Marketing Analytics – Week 6

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

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

SAS® Advisory Solutions Architect

Agenda

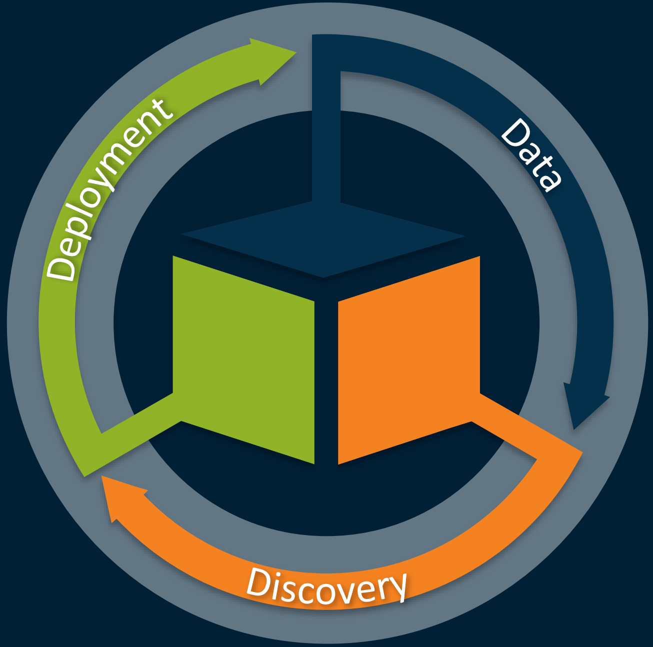
1. Introduction
2. Lecture & Demos: **Decision Trees**

Setting Expectations

- We will present important decision tree functionality within JMP but we cannot cover every feature.
- You will be introduced to decision trees, with the goal of understanding the analytical approach’s mechanics and concepts.

Getting Started

- Decision trees are used for explanatory data analysis, supervised segmentation, and predictive modeling
  - They are relatively easy-to-understand and explain, particularly to a non-technical audience



Business Applications

- Decision trees fall into two general categories
  1. **Classification trees**
  2. **Regression trees**
- If the target (dependent) variable is categorical (nominal or ordinal), then a classification tree is used to **predict the probability of a particular outcome based on a set of predictors**
- If the target (dependent) variable is continuous, then a regression tree is used to **predict the mean of the response**

Business Applications

An example of a representative business problem would be a charitable foundation studying potential donors:

Classification tree: Used to predict *whether* a person will contribute

Regression tree: Used to predict *how much* a person might contribute

Business Applications

Predictions & examples from everyday life:

1. Will a new movie be a blockbuster? How much will it earn?
2. Will a given candidate win an election?
3. What variables are most important in predicting housing prices?
4. What is the probability that a flight will be delayed?
5. What is the risk that an applicant will default on a loan?

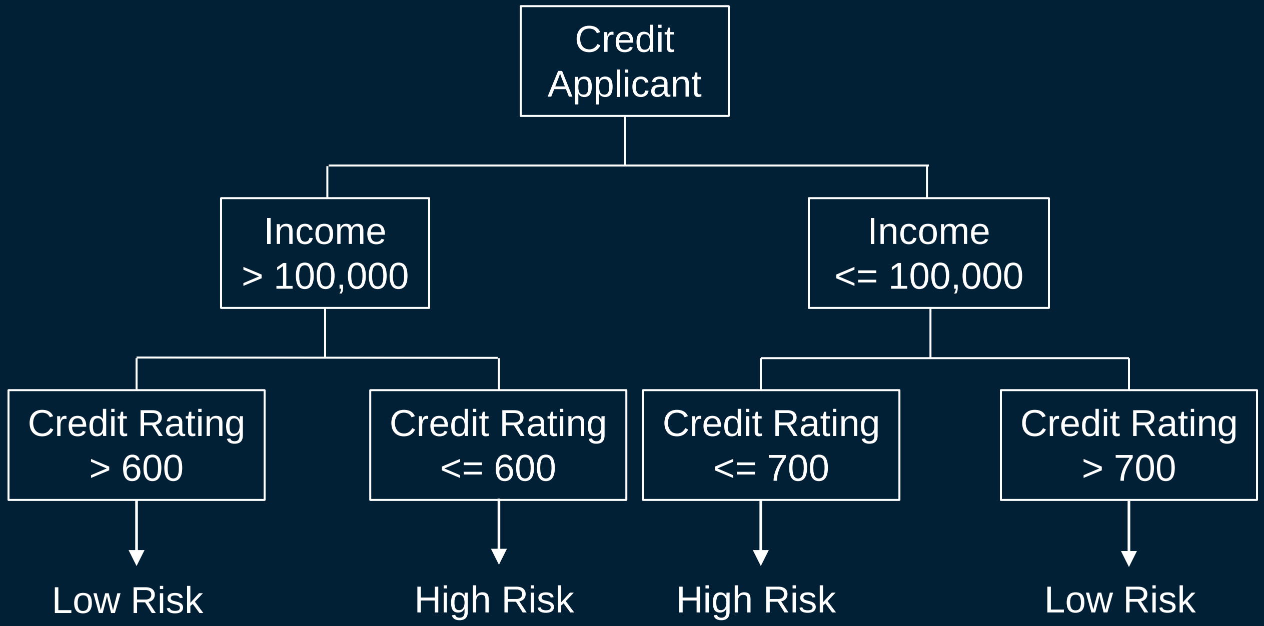
Notice anything similar to what we have studied thus far?

Previewing The End Result

Decision Tree Model:

- Consists of a set of conditional rules, based on decision thresholds
- The tree is essentially a **series of nested if-then statements** that lead to a classification or prediction
- Credit risk assessment example:
  - ❖ If income is > 100,000 and credit rating is > 600, then credit risk is low; but if credit rating is < 600, then credit risk is high
  - ❖ If income is <=100,000 and credit rating is < 700, then credit risk is high; but if credit rating is > 700, then credit risk is low

Previewing The End Result



Objectives

The objectives of recursive partitioning (or decision trees) are:

- Recursive partitioning refers to **segmenting the response into groups that are as homogeneous as possible while maximizing the difference in the response of the groups**
- Successive splits produce a **branched structure of rules and groups**, which is the model of the data
- The rules are readily **expressed in English making them easily understood**, and can also be **expressed in SQL** (to retrieve or score matching records)
- As models, the tree can be used for **classification, estimation, and prediction**, as well as being useful for data exploration and variable selection even when you plan to use a different technique to create the final model
- Different variants of the algorithm: **CART, CHAID, C5.0, C4.5**

Objectives

Diving deeper into the objectives:

- Recursive partitioning is a **machine learning method** that splits data into smaller and **smaller cells which are increasingly "pure"**
- The algorithm treats each cell **independently**
- To find a *new split*, the algorithm tests splits based on all available variables
- In doing so, the decision tree chooses the **most important variables** for the directed (or supervised) data mining task, thus explaining why they can be used for **variable selection as well as for building models**

SAS JMP Pro includes several advanced capabilities for recursive partitioning, including validation hold out sets by external column levels, Bootstrap Forest (or bagging, which averages many trees from random samples of cases), Boosted Tree (composite tree made from smaller trees), and K Nearest Neighbors

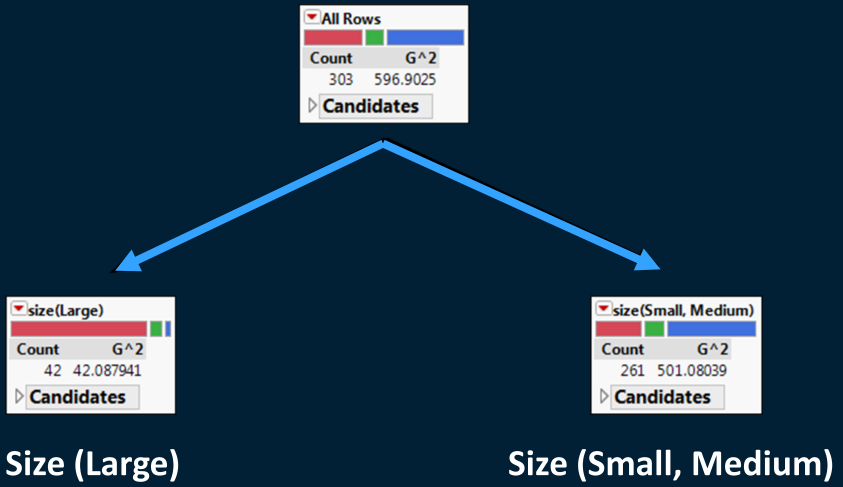
Decision Trees vs. Regression

What’s the difference?

- Regression models are **global**, and **do not do a good job of fitting data that has local characteristics**
- Trying to fit local phenomena in the input (predictor) space changes the values of the model everywhere
- Decision trees, on the other hand, **are local models**; they carve the input space into segments and produce a separate estimate for each one
- Regression models assume the relationship between a predictor and the target (dependent) is the **same everywhere**

Splitting Into New Groups

Consider an example in which you want to find predictors of the country from which a car originated. The Country levels are U.S. (red), Europe (green), and Japan (blue). The tree is fit to the data by recursive partitioning. This split produces two groups that are most different from one another and more homogenous within each group with respect to Country.



Finding The Best Split

Many different criteria may be used to evaluate potential splits:

- Comparing alternate splitting criteria **often leads to trees that look quite different from one another, but have similar performance**
- Different purity measures select different splits, but because all the measures (or predictors) are trying to capture the same idea, **the resulting models tend to behave similarly**

GEORGE MASON UNIVERSITY

Finding The Best Split

Many different criteria may be used to evaluate potential splits:

- Purity measures for evaluating splits for categorical target variables
  - Gini
  - **Entropy**
  - **Chi Square test (LogWorth)**
  - Incremental response
- Purity measures for evaluating splits for numeric (continuous) target variables:
  - Reduction in variance
  - F test

Split Metrics: Entropy & Logworth

- Candidate $\boldsymbol{G^2}$ measures the change in the ***entropy*** **(or information gain)**, which measures how disorganized a system is
  - ❖ Entropy is the change in –log likelihood (–L) over all the responses.
  - ❖ G^2 is twice the entropy (–2L).
  - ❖ **Larger $\boldsymbol{G^2}$ values are better.**

- Candidate **LogWorth** is the negative logarithm (base 10) of the *p*-value for the likelihood ratio chi-square.
  - ❖ The chi-square value measures how likely or unlikely the split is – the higher the value, the *less likely* the split is due to chance, and not being due to chance means the split is important
  - ❖ **Larger LogWorth values are better**.

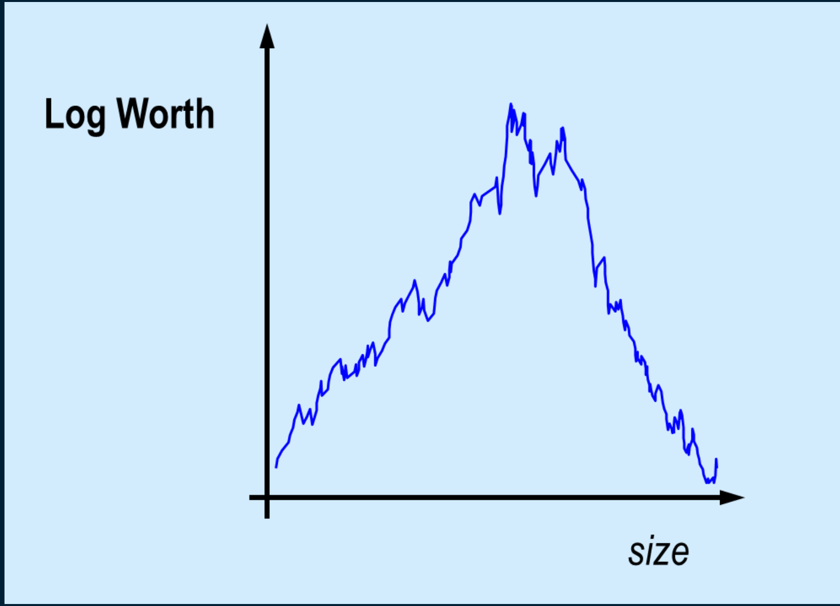
Split Metrics In SAS JMP

- The criterion in SAS JMP Pro for the best split on a categorical target variable is the **LogWorth** score

- For those of you who desire to read more on the statistical theory on splitting metrics, please review the Decision Tree supplemental content posted on Blackboard within the Discussion Board.

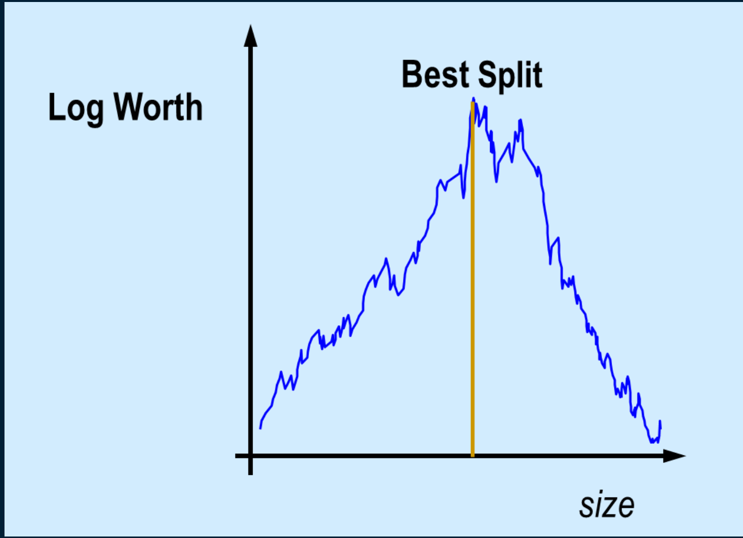
Calculate Split Metrics

The Partition platform in SAS JMP finds a set of cut points (continuous explanatory variable) or groupings (categorical explanatory variable) of X values that best predict a Y value **by exhaustively searching all possible cuts or groupings**



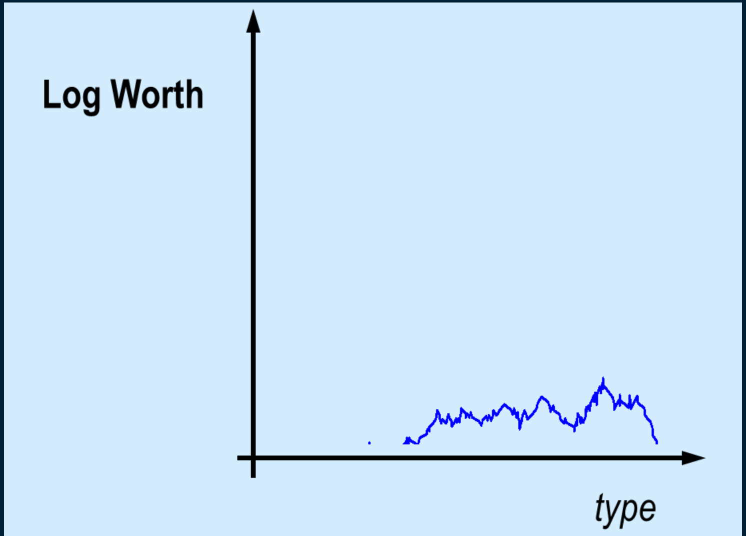
Finding The Best Split Value

After all of the possible splits have been calculated, choose the one that creates the largest metric for that variable. In this case, the categorical response was *Country*, the country where an automobile originated. The first independent variable considered was *Size* (size of the car). All of the possible splits using groupings of the levels of *Size* are considered. **The groupings that create the largest separation in the response are chosen as the best split for that variable.**



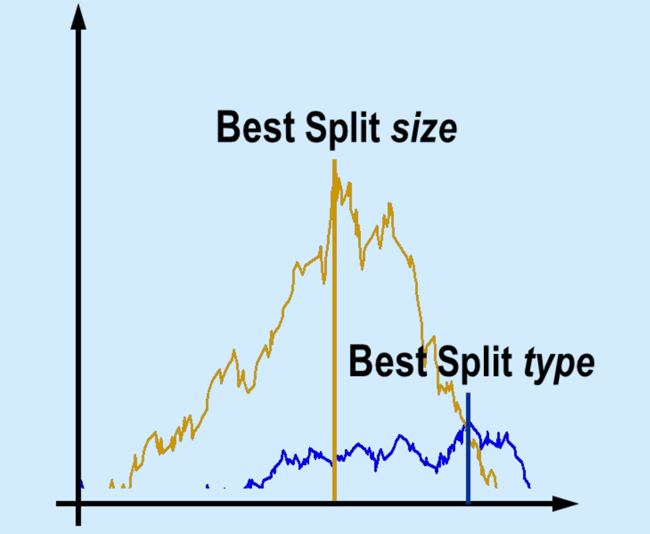
Evaluate Other Variables

The same process is repeated for each of the other independent predictor variables (Xs). For example, the variable *Type* is considered next.



Compare The Best Splits

In the case depicted on the left, the best split for *Size* results in a greater separation of the response levels than the best split for *Type*. The report shows the G^2 and LogWorth statistics for the best split for each independent variable. In the example, the best split for *Size* results in a greater separation of the responses than does any other best split.



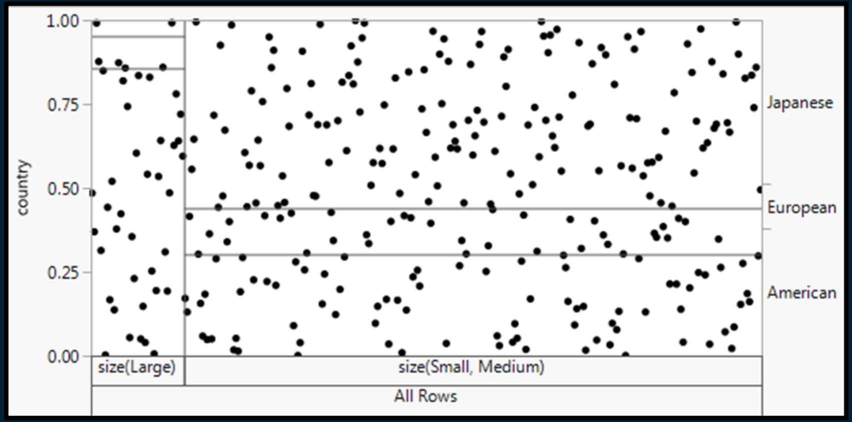
All Rows

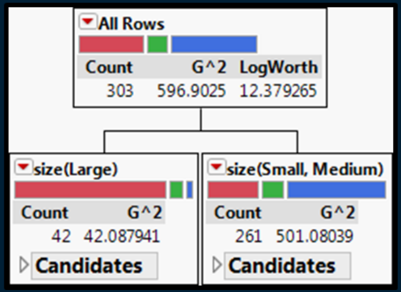
Count	G^2
303	596.9025

Candidates

Term	Candidate G^2		LogWorth	Cut Point
sex	0.31187579		0.06772297	Female
marital status	5.14007221		1.11615250	Married
age	13.53738248		1.79982282	36
size	53.73416567	*	12.37926451	Large
type	17.55577981		3.52887328	Sporty

Partition With Best Split

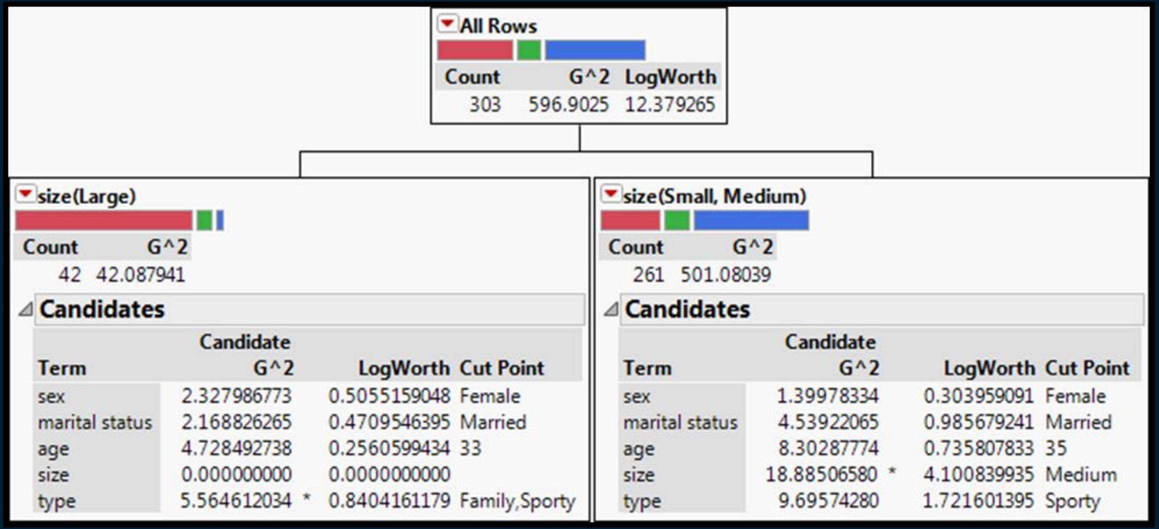




Before the first split, all the observed values are sorted based on the values of country. The vertical axis indicates the marginal probability of each value (American, European, and Japanese, in this case).

After this first split, a vertical line appears and indicates what groups are formed by the split. These groups are labeled with the variable (*Size*, in this case) and the level of the split. Within each group, the appropriate conditional probability is marked with horizontal lines.

Repeat Within Partitions



Now the process begins again: within each partition, each possible grouping into two levels is examined for the greatest separation of the responses. This is done for each independent variable (or predictor). Then, the best splits for all the independent variables are compared within each partition. Finally, the best splits for each of the two partitions are compared.

Other Considerations

Predictive modeling, under-fitting, and over-fitting:

- In predictive modeling, emphasis is put on selecting a model that is neither under-fit (too simple) nor over-fit (too complex)
  - ❖ Under-fitting is a situation where **too few splits are used and leads to bias in predictions**. The uncertainty could be reduced further.
  - ❖ Overfitting is a situation where **too many splits are used and leads to more variance in predictions.** The model incorporates features of random noise in the data, which will not be repeated again.
  - ❖ The decision tree algorithm makes it **very easy to over-fit models**
- One simple approach to preventing this from occurring is to **specify the minimum number of rows** that can be in any final tree node
  - ❖ The minimize size split setting in JMP has a default of 5, but it can be set to any integer value (number of rows) or any decimal value between 0 and 1 (proportion of rows)
- Another approach to prevent over-fitting is to use **model cross-validation**

Other Considerations

Introduction to **model cross-validation**:

- Cross-validation attempts to find the **optimum number of splits**
- The data used for the modeling is divided into groups (or partitions)
- One group is designated as the **holdout or validation set**
  - It is not used to train (fit) the model (tree)
  - It is used for predictions (as if it were future cases)
- The other group (**training set**) is used to train the model

Other Considerations

Further clarity on model cross-validation:

- The best test of a predictive model is **how well it predicts future responses.** Before new observations are in hand, you can divide your sample of cases into several groups.
- Use some of the groups to fit the model and use the rest of the groups to **independently evaluate the fit**.
- You can use cross-validation to **determine how well your partition makes predictions.**
- Cross-validation refers to withholding part of the data set from the partitioning. Then, when the tree is complete, the data that was withheld can be used to test the performance of the partition to see how well it predicts the dependent variable.

Other Considerations

Cross-validate in JMP with Validate Portion:

- A portion of the sample is randomly selected by the platform for the validation holdout data
- There is no universal rule for the size of the portion for the holdout set (rule of thumb: 25% to 50%)
- Some analysts create two holdout sets. One is used for cross-validation (model selection) and the other for testing after a model has been selected (model performance). Again, there is no specific size for the hold out data sets. The recommended sizes for each holdout set are:

Holdout Set	**Size**
Training	50%-80%
Validation	20%-50%
Testing	0%-30%

Other Considerations

Comparing Misclassification Rates On Training vs Validation Sets:

- The error rate on the validation **should be larger** than on the training
- A large difference in the error rates is a **symptom of an unstable model**
- In JMP, we have several **different diagnostic assessments to review** in helping us make the best decision

Other Considerations

Dealing with Missing Values:

- Missing values for predictor variables is a common issue in most modeling situations, and if not treated, causes the entire case to be excluded
- The default approach in the JMP partition platform (Informative Missing) is to consider that the **fact that a value is missing might actually be informative**

Other Considerations

Informative Missingness:

- For **continuous factors**, missing values are first considered on the low end of the split for each possible split evaluated
- Then they are considered to be on the high end of the split for each possible split
- The split with the best LogWorth is selected
- For **categorical factors**, a missing value code is used to indicate a missing value, and **appears as an additional level**

GEORGE MASON UNIVERSITY

Guided Demonstration & Exercise

SAS JMP Pro

Guided Demonstration: Credit Card Marketing

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

Data table description:

This file contains information on 18,000 current bank customers

Guided Demonstration: Credit Card Marketing

- Each customer receives an offer for a new credit card that has a particular type of reward program associated with it
- The type of mailing is varied, with some customers receiving a letter and others receiving postcards
- The reward type and mailing type are varied in a balanced random fashion across the prospective customer demographics
- After a set period of time, whether the individual has responded positively to the mailer and has opened a new credit card account is recorded

Guided Demonstration: Credit Card Marketing

The variables are:

- **Customer Number**: Unique identifier
- **Offer Accepted**: Did customer accept (Y) or reject (N) offer
- **Reward**: The type of reward program offered for the card
- **Mailer Type**: Letter or postcard
- **Income Level**: Low, Medium, or High
- **# of Bank Accounts Open**: How many non-credit card accounts are held by the customer
- **Overdraft Protection**: Does customer have overdraft protection on checking account(s) (Yes or No)

Guided Demonstration: Credit Card Marketing

The variables are:

- **Credit Rating**: Low, Medium, or High
- **# Credit Cards Held**: # of credit cards held at the bank
- **# Homes Owned**: # of homes owned by customer
- **Household Size**: Number of individuals in the family
- **Own Your Home**: Yes or No
- **Average Balance**: Average account balance (across all accounts over time)
- **Q1 Balance**: Average balance for Q1 in the last year
- **Q2 Balance**: Average balance for Q2 in the last year
- **Q3 Balance**: Average balance for Q3 in the last year
- **Q4 Balance**: Average balance for Q4 in the last year

Guided Demonstration: Credit Card Marketing

Define the Problem:

- We want to build a model that will provide insight into why some bank customers accept credit card offers
- Because the response is categorical (Yes or No) and we have a large number of potential predictor variables, we use the Partition (Decision Tree) Platform to build a classification tree for *Offer Accepted*

Guided Demonstration: Credit Card Marketing

Prepare for Analysis:

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

Guided Demonstration: Credit Card Marketing

To leverage the Columns Viewer:

- Go to **Cols > Column Viewer** and then select all variables, click Y, Columns, and OK.
- **After reviewing the distributions, what are some of your initial observations?**
  - ✓ N Categories, we see that each of our categorical variables has either two or three levels (low cardinality)
  - ✓ N Missing indicates with have 24 missing observations for each of the balance columns
  - ✓ The other statistics provide an idea of the centering, spread, and shapes of the continuous distributions

Guided Demonstration: Credit Card Marketing

Exploring one variable at a time:

- You can get to Distribution through the Analyze menu in the Data Table.
- 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?**
  - ✓ 5.68% of the 18,000 offers were accepted

GEORGE MASON UNIVERSITY

Guided Demonstration: Credit Card Marketing

Frequencies

Level	Count	Prob
Yes	1023	0.05683
No	16977	0.94317
Total	18000	1.00000

N Missing 0
2 Levels

Frequencies

Level	Count	Prob
Air Miles	6061	0.33672
Cash Back	5999	0.33328
Points	5940	0.33000
Total	18000	1.00000

N Missing 0
3 Levels

Frequencies

Level	Count	Prob
Letter	8853	0.49183
Postcard	9147	0.50817
Total	18000	1.00000

N Missing 0
2 Levels

- **Select the Yes level in Offer Accepted**, and examine distributions
- Our two experimental variables are Reward and Mailer Type
- **What observations can you make?**
- **What do you notice about Credit Rating and Income Level? (Hint: Nominal or Ordinal)**

Guided Demonstration: Credit Card Marketing

Exploring relationships between variables:

- Choose **Analyze > Fit Y by X >** select **Offer**, click Y, Columns, and select all other factors, click X, Factor, then OK.
- **What relationships are statistically significant, and potentially indicative of influencing the dependent variable of interest?**

Guided Demonstration: Credit Card Marketing

Random-seed add-in:

- JMP File Exchange: https://community.jmp.com/docs/DOC-6601
- Download it
- Set the random seed to **123**
- This will ensure we get the same results when we launch the partition platform (decision tree) and perform the modeling analysis

Guided Demonstration: Credit Card Marketing

Build the Model:

- Choose **Analyze > Predictive Modeling > Partition**
- Use Offer as the Y (response variable), and then select all other variables as X, Factor
- For Validation, assign a value of .30 for the Validation portion
- Click Okay

Guided Demonstration: Credit Card Marketing

Build the Model:

- To remove points from the graph, click on the red triangle and select **Display Options > Show Points**
- To show response rates in the tree nodes, select **Display Options > Show Split Count** from the top red triangle
- Open the Candidates drop down – **what will be the first split?**
- Close the Candidates menu, and **click Split three times**

Guided Demonstration: Credit Card Marketing

Build the Model:

- After each split, the model R-square updates accordingly – remember, R-square is a measure of how much variability in the response is being explained by the model
- Without a validation set, we can continue to split until the minimum split size is achieved in each node (JMP default is 5).
- However, additional splits are not necessarily beneficial and lead to more complex and potentially over-fit models

Guided Demonstration: Credit Card Marketing

Build the Model:

- Since we have a validation portion, **we click Go to automate the tree-building process**
- The tree with the maximum validation R-Square has 15 splits (results may vary if you didn’t use the random seed add-in)
- To summarize which variables are involved with the 15 splits, we turn on the **Column Contributions** (from the red triangle)
- This table indicates which variables are most important in terms of the overall contribution to the model

Guided Demonstration: Credit Card Marketing

Build the Model:

- Click **Show Fit Details** from the red triangle
- **What's different about the confusion matrix?**
- Remember, there are four possible outcomes in our classification:
  1. An accepted offer is correctly classified as an accepted offer
  2. An accepted offer is misclassified as not accepted
  3. An offer that was not accepted is correctly classified as not accepted
  4. An offer that was not accepted is misclassified as accepted
- **Do you see the problem?**
  1. False Negatives (FN): We predicted "lost", but the observation resulted in a "won" status. (Also known as a "Type II error.")
  2. True Positive Rate: When it's actually yes, how often does it predict yes?

Guided Demonstration: Credit Card Marketing

Build the Model:

- One important observation of the confusion matrix is that there were **few cases where the model predicted that the offer would be accepted**
- When the target (dependent) variable **is unbalanced** (there are far more observations in one level than the other), the model that is fit will **usually result in probabilities that are small for the under-represented category**
- In this example, the overall rate of Yes (offer accepted) is 5.68%, which is very close to the misclassification rate for the model
- However, when we review the Leaf Report (turn on under Red Triangle), we see that there are terminal branches in the tree that have much richer concentrations of Offer Accepted = Yes

Guided Demonstration: Credit Card Marketing

Build the Model:

- The fitted model has probabilities of Offer Accepted = Yes in the range of [0.0044, 0.6738]
- Recall that when JMP classifies rows with the model, it uses a default of Prob > 0.5 to make the classification decision
- In this example, only one of the predicted probabilities of Yes is > 0.5, and this one node has only 11 observations: 8 yes and 3 no under Response Counts
- All other rows are classified as Offer Accepted = No

Guided Demonstration: Credit Card Marketing

Build the Model:

- Receiver Operator Characteristic (ROC) Curve
  - ❖ Two additional measures of accuracy used when building classification models are Sensitivity and (1-Specificity)
  - ❖ Sensitivity is the true positive rate, and in our example, this is the **ability of the model to correctly classify Offer Accepted as Yes**
  - ❖ (1 – Specificity) is the false positive rate, and in our example, this occurs **when an Offer was not accepted, but was classified as Yes (accepted)**
  - ❖ A ROC graph is a two-dimensional plot of a classifier with false positive rate on the x axis against true positive rate on the y axis. As such, a ROC graph depicts relative trade-offs that a classifier makes between benefits (true positives) and costs (false positives).
  - ❖ **The ROC curve can be triggered from the red triangle**

GEORGE MASON UNIVERSITY

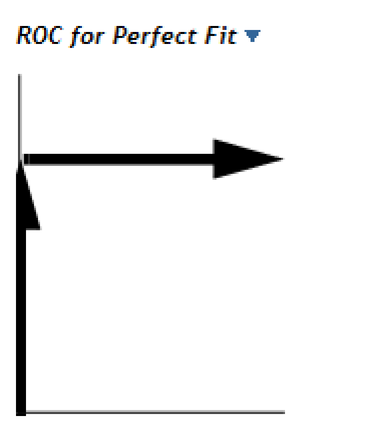
Guided Demonstration: Credit Card Marketing

Build the Model:

- Receiver Operator Characteristic (ROC) Curve
  - ❖ Conceptually, what the ROC Curve is measuring is the ability of the predicted probability formulas to rank an observation
  - ❖ **Here, we focus on the Yes outcome for the Offer Accepted response variable**
  - ❖ **Save the probability formula (red triangle) to the data table, and then sort the data table from the highest to lowest probability**
  - ❖ If this probability model can correctly classify the outcomes for Offer Accepted, we would expect to see more Yes response values at the top (where the probability for Yes is highest), and at the bottom of the sorted table, we would expect to see more No than Yes response values

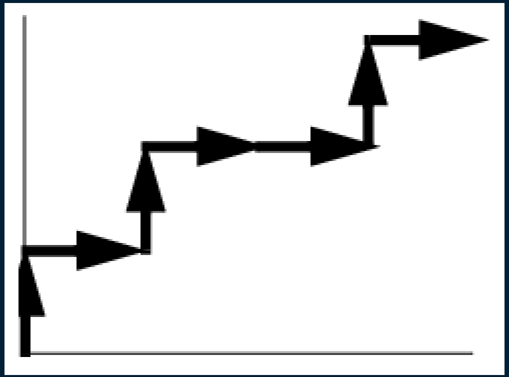
Guided Demonstration: Credit Card Marketing

If the model perfectly rank-orders the response values, then the sorted data has all the targeted values first, followed by all the other values. The curve moves all the way to the top before it moves at all to the right.



Guided Demonstration: Credit Card Marketing

If the model does not predict well, it wanders more or less diagonally from the bottom left to top right. In practice, the curve lifts off the diagonal. The area under the curve is the indicator of the goodness of fit; a value of 1 indicates a perfect fit.



Guided Demonstration: Credit Card Marketing

Build the Model:

- The Lift Curve
  - ❖ Lift is a measure of how much "richness" in the response we achieve by applying a classification rule to the data
  - ❖ **A Lift Curve (under the red triangle) plots the Lift on the Y-axis against the Portion (or decile) on the X-axis**
  - ❖ 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 top 10% of the rows corresponds to a portion of 0.1, and the top 20% of rows corresponds to a portion of 0.2, and so on…
  - ❖ **The lift for Offer Accepted = Yes, for a given portion, is simply the portion of Yes responses in this portion, divided by the overall proportion of Yes in the entire data table**

Guided Demonstration: Credit Card Marketing

Build the Model:

- The Lift Curve
  - ❖ **Make an interpretation of Model Lift at Portion = 0.15**
  - ❖ The lift at Portion = 0.15 is roughly 2.5
  - ❖ This means that rows in the data table that correspond to the top 15% of the model's predicted probabilities, the number of actual Yes outcomes is 2.5 times higher than we would expect if we had chosen 15% of the rows in the data set at random
  - ❖ If the model is not sorting the data well, then the lift will hover around 1.0 across all portion values

Guided Demonstration: Credit Card Marketing

Build the Model:

- The Lift Curve
  - ❖ Particularly useful if the overall predicted probabilities are lower than 0.5 for the outcome we wish to predict
  - ❖ Even though, in this example, the majority of the predicted probabilities of Offer Accepted = Yes were less than 0.2, the lift curve indicates that there are a threshold of values that we could use with this model to create a classifier rule that will be better than just guessing alone
  - ❖ This rule can be used to identify portions of our data that have a much richer number of customers that are likely to accept an offer

Guided Demonstration: Credit Card Marketing

Summary:

- For **categorical response models**, the misclassification rate, the confusion matrix, the ROC curve, and the lift curve all provide measures of **model accuracy, and each of these should be used to assess the quality of the prediction model**
- This model was created for **explanatory** rather than predictive purposes
- Based on this model, the four most important factors are: **Credit Rating, Mailer Type, Reward, and Income Level**
- How can this information be used? Your homework will address this.

Guided Demonstration: Credit Card Marketing

Alternate Cut-off Confusion Matrix Add-In

- JMP File Exchange: https://community.jmp.com/docs/DOC-6901
- Download it
- Designed to allow the user to specify a cut-off value or a range of values to generate a new confusion matrix for a binary response variable.

Guided Demonstration: Credit Card Marketing

Which is best? Well, it matters…

Confusion Matrix for Cut-off = 0.05

Target: Offer Accepted
Predictor: Partition

Confusion Rates (0.05)

	Predicted	
	No	Yes
Offer Accepted	Row %	Row %
No	54.44%	45.56%
Yes	18.48%	81.52%

Confusion Matrix for Cut-off = 0.1

Target: Offer Accepted
Predictor: Partition

Confusion Rates (0.1)

	Predicted	
	No	Yes
Offer Accepted	Row %	Row %
No	78.36%	21.64%
Yes	45.55%	54.45%

Confusion Matrix for Cut-off = 0.15

Target: Offer Accepted
Predictor: Partition

Confusion Rates (0.15)

	Predicted	
	No	Yes
Offer Accepted	Row %	Row %
No	96.57%	3.43%
Yes	83.68%	16.32%

Confusion Matrix for Cut-off = 0.25

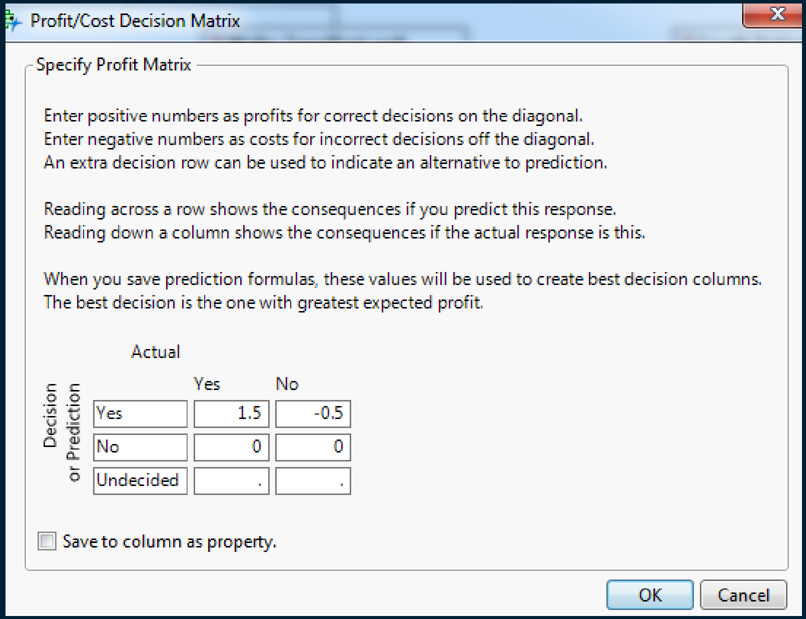
Target: Offer Accepted
Predictor: Partition

Confusion Rates (0.25)

	Predicted	
	No	Yes
Offer Accepted	Row %	Row %
No	99.95%	0.05%
Yes	99.02%	0.98%

Guided Demonstration: Credit Card Marketing

Something to look forward to…



Guided Demonstration: Credit Card Marketing

Let’s Walk Through An Example:

- Suppose we consider sending an offer to **1,000 people, 1% of whom respond on average (i.e. 10 responders)**
- Naively classifying everyone as a non-responder has a **misclassification (or error) rate of 1%, but would yield a profit of $0**
- Using a data mining approach, suppose we can produce this:

	Predict Class 0	**Predict Class 1**
Actual 0	970	20
Actual 1	2	8

* These classifications have a misclassification rate of 100 x (20+2)/1000 = **2.2%**

Guided Demonstration: Credit Card Marketing

Let’s Walk Through An Example:

- Now suppose that a **profit from one instance of Class 1 is $10 and the cost of sending the offer is $1**
- Classifying everyone as a non-responder may have a misclassification rate of 1%, but **yields a profit of $0.**
- Using the data mining approach, despite the higher misclassification rate of 2.2%, **yields a profit of $60.**

	Predict Class 0	**Predict Class 1**
Actual 0	$0	-$20
Actual 1	$0	$80

* Nothing is sent to the predicted non-responders (Class 0), so there are no costs or sales to that column

Guided Demonstration: Credit Card Marketing

Let's Walk Through An Example:

- Looking at terms of cost, when everyone is classified as a non-responder, there are no costs of sending the offer; **the only costs are the opportunity costs of failing to make sales to the ten responders (i.e. $100)**
- The cost (**actual costs of sending the offer, plus the opportunity costs of missed sales**) of using the data mining routine to select people to send the offer to is only $48, as follows:

	Predict Class 0	**Predict Class 1**
Actual 0	$0	$20
Actual 1	$20	$8

Guided Demonstration: Credit Card Marketing

Let's Walk Through An Example:

- However, this **does not improve** the actual classifications
- A better method is to **change the classification rules, hence the misclassification rates**, as previously discussed with the alternate cut-off confusion matrix example
- A popular performance measure that includes costs is the ***average misclassification cost***, which measures the average cost of misclassification per classified observation (we are looking for a classifier that **<u>minimizes</u> this quantity**)

Guided Demonstration: Credit Card Marketing

Let’s Do This In JMP – Specify Profit Matrix:

- Enables you to **specify profit or costs associated with correct or incorrect classification decisions**. Only available for categorical responses.
- When you define costs using the **Specify Profit Matrix option** and then select **Show Fit Details**, a **Decision Matrix report appears.**
- When you specify a profit matrix and select Save Columns > Save Prediction Formula from the report’s red triangle menu, **additional columns with formulas are saved to the data table**:
  - ✓ **Profit for <level>:** For each level of the response, a column gives the expected profit for classifying each observation into that level.
  - ✓ **Most Profitable Prediction for <column name>:** For each observation, gives the level of the response with the highest expected profit.
  - ✓ **Expected Profit for <column name>:** For each observation, gives the expected profit for the classification defined by the Most Profitable Prediction column.
  - ✓ **Actual Profit for <column name>:** For each observation, gives the actual profit for classifying that observation into the level specified by the Most Profitable Prediction column.

GEORGE MASON UNIVERSITY

Thank You