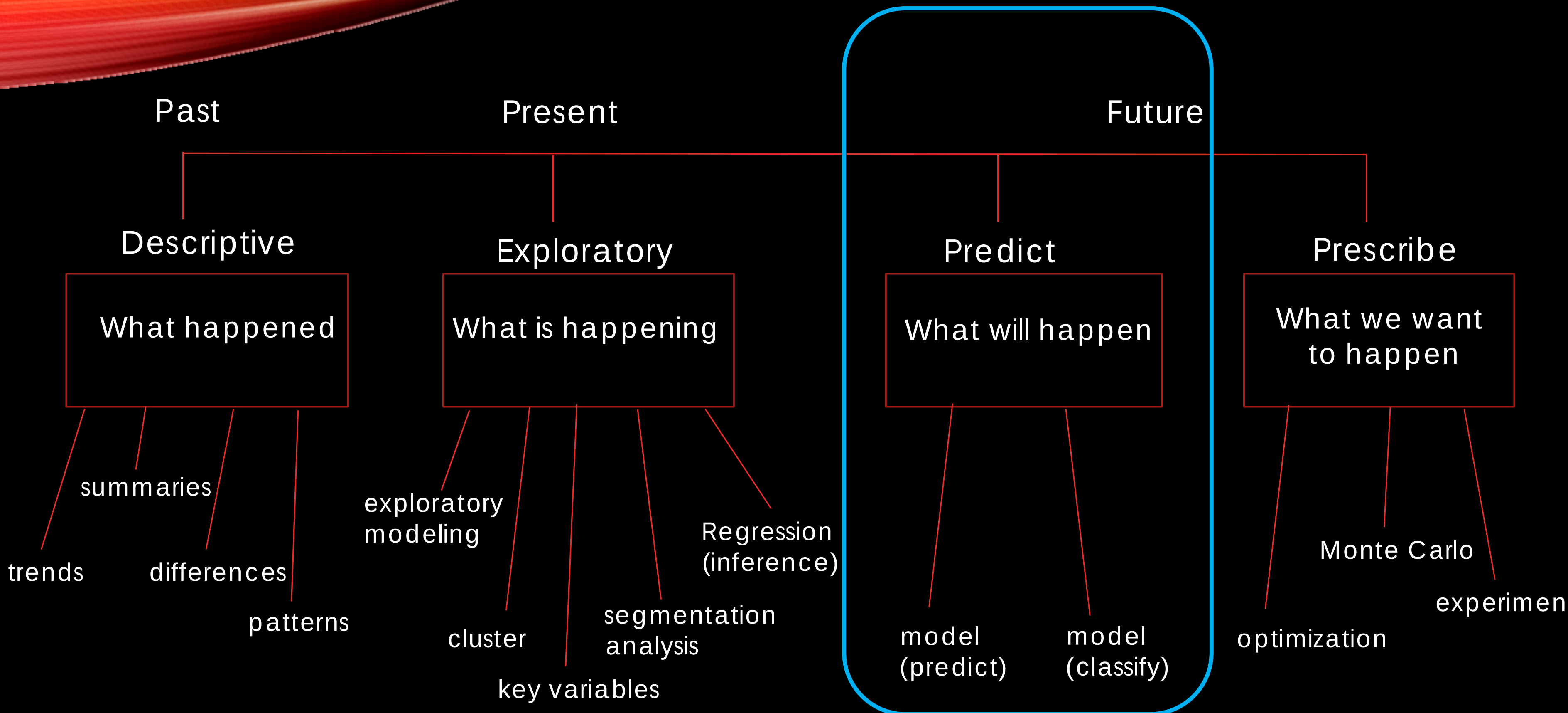
PREDICTIVE MODELING: MULTIPLE REGRESSION

Multiple Regression Predictive Modeling Learning Goals

1. Be able to explain the difference in goals of a predictive model vs an explanatory / exploratory model
2. Be able to explain the fundamental need of cross validation for predictive modeling
3. Be able to create a validation column for predictive modeling
4. Be able to use JMP to construct a MR predictive model
5. Be able to use the stepwise process in the construction of a MR predictive model
6. Be able to interpret the predictive performance results of a MR predictive model
7. Be able to compare performance of competing MR predictive models
8. Be able to explain to a manager the insights from a MR predictive model using the profiler and other graphical tools



Source: Jim Grayson, Ph.D. and Mia Stephens

Predict

What will happen

model (continuous response)

- Multiple Regression
- Regression Trees
- Bootstrap Forest
- Boosted Tree
- Neural Network

model (categorical response)

- Decision Trees
- Logistic Regression
- Discriminant Analysis
- Bootstrap Forest
- Boosted Tree
- Neural Network

Source: Jim Grayson, Ph.D. and Mia Stephens

Inference

"We are often interested in understanding the way that Y is affected as $X1, \ldots, Xp$ change. In this situation we wish to estimate f, but our goal is not necessarily to make predictions for Y. We instead want to understand the relationship between X and Y, or more specifically, to understand how Y changes as a function of $X1, \ldots, Xp$. Now $\hat{f}$ cannot be treated as a black box, because we need to know its exact form."

Excerpts from pages 19-20, An Introduction to Statistical Learning, James, et al (Springer)

Prediction

"… consider a company that is interested in conducting a direct-marketing campaign. The goal is to identify individuals who will respond positively to a mailing, based on observations of demographic variables measured on each individual. In this case, the demographic variables serve as predictors, and response to the marketing campaign (either positive or negative) serves as the outcome. The company is not interested in obtaining a deep understanding of the relationships between each individual predictor and the response; instead, the company simply wants an accurate model to predict the response using the predictors. This is an example of modeling for prediction."

Excerpts from pages 19-20, An Introduction to Statistical Learning, James, et al (Springer)

Inference & Prediction

"Finally, some modeling could be conducted both for prediction and inference. For example, in a real estate setting, one may seek to relate values of homes to inputs such as crime rate, zoning, distance from a river, air quality, schools, income level of community, size of houses, and so forth. In this case one might be interested in how the individual input variables affect the prices—that is, *how much extra will a house be worth if it has a view of the river?* This is an inference problem. Alternatively, one may simply be interested in predicting the value of a home given its characteristics: *is this house under- or over-valued?* This is a prediction problem."

Excerpts from pages 19-20, An Introduction to Statistical Learning, James, et al (Springer)

Prediction

"… consider a company that is interested in conducting a direct-marketing campaign. The goal is to identify individuals who will respond positively to a mailing, based on observations of demographic variables measured on each individual. In this case, the demographic variables serve as predictors, and response to the marketing campaign (either positive or negative) serves as the outcome. The company is not interested in obtaining a deep understanding of the relationships between each individual predictor and the response; instead, the company simply wants an accurate model to predict the response using the predictors. This is an example of modeling for prediction."

Excerpts from pages 19-20, An Introduction to Statistical Learning, James, et al (Springer)

"But, in general, we do not really care how well the method works on the training data. Rather, *we are interested in the accuracy of the predictions that we obtain when we apply our method to* ***previously unseen test data***."

page 30, An Introduction to Statistical Learning, James, et al (Springer)

- What is it?
- What can it do? (use cases)
- How does it work?
- JMP Mechanics
- Interpret results (statistically)
- Interpret results (application)
- How to apply the results
- How to understand the managerial implications

WHAT IS IT?

Multiple Linear Regression is an approach for predicting a quantitative response variable (Y) with one ore more predictor variables (X).

Y	X	Objective	Summary of Fit Measures	Statistical Significance Measure	Operational Significance	Mgt Insights
Continuous	Continuous or Categorical	Predictive	R Square, RASE (validation)	Prob > F (p-value)	Mean and Individual Confidence Limits	Profiler; Variable Importance

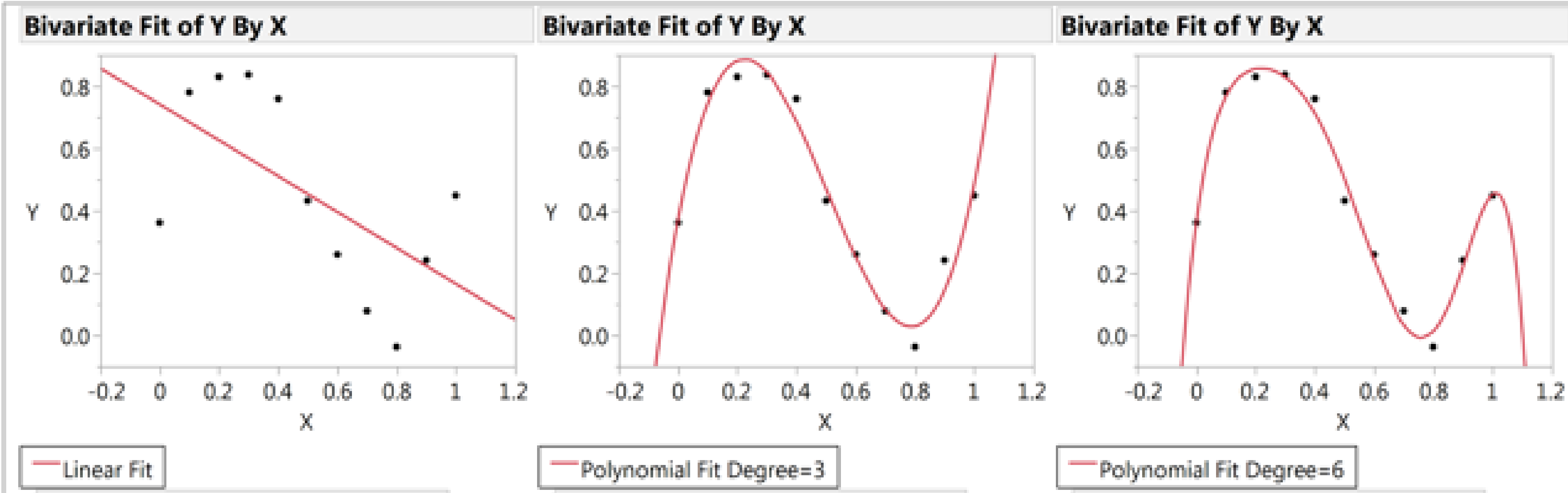
All performance in Predictive Analytics is based on the validation set and not the training set

HOW DOES IT WORK?

"A famous expression attributed to the imminent statistician George Box is "essentially, all models are wrong, but some are useful" (Box and Draper, 1987). Useful models give us accurate insights into the relationships between the inputs and outputs of the process, and allow us, for a given set of inputs, to make good predictions about the output. However, according to George Box, any statistical model we choose to describe the behavior of a process or make predictions about a process is, at best, an approximation. "

Draft version - Chapter 8, Building Better Models with JMP, Grayson, Gardner and Stephens (SAS)

Figure 8.1: Competing Models



Linear Fit

Summary of Fit

RSquare	0.37964
RSquare Adj	0.310711
Root Mean Square Error	0.257483
Mean of Response	0.453454
Observations (or Sum Wgts)	11

Polynomial Fit Degree=3

Summary of Fit

RSquare	0.972902
RSquare Adj	0.961288
Root Mean Square Error	0.06102
Mean of Response	0.453454
Observations (or Sum Wgts)	11

Polynomial Fit Degree=6

Summary of Fit

RSquare	0.986229
RSquare Adj	0.965571
Root Mean Square Error	0.057545
Mean of Response	0.453454
Observations (or Sum Wgts)	11

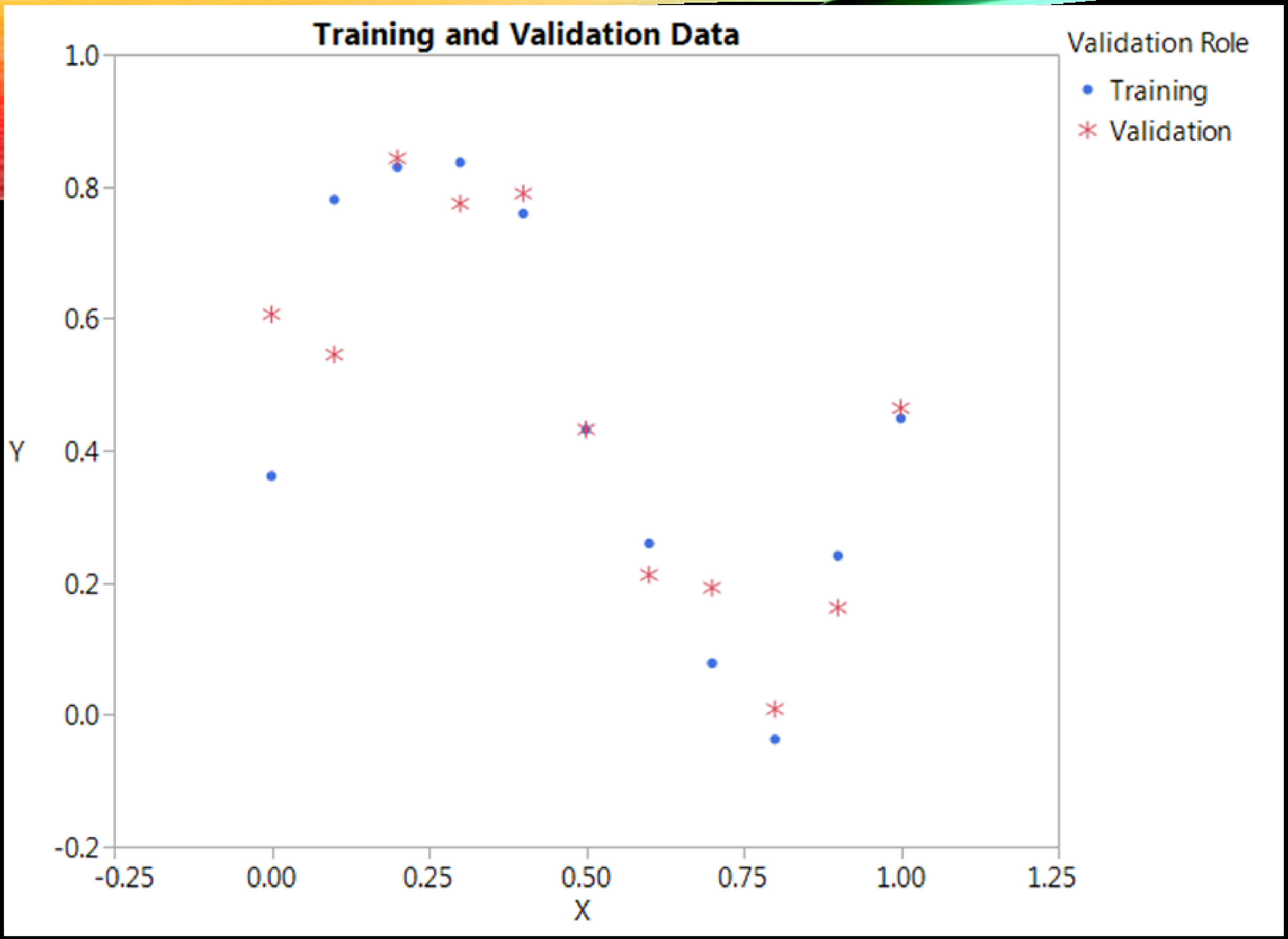
The fitted equations for the three models are:

$$\text{Model 1: } \hat{Y} = 0.742 - 0.576X$$

$$\text{Model 2: } \hat{Y} = 0.370 + 5.08X - 14.6X^2 + 9.59X^3$$

$$\text{Model 3: } \hat{Y} = 0.363 + 7.09X - 41.8X^2 + 137X^3 - 261X^4 + 245X^5 - 85.2X^6$$

Cross Validation Method: Hold Out Sample



Draft version - Chapter 8, Building Better Models with JMP, Grayson, Gardner and Stephens (SAS)

Tradeoff: Fitting Three Models

Measures of Fit for Y

Source Table	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
Polynomial Models Training Data	Predicted Y (6th Order)	Bivariate Polynomial Fit Degree=6		0.9862	0.0347	0.0276	11
Polynomial Models Training Data	Predicted Y (Cubic)	Bivariate Polynomial Fit Degree=3		0.9729	0.0487	0.0391	11
Polynomial Models Training Data	Predicted Y (Linear)	Bivariate Linear Fit		0.3796	0.2329	0.2029	11
Polynomial Models Validation Data	Predicted Y (6th Order)	Bivariate Polynomial Fit Degree=6		0.8130	0.1173	0.0853	11
Polynomial Models Validation Data	Predicted Y (Cubic)	Bivariate Polynomial Fit Degree=3		$\boxed{0.8429}$	0.1075	0.0807	11
Polynomial Models Validation Data	Predicted Y (Linear)	Bivariate Linear Fit		0.4727	0.1969	0.1779	11

Tradeoff: Bias v Variance

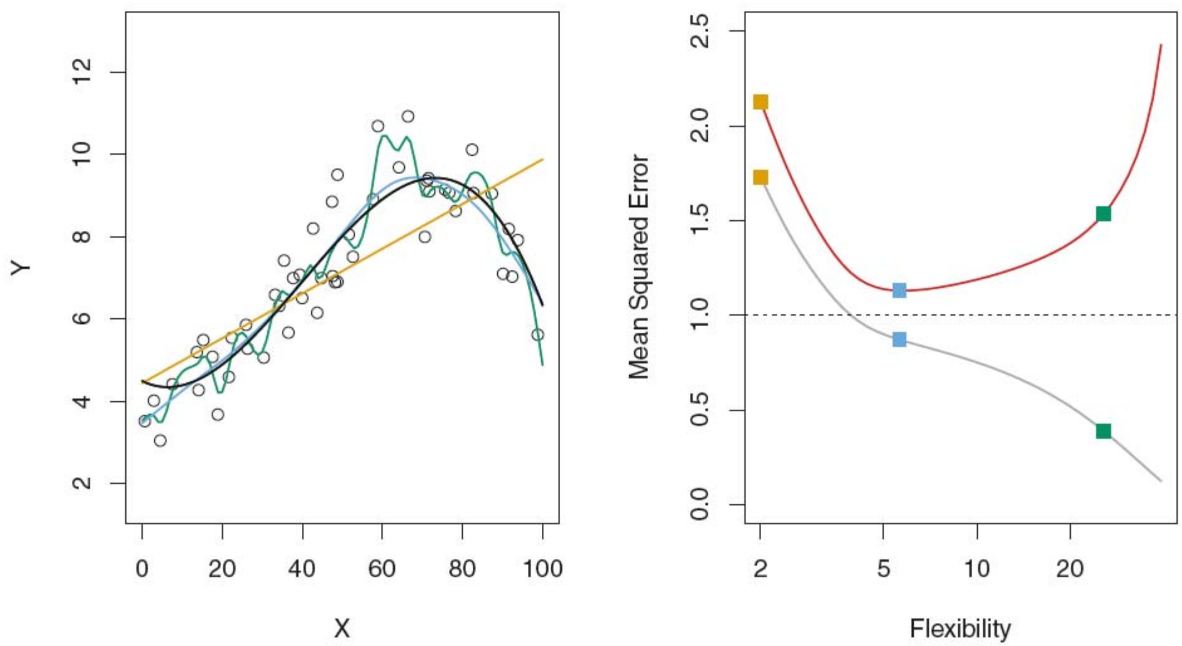


FIGURE 2.9. Left: *Data simulated from f, shown in black. Three estimates of f are shown: the linear regression line (orange curve), and two smoothing spline fits (blue and green curves).* Right: *Training MSE (grey curve), test MSE (red curve), and minimum possible test MSE over all methods (dashed line). Squares represent the training and test MSEs for the three fits shown in the left-hand panel.*

Figure 2.8, An Introduction to Statistical Learning, James, et al (Springer)

"Variance refers to the amount by which the f-hat would change if we estimated it using a different training data set.

… bias refers to the error that is introduced by approximating a real-life problem, which may be extremely complicated, by a much simpler model."

Excerpts from pages 34-35, An Introduction to Statistical Learning, James, et al (Springer)

Tradeoff: Flexibility v Interpretability

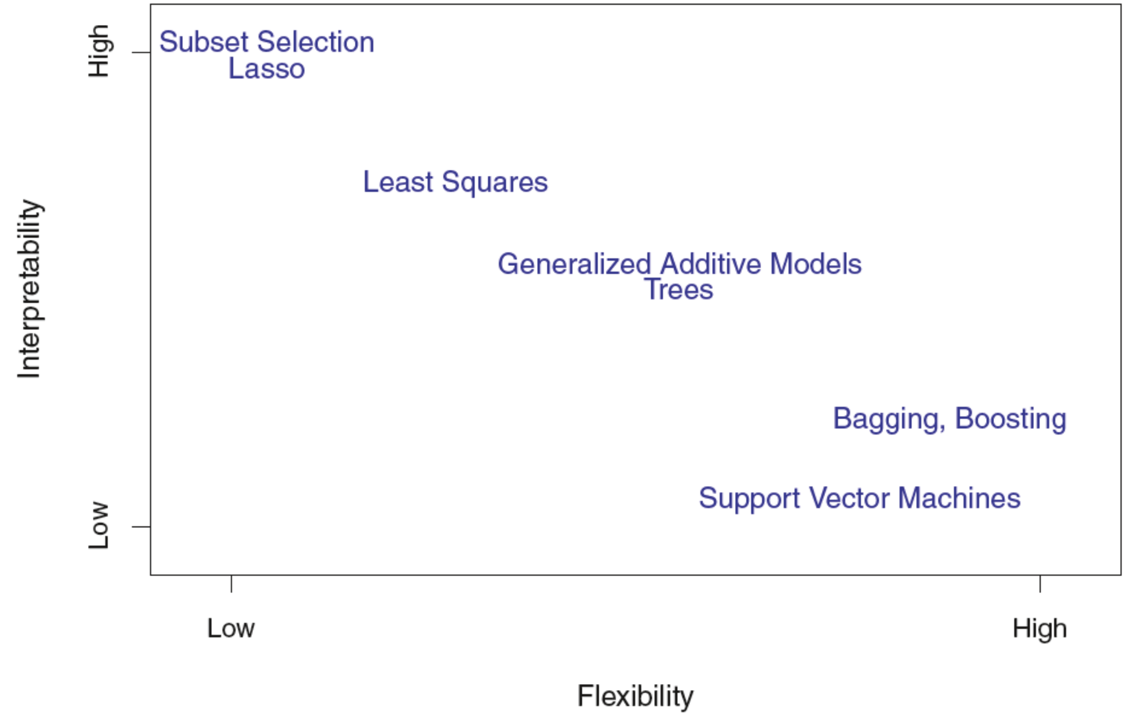


FIGURE 2.7. *A representation of the tradeoff between flexibility and interpretability, using different statistical learning methods. In general, as the flexibility of a method increases, its interpretability decreases.*

Figure 2.7, An Introduction to Statistical Learning, James, et al (Springer)

WHAT CAN IT DO? (USE CASES)

TECH 2/16/2012 @ 11:02AM | 2,665,449 views

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Andrew Pole had just started working as a statistician for Target in 2002, when two colleagues from the marketing department stopped by his desk to ask an odd question: “If we wanted to figure out if a customer is pregnant, even if she didn’t want us to know, can you do that? ”

http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

So Target started sending coupons for baby items to customers according to their pregnancy scores. Duhigg shares an anecdote — so good that it sounds made up — that conveys how eerily accurate the targeting is.

An angry man went into a Target outside of Minneapolis, demanding to talk to a manager:

http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

"My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"

The manager didn't have any idea what the man was talking about. He looked at the mailer. Sure enough, it was addressed to the man's daughter and contained advertisements for maternity clothing, nursery furniture and pictures of smiling infants. The manager apologized and then called a few days later to apologize again.

http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

On the phone, though, the father was somewhat abashed. "I had a talk with my daughter," he said. "It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology."

http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

The New York Times Magazine

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

How Companies Learn Your Secrets

Almost every major retailer, from grocery chains to investment banks to the U.S. Postal Service, has a "predictive analytics" department devoted to understanding not just consumers' shopping habits but also their personal habits, so as to more efficiently market to them. "But Target has always been one of the smartest at this," says Eric Siegel, a consultant and the chairman of a conference called Predictive Analytics World. "We're living through a golden age of behavioral research. It's amazing how much we can figure out about how people think now."

http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp

Increasing Marketable Universe

A software retailer captures $ 35 million in new revenue by increasing its marketable universe by a factor of four.

Situation: A well-known software vendor catering to creative professionals was seeking new prospects for its second most popular product, called Product X. Current practice was to market the new version of Product X to users and trial downloaders of earlier versions. The analytical task was to find a new revenue source for this product.

Jain, Piyanka; Sharma, Puneet (2014-11-05). Behind Every Good Decision: How Anyone Can Use Business Analytics to Turn Data Into Profitable Insight (Kindle Locations 1609-1616). AMACOM. Kindle Edition.

Increasing Marketable Universe con't

Action: The retailer had a database of 50 million prospects who had downloaded a free version of other products. We saw this as a great opportunity to mine for prospects who might have similar attributes to users of this product, although they had not shown any prior interest in Product X. Using logistic regression, we scored each prospect with an adoption probability score (to adopt Product X) based on the attributes of past users of Product X. The top 4 million prospects were then sent an offer for the latest version of Product X.

Impact: The model increased the retailer's addressable market by a factor of four and resulted in $ 35 million in new direct revenue. Additionally, this prospect set became a source of incremental revenue for all future releases of Product X.

Jain, Piyanka; Sharma, Puneet (2014-11-05). Behind Every Good Decision: How Anyone Can Use Business Analytics to Turn Data Into Profitable Insight (Kindle Locations 1609-1616). AMACOM. Kindle Edition.

Product Recommendation Engine

A payments company reduces marketing spend by 70 percent while collecting more than $ 20 million in incremental profits through a product recommendation engine.

Situation: A payments company had multiple payment products to fit the needs of its customers— the merchants. It was observed that merchants who owned more than one product were more engaged and had wider and deeper product usage that resulted in higher profitability. Because of this, the marketing team decided to offer every product to every merchant through marketing campaigns, but this resulted in confused merchants and high unsubscribes. So, the head of marketing wanted to figure out how to find the Next Best Product (NBP) recommendation for each merchant , thereby optimizing adoption as well as profitability.

Action: We used a multiclass decision tree to identify segments of merchants with higher adoption rates for certain products over others. We then combined the adoption data and historical profits to find the expected incremental profit (EIP) per product per segment. With this recommendation engine, the product with the highest EIP became the next best product recommendation for merchants with similar attributes to those in the particular segment. The model was then scored in the database to be used for outbound and inbound marketing.

Impact: Using the NBP recommendation engine, the marketing team now knew which offer to send to whom, resulting in a 70 percent reduction in marketing spending. Additionally, offering the right products using the NBP score resulted in sixfold increase in conversion and an increase of more than $ 20 million in profits just from the outbound marketing effort.

Jain, Piyanka; Sharma, Puneet (2014-11-05). Behind Every Good Decision: How Anyone Can Use Business Analytics to Turn Data Into Profitable Insight (Kindle Locations 1609-1616). AMACOM. Kindle Edition.

TECHNIQUE	DESCRIPTION	APPLICATION
Linear regression	Approach to model linear relationship between scalar dependent variable and one or more independent variables.	Customer lifetime value, cost of acquisition.
Logistic regression	Special case of linear regression where dependent variable is binary in nature.	Churn or attrition model, fraud detection model, and response model.
Decision tree	Type of tree diagram used to determine best classification of population (based on independent variable) to optimize prediction of dependent variable.	Cross-sell product prediction, customer segmentation.

Jain, Piyanka; Sharma, Puneet (2014-11-05). Behind Every Good Decision: How Anyone Can Use Business Analytics to Turn Data Into Profitable Insight (Kindle Locations 1726). AMACOM. Kindle Edition.

K-means clustering	Segmentation technique to partition population (number of observations) into clusters in which each observation belongs to the cluster with the nearest mean.	Unsupervised customer segmentation, that is, clustering for statistically similar attributes, but not driven by any target variable such as churn or customer lifetime value.
Time series forecasting	Techniques used to forecast future events based on known past values of same event.	Sales over time, forecasting.
Survival analysis	Technique used to predict time to event.	Hitting credit limit, customer tenure.
Neural networks	Generalizations of existing statistical models; black box; hard to understand but more powerful than other techniques.	Fraud detection, response model, and many others.

Jain, Piyanka; Sharma, Puneet (2014-11-05). Behind Every Good Decision: How Anyone Can Use Business Analytics to Turn Data Into Profitable Insight (Kindle Locations 1726). AMACOM. Kindle Edition.

JMP MECHANICS

Creating a Validation Column (Holdout Sample)

This page describes how to create a validation column in JMP®. Validation, or out-of-sample cross-validation, is used to assess the predictive ability of a model. Different methods for model validation are available in JMP. In **JMP Pro**, a validation column (example at right) can be used for automated model cross-validation in many modeling platforms.

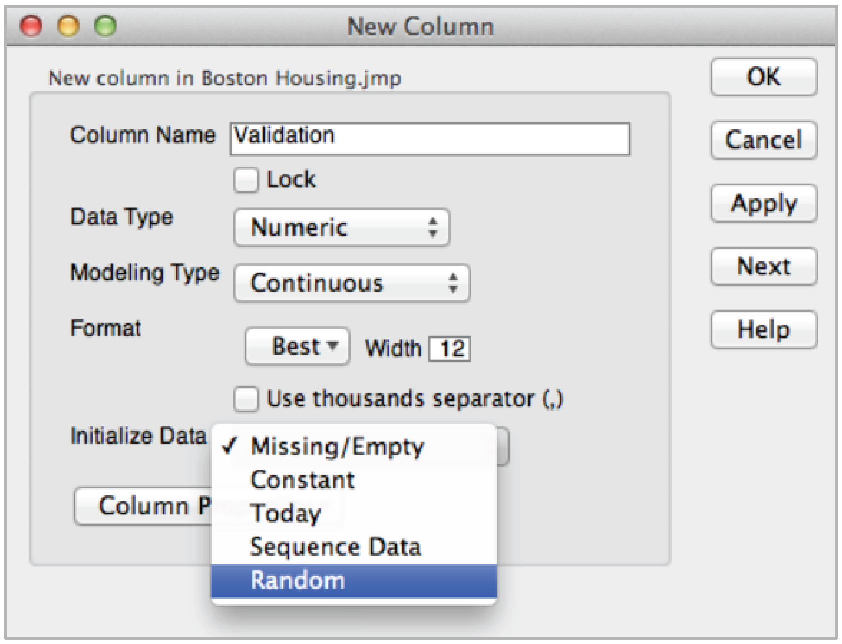
Validation
Test
Test
Train
Test
Train
Test
Train
Validate
Train
Train
Validate
Train

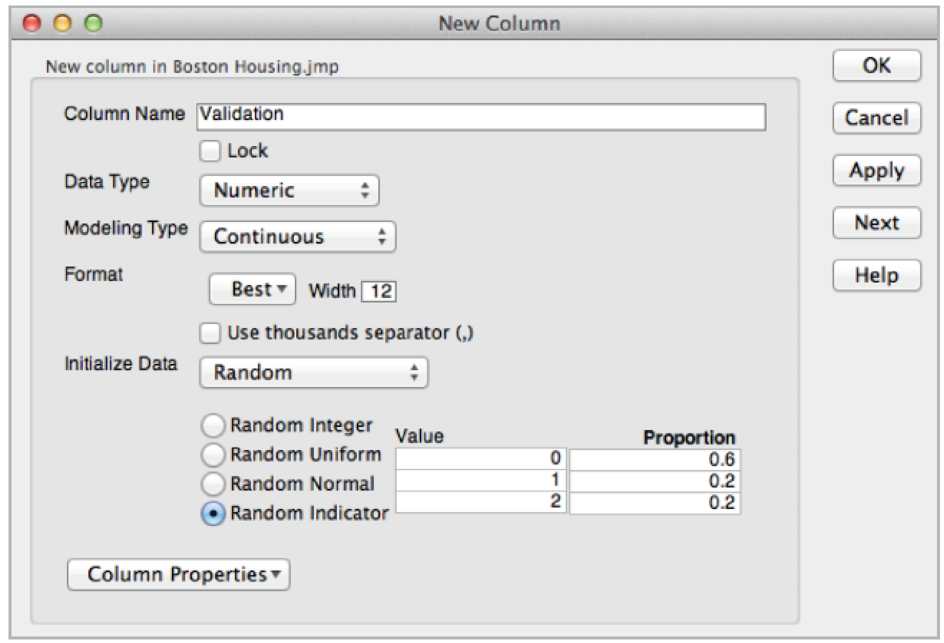
Creating a Validation Column (Train, Validate, Test)

1. From an open JMP data table, select **New Column** from the **Cols** menu.
2. In the resulting **New Column** window, change the **Column Name** to **Validation**.
3. Next to **Initialize Data**, click on the arrow and select **Random** (shown below, left).
4. Select **Random Indicator** (shown below, right). By default, the new column will contain 80% 0s, 20% 1s and 0% 2s.

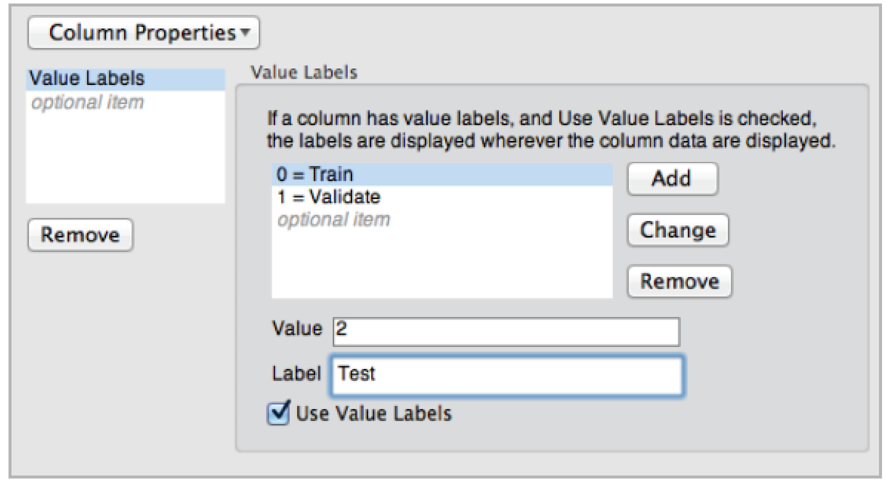
   - The 0s will be used to train the model.
   - The 1s will be used to validate the model.
   - The 2s (if created) will be used to test the model.

5. Change the default proportions as desired. In the example (below, right), 60% of the data will be used to train (develop) the model, 20% will be used to validate the model, and the remaining 20% will be used to test the final selected model.





6. To display the labels Train, Validate and Test rather than 0, 1 and 2, apply **value labels** (as shown, right) by selecting **Value Labels** under **Column Properties.** Enter the values and the desired labels, then click **Add**.
7. Click **Apply** to view the new column in the data table (to verify that the column will be created as desired). Then click **OK** to create the column.



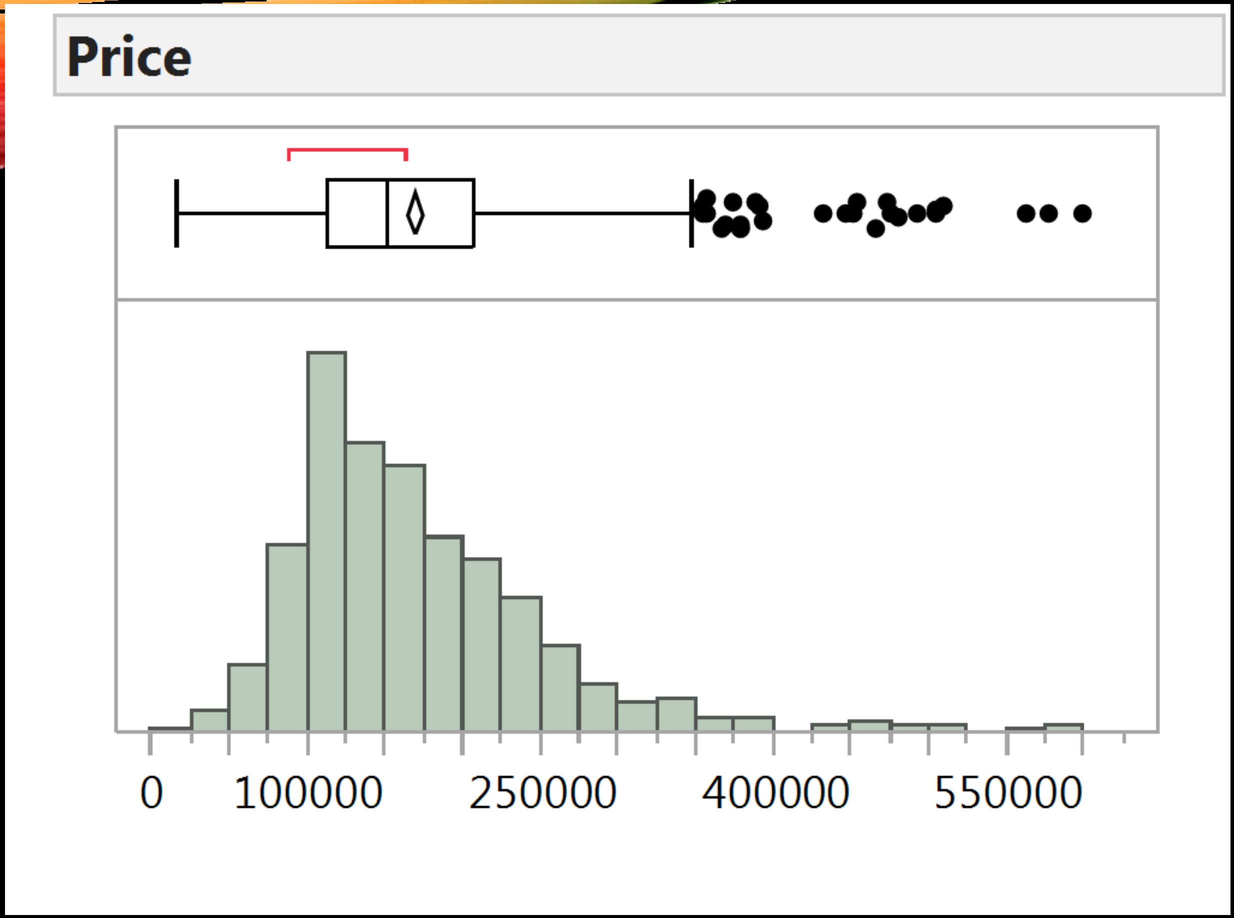
Notes: For information on validation options available in the different modeling platforms and model validation in JMP Pro, see the books ***Fitting Linear Models*** or ***Specialized Models*** (under **Help > Books**) or search for "validation" in the online documentation (jmp.com/support/help/).

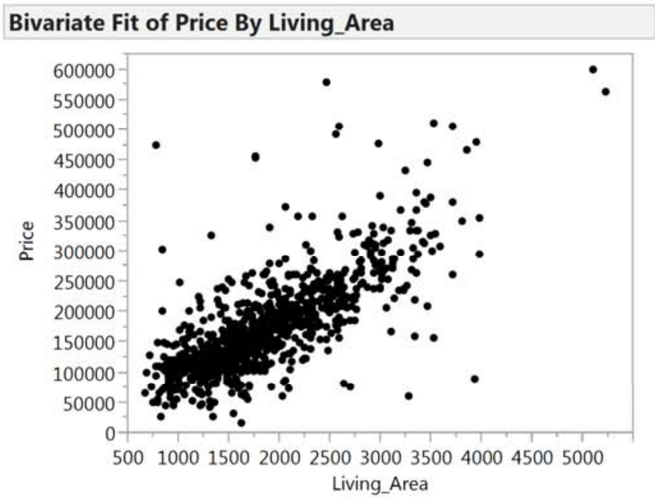
HOUSING PRICES - ZILLOW

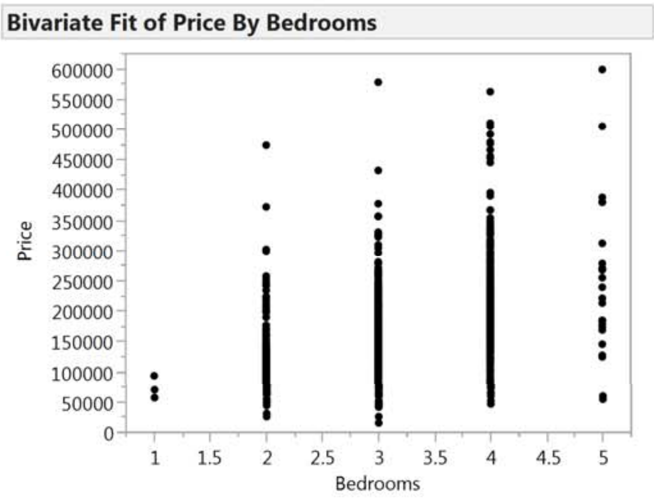
Price: price of house as sold in 2002
Living Area: size of living area in square feet
Bedrooms: number of bedrooms
Bathrooms: number of bathrooms (a half bath is a toilet and sink only
Age: the age of the house in years
Fireplaces: the number of fireplaces in the house

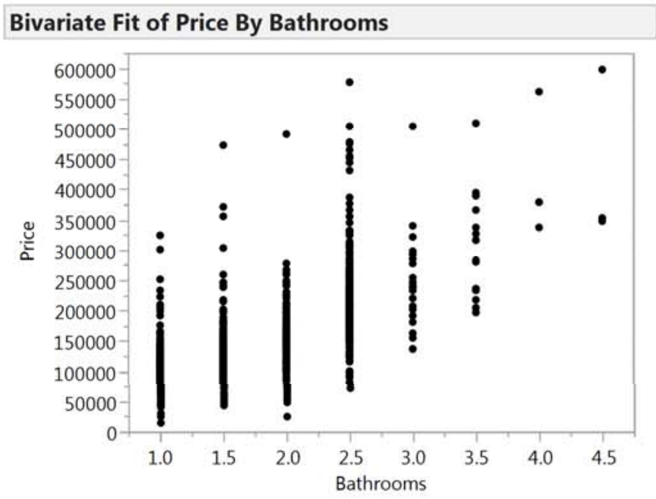
- Price
- Living_Area
- Bedrooms
- Bathrooms
- Fireplaces
- Age
- Validation *

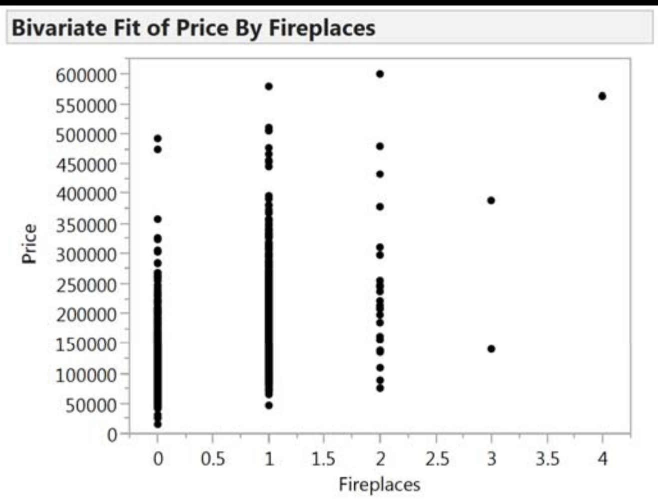
Price	Living_Area	Bedrooms	Bathrooms	Fireplaces	Age	Validation
142212	1982	3	1.0	0	133	Training
134865	1676	3	1.5	1	14	Training
118007	1694	3	2.0	1	15	Training
138297	1800	2	1.0	2	49	Training

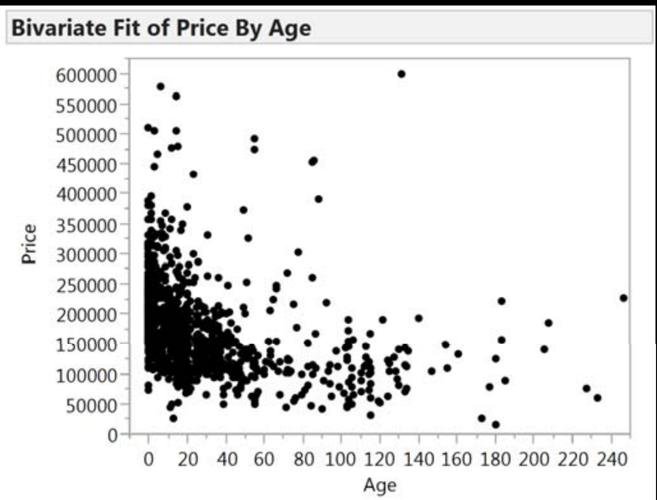

Price
0
100000
250000
400000
550000

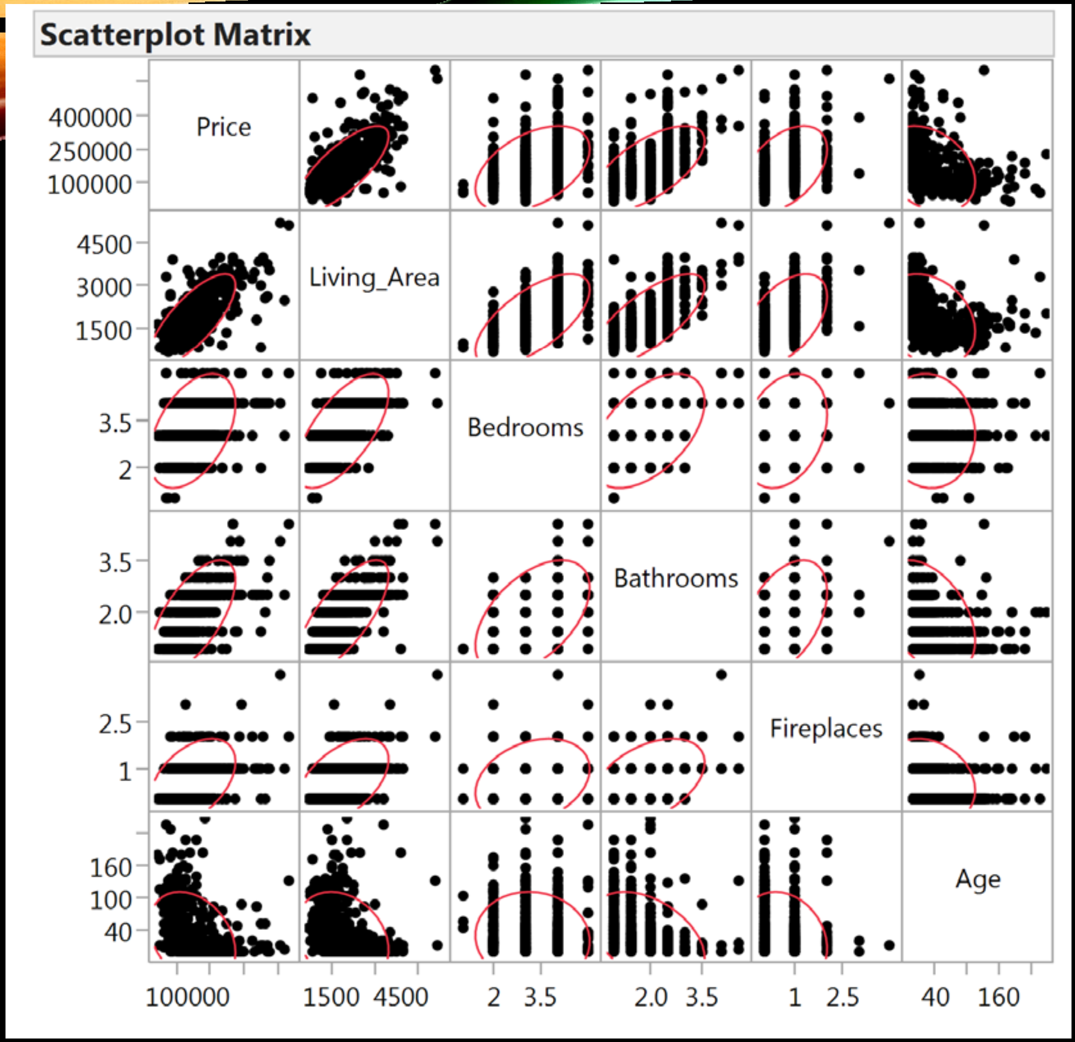

Bivariate Fit of Price By Living_Area
Price
Living_Area
600000
550000
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
500 1000 1500 2000 2500 3000 3500 4000 4500 5000

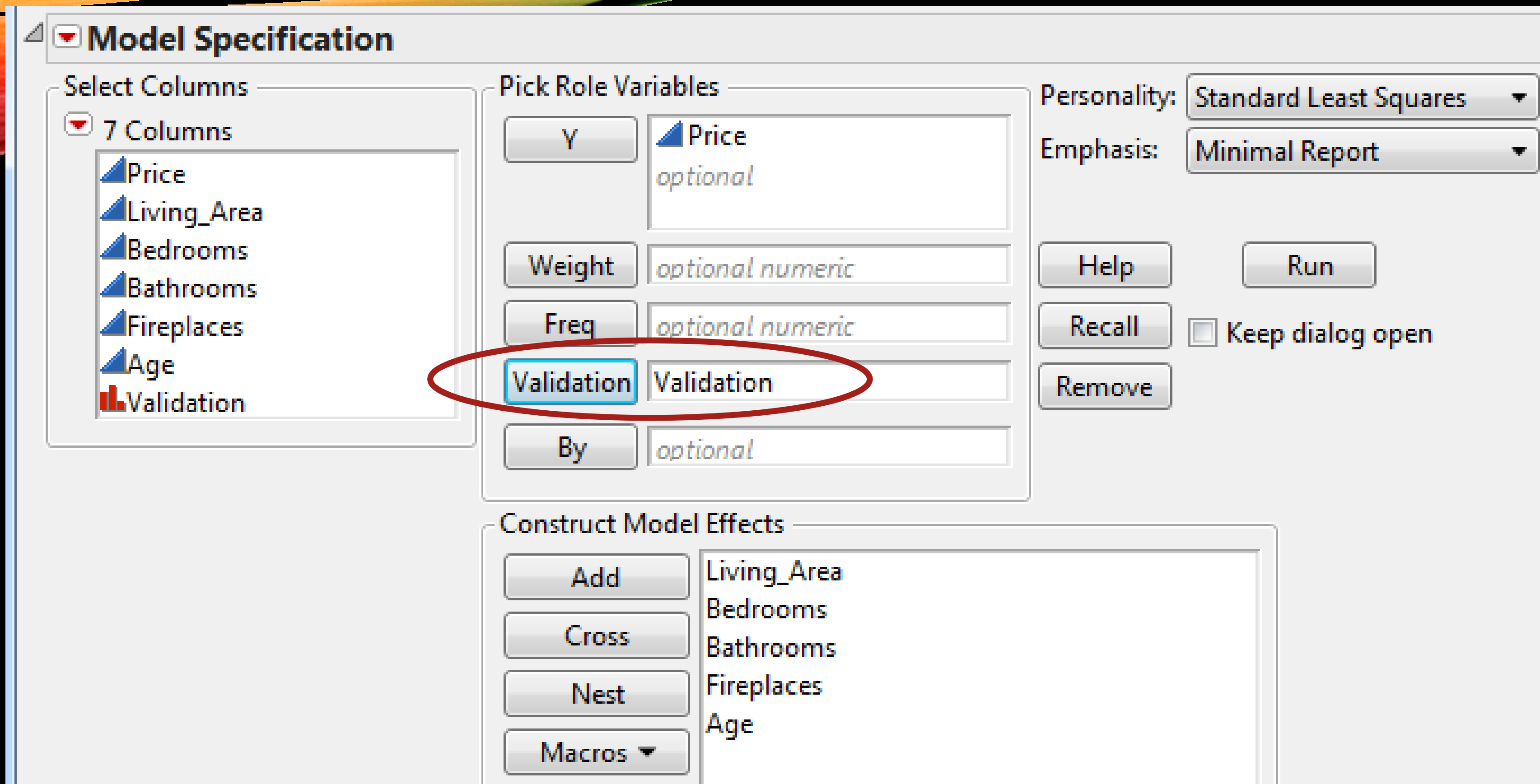

Bivariate Fit of Price By Bedrooms
Price
Bedrooms
1 1.5 2 2.5 3 3.5 4 4.5 5


Bivariate Fit of Price By Bathrooms
Price
Bathrooms
1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5


Bivariate Fit of Price By Fireplaces
Price
Fireplaces
0 0.5 1 1.5 2 2.5 3 3.5 4


Bivariate Fit of Price By Age
Price
Age
0 20 40 60 80 100 120 140 160 180 200 220 240


Scatterplot Matrix
Price
Living_Area
Bedrooms
Bathrooms
Fireplaces
Age
400000
250000
100000
4500
3000
1500
3.5
2
3.5
2.0
2.5
1
160
100
40
100000
1500 4500
2 3.5
2.0 3.5
1 2.5
40 160


Model Specification
Select Columns
7 Columns
Price
Living_Area
Bedrooms
Bathrooms
Fireplaces
Age
Validation
Pick Role Variables
Y
Price
optional
Weight
optional numeric
Freq
optional numeric
Validation
Validation
By
optional
Personality:
Standard Least Squares
Emphasis:
Minimal Report
Help
Run
Recall
Keep dialog open
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Living_Area
Bedrooms
Bathrooms
Fireplaces
Age

Response Price

Validation: Validation

Summary of Fit

RSquare	0.612365
RSquare Adj	0.609725
Root Mean Square Error	50096.33
Mean of Response	168540.7
Observations (or Sum Wgts)	740

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	2.91e+12	5.82e+11	231.9071
Error	734	1.8421e+12	2.5096e+9	**Prob > F**
C. Total	739	4.7521e+12		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	16723.131	9022.729	1.85	0.0642	.
Living_Area	72.440621	4.877772	14.85	<.0001*	3.2065334
Bedrooms	-6243.451	3413.333	-1.83	0.0678	1.8809382
Bathrooms	21704.214	4495.596	4.83	<.0001*	2.5524471
Fireplaces	7143.114	3988.53	1.79	0.0737	1.4757219
Age	-238.3752	62.59205	-3.81	0.0002*	1.3304611

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Living_Area	1	1	5.5352e+11	220.5573	<.0001*
Bedrooms	1	1	8396606115	3.3457	0.0678
Bathrooms	1	1	5.8496e+10	23.3085	<.0001*
Fireplaces	1	1	8049356606	3.2074	0.0737
Age	1	1	3.64e+10	14.5039	0.0002*

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.6124	49893	740
Validation Set	0.5711	45555	317

USING STEPWISE

Stepwise Regression Control

Stopping Rule: Max Validation RSquare

Direction: Forward

Enter All | Make Model | Remove All | Run Model

Go | Stop | Step

rows not used due to excluded rows or missing values.

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC	RSquare Validation	RMSE Validation
4.752e+12	739	80190.033	0.0000	0.0000	1155.5356	1	18815.43	18824.63	-0.001	69589.53

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	168540.716	1	0	0.000	1
☐	☐	Living_Area	0	1	2.72e+12	986.553	4e-138
☐	☐	Bedrooms	0	1	1e+12	197.320	6.8e-40
☐	☐	Bathrooms	0	1	2.07e+12	571.463	5.7e-94
☐	☐	Fireplaces	0	1	1.08e+12	216.575	3.6e-43
☐	☐	Age	0	1	5.94e+11	105.467	3.2e-23

Stepwise Fit for Price

Stepwise Regression Control

Stopping Rule: Max Validation RSquare

Direction: Forward

Enter All | Make Model | Remove All | Run Model

Go | Stop | Step

rows not used due to excluded rows or missing values.

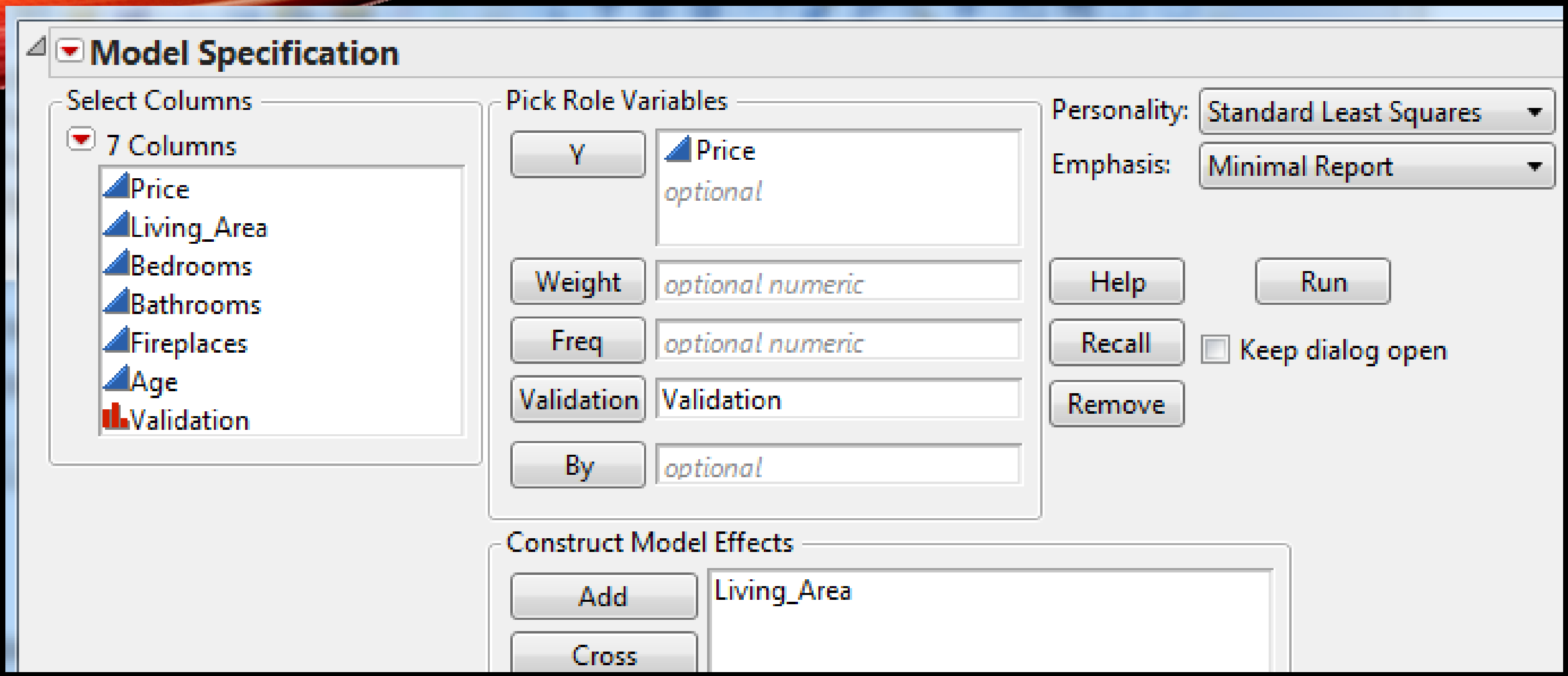
SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC	RSquare Validation	RMSE Validation
2.034e+12	738	52493.356	0.5721	0.5715	74.313899	2	18189.35	18203.14	0.5851	44803.95

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	5698.27941	1	0	0.000	1
☐	☑	Living_Area	89.6524657	1	2.72e+12	986.553	4e-138
☐	☐	Bedrooms	0	1	1.51e+10	5.513	0.01914
☐	☐	Bathrooms	0	1	1.24e+11	47.815	1e-11
☐	☐	Fireplaces	0	1	2.38e+10	8.731	0.00323
☐	☐	Age	0	1	1.17e+11	44.975	4e-11

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC	RSquare Validation
1	Living_Area	Entered	0.0000	2.72e+12	0.5721	74.314	2	18189.4	18203.1	0.5851
2	Bathrooms	Entered	0.0000	1.24e+11	0.5981	26.945	3	18144.9	18163.2	0.5785
3	Age	Entered	0.0000	4.98e+10	0.6086	9.1128	4	18127.3	18150.3	0.5596
4	Bedrooms	Entered	0.0488	9.801e+9	0.6107	7.2074	5	18125.5	18153	0.5638
5	Fireplaces	Entered	0.0737	8.049e+9	0.6124	6	6	18124.3	18156.4	0.5711
6	Best	Specific	.	.	0.5721	74.314	2	18189.4	18203.1	0.5851


Model Specification
Select Columns
7 Columns
Price
Living_Area
Bedrooms
Bathrooms
Fireplaces
Age
Validation
Pick Role Variables
Y
Price
optional
Weight
optional numeric
Freq
optional numeric
Validation
Validation
By
optional
Personality:
Standard Least Squares
Emphasis:
Minimal Report
Help
Run
Recall
Keep dialog open
Remove
Construct Model Effects
Add
Living_Area
Cross

Summary of Fit

RSquare	0.572063
RSquare Adj	0.571483
Root Mean Square Error	52493.36
Mean of Response	168540.7
Observations (or Sum Wgts)	740

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.7185e+12	2.718e+12	986.5530
Error	738	2.0336e+12	2.7556e+9	Prob > F
C. Total	739	4.7521e+12		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	5698.2794	5531.982	1.03	0.3033	.
Living_Area	89.652466	2.854316	31.41	<.0001*	1

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.5721	52422	740
Validation Set	0.5851	44804	317

INTERPRET RESULTS (APPLICATION)

Example of Cross Validation Results

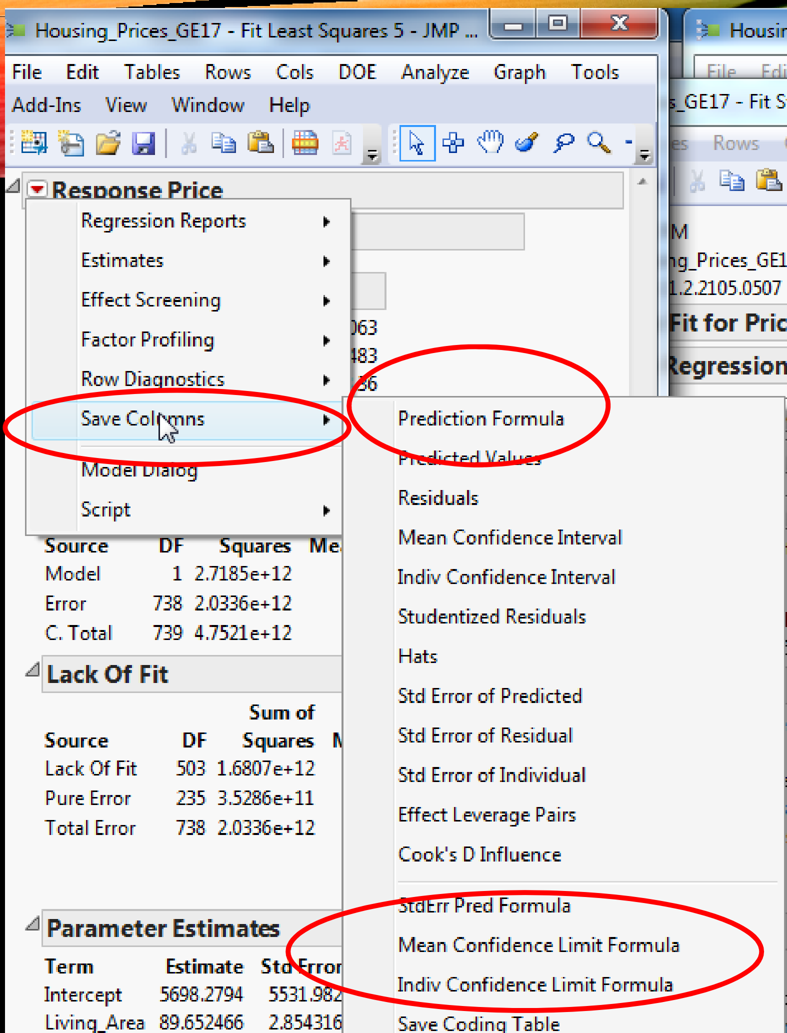
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	5698.2794	5531.982	1.03	0.3033	.
Living_Area	89.652466	2.854316	31.41	<.0001*	1

Crossvalidation

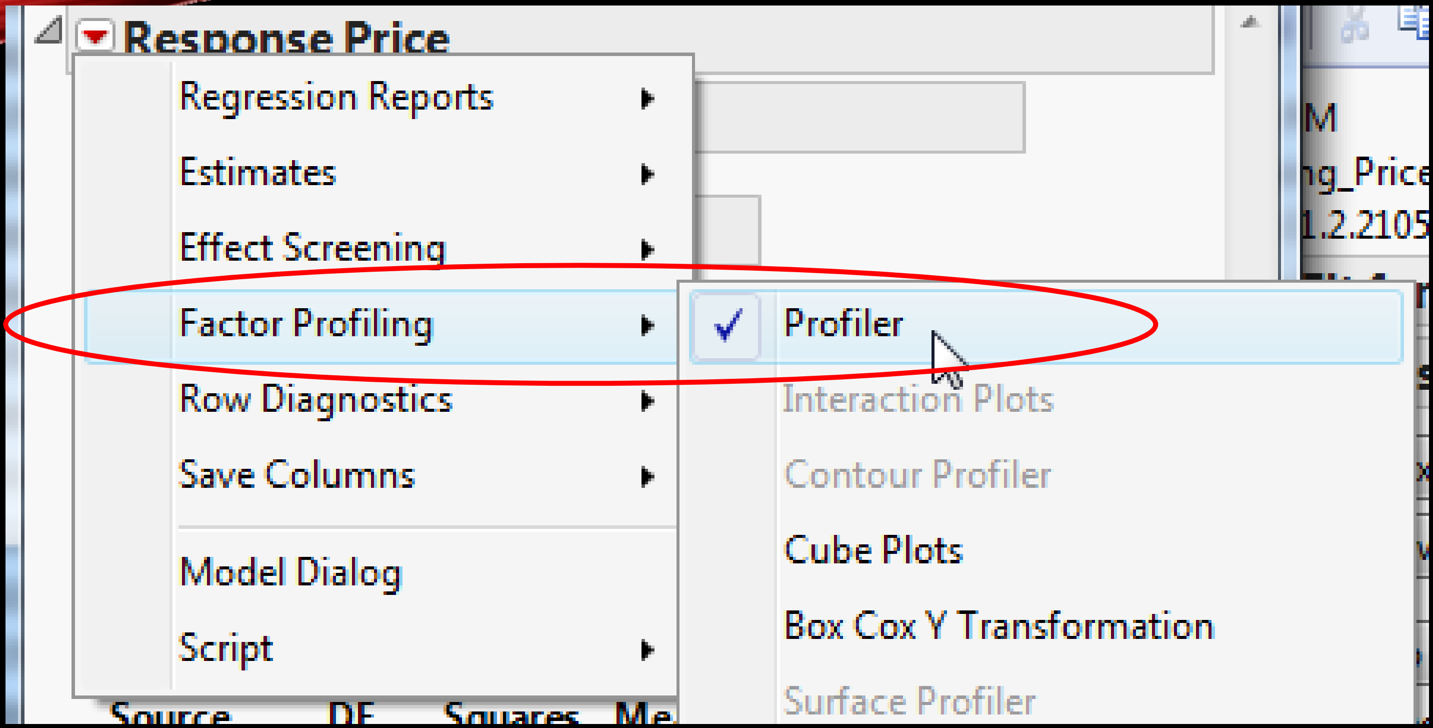
Source	RSquare	RASE	Freq
Training Set	0.5721	52422	740
Validation Set	0.5851	44804	317

HOW TO APPLY THE RESULTS



To use save Prediction Formula and Confidence Intervals (Mean and Individual)

HOW TO UNDERSTAND THE MANAGERIAL IMPLICATIONS



Response Price
Regression Reports
Estimates
Effect Screening
Factor Profiling
Row Diagnostics
Save Columns
Model Dialog
Script
Profiler
Interaction Plots
Contour Profiler
Cube Plots
Box Cox Y Transformation
Surface Profiler