## **jm**p

## **Model Comparison and Selection**

Use this platform to summarize and compare the performance of multiple statistical models that have been fit to data. For details on fitting different statistical models, see the appropriate guides.

## Model Comparison – Continuous Response

Example: We use the *Body Fat.jmp* data to predict Percent body fat. Formulas for several models, saved to the data table, are grouped under **Prediction Formulas** in the column panel. The models were chosen based on the model's performance on validation data.

- from the data table,select Analyze > Predictive Modeling > Model Comparison.
- 2. Select the prediction formula columns from **Select Columns**, and click **Y,Predictors**. If no columns are selected JMP will use all saved prediction formulas in the data table.
- 3. Select **Validation** from the **Select Columns** list, and click **By**. Then, click **OK**.

JMP provides the following statistics for comparing the performance of the different models (statistics for the Validation set are displayed):

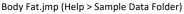
- RSquare (higher is better)
- RASE: Root Average Squared Error (like RMSE, but can be compared for competing models)
- AAE: Average Absolute Error (a measure of the magnitude of the errors)

## Model Comparison – Categorical Response

Example: In the *Equity.jmp* data, we predict Bad as a function of several predictors. In this example, the data was partitioned into Training, Validation and Test sets.

Create the predictive models of your choice (with Validation), then follow steps 1-3 above to generate model statistics for each value of the Validation column.

 The misclassification rates and other measures of performance are provided.



Model Comparison - JMP Pro		-	□ ×
Compares performance across models using prediction	on formula columns.		
Select Columns	Cast Selected Columns into Roles		Action
26 Columns	Y, Predictors 🚄 Pred Body Fat - 2nd Ord	ler Stepwise	ОК
Percent body fat	A Pred Body Fat - Vble R		Cancel
Age (years)	Pred Body Fat - Stepwise		cancer
Weight (lbs)	Pred Body Fat - Partition		
Height (inches)	Pred Body Fat - PLS		Remove
Neck circumference (cm)	Pred Body Fat - Bootstra	p Forest	Recall
Chest circumference (cm)	Pred Body Fat - Boosted		Recall
Abdomen circumference (cm)	Pred Body Fat - Neural (		Help
Hip circumference (cm)	📕 Pred Body Fat - Neural (		
Thigh circumference (cm)	Pred Body Fat - Boosted		
Knee circumference (cm)	Pred Body Fat - Ensemble	le Model	
Ankle circumference (cm)	optional		
Biceps (extended) circumference (cm)	Group optional		
Forearm circumference (cm)			
Wrist circumference (cm)			
⊿ Prediction Formulas (11/0)			
Pred Body Fat - 2nd Order Stepwise	Weight optional numeric		
Pred Body Fat - Variable Reduction PCA	weight		
Pred Body Fat - Stepwise	Freq optional numeric		
Pred Body Fat - Partition Pred Body Fat - PLS	By <b>L</b> Validation		
	optional		
Pred Body Fat - Bootstrap Forest Pred Body Fat - Boosted Tree	If you choose no Predictor columns, it will	find and	
Pred Body Fat - Boosted Tree Pred Body Fat - Neural (3,0,0)	analyze all predictors.		
Pred Body Fat - Neural (3,0,0) Pred Body Fat - Neural (4,0,0),(8,0,0)			
Pred Body Fat - Neural (4,0,0),(8,0,0)			
Pred Body Fat - Boosted Neural Pred Body Fat - Ensemble Model			
Validation			
- Yongation			

Predictors						
Measures of Fit for Percent body fat						
Predictor	Creator	.2.4.6.8	RSquare	RASE	AAE	Free
Pred Body Fat - 2nd Order Stepwise	Fit Least Squares		0.7156	4.6405	3.8349	73
Pred Body Fat - Variable Reduction PCA	Fit Least Squares		0.5209	6.0233	4.7128	73
Pred Body Fat - Stepwise	Fit Least Squares		0.7358	4.4725	3.7222	73
Pred Body Fat - Partition	Partition		0.6435	5.1956	4.1878	7.
Pred Body Fat - PLS	Partial Least Squares		0.6984	4.7789	4.0219	7.
Pred Body Fat - Bootstrap Forest	Bootstrap Forest		0.5874	5.5894	4.3478	7.
Pred Body Fat - Boosted Tree	Boosted Tree		0.6034	5.4804	4.4486	7.
Pred Body Fat - Neural (3,0,0)	Neural		0.6880	4.8609	3.7132	7.
Pred Body Fat - Neural (4,0,0),(8,0,0)	Neural		0.6186	5.3742	4.1153	7.
Pred Body Fat - Boosted Neural	Neural		0.7275	4.5422	3.4976	7.
Pred Body Fat - Ensemble Model	Neural		0.7063	4.7157	3.7557	7.

Equity.jmp (Help > Sample Data Folder)

Predictors								
Measures of Fit for BAD								
		Entropy	Generalized			Mean	Misclassification	
Creator	.2.4.6.8	RSquare	RSquare	Mean -Log p	RMSE	Abs Dev	Rate	
Partition		0.4042	0.5357		0.3082	0.1863	0.1166	11
Neural		0.5026	0.6343	0.2687		0.1567	0.1049	11
Fit Nominal Logistic		0.4626	0.5955	0.2903		0.1702	0.1174	1
Fit Generalized Two Stage Forward Selection		0.5478	0.6763		0.2698	0.1478	0.0931	1
Fit Generalized Forward Selection		0.5545	0.6823	0.2406	0.2681	0.1466	0.0914	1
Model Comparison Validation=	Test							
Predictors								
Measures of Fit for BAD								
		Entropy	Generalized			Mean	Misclassification	
Creator	.2.4.6.8	RSquare	RSquare	Mean -Log p	RMSE	Abs Dev	Rate	
Partition		0.3951	0.5177	0.3059	0.2984	0.1799	0.1107	1
Neural		0.5161	0.6391	0.2447	0.2663	0.1460	0.0940	1
Fit Nominal Logistic		0.4754	0.6000		0.2782	0.1584	0.0931	1
Fit Generalized Two Stage Forward Selection		0.5491	0.6697		0.2569	0.1367	0.0856	1
Fit Generalized Forward Selection		0.5449	0.6659	0 0200	0.2582	0.1369	0.0889	11

- Select **Decision Threshold** under the red triangle to launch an interactive tool to compare correct/incorrect classification rates using different cutoff criteria.
- **ROC curves**, **lift curves**, **model averaging** and other options are available from the red triangles.
- The **profiler** can also be shown for each model to provide a detailed view of the effect each of the factors has on the outcome and compare the predictions for various factor settings.

Model Comparison is also available from the Formula Depot.

Visit Predictive and Specialized Models > Model Comparison in JMP Help to learn more.