Tutorial #7: LC Segmentation with Ratings-based Conjoint Data

This tutorial shows how to use the Latent GOLD Choice program when the scale type of the dependent variable corresponds to a Rating as opposed to a Choice or Ranking. Ratings data can also be analyzed using the Regression module in Latent GOLD and the resulting parameter estimates will be identical. However, the Latent GOLD Choice program produces additional output, and can be used to 'fuse' rating and choice data, which is useful for situations where both types of data are available.

In this tutorial, we will reanalyze the conjoint data used previously in Latent GOLD Tutorial #3: (http://statisticalinnovations.com/products/LGtutorial3.pdf) "LC Regression with Repeated Measures". You will learn how to:

- Setup the data for the Latent GOLD Choice program using the 1-file format
- Examine the additional output not available in Latent GOLD Regression module.

By examining the additional 'Set Profile' Output generated by the LG Choice program we will see that the predicted ratings obtained using the standard aggregate (1-class) conjoint model fail to provide an adequate fit to the observed ratings. In contrast, we will see that the predictions generated by the 3-class model are quite good.

The Data

The data for this example are obtained from a hypothetical conjoint marketing study involving repeated measures. Respondents were asked to provide likelihood of purchase ratings for each of several different products that differ on type of fashion, quality and price. A partial listing of the data is shown below. The only difference between this (1-file) format and the format used in the earlier tutorial in conjunction with the Regression module of Latent GOLD is the inclusion of the additional variable called SET. For 'Choice Ratings' models, each set consists of a single product (scenario, profile) that is to be rated.

Table 1: Partial Listing of Conjoint Data

CONJOINT.SAV - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities AmosApps Window Help

1 : id 1

	id	sex	age	fashion	quality	price	rating	set
1	1	Male	25-39	Traditional	Low	Lower	Neutral	1
2	1	Male	25-39	Traditional	Low	Higher	Very Unlikely	2
3	1	Male	25-39	Traditional	High	Lower	Very Likely	3
4	1	Male	25-39	Traditional	High	Higher	Neutral	4
5	1	Male	25-39	Modern	Low	Lower	Somewhat Unlikely	5
6	1	Male	25-39	Modern	Low	Higher	Somewhat Unlikely	6
7	1	Male	25-39	Modern	High	Lower	Very Likely	7
8	1	Male	25-39	Modern	High	Higher	Very Likely	8
9	2	Female	16-24	Traditional	Low	Lower	Neutral	1
10	2	Female	16-24	Traditional	Low	Higher	Somewhat Unlikely	2
11	2	Female	16-24	Traditional	High	Lower	Very Likely	3
12	2	Female	16-24	Traditional	High	Higher	Very Likely	4

Data View Variable View

SPSS Processor is ready

As shown above, there are 8 records for each case (there are 400 cases in total); one record for each of the 2x2x2 = 8 sets which correspond to a complete factorial design of different scenarios for the purchase of a product:

- *FASHION (1 = Traditional; 2 = Modern)*
- *QUALITY (1 = Low; 2 = High)*
- *PRICE (1 = Lower; 2 = Higher)*

The dependent variable (*RATING*) is a rating of purchase intent on a five-point scale. Effects measuring the importance of each of these attributes on the likelihood-to-purchase rating are to be estimated.

Also included in the data file are two demographic variables which will be used as covariates in the model.

- *SEX (1 = Male; 2 = Female)*
- *AGE (1 = 16-24; 2 = 25-39; 3 = 40+).*

The Goal

Use Latent GOLD Choice to identify latent segments that differ with respect to the part-utility estimates of importance (effects) for the three attributes. We wish to determine the number of segments and see how these segments differ demographically.

Opening a Data File and Selecting the Type of Model

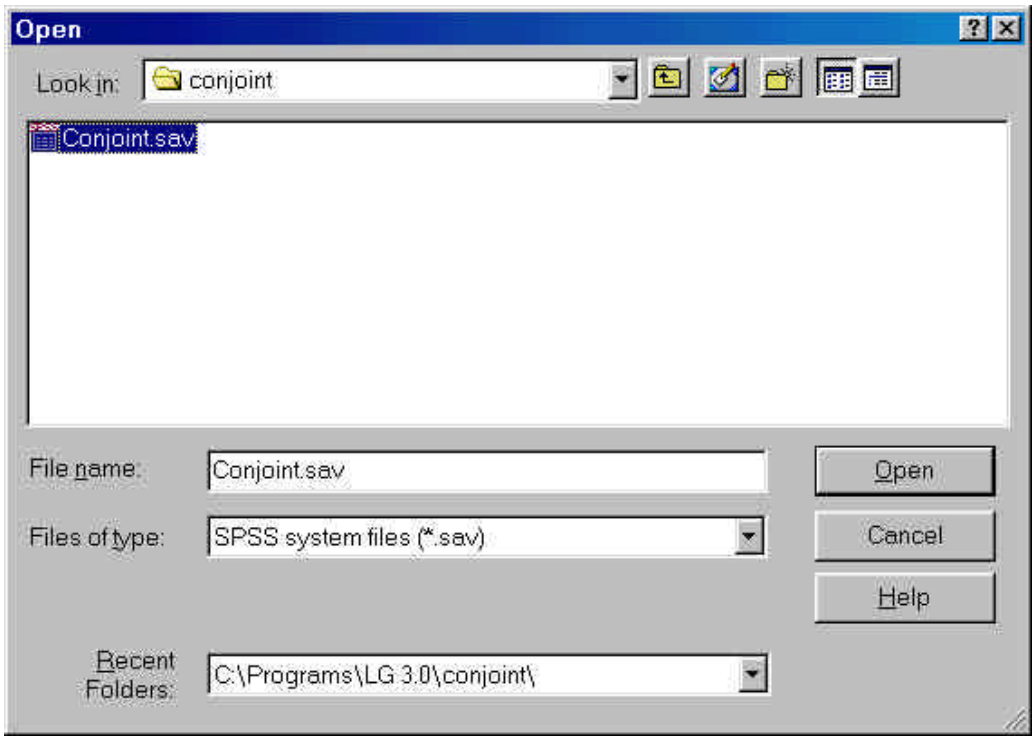
- To open the file, from the menus choose:

File
Open

- From the Files of type drop down list, select SPSS System Files if this is not already the default listing. All files with .sav extensions appear in the list (see Figure 1).

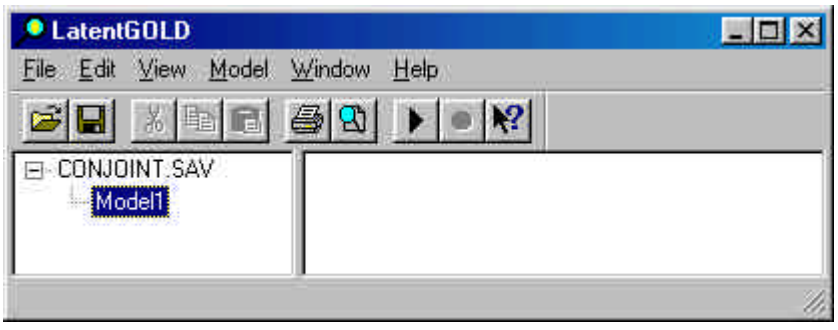
Note: If you copied the sample data file to a directory other than the default directory, change to that directory prior to retrieving the file.

Figure 1: Open Dialog Box



- Select conjoint.sav and click Open to open the Viewer window, shown in Figure 2

Figure 2: Viewer Window



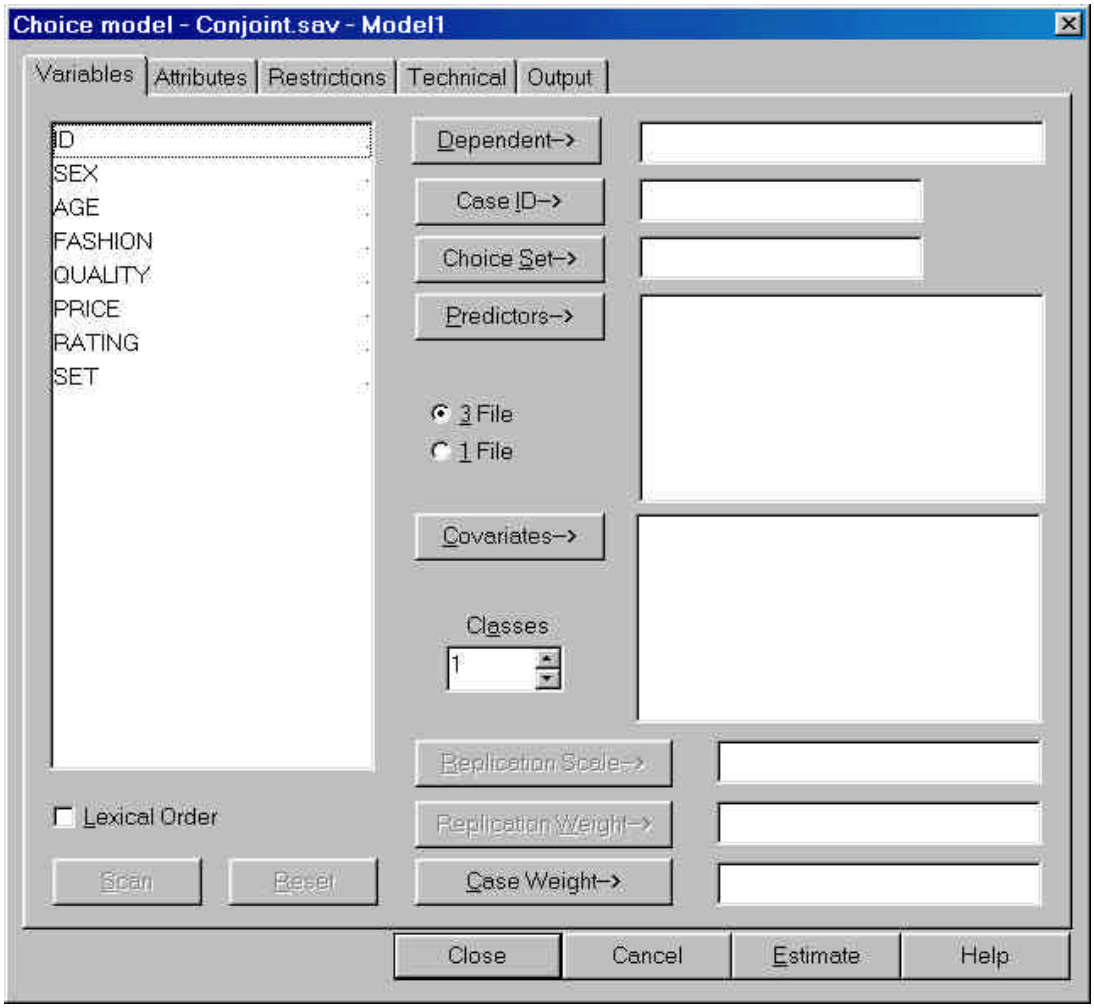
The Outline pane contains the name of the data file and a list of any models (and their output) previously estimated. Since we have not yet estimated any models, we will set up a new model (listed as 'Model1' above) and then estimate the parameters of that model. The Contents pane (currently empty) is where you view the various types of output following the estimation.

- Highlight 'Model1' if it is not already highlighted.
-
- If your Latent GOLD Choice is an add-on to Latent GOLD, right click to open the Model Selection menu (you may also double click the model name to open this menu or select the type of model from the Model menu) and select "Choice":

- Otherwise**,** if you have the stand-alone version of Latent GOLD Choice, simply double-click the model name to bring up the Model Analysis Dialog Box.

You should now see the Model Analysis Dialog Box, shown in Figure 3

Figure 3: Analysis Dialog Box for LC Choice Model



Selecting the Variables for the Analysis

For this analysis, *RATING* will be the **dependent** variable.

- Select *RATING* in the Variables List and click Dependent to move the variable to the Dependent box.

Dependent Variable Types

By default the scale type is set to 'Choice'. To set the scale type for the dependent variable to 'Rating'**:**

- Right-click on the dependent variable 'RATING' and select 'Rating' from the pop-up menu.

As illustrated above, the data contains repeated observations for each respondent (case). Therefore, we need to indicate which records belong to each case. This is accomplished using a **Case ID** variable, which contains a unique identification number for each case. All records belonging to the same case are assigned the same unique ID.

- Select *ID* in the Variables list and click Case ID to move the variable into the Case ID box.
- Select SET in the Variables list and click Choice Set to move the variable into the Choice Set box.

Next, we will select the demographic variables to be used as Covariates to predict class membership. These variables will be selected from those that appear in the Variables list box.

To estimate the model specifying *SEX* and *AGE* as covariates,

- Select *SEX* and *AGE* in the Variables list box.
- Click on Covariates to move them into the Covariates Box
- Right click on the variable names and select scale type Nominal

Specifying the File Format

- Click on "1-file" to select the 1-file format for this tutorial

Specifying the Number of Classes

The LC regression model simultaneously estimates a separate regression model for each class. A 1-class model estimates only a single regression model. It makes the standard homogeneity assumption that a single regression model holds true for all cases. In the current example, we will start by estimating a 1-class model and obtain a log-likelihood statistic to be used as a base. We will then estimate additional models by successively incrementing the number of classes by 1 and assess the significance of each additional class.

- In the box titled Classes (located below the Covariates pushbutton) type ‘1-4’ to request the estimation of 4 different LC Regression models – a 1-class model, a 2-class model, a 3-class model and a 4-class model

Specifying Attributes

Go to the Attributes Tab. Note that in the left-hand box the 3 remaining variables, FASHION, QUALITY and PRICE are listed, along with the variable called _Constants_.

- Highlight all of the variables, including _Constants_, and click Attributes to move them into the Attributes Box

- Highlight all of the variables in the Attributes Box and right-click to bring up the pop-up menu. Select Nominal from the pop-up menu.

Scanning the Data File

- Click Scan (located in the lower left of the Analysis dialog box) to scan the data file

The number of distinct categories (or values) along with the scaling option appears next to each variable in the Dependent and Predictors boxes.

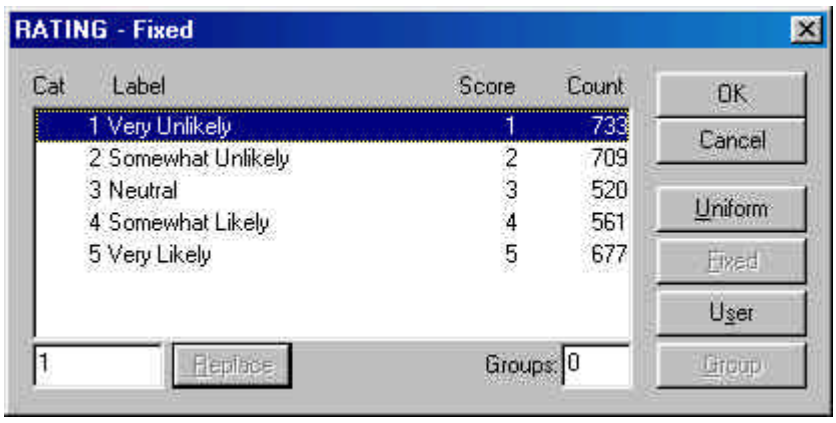
To view category labels, frequency counts and any scores assigned to any scanned variable,

- double click on the variable name in the Dependent or Predictor list box

The Variables dialog box opens (see Figure 4).

- Double click the dependent variable, *RATING*

Figure 4: Variables box for RATING



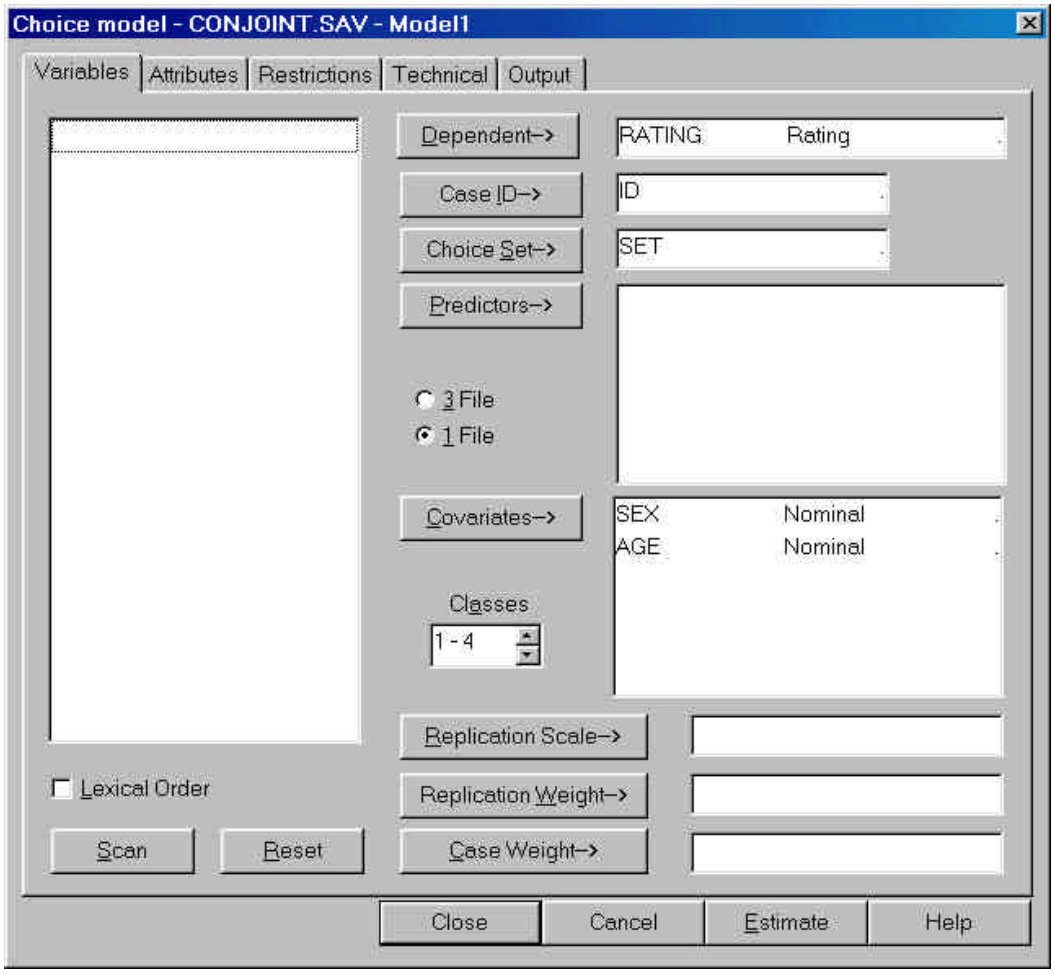
- Click OK to close the Variables dialog box and return to the Regression Analysis dialog box

Estimating the Model

Now that we have selected our variables and specified the models, we are ready to estimate the models.

Your analysis dialog box should look like Figure 5.

Figure 5: Regression Analysis Dialog Box with Initial Settings



- Click Estimate (located at the bottom right of the Analysis dialog box)

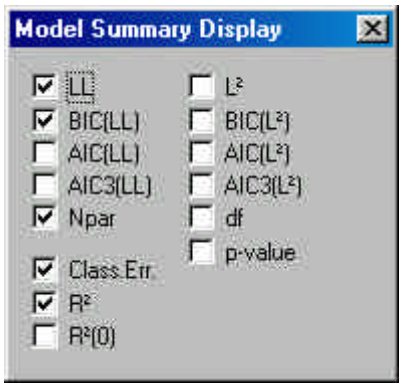
Viewing Output and Interpreting Results

For LC Choice models, several output files are produced. The Summary Output lists the models estimated and provides various summary statistics. The particular statistics displayed can be customized using the Model Summary Display.

To display only those statistics that are relevant,

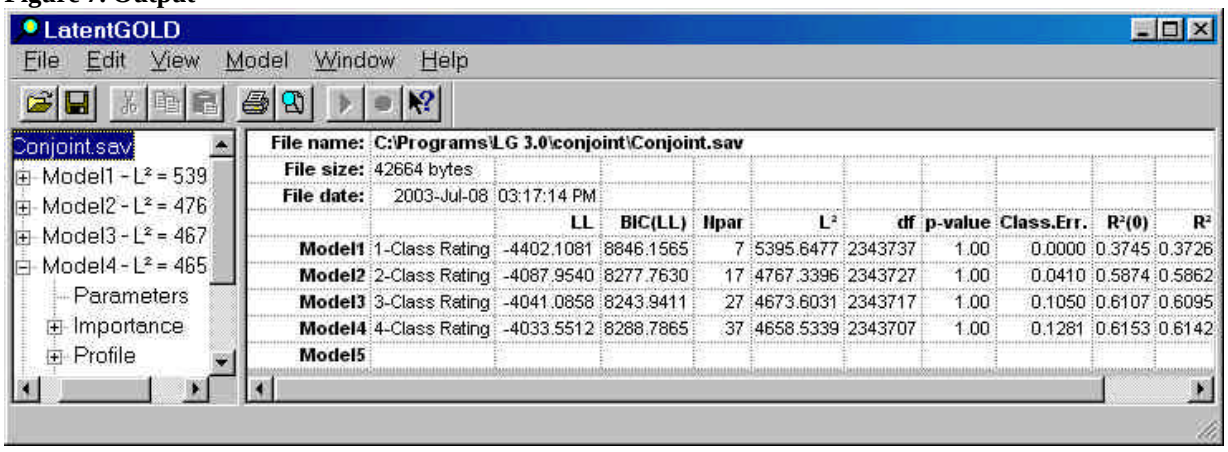
- In the Outline pane, click on the data file name, Conjoint.sav
- Right click in the Contents Pane to retrieve the Model Summary Display
- Customize the output in the Contents Pane to include only the statistics indicated in the Figure 6 below

Figure 6



Click on Conjoint.sav in the Outline Pane to bring up the summary output in the Contents Pane, which should look like Figure 7

Figure 7. Output



		LL	BIC(LL)	Npar	L²	df	p-value	Class.Err.	R²(0)	R²
Model1	1-Class Rating	-4402.1081	8846.1565	7	5395.6477	2343737	1.00	0.0000	0.3745	0.3726
Model2	2-Class Rating	-4087.9540	8277.7630	17	4767.3396	2343727	1.00	0.0410	0.5874	0.5862
Model3	3-Class Rating	-4041.0858	8243.9411	27	4673.6031	2343717	1.00	0.1050	0.6107	0.6095
Model4	4-Class Rating	-4033.5512	8288.7865	37	4658.5339	2343707	1.00	0.1281	0.6153	0.6142
Model5										

These statistics will assist us in determining the correct number of classes. It is important to determine the right number of classes because specifying too few ignores class differences, while specifying too many may cause the model to be unstable. While the log-likelihood increases each time the number of classes is increased, the minimum *BIC* value occurs for Model3, suggesting that the 3-class solution is the best of the four estimated models.

Note that the R^2 statistic increases from .37 for the 1-class model to .61 for the 3-class regression indicating much better predictions for the 3-class model. We will now

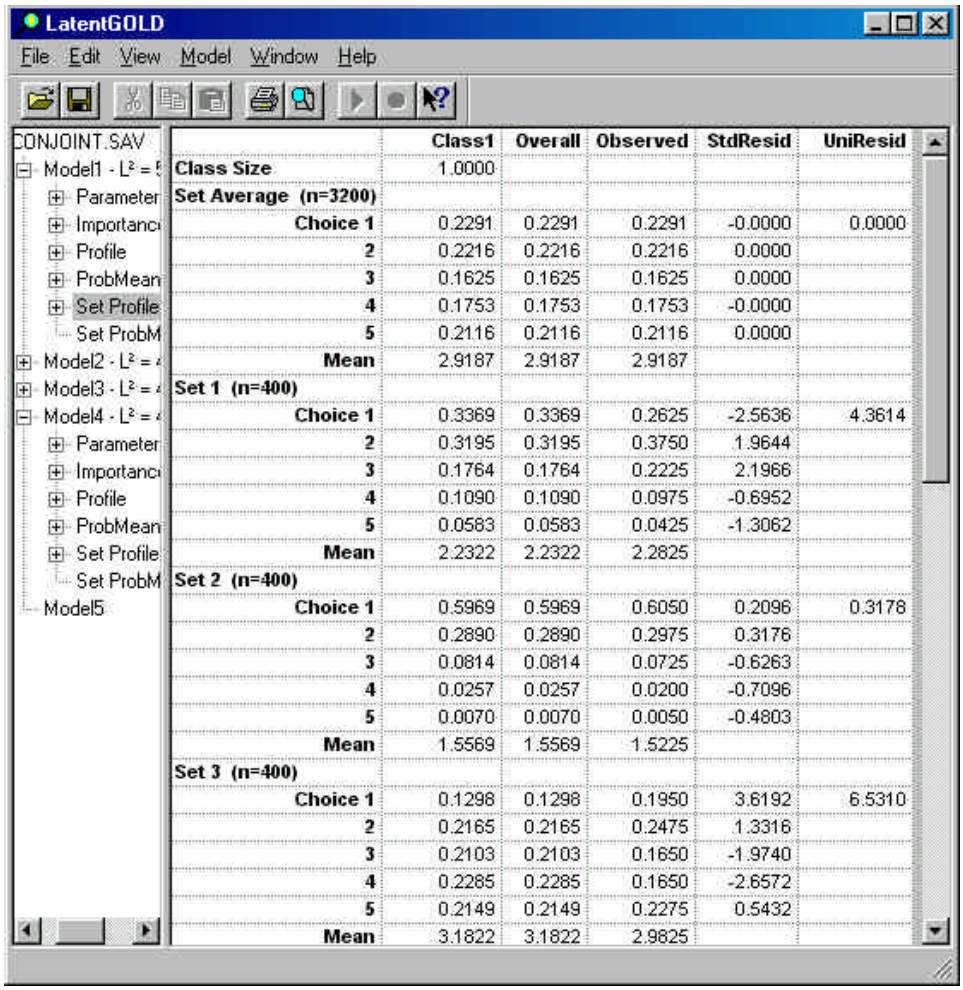
examine *new* output in the Latent GOLD Choice program that is not available in Latent GOLD 3.0 to compare the predictions 1-class and 3-class solutions in some detail. For discussion of the standard output, see the Regression tutorial #2.

Set Profile Output

- Click the + icon next to 'Model1' in the Outline pane to expand the listing of output for the 1-class model
- Click Set Profile

The Set Profile output presents *choice* probabilities predicted for each class (under the column headings 'Class1', 'Class2', etc.) as well as the *overall* probabilities for the entire sample (under the column heading 'Overall'). For ratings data, each choice probability corresponds to the probability of a particular rating (Very Unlikely, Somewhat Unlikely, etc.). The *overall* probabilities are compared with the corresponding proportions actually observed (under the 'Observed' column).

Additional columns contain Standardized Residuals which assess the difference between each *predicted* probability and the corresponding *observed* proportion, and a single test statistic for all probabilities for a given set (the uni-residual). This information is displayed at the top of the output file in aggregate for the average set (under the heading "Set Average"), followed by separate tables for each of the 8 sets (see **Figure 8**):.



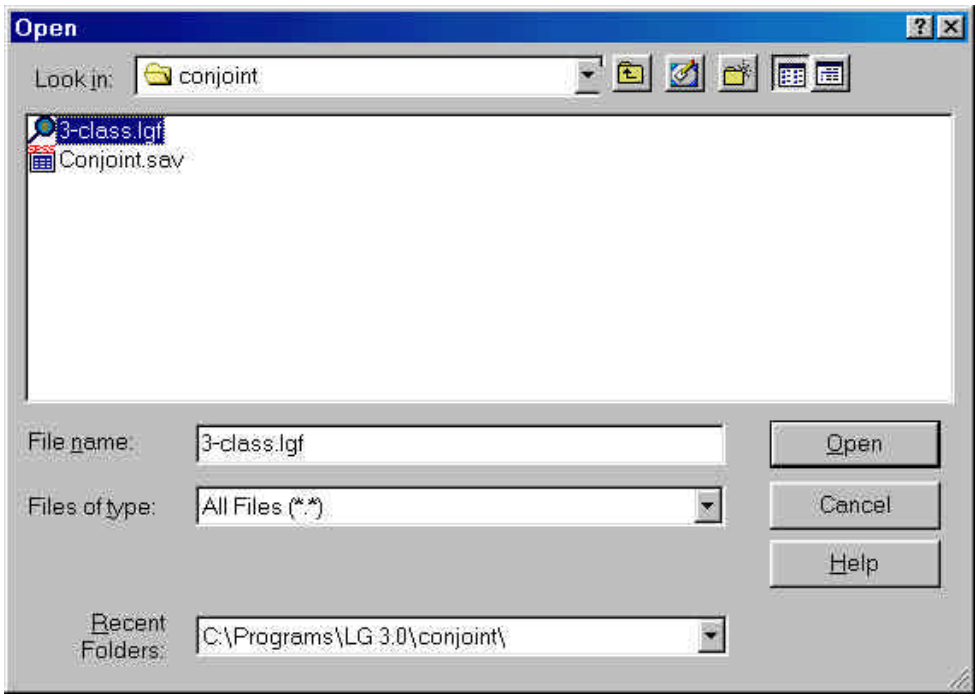
	Class1	Overall	Observed	StdResid	UniResid
Class Size	1.0000				
Set Average (n=3200)					
Choice 1	0.2291	0.2291	0.2291	-0.0000	0.0000
2	0.2216	0.2216	0.2216	0.0000	
3	0.1625	0.1625	0.1625	0.0000	
4	0.1753	0.1753	0.1753	-0.0000	
5	0.2116	0.2116	0.2116	0.0000	
Mean	2.9187	2.9187	2.9187		
Set 1 (n=400)					
Choice 1	0.3369	0.3369	0.2625	-2.5636	4.3614
2	0.3195	0.3195	0.3750	1.9644	
3	0.1764	0.1764	0.2225	2.1966	
4	0.1090	0.1090	0.0975	-0.6952	
5	0.0583	0.0583	0.0425	-1.3062	
Mean	2.2322	2.2322	2.2825		
Set 2 (n=400)					
Choice 1	0.5969	0.5969	0.6050	0.2096	0.3178
2	0.2890	0.2890	0.2975	0.3176	
3	0.0814	0.0814	0.0725	-0.6263	
4	0.0257	0.0257	0.0200	-0.7096	
5	0.0070	0.0070	0.0050	-0.4803	
Mean	1.5569	1.5569	1.5225		
Set 3 (n=400)					
Choice 1	0.1298	0.1298	0.1950	3.6192	6.5310
2	0.2165	0.2165	0.2475	1.3316	
3	0.2103	0.2103	0.1650	-1.9740	
4	0.2285	0.2285	0.1650	-2.6572	
5	0.2149	0.2149	0.2275	0.5432	
Mean	3.1822	3.1822	2.9825		

Figure 8: Set Profile Output

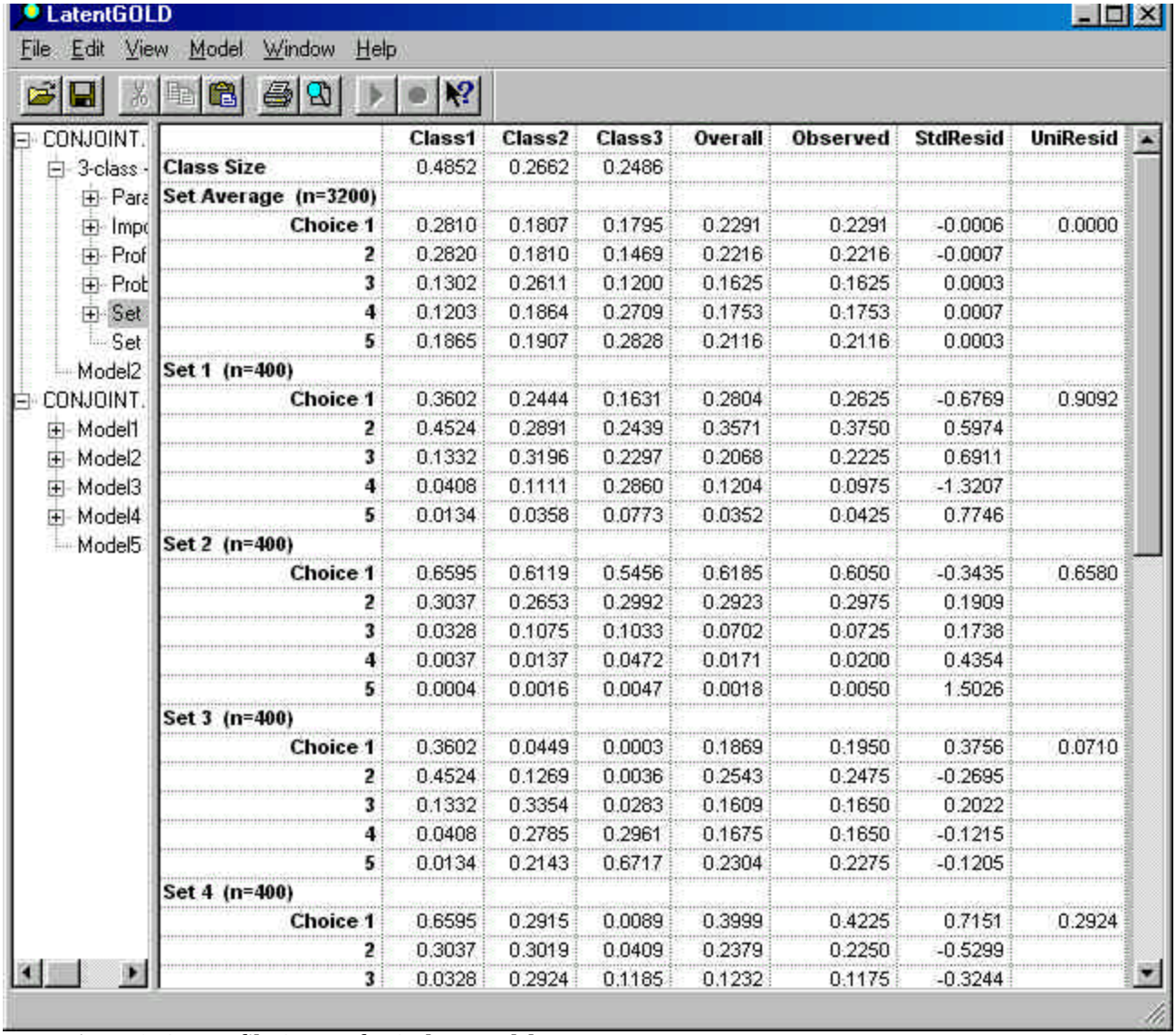
Notice that the predicted probabilities (and the mean rating) obtained under the 1-class model are very close to the corresponding observed values when aggregated over *all* sets ("Sets Average"). However, for several individual sets the differences between the predictions and the corresponding observed quantities fall outside statistical error limits. For example, for sets #1, #3, #4 and #8, the Uni-Residuals are all greater than 2, indicating that these predictions are significantly different from the observed quantities.

In contrast, the differences between the estimated probabilities and observed proportions obtained under the 3-class model are all within the accepted 95% confidence limits. Our *final* 3-class model obtained in Tutorial #2 included some restrictions. To retrieve that model, from the File Menu select

- Open '3-class.lgf' to retrieve the previously saved model named '3-class'



- Double-click on the model name '3-class'
- Click Estimate
- Click on the + icon to the left of the 3-class Model
- Click Set Profile to display the Set Profile output (see Figure 9)



	Class1	Class2	Class3	Overall	Observed	StdResid	UniResid
Class Size	0.4852	0.2662	0.2486				
Set Average (n=3200)							
Choice 1	0.2810	0.1807	0.1795	0.2291	0.2291	-0.0006	0.0000
2	0.2820	0.1810	0.1469	0.2216	0.2216	-0.0007	
3	0.1302	0.2611	0.1200	0.1625	0.1625	0.0003	
4	0.1203	0.1864	0.2709	0.1753	0.1753	0.0007	
5	0.1865	0.1907	0.2828	0.2116	0.2116	0.0003	
Set 1 (n=400)							
Choice 1	0.3602	0.2444	0.1631	0.2804	0.2625	-0.6769	0.9092
2	0.4524	0.2891	0.2439	0.3571	0.3750	0.5974	
3	0.1332	0.3196	0.2297	0.2068	0.2225	0.6911	
4	0.0408	0.1111	0.2860	0.1204	0.0975	-1.3207	
5	0.0134	0.0358	0.0773	0.0352	0.0425	0.7746	
Set 2 (n=400)							
Choice 1	0.6595	0.6119	0.5456	0.6185	0.6050	-0.3435	0.6580
2	0.3037	0.2653	0.2992	0.2923	0.2975	0.1909	
3	0.0328	0.1075	0.1033	0.0702	0.0725	0.1738	
4	0.0037	0.0137	0.0472	0.0171	0.0200	0.4354	
5	0.0004	0.0016	0.0047	0.0018	0.0050	1.5026	
Set 3 (n=400)							
Choice 1	0.3602	0.0449	0.0003	0.1869	0.1950	0.3756	0.0710
2	0.4524	0.1269	0.0036	0.2543	0.2475	-0.2695	
3	0.1332	0.3354	0.0283	0.1609	0.1650	0.2022	
4	0.0408	0.2785	0.2961	0.1675	0.1650	-0.1215	
5	0.0134	0.2143	0.6717	0.2304	0.2275	-0.1205	
Set 4 (n=400)							
Choice 1	0.6595	0.2915	0.0089	0.3999	0.4225	0.7151	0.2924
2	0.3037	0.3019	0.0409	0.2379	0.2250	-0.5299	
3	0.0328	0.2924	0.1185	0.1232	0.1175	-0.3244	

Figure 9: Set Profile Output for 3-class Model

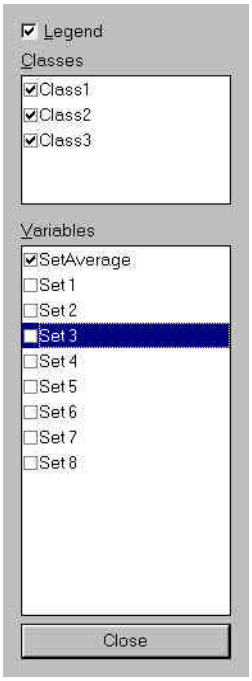
We see that all predictions are close to the corresponding observed quantities.

Set Profile: Profile Plot

Another new feature of Latent GOLD Choice is the Set Profile Prf-Plot, which provides a graphical representation of how the classes differ with respect to the predicted ratings.

- Click on the "+" symbol to the left of Set Profile and select Prf-Plot

- Right-click on the Plot to bring up the Plot Control Menu:



We will illustrate this display by plotting the mean rating predicted for the 2 sets corresponding to Traditional Lower Priced products. These sets (#1 and #3) differ only with respect to QUALITY (See Table 0-1).

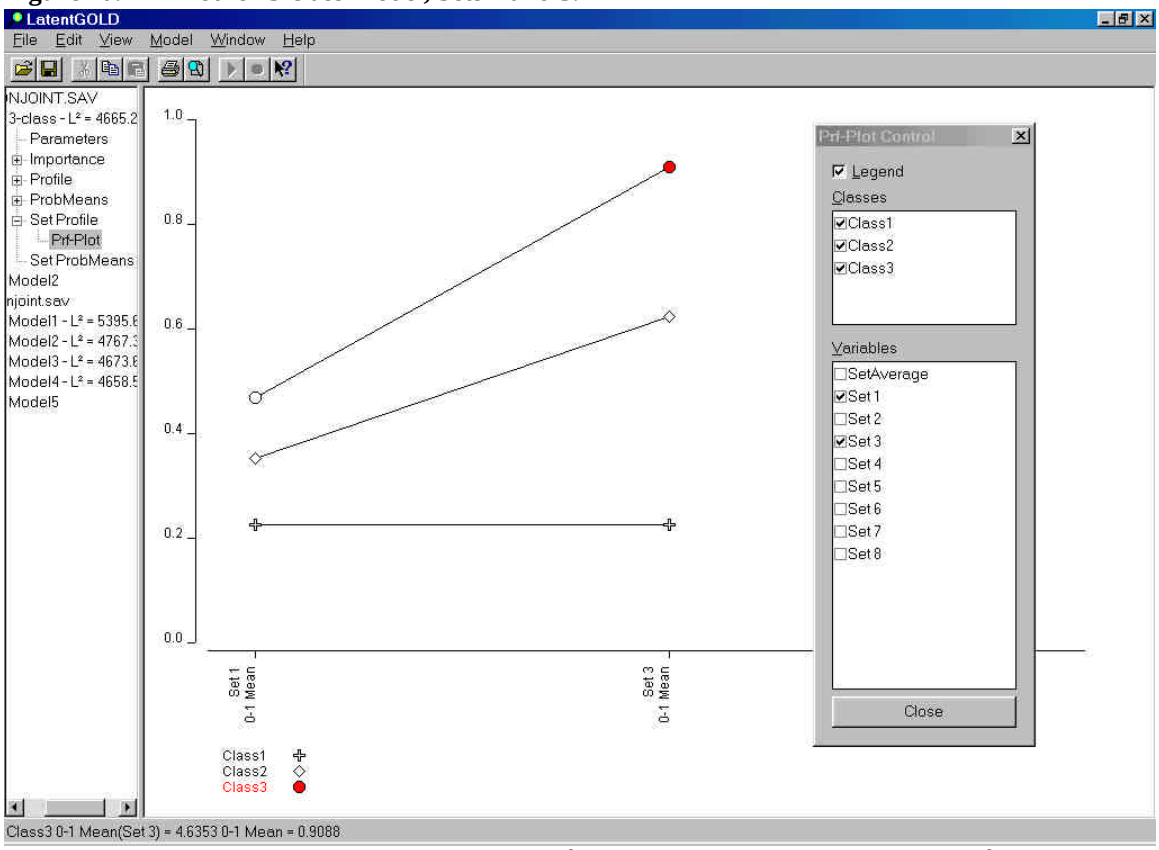
On the Plot Control Menu,

- click the box to the left of 'Set Average' to remove the checkmark and click 'Set 1' and also 'Set 3'

- Click on the circle symbol corresponding to Set 3 for Class 3

Your screen should now look like Figure 10:

Figure 10: Prf-Plot for 3-class Model, Sets 1 and 3.



Note that line associated with Class 1 is perfectly horizontal, indicating that for this class the mean rating is identical for both sets. This class is not influenced by Quality. Examining the slopes of the lines associated with the other 2 classes we see that Class 2 is influenced more by Quality, while Class 3 is the most affected.

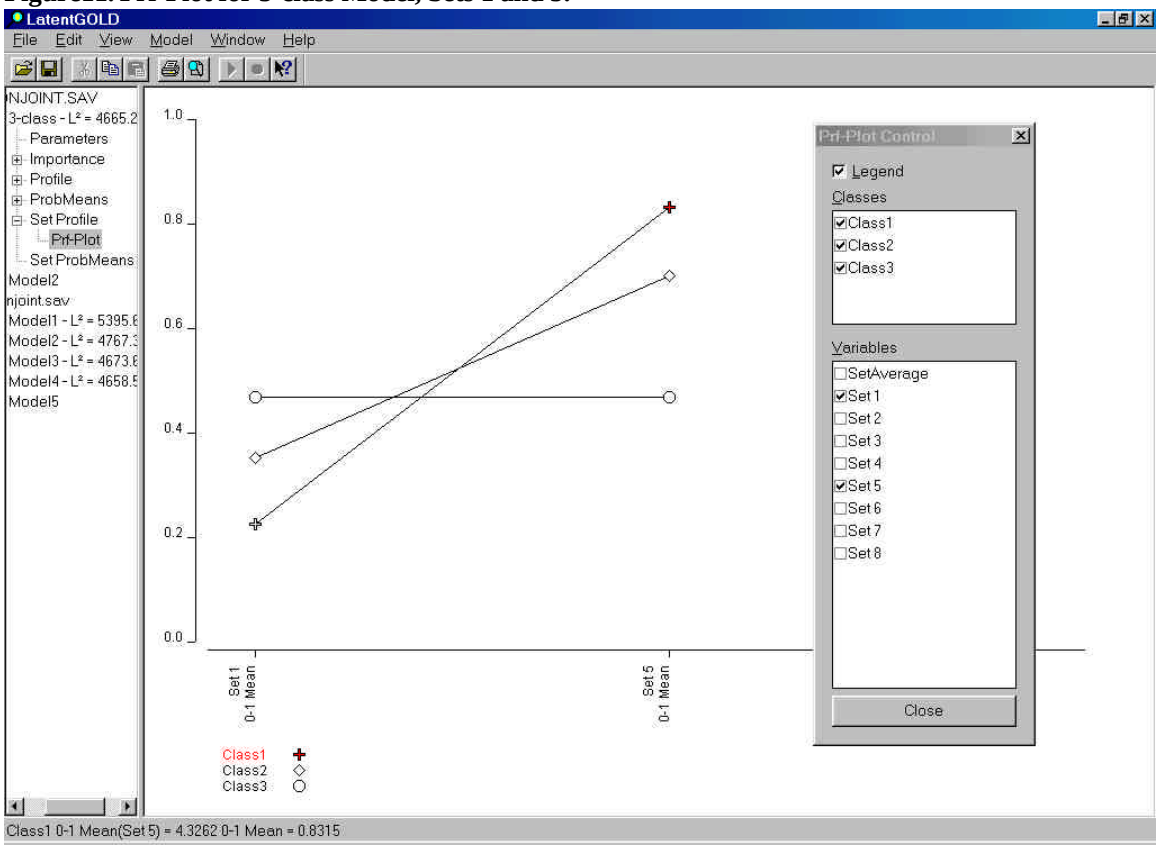
To fit all scores on the same plot, the scores are scaled to the range 0-1. To view the actual mean rating predicted for a given product, click on the desired symbol on the plot. On the Status bar at the bottom of the screen a message appears that indicates the predicted mean rating for this set is 4.6353, which corresponds to a 0-1 Mean of 0.9088.

Now, let us select Sets 1 and 5, which differ only with respect to FASHION. On the Plot Control Menu,

- click Set 3 to remove it from the plot
- select Set 5 to include this new set
- click on the symbol corresponding to Set 5 for Class 1 to highlight it

Your screen now looks like Figure 11:

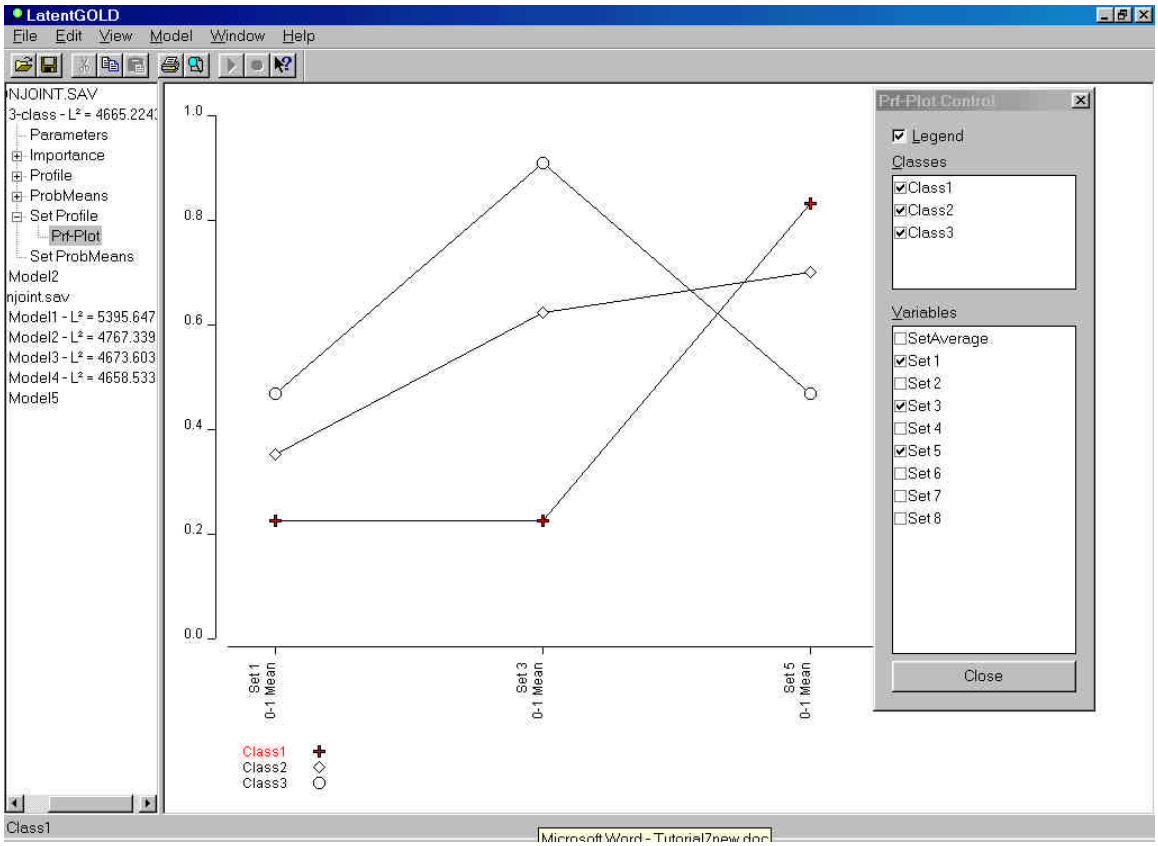
Figure11. Prf-Plot for 3-class Model, Sets 1 and 5.



By comparing the slopes of these lines, we see that Class 1 is the most influenced, Class 2 somewhat influenced and Class 3 not influenced at all by Fashion.

Now, go back to the Plot Control Menu and select Set 3 once again to view the effects of all 3 sets together, shown in Figure 12:

Figure 12: Prf-Plot for 3-class Model, Sets 1, 3 and 5.

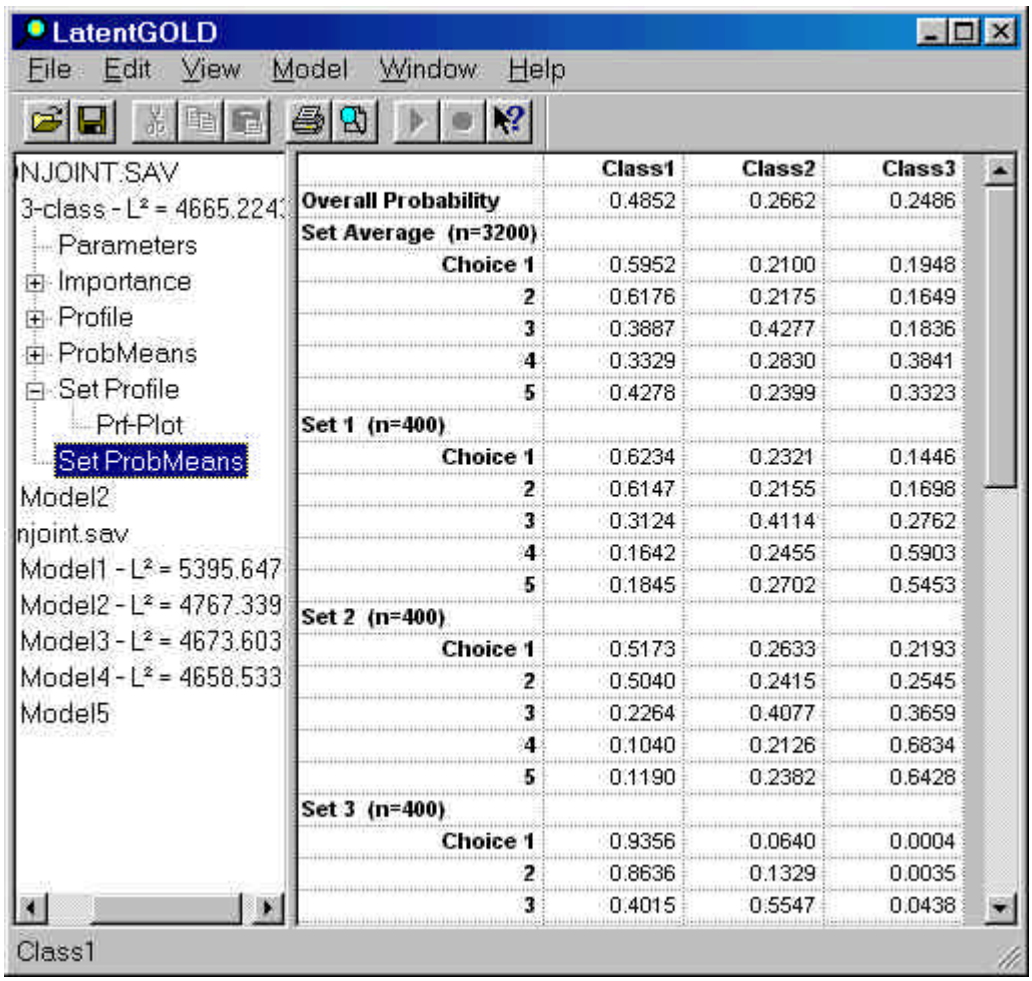


Set ProbMeans

The other additional output not available in the Regression module is the Set Probmeans output. In this file, the column percentages are converted to row percentages, which can then be displayed in the Tri-plot.

- Click on Set ProbMeans to bring up the output shown in Figure 13.

Figure 13: Set ProbMeans output for 3-class Model



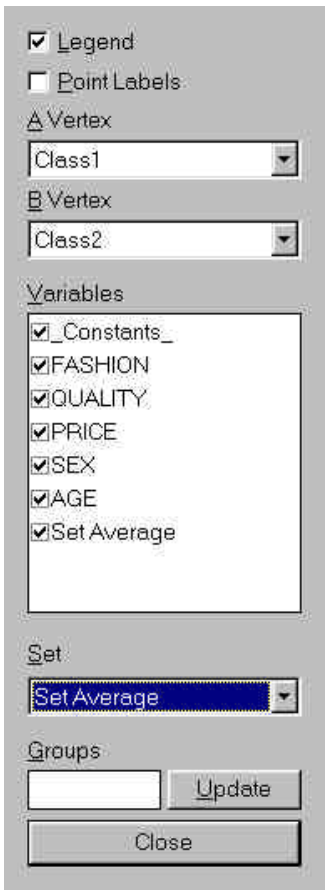
	Class1	Class2	Class3
Overall Probability	0.4852	0.2662	0.2486
Set Average (n=3200)			
Choice 1	0.5952	0.2100	0.1948
2	0.6176	0.2175	0.1649
3	0.3887	0.4277	0.1836
4	0.3329	0.2830	0.3841
5	0.4278	0.2399	0.3323
Set 1 (n=400)			
Choice 1	0.6234	0.2321	0.1446
2	0.6147	0.2155	0.1698
3	0.3124	0.4114	0.2762
4	0.1642	0.2455	0.5903
5	0.1845	0.2702	0.5453
Set 2 (n=400)			
Choice 1	0.5173	0.2633	0.2193
2	0.5040	0.2415	0.2545
3	0.2264	0.4077	0.3659
4	0.1040	0.2126	0.6834
5	0.1190	0.2382	0.6428
Set 3 (n=400)			
Choice 1	0.9356	0.0640	0.0004
2	0.8636	0.1329	0.0035
3	0.4015	0.5547	0.0438

For example, among those who gave the rating 1 (“Very Unlikely’) to Set 1, we see that most (62.34%) are in Class 1.

Probmeans: Tri-Plot

To include this information in the Tri-plot display:

- Click on the "+" symbol to the left of ProbMeans in the Outline Pane
- Select Tri-Plot
- Right-click on the plot to bring up the Plot Control Menu



- Click "point labels" to remove these from the display to make the graph a little easier to read
- In the Sets Box, select ‘Set 1’ to plot the Set 1 responses in segment space (i.e., with respect to Classes 1, 2 and 3).
- click the symbol for Set 1 to highlight it and make it more visible.

Your graph should now look like the graph in Figure 14:

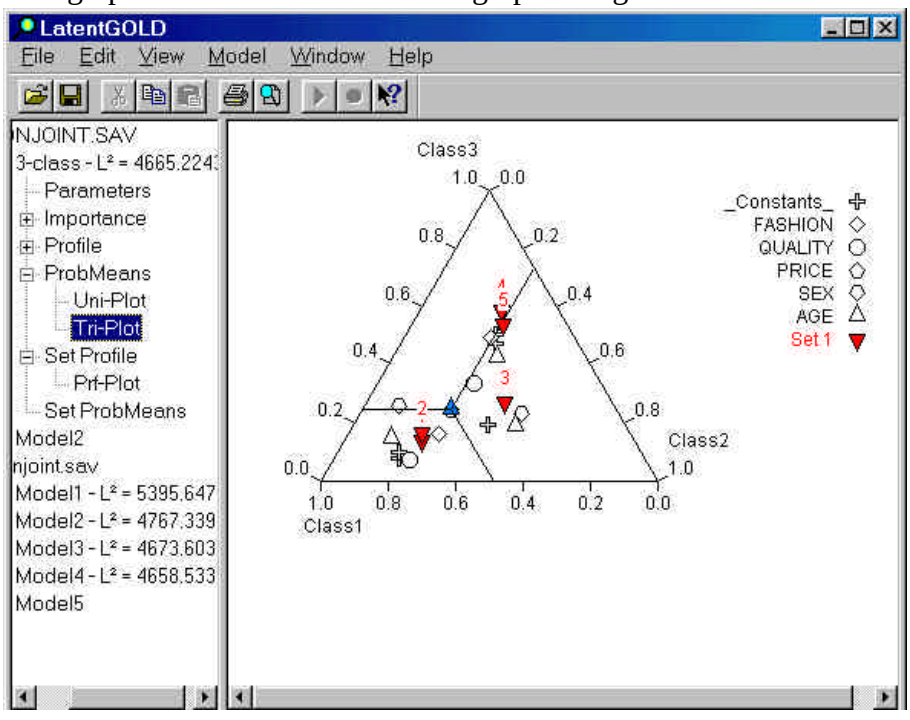


Figure 14: Tri-Plot for Set 1

Notice that the lower ratings (‘1’ and ‘2’) are closer to the lower left vertex associated with Class 1 while the higher ratings (‘4’ and ‘5’) are closer to the top vertex associated with Class 3.

- Now, select Set 3 (the Higher Quality product) on the Plot Control Menu and notice that these patterns are even more pronounced (Figure 15)

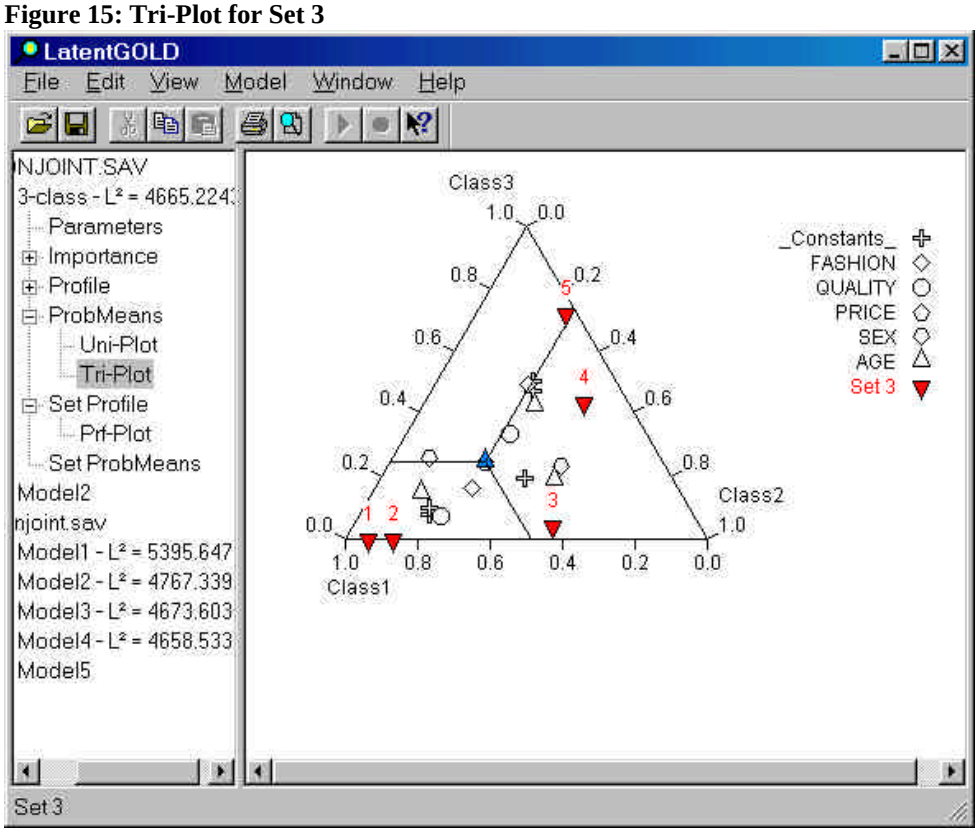



Figure 15: Tri-Plot for Set 3