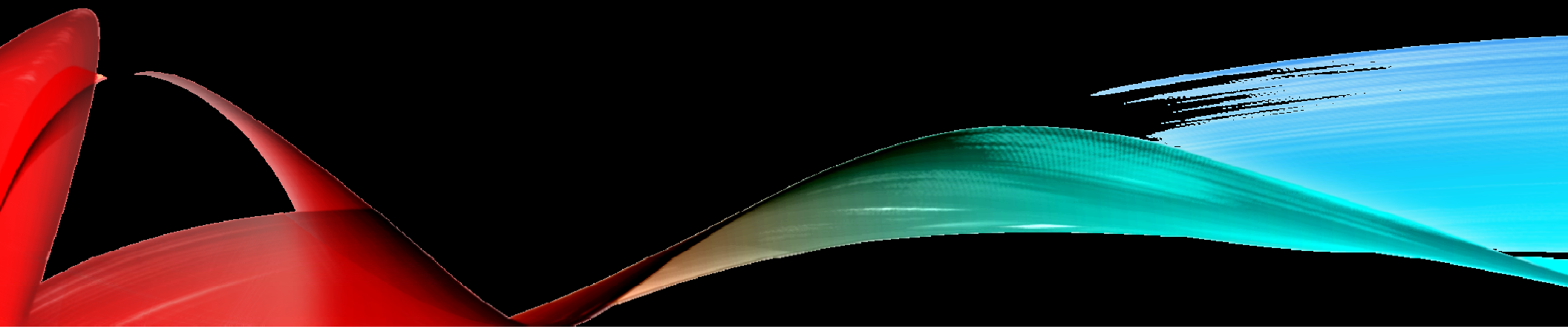
MULTIPLE REGRESSION – PART 2

CATEGORICAL VARIABLES



Fitness (Help>Sample Data)

- Name
- Sex
- Age
- Weight
- Oxy
- Runtime
- RunPulse
- RstPulse
- MaxPulse

Name	Sex	Age	Weight	Oxy	Runtime	RunPulse	RstPulse	MaxPulse
Donna	F	42	68.15	59.57	8.17	166	40	172
Gracie	F	38	81.87	60.06	8.63	170	48	186
Luanne	F	43	85.84	54.30	8.65	156	45	168
Mimi	F	50	70.87	54.63	8.92	146	48	155
Chris	M	49	81.42	49.16	8.95	180	44	185

Summary of Fit

RSquare	0.758967
RSquare Adj	0.741751
Root Mean Square Error	2.707204
Mean of Response	47.37581
Observations (or Sum Wgts)	31

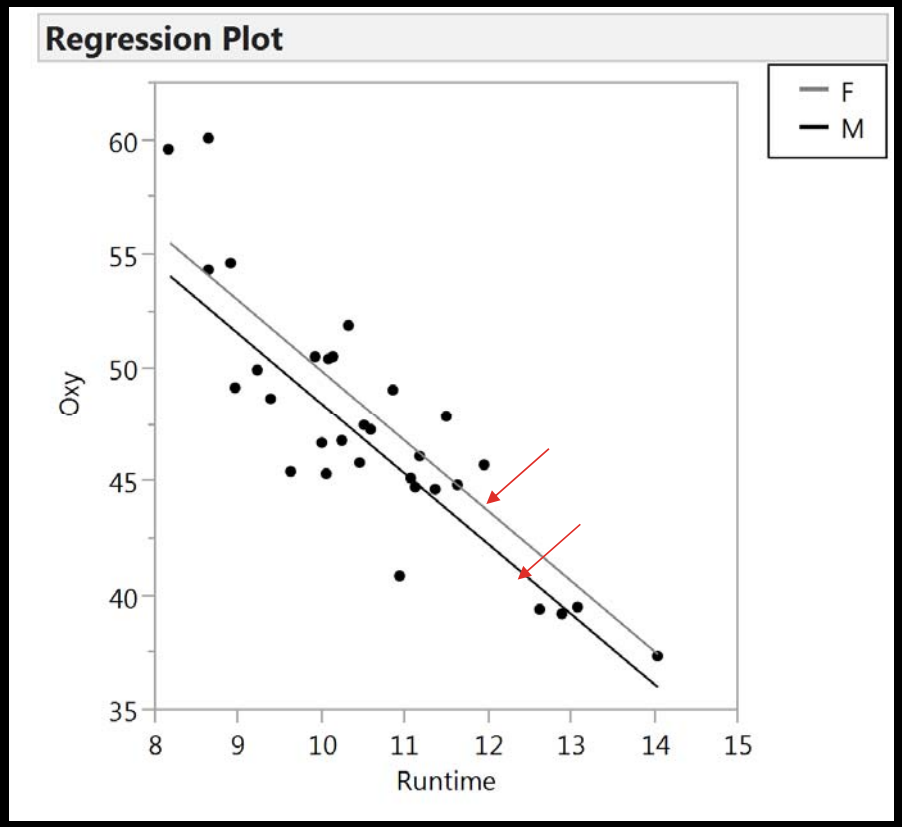
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	646.17085	323.085	44.0834
Error	28	205.21069	7.329	Prob > F
C. Total	30	851.38154		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	79.972837	4.215593	18.97	<.0001*
Sex[F]	0.7255551	0.539192	1.35	0.1892
Runtime	-3.081432	0.39485	-7.80	<.0001*

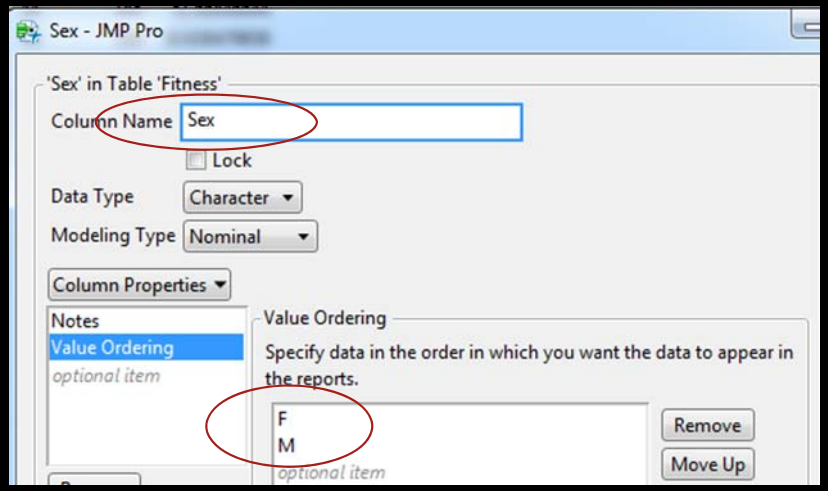
```
79.9728367824707
+ Match( Sex, "F" ⇒ 0.7255551433597, "M" ⇒ -0.7255551433597, else ⇒ . )
+ -3.0814318658011 * Runtime
```

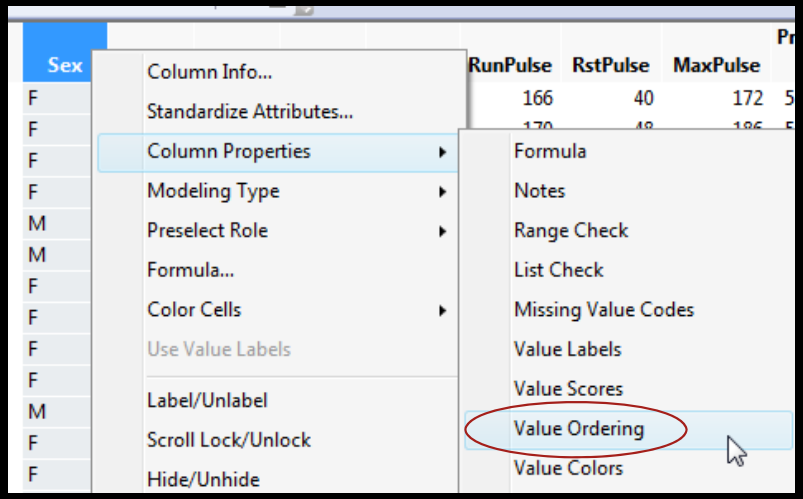


Value Ordering

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	79.972837	4.215593	18.97	<.0001*
Sex[F]	0.7255551	0.539192	1.35	0.1892
Runtime	-3.081432	0.39485	-7.80	<.0001*





Data: Roller Coasters (www.rcdb.com)
Bus Stat 3e JMP File: CE_18

Data Descriptions

- **Type** indicates what kind of track the roller coaster has. The possible values are "wooden" and "steel." (The frame usually is of the same construction as the track, hut doesn't have to be.)
- **Duration** is the duration of the ride in seconds.
- **Speed** is top speed in miles per hour.
- **Height** is maximum height above ground level in feet.
- **Drop** is greatest drop in feet.
- **Length** is total length of the track in feet.
- **Inversions** reports whether riders are turned upside down during the ride. It has the values "yes" or "no."

Fit a model with Duration (Y) and Length (X)

Linear Fit

Duration = 53.934828 + 0.0231084*Length

Summary of Fit

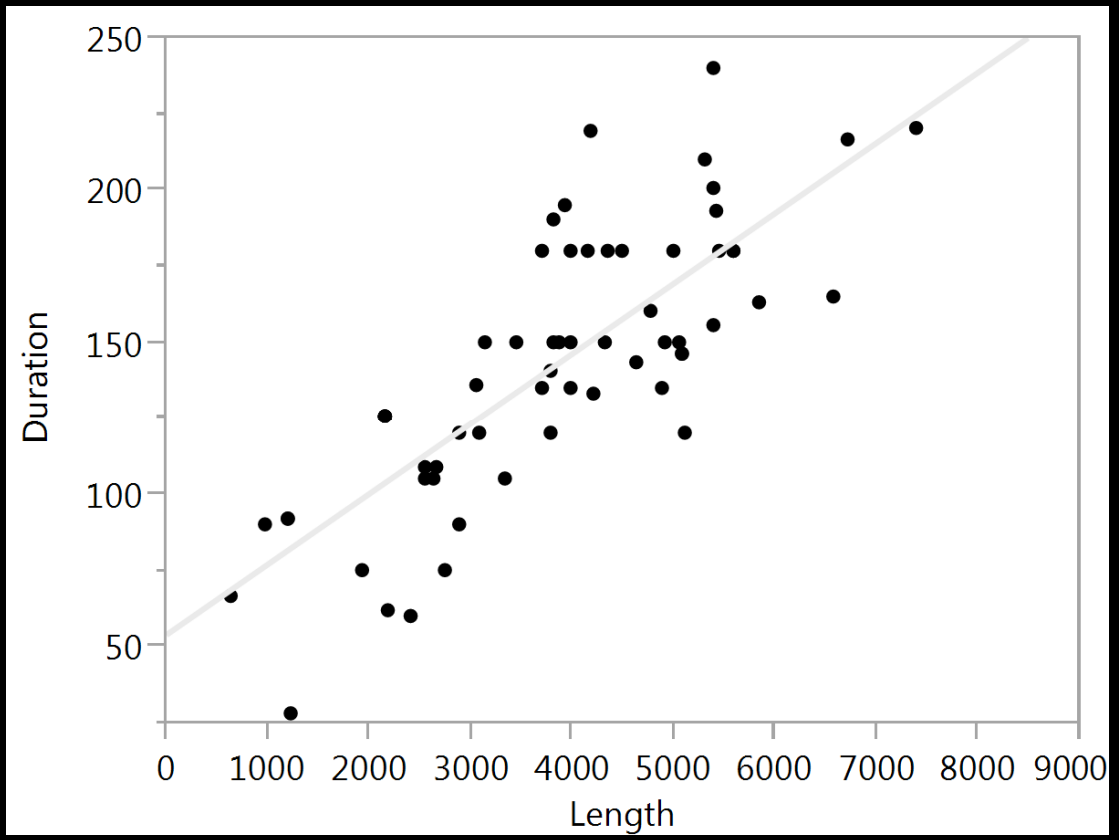
RSquare	0.620265
RSquare Adj	0.61404
Root Mean Square Error	27.23418
Mean of Response	142.2381
Observations (or Sum Wgts)	63

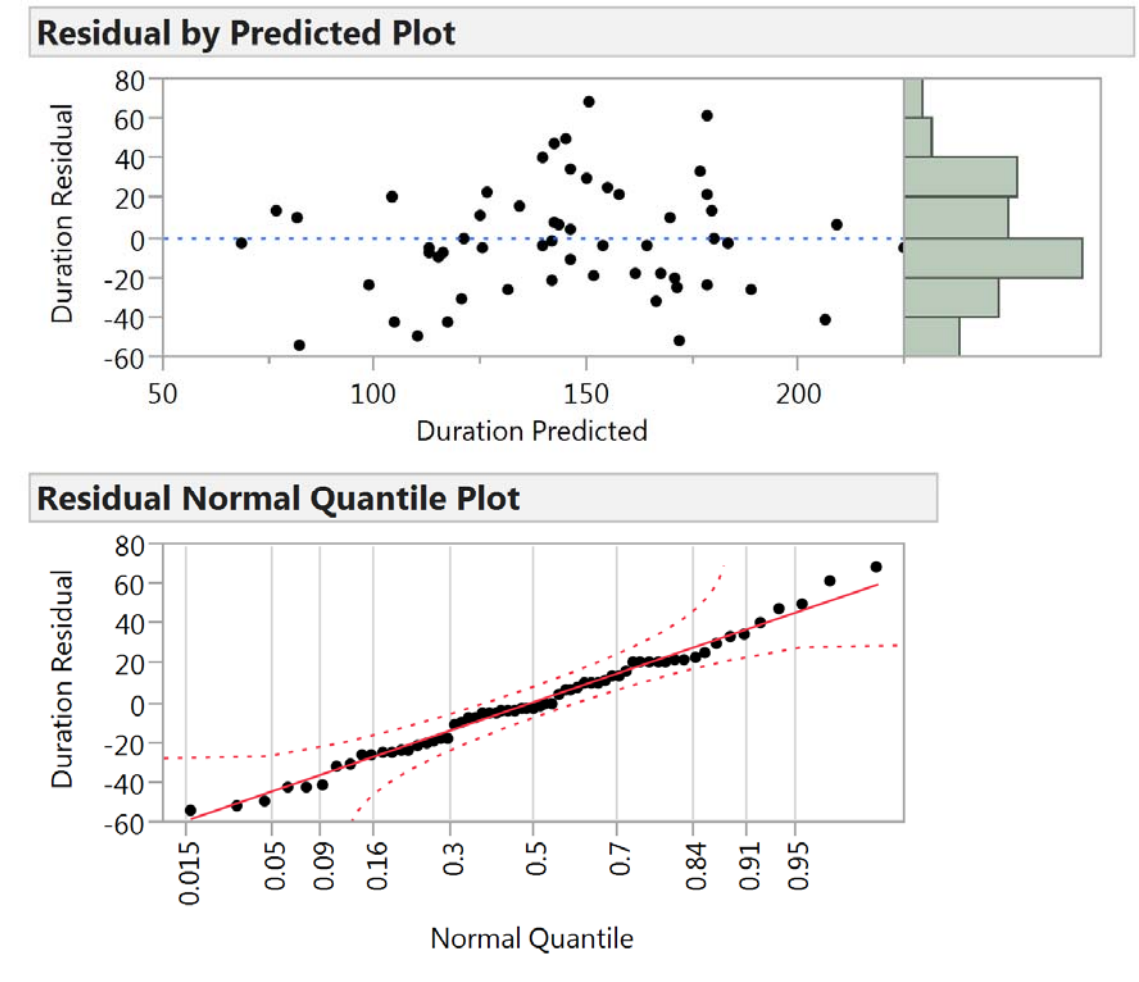
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	73901.71	73901.7	99.6382
Error	61	45243.72	741.7	Prob > F
C. Total	62	119145.43		<.0001*

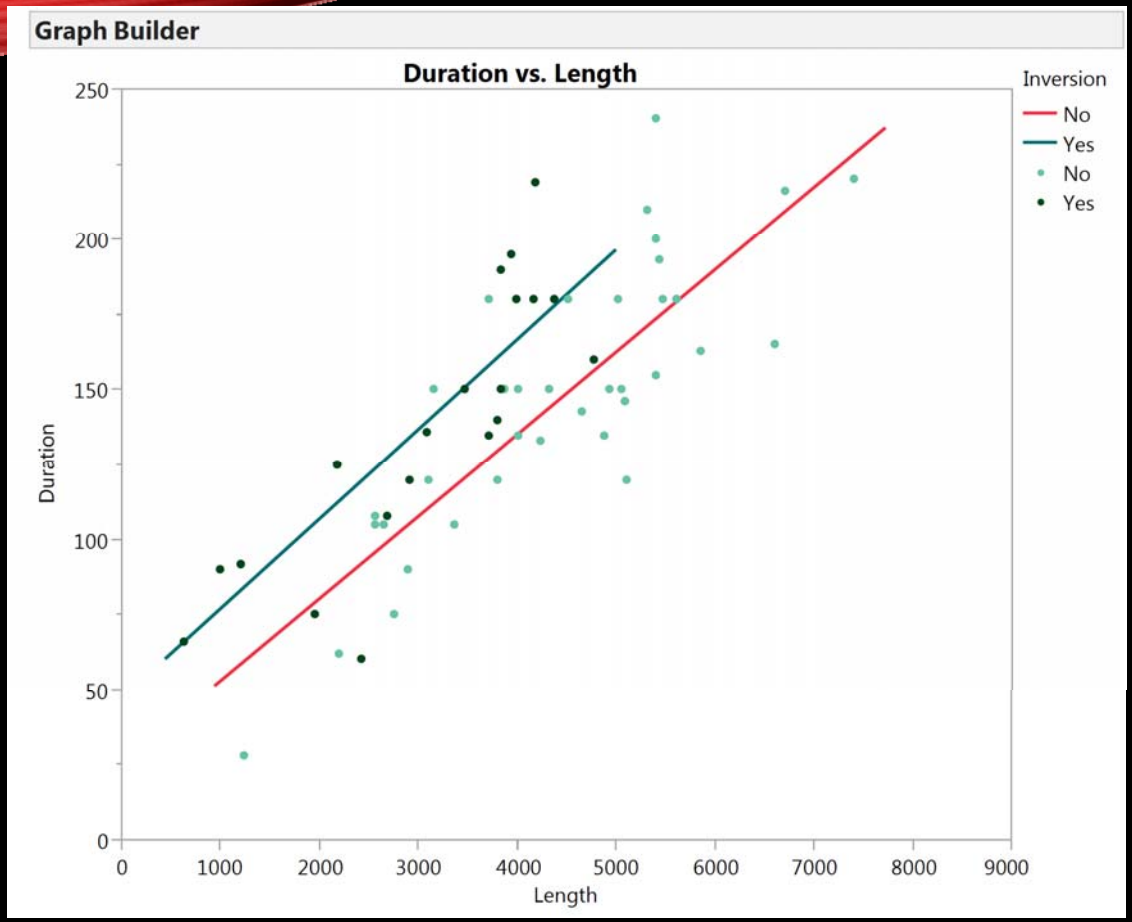
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	53.934828	9.488456	5.68	<.0001*
Length	0.0231084	0.002315	9.98	<.0001*




Residual by Predicted Plot
Duration Residual
80
60
40
20
0
-20
-40
-60
50
100
150
200
Duration Predicted
Residual Normal Quantile Plot
Duration Residual
80
60
40
20
0
-20
-40
-60
0.015
0.05
0.09
0.16
0.3
0.5
0.7
0.84
0.91
0.95
Normal Quantile

Customers like Inversion. Look at Duration and Length



Fit a model with Duration (Y) with Length and Inversion as (X)

Summary of Fit

RSquare	0.704218
RSquare Adj	0.694359
Root Mean Square Error	24.23531
Mean of Response	142.2381
Observations (or Sum Wgts)	63

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	83904.41	41952.2	71.4262
Error	60	35241.02	587.4	**Prob > F**
C. Total	62	119145.43		<.0001*

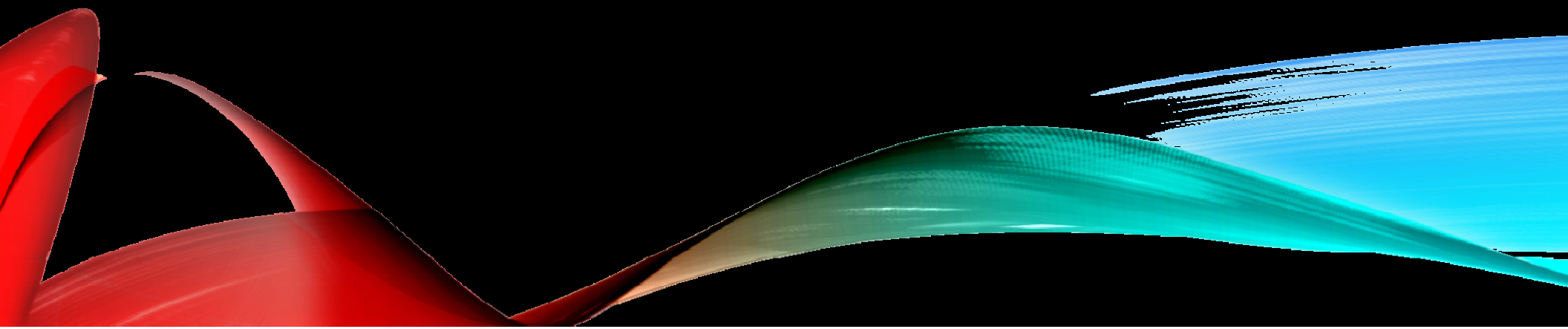
JMP Indicator Variables are +1 and -1

$$37.4320882497579 + 0.02823930214587 * Length + \text{Match}(Inversion, \text{"No"} \Rightarrow -15.041180332413, \text{"Yes"} \Rightarrow 15.041180332413, \text{else} \Rightarrow .)$$

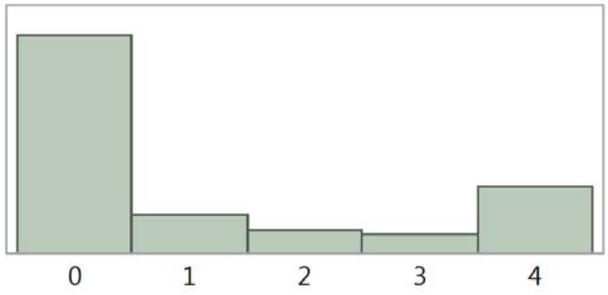
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	37.432088	9.342736	4.01	0.0002*	.
Length	0.0282393	0.002406	11.74	<.0001*	1.3642357
Inversion[No]	-15.04118	3.644784	-4.13	0.0001*	1.3642357

RE-CODING CATEGORICAL VARIABLES



SAV_ACCT



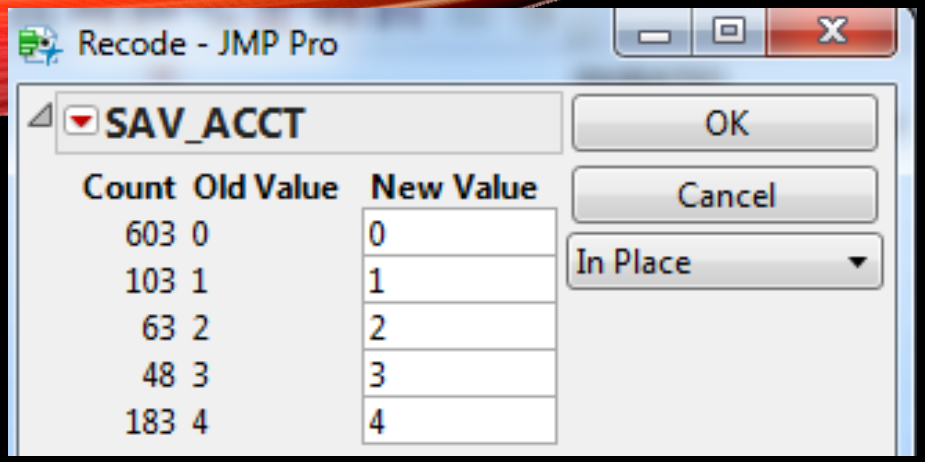
Frequencies

Level	Count	Prob
0	603	0.60300
1	103	0.10300
2	63	0.06300
3	48	0.04800
4	183	0.18300
Total	1000	1.00000
N Missing	0	
5 Levels		

Source: German Credit Case

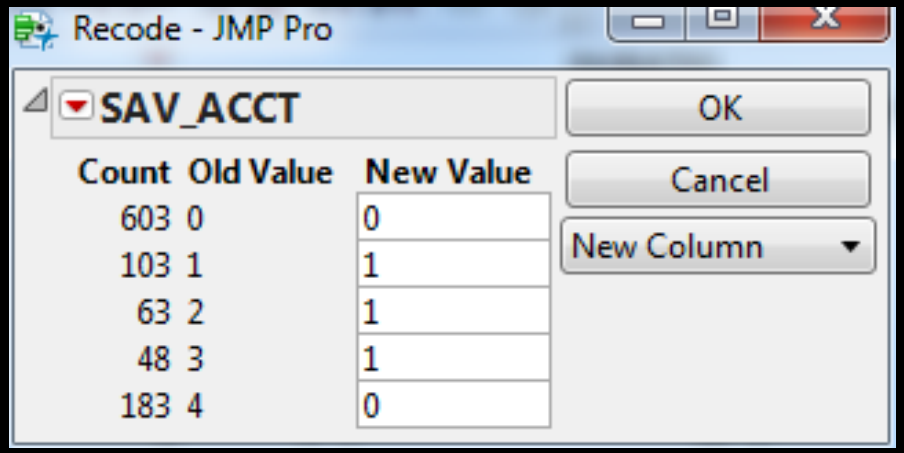
This variable is: Average balance in saving account

0: < 100DM
1: Between 100 and 500 DM
2: Between 500 and 1000 DM
3: More than 1000 DM
4: Unknown / No Savings Account

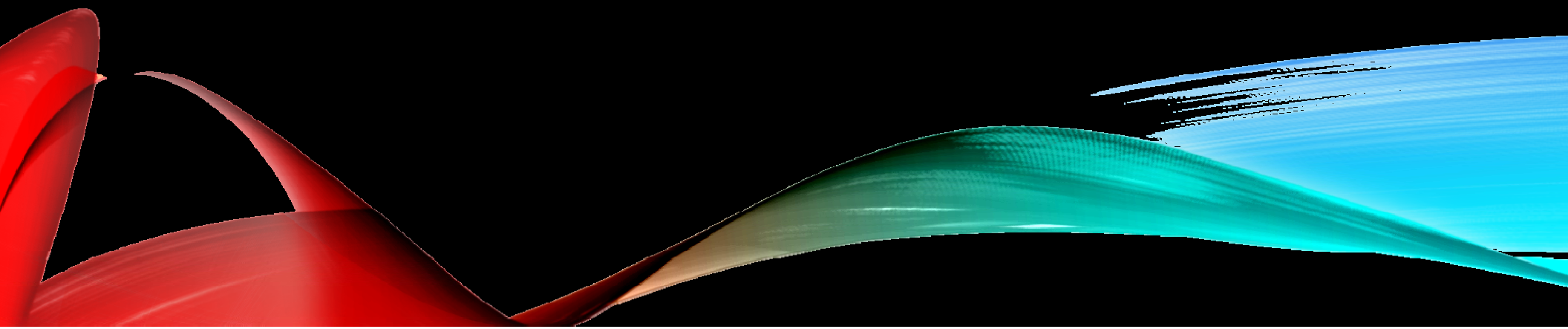


Cols > Recode:

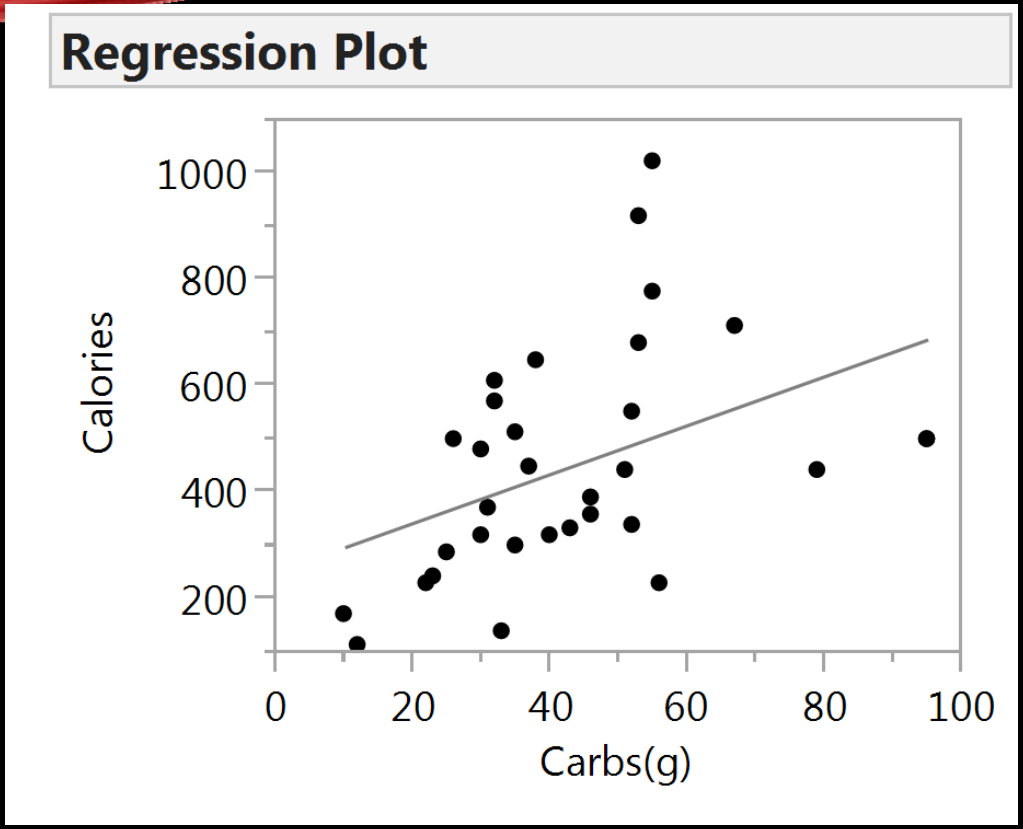
We can re-code and save as a new variable



INTERACTIONS



Data: Burger King
Bus Stat 3e JMP File: Burger_King_CE18



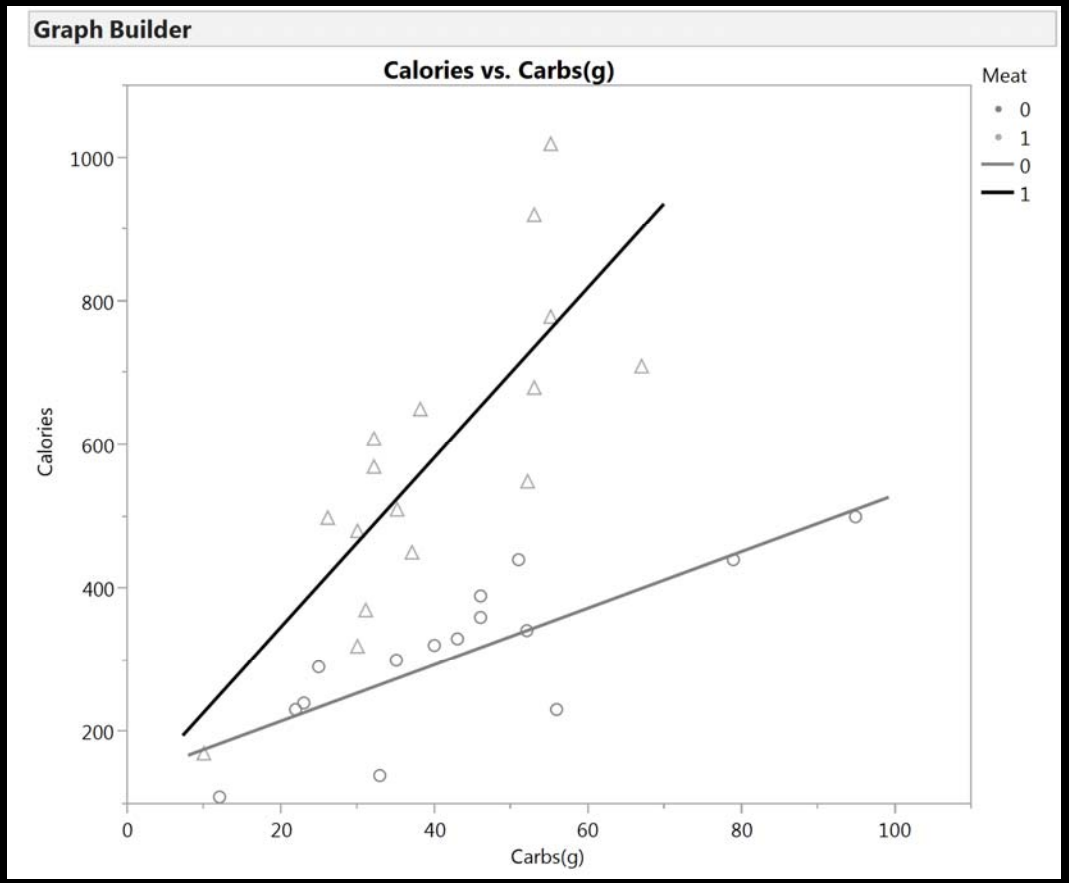
Calories and Carbs – notice how the data fans out as we move from left to right

What else is going on?

Data: Burger King
Bus Stat 3e JMP File: Burger_King_CE18

We explore using graph builder and look at two groups:

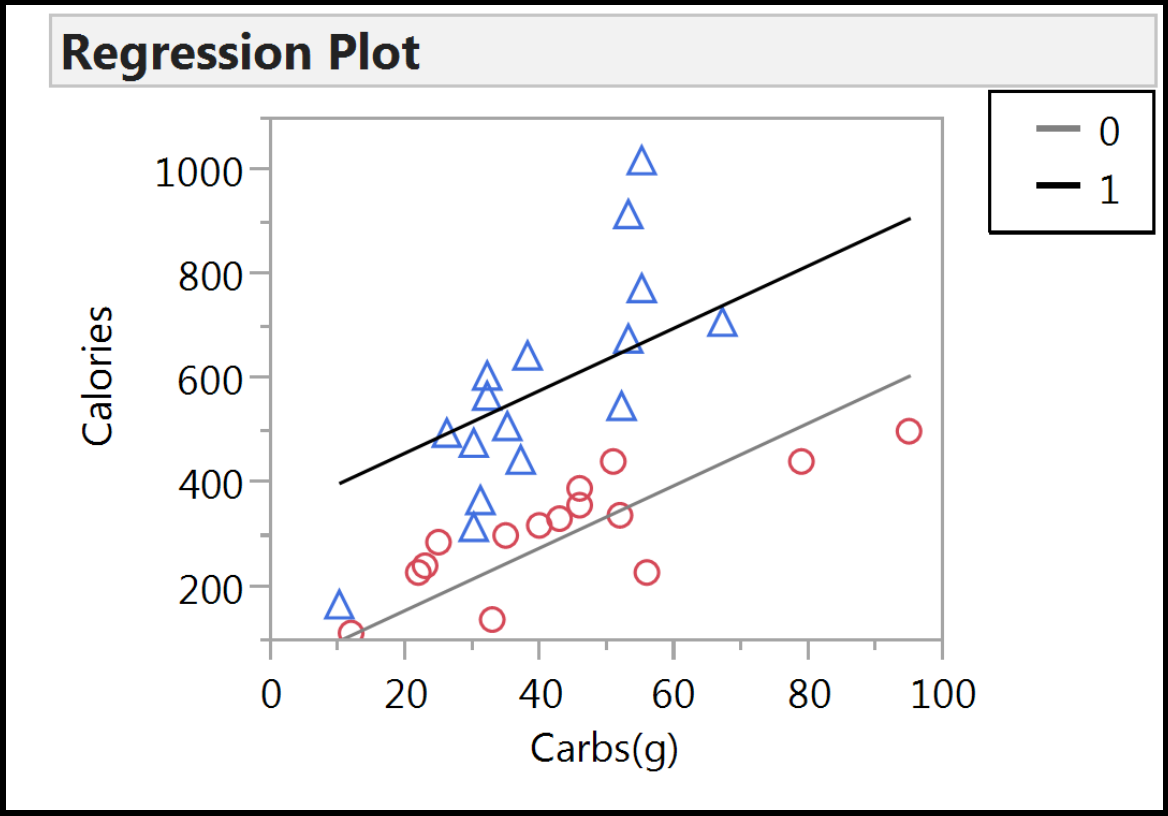
- Those with meat (including chicken and fish)
- Those without meat

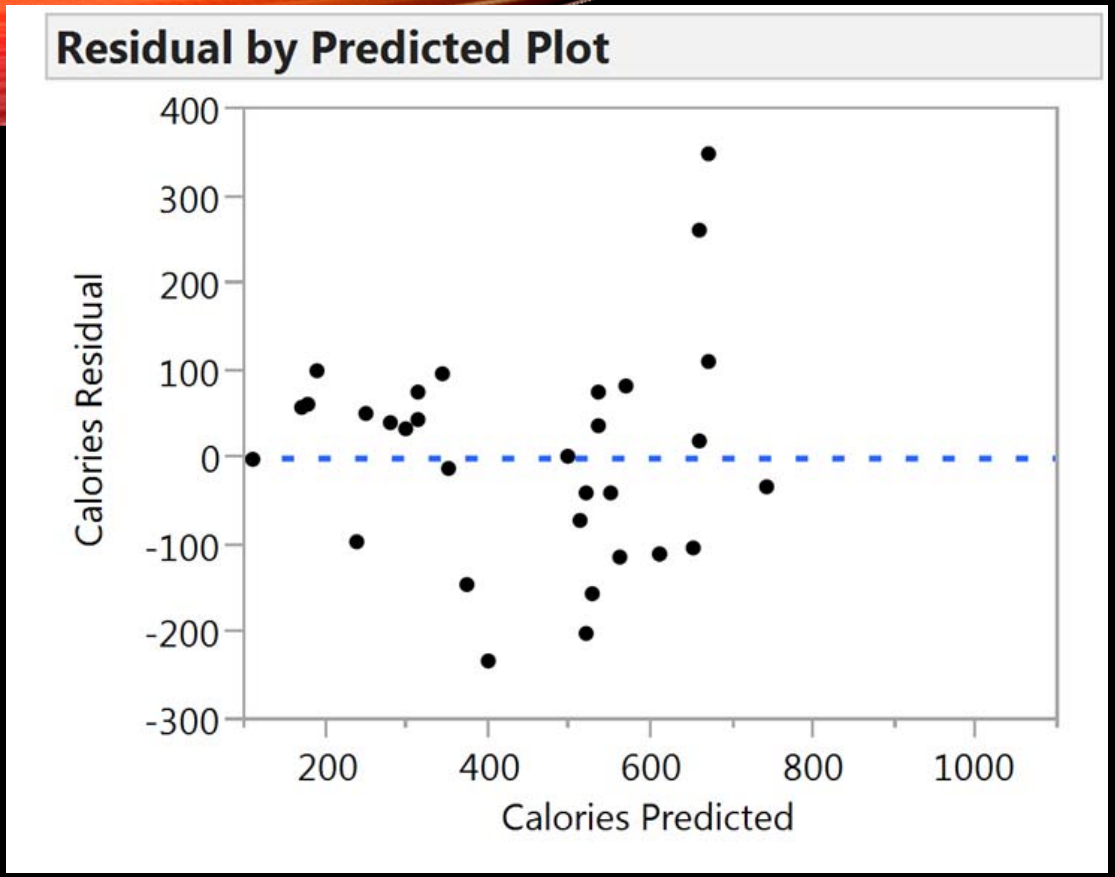


We develop a model with carbs and a categorical variable for meat (with and without)

Parameter Estimates

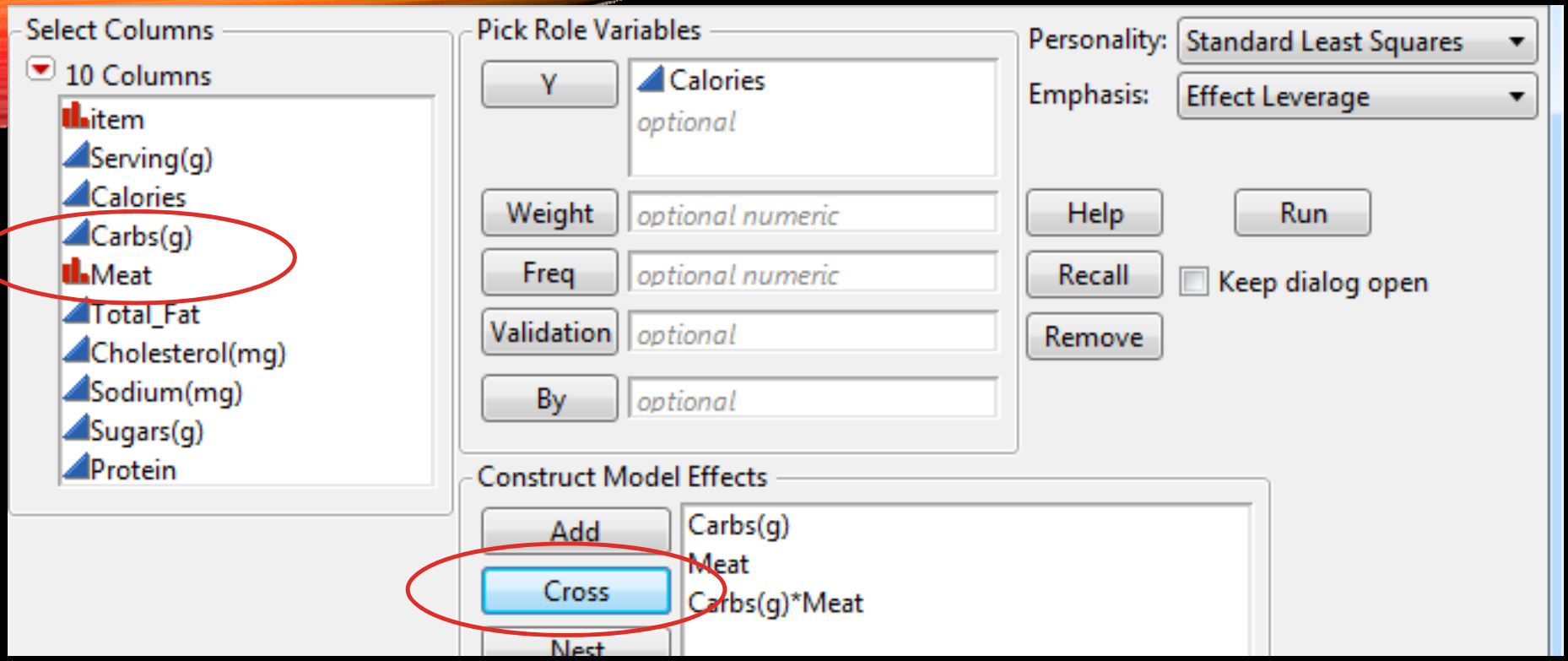
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	191.63292	54.23211	3.53	0.0014*	.
Carbs(g)	5.9882984	1.138852	5.26	<.0001*	1.0348528
Meat[0]	-150.9572	22.68943	-6.65	<.0001*	1.0348528





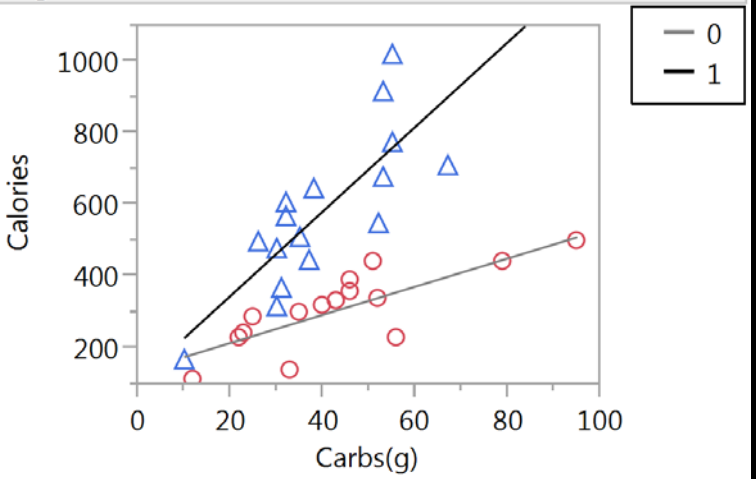
There is more to it than a categorical variable as we can see by the residual by predicted and also by the different slopes in the graph builder

We adjust for the slopes by introducing an interaction term which is the product of Carbs and Meat (Carbs * Meat)



Select from Column: Carbs(g) and Meat | Then select Cross (Model Effects)

Regression Plot



The interaction term allows a different slope for the two levels of Meat

Summary of Fit

RSquare	0.78057
RSquare Adj	0.75706
Root Mean Square Error	106.0395
Mean of Response	451.5625
Observations (or Sum Wgts)	32

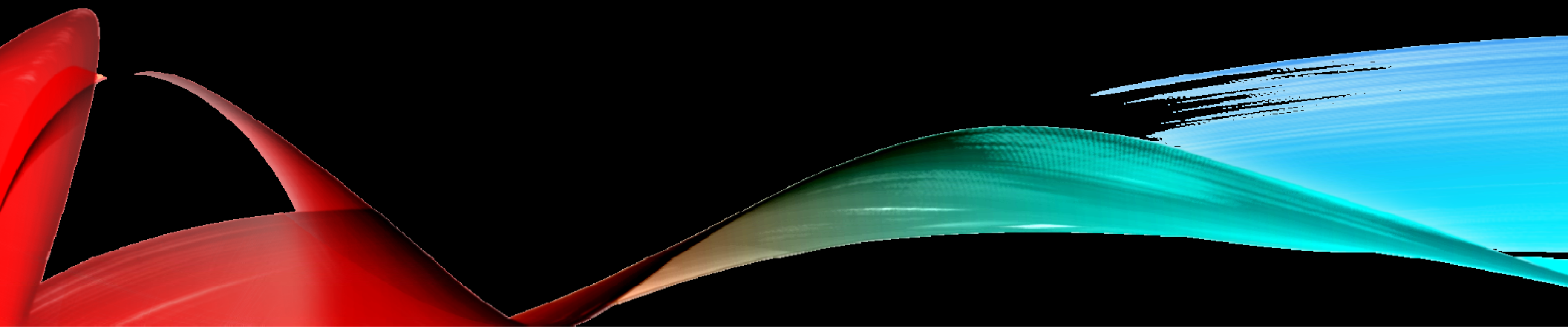
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	1119979.2	373326	33.2011
Error	28	314842.7	11244	**Prob > F**
C. Total	31	1434821.9		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	124.31704	49.23936	2.52	0.0175*	.
Carbs(g)	7.8708132	1.089746	7.22	<.0001*	1.3414587
Meat[0]	-157.8402	19.16409	-8.24	<.0001*	1.045179
(Carbs(g)-43.4063)*Meat[0]	-3.937648	1.089746	-3.61	0.0012*	1.2962797

QUADRATIC



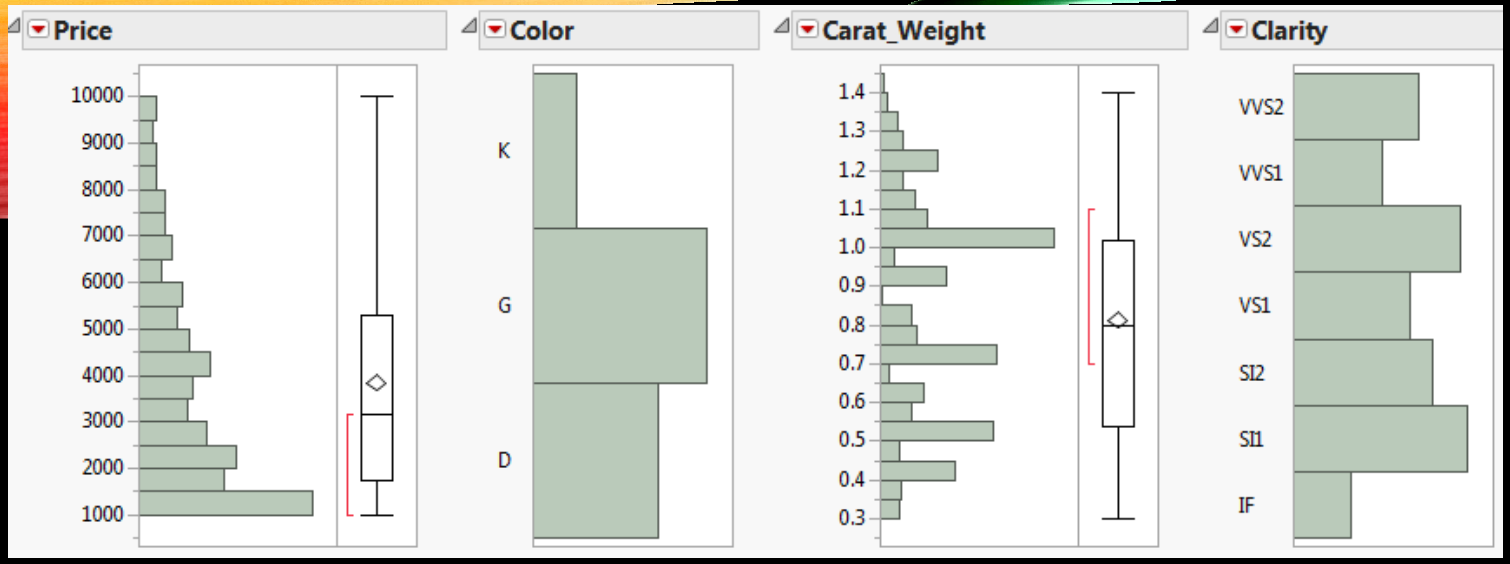
Data: Diamonds
Bus Stat 3e JMP File: Diamonds_DGK_FE18

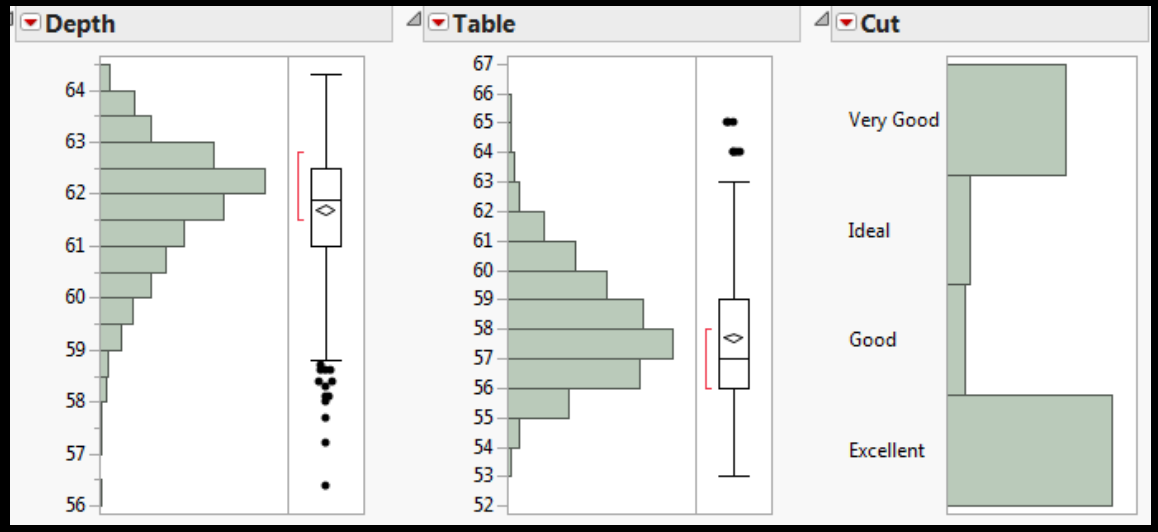
Cols > Column Viewer

Summary Statistics

7 Columns | Clear Select | Distribution

Columns	N Categories	Min	Max	Mean	Std Dev
Price	.	1000	9993	3845.866	2403.658
Color	3	.	.	.	.
Carat_Weight	.	0.3000	1.4000	0.81340	0.280201
Clarity	7	.	.	.	.
Depth	.	56.4000	64.3000	61.69626	1.209758
Table	.	53	65	57.71696	1.9451
Cut	4	.	.	.	.


Price
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
Color
K
G
D
Carat_Weight
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
Clarity
VVS2
VVS1
VS2
VS1
SI2
SI1
IF


Depth
64
63
62
61
60
59
58
57
56
Table
67
66
65
64
63
62
61
60
59
58
57
56
55
54
53
52
Cut
Very Good
Ideal
Good
Excellent

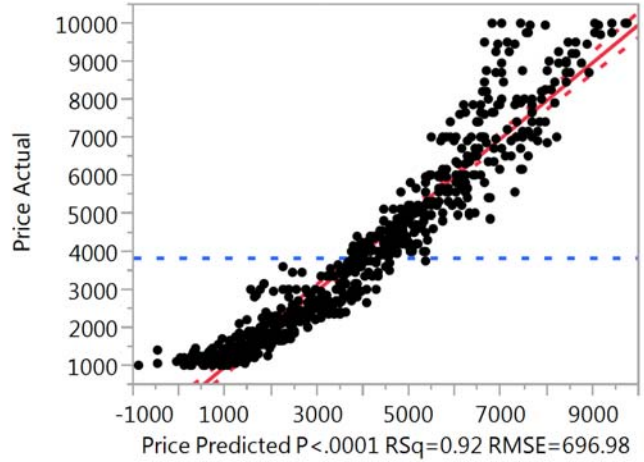
Summary of Fit

RSquare	0.917493
RSquare Adj	0.915919
Root Mean Square Error	696.9819
Mean of Response	3845.866
Observations (or Sum Wgts)	749

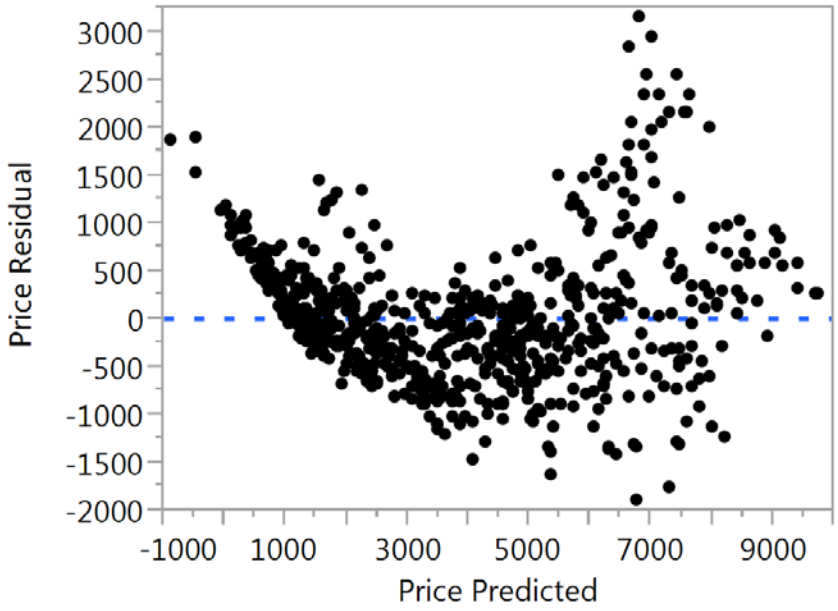
Analysis of Variance

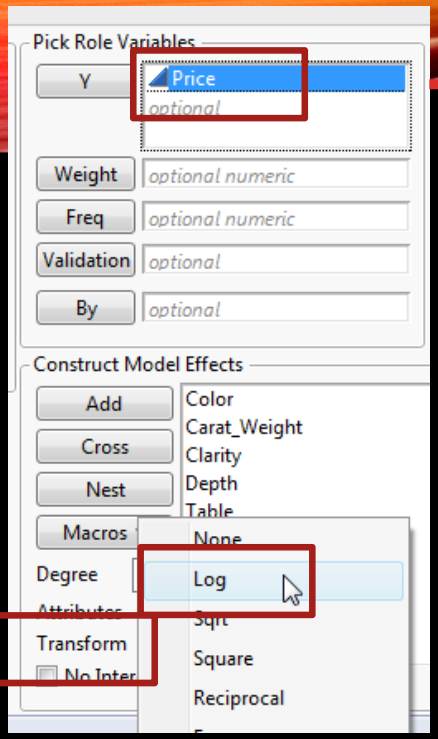
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	14	3965059902	283218564	583.0136
Error	734	356565331	485783.83	Prob > F
C. Total	748	4321625233		<.0001*

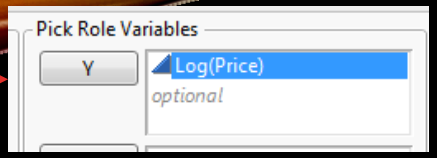
Actual by Predicted Plot



Residual by Predicted Plot



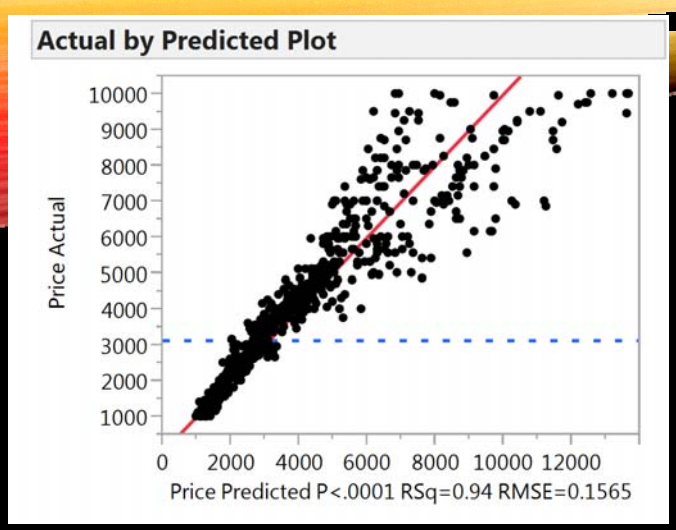


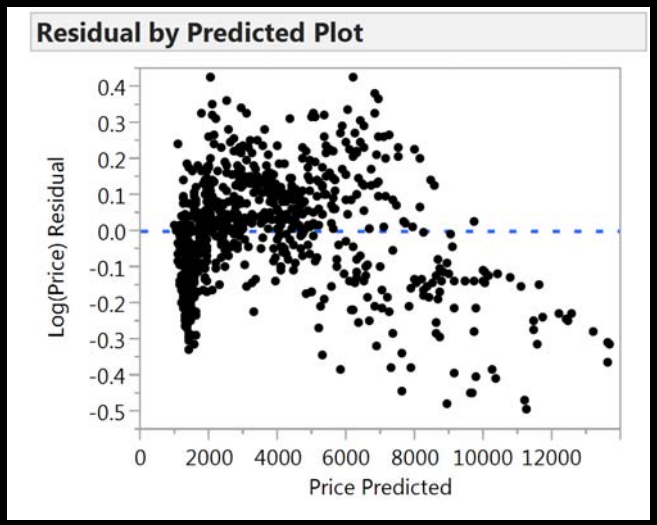


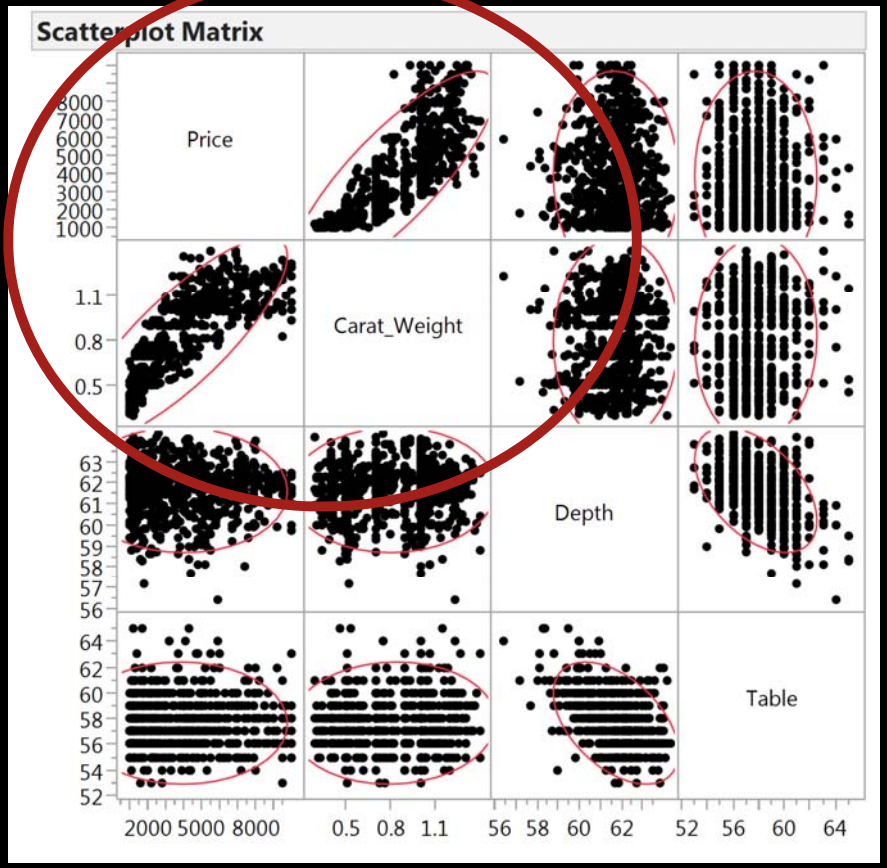
Transform the Price variable by taking Log (Price)

Effect Tests

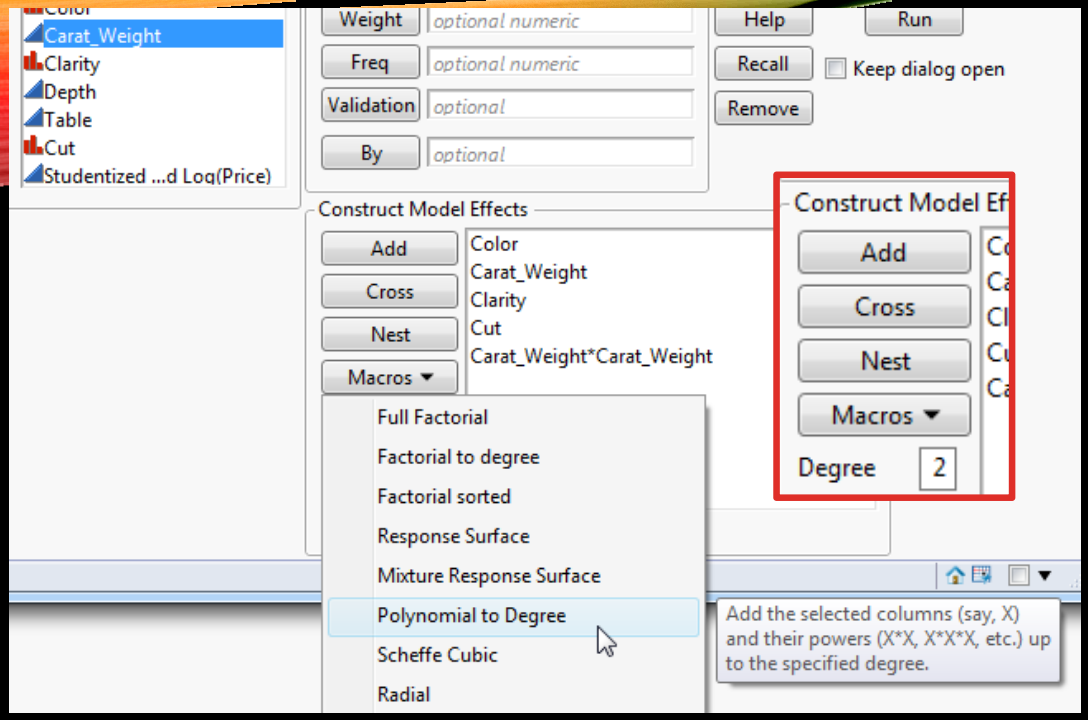
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Color	2	2	34.63436	706.5080	<.0001*
Carat_Weight	1	1	273.63704	11163.87	<.0001*
Clarity	6	6	25.87801	175.9622	<.0001*
Depth	1	1	0.00382	0.1560	0.6929
Table	1	1	0.07323	2.9878	0.0843
Cut	3	3	0.21731	2.9552	0.0318*

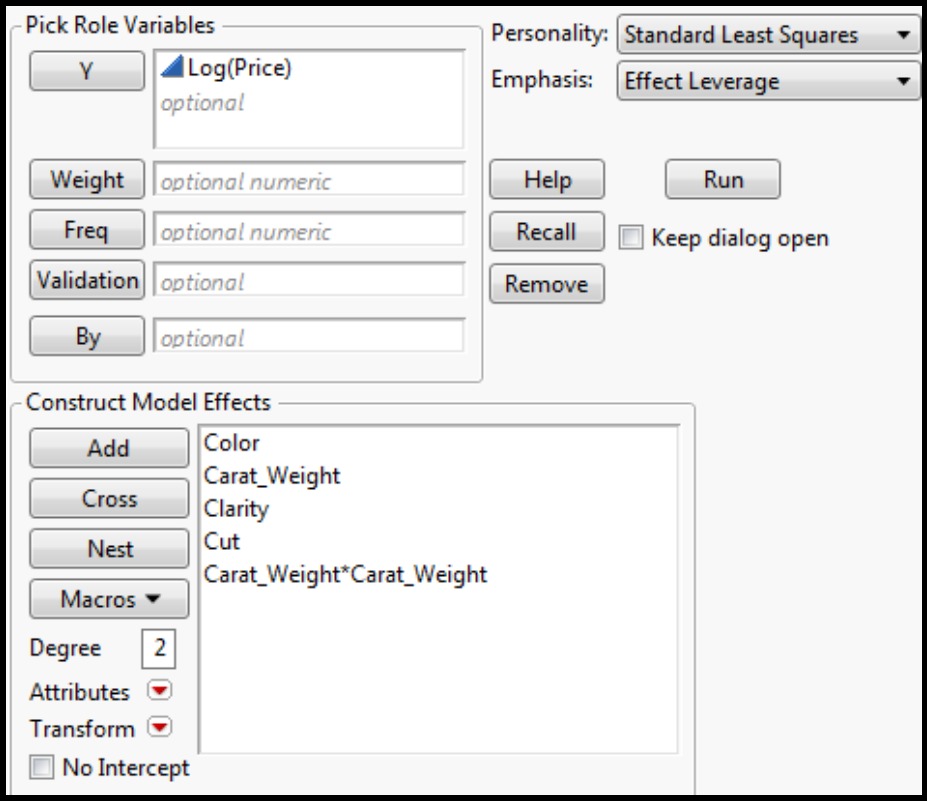






There could be some curvature of Carat_Weight with Price





We create a quadratic term for Carat_Weight

Summary of Fit

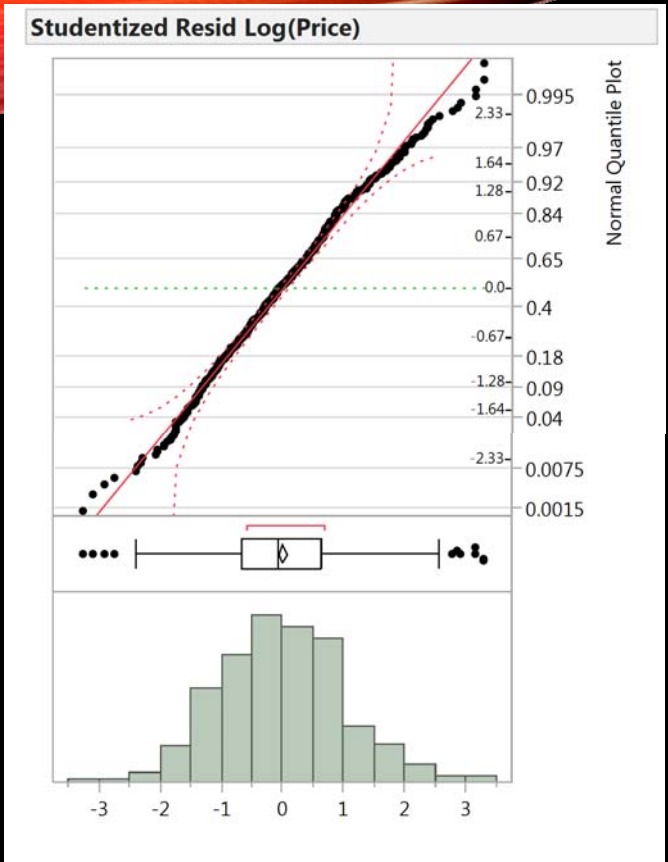
RSquare	0.97314
RSquare Adj	0.972665
Root Mean Square Error	0.109155
Mean of Response	8.048636
Observations (or Sum Wgts)	749

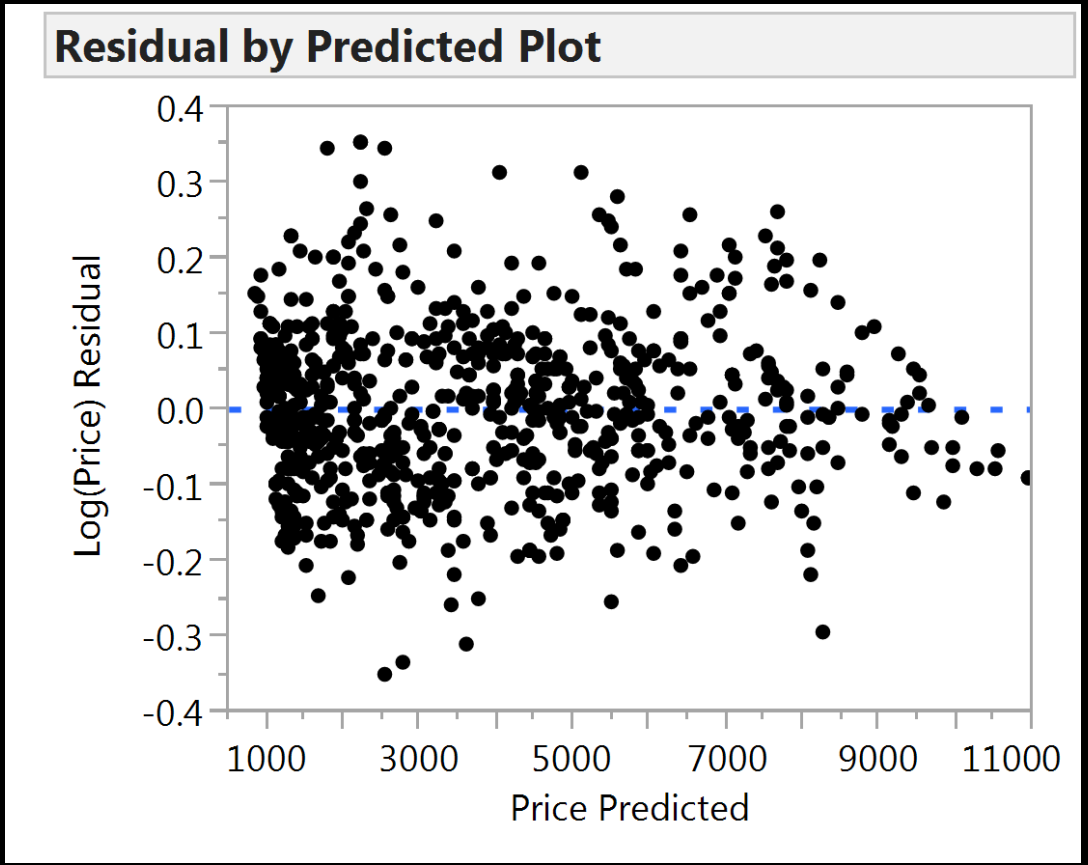
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	13	317.28394	24.4065	2048.396
Error	735	8.75746	0.0119	**Prob > F**
C. Total	748	326.04140		<.0001*

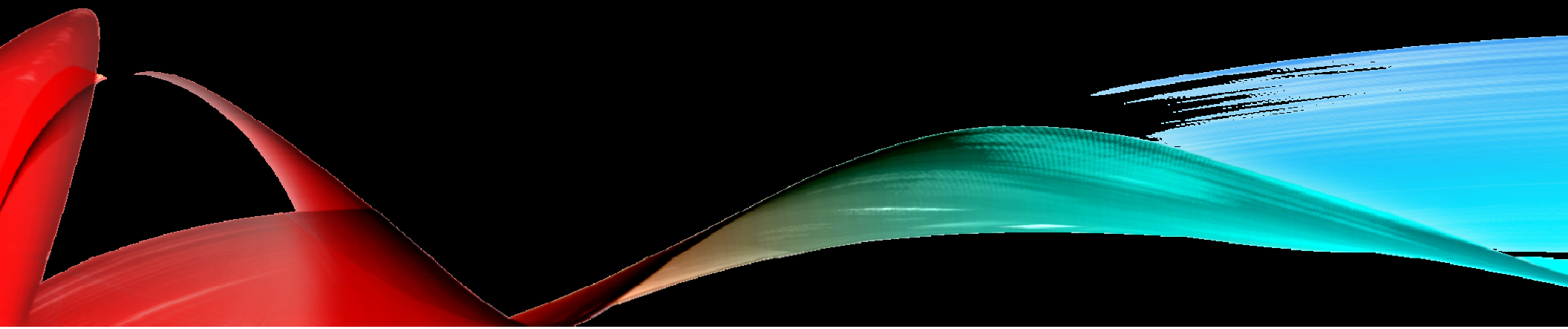
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	5.8094252	0.017071	340.31	<.0001*	.
Color[D]	0.3397239	0.006939	48.96	<.0001*	1.3145097
Color[G]	0.1206511	0.005837	20.67	<.0001*	1.0430003
Carat_Weight	2.8241661	0.018356	153.85	<.0001*	1.6608144
Clarity[IF]	0.3939524	0.014469	27.23	<.0001*	2.6973406
Clarity[SI1]	-0.266062	0.009108	-29.21	<.0001*	1.7797057
Clarity[SI2]	-0.419715	0.010324	-40.65	<.0001*	2.0378993
Clarity[VS1]	0.0225742	0.010166	2.22	0.0267*	1.8122294
Clarity[VS2]	-0.102391	0.008994	-11.38	<.0001*	1.6992538
Clarity[VVS1]	0.2150348	0.011975	17.96	<.0001*	2.197911
Cut[Excellent]	0.0217126	0.007274	2.99	0.0029*	2.8286118
Cut[Good]	-0.067848	0.013406	-5.06	<.0001*	3.7010049
Cut[Ideal]	0.0589345	0.012324	4.78	<.0001*	3.3523403
(Carat_Weight-0.8134)*(Carat_Weight-0.8134)	-1.611456	0.05763	-27.96	<.0001*	1.0696919


Studentized Resid Log(Price)
Normal Quantile Plot
0.995
0.97
0.92
0.84
0.65
0.4
0.18
0.09
0.04
0.0075
0.0015
2.33
1.64
1.28
0.67
0.0
-0.67
-1.28
-1.64
-2.33
-3
-2
-1
0
1
2
3


Residual by Predicted Plot
Log(Price) Residual
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
1000
3000
5000
7000
9000
11000
Price Predicted

TRANSFORMATIONS



Goals of Re-expression *

- Make the distribution of a variable (as seen in its histogram, for example) more symmetric.
- Make the spread of several groups (as seen in side-by-side boxplots) more alike
- Make the form of the scatterplot more nearly linear
- Make the scatter in a scatterplot or residual plot spread out evenly rather than following a fan shape

* pp. 528-530, Business Statistics, 3e, Sharpe, et al (Pearson)

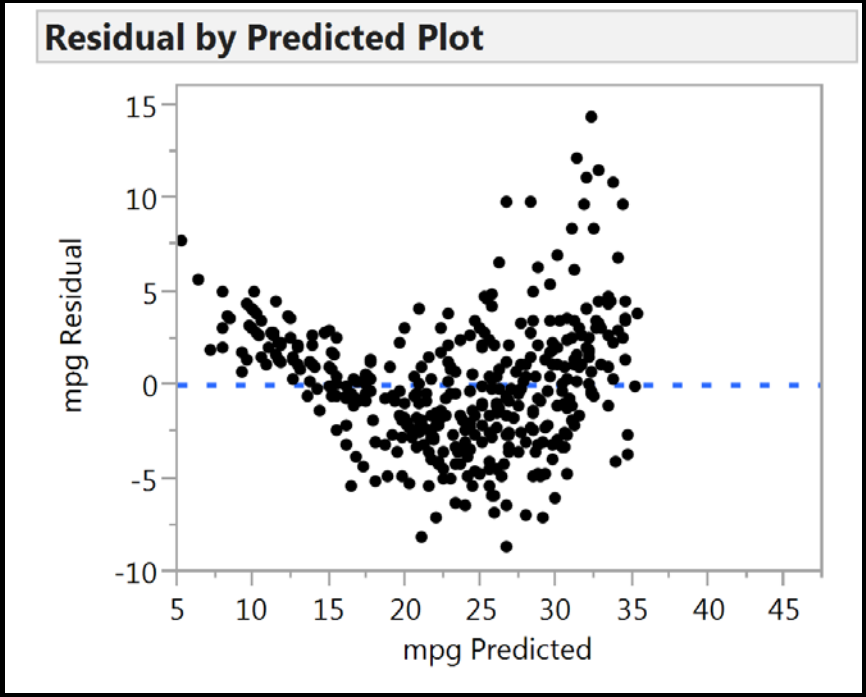
Table 4-5. Variance-stabilizing transformations

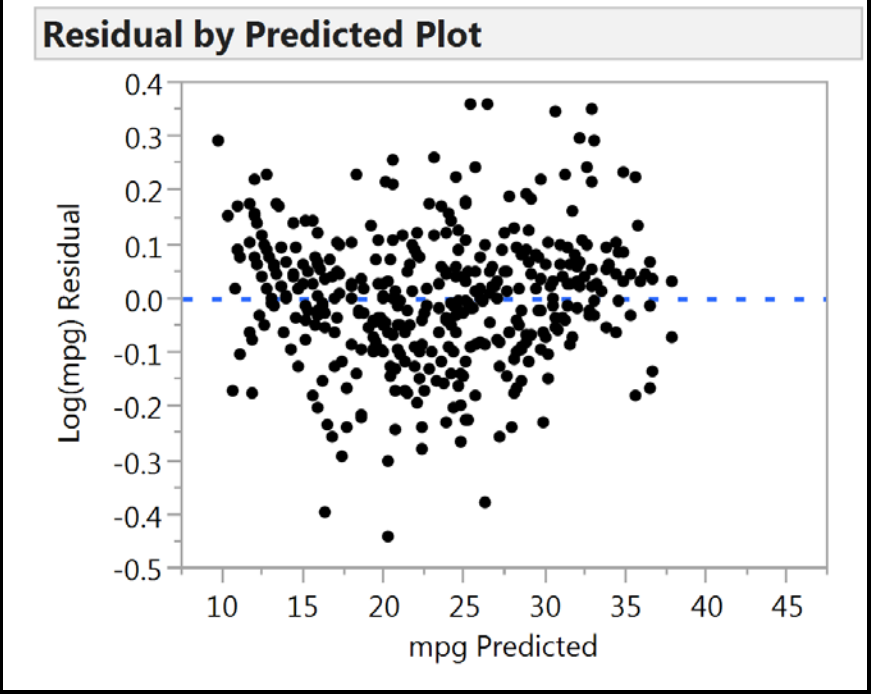
Power (a)	Transformation	Comment
0	None	Normal
0.5	Square root	Counts
1	Logarithm	Constant percent error
2	Inverse	Rate data

p. 82, DOE Simplified, Anderson and Whitcomb (Productivity Press)

See also p. 531, Business Statistics, 3e by Sharpe, De Veaux and Velleman

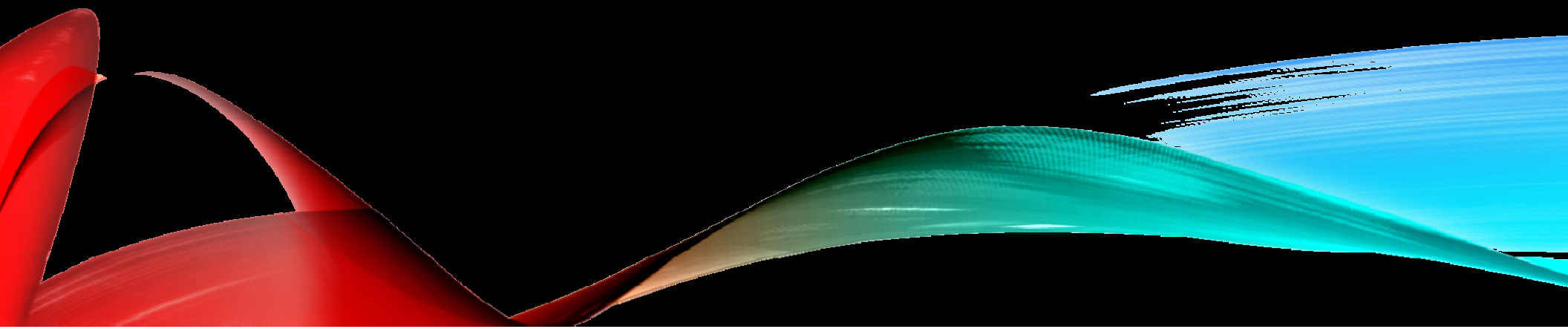
Impact of Log(mpg) – Auto Dataset





VARIABLE SELECTION / DIMENSION REDUCTION

Stepwise



Reducing the Number of Predictors

"However, there are several reasons for exercising caution before throwing all possible variables into a model.

- It may be expensive or not feasible to collect a full complement of predictors for future predictions.
- We may be able to measure fewer predictors more accurately (e.g., in surveys). The more predictors there are, the higher the chance of missing values in the data. If we delete or impute cases with missing values, multiple predictors will lead to a higher rate of case deletion or imputation.
- Parsimony is an important property of good models. We obtain more insight into the influence of predictors in models with few parameters.
- Estimates of regression coefficients are likely to be unstable, due to multicollinearity in models with many variables. (Multicollinearity is the presence of two or more predictors sharing the same linear relationship with the outcome variable.) Regression coefficients are more stable for parsimonious models. One very rough rule of thumb is to have a number of cases n larger than 5(p + 2), where p is the number of predictors.
- It can be shown that using predictors that are uncorrelated with the dependent variable increases the variance of predictions.
- It can be shown that dropping predictors that are actually correlated with the dependent variable can increase the average error (bias) of predictions."

Source: Data Mining for Business Intelligence, 2e by Shmueli, et al (Wiley), p. 127-8.

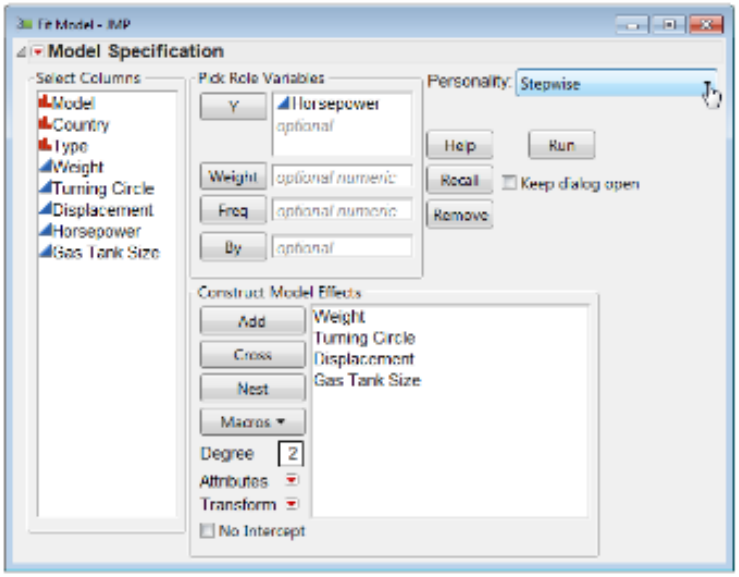
Stepwise Regression

Use for least squares or logistic regression variable selection, model comparison and model creation.

Stepwise Regression

1. From an open table, select **Analyze > Fit Model.**
2. Select a response variable from **Select Columns** and click **Y**.
3. Select predictor variables and click **Add**.
4. If desired, select a validation column **(JMP® Pro only)**.
5. Select **Stepwise** from the **Personality** drop-down menu.
6. In the resulting **Stepwise Fit** window (shown below):
   - Select a **Stopping Rule**.
   - Select the step **Direction** (forward, backward or mixed).
   - To run the regression automatically, click **Go**. To proceed manually, click **Step**.

Example: Car Physical Data.jmp (Help > Sample Data)

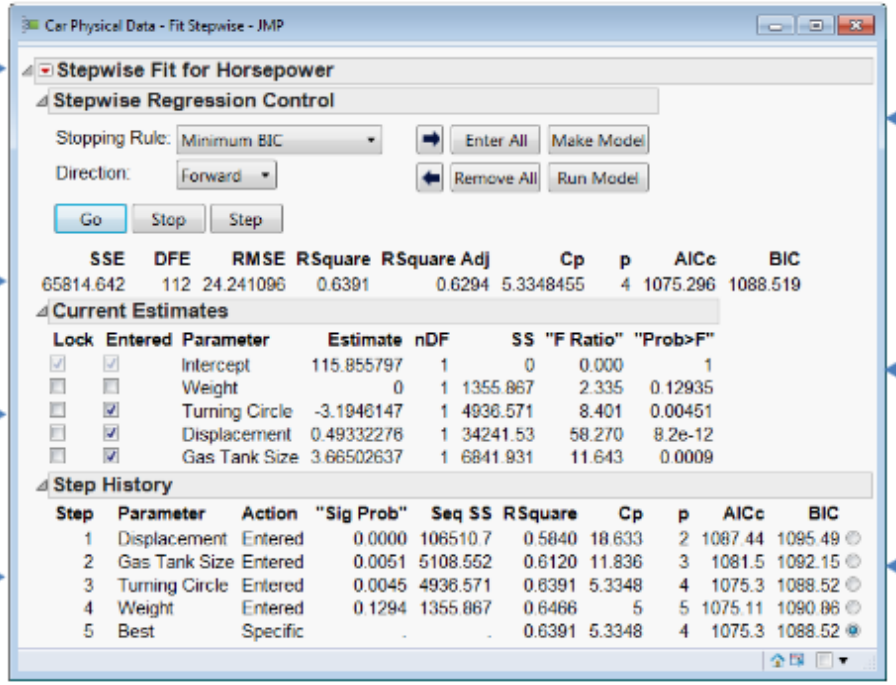


Click the **red triangle** for cross-validation, all possible models (all subsets regression), model averaging and other options.

Results for current model.

Check/uncheck **Entered** terms to change the model. **Locked** terms are used (or **not** used) in later steps.

Each time the model is changed, a new line is added to the **Step History** panel.



Use the **arrows** to step forward or backward.

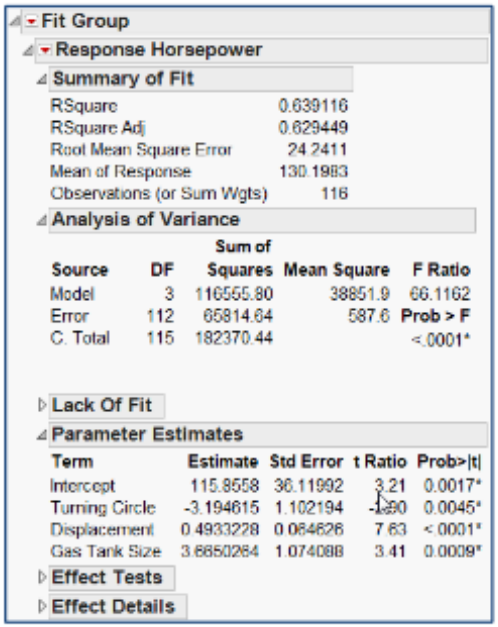
P-**values** are displayed under **Prob<F**.

Use the radio buttons to select a model.

Tips:

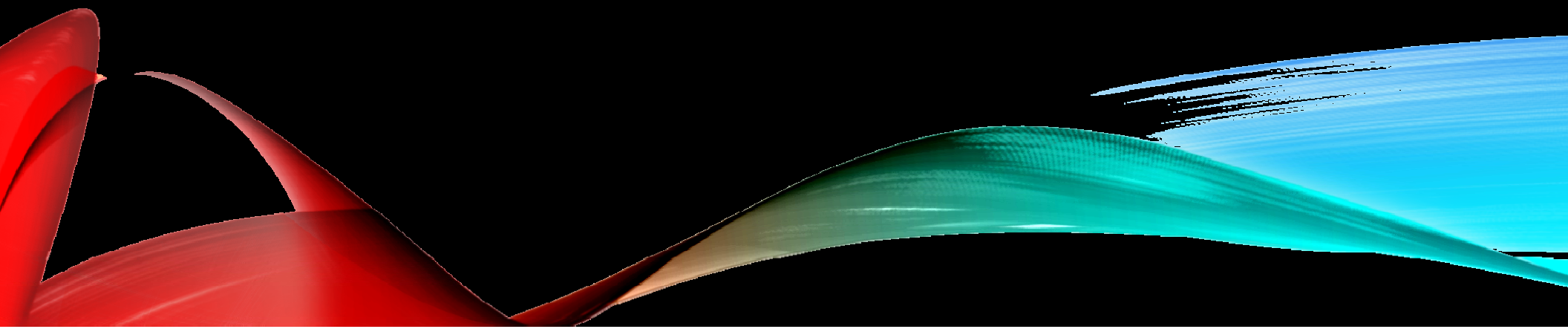
- For **Forward** regression, remove all terms, then click **Step** or **Go**.
- For **Backward** regression, enter all terms, then click **Step** or **Go**.
- The **Mixed** direction is only available with the p-value stopping rule.
- To run the model shown in the Current Estimates table, click **Run Model.** JMP generates a report, including fit statistics and information on parameter estimates and effect tests. See the **Multiple Linear Regression** or **Multiple Logistic Regression** one-page guides for details.

Note: For additional details search for "stepwise regression" in the JMP Help or in the ***Fitting Linear Models*** book (under **Help > Books**).



HOUSING PRICES EXAMPLE - AGAIN

Dataset: HousingPrices.jmp



A real estate company that manages properties around a ski resort in the United States wishes to improve its method for pricing homes. Sample data is obtained on a number of measures, including size of the home and property, location, age of the house, and a strength-of-market indicator.

The Data HousingPrices.jmp

The data set contains information on about 45 residential properties near a popular North American ski resort sold during a recent 12-month period, and is a representative sample of the full set of properties sold during that time period. The variables in the data set are:

Price: Selling price of the property (in thousands of dollars)

Beds: Number of bedrooms in the house

Baths: Number of bathrooms in the house

Square Feet: Size of the house in square feet

Miles to Resort: Miles from the property to the downtown resort area

Miles to Base: Miles from the property to the base of the ski resort's facing mountain

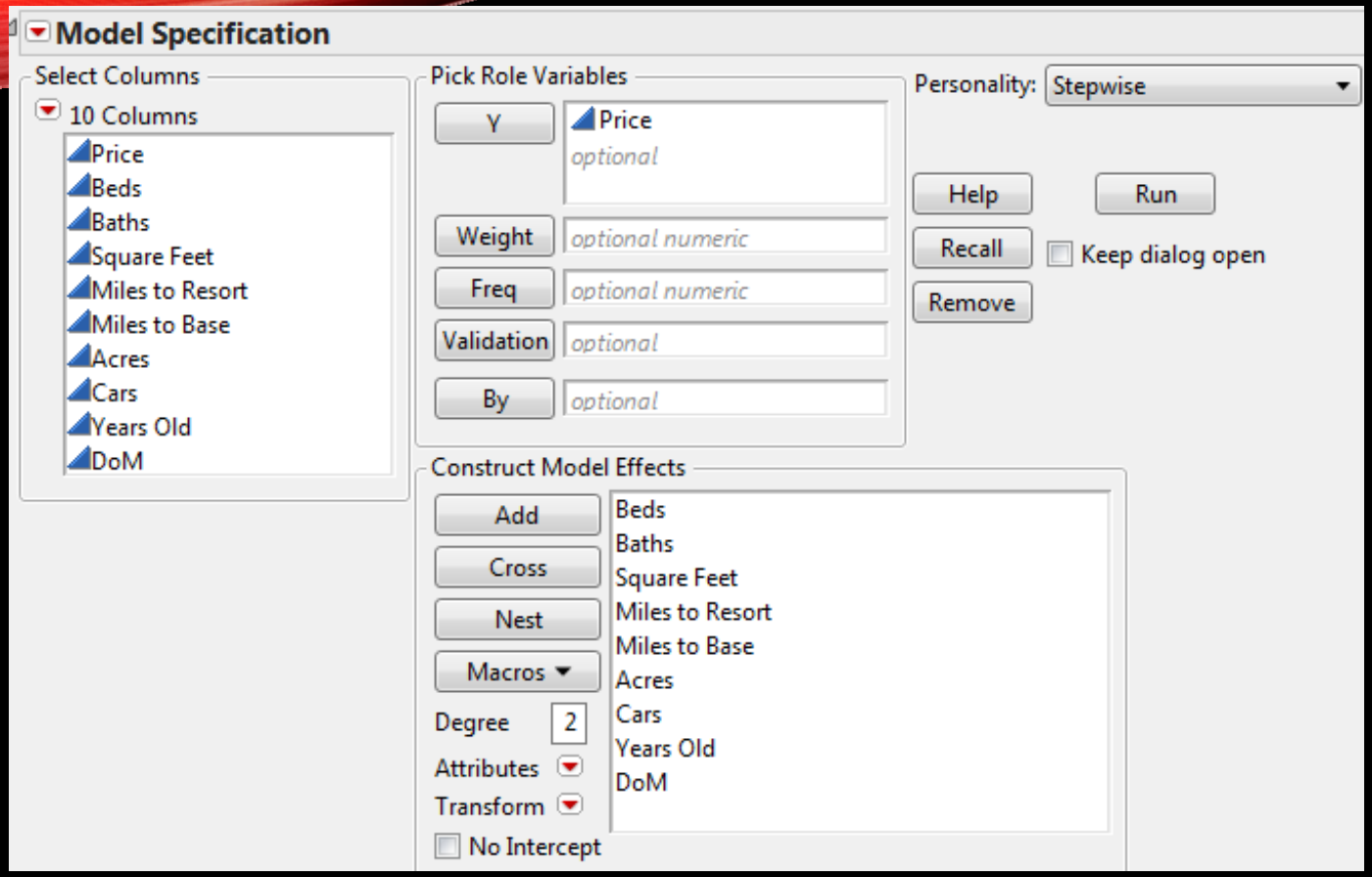
Acres: Lot size in number of acres

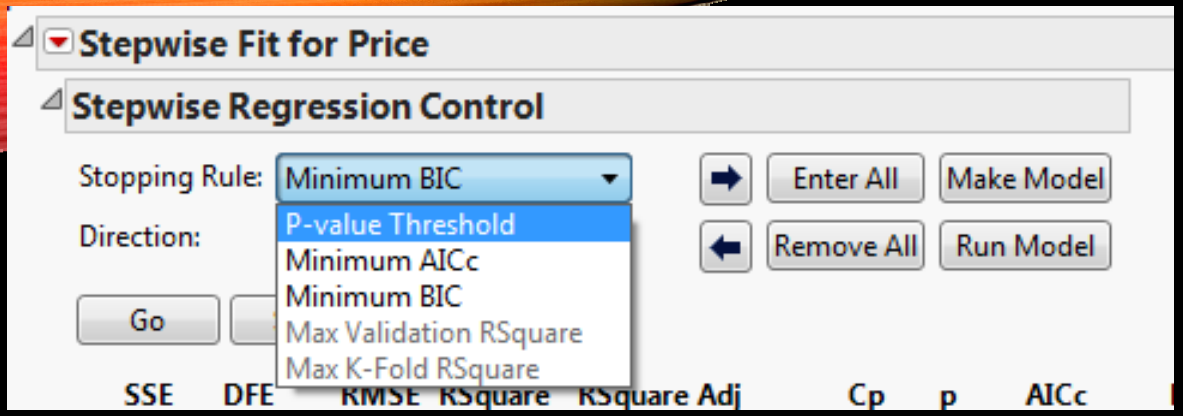
Cars: Number of cars that will fit into the garage

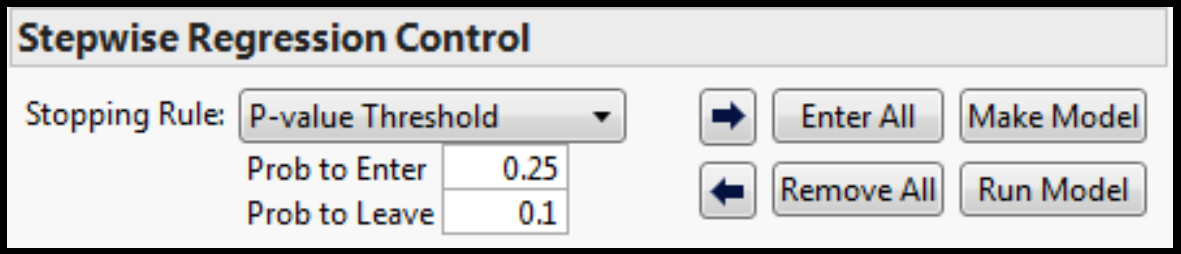
Years Old: Age of the house at the time it was listed in years

DoM: Number of days the house was on the market before it was sold

http://www.jmp.com/en_us/academic/case-study-library.html


Model Specification
Select Columns
10 Columns
Price
Beds
Baths
Square Feet
Miles to Resort
Miles to Base
Acres
Cars
Years Old
DoM
Pick Role Variables
Y
Price
optional
Weight
optional numeric
Freq
optional numeric
Validation
optional
By
optional
Personality:
Stepwise
Help
Run
Recall
Keep dialog open
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
Beds
Baths
Square Feet
Miles to Resort
Miles to Base
Acres
Cars
Years Old
DoM





This is equivalent to our manual process of deleting variables one at a time

"Two criteria for balancing under-fitting and over-fitting are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These criteria measure the information lost by fitting a given model. Therefore, models with smaller AIC and BIC values are considered better. AIC and BIC measure the goodness of fit (i.e. quality) of a model, but also include a penalty that is a function of the number of parameters in the model. As such, they can be used to compare various models for the same data set. It can be shown that Mallows's Cp is equivalent to AIC in the case of multiple linear regression. BIC applies a larger penalty than AIC does, and may lead to smaller models."

Source: Data Mining for Business Intelligence by Shmueli, et al (Wiley)

Stepwise Fit for Price

Stepwise Regression Control

Stopping Rule: P-value Threshold

Prob to Enter 0.25

Prob to Leave 0.1

Direction: Forward

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
153342.58	40	61.915785	0.8018	0.7820	1.3287813	5	507.9345	516.564

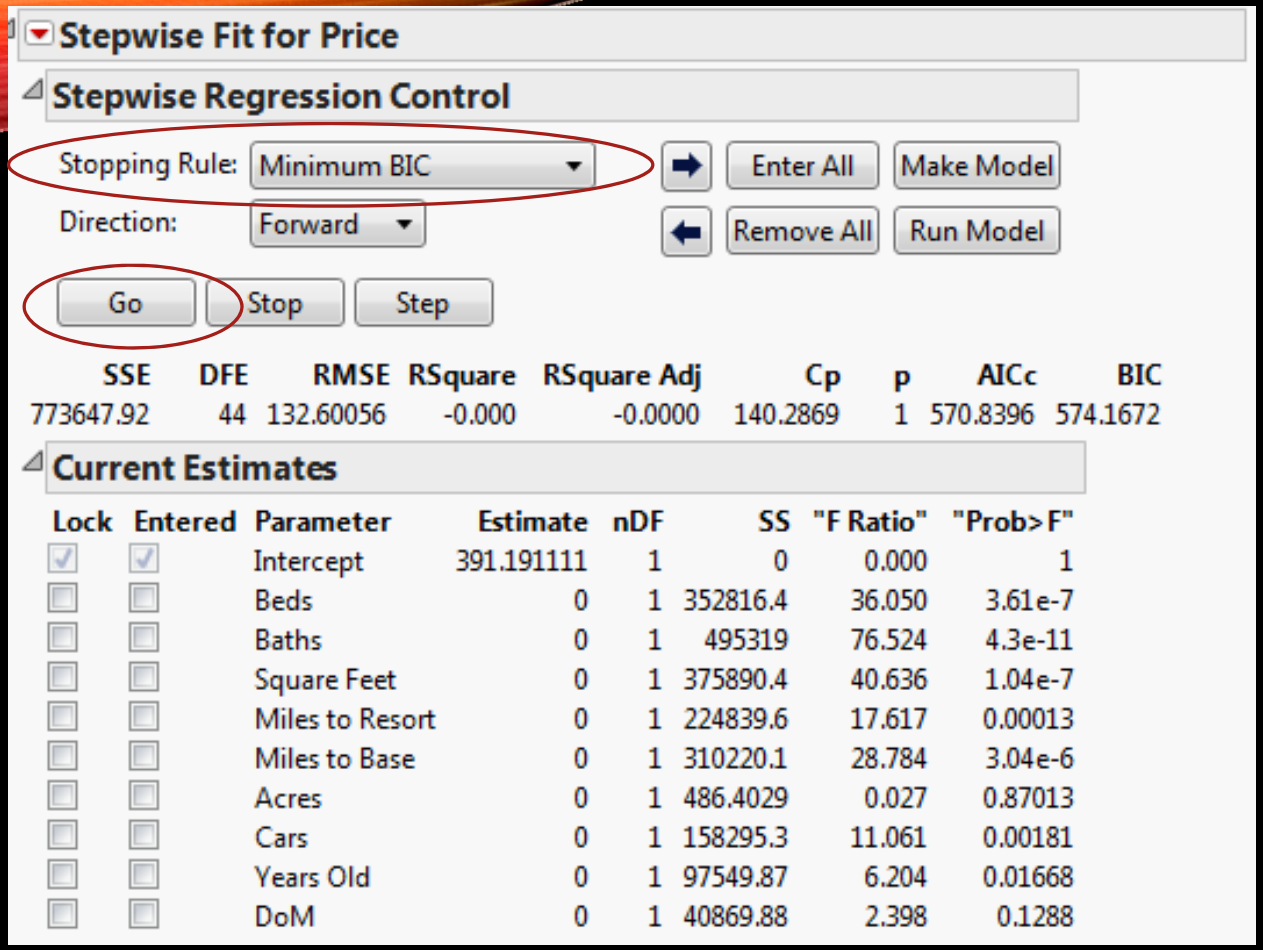
Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	197.150171	1	0	0.000	1
☐	☐	Beds	0	1	1148.056	0.294	0.59063
☐	☑	Baths	59.2080973	1	46387.91	12.100	0.00123
☐	☑	Square Feet	0.05112326	1	19597.28	5.112	0.02927
☐	☐	Miles to Resort	0	1	1735.229	0.446	0.508
☐	☑	Miles to Base	-3.7985104	1	90774.52	23.679	1.81e-5
☐	☑	Acres	5.00610917	1	46854.78	12.222	0.00117
☐	☐	Cars	0	1	311.3432	0.079	0.77968
☐	☐	Years Old	0	1	969.519	0.248	0.62118
☐	☐	DoM	0	1	408.8354	0.104	0.7485

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC	
1	Baths	Entered	0.0000	495319	0.6402	24.94	2	527.135	531.97	○
2	Miles to Base	Entered	0.0015	59830.74	0.7176	12.765	3	518.659	524.885	○
3	Acres	Entered	0.0021	45558.33	0.7765	3.9716	4	510.675	518.169	○
4	Square Feet	Entered	0.0293	19597.28	0.8018	1.3288	5	507.935	516.564	◉

This is equivalent to our manual process of deleting variables one at a time



Stepwise Fit for Price

Stepwise Regression Control

Stopping Rule: Minimum BIC | Enter All | Make Model

Direction: Forward | Remove All | Run Model

Go | Stop | Step

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
773647.92	44	132.60056	-0.000	-0.0000	140.2869	1	570.8396	574.1672

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	391.191111	1	0	0.000	1
☐	☐	Beds	0	1	352816.4	36.050	3.61e-7
☐	☐	Baths	0	1	495319	76.524	4.3e-11
☐	☐	Square Feet	0	1	375890.4	40.636	1.04e-7
☐	☐	Miles to Resort	0	1	224839.6	17.617	0.00013
☐	☐	Miles to Base	0	1	310220.1	28.784	3.04e-6
☐	☐	Acres	0	1	486.4029	0.027	0.87013
☐	☐	Cars	0	1	158295.3	11.061	0.00181
☐	☐	Years Old	0	1	97549.87	6.204	0.01668
☐	☐	DoM	0	1	40869.88	2.398	0.1288

We will use the BIC default

Select stopping rule and direction

Select “Go”

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
153342.58	40	61.915785	0.8018	0.7820	1.3287813	5	507.9345	516.564

Current Estimates

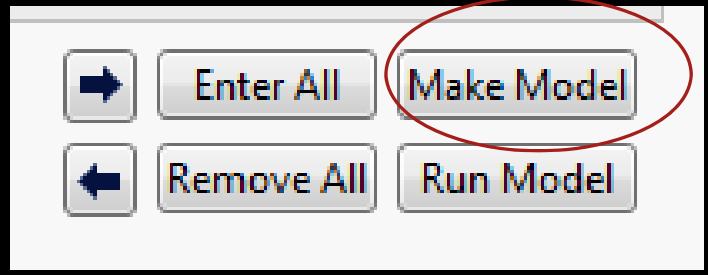
Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	197.150171	1	0	0.000	1
☐	☐	Beds	0	1	1148.056	0.294	0.59063
☐	☑	Baths	59.2080973	1	46387.91	12.100	0.00123
☐	☑	Square Feet	0.05112326	1	19597.28	5.112	0.02927
☐	☐	Miles to Resort	0	1	1735.229	0.446	0.508
☐	☑	Miles to Base	-3.7985104	1	90774.52	23.679	1.81e-5
☐	☑	Acres	5.00610917	1	46854.78	12.222	0.00117
☐	☐	Cars	0	1	311.3432	0.079	0.77968
☐	☐	Years Old	0	1	969.519	0.248	0.62118
☐	☐	DoM	0	1	408.8354	0.104	0.7485

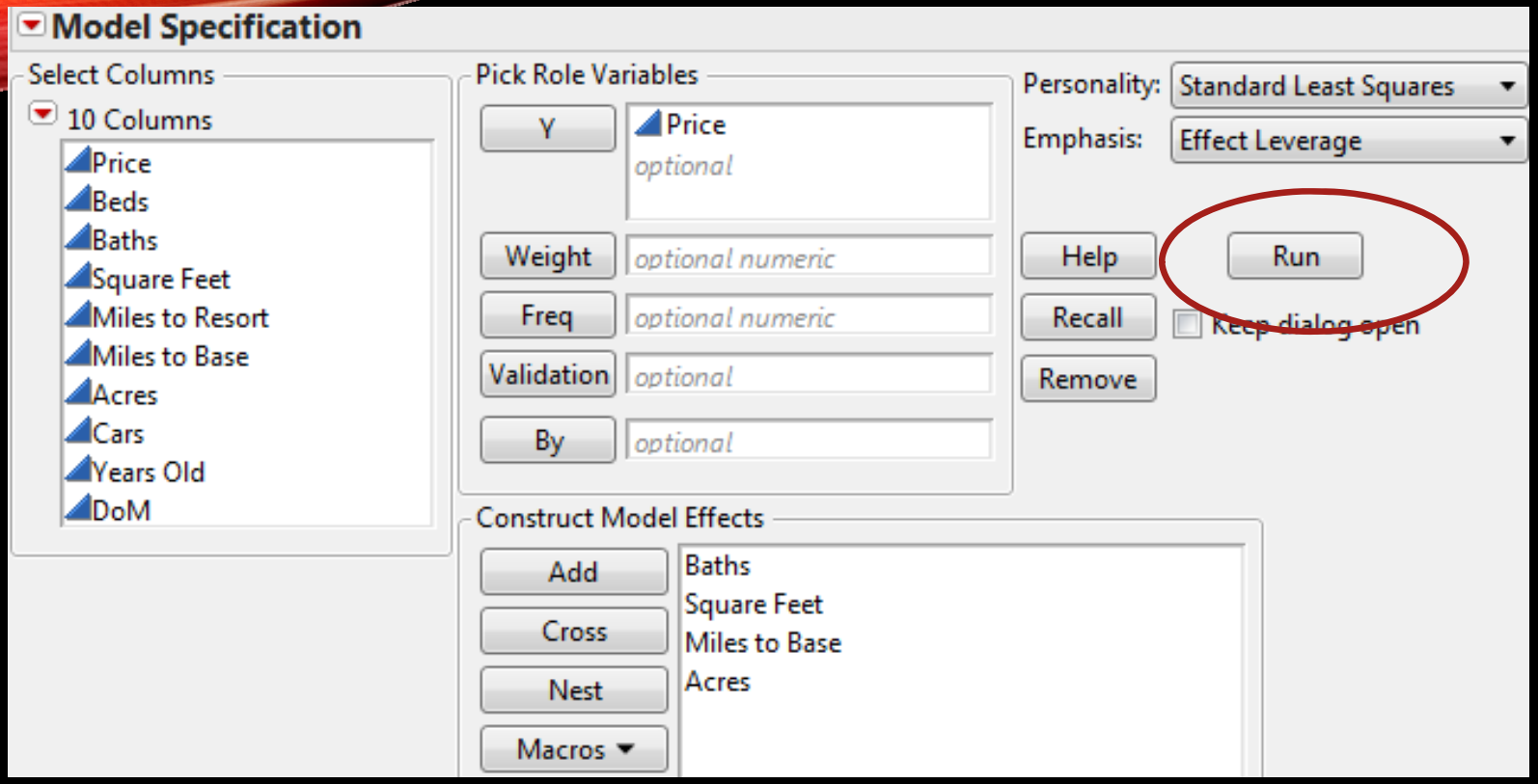
Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC	
1	Baths	Entered	0.0000	495319	0.6402	24.94	2	527.135	531.97	○
2	Miles to Base	Entered	0.0015	59830.74	0.7176	12.765	3	518.659	524.885	○
3	Acres	Entered	0.0021	45558.33	0.7765	3.9716	4	510.675	518.169	○
4	Square Feet	Entered	0.0293	19597.28	0.8018	1.3288	5	507.935	516.564	○
5	All	Removed	.	.	-0.000	140.29	1	570.84	574.167	○
6	Baths	Entered	0.0000	495319	0.6402	24.94	2	527.135	531.97	○
7	Miles to Base	Entered	0.0015	59830.74	0.7176	12.765	3	518.659	524.885	○
8	Acres	Entered	0.0021	45558.33	0.7765	3.9716	4	510.675	518.169	○
9	Square Feet	Entered	0.0293	19597.28	0.8018	1.3288	5	507.935	516.564	○
10	Miles to Resort	Entered	0.5080	1735.229	0.8040	2.9177	6	510.239	519.859	○
11	Years Old	Entered	0.5013	1816.724	0.8064	4.4873	7	512.669	523.123	○
12	Beds	Entered	0.7110	562.3105	0.8071	6.3541	8	515.643	526.76	○
13	Cars	Entered	0.5854	1245.656	0.8087	8.0589	9	518.594	530.19	○
14	DoM	Entered	0.8096	248.8224	0.8090	10	10	522.047	533.92	○
15	Best	Specific	.	.	0.8018	1.3288	5	507.935	516.564	◉

We see the final selected variables and the selection analysis results

If satisfied select Make Model




Model Specification
Select Columns
10 Columns
Price
Beds
Baths
Square Feet
Miles to Resort
Miles to Base
Acres
Cars
Years Old
DoM
Pick Role Variables
Y
Price
optional
Weight
optional numeric
Freq
optional numeric
Validation
optional
By
optional
Personality:
Standard Least Squares
Emphasis:
Effect Leverage
Help
Run
Recall
Keep dialog open
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Baths
Square Feet
Miles to Base
Acres

Creates our typical model results

Summary of Fit

RSquare	0.801793
RSquare Adj	0.781972
Root Mean Square Error	61.91579
Mean of Response	391.1911
Observations (or Sum Wgts)	45

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	620305.34	155076	40.4523
Error	40	153342.58	3834	**Prob > F**
C. Total	44	773647.92		<.0001*