
GenIQ-enhanced/Data-reused Regression

Bruce Ratner PhD
DM STAT-1 CONSULTING
| 800 DM STAT-1 | www.DMSTAT1.com

GenIQ[®]

The GenIQ Model

- The GenIQ Model is a machine learning alternative model to the statistical ordinary least squares, logistic regression models.
- GenIQ let's the data define the model – automatically
 1. **Data mines** for new variables
 2. Performs **variable selection**
 3. **Specifies the model** (equation) - so as to "**optimize the decile table,**" to fill the upper deciles with as much profit/many responses as possible.
- 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: The GenIQ-enhancer

- The GenIQ Model can be used on a final regression model to let GenIQs data mining prowess **enhance**, data permitting, the results of the final model.
- GenIQ let's the data define the model – automatically
 1. Data mines for new variables,
 2. Performs variable selection, and then
 3. Specifies the model equation - so as to "optimize the decile table," to fill the upper deciles with as much profit/many responses as possible.
- GenIQs Utility: **Enhance** a final regression model by running GenIQ with only one predictor: the final regression equation score.

Build OLS Regression Model

- Objective #1: Build a OLS regression model with data, Table 1. I build a OLS model for predicting Y1 using XX1.
- The OLS equation, where estimated Y, denoted **est_Y**, is:
 - ▶ **est_Y1** = 134.94743 + 0.10005*XX1

Table 1. The Data

ID	XX1	Y1
1	158.2	157.8
2	214.9	146.6
3	153.2	147.5
4	196.0	153.1
5	88.5	143.7
6	55.5	132.3
7	86.4	144.3
8	223.6	169.1
9	256.9	160.9
10	252.4	157.1
11	20.9	141.6
12	92.9	145.4

Build OLS Regression Model

OLS RESULTS: *The Y1 ranking is not perfect.*

The ranking of Y1 based on the OLS score **est_Y** is in Table 2, below. Clearly, the Y1 ordering is not perfect; assumingly, it seems **random!**

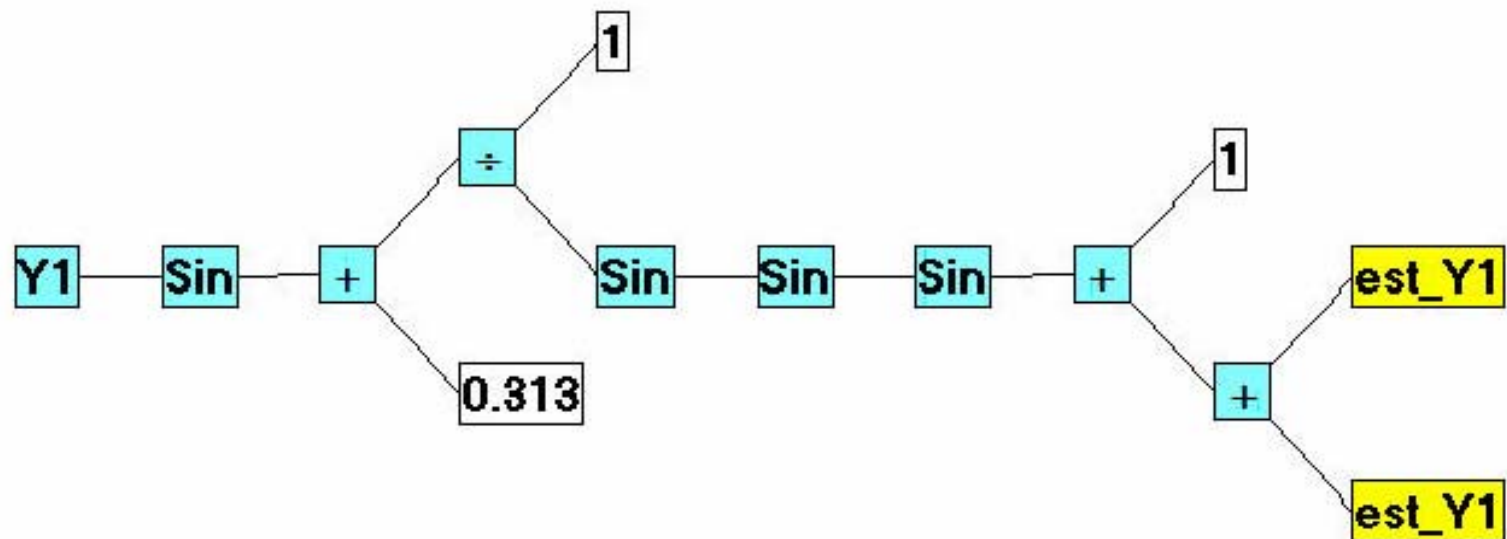
Table 2.

OLS Regression - est_Y1			
ID	XX1	Y1	est Y1
9	256.9	160.9	160.650
10	252.4	157.1	160.200
8	223.6	169.1	157.319
2	214.9	146.6	156.448
4	196.0	153.1	154.557
1	158.2	157.8	150.775
3	153.2	147.5	150.275
12	92.9	145.4	144.242
5	88.5	143.7	143.802
7	86.4	144.3	143.592
6	55.5	132.3	140.500
11	20.9	141.6	137.038

GenIQ-enhanced OLS: GenIQ Tree

- Running GenIQ with the OLS score:

▶ $est_Y1 = 134.94743 + 0.10005 * XX1$



GenIQ-enhanced OLS: GenIQ Code

- Running GenIQ with the OLS score:

- ▶ $est_Y1 = 134.94743 + 0.10005 * XX1$

```
x1 = .3125;  
  x2 = est_Y1;  
    x3 = est_Y1;  
  x2 = x2 + x3;  
    x3 = 1;  
  x2 = x2 + x3;  
  x2 = Sin(x2);  
  x2 = Sin(x2);  
  x2 = Sin(x2);  
    x3 = 1;  
  If x2 NE 0 Then x2 = x3 / x2; Else x2 = 1;  
x1 = x1 + x2;  
x1 = Sin(x1);  
GenIQvar1 = x1;
```

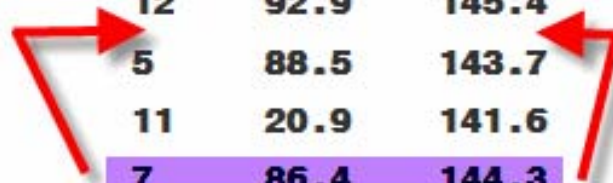
GenIQ-enhanced OLS: Results

The GenIQ-enhanced OLS Results: The Y1 ranking is *almost* perfect: ID # 7 should be ranked 9th. The GenIQ Model with only OLS score `est_Y1` produces GenIQ Model score `GenIQvar1`, in Table 3. GenIQ-enhancing usually produces perfect rankings. If not, then use: "Data Reuse."

Table 3.

GenIQ-enhanced OLS Regression - `GenIQvar1`

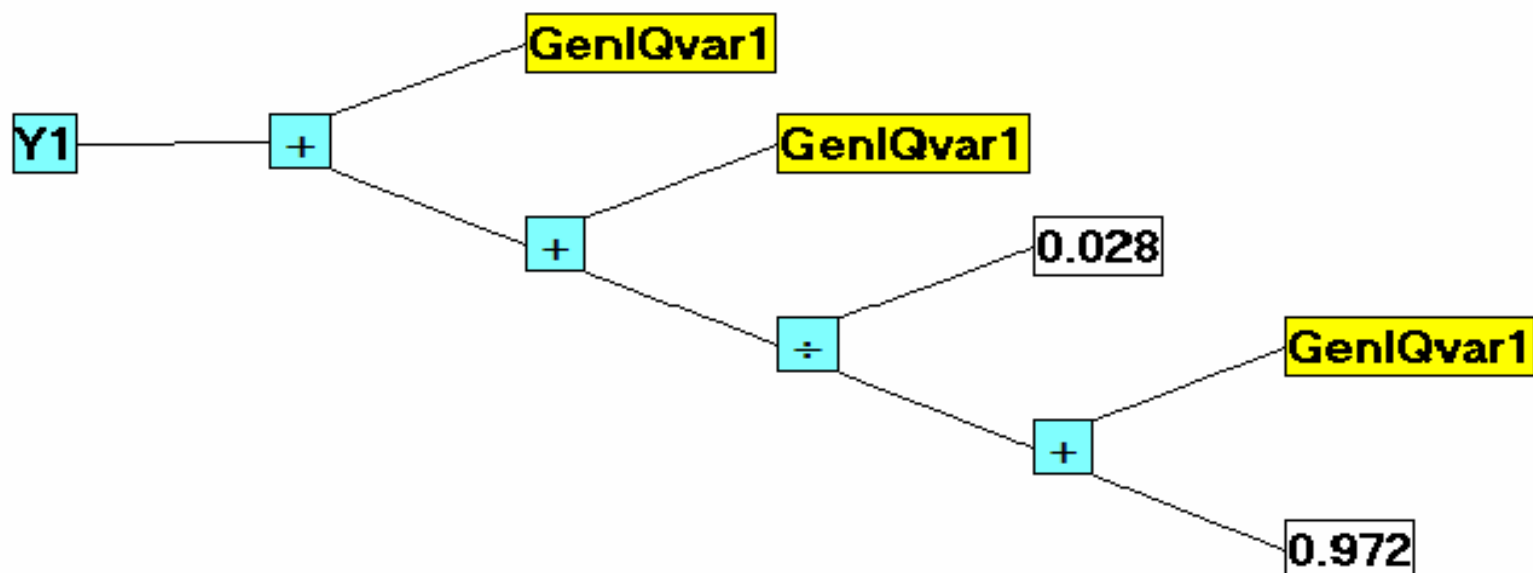
ID	XX1	Y1	GenIQvar1
8	223.6	169.1	0.99633
9	256.9	160.9	0.99383
1	158.2	157.8	0.97093
10	252.4	157.1	0.97091
4	196.0	153.1	0.96074
3	153.2	147.5	0.90391
2	214.9	146.6	0.46214
12	92.9	145.4	0.40553
5	88.5	143.7	-0.77044
11	20.9	141.6	-0.86257
7	86.4	144.3	-0.95786
6	55.5	132.3	-0.98445



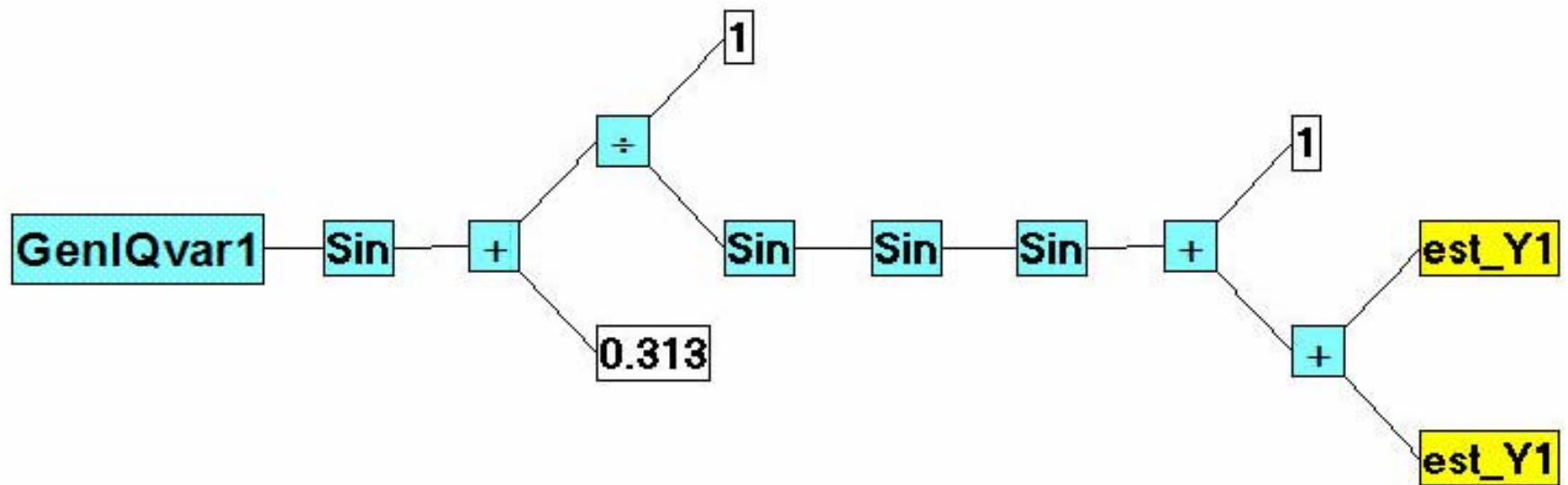
GenIQ-enhanced/Data-reused OLS

- Data Reuse: Adjusting for setting too small: number of generations and/or population size.
 - When building a GenIQ Model, it is recommended to append the new "genetically data-mined" variables (the branches of the GenIQ tree, including the model itself) to the original dataset, and then re-run GenIQ.
 - Re-run of GenIQ is repeated until no better model is evolved.
 - The implication of reusing variables that are statistically correlated is that multicollinearity is not an issue for GenIQ.
- GenIQ-enhanced/Data-reused OLS Results:
The Y1 ranking is perfect! I build a GenIQ Model using **est_Y1** and the **GenIQvar1** from the previous GenIQ-enhanced OLS.

GenIQ-enhanced/ Data-reused OLS: GenIQ Tree



GenIQ-enhanced/ Data-reused OLS: **GenIQvar1**



GenIQ-enhanced/ Data-reused OLS: Results

The GenIQ-enhanced/Data-reused OLS Results: *The Y1 ranking is perfect!*

The ranking of Y1 - based on the GenIQ Model with **est_Y1** and **GenIQvar1** - produces a GenIQ Model score **GenIQvar**, in Table 4, yielding perfect Y1 ordering. Note: the final model does not include **est_Y1**.

Table 4.

Data-Reuse GenIQ-enhanced OLS Regression - GenIQvar

ID	XX1	Y1	GenIQvar
8	223.6	169.1	2.00689
9	256.9	160.9	2.00190
1	158.2	157.8	1.95628
10	252.4	157.1	1.95623
4	196.0	153.1	1.93598
3	153.2	147.5	1.82275
2	214.9	146.6	0.94382
12	92.9	145.4	0.83140
7	86.4	144.3	0.08468
5	88.5	143.7	-1.40183
11	20.9	141.6	-1.46889
6	55.5	132.3	-4.19334

Thank you

Danke Xie xie

Khawp khun

Yum botie

Mahalo

Salamat

Juspaipaña

Spacibo Obrigada

Arigato