Ask the Expert

Variable Selection Using SAS Enterprise Guide and SAS Enterprise Miner

Goals

- Increase awareness of and comfort with capabilities in SAS® for doing variable selection
  - SAS® Enterprise Guide®
  - SAS® Enterprise Miner™
- Share resources for learning more

AGENDA

Variable Selection using SAS® Enterprise Guide® and SAS® Enterprise Miner™

- What is variable selection?
- Why is it important?
  - Why should it be on your list of activities when doing predictive modeling?
- How to do variable selection using
  - SAS Enterprise Guide
  - SAS Enterprise Miner

What?

Variable Selection or Variable Reduction

Variable selection is used to find a subset of the available inputs that accurately predict the output.

Why?

Variable Selection

Our life is frittered away by detail...Simplify, simplify.

Henry David Thoreau

Why Variable Selection?

- Smaller Data
  - Speed/Performance
    - Decreased Computation Time
    - Decreased Scoring Effort
  - Cost
    - Data Collection
    - Data Cleaning

- Other Statistical Reasons
  - Interpretability
  - Multicollinearity & Irrational Coefficients
  - Missing Data
  - Redundancy
  - Predictive Power
    - Destabilize the parameter estimates
    - Increase the risk of over fitting
    - Noise

The principle of Occam's Razor states that among several plausible explanations for a phenomenon, the simplest is best.

Before Variable Selection

Things to consider

Decide how you intend to use your model.

- Describe the relationship between variables
- Which predictors are statistically significant
- Model has reasonable goodness-of-fit
- Ability to predict

Ideally, the ultimate model would do all of these tasks, describe and predict, equally well. Rarely do we have that luxury in the real world of messy and uncooperative data to accomplish all of these.

Before Variable Selection

Tasks to complete

- Identify outliers and influential points - maybe exclude them at least temporarily.
- Add in any transformations of the variables that seem appropriate.
- Impute missing values

Variable Selection
Concepts

Variable Selection

- Regression based
- Criterion Based
- Variable Screening
- Variable Clustering

Variable Combination

- Principal Components uncorrelated linear combinations of ***all*** input variables

Variable Selection
Concepts

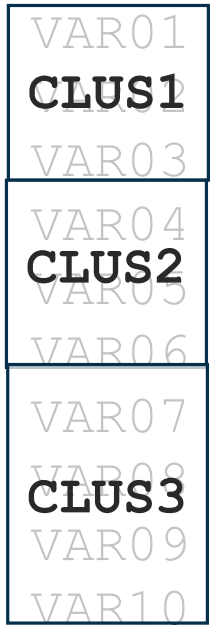
Input variables and TARGET:

VAR01
VAR02
VAR03
VAR04
VAR05
VAR06
VAR07
VAR08
VAR09
VAR10
TARGET

Variable Selection based on correlation with TARGET:

VAR01
VAR02
VAR04
VAR07
VAR09
TARGET

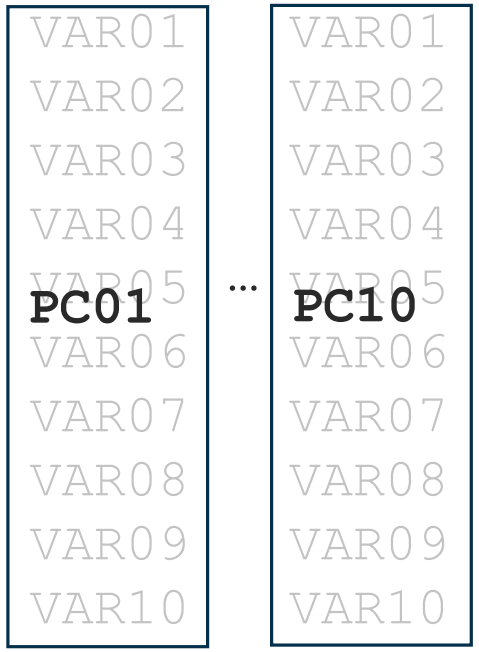
Cluster Scores based on **Variable Clustering**:



Best Variables based on **Variable Clustering**:

VAR02
VAR05
VAR09

Principal Components:



Variable Selection

Methods

- All Possible
  - Best subset selection methods
- Automatic Subset Selection Methods
  - Stepwise, Backward, Forward
- Criterion Based/Variable Screening
  - Variable Ranking, Correlations, Weight of Evidence

Which one is best?

Subset Selection Methods
Automatic and All Possible

- *Stepwise selection* considers adding and deleting predictors at each step of the process
- *Forward selection* begins with a simple regression model and adds, one at a time. However, once a predictor is in the equation, it is never deleted.
- *Backward selection* begins with the multiple regression model including all possible predictors and deletes, one at a time. Once a variable is deleted, it is never reconsidered for inclusion.
- *Best subsets* estimates one regression model for all possible combinations of the predictor variables and chooses the best model among them.

When five predictors (x's) are available for estimation, there are: 5 simple regression models, 10 different two-predictor models, 10 different three-predictor models, 5 different four-predictor models, and 1 five-predictor model, totaling 31 regressions. When ten predictors are available, there are 1,023 possible subsets.

Data

Donor_Raw_Data

People likely to donate to a charity

- Y=TARGET_B
- N = 19,372
- Variables = 50 (47 Inputs)

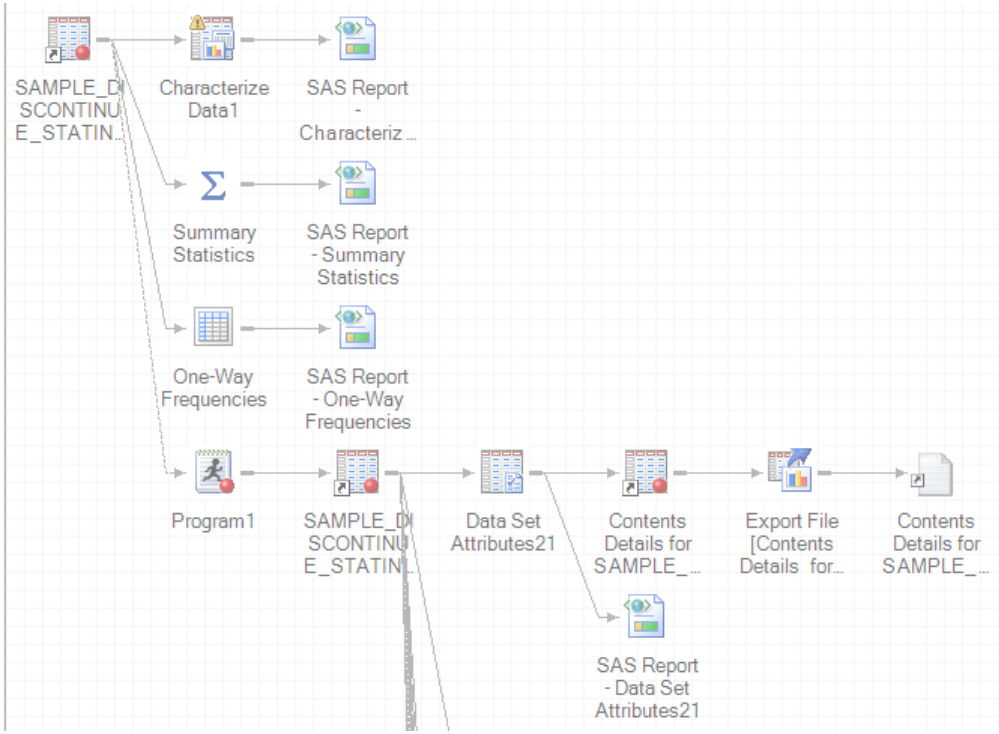
Column Descriptions
Download Data

Alphabetic List of Varia				
#	Variable	Type	Len	Form
37	CARD_PROM_12	Num	8	
8	CLUSTER_CODE	Char	2	
3	CONTROL_NUMBER	Char	8	
10	DONOR_GENDER	Char	3	
41	FILE_AVG_GIFT	Num	8	
42	FILE_CARD_GIFT	Num	8	
21	FREQUENCY_STATUS_97NK	Num	8	
9	HOME_OWNER	Char	3	
48	IM_DONOR_AGE	Num	8	2.
46	IM_INCOME_GROUP	Num	8	2.
50	IM_MONTHS_SINCE_LAST_PROM_RESP	Num	8	2.
44	IM_WEALTH_RATING	Num	8	2.
5	IN_HOUSE	Num	8	
36	LAST_GIFT_AMT	Num	8	
32	LIFETIME_AVG_GIFT_AMT	Num	8	
28	LIFETIME_CARD_PROM	Num	8	
30	LIFETIME_GIFT_AMOUNT	Num	8	
31	LIFETIME_GIFT_COUNT	Num	8	
33	LIFETIME_GIFT_RANGE	Num	8	
34	LIFETIME_MAX_GIFT_AMT	Num	8	
35	LIFETIME_MIN_GIFT_AMT	Num	8	
29	LIFETIME_PROM	Num	8	
14	MEDIAN_HOME_VALUE	Num	8	
15	MEDIAN_HOUSEHOLD_INCOME	Num	8	
40	MONTHS_SINCE_FIRST_GIFT	Num	8	
39	MONTHS_SINCE_LAST_GIFT	Num	8	
4	MONTHS_SINCE_ORIGIN	Num	8	
13	MOR_HIT_RATE	Num	8	
47	M_DONOR_AGE	Num	8	
45	M_INCOME_GROUP	Num	8	
49	M_MONTHS_SINCE_LAST_PROM_RESP	Num	8	
43	M_WEALTH_RATING	Num	8	
38	NUMBER_PROM_12	Num	8	
12	OVERLAY_SOURCE	Char	1	
16	PCT_OWNER_OCCUPIED	Num	8	
18	PEP_STAR	Num	8	
17	PER_CAPITA_INCOME	Num	8	
11	PUBLISHED_PHONE	Num	8	
20	RECENCY_STATUS_96NK	Char	5	
25	RECENT_AVG_CARD_GIFT_AMT	Num	8	
23	RECENT_AVG_GIFT_AMT	Num	8	
27	RECENT_CARD_RESPONSE_COUNT	Num	8	
24	RECENT_CARD_RESPONSE_PROP	Num	8	
26	RECENT_RESPONSE_COUNT	Num	8	
22	RECENT_RESPONSE_PROP	Num	8	
19	RECENT_STAR_STATUS	Num	8	
7	SES	Char	4	
1	TARGET_B	Num	8	
2	TARGET_D	Num	8	
6	URBANICITY	Char	4	

SAS® Enterprise Guide®

SAS® Enterprise Guide®

Methods available



- Regression
- Variable Screening
  - Correlation
- Variable Clustering
- Principal Components
- Weight of Evidence (WOE) and Information Value (IV)

SAS® Enterprise Guide®

First Things First

Impute missing values

Categorical

```
if donor_gender in ('U','A') then
    donor_gender='U';

if SES='?' then
    SES='5';

if URBANICITY='?' then
    URBANICITY='M';
```

Set Gender to Unknown, SES to Level 5 (Unknown), Urbanity to M (missing)

Continuous

```
proc HPIMPUTE data=Donor.donor_raw_data out=out1;
     input  wealth_rating income_group donor_age months_since_last_prom_resp;
     impute wealth_rating / method=random;
     impute income_group / method=random;
     impute donor_age /method=random;
     impute months_since_last_prom_resp / method=random;
 run;
```

MEAN, RANDOM, PMEDIAN or Constant Value

SAS® Enterprise Guide®

First Things First – New in SAS/Stat 14.1

Impute missing values – PROC SURVEYIMPUTE

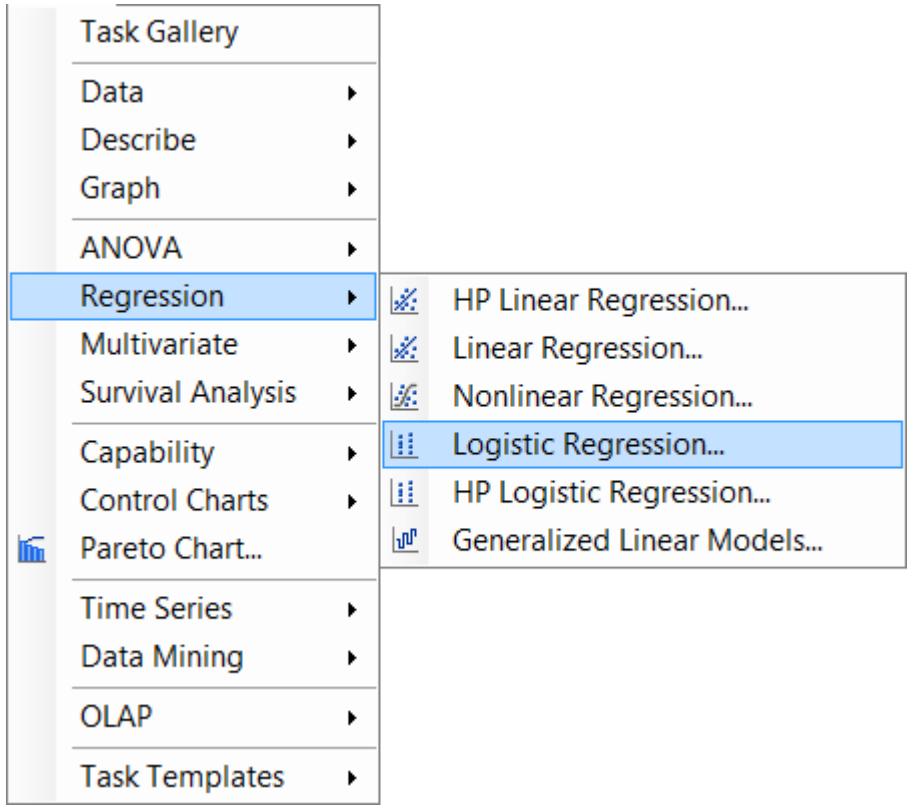
The SURVEYIMPUTE procedure imputes missing values of an item in a sample survey by replacing them with observed values from the same item. Imputation methods include single and multiple hot-deck imputation and fully efficient fractional imputation (FEFI)

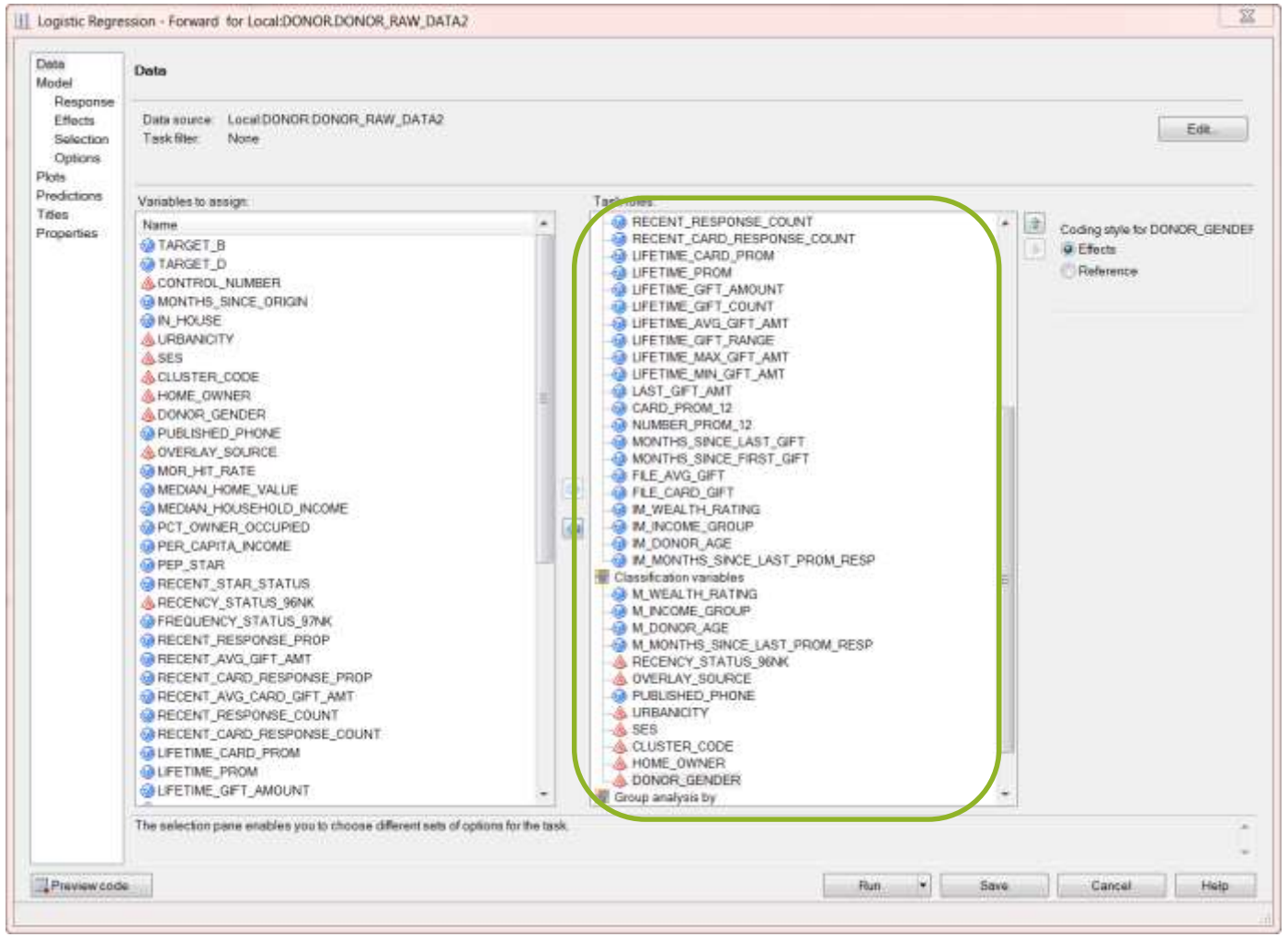
```
/* Joint imputation for Department and Response*/
proc surveyimpute data=SIS_Survey_Sub method=fefi varmethod=jackknife;
  class Department Response;
  var Department Response;
  strata State NewUser;
  cluster School;
  weight SamplingWeight;
  output out=SIS_Survey_Imputed outjkcoefs=SIS_JKCoefs;
run;
```

PROC SURVEYIMPUTE Documentation

SAS® Enterprise Guide®

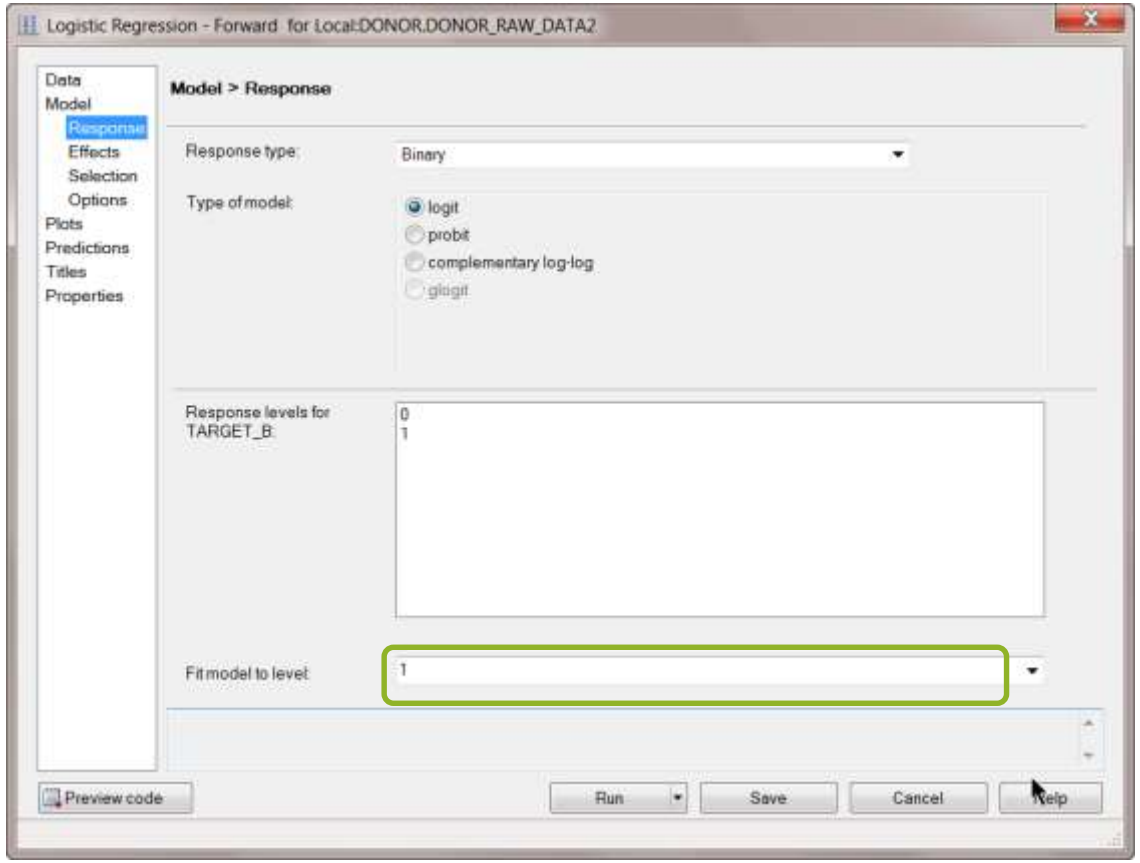
Logistic Regression

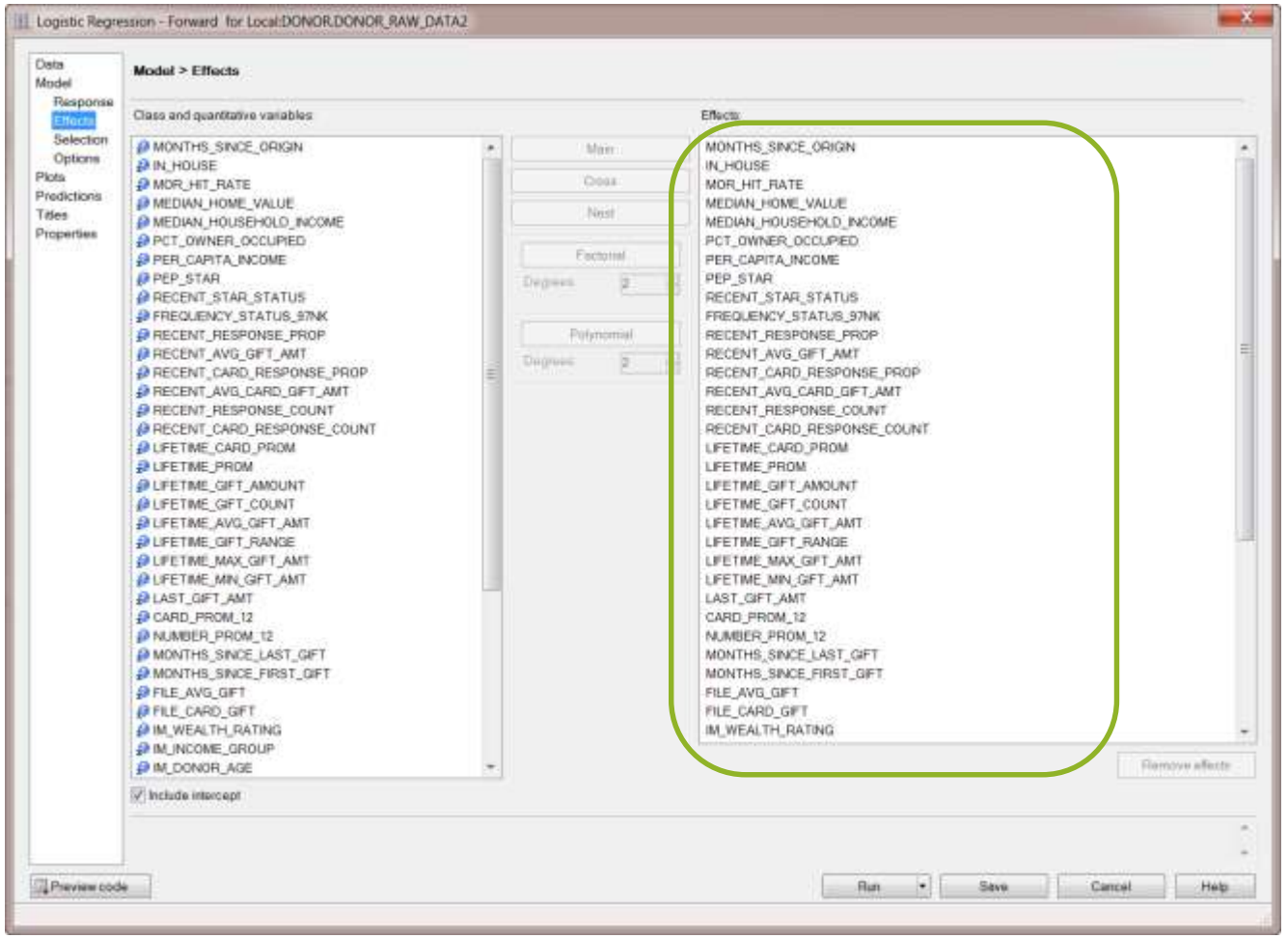




SAS® Enterprise Guide®

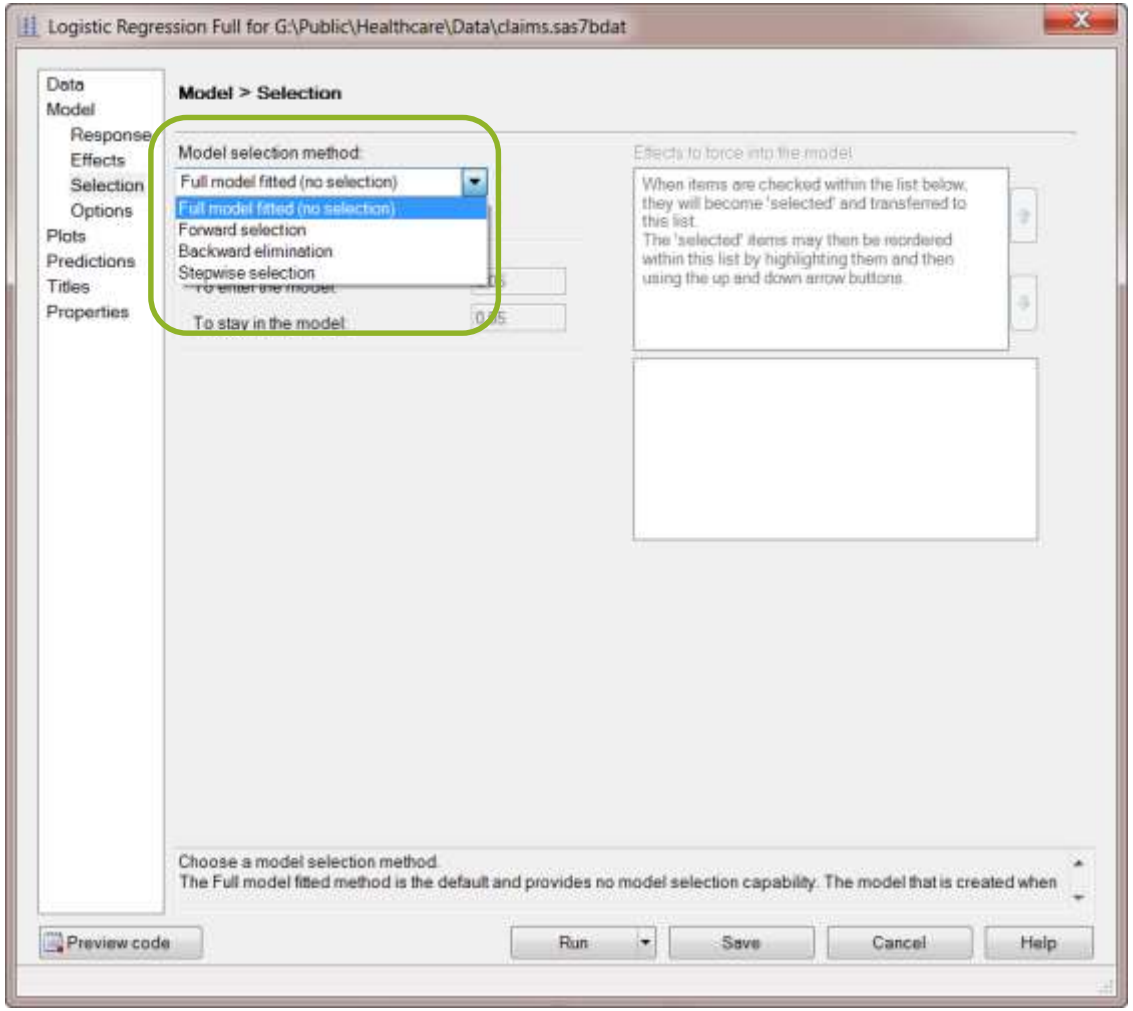
Logistic Regression





SAS® Enterprise Guide®

Logistic Regression



- Selection=None
- Selection=Forward
- Selection=Backward
- Selection=Stepwise
- Selection=Score

❖ Ones in BLUE available in SAS Enterprise Guide

Variable Selection Methods in Proc Logistic Documentation

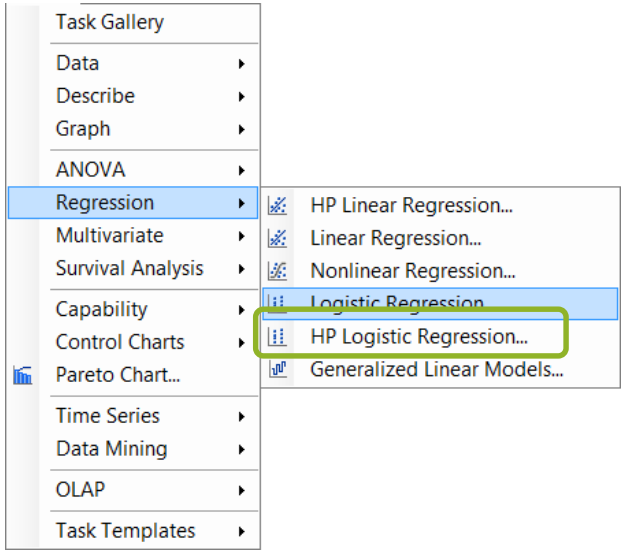
SAS® Enterprise Guide®

Logistic Regression

Variable Name	Stepwise	Backward	Forward
CLUSTER_CODE		*	
FREQUENCY_STATUS_97N	*	*	*
HOME_OWNER	*	*	*
IM_WEALTH_RATING	*	*	*
IN_HOUSE		*	
LIFETIME_CARD_PROM	*		*
M_WEALTH_RATING	*	*	*
MEDIAN_HOME_VALUE	*	*	*
MONTHS_SINCE_FIRST_GIFT	*	*	*
MONTHS_SINCE_LAST_GIFT	*	*	*
NUMBER_PROM_12		*	
PEP_STAR	*	*	*
RECENT_AVG_GIFT_AMT	*	*	*
RECENT_CARD_RESPONSE_COUNT			*
RECENT_CARD_RESPONSE_PROP	*	*	*
SES	*		*
Number of Variables	12	13	13

SAS® Enterprise Guide®

HPLogistic Regression



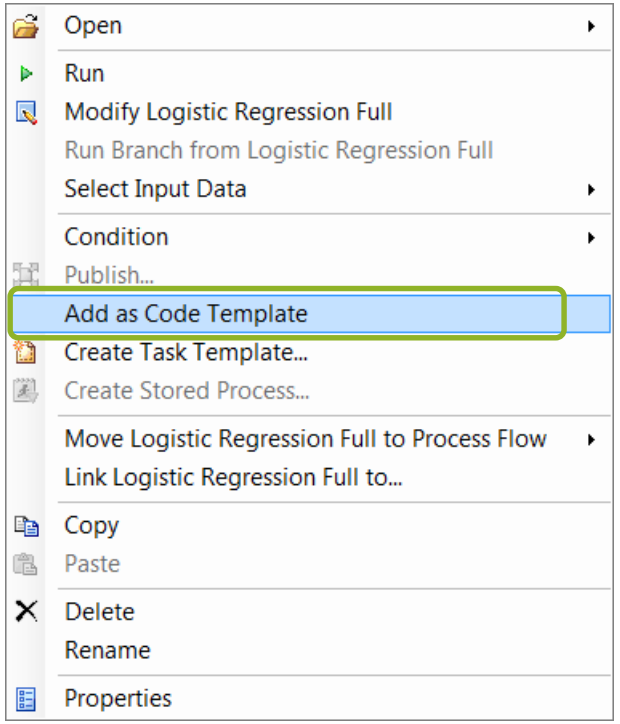
Variable Name	Stepwise	Backward	Forward	Backward Fast
CLUSTER_CODE		*		
FREQUENCY_STATUS_97N	*	*	*	*
HOME_OWNER	*	*	*	*
IM_WEALTH_RATING	*	*	*	*
IN_HOUSE		*		*
LIFETIME_CARD_PROM	*		*	
M_WEALTH_RATING	*	*	*	*
MEDIAN_HOME_VALUE	*	*	*	*
MONTHS_SINCE_FIRST_GIFT	*	*	*	*
MONTHS_SINCE_LAST_GIFT	*	*	*	*
NUMBER_PROM_12		*		*
PEP_STAR	*	*	*	*
RECENT_AVG_GIFT_AMT	*	*	*	*
RECENT_CARD_RESPONSE_COUNT			*	
RECENT_CARD_RESPONSE_PROP	*	*	*	
SES	*		*	
URBANICITY		*		*
Number of Variables	12	14	13	12

Available in SAS 9.4 and SAS Enterprise Guide 6.1 or higher

SAS® Enterprise Guide®

Logistic Regression – Use SELECTION=SCORE

Right Mouse Click on Logistic Regression Node



Change to SELECTION=SCORE

```
PROC LOGISTIC DATA=WORK.SORTTempTableSorted
        PLOTS(ONLY)=ALL
    ;
    CLASS M_WEALTH_RATING    (PARAM=EFFECT) M_INCOME_GROUP    (PARAM=EFFECT)
        HOME_OWNER   (PARAM=EFFECT) ;
    MODEL TARGET_B (Event = '1')=MONTHS_SINCE_ORIGIN IN_HOUSE MOR_HIT_RATE
        SELECTION=SCORE best=1
            LINK=LOGIT
    ;
RUN;
```

Only for numeric variables and 2 level categorical

SAS® Enterprise Guide®

Variable Selection Methods in SAS/Stat Proc Logistic

SELECTION = SCORE BEST=1

Regression Models Selected by Score Criterion

Number of Variables	Score Chi-Square	Variables Included in Model
1	365.4167	FREQUENCY_STATUS_97N
2	446.3751	FREQUENCY_STATUS_97N FILE_CARD_GIFT
3	518.9102	MEDIAN_HOME_VALUE FREQUENCY_STATUS_97N FILE_CARD_GIFT
4	565.0823	MEDIAN_HOME_VALUE FREQUENCY_STATUS_97N MONTHS_SINCE_LAST_GI FILE_CARD_GIFT
5	603.2741	MEDIAN_HOME_VALUE FREQUENCY_STATUS_97N RECENT_CARD_RESPONSE MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G
6	618.9703	MEDIAN_HOME_VALUE PEP_STAR FREQUENCY_STATUS_97N RECENT_CARD_RESPONSE MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G
7	630.1697	MEDIAN_HOME_VALUE PEP_STAR FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G
8	637.4636	MEDIAN_HOME_VALUE PEP_STAR FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_CARD_PROM MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G
9	643.0937	MEDIAN_HOME_VALUE PEP_STAR FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_CARD_PROM MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING
10	648.4047	MEDIAN_HOME_VALUE PEP_STAR FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_CARD_PROM MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING M_WEALTH_RATING
11	653.9334	IN_HOUSE MEDIAN_HOME_VALUE PEP_STAR FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING M_WEALTH_RATING
12	659.3394	IN_HOUSE MEDIAN_HOME_VALUE PEP_STAR FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING
13	662.7758	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PEP_STAR FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING
14	666.3360	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PEP_STAR FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_AVG_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING
15	669.5617	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_AVG_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING
16	671.6003	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PEP_STAR FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_AVG_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
17	674.8360	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_AVG_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
18	676.4634	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_AVG_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
19	677.8259	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_CARD_PROM LIFETIME_AVG_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE

SAS® Enterprise Guide®

Variable Selection Methods in SAS/Stat Proc Logistic

SELECTION = SCORE BEST=1 (continued)

20	679.1298	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_AVG_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
21	680.4874	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_AVG_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
22	681.8672	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
23	683.0187	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
24	683.8157	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
25	684.4014	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
26	684.8966	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
27	685.2634	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING IM_DONOR_AGE PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
28	685.7178	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
29	686.0816	MONTHS_SINCE_ORIGIN IN_HOUSE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM LAST_GIFT_AMT NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE

SAS® Enterprise Guide®

Variable Selection Methods in SAS/Stat Proc Logistic

SELECTION = SCORE BEST=1 (continued)

30	686.3773	MONTHS_SINCE_ORIGIN IN_HOUSE MOR_HIT_RATE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM LAST_GIFT_AMT NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE
31	686.6589	MONTHS_SINCE_ORIGIN IN_HOUSE MOR_HIT_RATE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM LAST_GIFT_AMT NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE M_MONTHS_SINCE_LAST_
32	686.8766	MONTHS_SINCE_ORIGIN IN_HOUSE MOR_HIT_RATE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM LAST_GIFT_AMT NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G FILE_CARD_GIFT IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE M_MONTHS_SINCE_LAST_
33	687.0585	MONTHS_SINCE_ORIGIN IN_HOUSE MOR_HIT_RATE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM LAST_GIFT_AMT CARD_PROM_12 NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G FILE_CARD_GIFT IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE M_MONTHS_SINCE_LAST_
34	687.1814	MONTHS_SINCE_ORIGIN IN_HOUSE MOR_HIT_RATE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_AVG_CARD_GIFT RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_MAX_GIFT_AM LAST_GIFT_AMT CARD_PROM_12 NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G FILE_CARD_GIFT IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE M_MONTHS_SINCE_LAST_
35	687.2716	MONTHS_SINCE_ORIGIN IN_HOUSE MOR_HIT_RATE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_AVG_CARD_GIFT RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_GIFT_RANGE LIFETIME_MAX_GIFT_AM LAST_GIFT_AMT CARD_PROM_12 NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G FILE_CARD_GIFT IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE M_MONTHS_SINCE_LAST_
36	687.3454	MONTHS_SINCE_ORIGIN IN_HOUSE MOR_HIT_RATE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_AVG_CARD_GIFT RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_GIFT_RANGE LIFETIME_MAX_GIFT_AM LAST_GIFT_AMT CARD_PROM_12 NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G FILE_CARD_GIFT IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE IM_MONTHS_SINCE_LAST PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE M_MONTHS_SINCE_LAST_
37	687.4036	MONTHS_SINCE_ORIGIN IN_HOUSE MOR_HIT_RATE MEDIAN_HOME_VALUE PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_RESPONSE_PROP RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_AVG_CARD_GIFT RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_GIFT_RANGE LIFETIME_MAX_GIFT_AM LAST_GIFT_AMT CARD_PROM_12 NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G FILE_CARD_GIFT IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE IM_MONTHS_SINCE_LAST PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE M_MONTHS_SINCE_LAST_
38	687.4314	MONTHS_SINCE_ORIGIN IN_HOUSE MOR_HIT_RATE MEDIAN_HOME_VALUE MEDIAN_HOUSEHOLD_INC PCT_OWNER_OCCUPIED PER_CAPITA_INCOME PEP_STAR RECENT_STAR_STATUS FREQUENCY_STATUS_97N RECENT_RESPONSE_PROP RECENT_AVG_GIFT_AMT RECENT_CARD_RESPONSE RECENT_AVG_CARD_GIFT RECENT_RESPONSE_COUN LIFETIME_CARD_PROM LIFETIME_PROM LIFETIME_GIFT_AMOUNT LIFETIME_GIFT_COUNT LIFETIME_AVG_GIFT_AM LIFETIME_GIFT_RANGE LIFETIME_MAX_GIFT_AM LAST_GIFT_AMT CARD_PROM_12 NUMBER_PROM_12 MONTHS_SINCE_LAST_GI MONTHS_SINCE_FIRST_G FILE_CARD_GIFT IM_WEALTH_RATING IM_INCOME_GROUP IM_DONOR_AGE IM_MONTHS_SINCE_LAST PUBLISHED_PHONE HOME_OWNER M_WEALTH_RATING M_INCOME_GROUP M_DONOR_AGE M_MONTHS_SINCE_LAST_

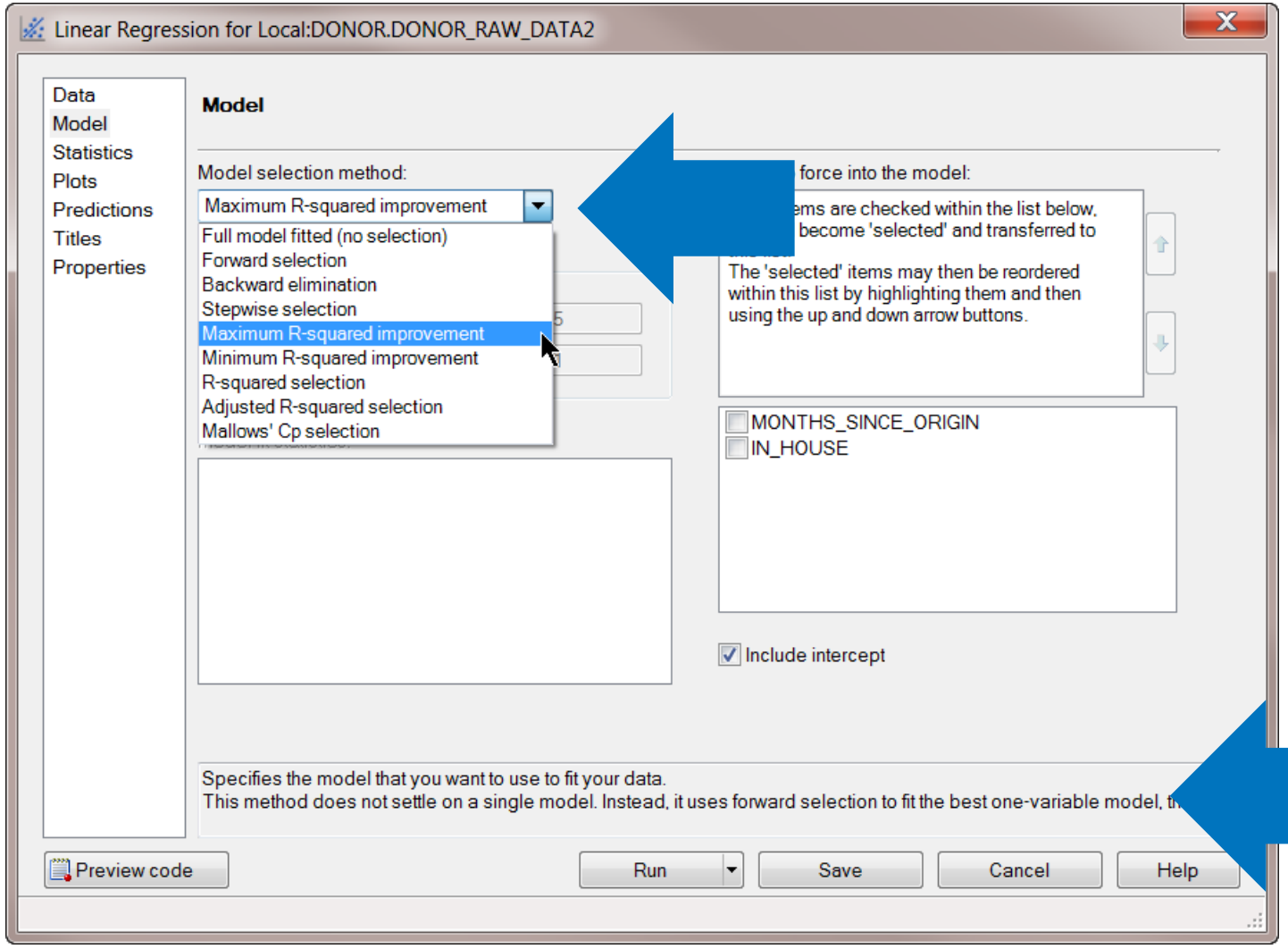
SAS® Enterprise Guide®

Variable Selection Methods in SAS/Stat Proc Logistic & PROC Reg

- PROC LOGISTIC selection methods
- PROC REG selection methods

SAS® Enterprise Guide®

Variable Selection Methods in SAS/Stat Proc REG



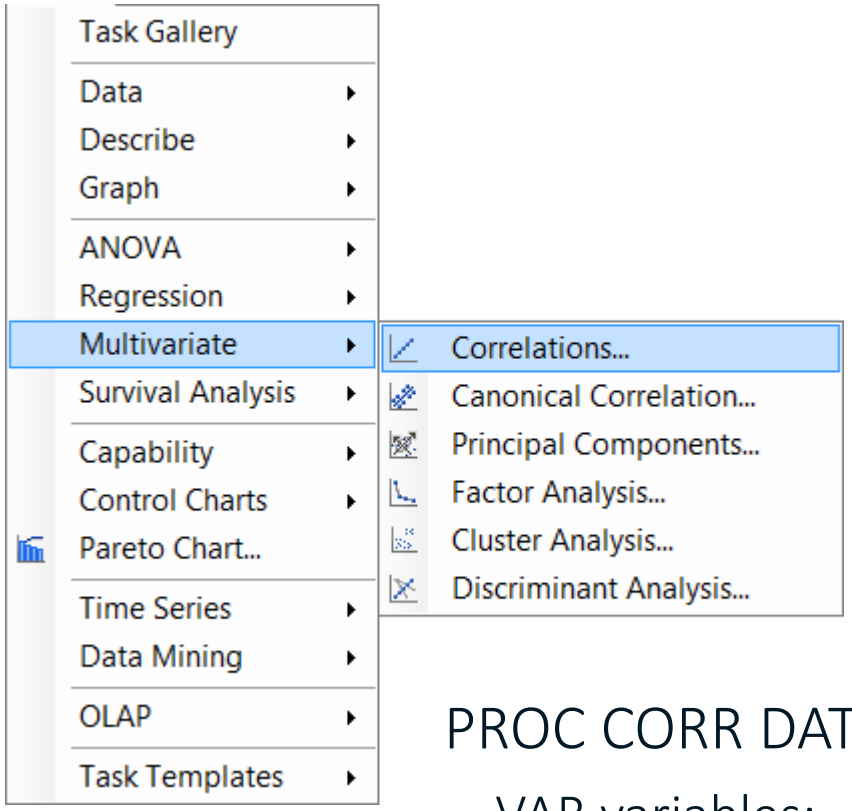
- Forward
- Backward
- Stepwise
- LASSO
- LARS
- MAXR
- MINR
- RSQUARE
- CP
- ADJRSQ
- SCORE

❖ Ones in BLUE available in SAS Enterprise Guide

Only for numeric variables

SAS® Enterprise Guide®

Variable Screening - Correlations



- Are input variables are correlated with Y
- Are variables correlated with each other

PROC CORR DATA=sas-data-set <options>;

VAR variables;

WITH target variable;

RUN;

Only for numeric variables

- Better for smaller datasets
- Becomes more complicated with more variables

SAS® Enterprise Guide®

Variable Screening - Correlations

Inputs correlated with Y or Target Variable

3 Correlation Types

- Pearson
- Spearman
- Hoeffding's D

	NAME	Pearson	HoeffdingsD	Spearman
1	FREQUENCY_ST...	0.1373431429	0.0020876447	0.135237261
2	RECENT_RESPO...	0.1287617475	0.002251795	0.1311443954
3	RECENT_CARD_...	0.1262411497	0.0019356539	0.1244696772
4	RECENT_RESPO...	0.1183428252	0.0017537614	0.1136453199
5	FILE_CARD_GIFT	0.1055518156	0.001714991	0.11346035
6	PEP_STAR	0.1053887583	0.0009170721	0.1053887583
7	RECENT_CARD_...	0.1009018322	0.0012633977	0.099171003
8	LIFETIME_GIFT_...	0.1000175219	0.001552333	0.1078897337
9	LIFETIME_PROM	0.0678464083	0.0005245559	0.0653438744
10	MONTHS_SINCE...	0.0665139194	0.0005743495	0.0681586622
11	LIFETIME_CARD...	0.0655853616	0.0004892737	0.0637318776
12	MONTHS_SINCE...	0.0627947298	0.0005144095	0.0652724858
13	MEDIAN_HOME_...	0.0503773695	0.0003425586	0.053581026
14	PER_CAPITA_IN...	0.0415280465	0.0002343356	0.0451772967
15	LIFETIME_GIFT_...	0.0413779133	0.0003856306	0.0562057102
16	IN_HOUSE	0.0409641241	-0.000033493	0.0409641241
17	NUMBER_PROM...	0.0399671882	0.0002309693	0.0365661724
18	CARD_PROM_12	0.038946534	0.0000903353	0.0330159639
19	MEDIAN_HOUSE...	0.0381904599	0.0001140488	0.0356059467
20	IM_WEALTH_RA...	0.0182176236	-5.846586E-6	0.0179018535
21	PCT_OWNER_O...	0.0157195221	-0.000030067	0.0128102791
22	MOR_HIT_RATE	0.0126887628	-5.085864E-6	0.0222912061
23	M_INCOME_GRO...	0.0105345021	-0.000058078	0.0105345021
24	M_WEALTH_RAT...	0.0099354244	-0.000050557	0.0099354244
25	IM_INCOME_GR...	0.0083238843	-0.000027112	0.0084115816
26	IM_DONOR_AGE	0.0080645908	-0.000037835	0.0096928761
27	RECENT_STAR_...	-0.001475228	0.0003268642	0.0655800215
28	PUBLISHED_PH...	-0.003218794	-0.000058081	-0.003218794
29	M_DONOR_AGE	-0.005731774	-0.000061829	-0.005731774
30	M_MONTHS_SIN...	-0.005855749	-0.000077509	-0.005855749
31	LIFETIME_GIFT_...	-0.006354095	0.0002130106	-0.02816312
32	IM_MONTHS_SIN...	-0.010747155	-0.000039109	-0.009734459
33	RECENT_AVG_C...	-0.016934647	0.0003283251	-0.019618095
34	LIFETIME_MAX_...	-0.03698973	0.0014569703	-0.103611422
35	LIFETIME_MIN_G...	-0.062755735	0.0010513765	-0.093529516
36	FILE_AVG_GIFT	-0.067106841	0.0016757138	-0.111206827
37	LIFETIME_AVG_...	-0.067106841	0.0016757138	-0.111206827
38	LAST_GIFT_AMT	-0.068220085	0.0019754045	-0.120494611
39	RECENT_AVG_G...	-0.074667909	0.0017353731	-0.111816464
40	MONTHS_SINCE...	-0.089854283	0.0008106037	-0.081184725

SAS® Enterprise Guide®

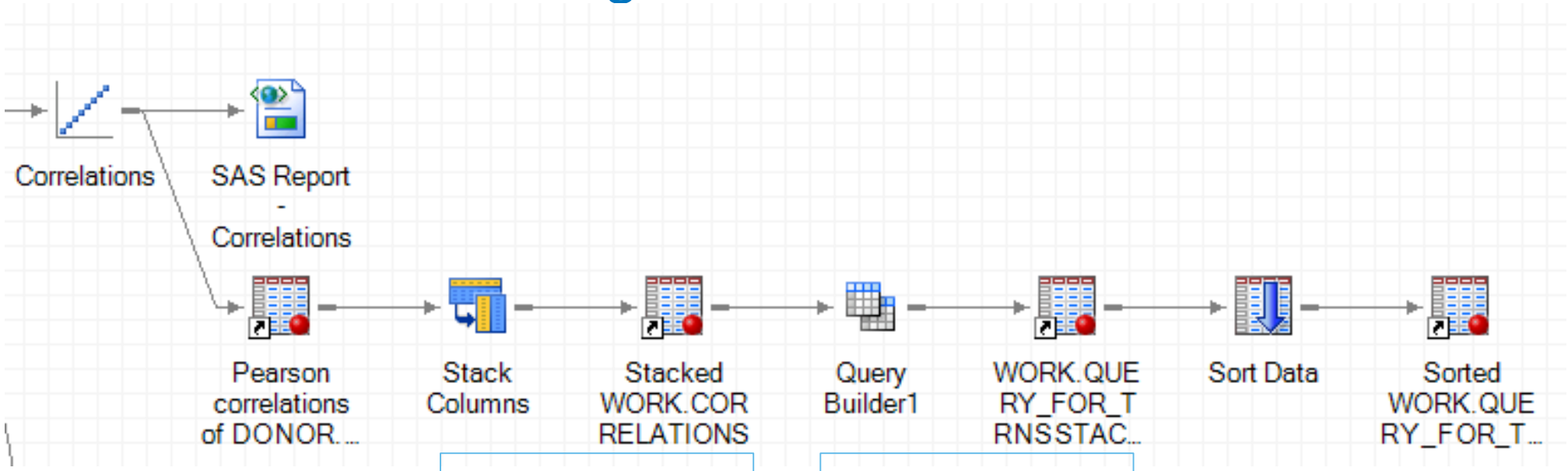
Variable Screening - Correlations

Input Variable Correlations

	MONTHS_SINCE_ORIGIN	IN_HOUSE	PUBLISHED_PHONE	MOR_HIT_RATE	MEDIAN_HOME_VALUE	MEDIAN_HOUSEHOLD_INCOME	PCT_C
MONTHS_SINCE_ORIGIN							
months in database	1.00000	0.15597	0.06562	0.07895	-0.04749	-0.03709	
IN_HOUSE							
Donated to In House program	0.15597	1.00000	0.00920	0.09704	0.04250	0.01666	
PUBLISHED_PHONE							
1 if telephone number is published	0.06562	0.00920	1.00000	0.21425	-0.07793	0.02436	
MOR_HIT_RATE							
known times responded to mailed solicitation	0.07895	0.09704	0.21425	1.00000	0.01287	0.04144	
MEDIAN_HOME_VALUE							
median home value (in $100)	-0.04749	0.04250	-0.07793	0.01287	1.00000	0.67968	
MEDIAN_HOUSEHOLD_INCOME							
median household income (in $100)	-0.03709	0.01666	0.02436	0.04144	0.67968	1.00000	
PCT_OWNER_OCCUPIED							
Pct owner-occupied housing in the neighborhood	0.03622	-0.01553	0.07110	0.01567	0.03751	0.44367	
PER_CAPITA_INCOME							
	-0.02553	0.03874	0.02488	0.04834	0.72910	0.81162	
PEP_STAR							
1 for STAR Donors	0.53430	0.10532	0.03205	0.03015	-0.05564	-0.04634	
RECENT_STAR_STATUS							
1 if STAR status last 4 years	0.31867	0.07269	0.02900	0.02499	-0.02784	-0.02081	
FREQUENCY_STATUS_97NK							
Frequency of donations last 12 months	0.05814	0.01491	-0.00044	-0.00296	-0.05875	-0.05426	
RECENT_RESPONSE_PROP							
Proportion responses to card promotions last 4 years	-0.10351	0.00011	-0.02126	-0.02392	-0.05055	-0.05238	
RECENT_AVG_GIFT_AMT							
average donation since 4 years ago	-0.07978	0.06040	-0.03417	0.00346	0.11213	0.09817	
RECENT_CARD_RESPONSE_PROP							
Proportion responses to card promotions	-0.19790	-0.00993	-0.02020	-0.01831	-0.01213	-0.01603	
RECENT_AVG_CARD_GIFT_AMT							
average donation since 4 years ago card promotion	-0.09871	0.05584	-0.02316	0.01324	0.08016	0.06855	

SAS® Enterprise Guide®

List High Correlation Values



Run Correlation Task

Stack Columns (keep _TYPE_= 'CORR')

Filter to NOT BETWEEN -.5 and .5 and NOT=1

Sort to DEDUP other ½ of matrix

SAS® Enterprise Guide®

List High Correlation Values

	NAME	ValueSource	ValueDescription	StackedValues
1	MONTHS_SINCE_LAST_GIFT	NUMBER_PROM_12	number of promotio...	-0.512796693
2	LIFETIME_GIFT_RANGE	RECENT_AVG_GIFT_AMT	average donation si...	0.5064235611
3	MONTHS_SINCE_ORIGIN	LIFETIME_GIFT_AMOUNT	total lifetime donatio...	0.5099871221
4	LIFETIME_PROM	NUMBER_PROM_12	number of promotio...	0.5175747784
5	RECENT_RESPONSE_COUNT	FILE_CARD_GIFT	lifetime average do...	0.5212052628
6	MONTHS_SINCE_FIRST_GIFT	LIFETIME_GIFT_AMOUNT	total lifetime donatio...	0.5253958563
7	LIFETIME_MIN_GIFT_AMT	LAST_GIFT_AMT	Ampunt most recen...	0.5301850241
8	PEP_STAR	MONTHS_SINCE_ORIGIN	months in database	0.534298978
9	RECENT_CARD_RESPONSE_COU...	FILE_CARD_GIFT	lifetime average do...	0.5364474964
10	RECENT_AVG_CARD_GIFT_AMT	FILE_AVG_GIFT	Same as LIFETIME ...	0.5373453476
11	RECENT_CARD_RESPONSE_PROP	RECENT_RESPONSE_COUNT	Number responses ...	0.5400208064
12	PEP_STAR	MONTHS_SINCE_FIRST_GIFT	months since the fir...	0.5416786665
13	LIFETIME_GIFT_COUNT	RECENT_RESPONSE_COUNT	Number responses ...	0.5431697343
14	LIFETIME_GIFT_AMOUNT	LIFETIME_GIFT_RANGE	Max-Min Donation	0.5525602182
15	FILE_CARD_GIFT	LIFETIME_GIFT_AMOUNT	total lifetime donatio...	0.5588773493
16	PEP_STAR	LIFETIME_PROM	total number of pro...	0.5622346794
17	NUMBER_PROM_12	IN_HOUSE	Donated to In Hous...	0.5751982001
18	PEP_STAR	LIFETIME_CARD_PROM	total number of card...	0.5891379889
19	LIFETIME_CARD_PROM	LIFETIME_GIFT_AMOUNT	total lifetime donatio...	0.6064173248
20	LIFETIME_GIFT_COUNT	PEP_STAR	1 for STAR Donors	0.6189178939
21	RECENT_AVG_CARD_GIFT_AMT	RECENT_AVG_GIFT_AMT	average donation si...	0.6303620135
22	RECENT_CARD_RESPONSE_COU...	FREQUENCY_STATUS_97NK	Frequency of donati...	0.634181588
23	NUMBER_PROM_12	CARD_PROM_12	Number card promo...	0.6443295283
24	LIFETIME_GIFT_AMOUNT	LIFETIME_GIFT_COUNT	total lifetime donatio...	0.6532508992
25	PEP_STAR	FILE_CARD_GIFT	lifetime average do...	0.6640462239
26	LIFETIME_GIFT_AMOUNT	LIFETIME_PROM	total number of pro...	0.6781956969
27	MEDIAN_HOUSEHOLD_INCOME	MEDIAN_HOME_VALUE	median home value...	0.679683473
28	RECENT_RESPONSE_PROP	FREQUENCY_STATUS_97NK	Frequency of donati...	0.7107479281
29	RECENT_AVG_GIFT_AMT	LIFETIME_MAX_GIFT_AMT	maximum donation...	0.7111412073
30	LIFETIME_GIFT_COUNT	MONTHS_SINCE_ORIGIN	months in database	0.7149145498
31	RECENT_RESPONSE_PROP	RECENT_CARD_RESPONSE_...	Number card respo...	0.7173048384
32	PER_CAPITA_INCOME	MEDIAN_HOME_VALUE	median home value...	0.7290997853
33	LIFETIME_GIFT_COUNT	MONTHS_SINCE_FIRST_GIFT	months since the fir...	0.729229891
34	LAST_GIFT_AMT	LIFETIME_MAX_GIFT_AMT	maximum donation...	0.7304301772
35	LIFETIME_MAX_GIFT_AMT	LIFETIME_AVG_GIFT_AMT	lifetime average do...	0.7308441815
36	MONTHS_SINCE_ORIGIN	FILE_CARD_GIFT	lifetime average do...	0.7434230542
37	FILE_CARD_GIFT	LIFETIME_PROM	total number of pro...	0.7450059902
38	MONTHS_SINCE_FIRST_GIFT	FILE_CARD_GIFT	lifetime average do...	0.7511540851
39	RECENT_CARD_RESPONSE_PROP	RECENT_RESPONSE_PROP	Proportion respons...	0.7521659279
40	RECENT_RESPONSE_COUNT	FREQUENCY_STATUS_97NK	Frequency of donati...	0.7708387266
41	LIFETIME_GIFT_COUNT	LIFETIME_CARD_PROM	total number of card...	0.7756398059
42	FILE_CARD_GIFT	LIFETIME_CARD_PROM	total number of card...	0.777063222
43	RECENT_CARD_RESPONSE_PROP	RECENT_CARD_RESPONSE_...	Number card respo...	0.7854865497
44	RECENT_AVG_GIFT_AMT	FILE_AVG_GIFT	Same as LIFETIME ...	0.7912924843
45	LIFETIME_GIFT_COUNT	LIFETIME_PROM	total number of pro...	0.7924570493
46	RECENT_AVG_GIFT_AMT	LAST_GIFT_AMT	Ampunt most recen...	0.8010516369
47	FILE_AVG_GIFT	LAST_GIFT_AMT	Ampunt most recen...	0.8039106569
48	LIFETIME_MIN_GIFT_AMT	LIFETIME_AVG_GIFT_AMT	lifetime average do...	0.8046675407
49	RECENT_RESPONSE_COUNT	RECENT_RESPONSE_PROP	Proportion respons...	0.8058316695
50	PER_CAPITA_INCOME	MEDIAN_HOUSEHOLD_INCOM...	median household i...	0.8116242147
51	RECENT_RESPONSE_COUNT	RECENT_CARD_RESPONSE_...	Number card respo...	0.8369734592
52	MONTHS_SINCE_ORIGIN	LIFETIME_PROM	total number of pro...	0.8603419303
53	MONTHS_SINCE_FIRST_GIFT	LIFETIME_PROM	total number of pro...	0.8707749446
54	LIFETIME_MAX_GIFT_AMT	LIFETIME_GIFT_RANGE	Max-Min Donation	0.8718370577
55	MONTHS_SINCE_ORIGIN	LIFETIME_CARD_PROM	total number of card...	0.9120626898
56	MONTHS_SINCE_FIRST_GIFT	LIFETIME_CARD_PROM	total number of card...	0.9168655202
57	FILE_CARD_GIFT	LIFETIME_GIFT_COUNT	total lifetime donatio...	0.9183076853
58	LIFETIME_CARD_PROM	LIFETIME_PROM	total number of pro...	0.9486531828
59	MONTHS_SINCE_ORIGIN	MONTHS_SINCE_FIRST_GIFT	months since the fir...	0.9878245914
60	LIFETIME_AVG_GIFT_AMT	FILE_AVG_GIFT	Same as LIFETIME ...	1

SAS® Enterprise Guide®

Variable Clustering

- Finds groups of variables that are as correlated as possible with each other
- And as uncorrelated as possible with other variables

```
PROC VARCLUS DATA=sas-data-set<options>;
   VAR variables;
   RUN;
```

Only for numeric variables

SAS® Enterprise Guide®

Variable Clustering

11 Clusters		R-squared with			
Cluster	Variable	Own Cluster	Next Closest	1-R**2 Ratio	Variable Label
Cluster 1	MONTHS_SINCE_ORIGIN	0.8330	0.0564	0.1770	months in database
	PEP_STAR	0.4911	0.1387	0.5909	1 for STAR Donors
	RECENT_STAR_STATUS	0.1795	0.0097	0.8285	1 if STAR status last 4 years
	LIFETIME_CARD_PROM	0.8982	0.1196	0.1156	total number of card promotions sent
	LIFETIME_PROM	0.8757	0.2476	0.1652	total number of promotions
	LIFETIME_GIFT_AMOUNT	0.4863	0.2765	0.7101	total lifetime donation amount (in $)
	LIFETIME_GIFT_COUNT	0.7947	0.1731	0.2483	total lifetime donation count
	MONTHS_SINCE_FIRST_GIFT	0.8483	0.0582	0.1610	months since the first donation
	FILE_CARD_GIFT	0.7886	0.2018	0.2648	lifetime average donation (in $)
Cluster 2	FILE_AVG_GIFT	0.9426	0.3148	0.0838	Same as LIFETIME_AVG_GIFT_AMT
	RECENT_AVG_GIFT_AMT	0.7694	0.3960	0.3818	average donation since 4 years ago
	RECENT_AVG_CARD_GIFT_AMT	0.4326	0.1948	0.7047	average donation since 4 years ago card promotion
	LIFETIME_AVG_GIFT_AMT	0.9426	0.3148	0.0838	lifetime average donation (in $)
	LIFETIME_MIN_GIFT_AMT	0.5927	0.1886	0.5020	minimum donation amount (in $)
	LAST_GIFT_AMT	0.7580	0.4036	0.4057	Ampunt most recent donation
Cluster 3	FREQUENCY_STATUS_97NK	0.6765	0.1341	0.3736	Frequency of donations last 12 months
	RECENT_RESPONSE_PROP	0.8340	0.0666	0.1779	Proportion responses to card promotions last 4 years
	RECENT_CARD_RESPONSE_PROP	0.6596	0.0281	0.3503	Proportion responses to card promotions
	RECENT_RESPONSE_COUNT	0.8270	0.1379	0.2006	Number responses last 4 years
	RECENT_CARD_RESPONSE_COUNT	0.8301	0.0886	0.1864	Number card responses last 4 years
Cluster 4	MEDIAN_HOME_VALUE	0.7729	0.0153	0.2306	median home value (in $100)
	MEDIAN_HOUSEHOLD_INCOME	0.8369	0.1023	0.1817	median household income (in $100)
	PER_CAPITA_INCOME	0.8719	0.0261	0.1316	
Cluster 5	IN_HOUSE	0.4705	0.0624	0.5647	Donated to In House program
	MOR_HIT_RATE	0.0139	0.0247	1.0110	known times responded to mailed solicitation
	CARD_PROM_12	0.5194	0.0782	0.5214	Number card promotions last 12 months
	NUMBER_PROM_12	0.8432	0.0976	0.1738	number of promotions last 12 months
	MONTHS_SINCE_LAST_GIFT	0.4829	0.0670	0.5542	months since the most recent donation
Cluster 6	M_INCOME_GROUP	0.5483	0.0021	0.4527	Imputation Indicator: 7 income groups
	M_DONOR_AGE	0.7114	0.0273	0.2967	Imputation Indicator: age as of last year's mail solicitation
	IM_DONOR_AGE	0.3645	0.0220	0.6498	age as of last year's mail solicitation
Cluster 7	LIFETIME_GIFT_RANGE	0.9359	0.1468	0.0751	Max-Min Donation
	LIFETIME_MAX_GIFT_AMT	0.9359	0.5481	0.1418	maximum donation amount (in $)
Cluster 8	IM_MONTHS_SINCE_LAST_PROM_RESP	0.6190	0.0027	0.3820	months since respose to a promotion
	M_MONTHS_SINCE_LAST_PROM_RESP	0.6190	0.0027	0.3820	Imputation Indicator: months since respose to a promotion
Cluster 9	IM_WEALTH_RATING	0.5898	0.0076	0.4133	10 possible wealth rating groups
	IM_INCOME_GROUP	0.5898	0.0025	0.4112	7 income groups
Cluster 10	PUBLISHED_PHONE	0.5355	0.0024	0.4656	1 if telephone number is published
	PCT_OWNER_OCCUPIED	0.5355	0.0658	0.4672	Pct owner occupied housing in the neighborhood
Cluster 11	M_WEALTH_RATING	1.0000	0.0231	0.0000	Imputation Indicator: 10 possible wealth rating groups

Inter-Cluster Correlations											
Cluster	1	2	3	4	5	6	7	8	9	10	11
1	1.00000	-0.22469	0.20509	-0.04047	0.32176	-0.02161	0.14909	0.00665	-0.00118	0.05487	-0.09834
2	-0.22469	1.00000	-0.30429	0.12904	0.00753	-0.00159	0.58067	0.00974	0.03033	-0.03752	0.00302
3	0.20509	-0.30429	1.00000	-0.05016	0.20698	0.03171	-0.16355	0.00608	-0.01261	-0.00531	0.00970
4	-0.04047	0.12904	-0.05016	1.00000	0.03052	-0.02237	0.07137	-0.01111	0.06588	0.16890	-0.00952
5	0.32176	0.00753	0.20698	0.03052	1.00000	-0.06841	0.17195	-0.05655	-0.00748	0.00557	-0.12039
6	-0.02161	-0.00159	0.03171	-0.02237	-0.06841	1.00000	-0.00719	-0.00198	-0.01628	-0.03637	0.15204
7	0.14909	0.58067	-0.16355	0.07137	0.17195	-0.00719	1.00000	0.00370	0.02402	-0.01095	-0.03083
8	0.00665	0.00974	0.00608	-0.01111	-0.05655	-0.00198	0.00370	1.00000	0.00194	0.01257	-0.04989
9	-0.00118	0.03033	-0.01261	0.06588	-0.00748	-0.01628	0.02402	0.00194	1.00000	0.01943	-0.04601
10	0.05487	-0.03752	-0.00531	0.16890	0.00557	-0.03637	-0.01095	0.01257	0.01943	1.00000	-0.00356
11	-0.09834	0.00302	0.00970	-0.00952	-0.12039	0.15204	-0.03083	-0.04989	-0.04601	-0.00356	1.00000

Number of Clusters	Total Variation Explained by Clusters	Proportion of Variation Explained by Clusters	Minimum Proportion Explained by a Cluster	Maximum Second Eigenvalue in a Cluster	Minimum R-squared for a Variable	Maximum 1-R**2 Ratio for a Variable
1	8.030593	0.2008	0.2008	5.691518	0.0000	
2	12.778018	0.3195	0.3094	3.159679	0.0004	0.9999
3	15.814253	0.3954	0.3296	2.420685	0.0004	1.0224
4	18.232069	0.4558	0.3296	2.170017	0.0004	1.0694
5	20.232465	0.5058	0.2658	1.644129	0.0015	1.0692
6	21.851074	0.5463	0.4201	1.434075	0.0015	1.0692
7	22.843821	0.5711	0.4201	1.238243	0.0015	1.0692
8	24.076294	0.6019	0.4201	1.170466	0.0015	1.0692
9	25.246192	0.6312	0.4201	1.073900	0.0015	1.0689
10	26.313209	0.6578	0.4201	1.008122	0.0139	1.0110
11	27.256888	0.6814	0.4660	1.005492	0.0139	1.0110
12	28.248913	0.7062	0.5355	0.928905	0.1795	0.8285

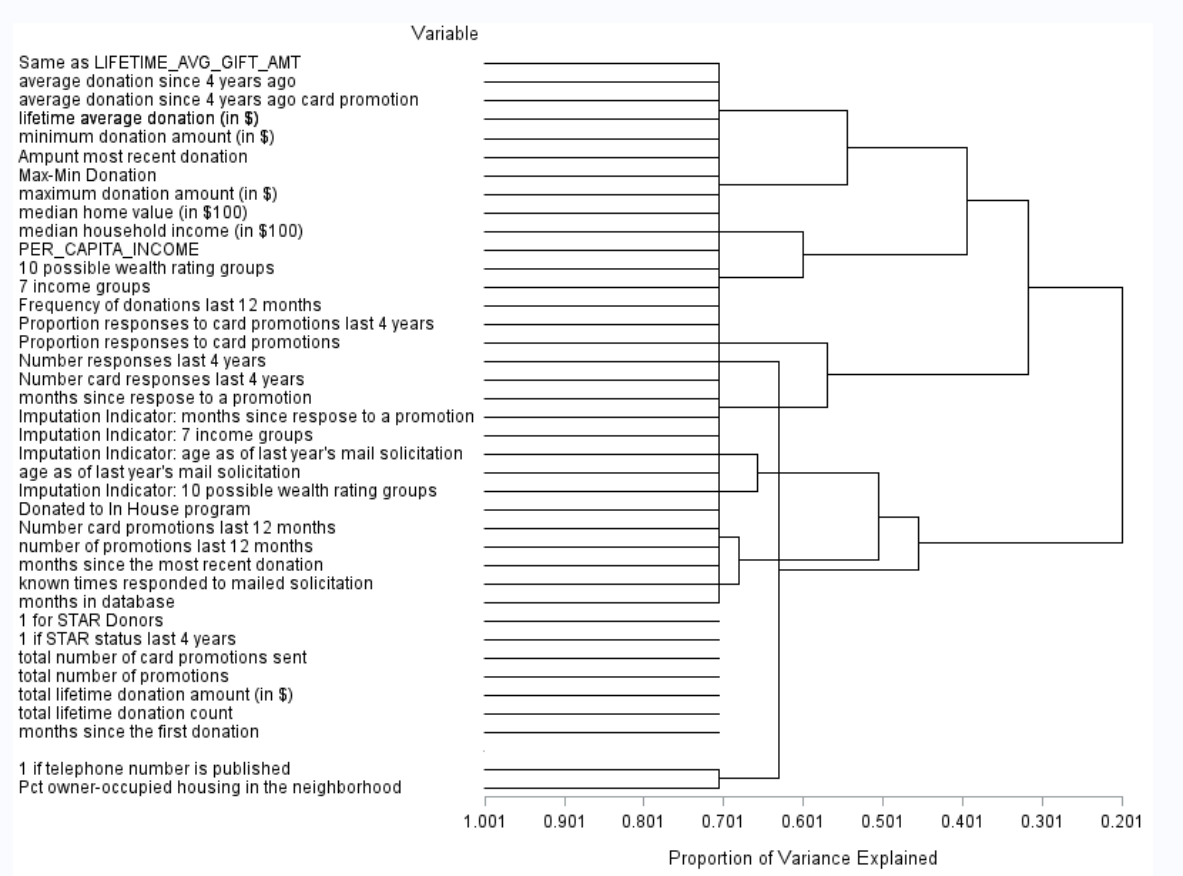
Select one variable from each cluster. If the cluster has several variables you can select multiple.

SAS® Enterprise Guide®

Variable Clustering

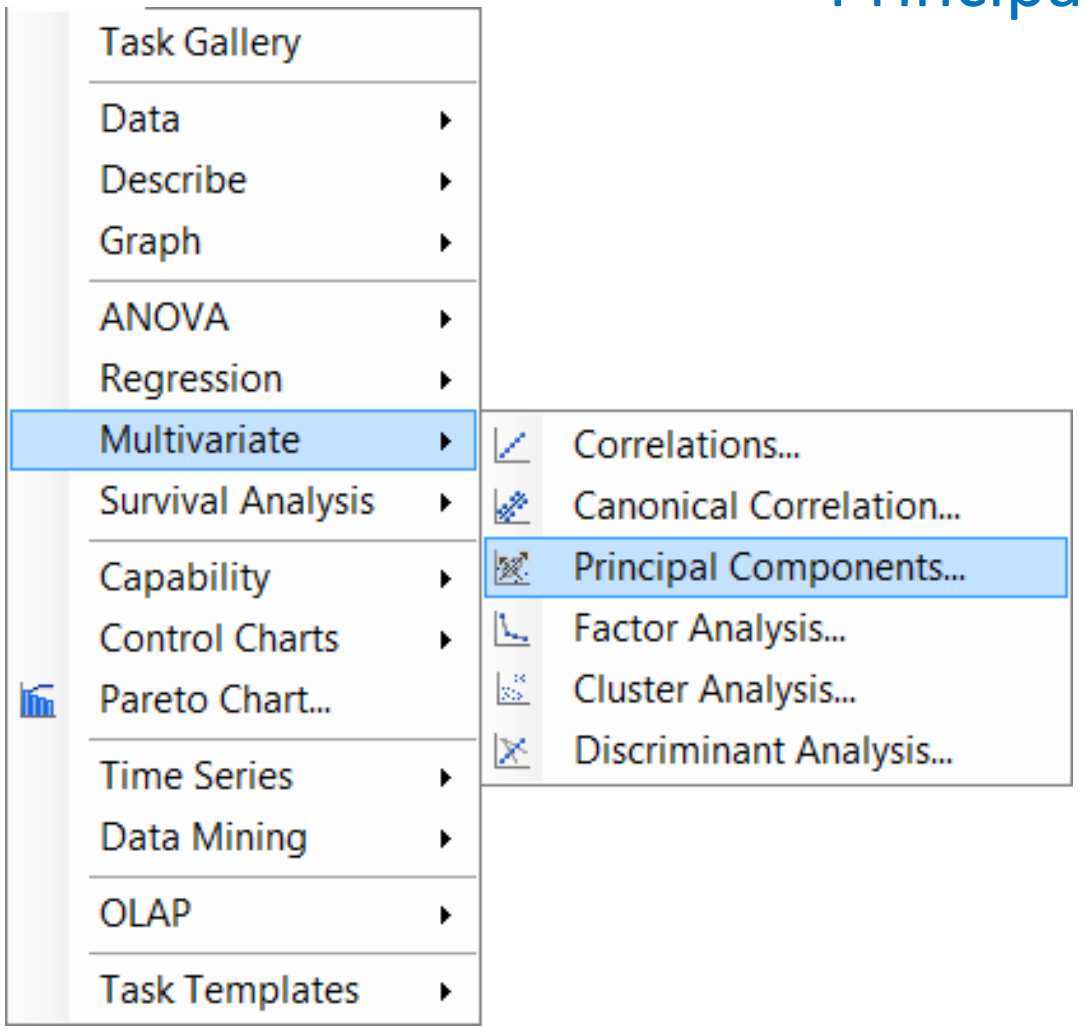
```
PROC VARCLUS DATA=sas-data-set outtree=tree;
   VAR variables;
   RUN;

PROC TREE horizontal haxis=axis1 vaxis=axis2;
     height _propor_;
     id _label_;
  run;
```

Using PROC TREE you can output a tree diagram of the Variable Clusters



SAS® Enterprise Guide®

Principal Components



- Uses all numeric variables
- Hard to interpret individual variables
- Called variable reduction or dimension reduction

Principal Component Analysis Chapter

Only for numeric variables

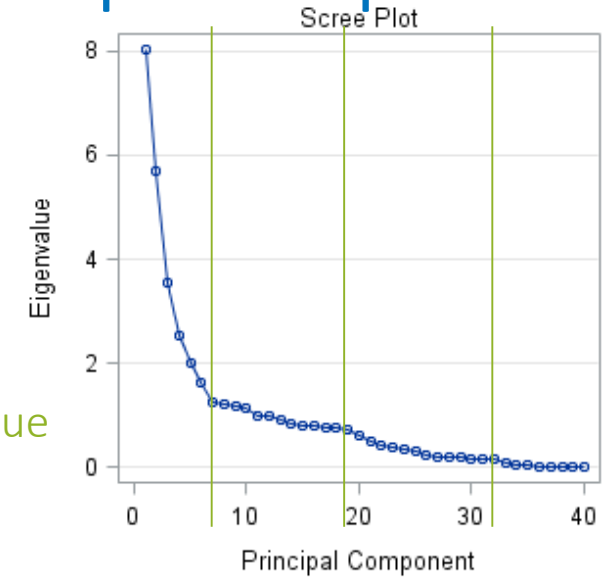
SAS® Enterprise Guide®

Principal Components

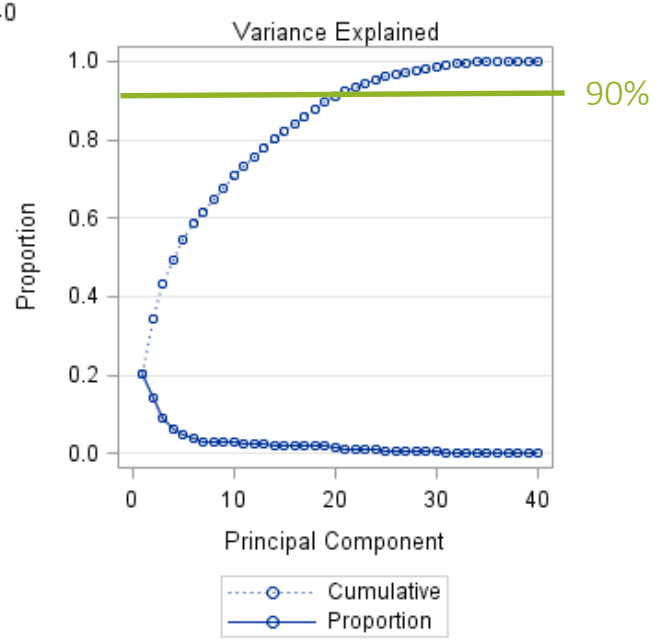
Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	8.03059329	2.33907514	0.2008	0.2008
2	5.69151815	2.14230954	0.1423	0.3431
3	3.54920861	1.00818165	0.0887	0.4318
4	2.54102696	0.53339914	0.0635	0.4953
5	2.00762781	0.38263517	0.0502	0.5455
6	1.62499265	0.36996067	0.0406	0.5861
7	1.25503198	0.03232792	0.0314	0.6175
8	1.22270406	0.02610604	0.0306	0.6481
9	1.19659802	0.03691036	0.0299	0.6780
10	1.15968766	0.15081975	0.0290	0.7070
11	1.00886792	0.01707500	0.0252	0.7322
12	0.99179291	0.08172904	0.0248	0.7570
13	0.91006387	0.06967036	0.0228	0.7797
14	0.84039351	0.02634725	0.0210	0.8008
15	0.81404626	0.02264887	0.0204	0.8211
16	0.79139739	0.02011892	0.0198	0.8409
17	0.77127847	0.01568968	0.0193	0.8602
18	0.75558879	0.03793811	0.0189	0.8791
19	0.71765069	0.09587683	0.0179	0.8970
20	0.62177385	0.13898703	0.0155	0.9125
21	0.48278683	0.04775995	0.0121	0.9246
22	0.43502687	0.04430503	0.0109	0.9355
23	0.39072185	0.03156800	0.0098	0.9453
24	0.35915385	0.06097790	0.0090	0.9542
25	0.29817595	0.04425927	0.0075	0.9617
26	0.25391667	0.04493867	0.0063	0.9680
27	0.20897800	0.00932102	0.0052	0.9733
28	0.19965698	0.01578272	0.0050	0.9783

Eigenvalue > 1

Proportion of Variance Explained > 90%



1. Eigenvalue - one criterion (keep any with eigenvalue >1)
2. The Scree Test (break between components)
3. Total Variance Explained (For example > 90%)



SAS® Enterprise Guide®

Weight of Evidence (WOE) and Information Value (IV)

Weight of Evidence (WoE) is a measure of how much an attribute in the data is related to the outcome.

$$WoE = [\ln(\frac{Relative\ Freq\ of\ Not\ Donating_I}{Relative\ Freq\ of\ Donating_i})\} * 100$$

Information Value (IV) is used to compare predictive power among variables.

$$IV = \sum(Relative\ Freq\ of\ Not\ Donating_i - Relative\ Freq\ of\ Donating_i) * \text{WOE}$$

Only for numeric variables

SAS® Enterprise Guide®

Weight of Evidence (WOE) and Information Value (IV)

```
proc hpbin data=sas-data-set numbin=5;
    input age/numbin=4;
    input all other variables;
    ods output Mapping=Mapping;
 run;
proc hpbin data=sas-data-set WOE BINS_META=Mapping;
    target Y/level=nominal order=desc;
run;
```

PROC HPBIN Documentation

YouTube Video on HPBIN

Available in SAS 9.4

SAS® Enterprise Guide®

Weight of Evidence (WOE) & Information Value

Weight of Evidence

Variable	Binned Variable	Range	Non-event Count	Non-event Rate	Event Count	Event Rate	Weight of Evidence	Information Value
MONTHS_SINCE_ORIGIN	BIN_MONTHS_SINCE_ORIGIN		0	0	0	0	0	0
		MONTHS_SINCE_ORIGIN < 31.400000	3931	0.78888	1052	0.21112	0.21959	0.01171
		31.400000 <= MONTHS_SINCE_ORIGIN < 57.800000	2440	0.75378	797	0.24622	0.02029	0.0000684
		57.800000 <= MONTHS_SINCE_ORIGIN < 84.200000	2430	0.75536	787	0.24464	0.02881	0.0001368
		84.200000 <= MONTHS_SINCE_ORIGIN < 110.600000	1973	0.73101	726	0.26899	-0.09885	0.00139
		110.600000 <= MONTHS_SINCE_ORIGIN	3755	0.71715	1481	0.28285	-0.16824	0.00797
IN_HOUSE	BIN_IN_HOUSE		0	0	0	0	0	0
		IN_HOUSE < 0.200000	13555	0.75498	4399	0.24502	0.02677	0.0006595
		0.200000 <= IN_HOUSE < 0.400000	0	0	0	0	0	0
		0.400000 <= IN_HOUSE < 0.600000	0	0	0	0	0	0
		0.600000 <= IN_HOUSE < 0.800000	0	0	0	0	0	0
		0.800000 <= IN_HOUSE	974	0.68688	444	0.31312	-0.31303	0.00771
PUBLISHED_PHONE	BIN_PUBLISHED_PHONE		0	0	0	0	0	0
		PUBLISHED_PHONE < 0.200000	7284	0.74861	2446	0.25139	-0.00739	0.0000275
		0.200000 <= PUBLISHED_PHONE < 0.400000	0	0	0	0	0	0
		0.400000 <= PUBLISHED_PHONE < 0.600000	0	0	0	0	0	0
		0.600000 <= PUBLISHED_PHONE < 0.800000	0	0	0	0	0	0
		0.800000 <= PUBLISHED_PHONE	7245	0.75140	2397	0.24860	0.00748	0.0000278
MOR_HIT_RATE	BIN_MOR_HIT_RATE		0	0	0	0	0	0
		MOR_HIT_RATE < 48.200000	14490	0.74984	4834	0.25016	-0.0008278	6.83721E-7
		48.200000 <= MOR_HIT_RATE < 96.400000	27	0.84375	5	0.15625	0.58779	0.0004855
		96.400000 <= MOR_HIT_RATE < 144.600000	0	0	0	0	0	0
		144.600000 <= MOR_HIT_RATE < 192.800000	0	0	0	0	0	0
		192.800000 <= MOR_HIT_RATE	12	0.75000	4	0.25000	0	0
MEDIAN_HOME_VALUE	BIN_MEDIAN_HOME_VALUE		0	0	0	0	0	0
		MEDIAN_HOME_VALUE < 1200.000000	10939	0.76113	3433	0.23887	0.06029	0.00266
		[illegible]						

Obs	Variable	IV
1	FREQUENCY_STATUS_97NK	0.09772
2	RECENT_CARD_RESPONSE_COUNT	0.07572
3	RECENT_RESPONSE_COUNT	0.07412
4	PEP_STAR	0.05982
5	RECENT_RESPONSE_PROP	0.05098
6	RECENT_CARD_RESPONSE_PROP	0.05052
7	MONTHS_SINCE_LAST_GIFT	0.04140
8	FILE_CARD_GIFT	0.03469
9	LIFETIME_GIFT_COUNT	0.02322
10	LIFETIME_CARD_PROM	0.02277
11	MONTHS_SINCE_ORIGIN	0.02128
12	CARD_PROM_12	0.02043
13	LIFETIME_PROM	0.01981
14	MONTHS_SINCE_FIRST_GIFT	0.01981
15	NUMBER_PROM_12	0.01597
16	MEDIAN_HOME_VALUE	0.01135
17	IN_HOUSE	0.00837
18	MEDIAN_HOUSEHOLD_INCOME	0.00725
19	RECENT_STAR_STATUS	0.00336
20	IM_INCOME_GROUP	0.00258
21	IM_WEALTH_RATING	0.00232
22	PCT_OWNER_OCCUPIED	0.00219
23	PER_CAPITA_INCOME	0.00219
24	LIFETIME_GIFT_AMOUNT	0.00199
25	LIFETIME_MIN_GIFT_AMT	0.00132
26	RECENT_AVG_CARD_GIFT_AMT	0.0008572
		0.0007176
		0.0007131
	[illegible]OM_RESP	0.0006711
		0.0006078

[Variable Reduction in SAS by Using Weight of Evidence and Information Value]

Information Value	
0.02-0.1	Weak
0.1-0.3	Medium
0.3-0.5	Strong
>0.5	Suspicious

Available in SAS 9.4

SAS® Enterprise Miner™

SAS® Enterprise Miner™

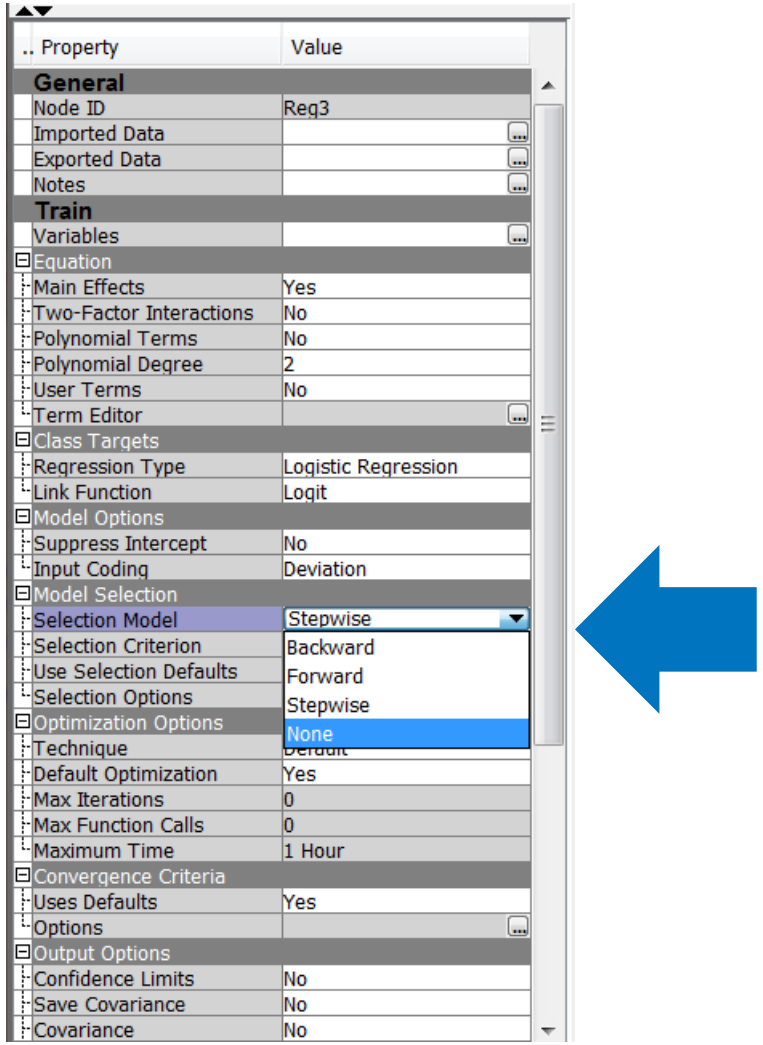
Methods available

- Regression
  - High Performance Regression
- Decision Trees
  - Random Forest
  - High Performance Tree
- Variable Selection
  - Stat Explore
  - Variable Selection
  - LARS/LASSO
  - High Performance Variable Selection
- Variable Clustering
- Principle Components
- Weight of Evidence (WOE)

SAS® Enterprise Miner™

Regression

- Full
- Stepwise
- Backward
- Forward



SAS® Enterprise Miner™

Logistic Regression

Variable Name	Stepwise	Backward	Forward
FREQUENCY_STATUS_97N	*	*	*
LIFETIME_CARD_PROM	*		
MEDIAN_HOME_VALUE	*	*	*
MONTHS_SINCE_FIRST_GIFT	*	*	
MONTHS_SINCE_LAST_GIFT		*	
PEP_STAR	*	*	*
RECENT_AVG_GIFT_AMT		*	
RECENT_CARD_RESPONSE_COUNT	*	*	
CARD_PROM_12	*	*	
IMP_INCOME_GROUP	*	*	
LIFETIME_GIFT_AMOUNT	*		
M_DONOR_AGE	*	*	
IMP_DONOR_AGE	*	*	
Number of Variables	11	11	3

Based on Validation Misclassification Rate – the forward model won

SAS® Enterprise Miner™

HP Regression

- Full
- Stepwise
- Backward
- Forward
- Fast Backward

Only for Interval Targets

- LAR
- LASSO

.. Property	Value
General	
Node ID	HPReg
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Equation	
Main Effects	Yes
Two-Factor Interactions	No
Polynomial Terms	No
Polynomial Degree	2
Suppress Intercept	No
Use Missing as Level	No
Modeling	
Regression Type	Logistic Regression
Link Function	Logit
Optimization Options	
Convergence Options	
Model Selection	
Selection Method	Forward
Selection Criterion	DEFAULT
Stop Criterion	DEFAULT
Selection Options	
Score	
Excluded Variables	Reject
Status	

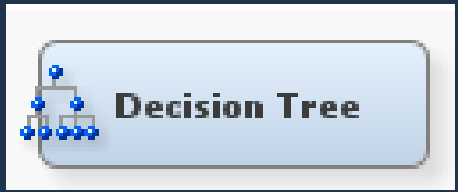
SAS® Enterprise Miner™

HP Regression

Variable Name	Stepwise	Backward	Forward	Fast Backward
FREQUENCY_STATUS_97N	*	*	*	*
MEDIAN_HOME_VALUE	*	*	*	*
MONTHS_SINCE_FIRST_GIFT			*	*
MONTHS_SINCE_LAST_GIFT	*		*	*
PEP_STAR	*	*	*	*
RECENT_AVG_GIFT_AMT				*
RECENT_CARD_RESPONSE_COUNT		*	*	*
CARD_PROM_12	*	*	*	*
IMP_INCOME_GROUP			*	*
LIFETIME_GIFT_AMOUNT			*	
M_DONOR_AGE		*	*	*
IMP_DONOR_AGE			*	*
RECENCY_STATUS_96NK		*		
PER_CAPITA_INCOME		*		*
Number of Variables	5	8	11	**12**

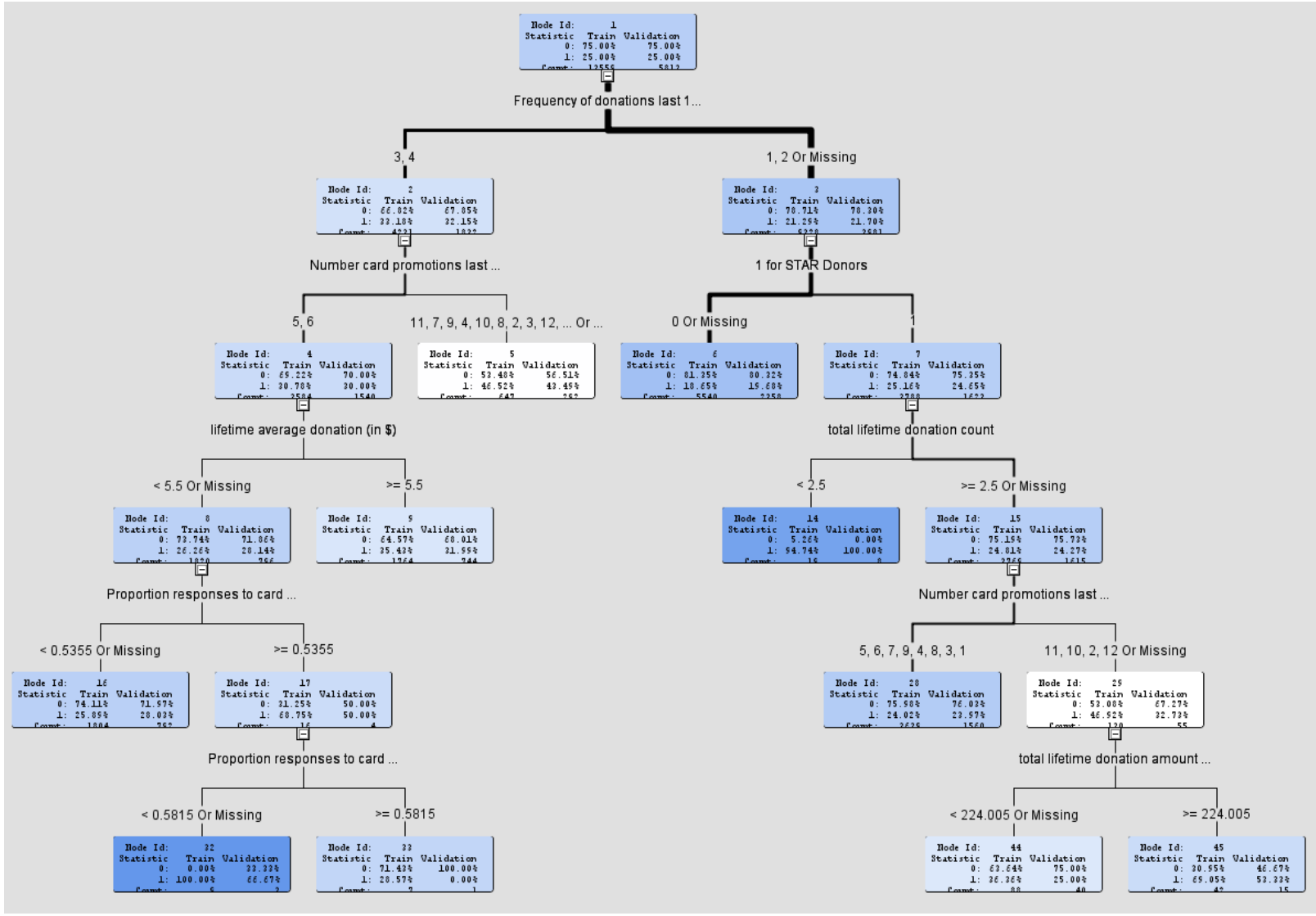
Based on Validation Misclassification Rate – the HP stepwise model won

SAS® Enterprise Miner™



.. Property	Value
Assessment Measure	Decision
Assessment Fraction	0.25
Cross Validation	
Perform Cross Validation	No
Number of Subsets	10
Number of Repeats	1
Seed	12345
Observation Based Importa	
Observation Based Importa	No
Number Single Var Importa	5
P-Value Adjustment	
Bonferroni Adjustment	Yes
Time of Bonferroni Adjustm	Before
Inputs	No
Number of Inputs	1
Depth Adjustment	Yes
Output Variables	
Leaf Variable	Yes
Interactive Sample	
Create Sample	Default
Sample Method	Random
Sample Size	10000
Sample Seed	12345
Performance	Disk
Variable Selection	Yes
Leaf Role	Rejected
Precision	4
Tree Precision	4
Class Target Node Color	Percent Correctly Classified
Interval Target Node Color	Average
Node Text	

Decision Tree



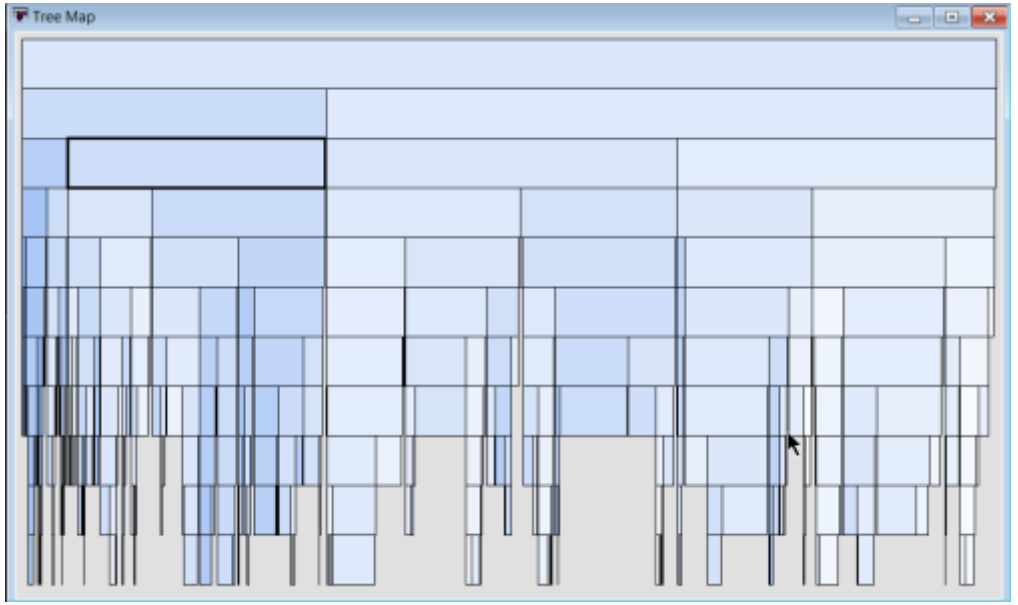
Based on Validation Misclassification Rate

Using Decision Tree for Variable Selection then Regression won

SAS® Enterprise Miner™

HP Tree

.. Property	Value
General	
Node ID	HPTree
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Splitting Rule	
Interval Target Criterion	Variance
Nominal Target Criterion	Entropy
Interval Bins	100
Minimum Distance	0.01
Significance Level	0.2
Bonferroni	No
Missing Values	Largest
Maximum Branch	2
Maximum Depth	10
Minimum Categorical Size	5
Leaf Size	5
Validation	
Create Validation	No
Validation	0.15
Partition Seed	12345
Split Search	
Exhaustive Search Compariso	500000
Fast Search Comparisons	1000000
Subtree	
Subtree Method	Assessment
Confidence	0.25
Nominal Target Assessment	Entropy
Assessment Threshold Value	1.0
Number of Leaves	1
Score	
Variable Selection	Yes
Node and Leaf Role	Segment
Report	
Nominal Target Node Color	Percent of Event
Interval Target Node Color	Average



SAS® Enterprise Miner™

Random Forest

- Predictive Model called a Forest
- Creates Several Trees
- Training Data sampled without replacement
- Input variables sampled

Available in EM 13.1

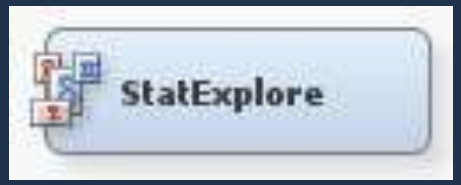
Property	Value
General	
Node ID	HPDMForest
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Tree Options	
Maximum Number of Trees	50
Seed	12345
Type of Sample	Proportion
Proportion of obs in each sam	0.6
Number obs in each sample	.
Splitting Rule Options	
Maximum Depth	50
Missing Values	Use In Search
Minimum Use In Search	1
Number vars to consider in sp	.
Significance Level	0.05
Max Categories in Split Searc	30
Minimum Category Size	5
Exhaustive	5000
Node Options	
Method for Leaf Size	Default
Smallest percentage of obs in	0.001
Smallest number of obs in no	5
Split Size	.
Score	
Variable Selection	Yes
Status	
Create Time	5/4/14 3:46 PM
Run ID	e0556ac7-ef0e-44e8-bdf9-c9
Last Error	
Last Status	Complete

Variable Importance

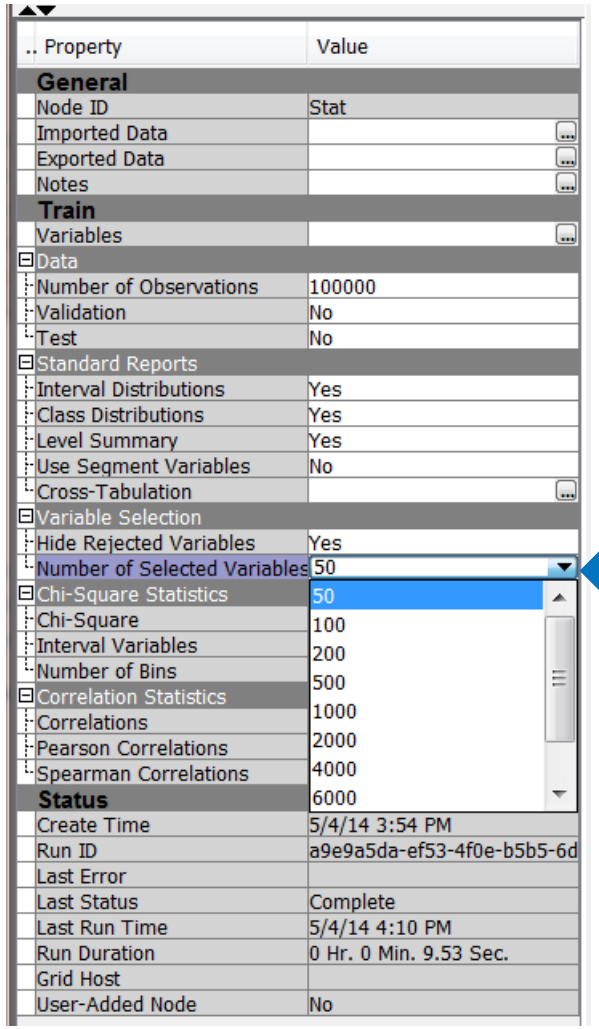
Variable Name	Number of Splitting Rules	Gini Reduction	Margin Reduction	OOB Gini Reduction	OOB Margin Reduction	Label
MONTHS SINCE LAST GIFT	62	0.000985	0.001971	0.00029	0.00097	months
FREQUENCY STATUS 97NK	53	0.001911	0.003822	0.00109	0.00236	Frequen
MEDIAN HOME VALUE	46	0.000578	0.001157	-0.00006	0.00033	median
RECENT RESPONSE PROP	43	0.001544	0.003088	0.00085	0.00183	Proporti
FILE CARD GIFT	36	0.000944	0.001888	0.00045	0.00109	lifetime a
PEP STAR	34	0.000721	0.001442	0.00037	0.000851	for ST
MEDIAN HOUSEHOLD INCOME	29	0.000291	0.000582	-0.00007	0.00007	median
IMP MONTHS SINCE LAST PROM RE	28	0.000366	0.000732	0.00001	0.00027	Imputed
RECENT CARD RESPONSE COUNT	28	0.000661	0.001323	0.00034	0.00086	Number
PER CAPITA INCOME	26	0.000351	0.000703	-0.00009	0.00012	
RECENT CARD RESPONSE PROP	25	0.000555	0.001110	0.00012	0.00049	Proporti
RECENT RESPONSE COUNT	25	0.001303	0.002605	0.00056	0.00140	Number
CARD PROM 12	23	0.000421	0.000842	0.00006	0.00042	Number
NUMBER PROM 12	22	0.000259	0.000517	0.00003	0.00018	number
RECENCY STATUS 96NK	22	0.000273	0.000546	0.00004	0.00021	recency
CLUSTER CODE	18	0.000370	0.000740	-0.00024	0.000065	4 Socio
FILE AVG GIFT	18	0.000324	0.000649	0.00004	0.00026	Same a
LIFETIME GIFT COUNT	18	0.000478	0.000957	0.00027	0.00058	total lifeti
MONTHS SINCE FIRST GIFT	16	0.000277	0.000554	0.00001	0.00016	months
RECENT AVG GIFT AMT	16	0.000224	0.000448	0.00008	0.00018	average
IMP INCOME GROUP	15	0.000160	0.000319	-0.00008	0.00004	Imputed
LIFETIME CARD PROM	15	0.000205	0.000410	-0.00002	0.00011	total nu
LIFETIME MAX GIFT AMT	15	0.000249	0.000499	0.00008	0.00025	maximu
MONTHS SINCE ORIGIN	14	0.000169	0.000337	0.00002	0.00013	months i
MOR HIT RATE	13	0.000085	0.000170	-0.00000	0.00005	known ti
IMP DONOR AGE	12	0.000093	0.000185	-0.00006	-0.00000	Imputed
LAST GIFT AMT	12	0.000162	0.000324	0.00003	0.00014	Amount
LIFETIME AVG GIFT AMT	12	0.000213	0.000425	0.00001	0.00013	lifetime a
RECENT STAR STATUS	12	0.000107	0.000214	-0.00004	0.00001	1 if STA
HOME OWNER	11	0.000054	0.000108	-0.00003	-0.00000	H if hom
IMP WEALTH RATING	11	0.000160	0.000320	-0.00010	-0.00000	Imputed
LIFETIME GIFT AMOUNT	11	0.000105	0.000209	-0.00002	0.00005	total lifeti
URBANICITY	11	0.000097	0.000194	-0.00007	-0.00002	Neighbor
M WEALTH RATING	9	0.000076	0.000151	-0.00000	0.00005	Imputati
SES	9	0.000070	0.000139	-0.00001	0.00004	Socioec
IN HOUSE	8	0.000056	0.000112	-0.00001	0.00003	Donated
LIFETIME PROM	8	0.000086	0.000172	-0.00003	0.00002	total nu
LIFETIME MIN GIFT AMT	5	0.000025	0.000050	-0.00002	-0.00001	minimu
M DONOR AGE	5	0.000022	0.000045	-0.00001	0.00000	Imputati
M INCOME GROUP	5	0.000020	0.000039	-0.00001	-0.00001	Imputati
DONOR GENDER	4	0.000020	0.000040	-0.00002	-0.00000	actual or
LIFETIME GIFT RANGE	4	0.000042	0.000083	-0.00001	0.00002	Max-Min
PUBLISHED PHONE	4	0.000022	0.000043	-0.00001	0.00001	1 if telep
RECENT AVG CARD GIFT AMT	4	0.000034	0.000069	-0.00000	0.00003	average
OVERLAY SOURCE	3	0.000027	0.000053	-0.00001	0.00001	data sou
PCT OWNER OCCUPIED	2	0.000009	0.000017	-0.00000	0.00000	Pct own
M MONTHS SINCE LAST PROM RESP	0	0.000000	0.000000	0.00000	0.00000	Imputati

SAS® Enterprise Miner™

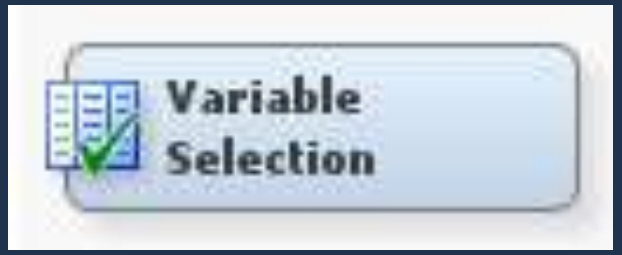
StatExplore



The StatExplore node is a multipurpose node that you use to examine variable distributions and statistics in your data sets. Use the StatExplore node to compute standard univariate statistics, to compute standard bivariate statistics by class target and class segment, and to compute correlation statistics for interval variables by interval input and target. You can also use the StatExplore node to reject variables based on target correlation.



SAS® Enterprise Miner™



Variable Selection

- R-square
- Chi-square
- Both

.. Property	Value
General	
Node ID	Varsel
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Max Class Level	100
Max Missing Percentage	50
Target Model	R and Chi-square
Manual Selector	
Rejects Unused Input	
Bypass Options	
Variable	
Role	
Chi-Square Options	
Number of Bins	50
Maximum Pass Number	6
Minimum Chi-Square	3.84
R-Square Options	
Maximum Variable Number	3000
Minimum R-Square	0.005
Stop R-Square	5.0E-4
Use AOV16 Variables	No
Use Group Variables	Yes
Use Interactions	No
Use SPD Engine Library	Yes
Print Option	Default
Score	
Hides Rejected Variables	Yes
Hides Unused Variables	Yes
Status	

Default
R-Square
Chi-Square
R and Chi-square
None

Variable Name	Reasons for Rejection	Role
CARD_PROM_12	Varsel:Small R-square value, Group variable preferred	Rejected
CLUSTER_CODE	Varsel:Small R-square value, Group variable preferred	Rejected
DONOR_GENDER	Varsel:Small R-square value, Small Chi-square value	Rejected
FILE_AVG_GIFT	Varsel:Small R-square value, Small Chi-square value	Rejected
FILE_CARD_GIFT	Varsel:Small R-square value	Rejected
FREQUENCY_STATUS_97NK		Input
G_CARD_PROM_12		Input
G_CLUSTER_CODE		Input
G_RECENCY_STATUS_96NK		Input
G_RECENT_CARD_RESPONSE_...		Input
G_RECENT_RESPONSE_COUNT		Input
HOME_OWNER	Varsel:Small R-square value, Small Chi-square value	Rejected
IMP_DONOR_AGE	Varsel:Small R-square value	Rejected
IMP_INCOME_GROUP	Varsel:Small R-square value	Rejected
IMP_MONTHS_SINCE_LAST_PRO...	Varsel:Small R-square value, Small Chi-square value	Rejected
IMP_WEALTH_RATING	Varsel:Small R-square value	Rejected
IN_HOUSE	Varsel:Small R-square value, Small Chi-square value	Rejected
LAST_GIFT_AMT	Varsel:Small R-square value	Rejected
LIFETIME_AVG_GIFT_AMT	Varsel:Small R-square value, Small Chi-square value	Rejected
LIFETIME_CARD_PROM	Varsel:Small R-square value, Small Chi-square value	Rejected
LIFETIME_GIFT_AMOUNT	Varsel:Small R-square value, Small Chi-square value	Rejected
LIFETIME_GIFT_COUNT	Varsel:Small R-square value	Rejected
LIFETIME_GIFT_RANGE	Varsel:Small R-square value, Small Chi-square value	Rejected
LIFETIME_MAX_GIFT_AMT	Varsel:Small R-square value, Small Chi-square value	Rejected
LIFETIME_MIN_GIFT_AMT	Varsel:Small R-square value, Small Chi-square value	Rejected
LIFETIME_PROM	Varsel:Small R-square value, Small Chi-square value	Rejected
MEDIAN_HOME_VALUE	Varsel:Small R-square value	Rejected
MEDIAN_HOUSEHOLD_INCOME	Varsel:Small R-square value, Small Chi-square value	Rejected
MONTHS_SINCE_FIRST_GIFT	Varsel:Small R-square value	Rejected
MONTHS_SINCE_LAST_GIFT		Input
MONTHS_SINCE_ORIGIN	Varsel:Small R-square value, Small Chi-square value	Rejected
MOR_HIT_RATE	Varsel:Small R-square value, Small Chi-square value	Rejected
M_DONOR_AGE	Varsel:Small R-square value, Small Chi-square value	Rejected
M_INCOME_GROUP	Varsel:Small R-square value, Small Chi-square value	Rejected
M_MONTHS_SINCE_LAST_PROM...	Varsel:Small R-square value, Small Chi-square value	Rejected
M_WEALTH_RATING	Varsel:Small R-square value, Small Chi-square value	Rejected
NUMBER_PROM_12	Varsel:Small R-square value, Small Chi-square value	Rejected
OVERLAY_SOURCE	Varsel:Small R-square value, Small Chi-square value	Rejected
PCT_OWNER_OCCUPIED	Varsel:Small R-square value, Small Chi-square value	Rejected
PEP_STAR		Input
PER_CAPITA_INCOME	Varsel:Small R-square value, Small Chi-square value	Rejected
PUBLISHED_PHONE	Varsel:Small R-square value	Rejected
RECENCY_STATUS_96NK	Varsel:Small R-square value, Group variable preferred, Smal...	Rejected
RECENT_AVG_CARD_GIFT_AMT	Varsel:Small R-square value, Small Chi-square value	Rejected
RECENT_AVG_GIFT_AMT	Varsel:Small R-square value, Small Chi-square value	Rejected

Effects Chosen for Target: TARGET_B

Effect	DF	R-Square	F Value	p-Value
Group: RECENT_RESPONSE_COUNT	5	0.020671	57.212566	<.0001
Group: CLUSTER_CODE	8	0.006350	11.049862	<.0001
Group: CARD_PROM_12	5	0.005002	13.994789	<.0001
Class: FREQUENCY_STATUS_97NK	3	0.002617	12.230202	<.0001
Class: PEP_STAR	1	0.002884	40.566767	<.0001
Group: RECENT_CARD_RESPONSE_COUNT	5	0.001201	3.382106	0.0047
Var: MONTHS_SINCE_LAST_GIFT	1	0.001234	17.395906	<.0001
Group: RECENCY_STATUS_96NK	3	0.000710	3.335562	0.0185

SAS® Enterprise Miner™

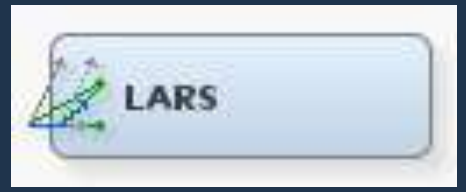
HP Variable Selection

- Fast Selection
- Lar
- Lasso

.. Property	Value
General	
Node ID	HPVS
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Pre-screening	No
Maximum Level	50
Maximum Missing Percent	50
Target Model	Sequential Selection
Unsupervised Selection	
Maximum Steps	200
Maximum Effects	200
Correlation Statistics	Correlation
Cumulative Variance Cutoff	0.99
Incremental Variance Cutoff	0.001
Supervised Selection	
Suppress Intercept	No
Selection Method	Fast Selection
Stop Criterion	SBC
Maximum Steps	200
Maximum Effects	200
Correlation Statistics	Correlation
Cumulative Variance Cutoff	0.99
Incremental Variance Cutoff	0.001
Score	
Hide Rejected Variables	No
Status	

Variable Name	Use	Role ▲	Measurement Level	Reason	Label
DATAOBS		ID	INTERVAL		Observation Number
IMP_INCOME_GROUP		INPUT	NOMINAL		Imputed: 7 income groups
MEDIAN_HOME_VALUE	Y	INPUT	INTERVAL		median home value (in $100)
MONTHS_SINCE_LAST_GIFT	Y	INPUT	INTERVAL		months since the most rece...
PEP_STAR	Y	INPUT	BINARY		1 for STAR Donors
RECENCY_STATUS_96NK	Y	INPUT	NOMINAL		recency status as of two yea...
RECENT_CARD_RESPON...	Y	INPUT	NOMINAL		Number card responses last...
RECENT_RESPONSE_PR...	Y	INPUT	INTERVAL		Proportion responses to car...
CARD_PROM_12	N	REJECTED	NOMINAL	Variance Explained (Supervi...	Number card promotions las...
CLUSTER_CODE	N	REJECTED	NOMINAL	Variance Explained (Supervi...	54 Socioeconomic Clusters
CONTROL_NUMBER	D	REJECTED	NOMINAL		unique identifier of each indiv...
DONOR_GENDER	N	REJECTED	NOMINAL	Variance Explained (Supervi...	actual or inferred gender
FILE_AVG_GIFT	N	REJECTED	INTERVAL	Variance Explained (Supervi...	Same as LIFETIME_AVG_GI...
FILE_CARD_GIFT	N	REJECTED	INTERVAL	Variance Explained (Unsuper...	lifetime average donation (in $)
FREQUENCY_STATUS_97...	N	REJECTED	NOMINAL	Variance Explained (Supervi...	Frequency of donations last ...
HOME_OWNER	N	REJECTED	BINARY	Variance Explained (Supervi...	H if homeowner, U is unknown
IMP_DONOR_AGE	N	REJECTED	INTERVAL	Variance Explained (Supervi...	Imputed: age as of last year'...
IMP_MONTHS_SINCE_LAS...	N	REJECTED	INTERVAL	Variance Explained (Supervi...	Imputed: months since resp...
IMP_WEALTH_RATING	N	REJECTED	NOMINAL	Variance Explained (Supervi...	Imputed: 10 possible wealth ...
IN_HOUSE	N	REJECTED	BINARY	Variance Explained (Supervi...	Donated to In House program
LAST_GIFT_AMT	N	REJECTED	INTERVAL	Variance Explained (Unsuper...	Ampunt most recent donation
LIFETIME_AVG_GIFT_AMT	N	REJECTED	INTERVAL	Variance Explained (Unsuper...	lifetime average donation (in $)
LIFETIME_CARD_PROM	N	REJECTED	INTERVAL	Variance Explained (Unsuper...	total number of card promoti...
LIFETIME_GIFT_AMOUNT	N	REJECTED	INTERVAL	Variance Explained (Unsuper...	total lifetime donation amoun...

SAS® Enterprise Miner™



LARS

.. Property	Value
General	
Node ID	LARS
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Modeling Techniques	
Use Class Inputs	Yes
Intercept	Yes
Variable Selection Method	LAR
Model Selection Criterion	
Path Stopping Criterion	
Maximum Steps	
Cross Validation Options	
Cross Validation	
CV Fold	5
Seed	12345
Reports	
Details	Summary
Score	
Excluded Variables	Reject
Status	
Create Time	5/4/14 3:56 PM
Run ID	b4fa064a-bb43-4664-9106-5c
Last Error	
Last Status	Complete
Last Run Time	5/4/14 4:10 PM
Run Duration	0 Hr. 0 Min. 29.12 Sec.
Grid Host	
User-Added Node	No

LASSO
LAR
Adaptive LASSO
None

- LASSO
- LAR
- Adaptive LASSO
- None

Selected Variables

Effect	Variable	Class Level	Standardized Estimate	Estimate
PEP STAR 0	PEP STAR	0	-0.04071	-0.035260
RECENT RES...	RECENT RES...		0.038388	0.144686
MONTHS SINC...	MONTHS SINC...		-0.03625	-0.003905
FREQUENCY ...	FREQUENCY ...	1	-0.02741	-0.023789
MEDIAN HOME...	MEDIAN HOME...		0.024523	0.000010904
FILE CARD GIFT	FILE CARD GIFT		0.020765	0.001948
RECENT CAR...	RECENT CAR...		0.010539	0.024549
RECENT AVG ...	RECENT AVG ...		-0.00849	-0.000360
RECENT CAR...	RECENT CAR...	5	0.007853	0.019191
IMP INCOME G...	IMP INCOME G...	1	-0.00659	-0.009748
RECENCY STA...	RECENCY STA...	F	-0.00208	-0.003347
INTERCEPT	INTERCEPT		0	0.300415

LAR and LASSO

Selected Variables

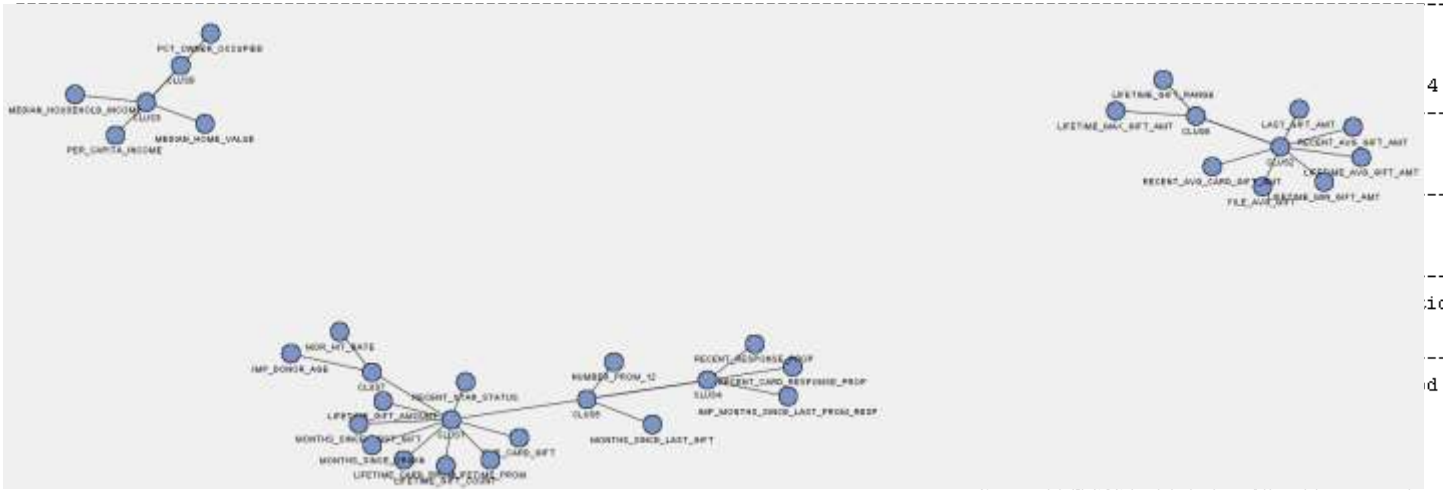
Effect	Variable	Class Level	Standardized Estimate	Estimate
FREQUENCY ...	FREQUENCY ...	1	-0.08367	-0.072607
RECENT CAR...	RECENT CAR...	0	-0.06976	-0.075421
CARD PROM ...	CARD PROM 12	6	-0.05369	-0.046585
RECENT CAR...	RECENT CAR...	1	-0.0513	-0.046918
MONTHS SINC...	MONTHS SINC...		-0.04985	-0.005370
MEDIAN HOME...	MEDIAN HOME...		0.04361	0.000019391
MONTHS SINC...	MONTHS SINC...		0.039524	0.000455
FREQUENCY ...	FREQUENCY ...	2	-0.03862	-0.040377
PEP STAR 0	PEP STAR	0	-0.03616	-0.031313
CARD PROM ...	CARD PROM 12	5	-0.03438	-0.038185
IMP INCOME G...	IMP INCOME G...	1	-0.02599	-0.038459
RECENT CAR...	RECENT CAR...	2	-0.02234	-0.023416
RECENT AVG ...	RECENT AVG ...		-0.00891	-0.000378
INTERCEPT	INTERCEPT		0	0.430865

Adaptive LASSO

SAS® Enterprise Miner™

Variable Clustering

8 Clusters		R-squared with Own Cluster	Next Closest	1-R**2 Ratio	Variable Label
Cluster 1	FILE_CARD_GIFT	0.7748	0.0908	0.2477	lifetime average donation (in $)
	LIFETIME_CARD_PROM	0.9106	0.0575	0.0949	total number of card promotions sent
	LIFETIME_GIFT_AMOUNT	0.5419	0.2295	0.5945	total lifetime donation amount (in $)
	LIFETIME_GIFT_COUNT	0.7920	0.0943	0.2296	total lifetime donation count
	LIFETIME_PROM	0.8928	0.1577	0.1272	total number of promotions
	MONTHS_SINCE_FIRST_GIFT	0.8669	0.0596	0.1415	months since the first donation
	MONTHS_SINCE_ORIGIN	0.8504	0.0577	0.1588	months in database
	RECENT_STAR_STATUS	0.1657	0.0164	0.8482	1 if STAR status last 4 years
Cluster 2	FILE_AVG_GIFT	0.9407	0.2602	0.0801	Same as LIFETIME_AVG_GIFT_AMT
	LAST_GIFT_AMT	0.7725	0.3512	0.3507	Ampunt most recent donation
	LIFETIME_AVG_GIFT_AMT	0.9407	0.2602	0.0801	lifetime average donation (in $)
	LIFETIME_MIN_GIFT_AMT	0.6216	0.1652	0.4533	minimum donation amount (in $)
	RECENT_AVG_CARD_GIFT_AMT	0.4128	0.1812	0.7172	average donation since 4 years ago card promotion
	RECENT_AVG_GIFT_AMT	0.7409	0.3462	0.3963	average donation since 4 years ago
Cluster 3	MEDIAN_HOME_VALUE	0.7724	0.0154	0.2311	median home value (in $100)
	MEDIAN_HOUSEHOLD_INCOME	0.8350	0.1954	0.2051	median household income (in $100)
	PER_CAPITA_INCOME	0.8749	0.0440	0.1308	



Selecting best variable

CLUS1	LIFETIME_CARD_PROM	total number of card promoti...
CLUS2	FILE_AVG_GIFT	Same as LIFETIME_AVG_G...
CLUS3	PER_CAPITA_INCOME	
CLUS4	RECENT_RESPONSE_P...	Proportion responses to car...
CLUS5	NUMBER_PROM_12	number of promotions last 1...
CLUS6	LIFETIME_GIFT_RANGE	Max-Min Donation
CLUS7	MOR_HIT_RATE	known times responded to ...
CLUS8	PCT_OWNER_OCCUPIED	Pct owner-occupied housin...

The **Best Variables** property exports the variables in each cluster that have the minimum R-square ratio values.

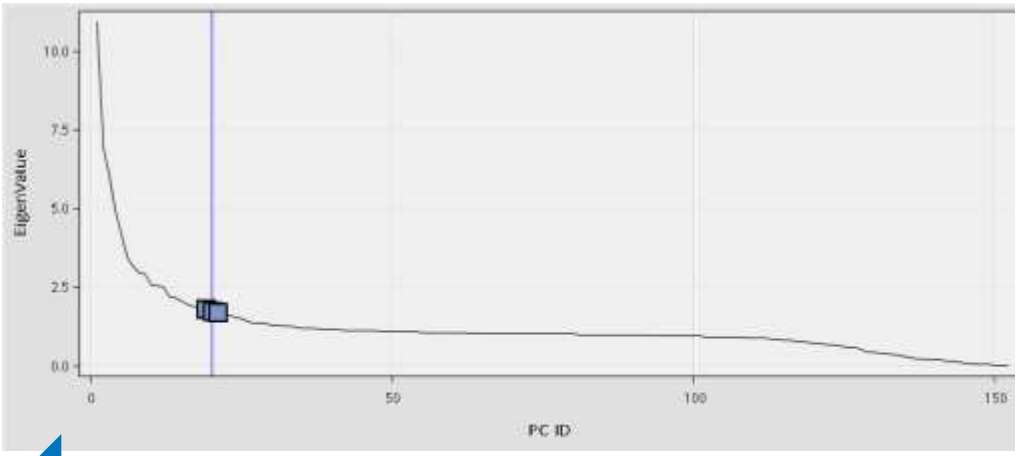
Cluster components

CLUS1	CLUS1	Cluster 1	1CLUS5
CLUS2	CLUS2	Cluster 2	1CLUS6
CLUS3	CLUS3	Cluster 3	1CLUS8
CLUS4	CLUS4	Cluster 4	1CLUS5
CLUS5	CLUS5	Cluster 5	1CLUS4
CLUS6	CLUS6	Cluster 6	1CLUS2
CLUS7	CLUS7	Cluster 7	1CLUS1
CLUS8	CLUS8	Cluster 8	1CLUS3

The **Cluster Component** property exports a linear combination of the variables from each cluster. **Cluster Component** is the default setting for the **Variable Selection** property.

SAS® Enterprise Miner™

Principle Components



.. Property	Value
General	
Node ID	PRINCOMP
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Eigenvalue Source	Correlation
Interactive Selection	Covariance / Correlation / Uncorrected
Print Eigenvalue Source	
Score	
Princomp Prefix	PC
Eigenvalue Cutoff	
Cumulative	0.99
Increment	0.001
Max Number Cutoff	
Apply Maximum Number	Yes
Maximum Number	20
Reject Original Input Variable	Yes
Hide Rejected Variables	Yes

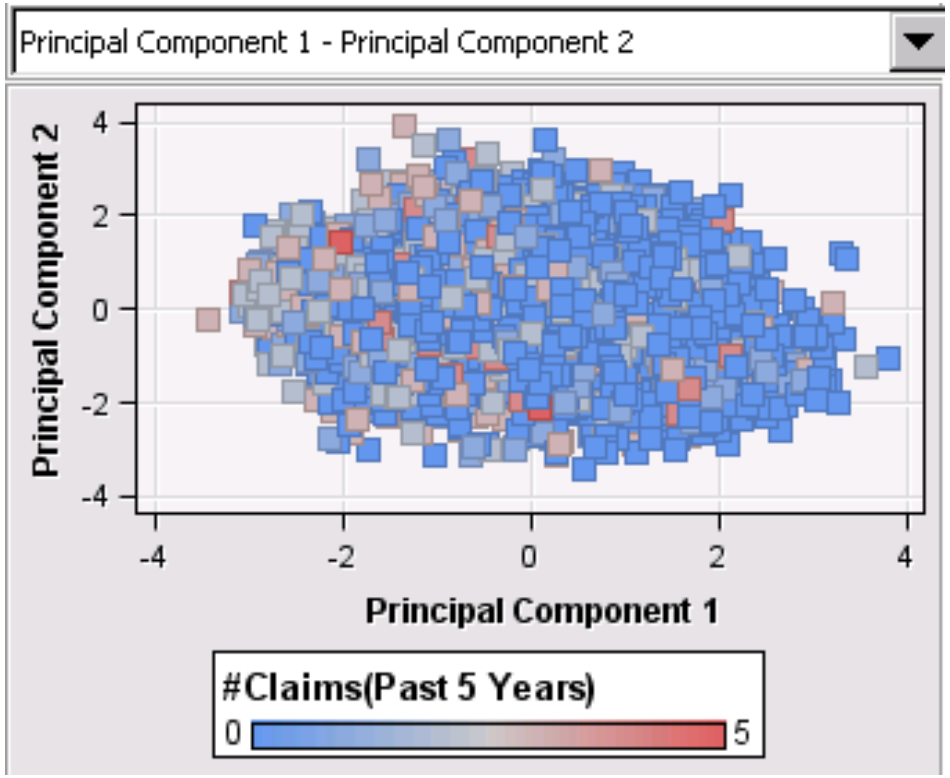
```
*------------------------------------------------------------*
Summary of Exported Principal Components
*------------------------------------------------------------*

Total number of input variables: 47
Maximum number cutoff of principal components: 20
Cumulative proportional eigenvalue cutoff: 0.99
Proportional eigenvalue increment cutoff: 0.001
Number of the selected principal components: 20
Total variation explained by the selected principal components: 0.3735688938
```

PC_1	Input	Interval
PC_10	Input	Interval
PC_11	Input	Interval
PC_12	Input	Interval
PC_13	Input	Interval
PC_14	Input	Interval
PC_15	Input	Interval
PC_16	Input	Interval
PC_17	Input	Interval
PC_18	Input	Interval
PC_19	Input	Interval
PC_2	Input	Interval
PC_20	Input	Interval
PC_3	Input	Interval
PC_4	Input	Interval
PC_5	Input	Interval
PC_6	Input	Interval
PC_7	Input	Interval
PC_8	Input	Interval
PC_9	Input	Interval
TARGET_B	Target	Binary
dataobs	ID	Interval

SAS® ENTERPRISE MINER™

HP Principal Components

Perform principal component analysis for data dimension reduction, a frequent intermediate step in the data mining process



New in SAS® Enterprise Miner™ 13.1

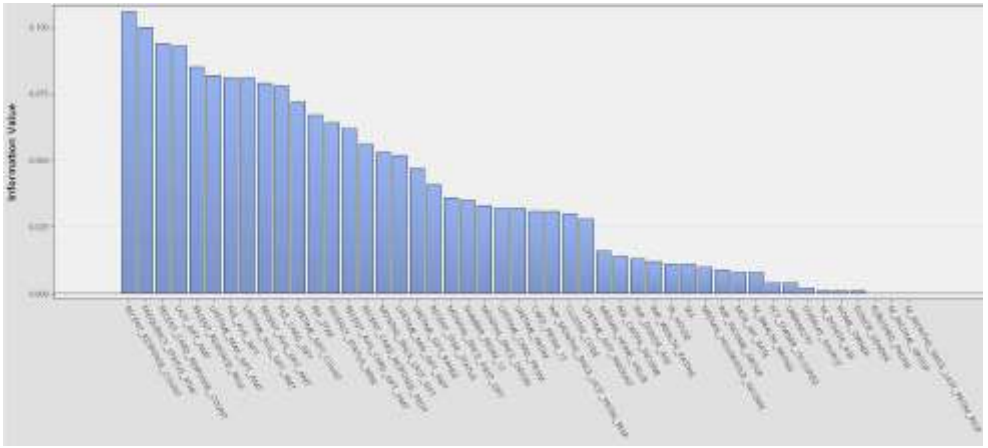
SAS® Enterprise Miner™

Available in Credit Scoring Add-On

Weight of Evidence (WOE) and Information Value (IV)

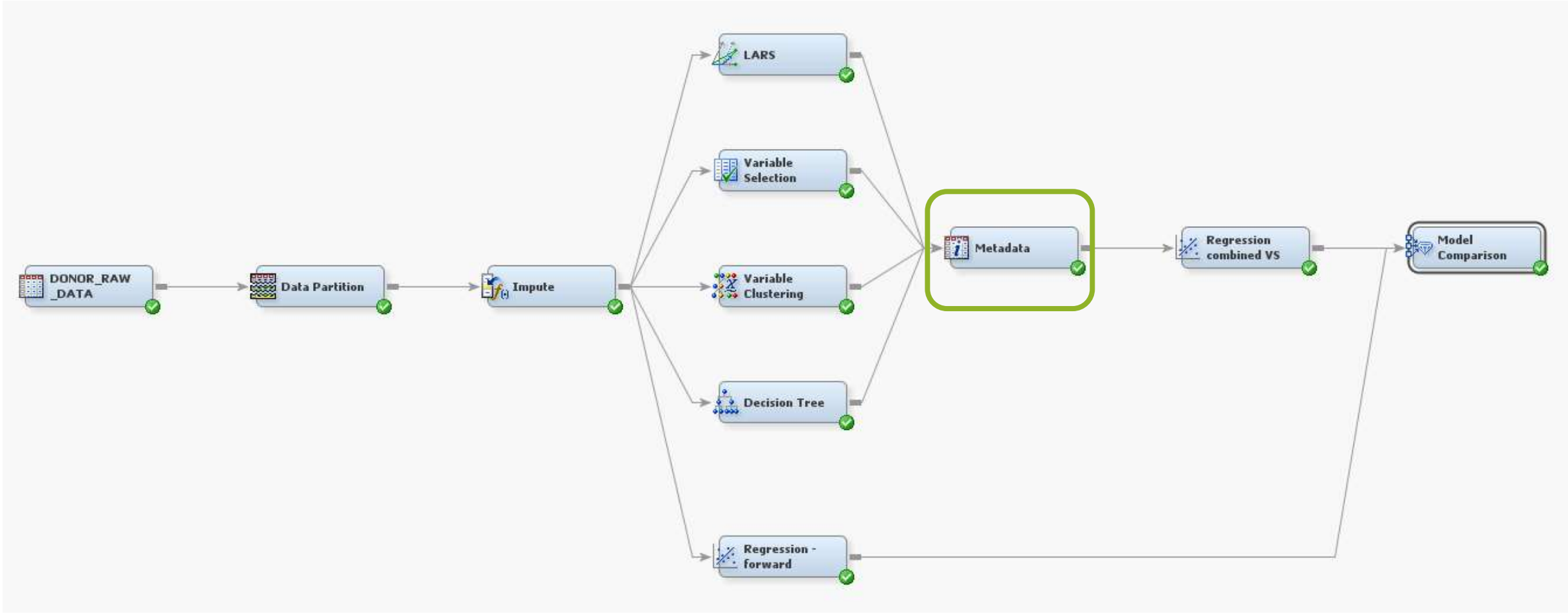
.. Property	Value
Interval Target Options	
Binary Transformation	Weighting Method
Use Target Min/Max for Scaling	No
Minimum for Scaling	0
Maximum for Scaling	1
Allow Out-of-Range Values	No
Random Seed	12345
Cutoff Value	0.20
Pre-Defined Groupings	
Use Frozen Groupings	No
Import Grouping Data	No
Import Data Set	
Use Pre-Defined WOE values	None
Interval Variable Binning Options	
Apply Level Rule	No
Binning Method	Quantile
Number of Bins	20
Special Code Options	
Use Special Codes	No
Special Codes Data Set	
Grouping Options	
Interval Grouping Method	Optimal Criterion
Ordinal Grouping Method	Optimal Criterion
Tree Based Grouping Options	
Constrained Optimal Options	
Advanced Constrained Optimal	
Maximum Number of Groups	5
Significant Digits	2
Apply Restrictions	Yes
Type	Percent
Percent	5.0
Count	.
Adjust WOE	Yes
Adjustment Factor	0.5
Score	
Group Level	Ordinal
Variable Selection Method	Information Value
Gini Cutoff	20.0

Variable	Information Value
RECENT RESPONSE COUNT	0.106
FREQUENCY STATUS 97NK	0.1
RECENT CARD RESPONSE COUNT	0.094
LAST GIFT AMT	0.093
RECENT RESPONSE PROP	0.085
LIFETIME MAX GIFT AMT	0.082
FILE AVG GIFT	0.081
LIFETIME AVG GIFT AMT	0.081
RECENT AVG GIFT AMT	0.079
FILE CARD GIFT	0.078
LIFETIME GIFT COUNT	0.072
PEP STAR	0.067
RECENCY STATUS 96NK	0.064
RECENT AVG CARD GIFT AMT	0.062
RECENT CARD RESPONSE PROP	0.056
MONTHS SINCE LAST GIFT	0.053
LIFETIME MIN GIFT AMT	0.052
LIFETIME GIFT RANGE	0.047
RECENT STAR STATUS	0.041
MONTHS SINCE FIRST GIFT	0.036
NUMBER PROM 12	0.035
MONTHS SINCE ORIGIN	0.033
LIFETIME CARD PROM	0.032
LIFETIME PROM	0.032
CARD PROM 12	0.031
IMP MONTHS SINCE LAST PROM	0.031
CLUSTER CODE	0.03
LIFETIME GIFT AMOUNT	0.028
MEDIAN HOME VALUE	0.016
PER CAPITA INCOME	0.014
IMP DONOR AGE	0.013

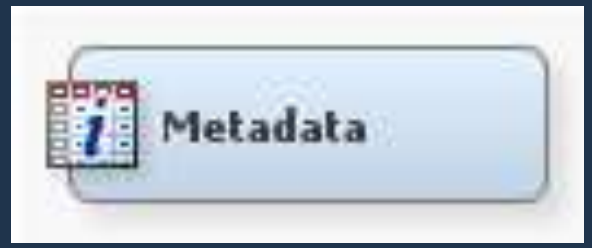


SAS® Enterprise Miner™

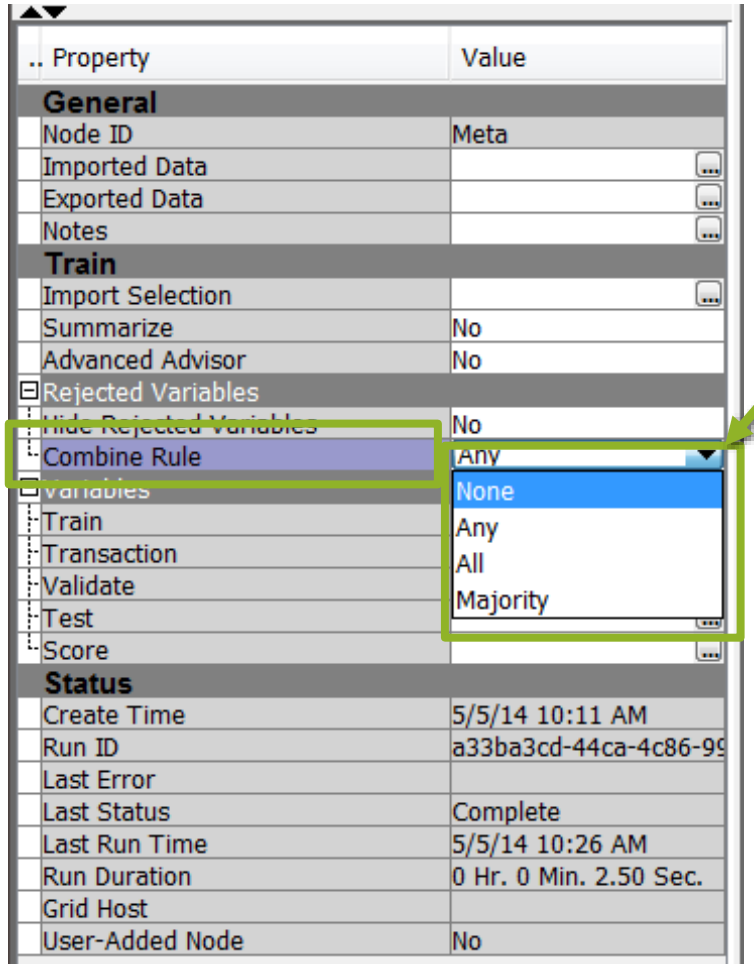
Use Metadata Node For Combining



SAS® Enterprise Miner™



Metadata



None — The role of input and rejected variables is based on the active metadata.

Any — A variable is set to Rejected if it is rejected in at least one of the incoming metadata sources.

All — A variable is rejected only if it is rejected in all of the incoming metadata sources.

Majority — A variable is rejected if it is rejected in the majority of the incoming metadata sources. If there is a tie, the rejection is based on the active metadata source.

Resources for Variable Selection

Model Selection

using SAS® Enterprise Guide® and SAS® Enterprise Miner™

http://support.sas.com/training/askexpert.html

Analytics

SAS Enterprise Miner: Getting Started

In this session, you learn how to use SAS Enterprise Miner to create a project, define a data source, create a basic flow process, explore and modify data, perform basic modeling, and score new data.

Register for live or on-demand event

Data Mining Tasks with SAS Enterprise Guide

This session covers SAS Rapid Predictive Modeler and model scoring. It also covers recency, frequency, and monetary analysis.

Register for live or on-demand event

Data Preparation Best Practice Approaches for Deriving Better Insights New Topic!

Creating consistent, reliable data that is ready for analytics can be a time-intensive task including many different techniques. Explore data preparation ideas to help drive better insights from analytics, moving from basic best practices to advanced options. This webinar is for data scientists, data engineers, data stewards and analysts.

Register for live or on-demand event

Ensemble Models and Partitioning Algorithms in SAS Enterprise Miner

This session presents various ensemble models based on partitioning algorithms in SAS Enterprise Miner. These include decision trees, bagging, boosting, gradient boosting, random forests, and ensemble trees.

Register for live or on-demand event

Introduction to Survival Data Mining New Topic!

This webinar will discuss topics in survival data mining such as building models for time-dependent outcomes derived from customer event histories, accounting for competing risks, and incorporating time-dependent covariates. The demonstration will be in SAS/STAT 14.3 and will show how to fit a multinomial logistic regression model with splines for right censored data. The demonstration will also show how to produce graphics to illustrate the results of the model.

Register for live or on-demand event

Model Selection Techniques in SAS Enterprise Guide & SAS Enterprise Miner

This presentation will answer the what, why and how on model selection. We will take a look at criterion and techniques used for model selection. Why is it important? And why should it be on your list of activities when doing predictive modeling? How to do model selection using SAS Enterprise Guide and SAS Enterprise Miner? This presentation will include examples for both SAS products.

Register for live or on-demand event

Resources

SAS Courses

- Predictive Modeling Using Logistic Regression
- Applied Analytics Using SAS Enterprise Miner
- SAS Enterprise Miner High-Performance Data Mining Nodes
- Data Mining Techniques: Theory and Practice
- Predictive Modeling Using SAS High-Performance Analytics Procedure
- Applied Clustering Techniques

For a complete list of courses, please see
https://support.sas.com/edu/courses.html?ctry=us

Resources

Videos

- The HPBIN Procedure
- Introducing the HPGENSELECT Procedure
- Introducing PROC QUANTSELECT
- What's New in SAS Enterprise Miner
- Interval Target Scorecards – Interactive Binning Node
- The New HP GLM Node in SAS Enterprise Miner
- Tutorials for SAS programming, Enterprise Guide, Analytics

Resources

Additional Reading

- Graphs Useful For Variable Selection in Predictive Modeling Predictive Models Based on Reduced Input Space That Uses Rejected Variables
- Variable Reduction in SAS by Using Weight of Evidence and Information Value
- Combining Decision Trees with Regression in Predictive Modeling with SAS® Enterprise Miner™
- Variable Reduction for Modeling using PROC VARCLUS
- Applications of the GLMSELECT Procedure for Megamodel Selection
- Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy
- Model Variable Selection Using Bootstrapped Decision Tree in Base SAS®
- SAS® Code for Variable Selection in Multiple Linear Regression Models Using Information Criteria Methods with Explicit Enumeration for a Large Number of Independent Regressors
- On Bayesian Model and Variable Selection using MCMC
- Recreating the SELECTION=SCORE Model Specification with the BEST=*n* Effect Selection Option for PROC SURVEYLOGISTIC
- An Overview of Machine Learning with SAS Enterprise Miner
- How to Apply the VIF Regression Algorithm in SAS Enterprise Miner
- Extension Node to the Rescue of the Curse of Dimensionality via Weight of Evidence (WOE) Recoding

Questions?

Thank you for your time and attention!

Connect with me:
LinkedIn: https://www.linkedin.com/in/melodierush
Twitter: @Melodie_Rush

sas.com