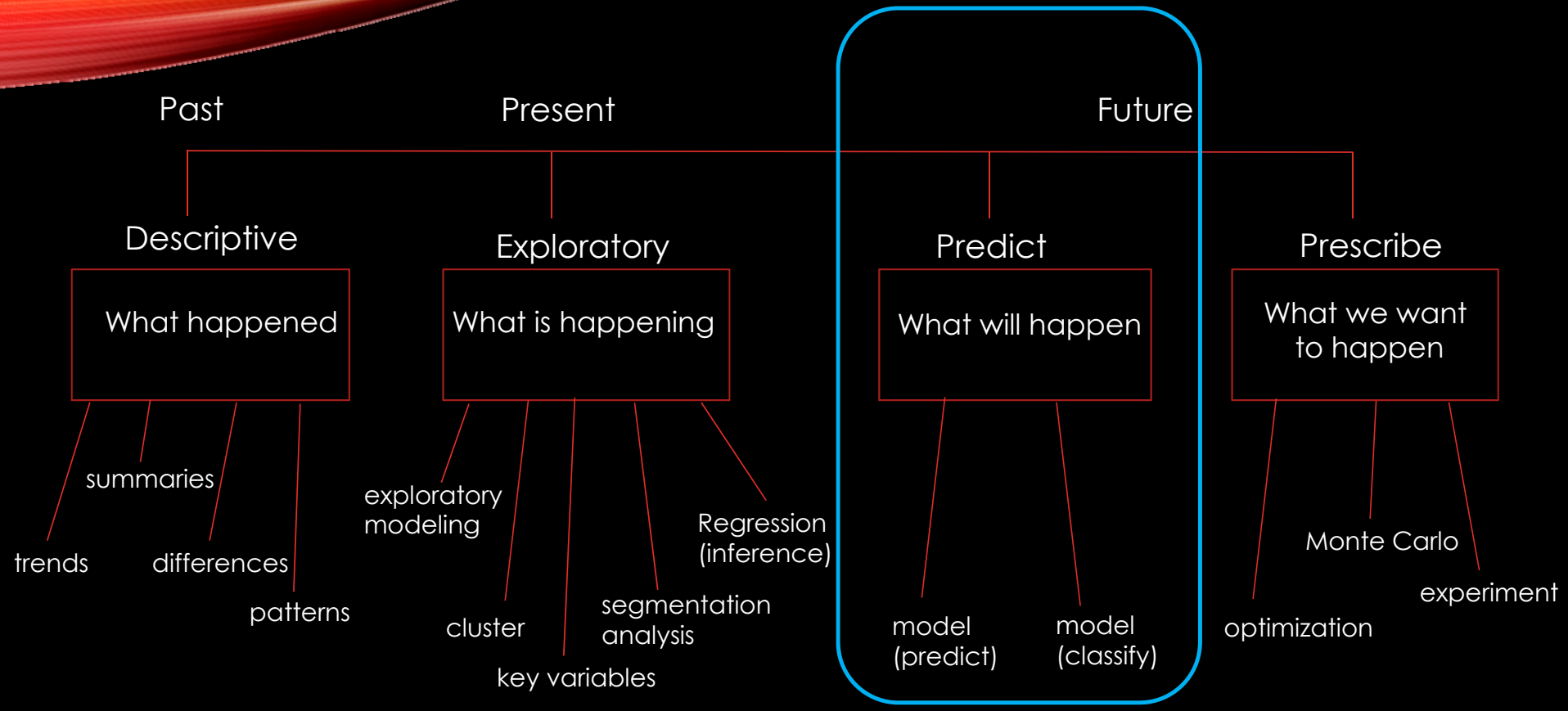
LOGISTIC REGRESSION

- Classification
- Predictive



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

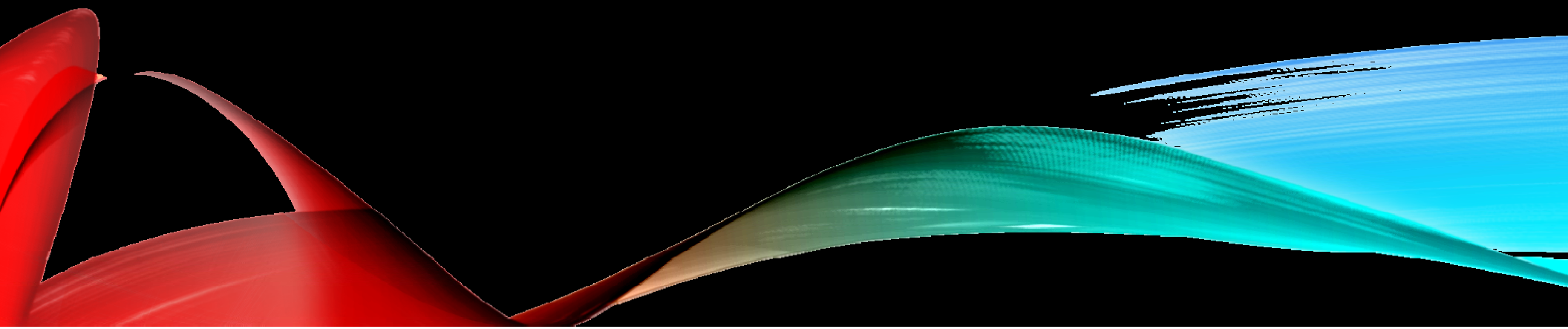
Source: Jim Grayson, Ph.D. and Mia Stephens

Logistic Regression Learning Goals

1. Be able to explain the usefulness of a logistic regression model
2. Be able to explain the fundamental concepts of logistic regression
3. Given LR parameters be able to
    a. construct the LOGIT, ODDS and Probabilities
    b. interpret the parameters of a LR model using the signs of the parameters and the unit and range odds ratios
4. Be able to use JMP to construct and use a logistic regression model for classification
5. Be able to evaluate the predictive performance of a LR model using the misclassification rate, RMSE/MAD and the Confusion Matrix
6. Be able to apply variable selection (stepwise) for a LR model
7. Be able to explain to a manager the insights from a LR model using the profiler

- 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?

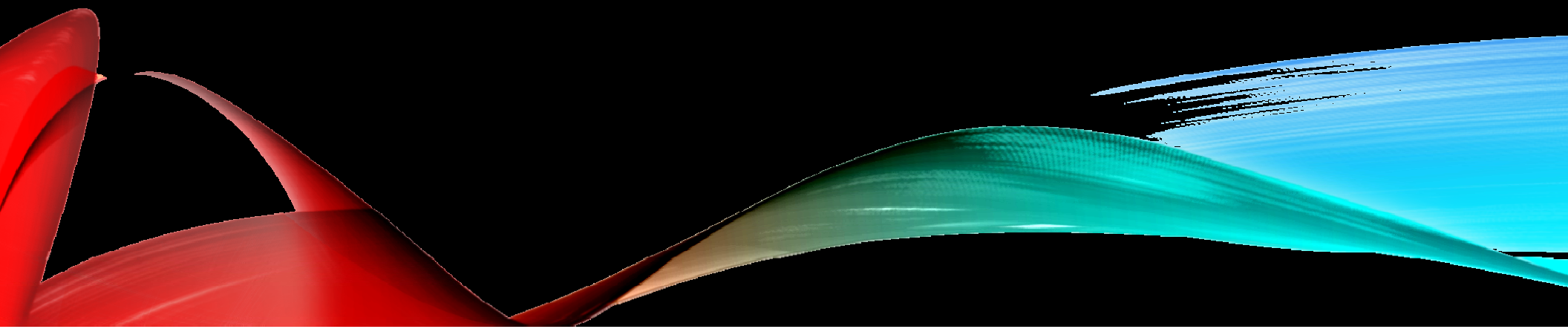


Logistic Regression is an approach for predicting a categorical response variable (Y) with continuous or categorical predictor variables (X).

Y	X	Objective	Predictive Accuracy	Statistical Significance Measure	Model Fit
Categorical	Continuous or Categorical	Prediction or Classification	Misclassification Rate	Prob > Chi-Square (p-value)	RMSE, MAD, ROC Curve

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

WHAT CAN IT DO? (USE CASES)

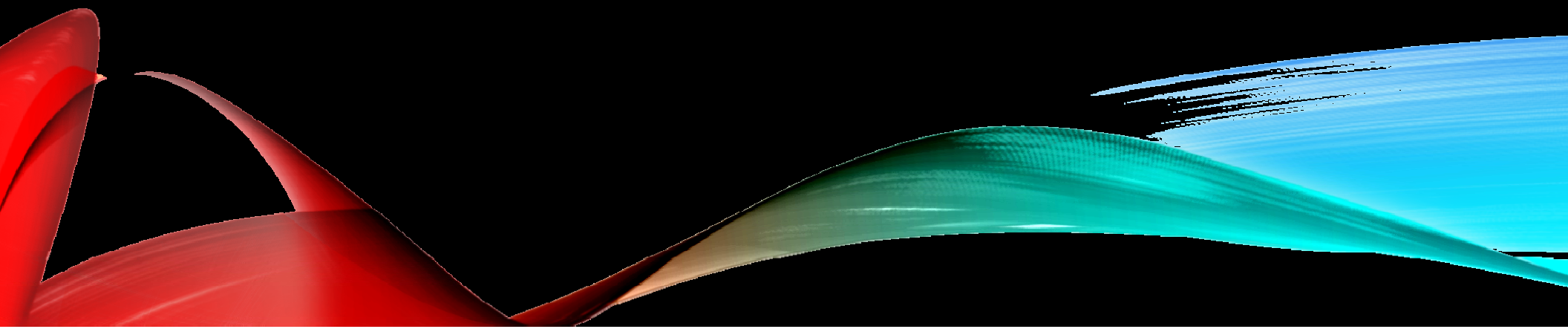


Logistic regression, among other modeling tools, can be used to model the probability that an event will occur. What's the probability someone will win an Oscar, or an election, or a tennis match, based on other events that have transpired? Here are some other example uses of logistic regression:

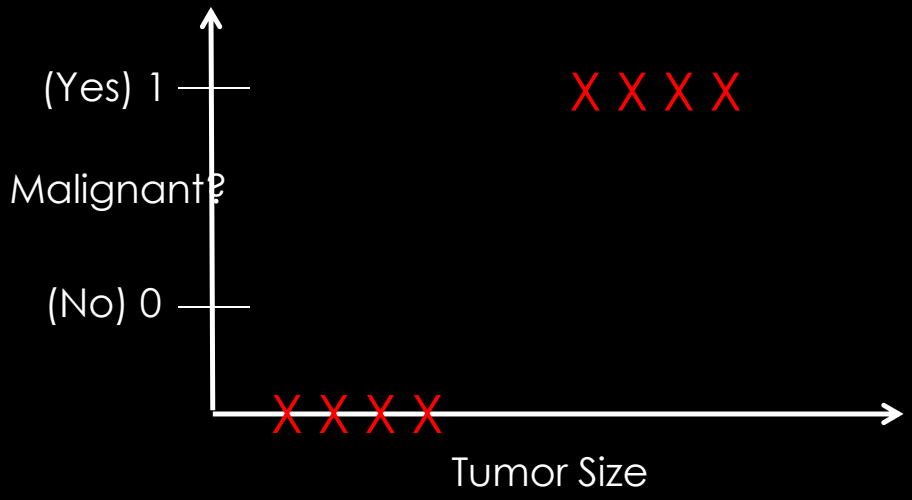
- Identify fraudulent checks submitted to a bank
- Determine causes of defective items in an assembly line
- Predict flight delays
- Understand reasons behind flight risk for employees
- Determine which content to display on a website based on mouse clicks

DRAFT Chapter 5: Logistic Regression from Building Better Models using JMP by Grayson, Gardner and Stephens (SAS Press).

HOW DOES IT WORK?



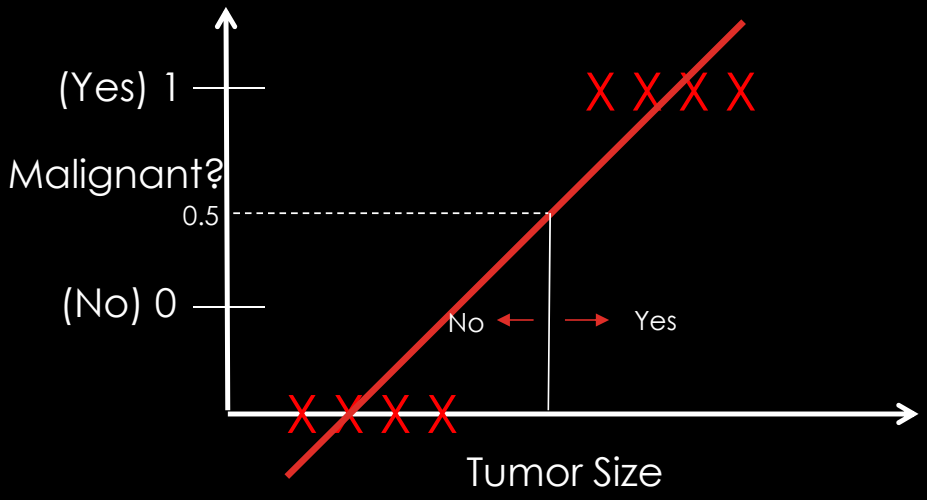
HOW DOES IT WORK?



Based on an illustration given by Andrew Ng, Stanford University

We want to classify a tumor as
0: Benign
1: Malignant

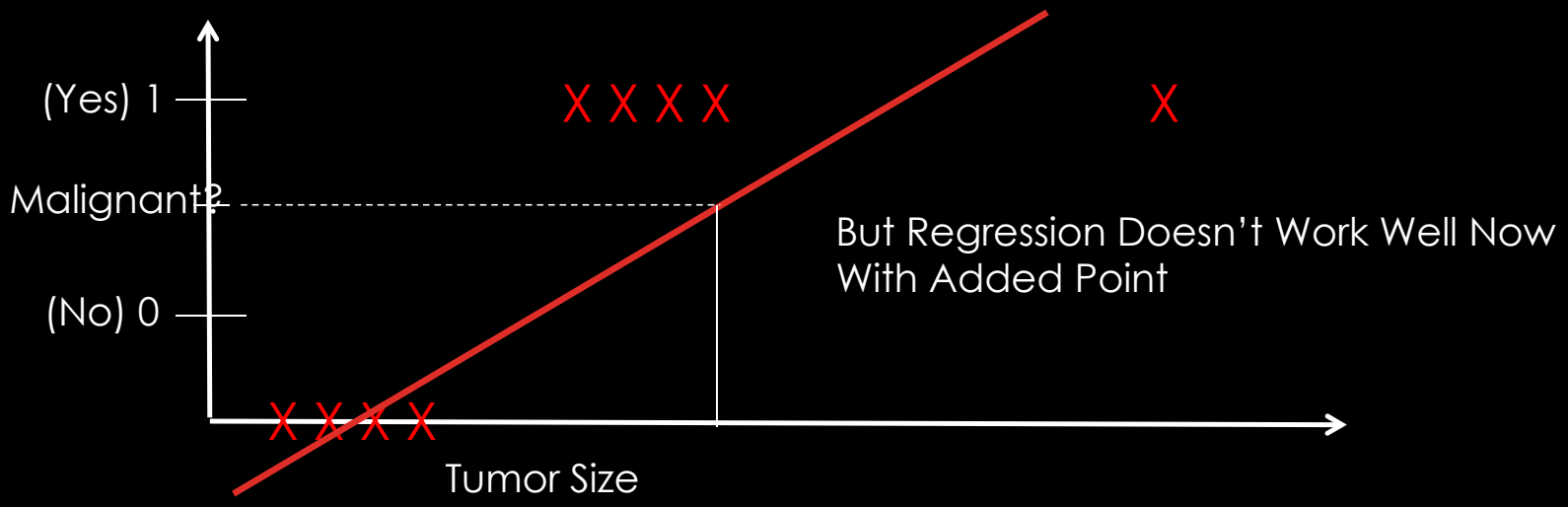
HOW DOES IT WORK?



Based on an illustration given by Andrew Ng, Stanford University

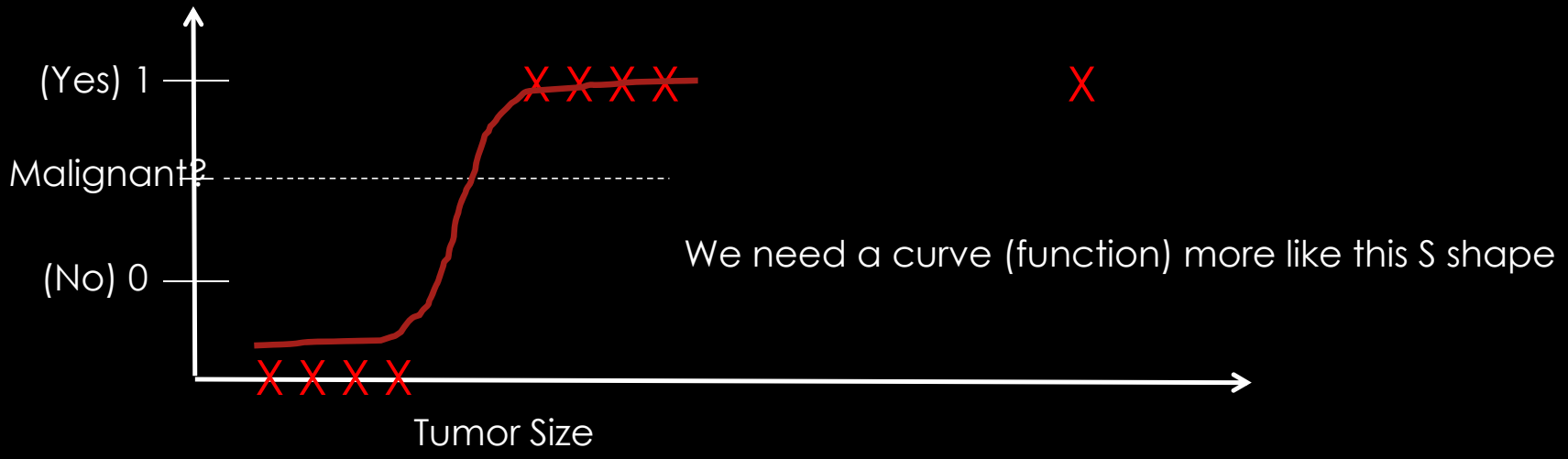
Try Regression

HOW DOES IT WORK?



Based on an illustration given by Andrew Ng, Stanford University

HOW DOES IT WORK?



Based on an illustration given by Andrew Ng, Stanford University

HOW DOES IT WORK?

The coefficients are not fit using least squares, but a method called ***maximum likelihood***.

The basic idea is to try to fit coefficients so that the for the training data the response (classification) prediction is as close as possible to the observed classification.

Fitting Coefficients Using Maximum Likelihood

"Although we could use (non-linear) least squares to fit the model (4.4), the more general method of *maximum likelihood* is preferred, …

…

In other words, we try to find $\hat{\beta}_0$ and $\hat{\beta}_1$ such that plugging these estimates into the model for $p(X)$, given in (4.2), yields a number close to one for all individuals who defaulted, and a number close to zero for all individuals who did not. This intuition can be formalized using a mathematical equation called a *likelihood function*:"

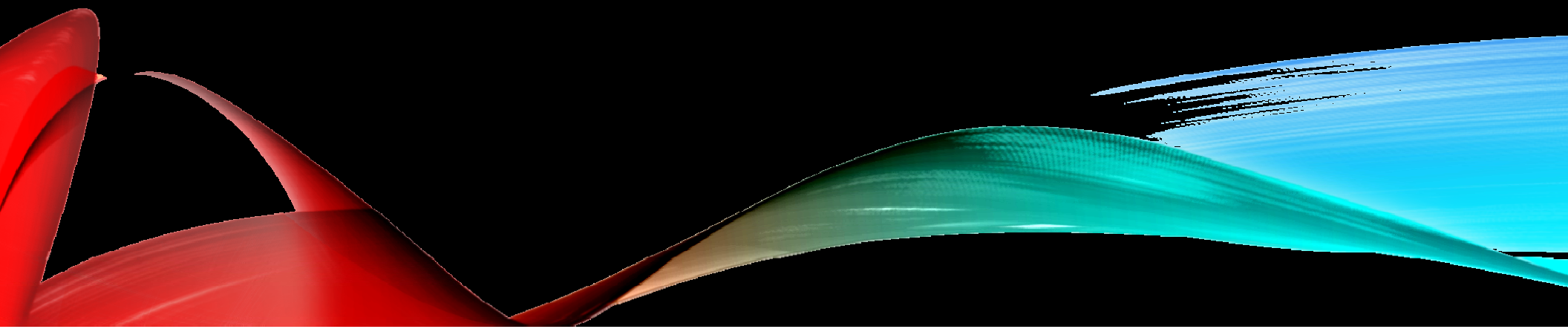
$$\ell(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_{i'}=0} (1 - p(x_{i'})). \qquad (4.5)$$

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}. \qquad (4.2)$$

Source: An Introduction to Statistical Learning, James, et al (Springer), p. 132-3.

JMP MECHANICS

Simple Logistic Regression



Simple Logistic Regression

Logistic regression is used to predict the probability of the occurrence of an event.

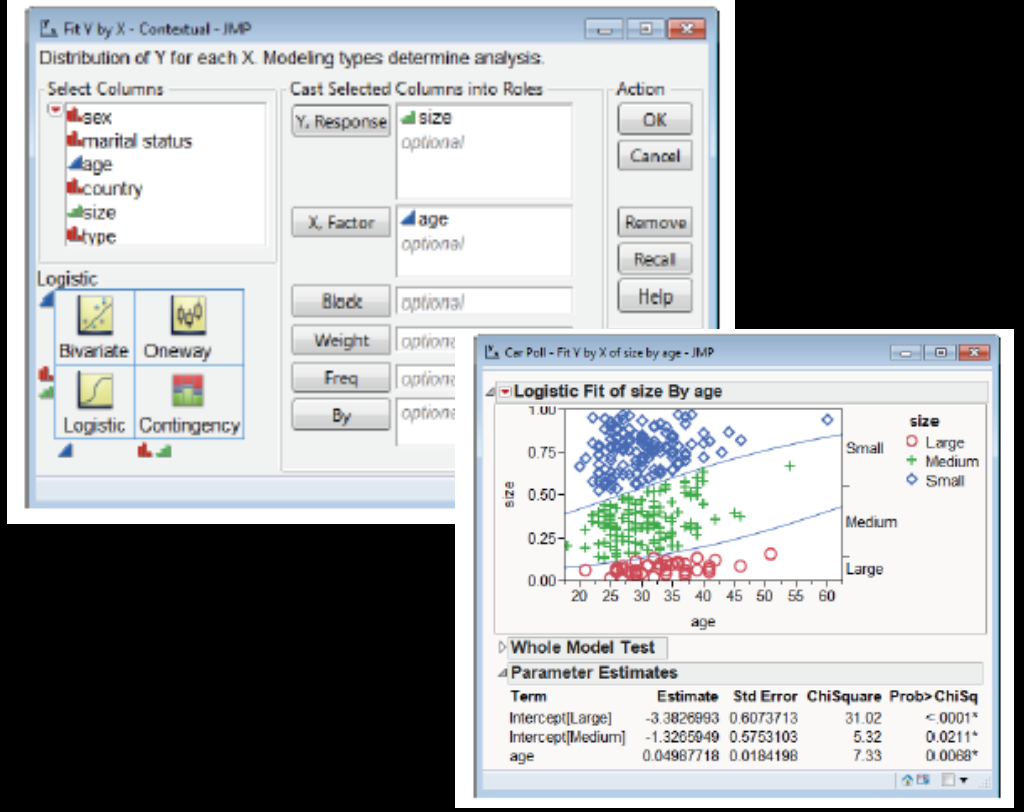
Logistic Regression Using Fit Y by X

1. From an open JMP® data table, select **Analyze > Fit Y by X.**
2. Click on a categorical variable from **Select Columns**, and click **Y, Response** (nominal variables have red bars, ordinal variables have green bars).
3. Click on a continuous variable, and click **X, Factor** (continuous variables have blue triangles).
4. Click **OK** to run the analysis.

By default, JMP will provide the following results:

- The logistic plot, with curves of cumulative predicted (fitted) probabilities.
- The whole model test for model significance.
- Parameter estimates for the fitted model.

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



Simple Logistic Regression

Logistic regression is used to predict the probability of the occurrence of an event.

Tips:

- When the response is nominal, a nominal logistic model will be fit. When the response is ordinal, as in this example, an ordinal logistic model will be fit.
- To color points and add a legend, right-click in the graph and select **Row Legend**. Select a variable under **Mark by Column**, and select **Markers** to change the marker, and click **OK**.

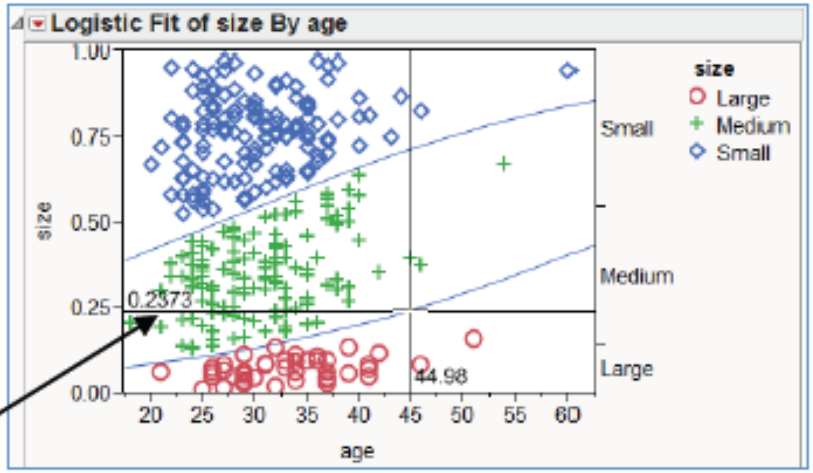
- To save the **probability formula** or request other options, click on the **top red triangle** and select the option.
- To find the fitted probability for a given value of X, select the **cross-hair** tool () from the toolbar or use the keyboard shortcut (C), and click on the graph.

Simple Logistic Regression

Logistic regression is used to predict the probability of the occurrence of an event.

Interpretation (for this example, X = buying age and Y = car size):

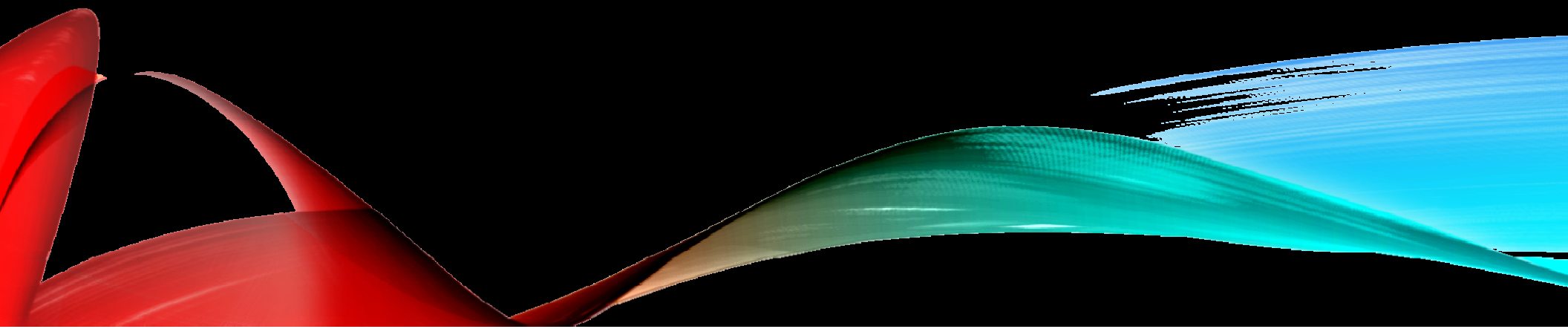
- The **bottom curve** represents the predicted probability that for a given age, someone will buy a **large car.**
- The **second curve** represents the probability that someone will buy a **large or medium car.**
- The **distance between the two curves** represents the probability that someone will buy a **medium car.**
- The **distance between 1.00 and the top curve** represents the probability that someone will buy a **small car.**
- The cross-hairs show that the predicted probability that someone aged 44.98 years will purchase a large car is 0.2373.



Notes: Simple nominal and ordinal logistic regression can also be performed from **Analyze > Fit Model**. For more details see the book ***Basic Analysis*** (under **Help > Books**) or search for "simple logistic regression" in the JMP Help.

SIMPLE LOGISTIC REGRESSION

Lost Sales



EXAMPLE: LOST SALES

In many industries throughout the world, suppliers compete for business by submitting quotes for work, services or products. A key criterion used to determine the winning quote is the dollar amount of the quote, but other factors include expected quality, estimated delivery time of the product, or quoted completion time of the work.

The focus of this case is a supplier of equipment to the automotive industry. The products of interest in this case are various precision metal components used in a range of automotive applications, such as braking systems, drive trains, and engines. Some of the products will be used in the manufacture or assembly of new automobiles (i.e. original equipment), while others will be used as replacement parts in automobiles already on the road (i.e. aftermarket). Lost Sales, JMP Case Study Library | www.jmp.com/en_us/academic/case-study-library/multiple-regression.html

EXAMPLE: LOST SALES

The data set contains 550 records for quotes provided over a six month period. The variables in the data set are:

• Quote	The quoted price, in dollars, for the order
• Time to Delivery	The quoted number of calendar days within which the order is to be delivered
• Part Type	OE = original equipment; AM = aftermarket
• Status	Whether the quote resulted in a subsequent order within 30 days of receiving the quote: Lost = the order was not placed; Won = the order was placed.

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

Status

Frequencies

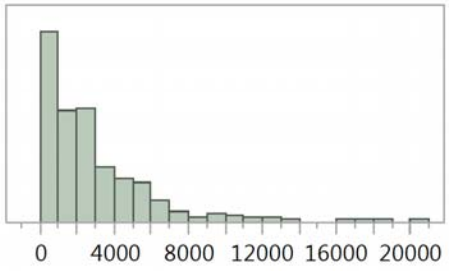
Level	Count	Prob
Won	278	0.50545
Lost	272	0.49455
Total	550	1.00000

N Missing 1

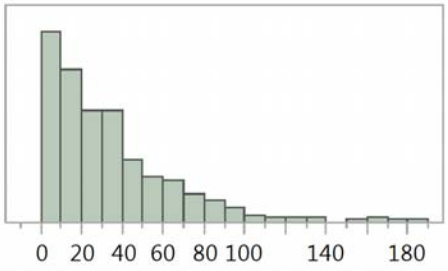
2 Levels

Distributions

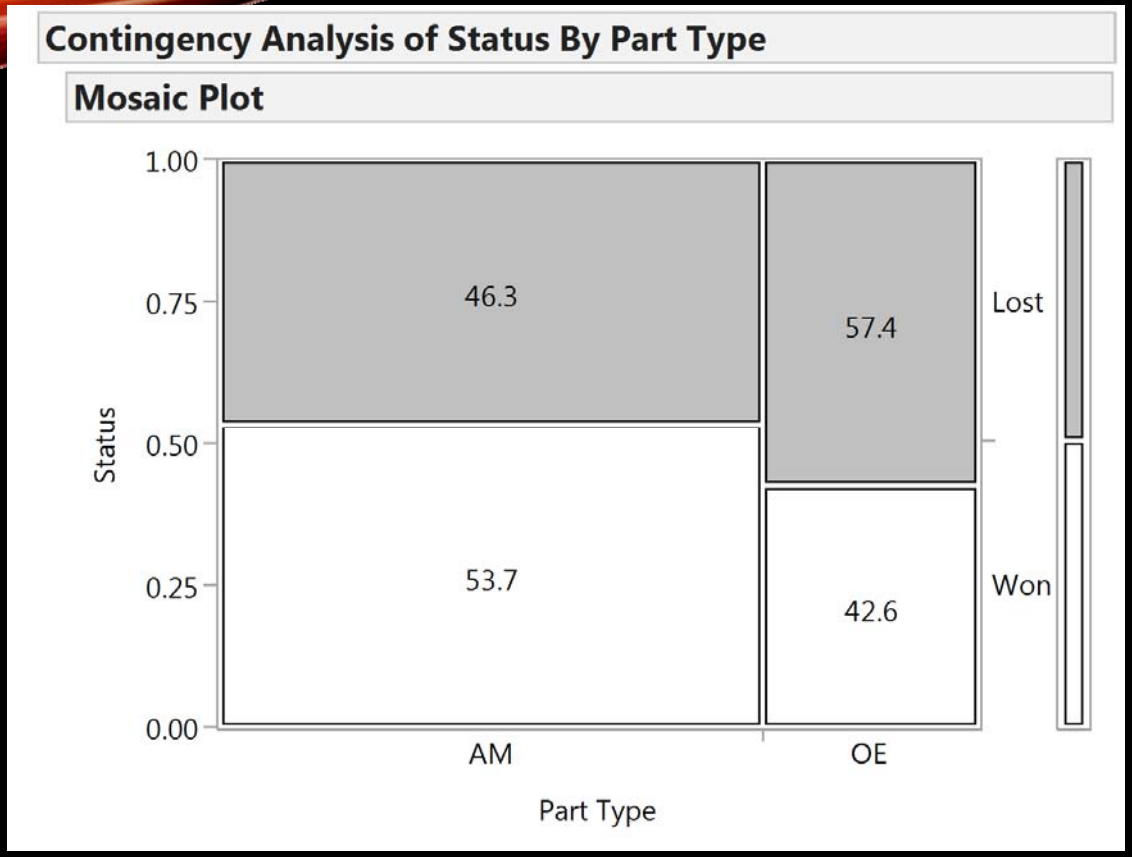
Quote



Time to Delivery

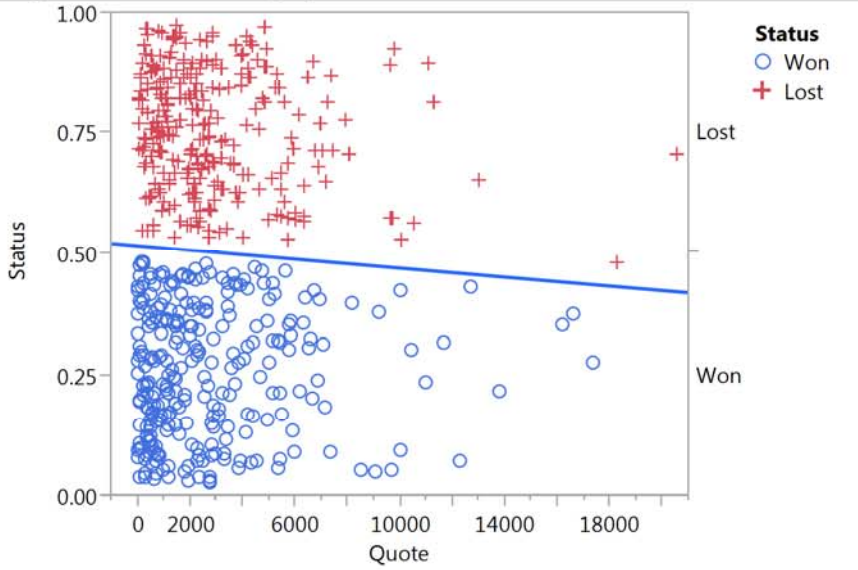


Part Type


Contingency Analysis of Status By Part Type
Mosaic Plot
1.00
0.75
0.50
0.25
0.00
Status
46.3
57.4
53.7
42.6
Lost
Won
AM
OE
Part Type

Fit Group

Logistic Fit of Status By Quote

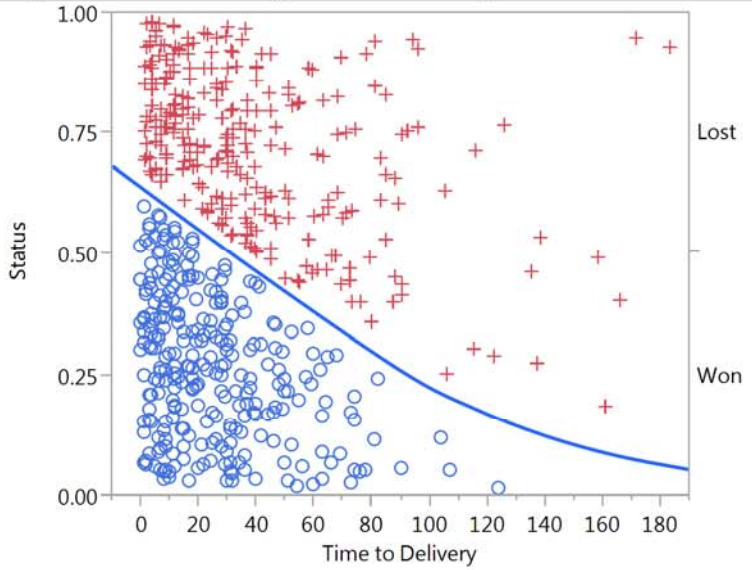


Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.0741106	0.1190632	0.39	0.5336
Quote	-1.8813e-5	0.0000299	0.40	0.5293

For log odds of Won/Lost

Logistic Fit of Status By Time to Delivery

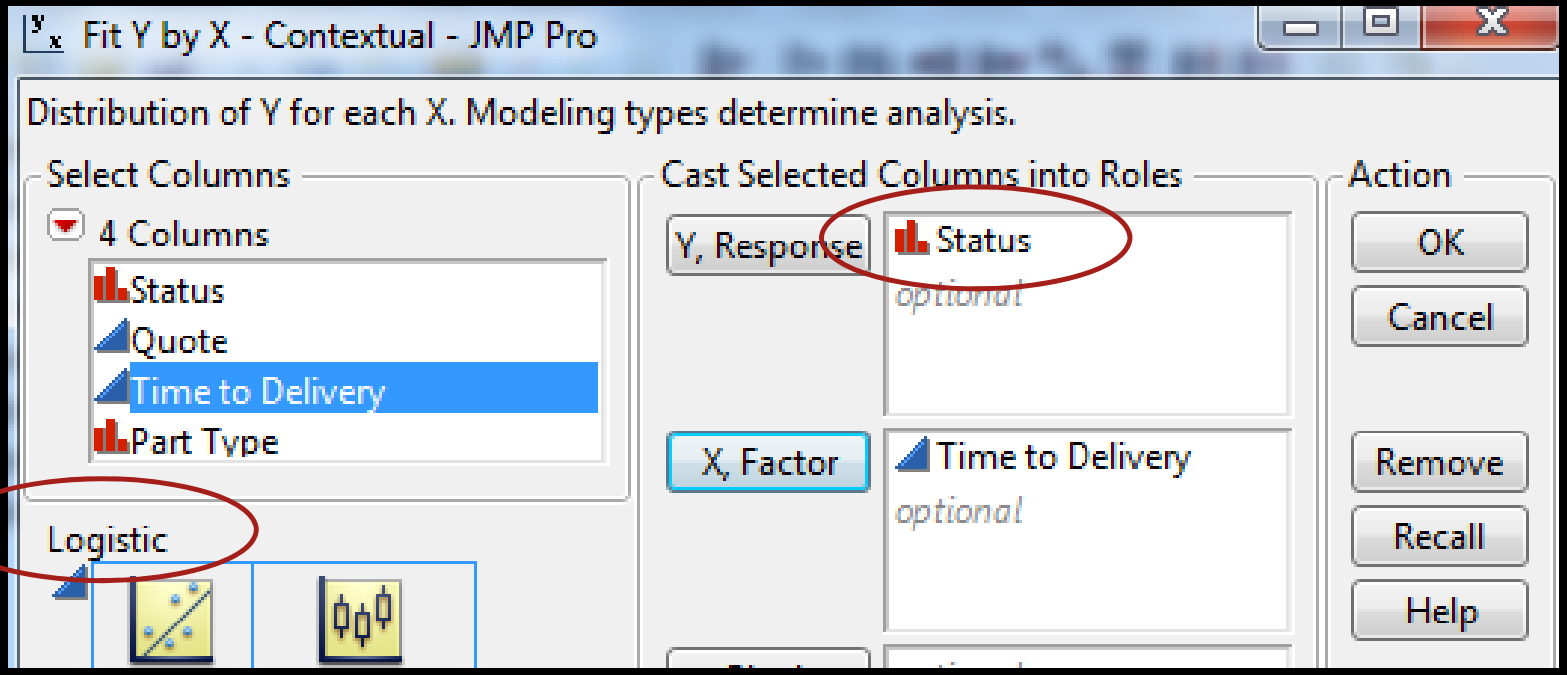


Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.58291532	0.1348966	18.67	<.0001*
Time to Delivery	-0.0181341	0.0034641	27.40	<.0001*

For log odds of Won/Lost

Analyze > Fit Y by X



Lost Sales

Model Significance

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	16.13198	1	32.26396	<.0001*
Full	365.06624			
Reduced	381.19822			

RSquare (U)	0.0423
AICc	734.154
BIC	742.752
Observations (or Sum Wgts)	550

Measure	Training	Definition
Entropy RSquare	0.0423	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0760	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.6638	$\sum$ -Log(ρ[j])/n
RMSE	0.4861	$\sqrt{\sum(y[j]-\rho[j])^2/n}$
Mean Abs Dev	0.4723	$\sum \lvert y[j]-\rho[j]\rvert/n$
Misclassification Rate	0.4091	$\sum$ (ρ[j]≠ρMax)/n
N	550	n

Null hypothesis:
Probability of Won/Lost does not depend on Time to Delivery

Alternate:
There is an association between Won/Lost and Time to Delivery

Lost Sales

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	16.13198	1	32.26396	<.0001*
Full	365.06624			
Reduced	381.19822			

RSquare (U)	0.0423
AICc	734.154
BIC	742.752
Observations (or Sum Wgts)	550

Model Fit

Measure	Training	Definition
Entropy RSquare	0.0423	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0760	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.6638	∑ -Log(ρ[j])/n
RMSE	0.4861	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.4723	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.4091	∑ (ρ[j]≠ρMax)/n
N	550	n

Here are the details. Logistic regression is, of course, estimated by maximizing the likelihood function. Let L_0 be the value of the likelihood function for a model with no predictors, and let L_M be the likelihood for the model being estimated. McFadden's R^2 is defined as

$$R^2_{McF} = 1 - \ln(L_M) / \ln(L_0)$$

where ln(.) is the natural logarithm.

http://www.statisticalhorizons.com/r2logistic

What's the Best R-Squared for Logistic Regression? | February 13, 2013 By Paul Allison

Lost Sales

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	16.13198	1	32.26396	<.0001*
Full	365.06624			
Reduced	381.19822			

RSquare (U)	0.0423
AICc	734.154
BIC	742.752
Observations (or Sum Wgts)	550

Model Fit

Measure	Training	Definition
Entropy RSquare	0.0423	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0760	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.6638	∑ -Log(ρ[j])/n
RMSE	0.4861	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.4723	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.4091	∑ (ρ[j]≠ρMax)/n
N	550	n

The rationale for this formula is that $\ln(L_0)$ plays a role analogous to the residual sum of squares in linear regression. Consequently, this formula corresponds to a proportional reduction in "error variance". It's sometimes referred to as a "pseudo" R^2.

http://www.statisticalhorizons.com/r2logistic

What's the Best R-Squared for Logistic Regression? | February 13, 2013 By Paul Allison

Lost Sales

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	16.13198	1	32.26396	<.0001*
Full	365.06624			
Reduced	381.19822			

RSquare (U)	0.0423
AICc	734.154
BIC	742.752
Observations (or Sum Wgts)	550

Measure	Training	Definition
Entropy RSquare	0.0423	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0760	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.6638	$\sum$ -Log(p[j])/n
RMSE	0.4861	$\sqrt{\sum(y[j]-p[j])^2/n}$
Mean Abs Dev	0.4723	$\sum \lvert y[j]-p[j] \rvert /n$
Misclassification Rate	0.4091	$\sum (p[j] \neq pMax)/n$
N	550	n

Model Accuracy

Prediction

Lost Sales

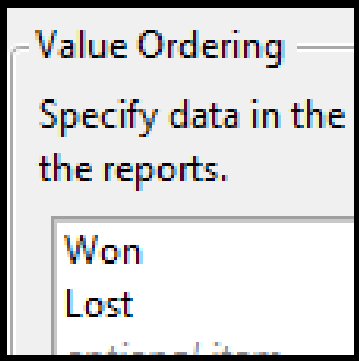
Model Significance

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.58291532	0.1348966	18.67	<.0001*
Time to Delivery	-0.0181341	0.0034641	27.40	<.0001*

For log odds of Won/Lost

Log Odds Success/Failure – this ordering is dependent on the response variable "value ordering" | see Column > Column Properties > Value Ordering



Lost Sales

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.58291532	0.1348966	18.67	<.0001*
Time to Delivery	-0.0181341	0.0034641	27.40	<.0001*

For log odds of Won/Lost

For logistic regression, the probability of an event, *p*, is related to predictive factors, $(X_1, X_2, \ldots, X_k)$ by the mathematical relationship where *log(p/(1-p))* is called the ***log-odds*** **or** ***logit***.

$$\log(p/(1-p)) = \beta_0 + \beta_0 X_1 + \ldots + \beta_k X_k$$

Excerpt from Chapter 5: Logistic Regression from Building Better Models using JMP (Draft) by Grayson, Gardner and Stephens (SAS Press).

Lost Sales

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.58291532	0.1348966	18.67	<.0001*
Time to Delivery	-0.0181341	0.0034641	27.40	<.0001*

For log odds of Won/Lost

The right side of this equation looks a lot like our multiple linear regression model (without the error). Rearranging this formula to solve for *p* directly, we have

$$p = 1 / (1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)})$$

Excerpt from Chapter 5: Logistic Regression from Building Better Models using JMP (Draft) by Grayson, Gardner and Stephens (SAS Press).

Interpreting Logistic Regression Coefficients and Odds

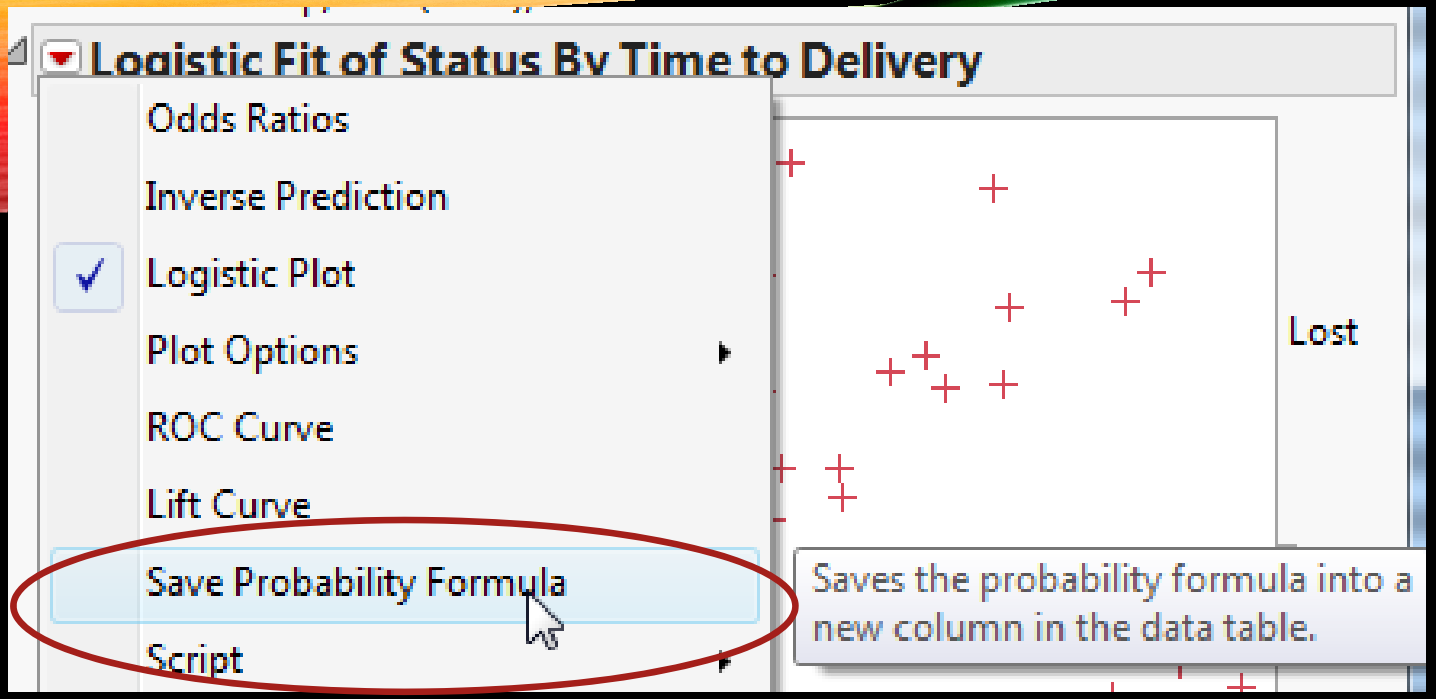
"The quantity $p(X)/[1-p(X)]$ is called the *odds*, and can take on any value between 0 and ∞. Values of the odds close to 0 and ∞ indicate very low and very high probabilities of default, respectively. For example, on average 1 in 5 people with an odds of 1/4 will default, since $p(X) = 0.2$ implies an odds of $0.2 / 1-0.2 = 1/4$. Likewise on average nine out of every ten people with an odds of 9 will default, since $p(X) = 0.9$ implies an odds of $0.9 / 1-0.9 = 9$."

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq	Odds Ratio	Unit Odds Ratio
Intercept	0.58291532	0.1348966	18.67	<.0001*	.	.
Time to Delivery	-0.0181341	0.0034641	27.40	<.0001*	0.98202929	0.03620537

For log odds of Won/Lost

Source: An Introduction to Statistical Learning, James, et al (Springer), p. 132-3.



Lin[Won]	Prob[Won]	Prob[Lost]	Most Likely Status
0.2927690673	0.5726739109	0.4273260891	Won
-0.251255161	0.4375145854	0.5624854146	Lost

Lin[Won]	Prob[Won]	Prob[Lost]	Most Likely Status

$$0.58291532221625 + -0.0181341409344 * \textit{Time to Delivery}$$

$$\frac{1}{1 + \text{Exp}\left[\textit{Lin[Won]}\right]}$$

$$\frac{1}{1 + \text{Exp}\left[-\textit{Lin[Won]}\right]}$$

$$\text{Match}\left[\text{Maximum}\left[\textit{Prob[Won]}, \textit{Prob[Lost]}\right],\ \textit{Prob[Won]} \Rightarrow \text{"Won"},\ \textit{Prob[Lost]} \Rightarrow \text{"Lost"},\ \text{else} \Rightarrow \text{""}\right]$$

$$p = 1 / (1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)})$$

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.58291532	0.1348966	18.67	<.0001*
Time to Delivery	-0.0181341	0.0034641	27.40	<.0001*

For log odds of Won/Lost

Assume time to delivery = 90. From this logistic regression model estimate the following four quantities using EXCEL:

Logit [same as Linear Model]

Odds [Exp(Linear Model)]

Probability of a Won [Odds/(1+Odds)]

Classification of Won/Lost [maximum probability]

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.58291532	0.1348966	18.67	<.0001*
Time to Delivery	-0.0181341	0.0034641	27.40	<.0001*

For log odds of Won/Lost

Assume time to delivery = 90.

Logit — Logit = intercept + time to delivery * value = 0.5829 - 0.018*90

Odds — Exp (logit) = Exp(0.5829 - 0.018*90) = 0.3502

Probability of a Won — Odds/1+Odds = 0.3502 / 1 + 0.3502 = 0.2594

Classification of Won/Lost — Prob Won = 0.2594 then Prob Lost = 1-0.2594
Therefore, status = LOST

	A	B	C	D	E	F
4		Time to Delivery	90			
5						
6			Input variables		Coefficient	
7			Intercept		0.58291532	
8			Time to Delivery		-0.0181341	
9						
10						
11		**logit**	**= -1.0492**		**=E7+E8*C4**	
12						
13		**Odds**	**= 0.3502**		**=EXP(C11)**	
14						
15		**p**	**= 0.2594**		**=C13/(1+C13)**	
16						
17		**p**	**= 0.2594**		**=1/(1+EXP(-(E7+E8*C4)))**	
18						

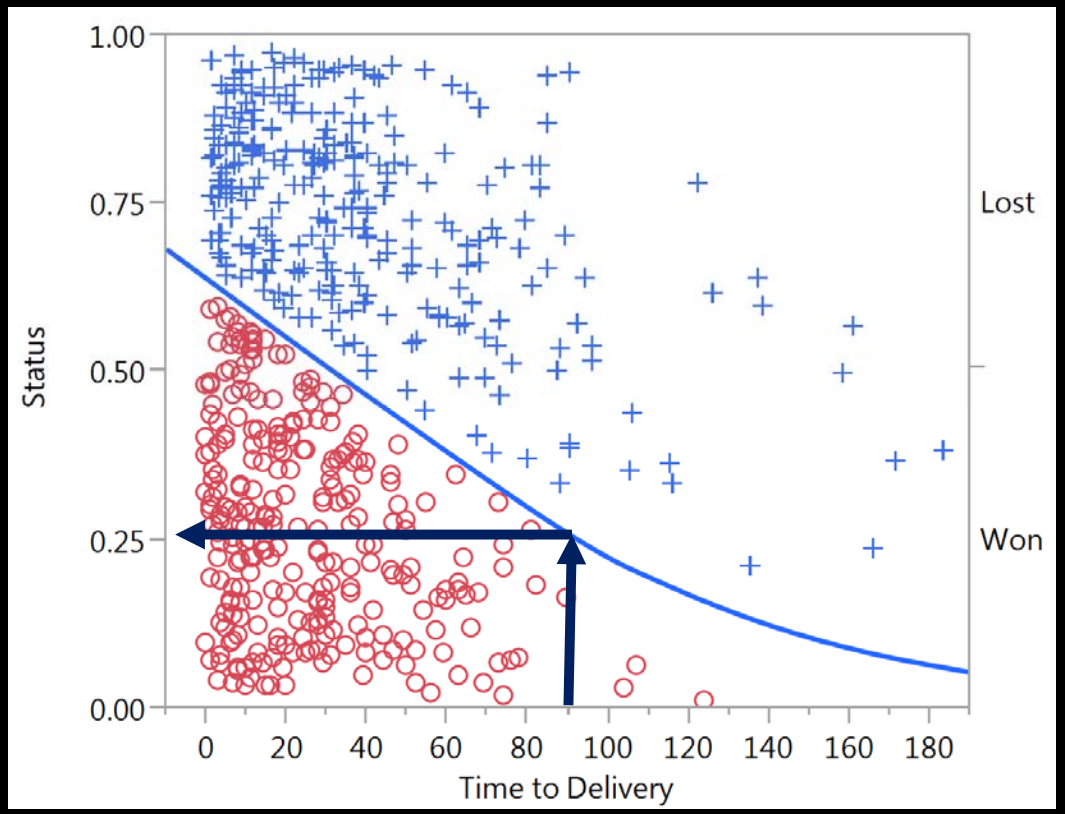
	Status	Quote	Delivery	Part Type	Validation	Lin[Won]	Prob[Won]	Prob[Lost]	Status
551		•	90		•	-1.049157362	0.2593869433	0.7406130567	Lost

Interpreting Logistic Regression Coefficients and Odds

"… in a linear regression model, β_1 gives the average change in *Y* associated with a one-unit increase in *X*. In contrast, in a logistic regression model, increasing *X* by one unit changes the log odds by β_1 (4.4), or equivalently it multiplies the odds by $e^{\beta 1}$(4.3). However, because the relationship between *p*(*X*) and *X* in (4.2) is not a straight line, β_1 does *not* correspond to the change in *p*(*X*) associated with a one-unit increase in *X*. The amount that *p*(*X*) changes due to a one-unit change in *X* will depend on the current value of *X*. But regardless of the value of *X*, if β_1 is positive then increasing *X* will be associated with increasing *p*(*X*), and if β_1 is negative then increasing *X* will be associated with decreasing *p*(*X*). The fact that there is not a straight-line relationship between *p*(*X*) and *X*, and the fact that the rate of change in *p*(*X*) per unit change in *X* depends on the current value of *X*, can also be seen by inspection of the right-hand panel of Figure 4.2."

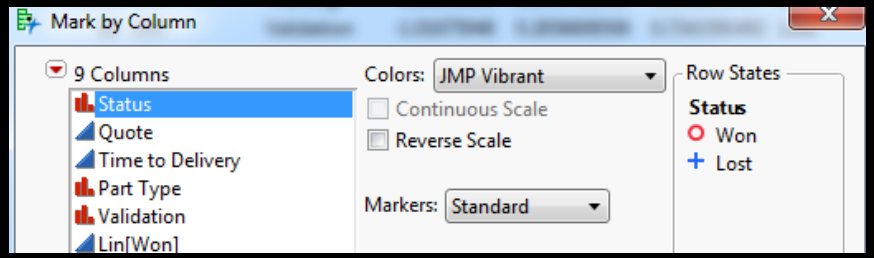
Source: An Introduction to Statistical Learning, James, et al (Springer), p. 132-3.

VISUALIZING LOGISTIC MODEL FIT

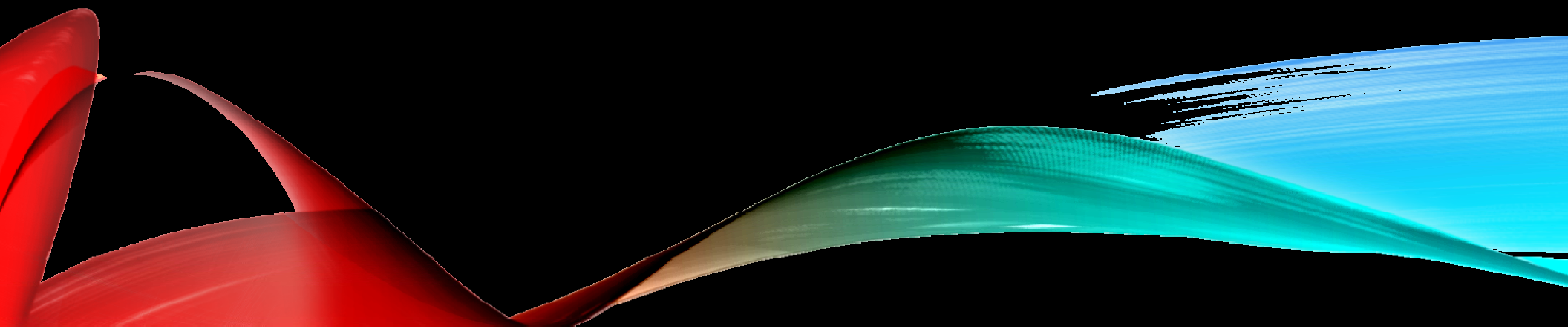


For a Time to Delivery of 90 days,
there is about a 25% chance of "Won" and a 75% chance of "Lost"

- To color points and add a legend, right-click in the graph and select **Row Legend**. Select a variable under **Mark by Column**, and select **Markers** to change the marker, and click **OK**.

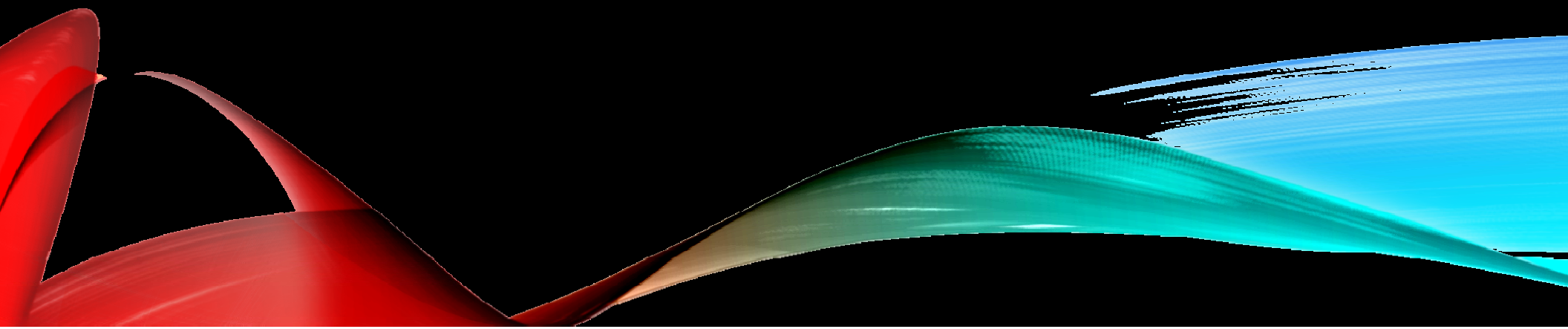


MULTIPLE LOGISTIC REGRESSION



JMP MECHANICS

Multiple Logistic Regression



Multiple Logistic Regression

Multiple logistic regression is used to predict the probability of the occurrence of an event using more than one explanatory variable.

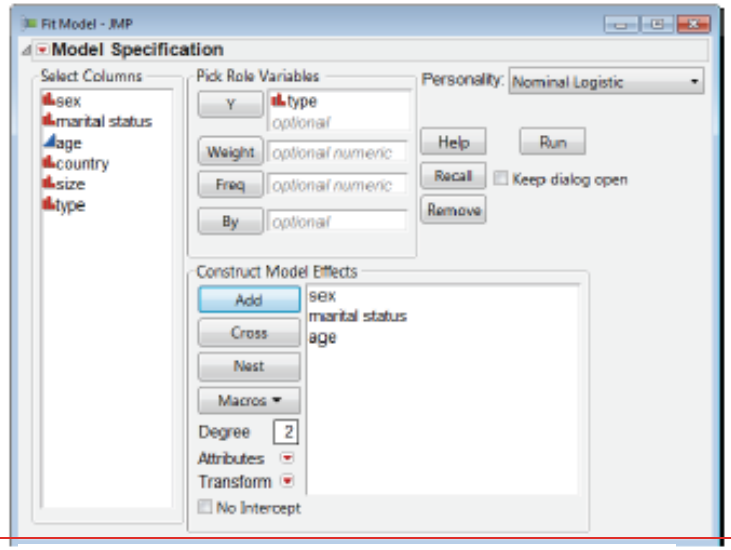
Multiple Logistic Regression Using Fit Model

1. From an open JMP® data table, select **Analyze > Fit Model**.
2. Click on a categorical variable from **Select Columns**, and click **Y** (nominal variables have red bars, ordinal variables have green bars).
3. Choose explanatory variables from **Select Columns**, and click **Add**.
4. Click **Run Model**.

By default, JMP will provide the following results:

- The Iterations history (not shown).
- The Whole Model Test.
- Lack of Fit (not shown).
- Parameter Estimates for the model.
- Effect Likelihood Ratio Tests (not shown).

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



Tips:

- When the response is ordinal, an ordinal logistic model will be fit. When the response is nominal, as in this example, a nominal logistic model will be fit.
- To save the predicted probabilities to the data table, click on the **top red triangle**, select **Save Probability Formula**.

Multiple Logistic Regression

Multiple logistic regression is used to predict the probability of the occurrence of an event using more than one explanatory variable.

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	22.99752	6	45.99504	<.0001*
Full	280.19796			
Reduced	303.19548			

RSquare (U)	0.0759
AICc	576.886
BIC	606.106
Observations (or Sum Wgts)	303

Measure	Training	Definition
Entropy RSquare	0.0759	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.1628	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.9247	$\sum$ -Log(p[j])/n
RMSE	0.5827	$\sqrt{\sum(y[j]-p[j])^2/n}$
Mean Abs Dev	0.5462	$\sum \lvert y[j]-p[j]\rvert/n$
Misclassification Rate	0.4290	$\sum$ (p[j]≠pMax)/n
N	303	n

Multiple Logistic Regression

Multiple logistic regression is used to predict the probability of the occurrence of an event using more than one explanatory variable.

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.21751807	0.8942808	0.06	0.8078
sex[Female]	0.10271474	0.1674203	0.38	0.5395
marital status[Married]	0.23592617	0.1815837	1.69	0.1939
age	0.02719008	0.0281513	0.93	0.3341
Intercept	3.0890961	1.0026577	9.49	0.0021*
sex[Female]	-0.0515722	0.1822018	0.08	0.7771
marital status[Married]	-0.3754868	0.1868537	4.04	0.0445*
age	-0.0779815	0.0329797	5.59	0.0181*

For log odds of Family/Work, Sporty/Work

jmp.com/learn rev 02/2014

Multiple Logistic Regression

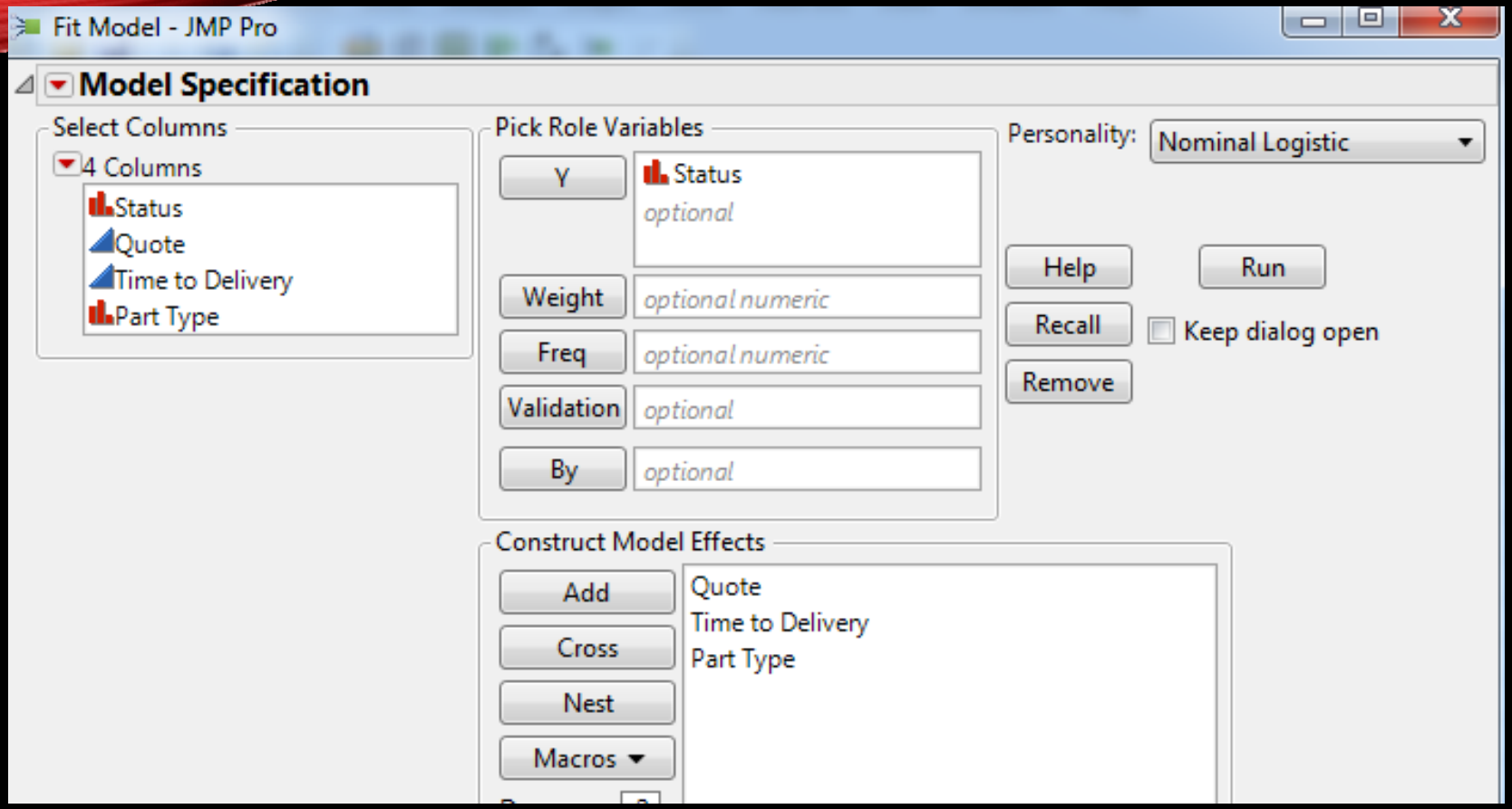
Multiple logistic regression is used to predict the probability of the occurrence of an event using more than one explanatory variable.

To view the effect of an explanatory variable on the predicted probabilities, click on the **top red triangle** and select **Profiler**.

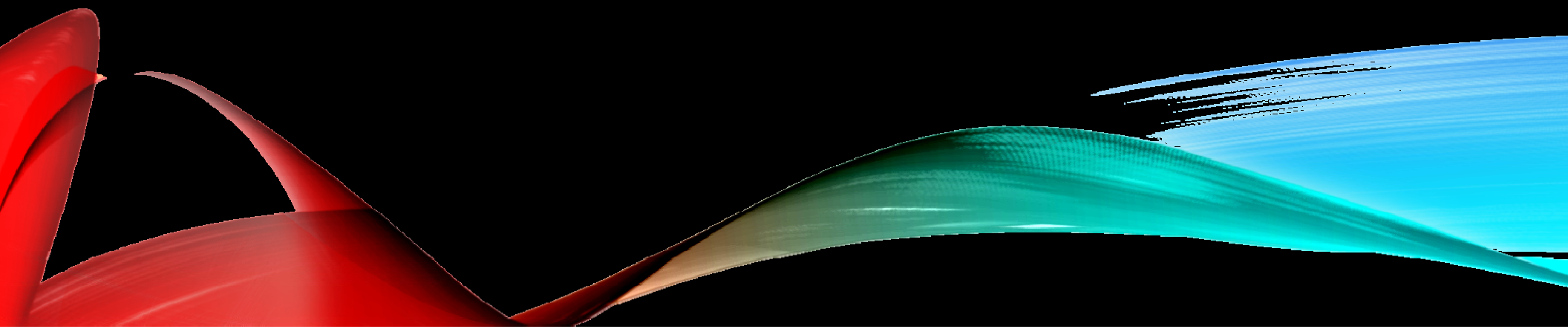
In the **Prediction Profiler**, click and drag the vertical red line for a variable to change the level or value. The predicted probabilities are displayed.

Note: For more details on logistic regression, see the book ***Fitting Linear Models*** (under **Help > Books**) or search for "multiple logistic regression" in the JMP Help.

Lost Sales: Multiple Regression Model



INTERPRET RESULTS (STATISTICALLY)



Lost Sales

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	19.28543	3	38.57086	<.0001*
Full	361.91279			
Reduced	381.19822			

Model significant?

RSquare (U)	0.0506
AICc	731.899
BIC	749.065
Observations (or Sum Wgts)	550

Model fit?

Measure	Training	Definition
Entropy RSquare	0.0506	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0903	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.6580	∑ -Log(ρ[j])/n
RMSE	0.4832	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.4667	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.4018	∑ (ρ[j]≠ρMax)/n
N	550	n

Examine estimates

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.54119866	0.1667172	10.54	0.0012*
Quote	-0.0000193	3.0733e-5	0.39	0.5299
Time to Delivery	-0.0183689	0.003483	27.81	<.0001*
Part Type[AM]	0.23555514	0.0983823	5.73	0.0167*

For log odds of Won/Lost

Predictors significant?

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Quote	1	1	0.39541571	0.5295
Time to Delivery	1	1	32.790892	<.0001*
Part Type	1	1	5.78404388	0.0162*

Model predictive?

USEFULNESS & PREDICTIVE ABILITY

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	19.28543	3	38.57086	<.0001*
Full	361.91279			
Reduced	381.19822			

RSquare (U)	0.0506
AICc	731.899
BIC	749.065
Observations (or Sum Wgts)	550

Measure	Training	Definition
Entropy RSquare	0.0506	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.0903	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.6580	$\sum$ -Log(ρ[j])/n
RMSE	0.4832	$\sqrt{\sum(y[j]-\rho[j])^2/n}$
Mean Abs Dev	0.4667	$\sum \lvert y[j]-\rho[j]\rvert/n$
Misclassification Rate	0.4018	$\sum$ (ρ[j]≠ρMax)/n
N	550	n

We look at p-value for overall model significance.

We look at RSquare for a measure of fit

We look at RMSE and Mean Abs Dev for a measure of error.

We look at misclassification rate for a measure of predictive ability

ASSESS AND UNDERSTAND MODEL

"We interpret coefficients for logistic regression differently than linear regression coefficients.* Increasing an X by one unit changes the log odds by b_1 or it multiplies the odds by e^{b1} but the rate of change depends on the current value of X.

In general: if a b is positive, then increasing X will be associated with increasing p(x), and if b is negative, then increasing X will be associated with a decrease in p(x)."

*Chapter 4: Classification, An Introduction to Statistical Learning by James, et al (Springer)

Lost Sales

Parameter Estimates

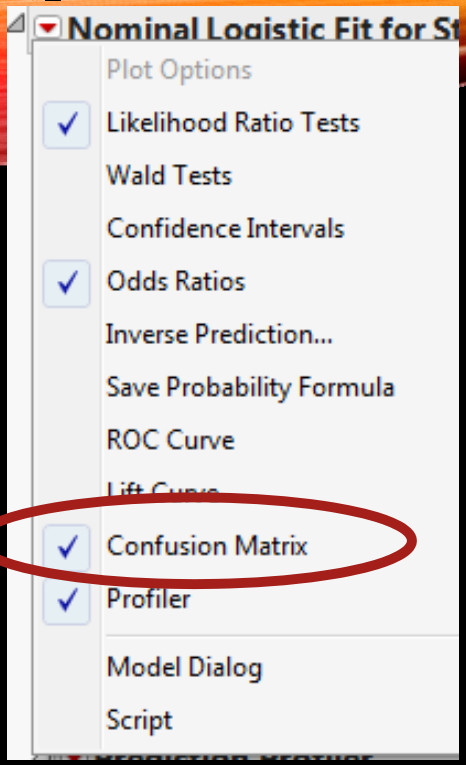
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.54119866	0.1667172	10.54	0.0012*
Quote	-0.0000193	3.0733e-5	0.39	0.5299
Time to Delivery	-0.0183689	0.003483	27.81	<.0001*
Part Type[AM]	0.23555514	0.0983823	5.73	0.0167*

For log odds of Won/Lost

Unit Odds Ratios

Per unit change in regressor

Term	Odds Ratio	Lower 95%	Upper 95%	Reciprocal
Quote	0.999981	0.99992	1.000041	1.0000193
Time to Delivery	0.981799	0.974896	0.988314	1.0185386



Confusion Matrix

Actual		Predicted
Training	**Won**	**Lost**
Won	193	85
Lost	136	136

USEFULNESS & PREDICTIVE ABILITY

Misclassification Rate	0.4018

Confusion Matrix

Training

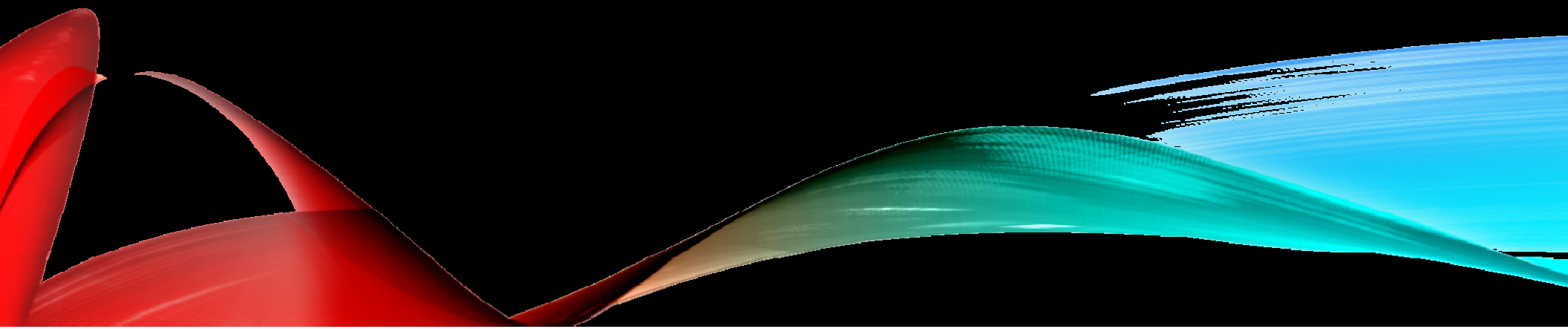
Actual Status	Predicted Won	Predicted Lost
Won	193	85
Lost	136	136

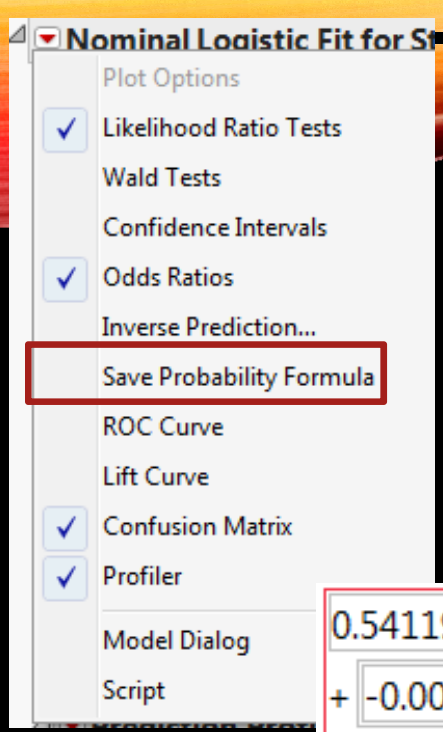
We look at the confusion matrix (off diagonal cells to understand predictive ability).

For example, our model predicted a Lost 85 times [or 85/(193+85) = 30 percent] when in reality it was a Win.

Similarly, our model predicted a Win 136 times when the reality was a Lost, or 50% of the cases.

INTERPRET RESULTS (APPLICATION)





Lin[Won]	Prob[Won]	Prob[Lost]	Most Likely Status
0.7094655787	0.6702830615	0.3297169385	Won
0.0099710991	0.5024927541	0.4975072459	Won

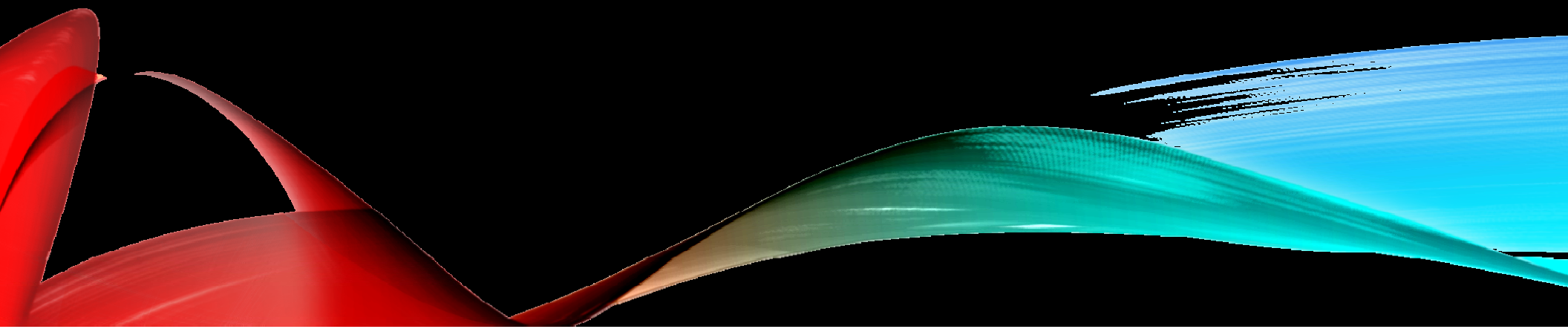
$$\frac{1}{\left(1 + \text{Exp}\left(-\textit{Lin[Won]}\right)\right)}$$

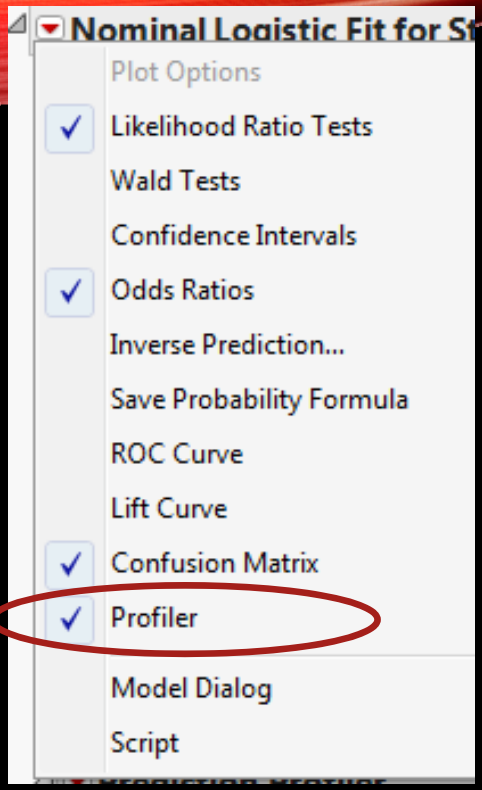
$$\frac{1}{\left(1 + \text{Exp}\left(\textit{Lin[Won]}\right)\right)}$$

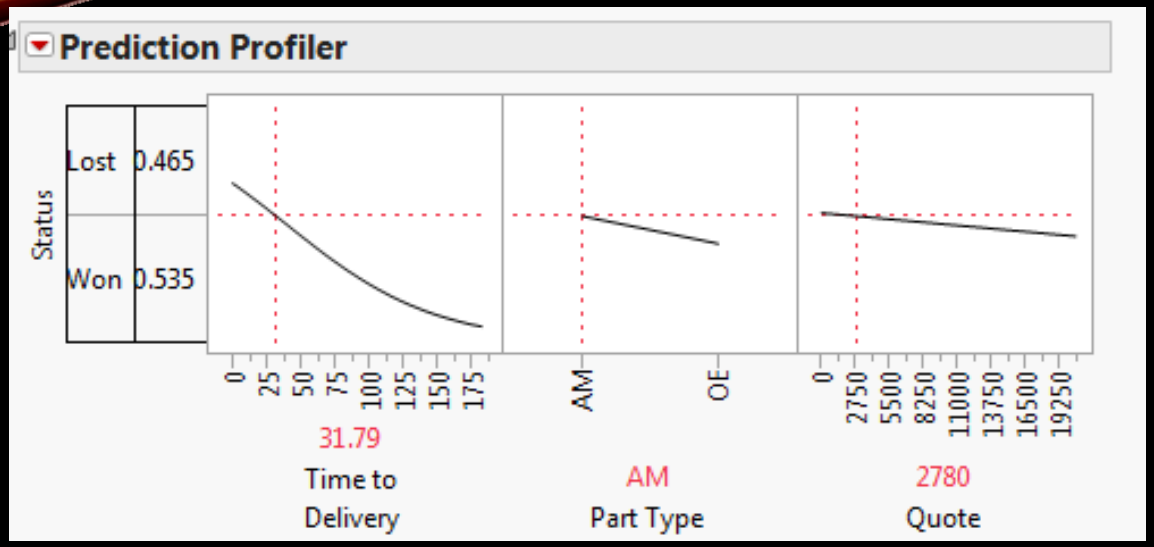
$$0.54119865783093 + -0.0000193051213 * \textit{Quote} + -0.0183688946473 * \textit{Time to Delivery} + \text{Match}\left(\textit{Part Type}, \text{"AM"} \Rightarrow 0.23555513631327,\ \text{"OE"} \Rightarrow -0.2355551363133,\ \text{else} \Rightarrow .\right)$$

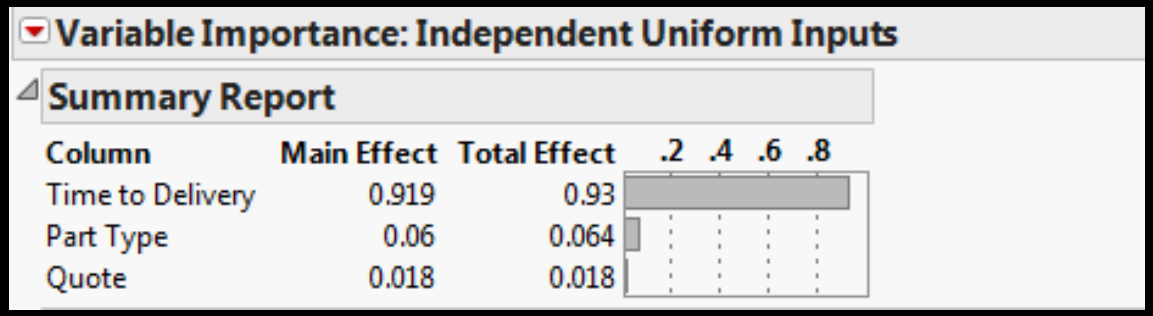
$$\text{IfMax}\left(\textit{Prob[Won]} \Rightarrow \text{"Won"},\ \textit{Prob[Lost]} \Rightarrow \text{"Lost"},\ \text{else} \Rightarrow \text{""}\right)$$

HOW TO UNDERSTAND THE MANAGERIAL IMPLICATIONS

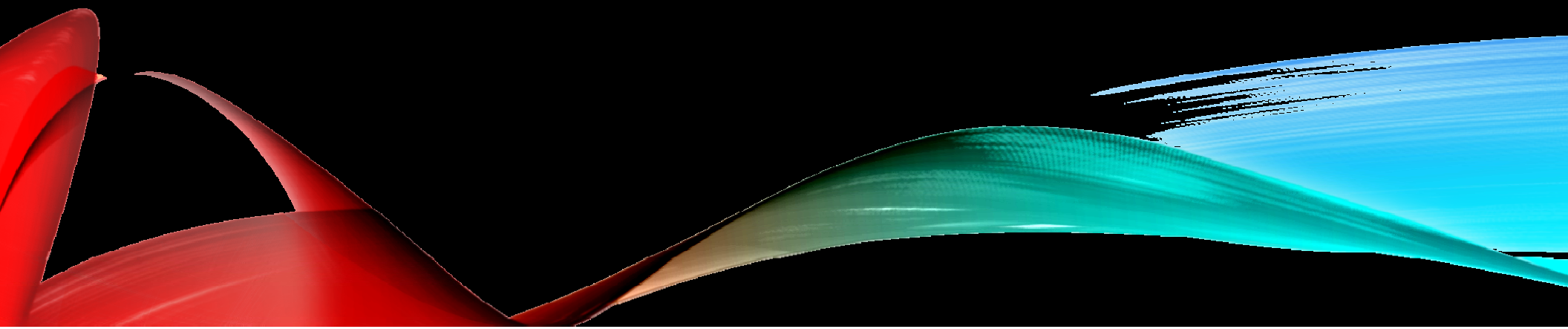



Nominal Logistic Fit for St
Plot Options
Likelihood Ratio Tests
Wald Tests
Confidence Intervals
Odds Ratios
Inverse Prediction...
Save Probability Formula
ROC Curve
Lift Curve
Confusion Matrix
Profiler
Model Dialog
Script


Prediction Profiler
Status
Lost 0.465
Won 0.535
0 25 50 75 100 125 150 175
AM OE
0 2750 5500 8250 11000 13750 16500 19250
31.79
Time to Delivery
AM
Part Type
2780
Quote


Variable Importance: Independent Uniform Inputs
Summary Report
Column Main Effect Total Effect
Time to Delivery 0.919 0.93
Part Type 0.06 0.064
Quote 0.018 0.018
.2 .4 .6 .8

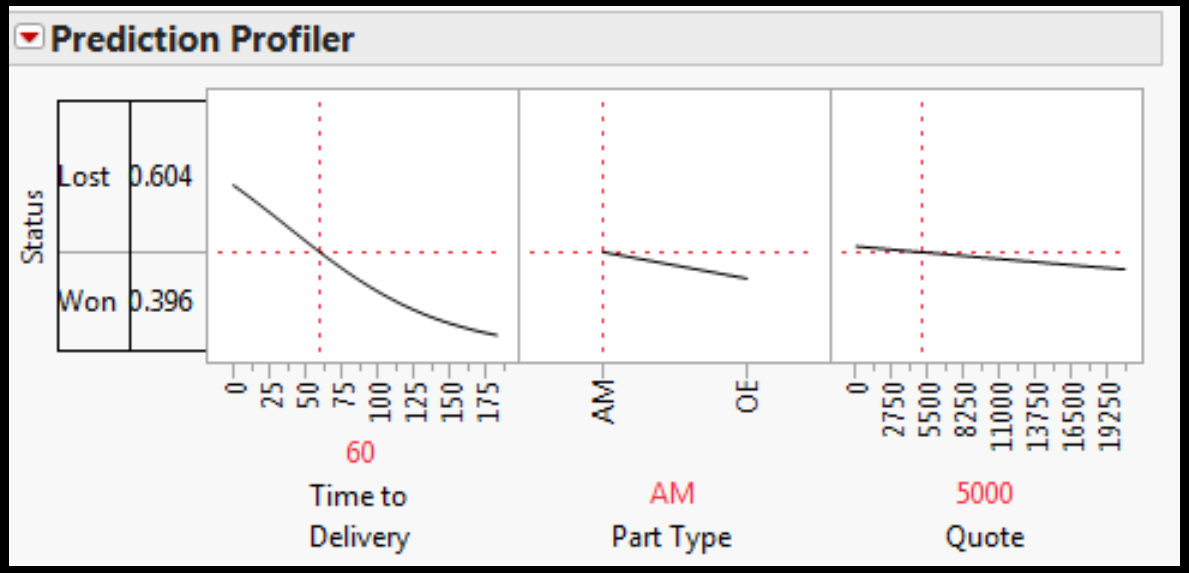
HOW TO APPLY THE RESULTS



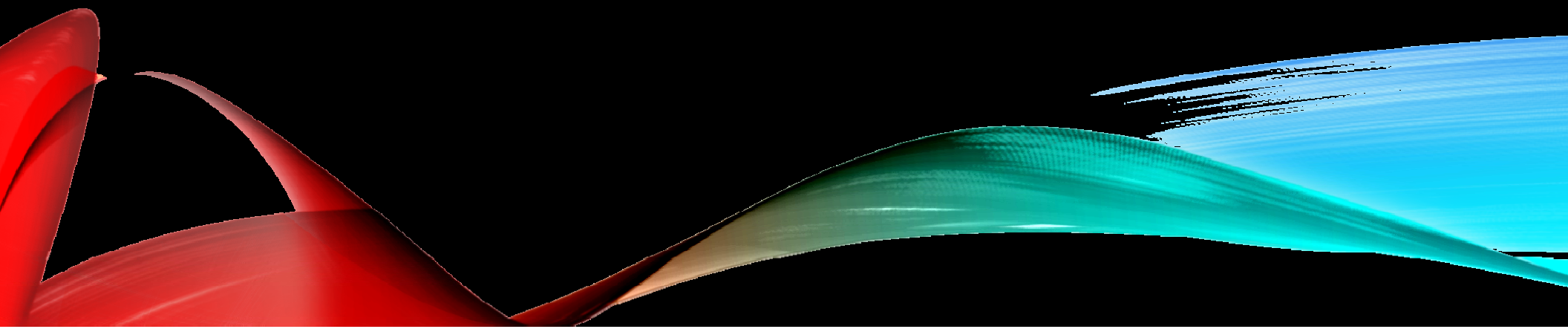
USING PROFILER TO PREDICT

What is the estimated probability of a Win for a proposal of:

Quote = 5000
Time to Delivery = 60
Part Type = AM

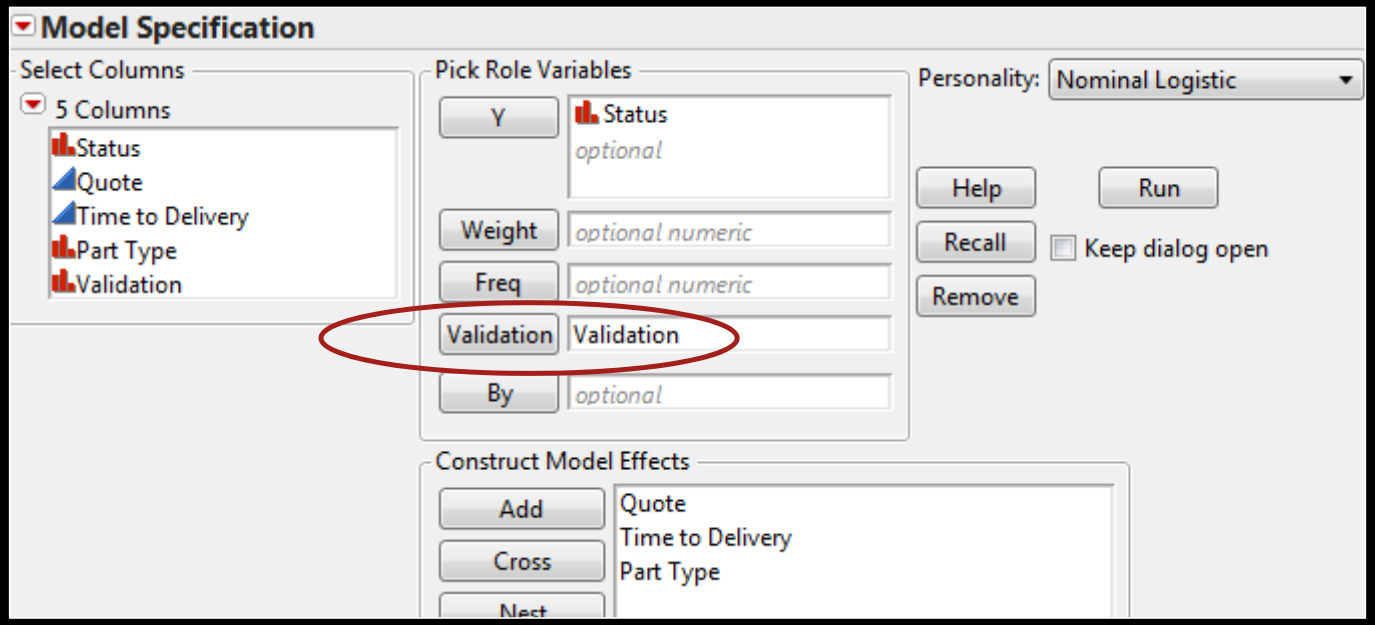


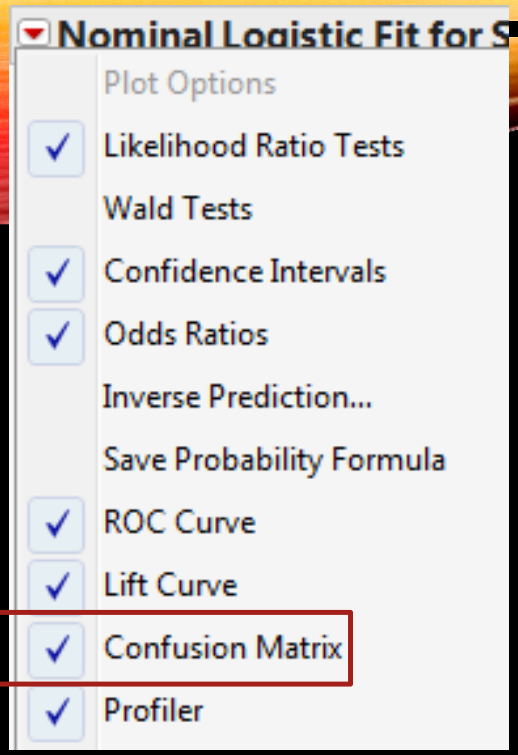
LOGISTIC REGRESSION PREDICTIVE MODELING



Developing a **validation variable** and then using the "validation" portion to evaluate predictive ability of the model.

	Status	Quote	Time to Delivery	Part Type	Validation
1	Lost	1153	16	OE	Validation
2	Lost	2313	46	AM	Training
3	Lost	2681	63	OE	Validation
4	Lost	845	16	OE	Validation
5	Lost	2213	61	AM	Training



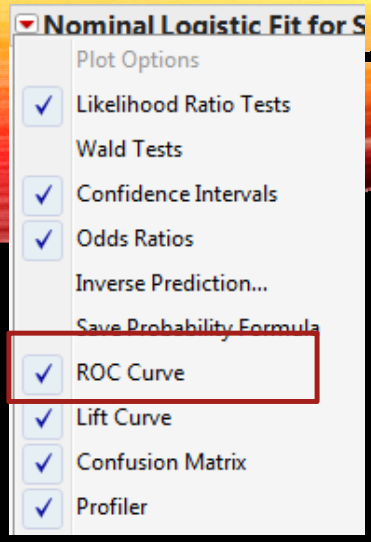


Measure	Training	Validation	Definition
Entropy RSquare	0.0748	-0.006	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.1312	-0.011	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.6399	0.6959	∑ -Log(ρ[j])/n
RMSE	0.4741	0.5013	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.4495	0.4730	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.3758	0.4636	∑ (ρ[j]≠ρMax)/n
N	330	220	n

Confusion Matrix

Actual		Predicted
Training	**Won**	**Lost**
Won	125	49
Lost	75	81

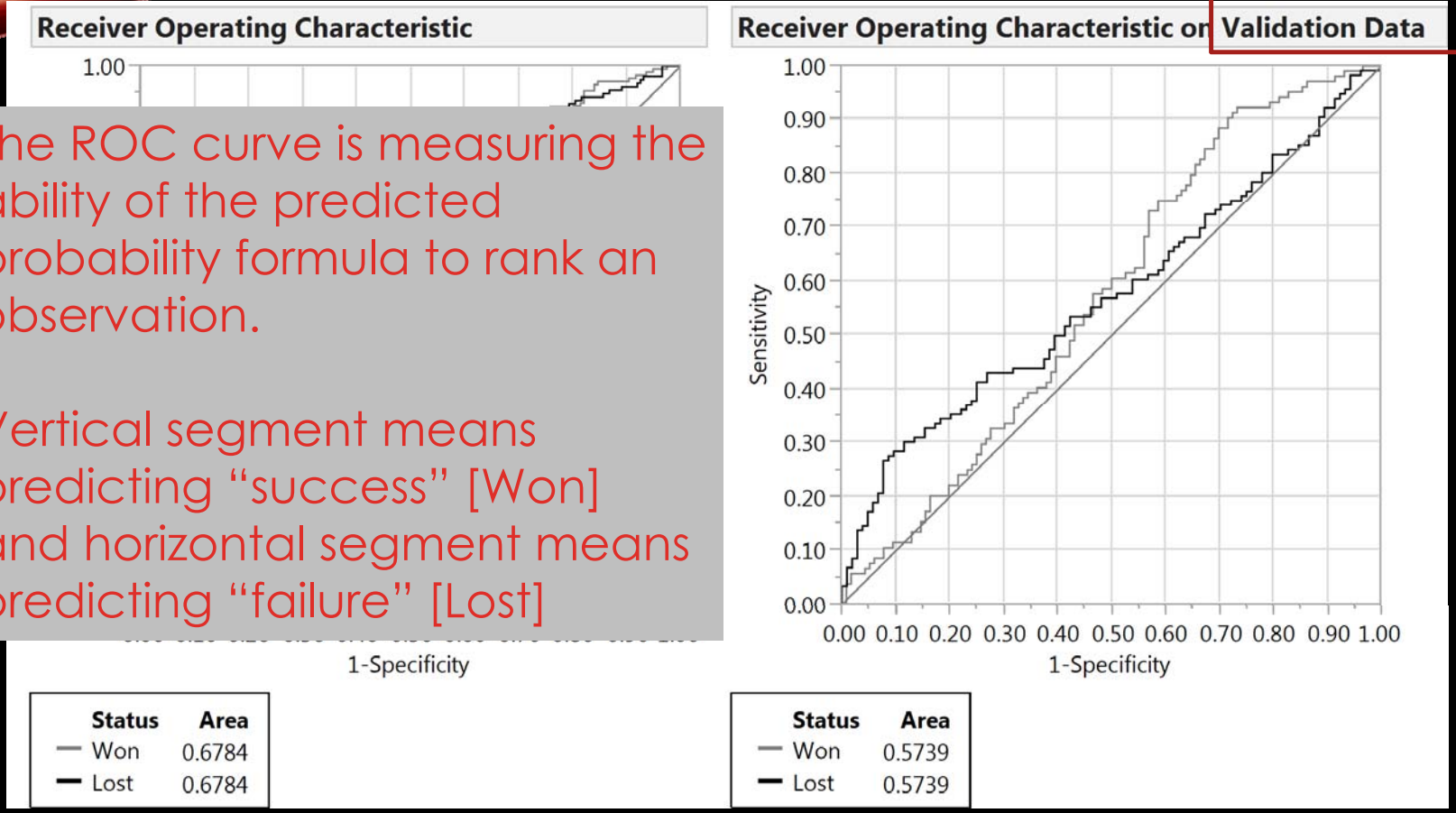
Actual		Predicted
Validation	**Won**	**Lost**
Won	65	39
Lost	63	53



Look at AUC (area under curve) – Higher better

The ROC curve is measuring the ability of the predicted probability formula to rank an observation.

Vertical segment means predicting "success" [Won] and horizontal segment means predicting "failure" [Lost]



Constructing an ROC Curve

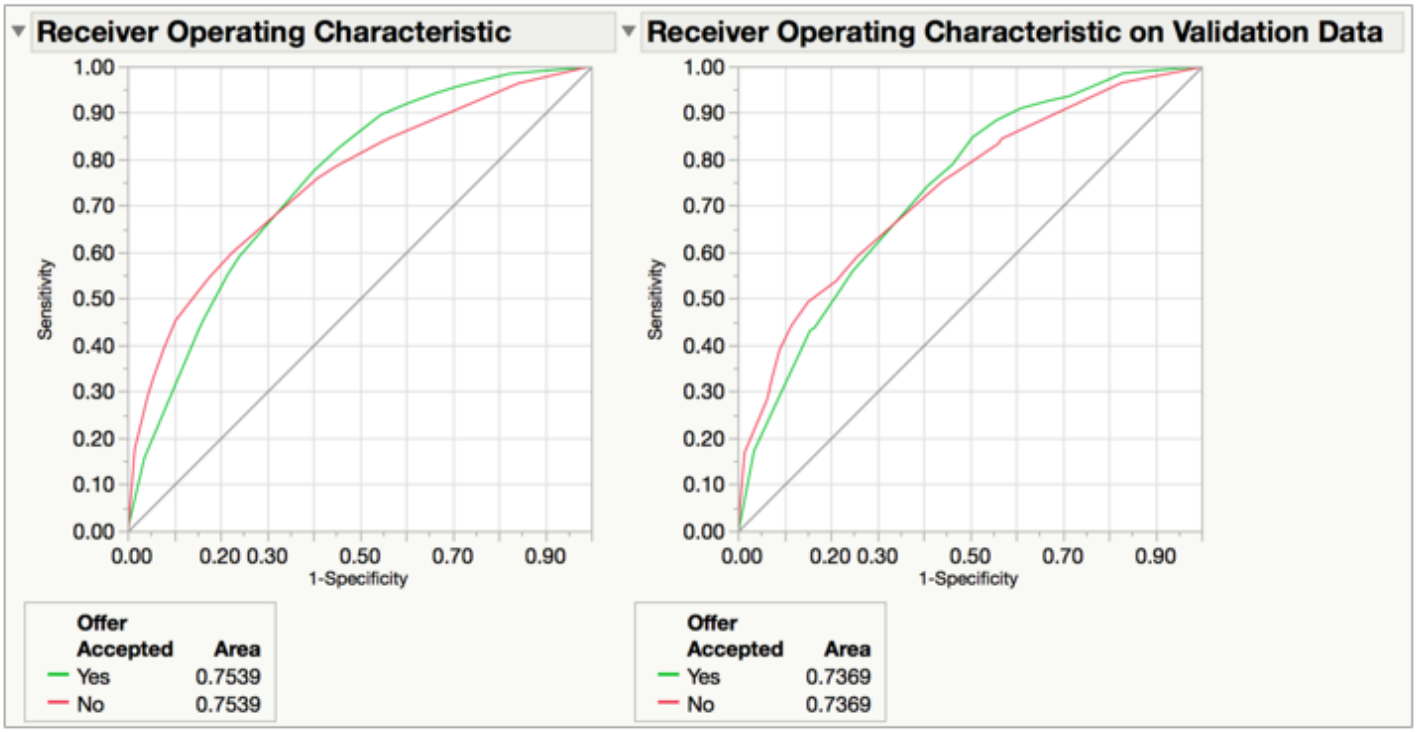
Here is a practical algorithm to quickly draw an ROC curve after the table has been sorted by the predicted probability. We walk through the algorithm for **Offer Accepted=Yes**, but this is done automatically in JMP for each response category.

For each observation in the sorted table, starting at the observation with the highest probability **Offer Accepted = Yes**:

- If the observed response value is **Yes**, then a vertical line segment (increasing, along the Sensitivity axis) is drawn. The length of the line segment is 1/(total number of **Yes** responses in the table)

If the observed response value is No, then a horizontal line segment (increasing, along the 1-Specificity axis) is drawn. The length of the line segment is 1/(total number of "No" responses in the data table)

Figure 6.30: Credit, ROC Curve for Offer Accepted

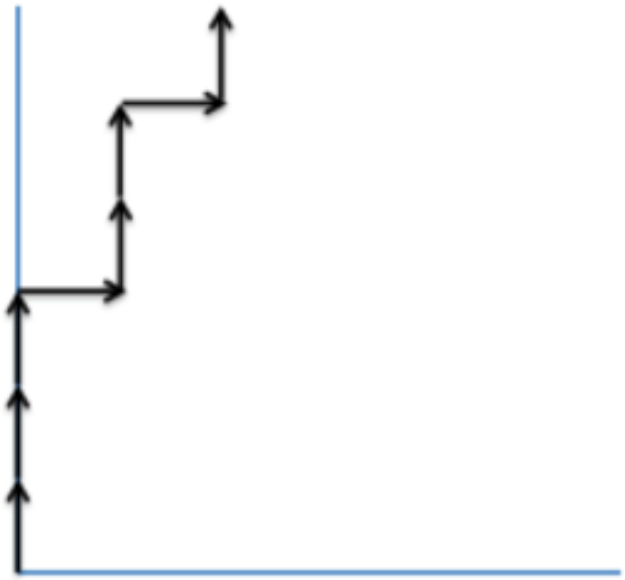


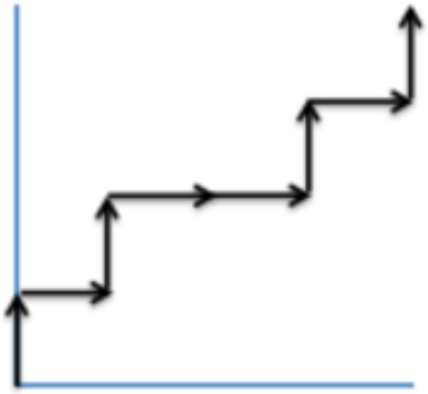
Simple ROC Curve Examples

We use a simple example to illustrate. Suppose we have a data table with only 8 observations. We sort these observations from high to low based on the probability that the **Outcome=Yes**. The sorted actual response values are **Yes**, **Yes**, **Yes**, **No**, **Yes**, **Yes**, **No**, and **Yes**. This results in the ROC curve on the left of Figure 6.31 (arrows are added to show the steps in the ROC curve construction). The first 3 line segments are drawn up because the first three sorted values have **Outcome=Yes.**

Now, suppose we have different probability model that we use to rank the observations, resulting in the sorted outcomes **Yes**, **No**, **Yes**, **No**, **No**, **Yes**, **No**, and **Yes**. The ROC curve for this situation is shown on the right of Figure 6.31. The first ROC curve moves "up" faster than the second curve. This is an indication that the first model is doing a better job as separating the **Yes** responses from the **No** responses, based on the predicted probability.

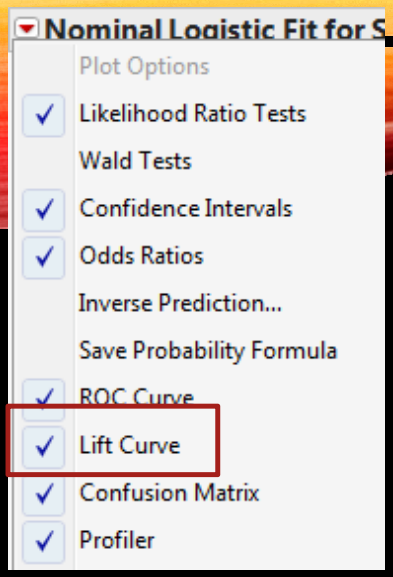
Figure 6.31: ROC Curve Examples

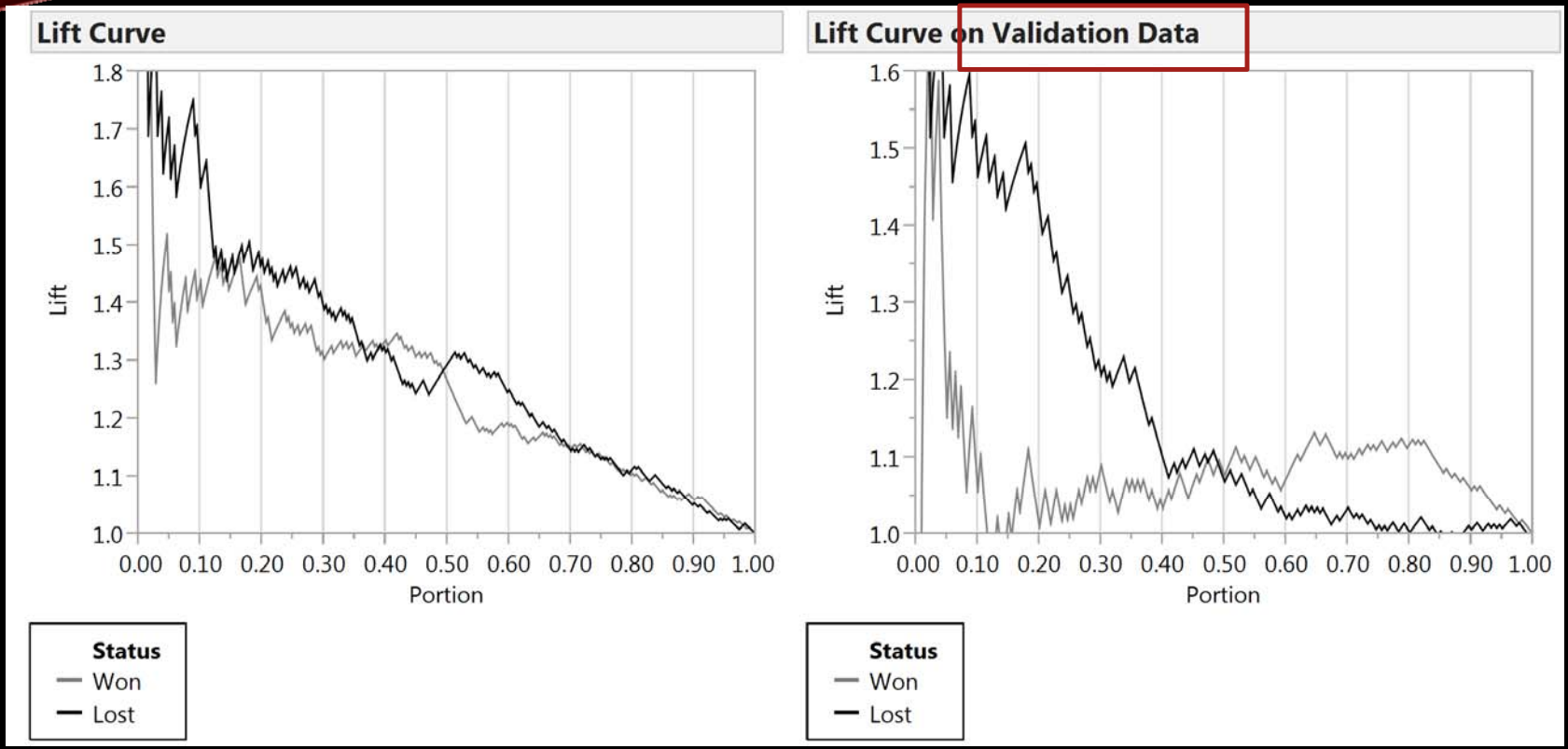




Referring back to the example ROC curve in Figure 6.30, we see that JMP Pro has also displayed a diagonal reference line on the chart, which represents the *Sensitivity = 1-Specificity* line. If a probability model cannot sort the data into the correct response category, then it may be no better than just sorting at random. In this case, the ROC curve for a "random ranking" model would be similar to this diagonal line. A model that sorts the data perfectly, with all the **Yes** responses at the top of the sorted table, would have an ROC Curve that goes from the origin of the graph straight up to sensitivity = 1, then straight over to 1-specificity = 1. A model that sorts perfectly can be made into a classifier rule that classifies perfectly - that is, a classifier rule that has a sensitivity of 1.0 and 1-specificity of 0.0.

The *area under the curve*, or *AUC* (labeled **Area** in Figure 6.30) is a measure of how good our model does at sorting the data. The diagonal line, which would represent a random sorting model, has an AUC of 0.5. A perfect sorting model has an AUC of 1.0. The area under the curve for **Offer Accepted** = **Yes** is 0.7369 (see Figure 6.30), indicating that the model predicts better than the random sorting model.

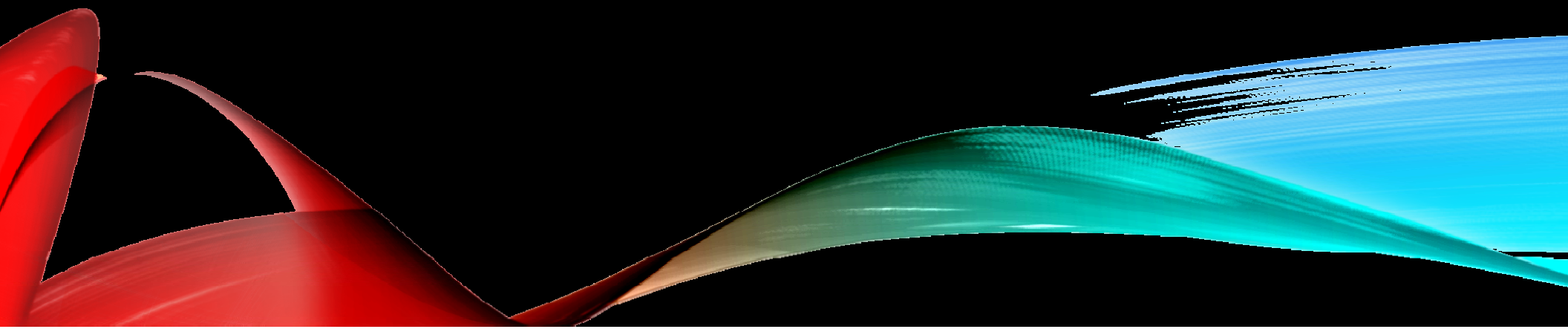

Nominal Logistic Fit for S
Plot Options
Likelihood Ratio Tests
Wald Tests
Confidence Intervals
Odds Ratios
Inverse Prediction...
Save Probability Formula
ROC Curve
Lift Curve
Confusion Matrix
Profiler

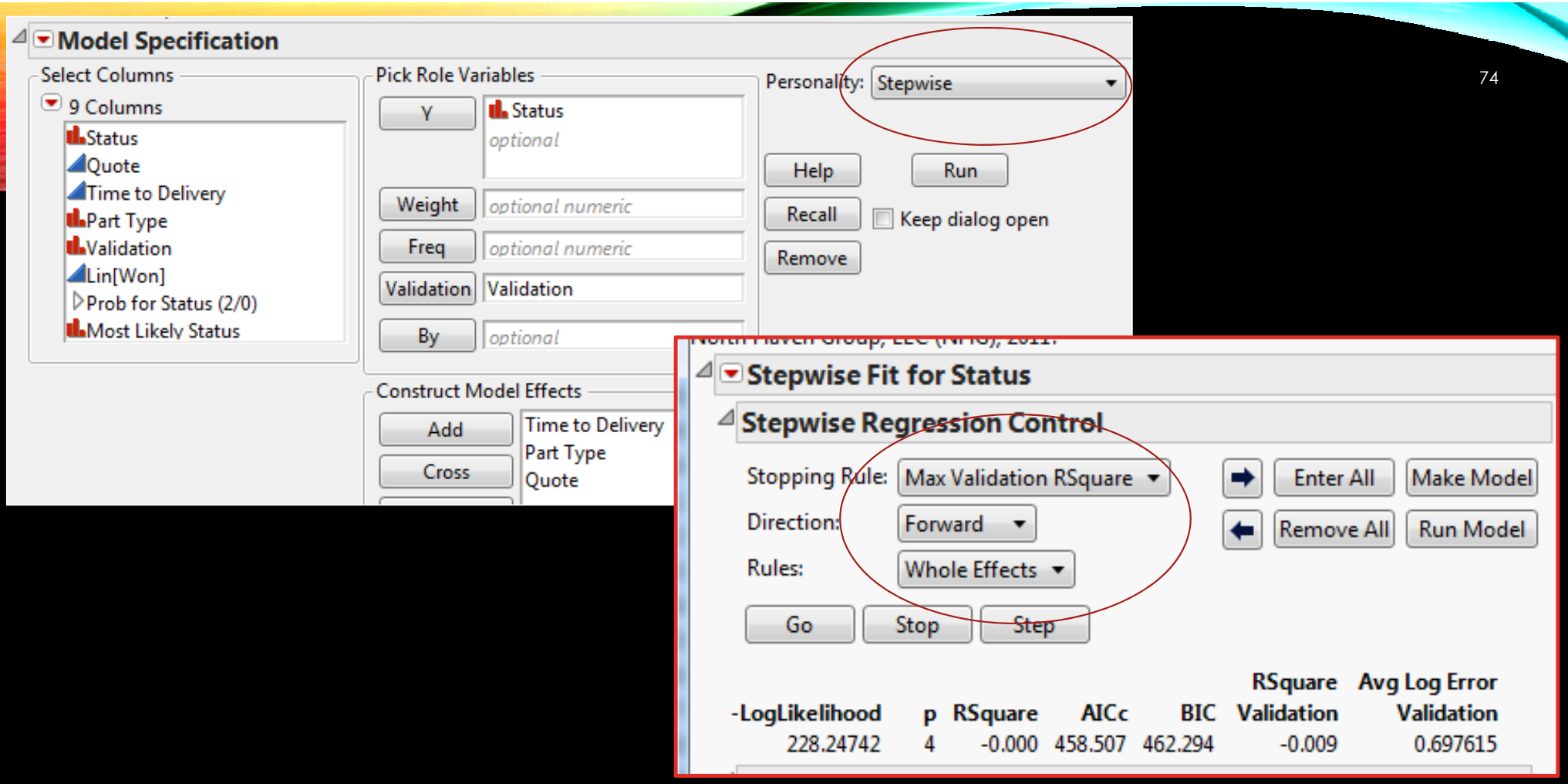

Lift Curve
Lift
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.00 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00
Portion
Status
Won
Lost
Lift Curve on Validation Data
Lift
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.00 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00
Portion
Status
Won
Lost

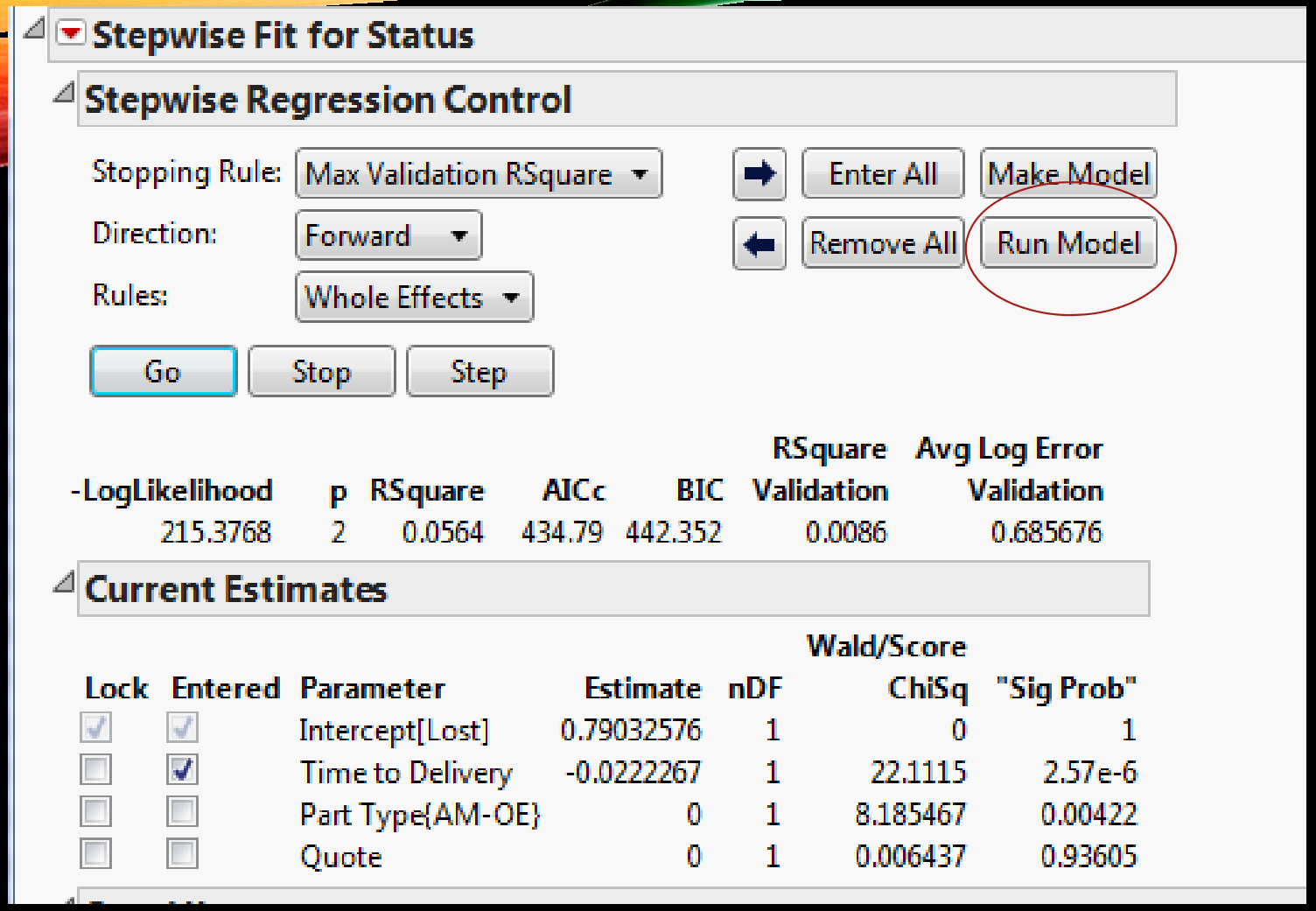
Another measure of how well a model can sort outcomes is the model ***lift***. As with the ROC curve, we examine the table that is sorted in descending order of the predicted probability. For each sorted row, we calculate the sensitivity, and we divide that by the proportion of values in the table where **Outcome=Yes**. This value is the model lift.

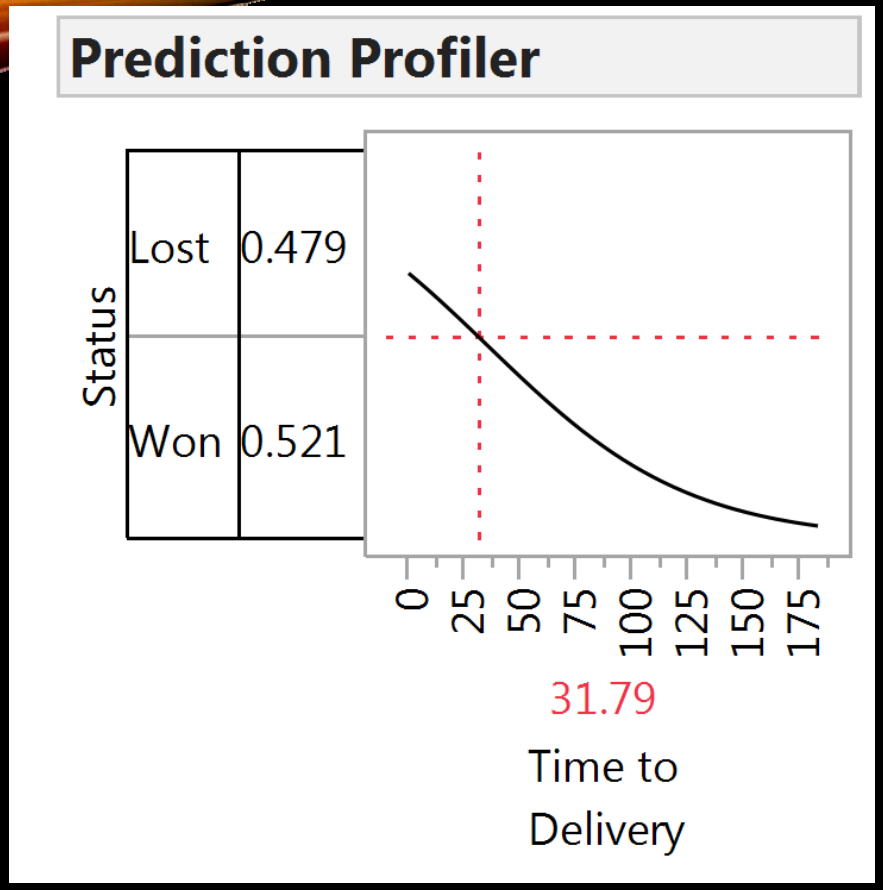
Lift is a measure of how much "richness" in the response we achieve by applying a classification rule to the data. A **Lift Curve** plots the **Lift** (on the y-axis) against the **Portion** (on the x-axis). Again, if we consider the data table that has been sorted by the predicted probability of a given outcome, as we go down the table from the top to the bottom, the portion is just the relative position of the row we are considering. The top 10% of the rows in the sorted table corresponds to a portion of 0.1, the top 20% of the rows corresponds to a portion of 0.2, and so on. The lift for **Outcome=Yes**, for a given portion, is simply the proportion of **Yes** responses in this portion, divided by overall proportion of **Yes** responses in the entire data table.

STEPWISE




Model Specification
Select Columns
9 Columns
Status
Quote
Time to Delivery
Part Type
Validation
Lin[Won]
Prob for Status (2/0)
Most Likely Status
Pick Role Variables
Y
Status
optional
Weight
optional numeric
Freq
optional numeric
Validation
Validation
By
optional
Construct Model Effects
Add
Cross
Time to Delivery
Part Type
Quote
Personality:
Stepwise
Help
Run
Recall
Keep dialog open
Remove
Stepwise Fit for Status
Stepwise Regression Control
Stopping Rule:
Max Validation RSquare
Direction:
Forward
Rules:
Whole Effects
Enter All
Make Model
Remove All
Run Model
Go
Stop
Step
-LogLikelihood
p
RSquare
AICc
BIC
RSquare Validation
Avg Log Error Validation
228.24742
4
-0.000
458.507
462.294
-0.009
0.697615


Stepwise Fit for Status
Stepwise Regression Control
Stopping Rule: Max Validation RSquare
Direction: Forward
Rules: Whole Effects
Enter All
Make Model
Remove All
Run Model
Go
Stop
Step
-LogLikelihood p RSquare AICc BIC RSquare Validation Avg Log Error Validation
215.3768 2 0.0564 434.79 442.352 0.0086 0.685676
Current Estimates
Lock Entered Parameter Estimate nDF Wald/Score ChiSq "Sig Prob"
Intercept[Lost] 0.79032576 1 0 1
Time to Delivery -0.0222267 1 22.1115 2.57e-6
Part Type{AM-OE} 0 1 8.185467 0.00422
Quote 0 1 0.006437 0.93605


Prediction Profiler
Status
Lost
0.479
Won
0.521
0
25
50
75
100
125
150
175
31.79
Time to
Delivery