Verification System Analysis - Macro

(1) Use last year's merged payment file
Eligibles who applied for before the 1st of January of the 2nd year
Eligibles selected randomly or immune
Applicants from 1999-2000 & 2000-01

(2) Eliminate applicants from non-participating institutions

(3) Merged Application/Payment File

First Work File
(4) Selection of splits can be cross-validated from one to another

Second Work File
(4a) Selection of splits can be cross-validated from one to another

(5) Identify the first selected transaction / first transaction with an EFC (called target transaction)

(6) Use formula for taxes (formula of processing system)
Creates variable defined by discrepancies between expected and reported taxes

Dependent Students Cluster
(7) (Dependent and independent students treated separately)

Partitions
(8) (Partitions are created by using application variables)

Partition By Income (8a)
Partition By Household Size (8b)
Partition By Taxes Paid (8c)
Partition By Tax Discrepancy (8d)
Partition By Marital Status (8e)
Partition By Tax Form (8f)
Partition By Tax Estimate (8g)
Partition By Tax Payment Status (8h)
Partition By Nontaxable Income (8i)
Partition By Marital Status Effective Date (8j)
Partition By Receipt of Application Date (8k)
Partition By Control of First Choice School (8ll)

Continous Variables (9)
(partitions identified as above or below some threshold, with multiple thresholds defined for each variable)

Example First Partition
Dependent Student (10)

Group One (11)
- dichotomous variable
- total of 147 partitions of dependent students defined

Group Two (12)
- those with very small percentage, removed and replaced with other partitions

Each partition defined into two clusters (13)
Index calculated for each cluster
Partition considered if:
- at least 400 applicants in selected & immune group
- in each of two clusters of partition

First Split (14)
Partition with largest differences of indices/ root mean square of standard deviation (must be greater than 2.36) became the first split, creating two clusters

First Split / First Cluster (15)
Selection Process
Partition with largest difference in index between two clusters

First Split/ Second Cluster (16)
Same Selection Process

Second Split/ First Cluster (17)

Second Split/ Second Cluster (18)

NOTE: If the index is the same in both samples is higher than the other clusters, then the split is retained. Otherwise, split is eliminated and the parent cluster is terminal cluster. Terminal clusters = Targeted Selection Criteria. (19)

NOTE: Model is programmed into the other work file, which serves as a replicate sample. (20)

NOTES: (21)
1. Model adequate if no terminal terminal cluster > 15% of total applicants.
2. Trees produced were examined, only one was acceptable.
3. For Independents, only change in root mean square formula.
4. Two work files were combined for an index and savings for each TSC. (targeted selection criteria)
5. Specifications written and tested.
6. Index , Cumulative Index, Savings and Cumulative Savings Tables produced.

Annotations:

(1)

- sets the target for the next year going forward.
- need prior application cycle for 1999-2000 and 2000-2001.
- need only the payment file for Award Year 2000-2001.
- focus is on PELL eligibles
- applications are run on a 104 week cycle (2 year cycle).

(4)

- second work file serves as the control file.
- selected transaction is the 1st transaction with an EFC.
- the process takes the 1st transaction to calculate an EFC, just one action.

(6)

- formula uses AGI repeated through IRS tables; creating a match.
- formula uses variables that impact your marginal tax amount.

(9)

- continuous variables are variables that include a max./min. threshold.

(11)

- group one can be used for analysis.
- you, as an applicant, fit into either one of the 147 partitions or not.

(12)

- group two can be used as a control file.

(19)

- some TSCs will repeat from one year to another.
- change of tax laws creates difficulty in year-by-year data comparisons.