

---

# Genetically Data-mined Variables

## *~ Illustration ~*

Bruce Ratner PhD  
**DM STAT-1** CONSULTING  
1 800 DM STAT-1 [www.DMSTAT1.com](http://www.DMSTAT1.com)

GenIQ®

## ~ The GenIQ Model ~

The **GenIQ Model**© is a machine learning alternative model to the statistical ordinary least squares and logistic regression models. **GenIQ** let's the data define the model – automatically data mines for new variables, performs variable selection, and then specifies the model equation – so as to "optimize the decile table," to fill the upper deciles with as much profit/many responses as possible.

In this illustration, GenIQs optimizing of the deciles is equivalent to predicting the best possible ranking (permitted by the data) of the target variable based on the GenIQ Model score **GenIQvar**.

**OBJECTIVE #1:** To build a **binary response** model with data from Table 1.

I built a logistic regression model (LRM) for predicting the target variable Response using two predictor variables, XX1 and XX2.

The LRM equation is:

$$\text{Logit of Response (=Yes)} = 0.1978 - 0.0328 * \text{XX1} + 0.0308 * \text{XX2}$$

Table 1

<u>ID</u>	<u>XX1</u>	<u>XX2</u>	<u>Response</u>
1	31	38	Yes
2	12	30	No
3	35	21	Yes
4	23	30	No
5	45	37	No
6	16	13	No
7	45	5	Yes
8	30	30	Yes
9	6	10	Yes
10	30	10	No

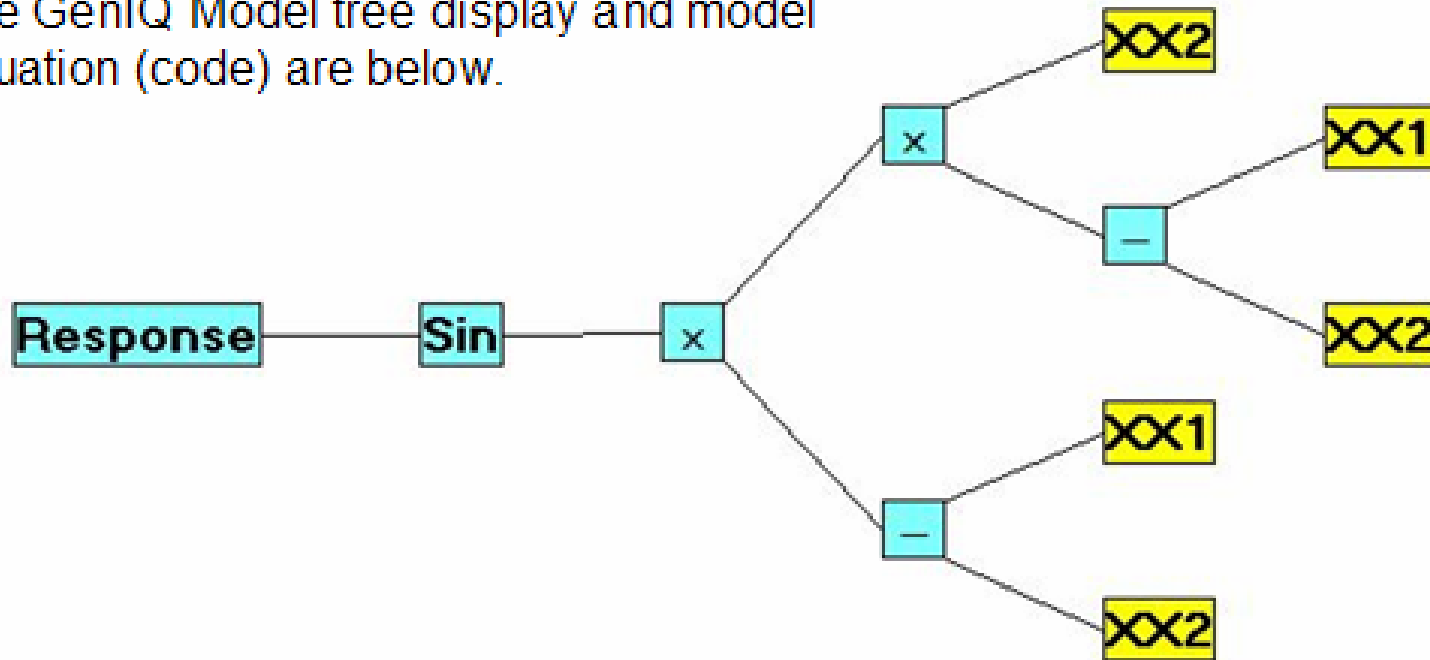
**LRM RESULTS:** The Response **ranking is not perfect.**

The ranking of Response based on the LRM probability score, Prob\_of\_Response, which is obtained directly from the LRM logit score, is in Table 2, below. Clearly, the Yes-No ordering is not perfect; amusingly, it seems random!

**Table 2**

<b>ID</b>	<b>XX1</b>	<b>XX2</b>	<b>Response</b>	<b>Prob_of_Response</b>
7	45	5	Yes	0.75472
10	30	10	No	0.61728
3	35	21	Yes	0.57522
5	45	37	No	0.53452
6	16	13	No	0.48164
8	30	30	Yes	0.46556
9	6	10	Yes	0.42336
1	31	38	Yes	0.41299
4	23	30	No	0.40913
2	12	30	No	0.32557

I built a **GenIQ Model to predict Response** using XX1 and XX2 with data in Table 1. The GenIQ Model tree display and model equation (code) are below.



```

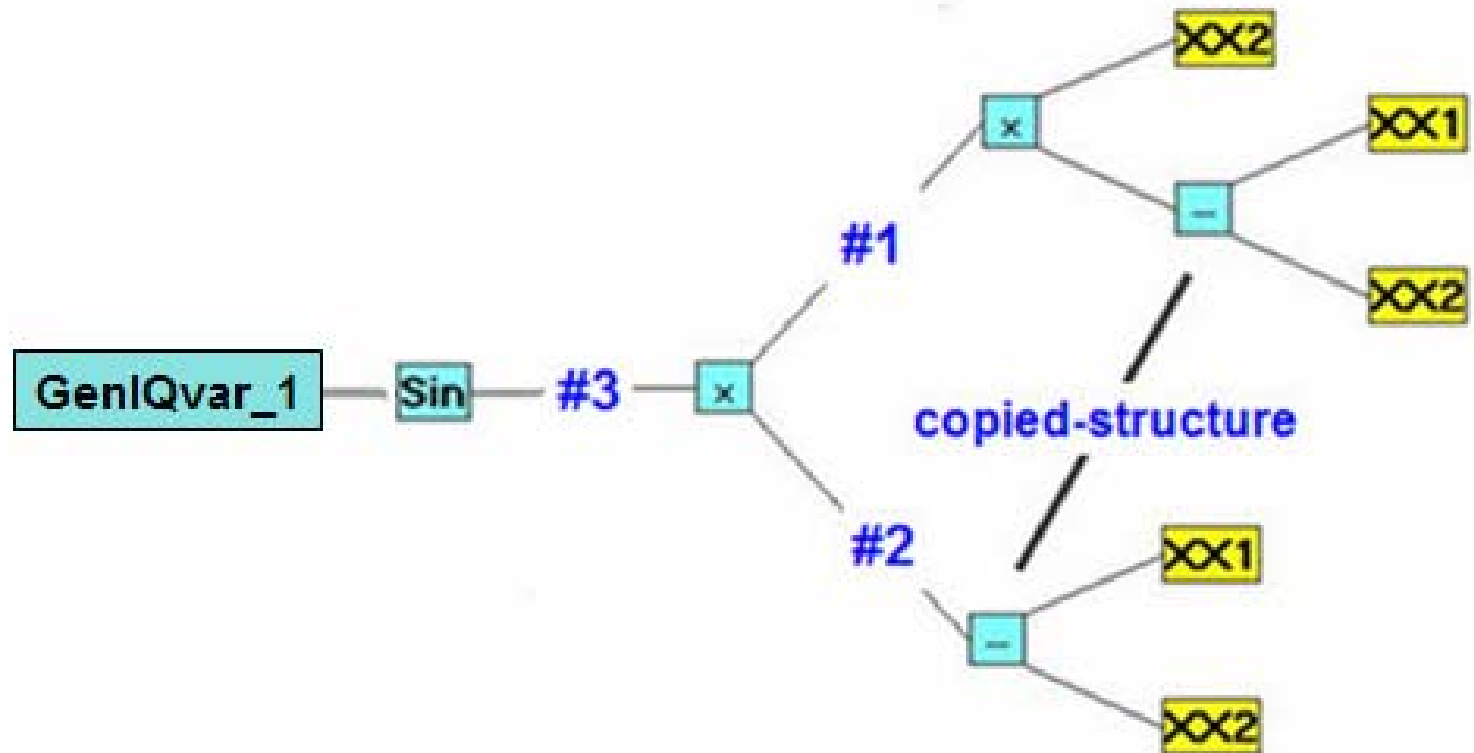
x1 = XX2;
    x2 = XX1;
x1 = x2 - x1;
    x2 = XX2;
        x3 = XX1;
    x2 = x3 - x2;
        x3 = XX2;
    x2 = x2 * x3;
x1 = x1 * x2;
x1 = Sin(x1);
GenIQvar = x1;

```

The tree and code are equivalent forms of GenIQs ability of extracting nonlinear dependencies within the data. An additional form can be expressed as:  
 Sine  $[(X2 * (X1 - X2)) * (X1 - X2)]$

# Genetically Data-mined Variables

Data Resue: Append GD-m Variables  
to Original Dataset



## GenIQ RESULTS: The Response ranking is perfect!

The ranking of Response based on the GenIQ Model score **GenIQvar** (which is a *unitless number*: the larger the value the greater the probability of response) is in Table 3, below. The Yes-No ordering is the best possible: All the Yeses are followed by all the Noes (or Nos). (*Unabridged*. Merriam-Webster, 2002)

**Table 3**

<b>ID</b>	<b>XX1</b>	<b>XX2</b>	<b>Response</b>	<b>GenIQvar</b>
<b>7</b>	<b>45</b>	<b>5</b>	<b>Yes</b>	<b>0.99784</b>
<b>1</b>	<b>31</b>	<b>38</b>	<b>Yes</b>	<b>0.82173</b>
<b>3</b>	<b>35</b>	<b>21</b>	<b>Yes</b>	<b>0.49134</b>
<b>9</b>	<b>6</b>	<b>10</b>	<b>Yes</b>	<b>0.21943</b>
<b>8</b>	<b>30</b>	<b>30</b>	<b>Yes</b>	<b>0.00000</b>
<b>2</b>	<b>12</b>	<b>30</b>	<b>No</b>	<b>-0.08756</b>
<b>4</b>	<b>23</b>	<b>30</b>	<b>No</b>	<b>-0.26226</b>
<b>10</b>	<b>30</b>	<b>10</b>	<b>No</b>	<b>-0.68350</b>
<b>5</b>	<b>45</b>	<b>37</b>	<b>No</b>	<b>-0.68955</b>
<b>6</b>	<b>16</b>	<b>13</b>	<b>No</b>	<b>-0.68970</b>



# GenIQ Model

**If you want to learn more about the GenIQ Model,  
then the best way is to attend a  
*free, at-your-convenience*  
GenIQ Model Webcast.**

**If interested, click [here](#).**

**Hope to have you attend.  
I promise not to waste your time,  
because I do not want to waste mine.**

A handwritten signature in black ink that reads "Bruce".