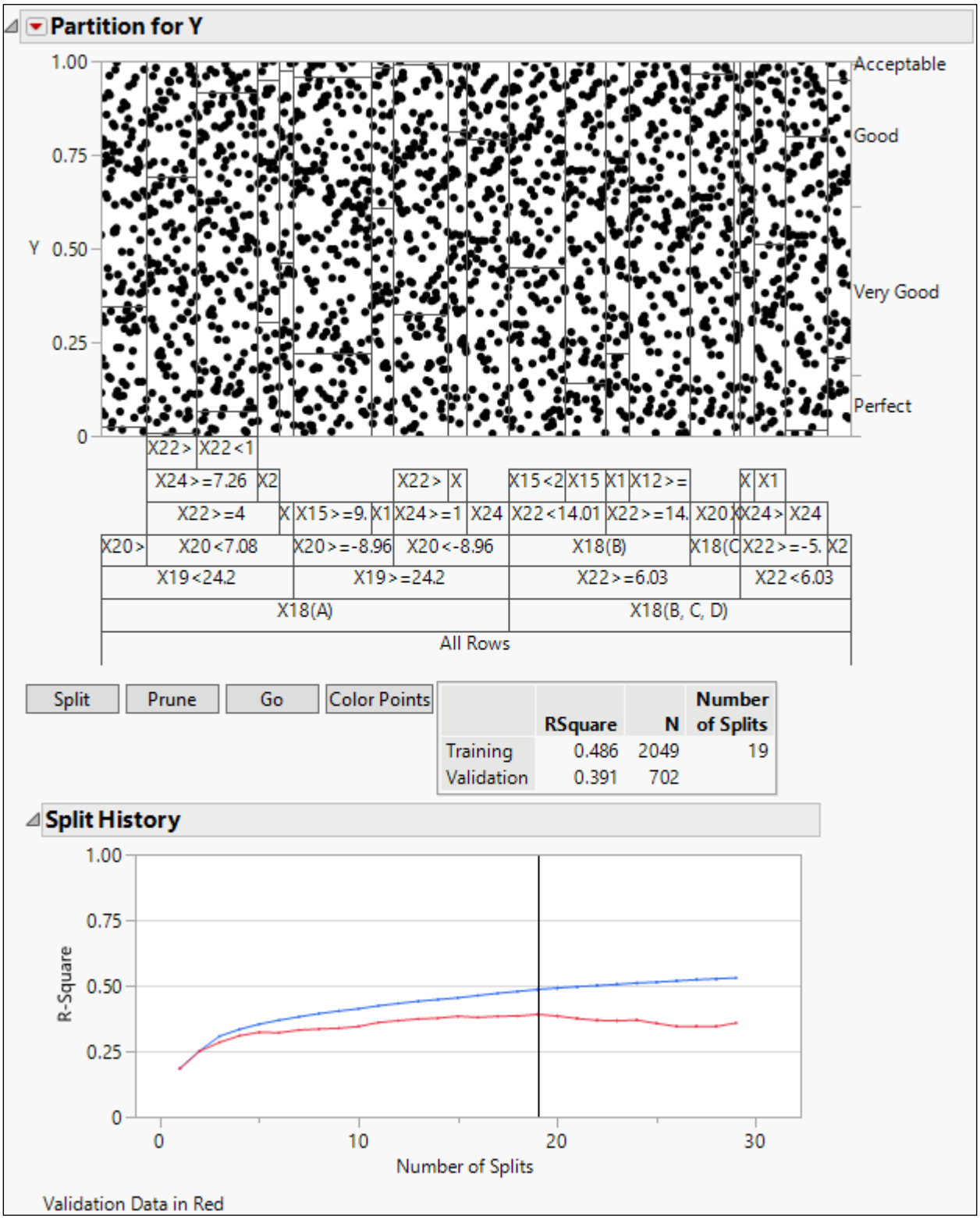


5) Click **OK**.



6) Click **Go**.

7) If needed, to remove the tree from the output, click the red triangle next to **Partition for Y** and select **Display Options** > **Show Tree**.



8) Click the red triangle next to **Partition for Y** and select **Show Fit Details**.

Fit Details

Measure	Training	Validation	Definition
Entropy RSquare	0.4856	0.3907	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.7292	0.6398	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.5419	0.6453	∑ -Log(ρ[j])/n
RASE	0.4226	0.4513	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.3430	0.3672	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.2479	0.2749	∑ (ρ[j]≠ρMax)/n
N	2049	702	n

Confusion Matrix

Training

Actual	Predicted Count			
Y	Perfect	Very Good	Good	Acceptable
Perfect	170	159	3	0
Very Good	57	733	137	0
Good	1	140	629	7
Acceptable	0	0	4	9

Actual	Predicted Rate			
Y	Perfect	Very Good	Good	Acceptable
Perfect	0.512	0.479	0.009	0.000
Very Good	0.061	0.791	0.148	0.000
Good	0.001	0.180	0.810	0.009
Acceptable	0.000	0.000	0.308	0.692

Validation

Actual	Predicted Count			
Y	Perfect	Very Good	Good	Acceptable
Perfect	51	59	2	0
Very Good	22	262	52	0
Good	0	52	193	2
Acceptable	0	0	4	3

Actual	Predicted Rate			
Y	Perfect	Very Good	Good	Acceptable
Perfect	0.455	0.527	0.018	0.000
Very Good	0.065	0.780	0.155	0.000
Good	0.000	0.211	0.781	0.008
Acceptable	0.000	0.000	0.571	0.429

Your results will vary due to the random nature of the holdout procedure. The misclassification rate on the validation set is about 27%. The misclassification rate of the Acceptable group on the validation set is about 57%. About half the observations in the Acceptable group were misclassified as Good.