

Creating a Validation Column (Holdout Sample)

■ Make Validation Column

Specify how to allocate rows to Training, Validation and Test sets.

0.6

0.2

0.2

Formula Random Formula column with a random function.

Fixed Random Column with no formula. Seed:

Stratified Random Column with sets that are balanced across levels of selected columns.

Cutpoint Column with holdback sets based on time series cutpoints

Grouped Random Column with each level of the grouping column assigned entirely to

after the model is chosen.

Enter either rates or counts.

Validation Set Test Set

Cancel Help

Initialize Data

Random

Random Integer Value
Random Uniform

Random Indicator

New Column Name Validation

Choose a method to create the holdback sets:

Training Set

A validation column divides the rows of the data table into a training set to estimate the model; a

alidation set to help choose a model that predicts well; and sometimes a test set to check prediction

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

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

0.2

Creating a Validation Column (Train, Validate, Test) in JMP Pro

- From an open JMP data table, select Analyze >
 Predictive Modeling > Make Validation Column.
- In the resulting window, enter values (counts or proportions) indicating how the data will be allocated to the training, validation and test sets.

The observations in the Training Set will be used to train the model. The observations in the Validation Set will be used to validate the model, and the observations in the Test set will be used to test the model.

Note that for some algorithms, the validation set is used for model selection and to control overfitting.

 Select a method for partitioning the data. The available methods, including Fixed Random and Stratified Random, are described in the dialog window.

A new column is created, populated with the values Train, Validation, and Test, in the proportions (or counts) specified above. Note that the resulting proportions will be approximate equal to those specified.

For reproducibility, enter a value greater than 1 in the Seed field and use Fixed Random.

Creating a Validation Column – Another Way

- 1. From an open JMP data table, select **New Column** from the **Cols** menu.
- 2. In the resulting **New Column** window, change the **Column Name** to **Validation**.
- 3. Next to Initialize Data, click on the arrow and select Random.
- 4. Select **Random Indicator**. By default, the new column will contain 80% 0s (train), 20% 1s (validate) and 0% (test)s.
- 5. Change the default proportions as desired.
- 6. To display the labels Train, Validate and Test rather than 0, 1 and 2, select **Value Labels** under **Column Properties** and enter the values and the desired labels, then click **Add**.
- 7. Click **Apply** to view the new column in the data table (to verify that the column will be created as desired). Then click **OK** to create the column.

Notes: For information on creating validation columns and validation options available in the different modeling platforms and model validation in JMP Pro, see the books *Fitting Linear Models* or *Predictive and Specialized Models* (under Help > Books) or search for "validation" in the JMP Help.