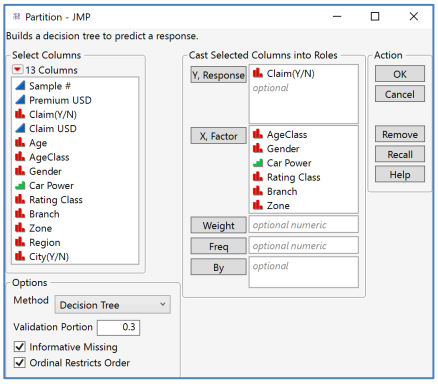
Classification Trees (Partition)

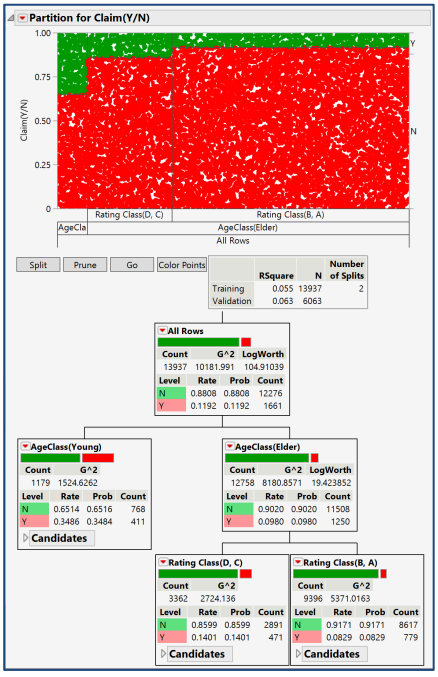
Use to build a partition-based model (Decision Tree) that identifies the most important factors that predict a categorical outcome (classify) and use the resulting tree to make predictions for new observations.

Classification Trees

Auto Raw Data.jmp (Help > Sample Data Folder)



1. From an open JMP® table, select **Analyze > Predictive Modeling > Partition.**
2. Select a nominal or ordinal response variable from **Select Columns** and click **Y, Response**.
3. Select explanatory variables and click **X, Factor**.
4. If desired, enter the **Validation Portion** or select a validation column and click **Validation** (**JMP Pro only**). A randomly selected validation set of 30% was used in this illustration. Thus 70% for training.
5. In **JMP Pro only**, select the tree **Method**: **Decision Tree** (the default), **Bootstrap Forest**, **Boosted Tree**, **K Nearest Neighbors** or **Naive Bayes**. JMP Pro also allows you to specify a validation column.
6. Click **OK.**
   JMP initially displays a graph showing the proportion of observations in each response level.



7. Click the **Split** button. The observations will be split into two nodes, or leaves. The graph will update to reflect the split and a tree diagram describing the split in more detail will be created.

   Note: Click on the **top red triangle** and select **Display Options > Show Split Counts** to show **Counts, Rates** (proportion of observations) and **Probs** (predicted probabilities) in each leaf.

8. Click **Split** to continue additional splits. Click **Prune** to remove a split. If a validation set is used, click **Go** to perform automatic splitting and pruning optimizing the fit on the validation data. Here only two splits were performed.

Interpretation for the first two splits (Response is Claim (Y/N)):

- There are 1,179 obs in the left leaf, corresponding to AgeClass(Young). 768 of those (65.1%) are Claim(Y/N) = N and 411 (34.8%) are Claim(Y/N) = Y.
- There are 12,758 obs in the right leaf, corresponding to AgeClass(Elder). The response rate is 90.2% for N and 9.8% for Y in that node.
- For the 12,758 observations in the AgeClass(Elder), the second split, based on the Rating Class variable, is at Rating Class (D,C) and Rating Class (B, A).

Notes:

For additional options, such as **Column Contributions**, **ROC** and **Lift Curves**, click the **top red triangle**. Other options, such as **Save Prediction Formula** are available from the **top red triangle > Save Columns**. Select **Decision Threshold** under the **top red triangle** to display correct and incorrect classification rates for the model including the ability to evaluate those rates under different cutoff values. The default is 50%.

Visit **Predictive and Specialized Models > Partition Models** in **JMP Help** to learn more.