

Creating a Validation Column (Holdout Sample)

Use to subset the data into a set used to build a model (training) and a set used to evaluate a model's predictive performance (validation). If multiple models are fit, the best performer on the validation data is often the one chosen. At times, a third set is used (test) to evaluate the chosen model's predictive performance on new data. This is considered to be a more accurate means to evaluate a model's future performance as the test set was neither used in the model building nor selection process.

Using a validation column is particularly useful in building models that have a tendency to overfit the data. Some modeling platforms in JMP provide the option to specify the validation portion when fitting the model and thus creating a validation column is not necessary.

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

1. From an open JMP data table, select **Analyze > Predictive Modeling > Make Validation Column**.
2. Stratification, Grouping, and Cutpoint columns can be used to tailor the partitioning. If a simple validation column is desired, Click **OK**.
3. In the resulting window, enter values (counts or proportions) indicating how the data will be allocated to the training, validation and test sets. Choose a Random Seed in order to reproduce the same random assignment if desired. Click **OK**.

A new column is created, populated with the values 0, 1, and 2 in the proportions (or counts) specified.

Make Validation Column
Random Validation Column

Randomly partitions the rows of the data table into a training set to estimate the model, a validation set to choose a model by comparing the predictive performance of several candidate models, and an optional test set to independently evaluate performance after the model is chosen.

Specify rates or relative rates

	Adjusted Rates	Row Counts
Training Set	0.6	3576
Validation Set	0.3	1788
Test Set	0.1	596
Excluded Rows		0
Total Rows		5960

Options

New Column Name: Validation

Validation Column Type: Fixed

Random Seed: []

Go, Cancel, Help

- 3,576 (60%) of the observations (Training set) will be used to build (train) the model.
- 1,758 (30%) of the observations (Validation set) will be used to validate and select the best model.
- 596 (10%) of the observations (Test set) will be used to test the chosen model's performance on new data.

Creating a Validation Column in JMP

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**. Type in the desired proportions. Here we chose 50% 0s (train), 30% 1s (validate) and 20% 2s (test).
5. To display the labels Train, Validate and Test rather than 0, 1 and 2, right click on the column and select **Column Properties > Value Labels**. Enter the **value** and the desired **label** and click **Add** one value at a time.
6. 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.

Initialize Data: Random

Random Integer
 Random Uniform
 Random Normal
 Random Indicator

Value	Proportion
0	0.5
1	0.3
2	0.2

Column Properties

Value Labels

If a column has value labels, and Use Value Labels is checked, the labels are displayed wherever the column data are displayed.

0 = Train [Add]
1 = Validate [Change]
2 [Remove]

Allow Ranges

Value: 2
Label: Test

Use Value Labels

Visit **Predictive and Specialized Models > Make Validation Column** in **JMP Help** to learn more.