Marketing Analytics – Week 5

Suneel Grover, M.B.A., M.S.

George Mason University (Dept. of Marketing – School of Business) – Adjunct Professor

SAS® Advisory Solutions Architect

GEORGE MASON UNIVERSITY

Agenda

1. Introduction
2. Lecture & Demos: **Logistic Regression**

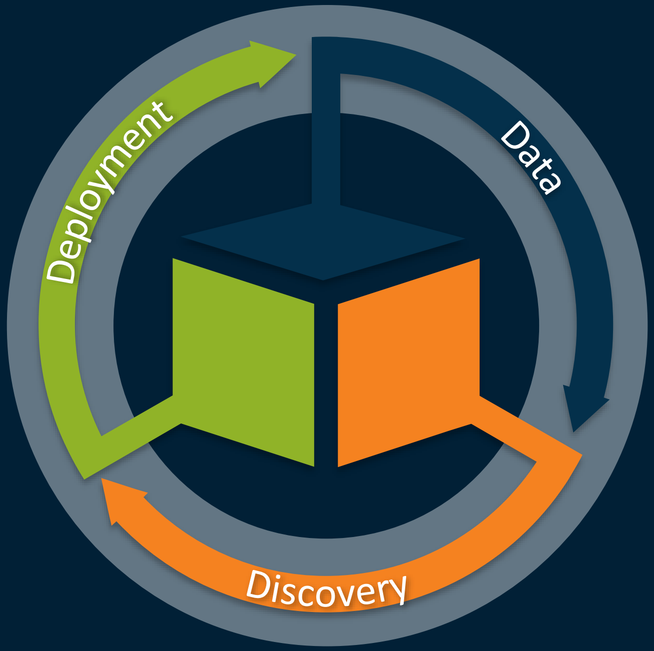
Setting Expectations

- We will present important logistic regression functionality within JMP but we cannot cover every feature.

- You will be introduced to logistic regression, with the goal of understanding the analytical approach's mechanics and concepts

Getting Started

- Logistic regression can be used to model the probability that an event will occur
  - Used to identify circumstances that make it more likely that a given event will occur



Business Applications

Predictions & examples from everyday life:

1. Probability someone will win an Oscar, an election, or a soccer match, based on other events that have transpired
2. Identify potential fraudulent checks or bank transactions
3. Predict flight delays
4. Understand reasons why employees leave for another job
5. Determine which content to display on a website or mobile app based on visitor interactions

Previewing The End Result

Logistic Regression Model

- Predict whether a customer will respond to a new product offer
- The response (dependent) is a yes/no **categorical** variable, and we want to predict the probability (*p*) that the customer will respond, based on two predictor variables: customer age; and the number of days since last purchase:

$\text{Log}(\hat{p} / 1 - \hat{p}) = -3.298 + 0.08365\ (\text{Age}) + 0.00364\ (\text{Days since last Purchase})$

Previewing the End Result

Customer Age: 50 years old

Days since last Purchase: 100 days

Predicted probability of purchase:

-3.298 + (0.08365 * 50) + (0.00364 * 100) = **0.78**

Objectives

The objectives of logistic regression are as follows:

- Allows us to fit a regression model to a set of data to develop a **“typical looking”** regression equation with coefficients
- Logistic differs from multiple linear regression (least squares regression) in several ways:

  - ❖ The response is **categorical**, often binary (yes or no, 0 or 1)
  - ❖ For a binary response, instead of predicting the event directly, we build a model to **predict the probability of one level of the event (either yes or no) occurring**
  - ❖ Probability is **bounded, [0, 1]**, but the response in a linear regression model is unbounded, $(-\infty, \infty)$.
  - ❖ As a result, the mathematical method is **more complicated**

Objectives

When comparing linear regression to logistic, keep this mind:

- Linear regression is the equation for a **straight line**
- Among the properties of a straight line is that it **goes on forever**, continuously in both directions with no minimum or maximum (except when the line is parallel with the x-axis)
- These properties make linear regression well-suited to **estimating continuous quantities** that take on a wide range of values
- These properties make linear regression unsuitable **for modeling binary outcomes, such as yes/no, or good/bad**
- Because binary problems are extremely common, **it is not surprising that statisticians have found a way to adapt regression models with logistic**

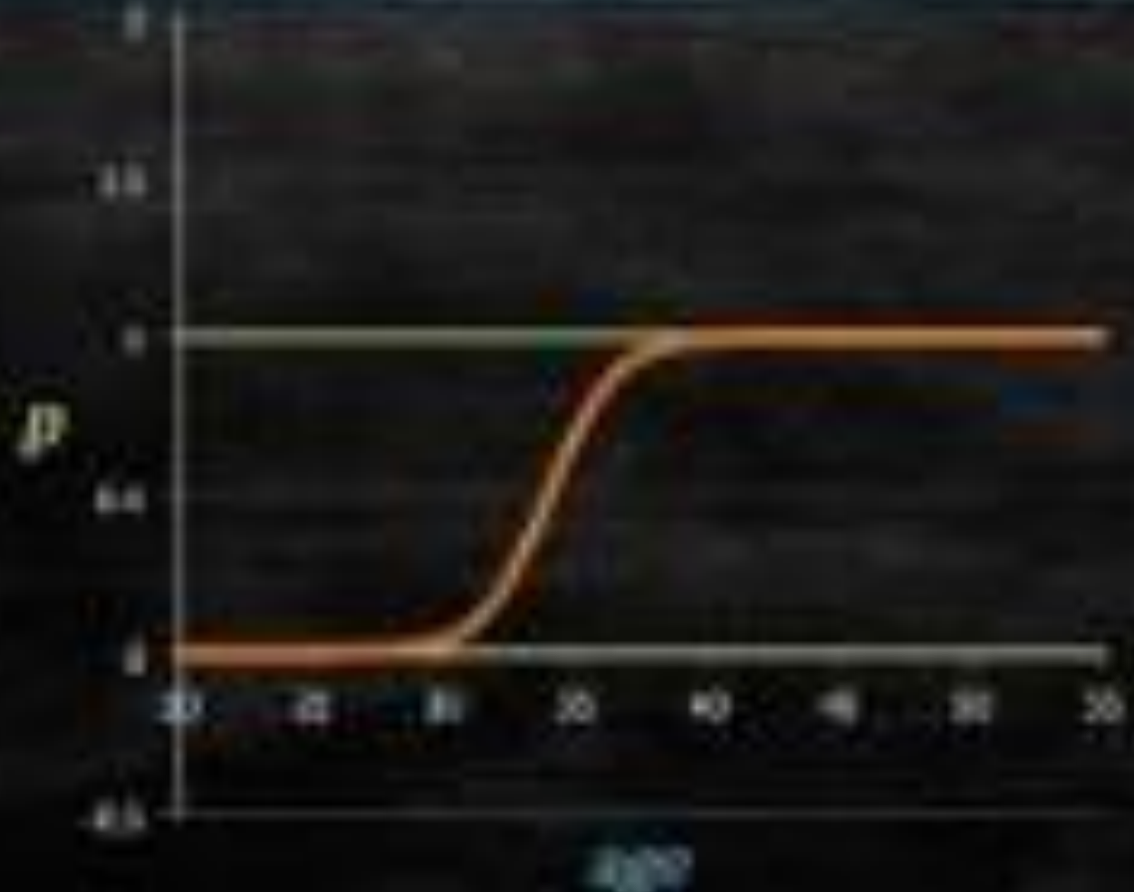
Modeling Binary Outcomes

- Modeling binary outcomes does not seem like an estimation task
  - There are two categories, and the task is to assign each record to one or the other
  - Surely, this is a classification task – **and it is!**
- However, the task can be restated as:

"What is the probability that the record belongs to class one?

Because probabilities are numbers, **<u>the problem is now an estimation task.</u>**

Logistic Model Plot



How the Algorithm Works

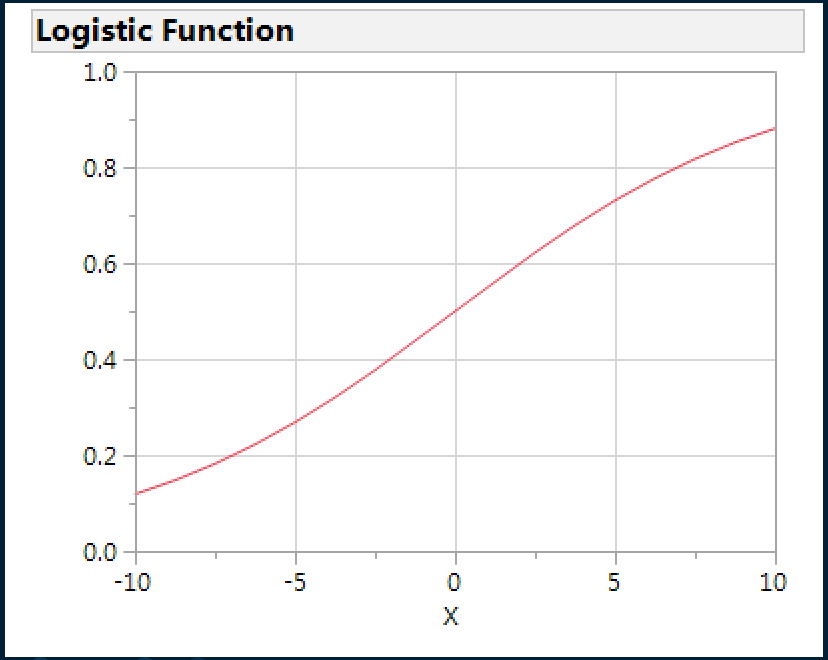
For logistic regression, the probability of an event, *p*, is related to predictive factors (X_1, X_2, …, X_k) by the mathematical relationship:

$$\boldsymbol{Log\ (p\ /\ (1 - p)) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k}$$

Important Takeaways:

- *Log (p / (1 – p))* is sometimes referred to as the **log-odds** or **logit**
- The right-side of this equation is similar to multiple linear regression (without the error)
- Rearranging this formula to solve for *p* directly: $\boldsymbol{p = 1\ /\ (1 + e^{-(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)})}$

How the Algorithm Works



Interpreting the Logistic Model:

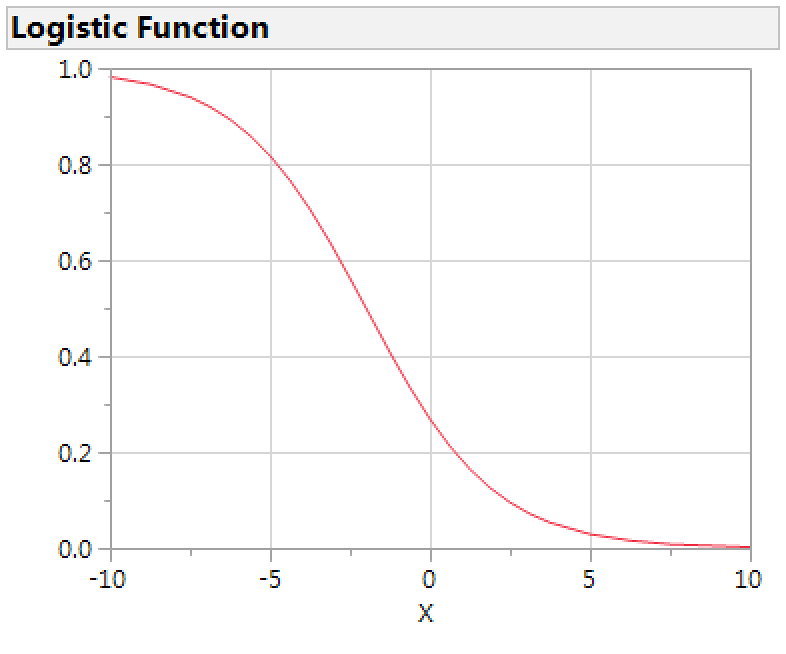
- The intercept β_0 of the logit is 0, and the slope β_1 is 0.2
- The model is written as:

$$\log(p / (1 - p)) = \beta_0 + \beta_1 X_1 = 0 + 0.2 * X$$

$$p = 1 / (1 + e^{-(0 + 0.2X)})$$

- At the intercept, the probability is 0.5, and the slope is positive, so an increase in the value of X results in an increase in the probability of the event

How the Algorithm Works



Interpreting the Logistic Model:

- The intercept β_0 of the logit is -1, and the slope β_1 is - 0.5
- The model is written as:

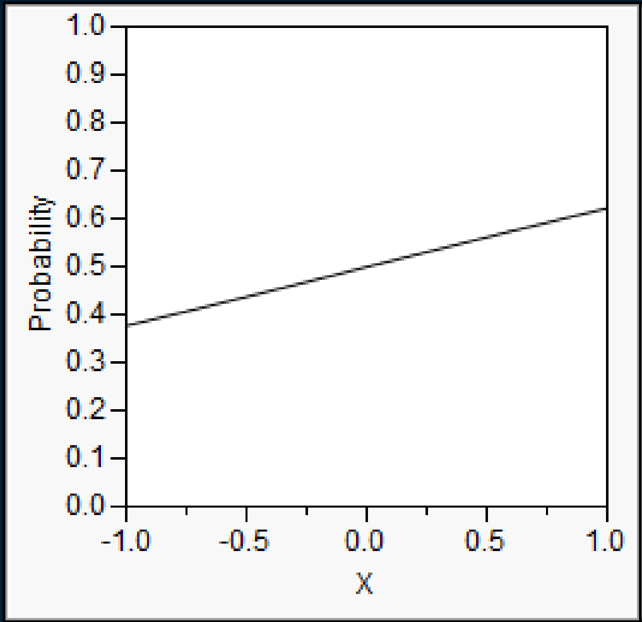
$$\log\left(p / (1 - p)\right) = \beta_0 + \beta_1 X_1 = -1 - 0.5 * X$$

$$p = 1 / (1 + e^{1 + 0.5X)})$$

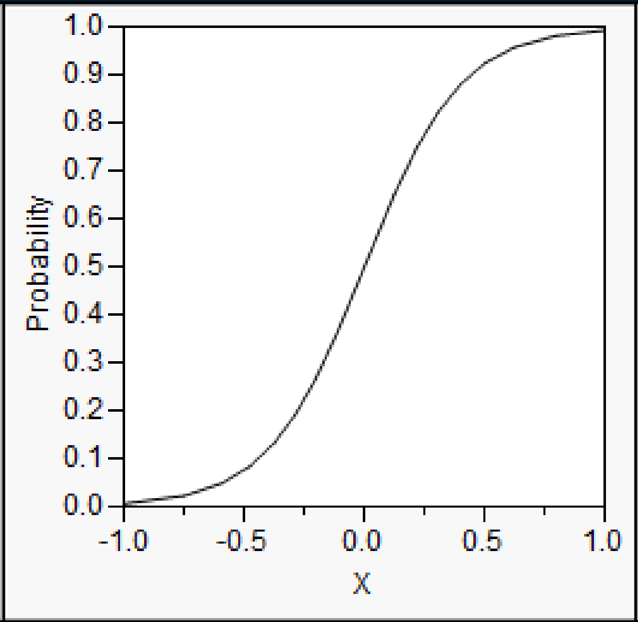
- An increase in the value of X results in a decrease in the probability of the event

* The intercept β_0 is the *log of the odds* of the predicted event level, only when $X_1 = 0$

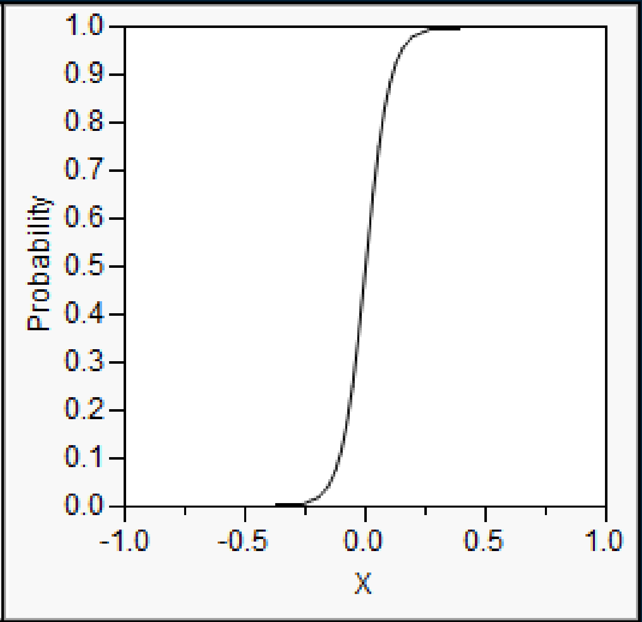
How the Algorithm Works



Weak Relationship



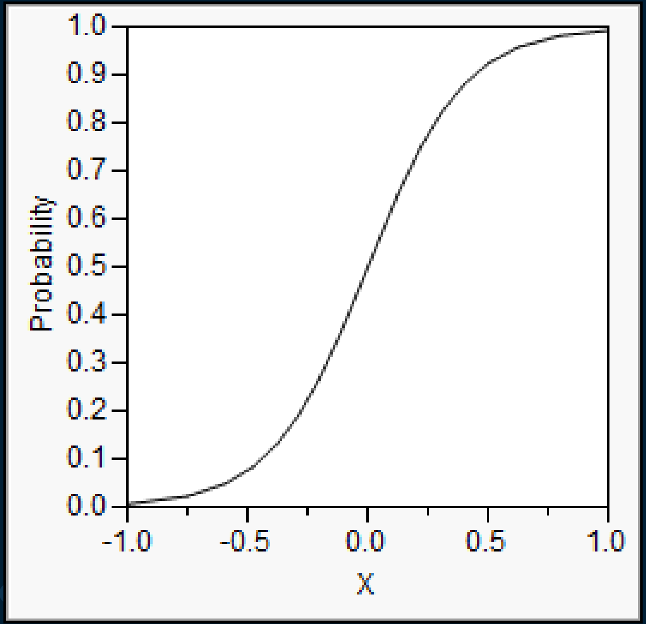
Strong Relationship



Very Strong Relationship

GEORGE MASON UNIVERSITY

Asymptotic Properties



Strong Relationship

- The graph shows that as the explanatory variable initially increases, **the probability of a response slowly increases**.
- The **probability increases at a faster rate near the middle of the curve**, where the relationship is nearly linear.
- Finally, further increases in the predictor **produce little increase** in the response.
- The nonlinear relationship between the probability of the response and the predictor is solely due to the constrained scale of the probabilities.

Method Of Maximum Likelihood

Logistic regression models are typically fit using the **method of maximum likelihood**, which is more general than the least squares approach used in linear regression.

- The basic idea is to try to fit the coefficients so that the response prediction (estimated classification) is as close as possible to the observed classification.
- For logistic regression, the likelihood function has exactly one optimal value.
- A numerical optimization technique is used to maximize the likelihood.

Method Of Maximum Likelihood vs. Least Squares

Least Squares: The idea of least squares is that we choose parameter estimates that **minimize the average squared difference between observed and predicted values**.

Method of Maximum Likelihood: Searches over all possible sets of parameter values for a specified model to find the set of values for which the observed sample was most likely. That is, we find the **set of parameter values that, given a model, were most likely to have given us the data that we have in hand.**

Method Of Maximum Likelihood vs. Least Squares

By way of analogy, imagine that you are in a jury for a civil trial. Four things are presented to you during the trial:

1) Charges that specify the purpose of the trial
2) Prosecution's version of the truth
3) Defendant's version of the truth
4) Evidence

Your task on the jury is to decide, in the context of the specified charges and given the evidence presented, which of the two versions of the truth most likely occurred. You are asked to choose which version of the truth was most likely to have resulted in the evidence that was observed and presented.

Method Of Maximum Likelihood vs. Least Squares

Analogously, in logistic regression analysis with maximum likelihood, we are given:

1) A specified conceptual, mathematical, and statistical model
2) One set of values for the parameters of the model
3) Another set of values for the parameters of the model
4) Observed data.

We want to find the set of values for the parameters (or predictors) of the model that are most likely to have resulted in the data that were actually observed. **We do this by searching over all possible sets of values for the parameters, not just two sets.**

Method Of Maximum Likelihood vs. Least Squares

In linear regression, we measure the fit of the model to the data using the sum of squared errors.

In logistic regression with maximum likelihood, the likelihood **measures the fit of the model to the data**. Therefore, we want to choose parameter values **that maximize the likelihood**.

Industry Applications

A fitted logistic regression model can be applied in three ways:

1. Explain or understand how the probability of an event is influenced by the various factors
2. Make predictions of the probability of an event
3. Use the model to build a classifier rule based on the predicted probability

$$\text{Outcome} = \begin{cases} \text{Yes} & \text{if } p > T \\ \text{No} & \text{otherwise} \end{cases}$$

The value of T is often 0.5, but sometimes other values are used to obtain the desired level of correct classification.

Guided Demonstration & Exercise

SAS JMP Pro

Guided Demonstration: Lost Sales Opportunities

- Open the file called **Lost Sales BBM.jmp**
- Launch JMP Pro and open this file (posted on Blackboard)

Data table description:

This file contains 550 records for quotes provided over a six-month period.

Guided Demonstration: Lost Sales Opportunities

- A supplier in the automotive industry wants to **increase sales and expand its market position**
- The sales team provides quotes to prospective customers, and orders are either **won (customer places order) or lost (customer does not place order)**
- To increase sales and expand, the sales team **needs to understand why orders are lost, and are there situations that make it more or less likely that a customer will or will not place an order**

Guided Demonstration: Lost Sales Opportunities

Define the Problem:

- We want to build a model that will provide insight into why some orders are won and others are lost
- Understanding the factors related to winning and losing is strategically very important
- Ultimately, this will lead to increased revenue, growth opportunities, and hopefully, better profits

Guided Demonstration: Lost Sales Opportunities

Prepare for Analysis:

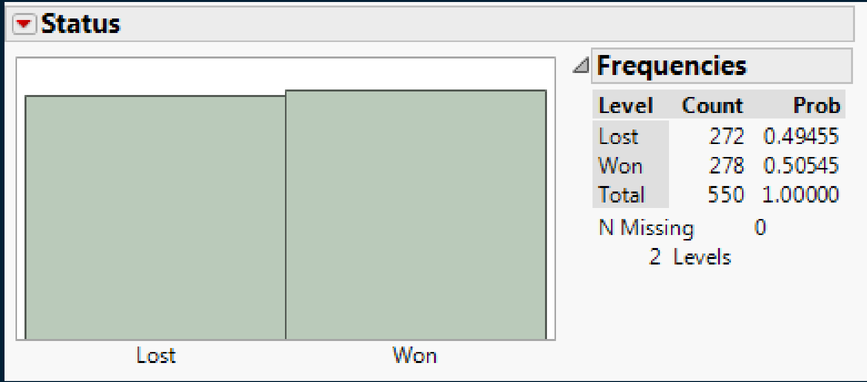
- Always begin by getting to know your data
- In this demonstration, we will explore **variable distributions**, as well as investigate relationships between the response and potential predictor variables through the **Fit Y by X platform**

Guided Demonstration: Lost Sales Opportunities

Exploring one variable at a time:

- Go to **Analyze > Distribution** and then select all variables, click Y, Columns, and OK.
- **After reviewing the distributions, what are some of your initial observations?**
  - ✓ Our Win rate is 50.5%
  - ✓ Quote and Time to Delivery are right-skewed

Guided Demonstration: Lost Sales Opportunities



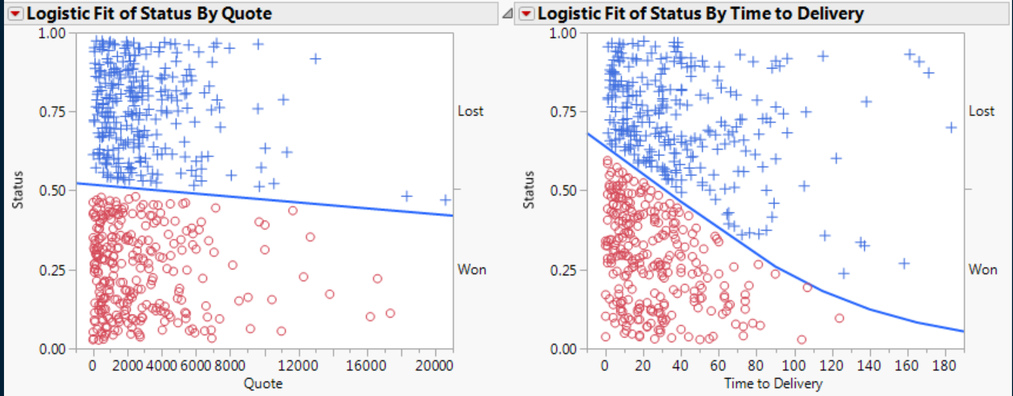
- JMP plots data in alpha-numeric order, so Lost appears first in the graph
  - ✓ Because of this, our future logistic model will predict the probability that an order is Lost rather than Won
- To change this, we adjust the Value Ordering column property for Status
- **Column Properties > Value Ordering, select Won, and click Move Up**

Guided Demonstration: Lost Sales Opportunities

Exploring relationships between variables:

- Choose **Analyze > Fit Y by X >** select **Status**, click Y, Columns, and select all other factors, click X, Factor, then OK.
- **What observations can you make regarding the mosaic plot?**
- **Let's interpret the simple logistic regressions…**

Guided Demonstration: Lost Sales Opportunities



- To aid in interpretation, each point in the logistic graph is plotted at the observed value of the predictor (the x-axis), and on either side of the logistic line (depending on whether the order was won or lost)
- The points for won orders are below the line, and lost orders are above the line
- Using a row legend for Status with colored markers eases interpretation

Guided Demonstration: Lost Sales Opportunities

Build the Model:

- Choose **Analyze > Fit Model**
- Use Status as the Y (response variable), and then select all other variables as model effects. Select Target Level to **Won**.
- Since our Y variable (dependent) is categorical, the Personality automatically changes to Nominal Logistic

Guided Demonstration: Lost Sales Opportunities

Build the Model:

- The overall model is highly significant (whole model test)
- Only Time to Delivery and Part Type are significant at the 0.05 level
- Quote is not a significant predictor of Status, so we try fitting a simpler model without Quote
- We do this because, if possible, we would like to have the simplest model that explains the variability in the data well (**principle of parsimony**)

Guided Demonstration: Lost Sales Opportunities

Build the Model:

- There are other measures of how “good” the model is (or isn’t) that we should consider
- Misclassification Rate = 0.4018
  - ❖ This means that the model-based classifications are incorrect for over 40% of the data
  - ❖ To put this in perspective, if we were to categorize orders that are won or lost at random, the misclassification rate would be around 0.50 on average
  - ❖ Our misclassification rate is somewhat better than choosing at random

Guided Demonstration: Lost Sales Opportunities

Build the Model:

- For this model, there are four possible classification outcomes
  1. A won sale is correctly classified as won
  2. A won sale is incorrectly classified as lost
  3. A lost sale is correctly classified as lost
  4. A lost sale is incorrectly classified as won
- Classifications are based on the predicted probability that an order is won. If the probability is > 0.50, then the order is classified as won. Otherwise, it is classified as lost.

Guided Demonstration: Lost Sales Opportunities

Build the Model:

- The confusion matrix provides a breakdown of the types of correct and incorrect classifications (**select Confusion Matrix from the red triangle**)
- **What is the Confusion Matrix interpretation?**

Guided Demonstration: Lost Sales Opportunities

What can we learn from this matrix?

- There are two possible predicted classes: “won" and “lost"
- The classifier made a total of 550 predictions
- Out of those 550 cases, the classifier predicted “won" 329 times, and “lost" 221 times.
- In reality, 278 observations in the sample “won”, and 272 “lost”.

Guided Demonstration: Lost Sales Opportunities

Confusion matrix basic terms, which are whole numbers (not rates):

- **True Positives (TP):** These are cases in which we predicted “won”, and the observation resulted in a “won” status.
- **True Negatives (TN):** We predicted “lost”, and the observation resulted in a “lost” status.
- **False Positives (FP):** We predicted “won”, but the observation resulted in a “lost” status. (Also known as a "Type I error.")
- **False Negatives (FN):** We predicted “lost”, but the observation resulted in a “won” status. (Also known as a "Type II error.")

Guided Demonstration: Lost Sales Opportunities

This is a list of rates that are often computed from a confusion matrix for a binary classifier:

- **Accuracy:** Overall, how often is the classifier correct?
- **Misclassification Rate:** Overall, how often is it wrong?
- **True Positive Rate:** When it's actually yes, how often does it predict yes?
- **False Positive Rate:** When it's actually no, how often does it predict yes?
- **Specificity:** When it's actually no, how often does it predict no?
- **Precision:** When it predicts yes, how often is it correct?
- **Prevalence:** How often does the yes condition actually occur in our sample?

Guided Demonstration: Lost Sales Opportunities

Build the Model:

- If we decide to use the model to predict the probability that an order is won, we would use the parameter estimates to develop a prediction equation.
- **Let's interpret the estimates:**
  - ❖ Time to Delivery: A value of – 0.0183 indicates that as the quoted time to deliver increases, the predicted probability that an order is won decreases .
  - ❖ Part Type [AM]: A value of 0.2379 for aftermarket, and a value of – 0.2379 for Part Type [OE], indicating that the probability that an order is won is higher for aftermarket than for original equipment.

From y^* to p

- If we have: $y^* = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \, age$
- Then: $p = \dfrac{\exp(\beta_0 + \beta_1 \, age)}{\exp(\beta_0 + \beta_1 \, age) + 1} = \dfrac{e^{\beta_0 + \beta_1 \, age}}{e^{\beta_0 + \beta_1 \, age} + 1}$
- Or simply put: $p = \dfrac{\exp(y^*)}{\exp(y^*) + 1} = \dfrac{e^{y^*}}{e^{y^*} + 1}$

Guided Demonstration: Lost Sales Opportunities

Build the Model:

- The logistic regression model used to predict the probability an order is “won” is a function of these parameter estimates
- We can save the probability formula to the data table using the **Save Probability Formula from the red triangle**
- The estimated logit (or log-odds) for this model is saved as a formula in a new column labeled Lin(Won)

Guided Demonstration: Lost Sales Opportunities

Build the Model:

- To calculate the predicted probability of a won order, for given values of our predictors, we simply plug the values into the formula of the logit, and then use the resulting value to calculate the Prob(Won)
- For example, if the quoted delivery time is 30 days for an aftermarket order, we substitute “30” for time for delivery and use the value 0.2379 for the aftermarket effect on the logit

Lin(Won) = 0.4856 – 0.0183*30 + 0.2379 = 0.1732

- Then, we plug this value into the Prob(Won) formula:

Prob(Won) = $1 / (1 + e^{-\text{Lin(Won)}}) = 1 / (1 + e^{-0.1732}) = 0.5431$

Guided Demonstration: Lost Sales Opportunities

Build the Model:

- JMP will calculate this for us for all observations in the data table (after saving the probability formula)
- It will also calculate Prob(Won) and Prob(Lost) **if we plug those values into a new row in the data table**
- To gain a better understanding of this model, we use the prediction profiler (in a similar context in how it was used for linear regression)
- **Return to the Nominal Logistic Fit window, and select Profiler from the red triangle**

Guided Demonstration: Lost Sales Opportunities

Implement the Model:

- The model has driven us to an interesting insight: reducing the promised delivery time could lead to an increased probability of success (i.e., winning the order)
- The next step is to determine how to put our model, and this new insight, to use

Example: Management can now look for ways to reduce the amount of time it takes to fill an order. Conducting an economic trade-off analysis that is needed to determine how much investment should be made in capacity increase, cycle time reduction, or other business process improvements.

Guided Demonstration: Lost Sales Opportunities

Assumptions of Logistic Regression:

Logistic regression does not make many of the key assumptions of linear regression and general linear models that are based on ordinary least squares algorithms – **particularly regarding linearity, normality, homoscedasticity, and measurement level**.

1. It <u>does not</u> need a linear relationship between the dependent and independent variables. Logistic regression can handle all sorts of relationships, because it applies a non-linear log transformation to the predicted odds ratio.
2. The independent variables <u>do not</u> need to be multivariate normal – although multivariate normality yields a more stable solution. Also, the error terms (the residuals) do not need to be multivariate normally distributed.
3. Homoscedasticity <u>is not</u> needed. Logistic regression does not need the residual variances to be heteroscedastic for each level of the independent variables.
4. It <u>can handle ordinal and nominal data</u> as independent variables. The independent variables do not need to be numeric.

Guided Demonstration: Lost Sales Opportunities

Assumptions of Logistic Regression:

In logistic regression, here is what you need to check for:

1. Binary logistic regression <u>requires</u> the dependent variable to be binary and ordinal logistic regression requires the dependent variable to be ordinal.
2. Since logistic regression assumes that P(Y=1) is the probability of the event occurring, it is necessary that the dependent variable is coded accordingly. That is, for a binary regression, the factor level 1 of the dependent variable should represent the desired outcome.
3. The model should be fitted correctly. Neither over fitting nor under fitting should occur. That is only the meaningful variables should be included, but also all meaningful variables should be included.

Thank You