

TUTORIAL 2: USING SI-CHAID TO IDENTIFY PROFITABLE SEGMENTS

This tutorial shows how to use the CHAID ordinal algorithm to segment based on profitability scores. We will again use the magazine subscription data set, *subscribe.sav*, used previously in Tutorial 1. However, our dependent variable will now be RESP3, coded 1 (paid responder), 2 (unpaid responder) and 3 (nonresponder). We'll compare a default nominal CHAID segmentation of RESP3 to the ordinal CHAID analysis that takes into account the gain (or loss) associated with each response group. For simplicity, we utilize the SI-CHAID option settings used in Magidson (1993).

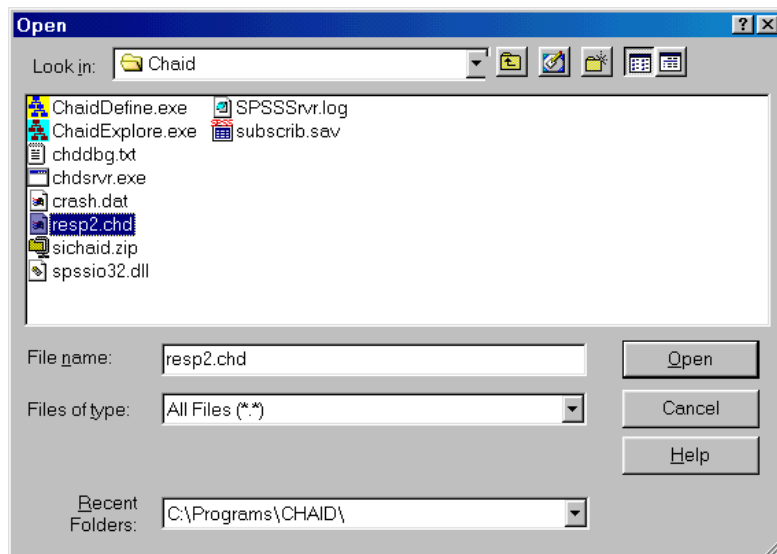
The Data

For this Tutorial, we will be using the same data file as for *Tutorial 1: Beginning a CHAID Analysis*. The file *subscribe.sav* contains information about a direct marketing promotion used to encourage people to subscribe to a magazine. Households that were sent the promotion were categorized as paid responders, unpaid responders, or nonresponders. The data and analyses are described in more detail in Magidson (1993).

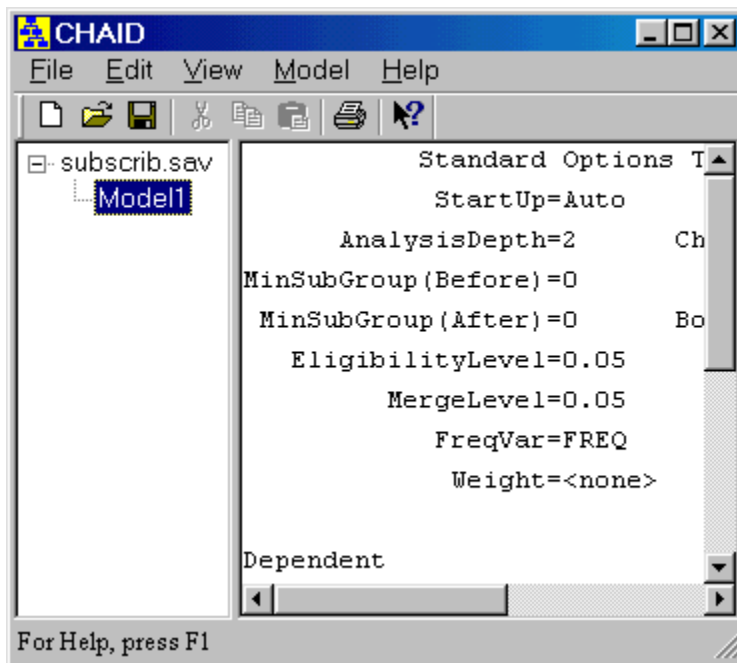
Modifying the Previous Analysis File

If your analysis file from tutorial #1 is not still open, re-open it:

- Open the Define program
- Select Open from the File Menu
- From the files listed select 'resp2.chd' and click the Open button



Your earlier analysis file is retrieved:

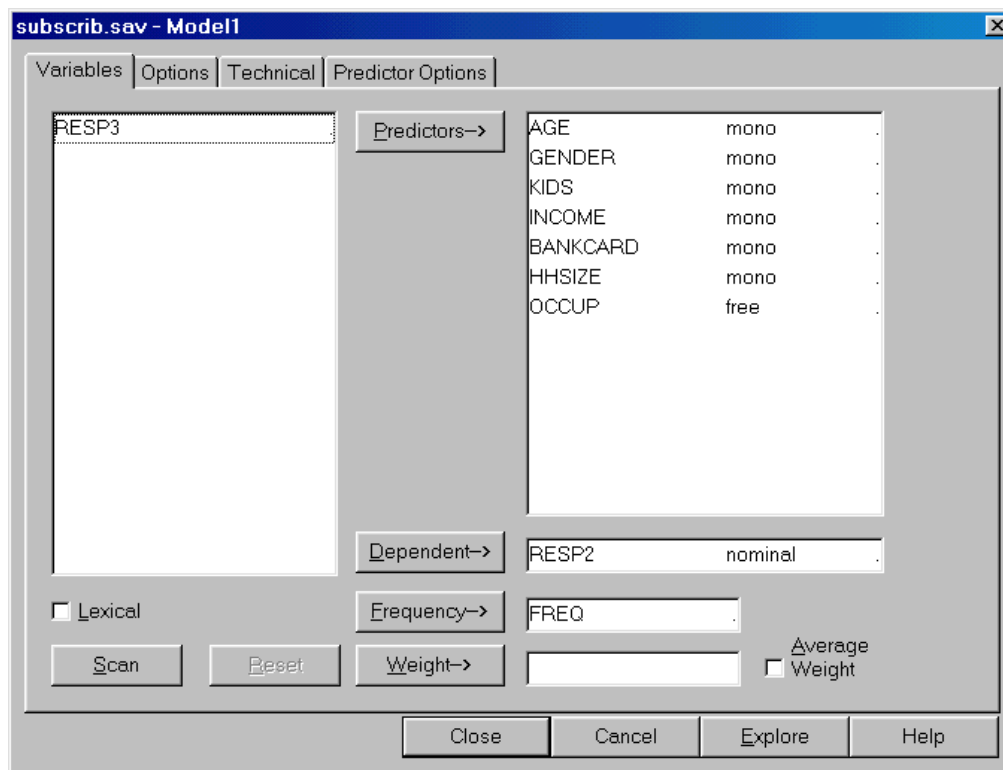


➤ Right-click on 'Model1' and select 'Edit'

or alternatively,

➤ double-click on 'Model1'

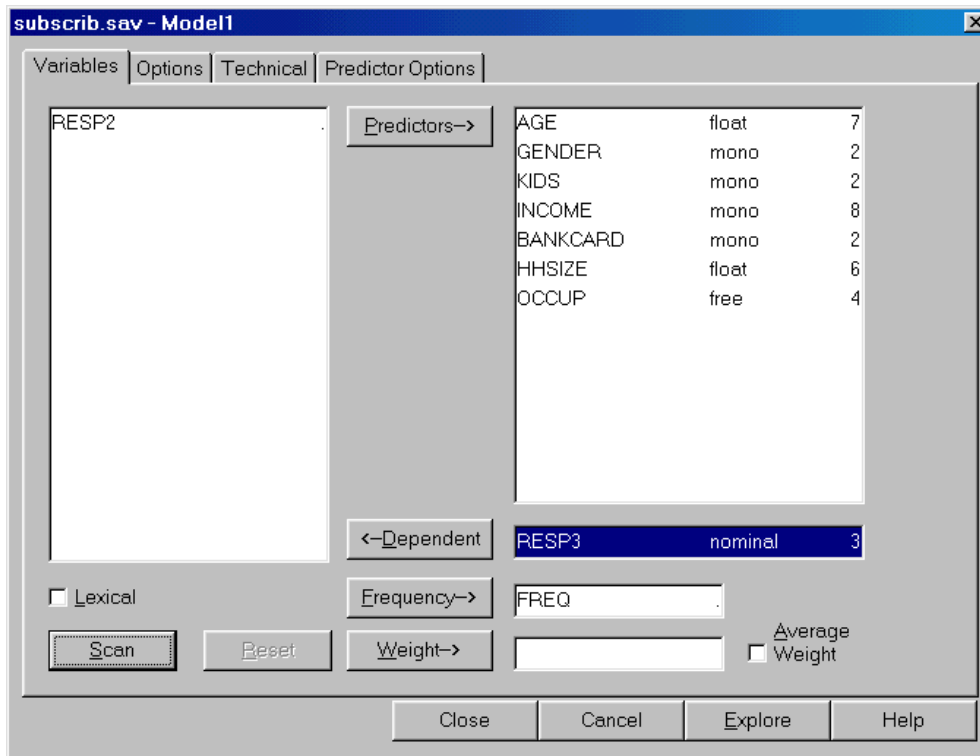
to enter the Variables tab of the Model Analysis Dialogue Box:



To change the dependent variable from Resp2 to Resp3 and re-scan the data file:

- Click on Resp2
- Click the Dependent button
- Select Resp3 from the Variables box
- Click the Dependent button
- Click Scan

The Model Analysis Dialog Box should now look like this:

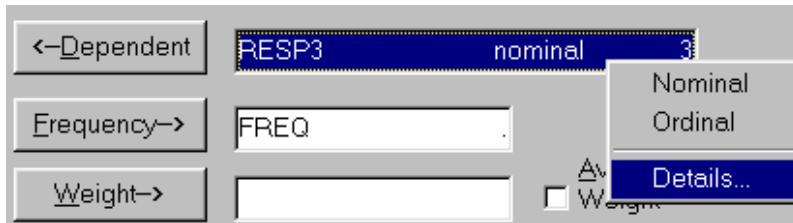


Assigning Category Scores

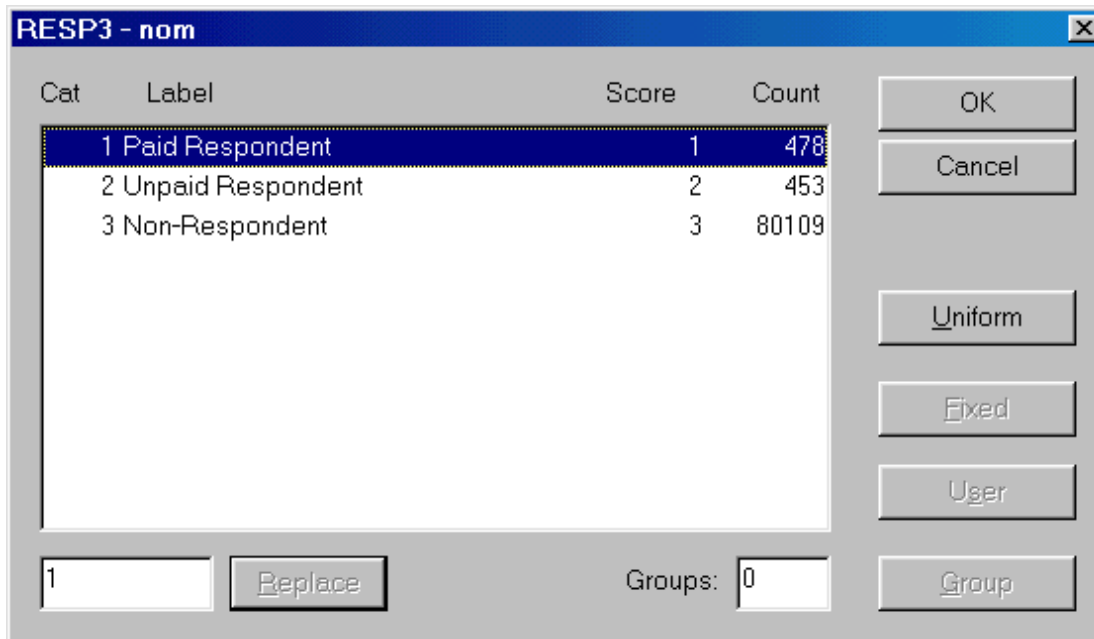
Nominal Method

Before growing the new tree, we will assign profitability scores to the categories of the dependent variable for future use. Although the standard CHAID algorithm (the 'nominal' algorithm) does not utilize these scores to grow the tree, the scores may still be used by the gains chart to identify which of the resulting segments are most profitable. Later, we will compare results from the nominal segmentation to the segmentation obtained from the ordinal algorithm.

- Right-click on RESP3 in the dependent box of the Model Analysis Dialog Box
- In the pop-menu, select Details



Clicking Details will bring up the Edit Scores Box



(Alternatively, double-clicking on Resp3 would also get us to this screen)

The first category (Paid Respondent) is highlighted. The default scores correspond to the integer codes used in the SPSS file – 1,2 and 3. To change the score for Paid Respondents,

- double-click on the 'Paid Respondent' label

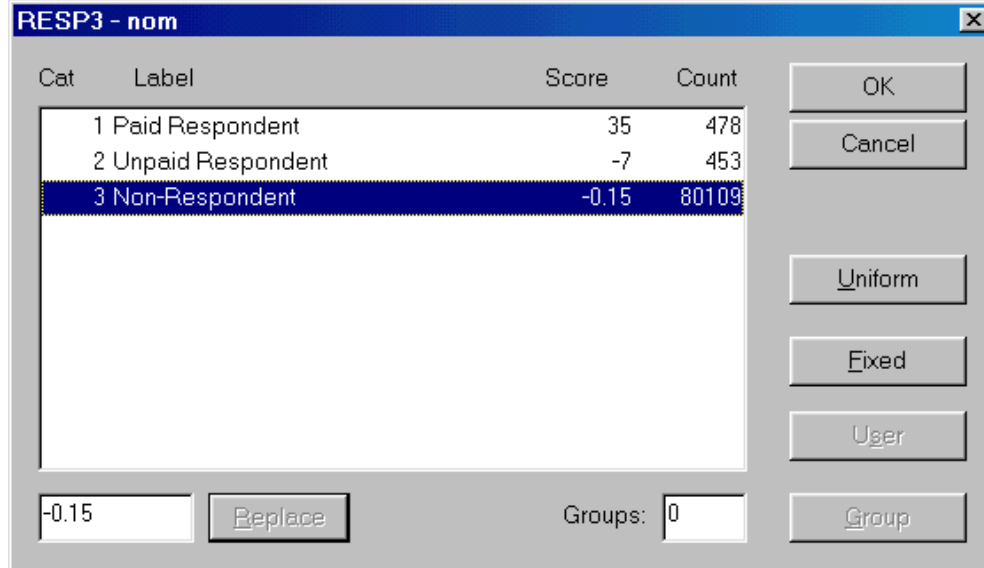
the score '1' is highlighted in the Edit Scores box

- replace the score '1' with the score '35' and click the Replace button

Now repeat these steps for the other categories:

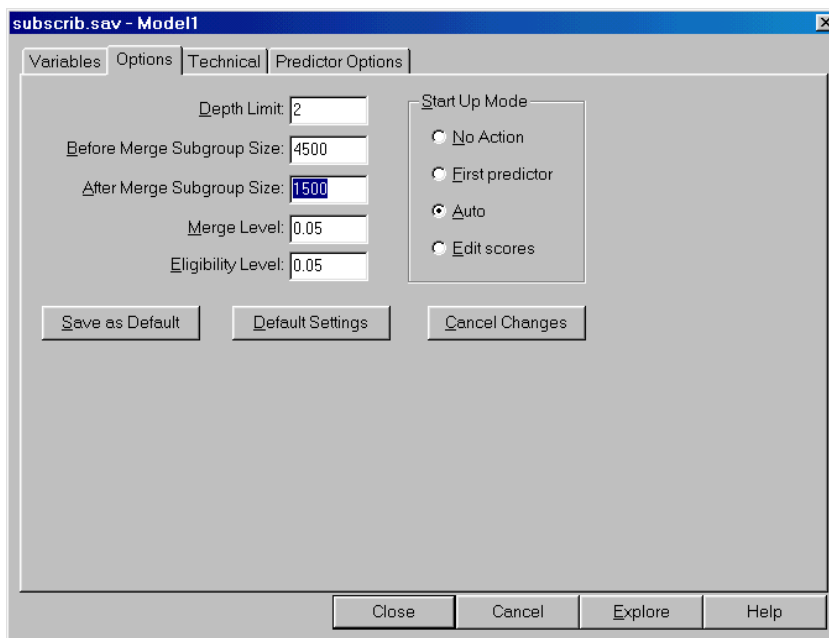
- double-click on the second category ('Unpaid Respondent').
- replace the score '2' with the score '-7' and click the Replace button
- double-click on the third category ('Nonresponder').
- replace the score '3' with the score '-0.15' and click the Replace button

Your screen should now look like this:



- Click OK to return to the Model Analysis Dialog Box
- Now, go to the Options Tab
- Change the "Before Merge Subgroup Size" to '4500' and the "After Merge Subgroup Size" to '1500'. These were the settings used in the Magidson (1994) article.

The Options Tab should now look like this:

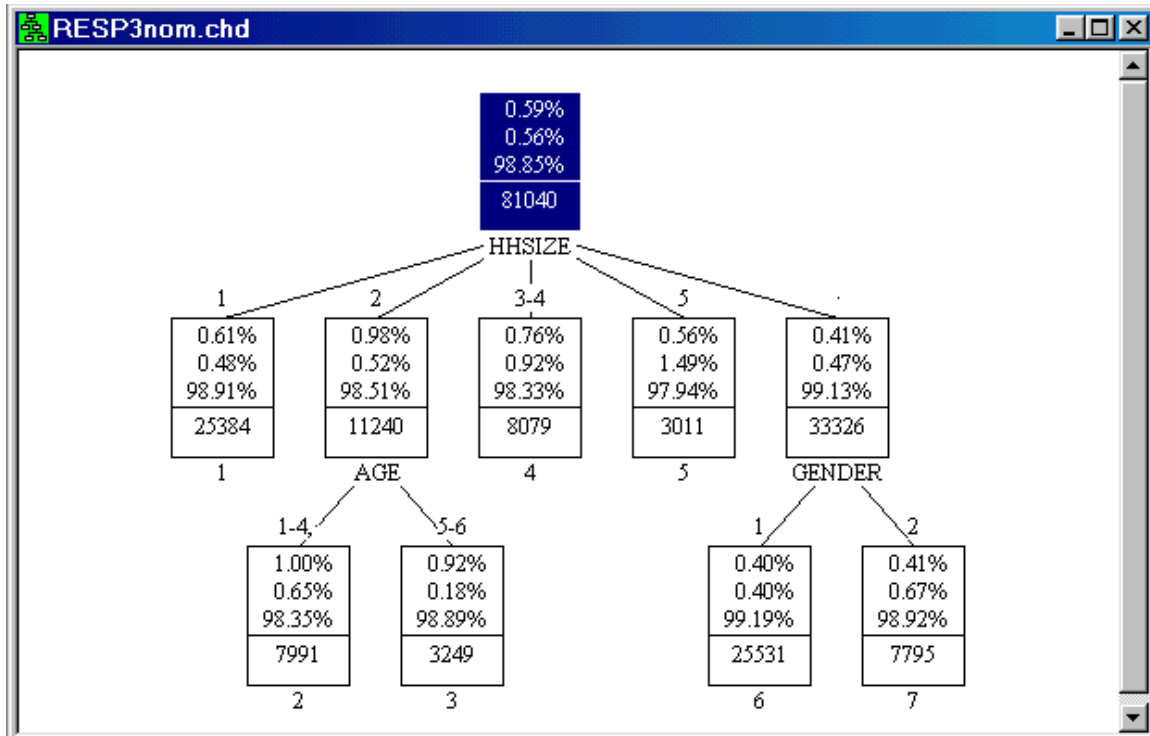


To save the new analysis file and grow the tree:

- Click Explore
- In the File name box type *RESP3nom.chd* to override the suggested filename
- Click the Save button

This tells SI-CHAID to save your analysis settings to an analysis file with the name *RESP3nom.chd*. All printed and saved output will be prefixed by the name *RESP3nom*. Later, we will create another analysis file with named *RESP3ord.chd* corresponding to the ordinal algorithm.

After you click Save, SI-CHAID automatically opens ChaidExplore and generates the following 7-segment tree:



Notice that this RESP3nom solution differs from our earlier 6-segment RESP2 solution (recall *Tutorial 1: Beginning a CHAID Analysis*). For example, while HHSIZE is still used for the first split, it is now merged into five categories instead of four. In our earlier analysis, HHSIZE categories 2 and 3 were merged. Now category 2 is a separate category and categories 3 and 4 are merged.

To obtain a gains chart for this segmentation,

- Select 'New Gains' from the Windows menu.

The gains chart appears as follows:

Id	size	% of all resp	%resp	score	index	Cum. size	% of all resp	%resp	score	index
2	7991	9.9	80	16.7	1.00	7991	9.9	80	16.7	1.00
3	3249	4.0	30	6.3	0.92	11240	13.9	110	23.0	0.98
4	8079	10.0	61	12.8	0.76	19319	23.8	171	35.8	0.89
1	25384	31.3	155	32.4	0.61	44703	55.2	326	68.2	0.73
5	3011	3.7	17	3.6	0.56	47714	58.9	343	71.8	0.72
7	7795	9.6	32	6.7	0.41	55509	68.5	375	78.5	0.68
6	25531	31.5	103	21.5	0.40	81040	100.0	478	100.0	0.59

The most profitable of these 7 segments (at the top of the list) is segment #3. The expected profit of \$.16 from mailing each household in this segment is computed by SI-CHAID as follows:

$$.0092 \times (\$35) + .0018 \times (-\$7) + .9889 \times (-\$0.15) = \$0.16$$

- Click the X in the upper right of the gain-chart to close it

To display the expected profit in each node of the tree rather than the percentages for paid, unpaid and non-responders:

- Right click in any node of the tree diagram
- Select 'node items' from the pop-up menu
- Click the box to the left of 'Score'

A check-mark appears in this box.

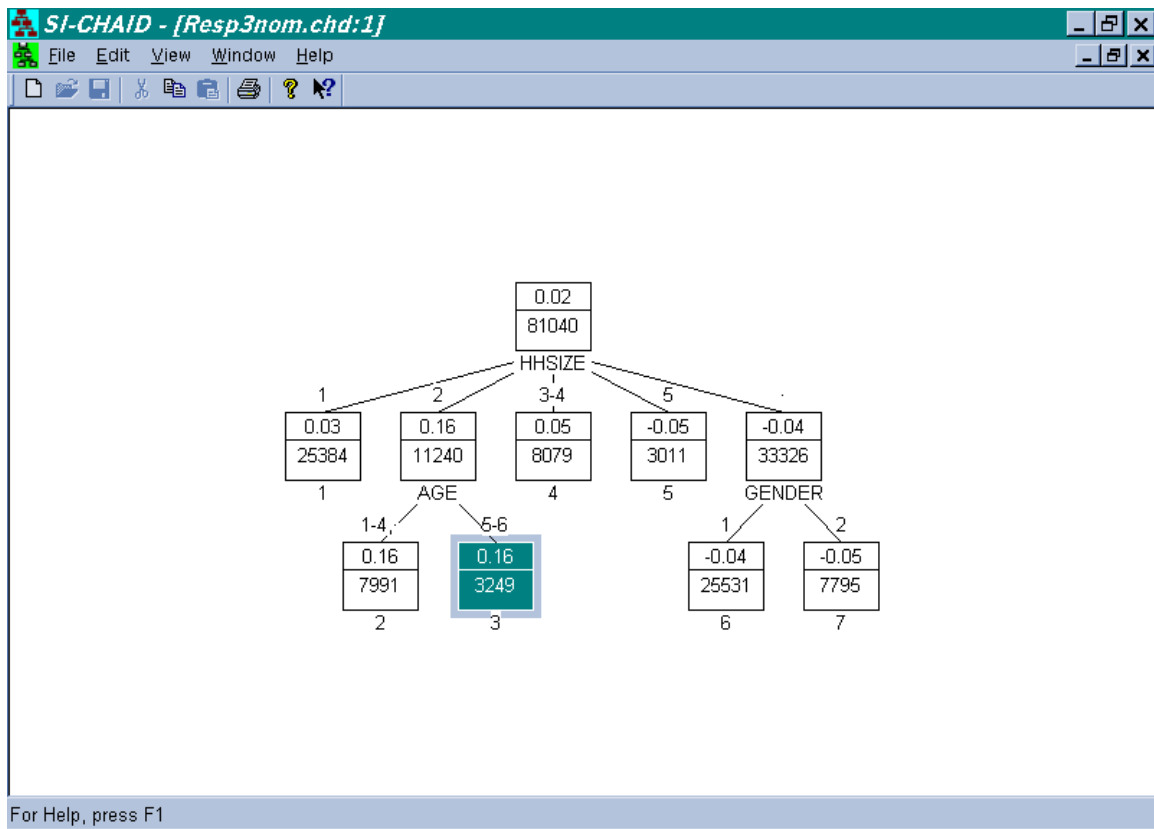
To remove the percentages from each node of the tree:

- Click the box to the left of 'Percents'

The check-mark disappears from this box.

- Click 'Close'

The revised tree display is as follows:



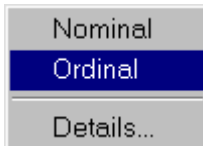
Ordinal Method

We will now reanalyze these data using the same category scores but we will use the ordinal method which treats the dependent variable as ordinal.

- Return to ChaidDefine.exe and double-click on "Model 1" in the left pane.

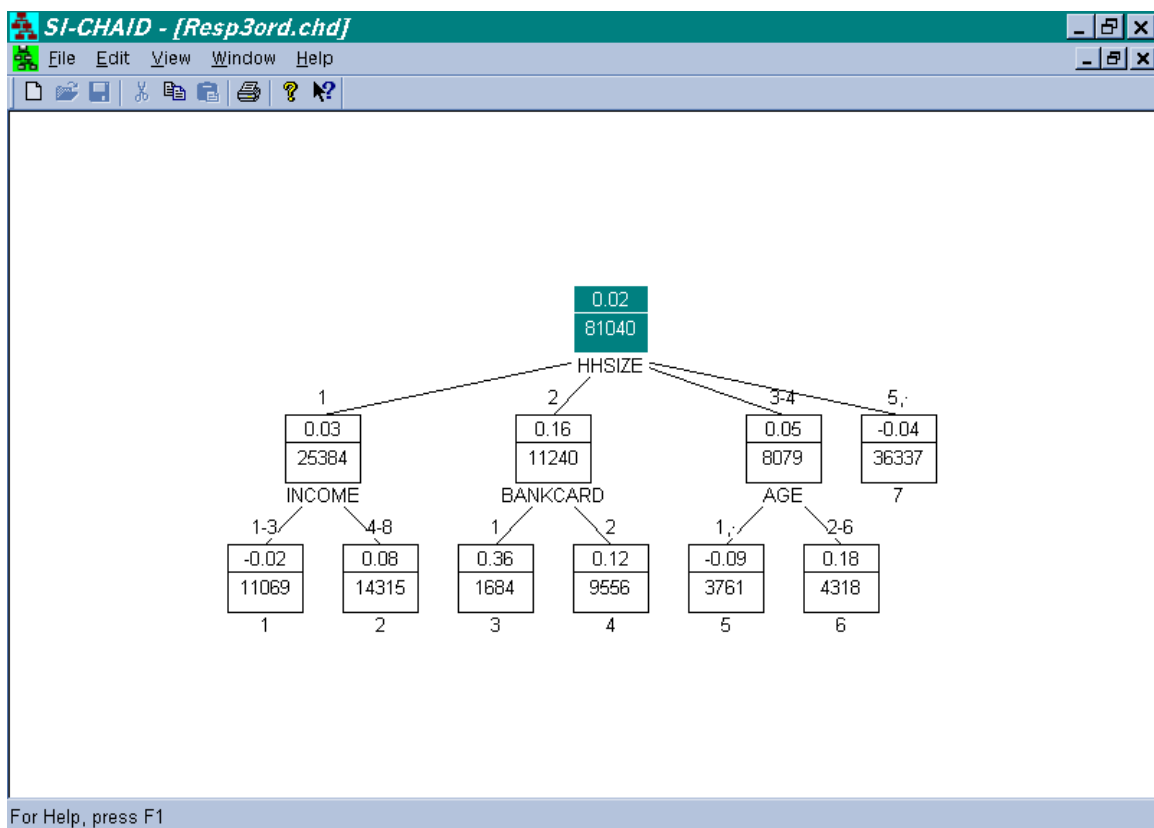
The Model Analysis Dialog Box pops up

- Right-click on RESP3 in the Dependent variable box and select Ordinal from the pop-up menu



- Click Explore
- Enter the filename RESPord.chd so as to not replace our earlier analysis file RESPnom.chd
- Click Save

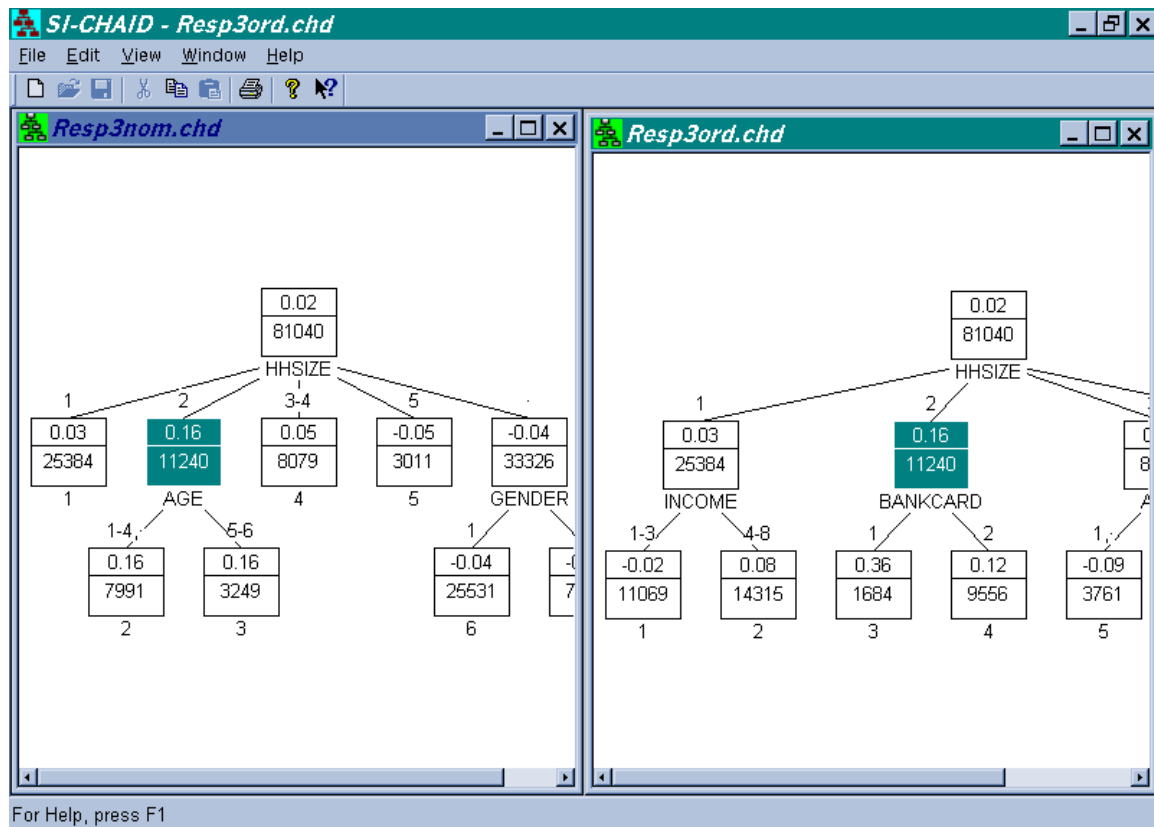
The following tree diagram is displayed:



To display the Nominal and Ordinal segmentation trees side-by-side:

- Select 'Tile Vertical' from the Windows menu

Note that two-person households are now split based on whether they own a bankcard rather than based on Age, and that the expected gain for two-person households that own a bankcard (0.36) is three times greater than the expected gain for two-person households that do not own a bankcard (0.12).

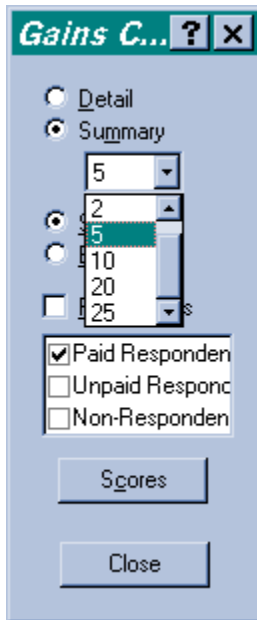


- Return to the nominal segmentation and click on the node corresponding to HHSIZE =2
- Right-click and choose 'Select'

Notice that only a single predictor, AGE, is listed as a candidate for splitting this subgroup using the nominal method. The nominal test of significance is not powerful enough to identify the important BANKCARD effect. By taking into account the profitability scores, the ordinal test of significance utilizes only a single degree of freedom. Thus, it provides a more powerful test of significance and a better segmentation here than the nominal method (For further details, see Magidson, 1994).

To compare gains charts from the different segmentations:

- Click in the Window of the nominal segmentation tree to make it active
- Select New Gains from the Windows menu
- Right-click on this gains chart and select Gains Items from the pop-up menu
- Select Summary to display the quantile format and change the default to 5 percentile units
- Click OK to close this Window



Repeat these steps to obtain a corresponding gains chart for the ordinal segmentation tree:

- Click in the Window of the ordinal segmentation tree to make it active
- Select New Gains from the Windows menu
- Right-click on this gains chart and select Gains Items from the pop-up menu
- Select Summary to display q quantile format and change the default to 5 percentile units
- Click OK to close this Window

- Rearrange the gains Windows to present them side-by-side:

tile	size	score	index
5	4052	0.16	833
10	8104	0.16	833
15	12156	0.15	791
20	16208	0.13	663
25	20260	0.11	580
30	24312	0.10	512
35	28364	0.09	463
40	32416	0.08	426
45	36468	0.08	397
50	40520	0.07	374

tile	size	score	index
5	4052	0.25	1315
10	8104	0.20	1048
15	12156	0.17	915
20	16208	0.16	839
25	20260	0.14	751
30	24312	0.13	692
35	28364	0.12	650
40	32416	0.11	577
45	36468	0.09	499
50	40520	0.08	436

Comparison of these gains charts show that the ordinal segmentation would be expected to outperform the nominal segmentation for mailings involving profitable segments (less than 50% of all cases). Hence, by taking into account the profitability scores, the ordinal algorithm provides a more profitable segmentation.

