GenIQ, Ltd
Genetic Modeling & Data Mining Software **BRUCE RATNER, PhD**
574 Flanders Drive North Woodmere, NY 11581 br@dmstat1.com
516.791.3544 fax 516.791.5075 www.GenIQModel.com

Minified Primer on the Confusion Matrix

Bruce Ratner, Ph.D.

Assessment of the predictiveness of a binary classification model involves the 2x2 table of actual versus predicted outcomes, also known as the *confusion matrix*, in Table 1. I focus the database-marketing logistic regression model; accordingly, I use the binary (dichotomous) target variable Response, which assumes 0 and 1. The treatment of this topic can easily be extended to a polychotomous (multinomial) target variable.

Table 1: Confusion Matrix
Response Classification Model Results

		Predicted		Total
		0	1	
Actual	1	a	b	a + b
	0	c	d	c + d
Total		a + c	b + d	a + b + c + d

The matrix entries, based on the classification model with Prob (Response = 1) = b / (a + b +c + d):

- ***a*** is the number of **incorrect** predictions - actual Response = 1 and predicted Response = 0
- ***b*** is the number of **correct** predictions - actual Response = 1 and predicted Response = 1
- ***c*** is the number of **correct** predictions - actual Response = 0 and predicted Response = 0
- ***d*** is the number of **incorrect** predictions - actual Response = 0 and predicted Response = 1

There are six standard terms, drawn from the confusion matrix, that partially render the matrix *confusing*. I present three terms, which are important for evaluation of predictive power of a database-marketing logistic regression model. The three remaining terms are found in any standard text on statistical classification models.

- *Accuracy* (A) is the proportion of the total number of correct predictions:

$$A = (b + c) / (a + b+ c +d)$$

- *Recall* (R) is the proportion of 1's that are correctly predicted among the actual 1s:

$$R = b / (a + b)$$

- *Precision* (P) is the proportion of the predicted 1s that are correct among the predicted 1s:

$$P = b / (b+ d)$$