Step 1:
Launch GenIQ.
Click "Start a New Project."
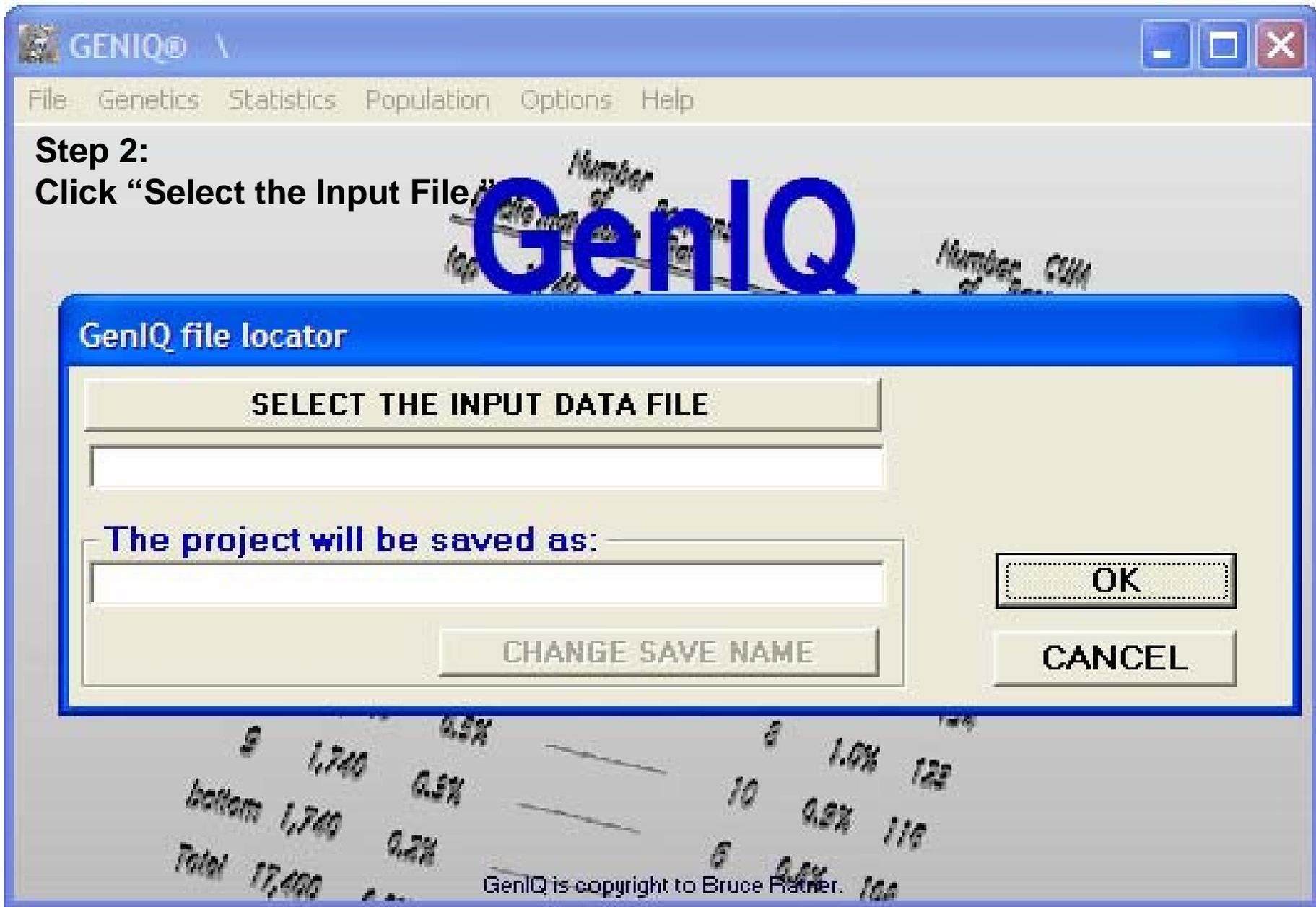
Step 2:
Click "Select the Input File."

GENIQ® \

File   Genetics   Statistics   Population   Options   Help

GenIQ file locator

SELECT THE INPUT DATA FILE

The project will be saved as:

CHANGE SAVE NAME

OK

CANCEL

GenIQ is copyright to Bruce Ratner.

Step 4:
Click "OK."

LOADING FILE NOW.

Identify VARS

- ◉ Target Y
- ○ Predictor X
- ○ Deselect

Set holdout %

◀ ▶ 50

VARIABLE NAME	Abr	USAGE	TYPE	Comment
response	x1	Y	Numeric	OK
hi_balance	x2	X	Numeric	OK
no_bal_decr	x3	X	Numeric	OK
no_of_trans	x4	X	Numeric	6 missing
recency_mos	x5	X	Numeric	OK
mos_on_file	x6	X	Numeric	OK
tranx_active	x7	X	Classes[3]	OK

Step 5:
A. **Select Identification of Variables:**
   Target Variable, and Predictor Variables.
B. **Select Handling of Data:**
   Missing Data, and % for Holdout.
C. **Then, click "OK."**

OK

CANCEL

HOW SHOULD MISSING DATA BE TREATED

- ○ Quasi Complete-Case Analysis
  (Casewise Deletion on NumVars)
- ○ Complete-Case Analysis
  (Casewise Deletion)
- ◉ All-Case Analysis
  (Genetic Imputation)

Step 6:
Accept or change sample response rate.
Then click "OK."

Step 7:
Click "SET DEFAULTS."
Then click "OK."

GENETIC ALPHABET SELECTOR

| SET DEFAULTS | CLEAR | O.K. | CUSTOM FUNCTIONS | HELP |

ARITHMETIC operations

+	−	x	÷	∧	\=\
Abs	Fra	Int	Mod	Sqn	
Ln	Exp	Log	10^x	Exp-	
√	Crt	Lsr	Bgr	Atn2	

CIRCULAR functions

Sin	Cos	Tan
Asin	Acos	Atn
Csec	Sec	Cot
Acsc	Asec	Acot

NUMERIC MATERIAL

0.1	2	Pi	Rand
1	3	e	
10	4	True	ℜ
100	5	False	

LOGICALS   {CRISP, fuzzy}

<	≤	=	>	≥	≠	≅
AND	OR	NOT	EQV	IMP	XOR	
and	or	not	eqv	imp	xor	
Btwn	ifT	if<=	Xltr			

HYPERBOLIC functions

Sinh	Cosh	Tanh
Asnh	Acsh	Atnh
Csch	Sech	Coth
Acsch	Asch	Acth

CUSTOM functions

Fn	Fn	Fn	Fn
Fn	Fn	Fn	Fn
Fn	Fn	Fn	Fn
Fn	Fn	Fn	Fn

Bruce Ratner, Ph.D.

Step 8:
Click "Run GenIQ" bar at the bottom.

POPULATION AND BREEDING CONTROLS

Genetic POPULATION size

Models/Gen. ◄ | | ► | 250

Stop when

Decile 1 Cum Lift is > 999

OR generations = ◄ | | ► | 100

BREEDING PROBABILITIES

PROBABILITIES		%
Crossover	◄►	80
Terminal Selection	◄►	20
Mutation (I/R/P)	◄►	20
Editing	◄►	2

Set default values

GenIQ FITNESS CONTROL

SELECTION METHOD

○ Fitness only

○ Fitness & beauty

◉ Tournament

FITNESS BASED ON

◉ GenIQ Fitness Function 2.0

Run GenIQ

Bruce Ratner, Ph.D.   8

Step 9:
Let the GenIQ Model, displayed in tree-form in the top-right window, evolve. Click "PAUSE" when you are satisfied with the corresponding 'optimized' decile table, as GenIQs objective is to optimized the decile table.

RESPONSE DECILE ANALYSIS

TRAINING SET sample size = 4597
Population response = 8.70%

DECILE	Predicted Response	Resp Rate%	Cum RespR%	Cum Lift
Top	75	16.31	16.30	187
2nd	48	10.44	13.38	154
3rd	40	8.70	11.82	136
4th	47	10.22	11.42	131
5th	34	7.40	10.62	122
6th	34	7.40	10.08	116
7th	41	8.92	9.91	114
8th	29	6.31	9.46	109
9th	27	5.87	9.06	104
Bottom	25	5.44	8.70	100
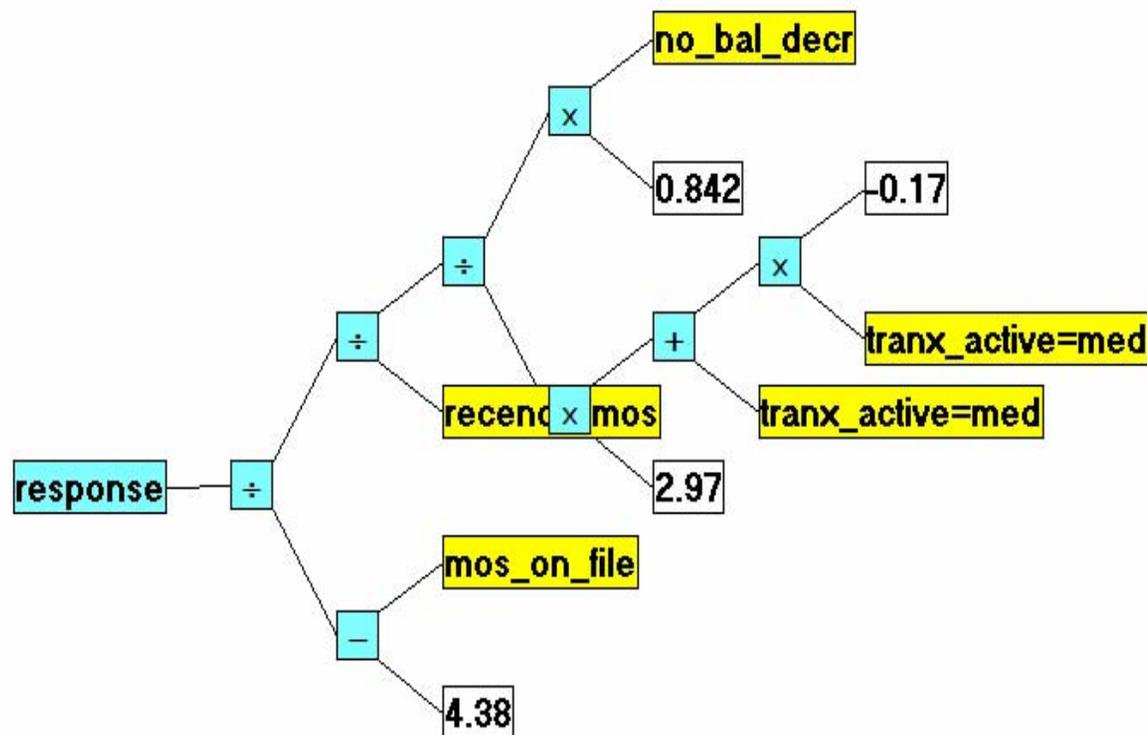
- **The GenIQ Model is conveniently displayed in tree-form; but, it is actually a *computer program!* Accordingly, it is the computer code that represents the final GenIQ Model, i.e., serves as the model equation.**
- **For the GenIQ Model presented, the model equation/computer code is:**

```
x1 =  4.38;
     x2 = mos_on_file;
   x1 = x2 - x1;
     x2 = recency_mos;
        x3 =  2.97;
           If tranx_active = "med" Then x4 = 1; Else x4 = 0;
              If tranx_active = "med" Then x5 = 1; Else x5 = 0;
                 x6 = -.17;
              x5 = x5 * x6;
           x4 = x4 + x5;
        x3 = x3 * x4;
           x4 =  .842;
              x5 = no_bal_decr;
           x4 = x4 * x5;
        If x3 NE 0 Then x3 = x4 / x3; Else x3 = 1;
     If x2 NE 0 Then x2 = x3 / x2; Else x2 = 1;
   If x1 NE 0 Then x1 = x2 / x1; Else x1 = 1;
GenIQvar = x1;
GenIQ_Probability = 1 / (1 + Exp(-(-2.688323 + GenIQvar * 88.51257)));
```

GenIQ Predictive Analytics

- The previous nine steps/screen-shots indicate how simple predictive analytics can be. Of course, there is much to be said about each step. But, the message it clear: Building a regression model with GenIQ requires virtually no statistical background, as the GenIQ provides a machine learning solution to the classical statistical problem of building predictive regression models.

- For more about GenIQ, call Bruce Ratner at 516.791.3544

- Benefits of GenIQ, and GenIQ as a Data Miner are on the following slides.

The GenIQ Model© Software: Benefits

<u>BENEFITS</u>
The following features of the GenIQ Model© directly translate into significant value-added benefits of the GenIQ Software.

1. The GenIQ Model has <u>maximum model lift</u>, as its best-of-generation solution estimates the rank-order likelihood of response or contribution of profit.
   Value-added: Greater response/profit – increased cost efficiency in DM programs.

2. The GenIQ Model is <u>more reliable</u> than the logistic and ordinary regression models because it lets the data suggest the equation form without any assumptions. The regression models require the rigid assumption of the equation form (sum of weighted predictor variables), and a host of other untenable assumptions.
   Value-added: Greater model reliability – increased confidence in the predicted DM program results.

3. The GenIQ Model is <u>automatically built</u> in a timeframe favorably commensurate with the input file size. Model validation is quickly and easily performed.
   Value-added: Greater rate per model built – increased human capital.

4. The GenIQ Model can serve as a <u>productivity tool</u> for data analysts who build models in the DM Space where maximizing model lift is the bottom-line objective.
   Value-added: Greater productive labor – increased quantity of models built.

5. GenIQ Model can serve as a <u>specialty tool</u> for experts who want control over the individual processes: selecting among the important original variables, assessing the predictiveness of the newly data mined variables, and developing the best model with maximum model lift.
   Value-added: Greater "creativity" afforded – increased quality of models built and individuality of human capital.

6. The GenIQ Model provides <u>automatic data mining</u> – an inherent by-product of the genetic programming methodology. Data analysts can export the genetically constructed variables into an independent model application.

    Value-added: Presents a hybrid statistics-ML paradigm that yields a utile alternative for DM modeling. (See article *A Hybrid Statistics-Machine Learning Paradigm for DM Modeling*.)

7. The GenIQ Model provides a <u>unique variable selection</u> of important predictor variables, as it provides the ranking of the relationship between each predictor variable with the target variable – accounting for the presence of the other predictor variables jointly considered. This is in stark contrast to the statistical correlation coefficient, which provides the ranking of the linear-relationship between each predictor variable with the target variable – without considering the other predictor variables.

    Value-added: Presents a hybrid statistics-ML paradigm that yields a utile alternative for DM modeling. (See article *A Hybrid Statistics-Machine Learning Paradigm for DM Modeling*.)

8. "The GenIQ Model expands to fill the time available for model completion – <u>guaranteeing the best model for the allotted time</u>."

    Value-added: See The GenIQ-Parkinson's Law.

<u>WHO & WHY</u>

The standard statistical models of logistic and ordinary regression for binary and continuous target variables, respectively, have two weaknesses:

1. They do not address the DM objective of maximizing model lift – the customary performance measure used in the DM Space.
2. They do not have built-in data mining capability – the intelligence for constructing new variables with incremental predictiveness beyond the original variables.

Thus, Dr. Ratner invented the assumption-free, nonparametric GenIQ Model©, which addresses the weaknesses of the statistical DM problem-solution, by using the machine learning (ML) method of genetic programming (GP).

WHAT

GenIQ Software is the PC-based implementation of the GenIQ Model for building:

1. Classification model – the target variable is a binary (e.g., yes-no response).
2. Prediction model – the target variable is continuous (e.g., many-valued profit).

GenIQ Software is flexible during the data input and output processes:

1. Imports unlimited records and candidate predictor variables of any type (categorical, ordinal, and continuous).
2. Reduces the mandatory exploratory data analysis to only checking for implausible and impossible values – due to the consequences of features #1 and #5 above.
3. Eliminates the concern for finding outliers, and assessing their effects.
4. Accommodates efficiently the problem of missing data with genetic imputation.
5. Builds a resistant-to-overfitting GenIQ Model equation – due to a "smoothing" component in the fitness function.
6. Exports effortlessly the GenIQ Model equation into any independent application for error-free scoring of the model to an external file or database.

HOW

GenIQ Software has only four simple screens:

1. Input File – to indicate the input data location and file type (e.g., SAS©, SPSS©, SQL, Excel, text, and comma delimited).
2. Identify Variables – to identify the target and candidate predictor variables.
3. Genetic Parameters – to scan the default control settings for the run. Fiddling with the settings is rarely needed, as the GP-approach to maximizing model lift is quite resistant.
4. GenIQ Model – to view the model in its signature "tree" display, not the typical ML indiscernible "black box."

DISADVANTAGE

The GenIQ Model – represented as a tree – is difficult to interpret, in part, because it has no coefficients. Tyros and experienced analysts when interpreting a model unwittingly seek the regression coefficients, as they are the means to interpret the everyday logistic and ordinary regression models.

Non-DISADVANTAGE

The GenIQ Model – represented as a set of "branches" – lends itself to interpretation even without coefficients. The branches, which are actually small models that are defined with a few variables, can be explained with an understanding of the content domain of the variables. Viewing the GenIQ Model as such, data analysts and end-users acquire a comfort level for using the model that is otherwise difficult to understand and accept.

WHEN

Now, call Bruce at 516.791.3544, or e-mail him at br@dmstat1.com to set up a date and time for a ten minute e-demonstration of the GenIQ Software, after which you can get the GenIQ demo for a 2 week trial.

The GenIQ Model©: A Unique Data Mining Tool

The GenIQ Model© is an *evolutionary* advance in data mining methodology developed and patented by Bruce Ratner, President and Founder of DM STAT-1 CONSULTING, the leading firm for analysis and modeling in the DM Industry. GenIQ offers exceptional predictions with minimal error variance, and a unique feature accommodating dirty and incomplete data. GenIQ can handle both classification (e.g., target yes-no response variable) and regression (e.g., target continuous sales variable) problems with categorical, ordinal and continuous candidate predictor variables.

GenIQ is designed for the optimization of the ubiquitous decile analysis (gains chart). When GenIQ achieves this goal - for either a simple or complex model - the visual displays produced are easy on the analyst's eyes for understanding the impact of any relevant predictor variable or pair of predictor variables on the target variable, thus revealing the underlying data structure.

GenIQ is a tool to be used virtually without data preparation - except for insuring there are no impossible or improbable values (e.g., age of 120 years, or a boy named Sue). GenIQ quickly leads to a detailed understanding of the value of the data, i.e., the identification of the key-drivers of the target variable. The GenIQ model output looks like a *tree*, not like a CHAID or CART tree, but like itself! Actually, it is technically a computer program, thus the GenIQ Model is a set of computer code. Each branch, which defined by two or more variables tied-together by one or more functions, is the identification of genetic-evolved key-drivers of the target variable. This is the unique data mining feature of the GenIQ Model in the tree below.
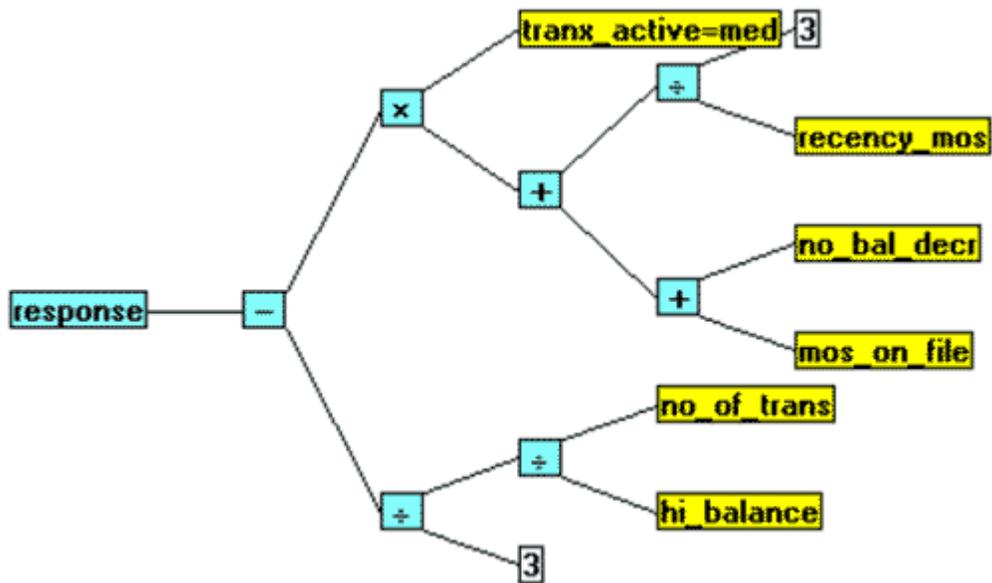
Thus, the following pieces of structure mined by GenIQ for predicting response are:

1. Structure_1 (mini-model #1) = **3 / recency_mos**
2. Structure_2 (mini-model #2) = **no_bal_decr + mos_on_file**
3. Structure_3 (mini-model #3) = **no_of_trans / hi_balance**
4. Structure_4 (mini-model #4) = **Structure_3 / 3**
5. Structure_5 (mini-model #5) = **Structure_1 / Structure_2**
6. Structure_6 (mini-model #6) = **tranx_active = med / Structure_3**
7. GenIQ Model (Super-structure) = **Structure_6 - Structure_4**