7.3 Tutorial #3: LC Regression with Repeated Measures

DemoData = 'conjoint.sav'

This tutorial shows how to develop Latent Class (LC) Regression models using the sample data file "conjoint.sav". You will learn how to:

- Select the dependent variable and specify its scale type
- Distinguish predictors from covariates
- Impose restrictions on the predictor effects
- Specify covariates as active or inactive
- Determine the number of latent classes (i.e., segments)
- Examine R^2 and various other information related to model prediction

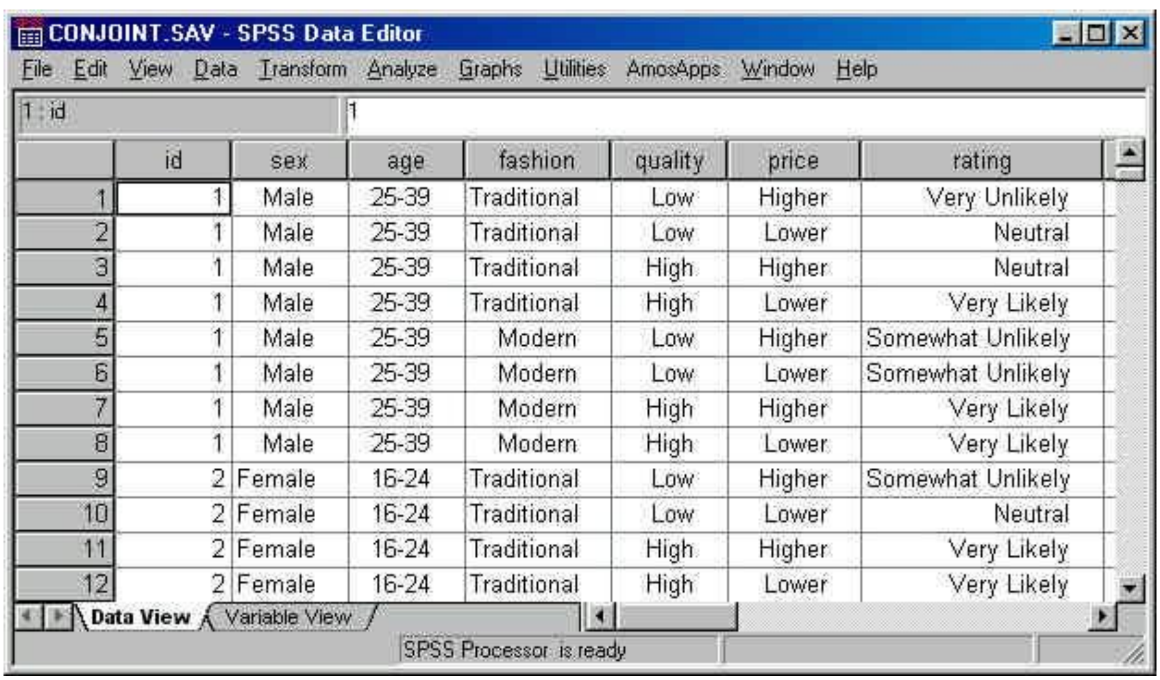
In addition, this example illustrates several advanced options in the LC Regression Module. You will learn how to:

- Use the optional case ID variable to specify repeated observations
- Explore the Profile and ProbMeans output
- Use demographic variables as covariates to predict segment membership
- Obtain predictions based solely on the covariates
- Classify cases into latent segments

The Data

The data for this example are obtained from a hypothetical conjoint marketing study involving repeated measures where respondents were asked to provide likelihood of purchase ratings under each of several different scenarios. A partial listing of the data is shown in Figure 7-55.

Figure 7-55: Partial Listing of Conjoint Data



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

As suggested in Figure 7-55, there are 8 records for each case (there are 400 cases in total); one record for each cell in this 2x2x2 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. The three attributes listed above will be used as predictor variables in the model.

We will also include the two demographic variables as covariates, in a second model.

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

The Goal

Use Latent GOLD to identify latent segments differing with respect to the estimate of importance attached to each of the three attributes, which influence an individual's purchase decision. The LC regression model allows for the fact that these estimates may differ for different segments. That is, for one segment, price and only price may influence the decision, while a second segment may be influenced by quality and modern appearance, but is price insensitive. We will treat *RATING* as an *ordinal* dependent variable and compare several models to determine the number of segments. We will then show how to describe the demographic differences between these segments and to classify each respondent into that segment which is most likely.

Estimating an LC Regression Model

Opening a Data File and Selecting the Type of Model

For this example, the data file being used is an SPSS system file.

- 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 7-56).

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.

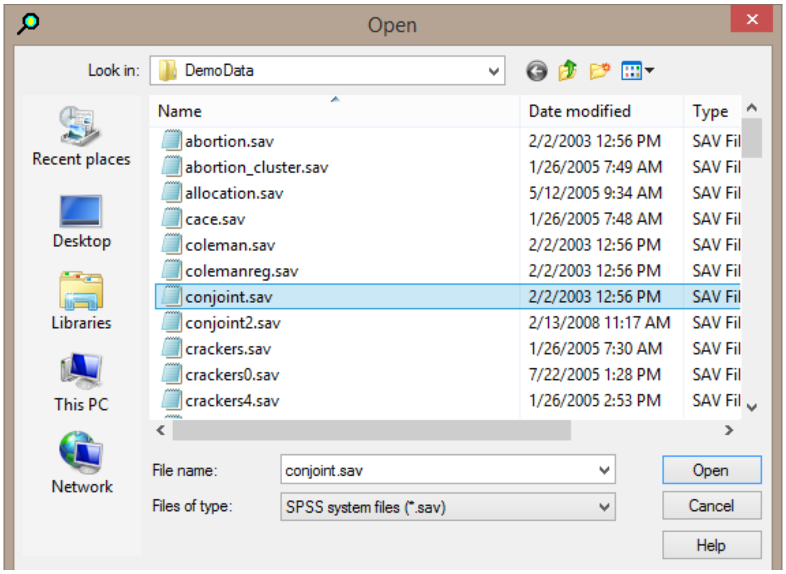


Figure 7-56: Open Dialog Box

- ➢ Select conjoint.sav and click Open to open the Viewer window.
- ➢ Highlight 'Model1' if it is not already highlighted.
- ➢ 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).
- ➢ Select Regression and the LC Regression analysis dialog box, which contains 6 tabs, will open (see Fig. 7-57).

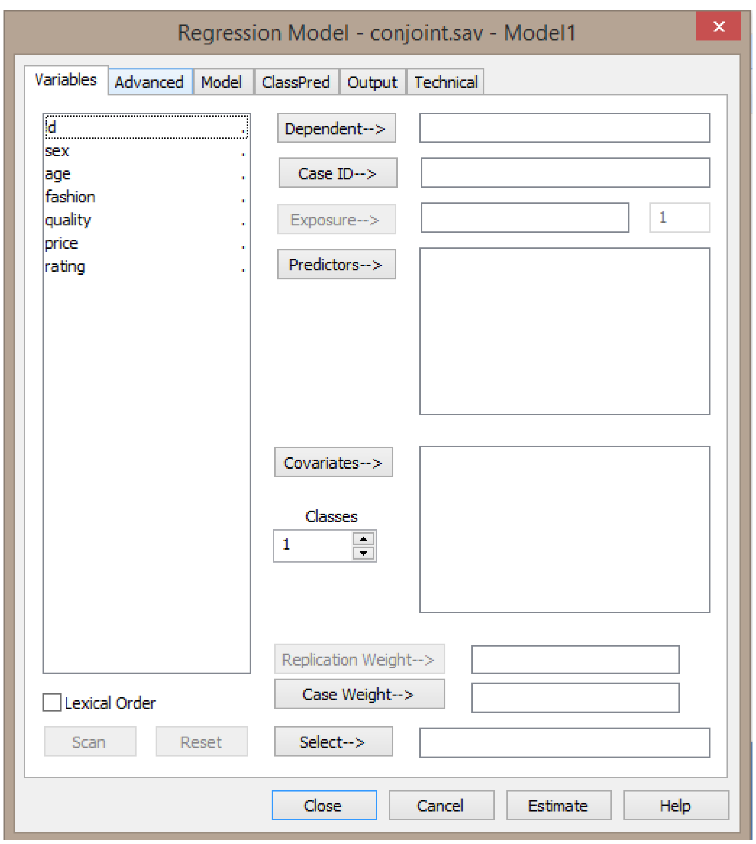


Figure 7-57: Analysis Dialog Box for LC Regression Model

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.

We also need to indicate the dependent variable scale type. For this example, we will use the default scale type (Ordinal-Fixed) which takes into account the natural ordering between the 5 levels of purchase intent. By default, the fixed scores on the data file (1, 2, 3, 4 and 5) are used which order the levels and establish equal distance between adjacent levels.

As explained 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.

Next, we will select the Predictors. Predictors are used as independent variables in the regression model. In the current example, we use the product attributes *FASHION*, *QUALITY* and *PRICE* as predictors.

- Select *FASHION*, *QUALITY* and *PRICE* in the Variables list and click Predictors to move the variables into the Predictors box.

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, which successively increment the number of classes by 1 and assess the significance of each additional class.

One assessment consists of a check of whether the change in the log-likelihood for each pair of successive models fails to decrease by a significant amount as determined by the *BIC* statistic. (The model having the lowest *BIC* might then be selected.) A second assessment is to utilize the p-value associated with the L^2 fit statistic.

- 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.

Scanning the Data File

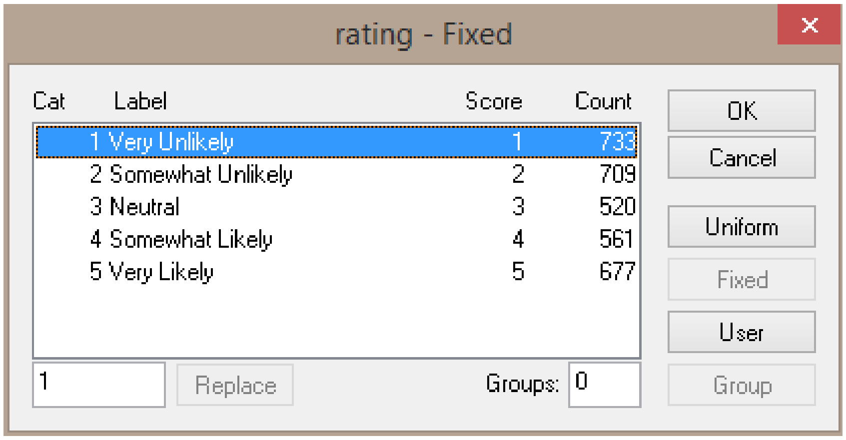
- Click Scan (located in the lower left of the Analysis dialog box) to scan the data file.

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 will open (see Figure 7-58).

- ➢ Double click the dependent variable, *RATING*.

Figure 7-58: Variables box for RATING



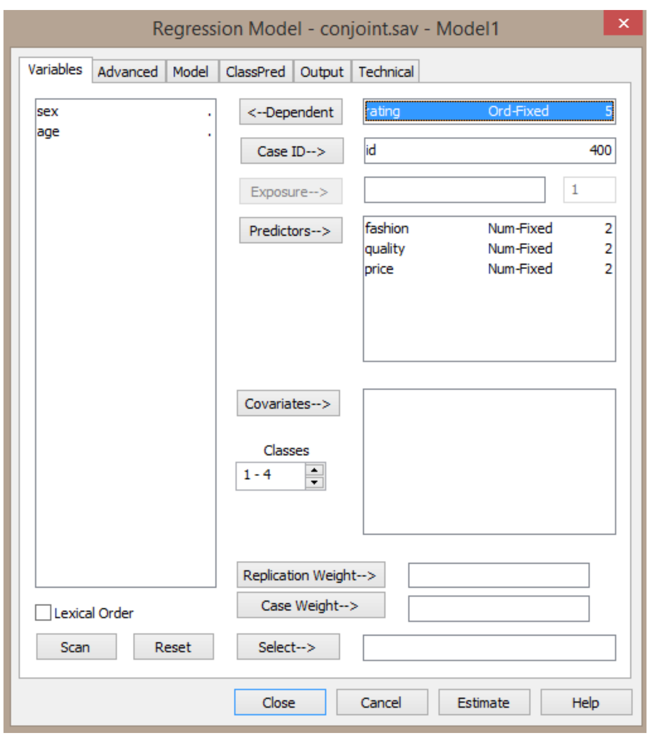
- ➢ Click OK to close the Variables dialog box and return to the Regression Analysis dialog box.

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 7-59.

Figure 7-59: 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 Regression models, several output files are produced. To view a summary of the models estimated (Figure 7-60),

- ➢ Click on the data file name, conjoint.sav in the Outline pane.

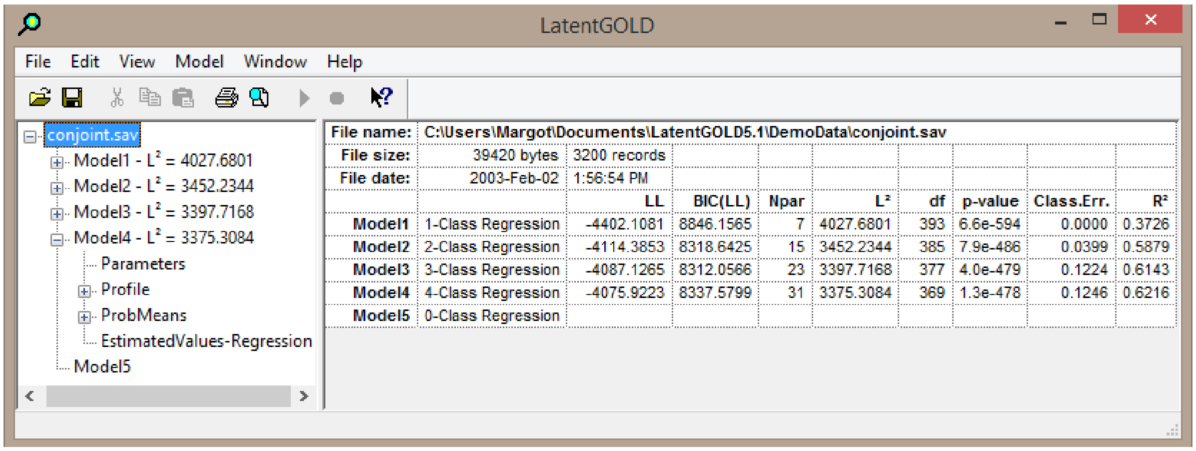


Figure 7-60: Summary of Models Estimated

This output reports statistics that will assist you in determining the correct number of classes -- the **log-likelihood** (*LL*) values, the *BIC* values, and the number of parameters in the estimated models. 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. The R^2 increases from .37 for the 1-class model to .61 for the 3-class regression.

Note 1: Occasionally, you might obtain a local (suboptimal) solution. For these data, it is possible to obtain a local solution for the 4-class model, obtaining *LL* = -4080.318 instead of –4075.922. If this occurs, click Estimate to re-estimate the 4-class model (see section 6.6 in the *Technical Guide* for a discussion of preventing local solutions).

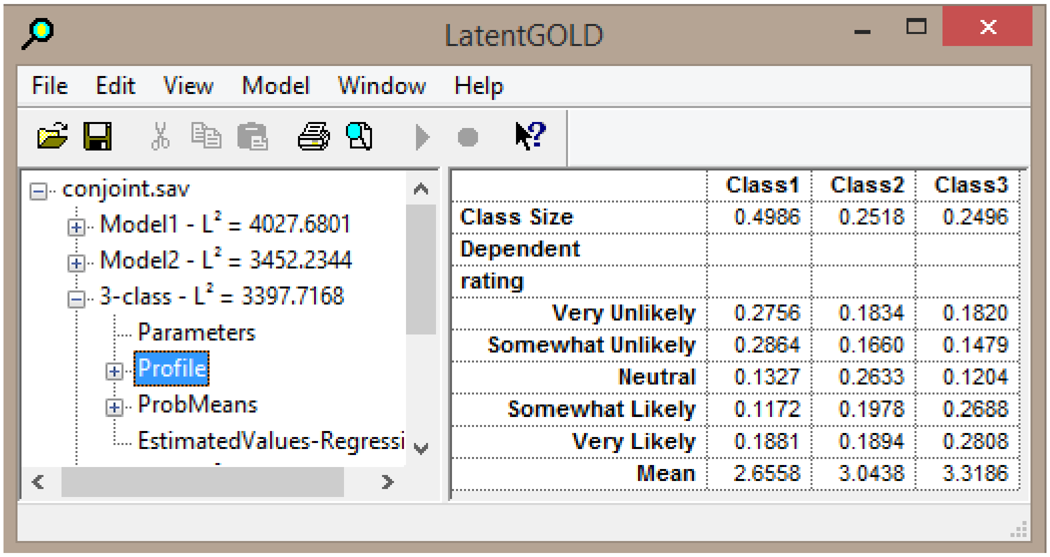
Note 2: Notice that the p-values based on the model L^2 and reported degrees of freedom (df) are not valid assessments of fit because we are dealing with sparse data.

We will now examine the detailed output for the 3-class solution.

Profile Output

- ➢ Rename Model3 to "3-class" by clicking on its name.
- ➢ Click the + icon next to ‘3-class’ in the Outline pane to expand the listing of output for this model.
- ➢ Click Profile.

The Profile output contains information on the class sizes, the class-specific (marginal) probabilities and means of the dependent variable (see Figure 7-61).



	Class1	Class2	Class3
Class Size	0.4986	0.2518	0.2496
Dependent			
rating			
Very Unlikely	0.2756	0.1834	0.1820
Somewhat Unlikely	0.2864	0.1660	0.1479
Neutral	0.1327	0.2633	0.1204
Somewhat Likely	0.1172	0.1978	0.2688
Very Likely	0.1881	0.1894	0.2808
Mean	2.6558	3.0438	3.3186

Figure 7-61: Profile Output for 3-Class Model

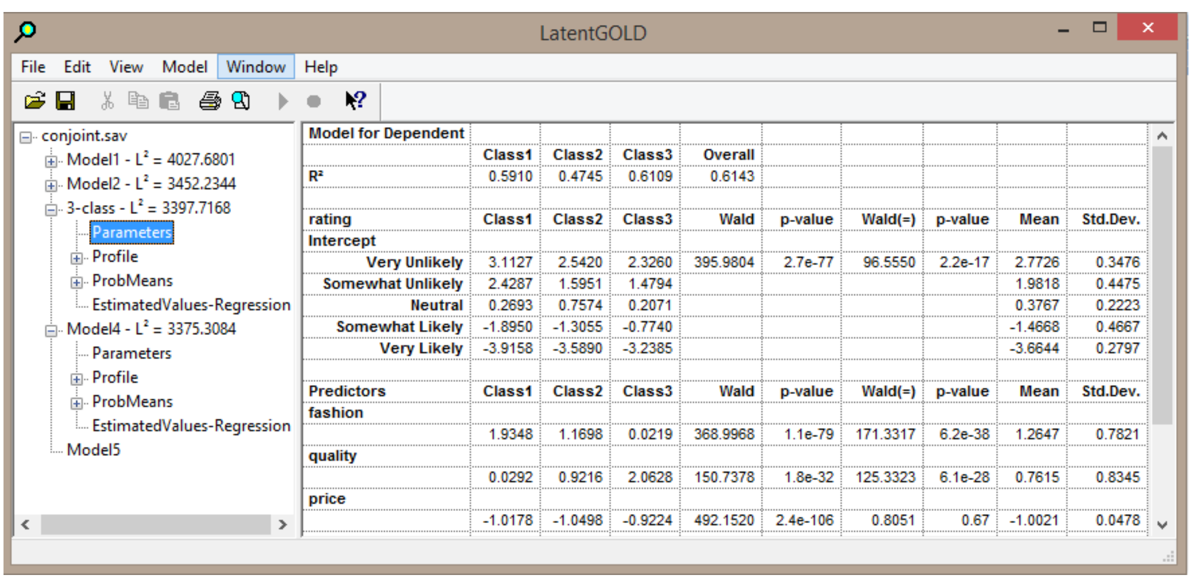
The classes are always ordered from high to low according to their size. It can be seen from the first row of the table that segment 1 contains about 50% of the subjects (.4986), segment 2 contains about 25% and segment 3 contains the remaining 25%.

Examination of class-specific probabilities shows that overall, segment 1 is least likely to buy (only 27.56% are *Very Unlikely* to buy) and segment 3 is most likely (28.08% are *Very Likely* to buy). Later in this tutorial, we will show how to classify each case into the most appropriate segment.

Parameters Output

Next, we will view the Parameters output (see Figure 7-62).

- ➢ For the '3-class' model, click Parameters.



Model for Dependent									
	Class1	Class2	Class3	Overall					
R²	0.5910	0.4745	0.6109	0.6143					
rating	Class1	Class2	Class3	Wald	p-value	Wald(=)	p-value	Mean	Std.Dev.
Intercept									
Very Unlikely	3.1127	2.5420	2.3260	395.9804	2.7e-77	96.5550	2.2e-17	2.7726	0.3476
Somewhat Unlikely	2.4287	1.5951	1.4794					1.9818	0.4475
Neutral	0.2693	0.7574	0.2071					0.3767	0.2223
Somewhat Likely	-1.8950	-1.3055	-0.7740					-1.4668	0.4667
Very Likely	-3.9158	-3.5890	-3.2385					-3.6644	0.2797
Predictors	Class1	Class2	Class3	Wald	p-value	Wald(=)	p-value	Mean	Std.Dev.
fashion									
	1.9348	1.1698	0.0219	368.9968	1.1e-79	171.3317	6.2e-38	1.2647	0.7821
quality									
	0.0292	0.9216	2.0628	150.7378	1.8e-32	125.3323	6.1e-28	0.7615	0.8345
price									
	-1.0178	-1.0498	-0.9224	492.1520	2.4e-106	0.8051	0.67	-1.0021	0.0478

Figure 7-62: Parameters Output for 3-Class Model

The beta parameter for each predictor is a measure of the influence of that predictor on *RATING*. The beta effect estimates under the column labeled Class 1 suggest that segment 1 is influenced in a positive way by products for which *FASHION* = *Modern* (beta = 1.9348), in a negative way by *PRICE* (beta = -1.0178), and not at all by *QUALITY* (beta is approximately 0). We also see that segment 2 is influenced by all 3 attributes, having a preference for those product choices that are *modern* (beta = 1.1698), and *higher quality* (beta = .9216), but, like segment 1, their preference also decreases as a function of price (beta = -1.0498). Members of segment 3 prefer *higher quality products* (beta = 2.0628), but their preference also decreases as a function of price (beta = -.9224), and they are not influenced by *FASHION*.

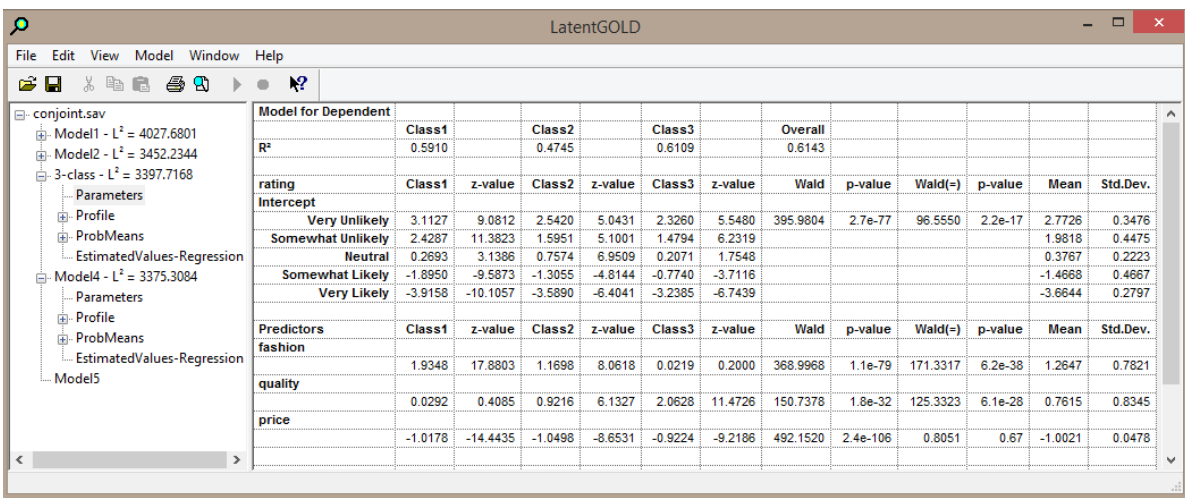
Note that *PRICE* has more or less the same influence on all three segments. The Wald (=) statistic indicates that the differences in these beta effects across classes are not significant (the *p*-value = .67 which is much higher than .05, the standard level for assessing statistical significance). This means that all 3 segments exhibit price sensitivity to the same degree. This is confirmed when we estimate a model in which this effect is specified to be class-independent (see next section). The *p*-value for the Wald statistic for *PRICE* is 2.4×10^{-106} indicating that the amount of price sensitivity is highly significant.

With respect to the effect of the other two attributes we find large between-segment differences. The predictor *FASHION* has a strong influence on segment 1, a less strong effect on segment 2, and virtually no effect on segment 3. *QUALITY* has a strong effect on segment 3, a less strong effect on segment 2, and virtually no effect on segment 1. The fact that the influence of *FASHION* and *QUALITY* differs significantly between the 3 segments is confirmed by the significant *p*-values associated with the Wald(=) statistics for these attributes. For example, for *FASHION*, the *p*-value = 6.2×10^{-36}.

In summary, segment 1 could be labeled the "Fashion-Oriented Segment", segment 3 the "Quality-Oriented Segment", and segment 2 is the segment that takes into account all 3 attributes in their purchase decision.

To test each individual class-specific beta for statistical significance we can append standard errors, Z-statistics, or both to the output.

Right click on the output in the Contents Pane and choose Z Statistic:The Wald statistics are replaced by the Z-statistics for the betas. Notice that the absolute values of the *z*- score associated with QUALITY for class 1 and with FASHION for class 3 fall under 2, and hence are not significant at the .05 level.



Model for Dependent												
	Class1		Class2		Class3		Overall					
R²	0.5910		0.4745		0.6109		0.6143					
rating	Class1	z-value	Class2	z-value	Class3	z-value	Wald	p-value	Wald(=)	p-value	Mean	Std.Dev.
Intercept												
Very Unlikely	3.1127	9.0812	2.5420	5.0431	2.3260	5.5480	395.9804	2.7e-77	96.5550	2.2e-17	2.7726	0.3476
Somewhat Unlikely	2.4287	11.3823	1.5951	5.1001	1.4794	6.2319					1.9818	0.4475
Neutral	0.2693	3.1386	0.7574	6.9509	0.2071	1.7548					0.3767	0.2223
Somewhat Likely	-1.8950	-9.5873	-1.3055	-4.8144	-0.7740	-3.7116					-1.4668	0.4667
Very Likely	-3.9158	-10.1057	-3.5890	-6.4041	-3.2385	-6.7439					-3.6644	0.2797
Predictors	Class1	z-value	Class2	z-value	Class3	z-value	Wald	p-value	Wald(=)	p-value	Mean	Std.Dev.
fashion												
	1.9348	17.8803	1.1698	8.0618	0.0219	0.2000	368.9968	1.1e-79	171.3317	6.2e-38	1.2647	0.7821
quality												
	0.0292	0.4085	0.9216	6.1327	2.0628	11.4726	150.7378	1.8e-32	125.3323	6.1e-28	0.7615	0.8345
price												
	-1.0178	-14.4435	-1.0498	-8.6531	-0.9224	-9.2186	492.1520	2.4e-106	0.8051	0.67	-1.0021	0.0478

Figure 7-63: Parameters Output with Z-values

Restricting Certain Effects to be Zero or Class Independent

In the Parameters output above we saw that the beta estimates associated with *PRICE* are approximately equal for all 3 classes. To test the null hypothesis of equality, we used the Wald(=) statistic. The low value of .67 was too small to reject this null hypothesis. We also showed that 2 of the betas were not significantly different from 0. We will now show how to obtain a more parsimonious model by imposing zero restrictions on the 2 betas and by restricting the betas associated with *PRICE* to be equal across segments. This is accomplished using the Model Tab.

To specify the *PRICE* effects to be class independent,

- Double click '3-class' to open the Analysis Dialog Box for this model
- Left click on the Model Tab
- In the row for PRICE, right click on the Class Independent column and select 'Yes'.

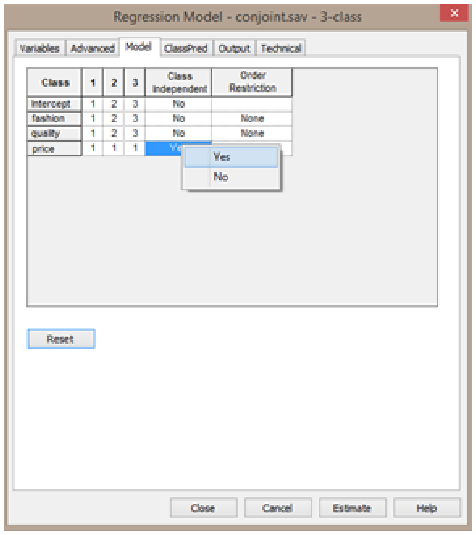


Figure 7-64. Setting the Class-Independent effects for the variable PRICE

To specify the betas to be zero,

- Right click on the cells corresponding to the betas to be set to zero
- Select No Effect

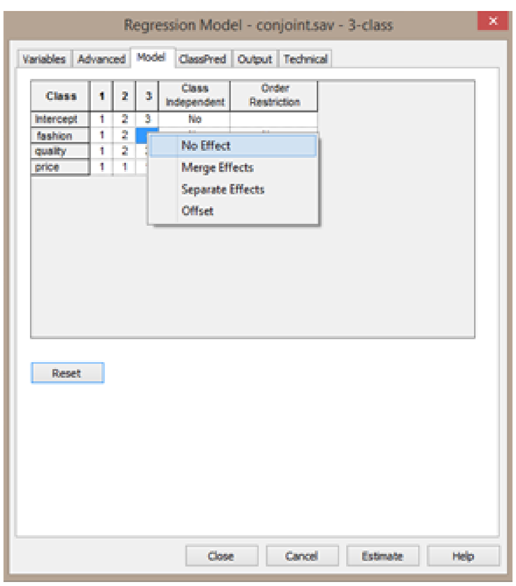


Figure 7-65. Specifying betas to 0 for variable FASHION

- Click on Variables to return to the Variables Tab.

An '=' will appear to the right of the variable PRICE to indicate the class independent restriction.

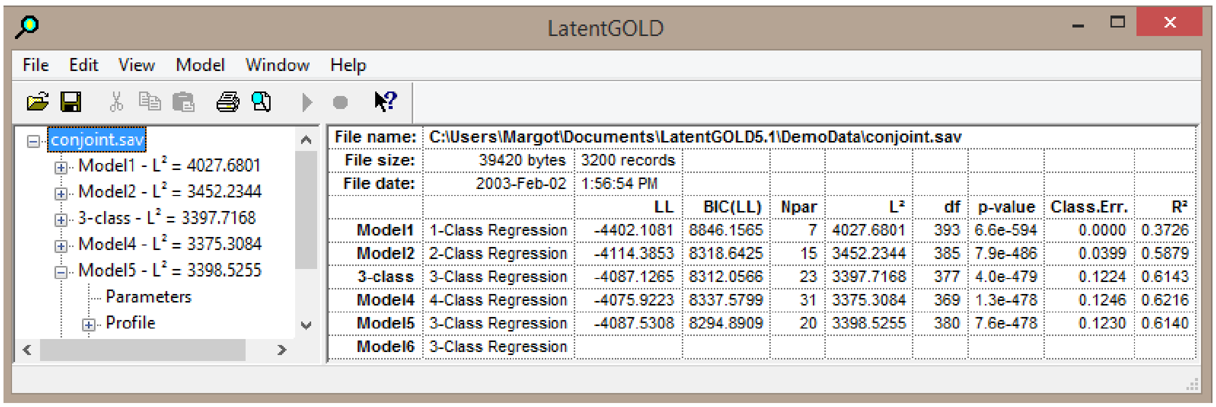
- ➢ Click Estimate to re-estimate this model. The

output for this new model will be listed as Model5.

Viewing Output and Interpreting Results

To view the summary output,

- ➢ Click on the data file name, conjoint.sav in the Outline pane.



File name:	C:\Users\Margot\Documents\LatentGOLD5.1\DemoData\conjoint.sav								
File size:	39420 bytes	3200 records							
File date:	2003-Feb-02	1:56:54 PM							
		LL	BIC(LL)	Npar	L^2	df	p-value	Class.Err.	R^2
Model1	1-Class Regression	-4402.1081	8846.1565	7	4027.6801	393	6.6e-594	0.0000	0.3726
Model2	2-Class Regression	-4114.3853	8318.6425	15	3452.2344	385	7.9e-486	0.0399	0.5879
3-class	3-Class Regression	-4087.1265	8312.0566	23	3397.7168	377	4.0e-479	0.1224	0.6143
Model4	4-Class Regression	-4075.9223	8337.5799	31	3375.3084	369	1.3e-478	0.1246	0.6216
Model5	3-Class Regression	-4087.5308	8294.8909	20	3398.5255	380	7.6e-478	0.1230	0.6140
Model6	3-Class Regression								

Figure 7-66: Summary Output

The new model estimated has been added to the bottom of the list as Models 5. For Model 5, the fit is almost identical to Model 3 and we obtain a lower (better) *BIC* value.

Parameters Output

- ➢ Click Parameters for Model5 to view the results containing the desired restrictions.

For *PRICE*, which we specified as class independent, note that the Wald(=) is now zero because the betas have been restricted to be exactly equal to each other across classes.

Adding Covariates

There is one important topic left with respect to the specification of LC regression models; that is, the use of covariates. In the Covariates list box, we can specify variables that we want to use to predict class membership. For this example we will re-estimate the 3-class model, this time including *SEX* and *AGE* as covariates.

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

- ➢ In the Outline pane, double click Model5 to open the Analysis dialog box for this model. Latent GOLD has maintained our previous settings (we will keep *PRICE* set as class independent).
- ➢ Select *SEX* and *AGE* in the Variables list box.
- ➢ Click Covariates to move these variables to the Covariates box.
- ➢ Right click on the variable names and select scale type Nominal.

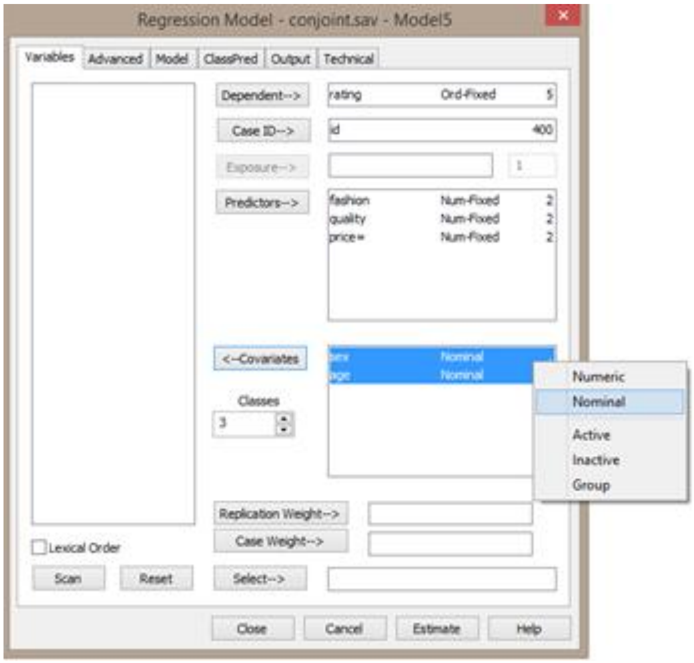


Figure 7-67: Specification of Model with Covariates

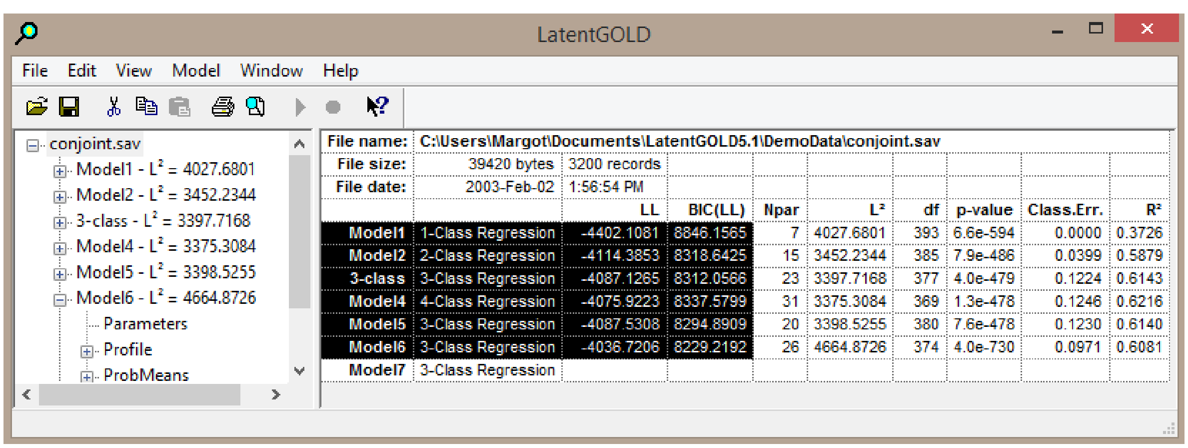
- ➢ Click Estimate. The output for this new model will be listed as Model6.

Viewing Output and Interpreting Results

To view the summary output,

- ➢ Click on the data file name, conjoint.sav in the Outline pane.

Figure 7-68: Summary Output

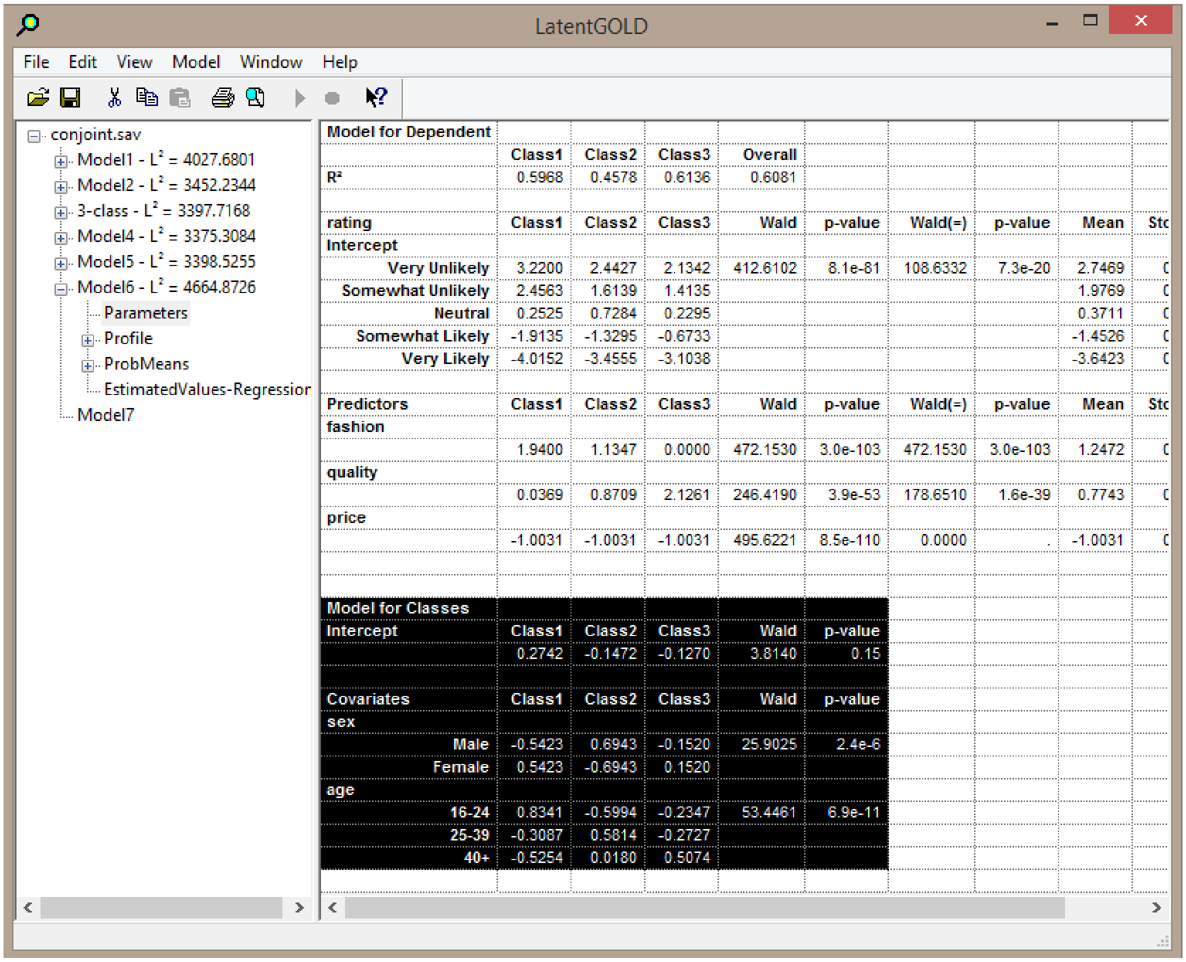


The new model has been added to the bottom of the list as Model 6. Note that this model is even better than our previous 3-class models as indicated by the lower *BIC* value.

Parameters Output

➢ Click Parameters for Model6 to view the results reported in Figure 7-69.

Figure 7-69: Parameters Output for 3-Class Model with Covariates



Model for Dependent	Class1	Class2	Class3	Overall
R²	0.5968	0.4578	0.6136	0.6081

rating	Class1	Class2	Class3	Wald	p-value	Wald(=)	p-value	Mean
Intercept								
Very Unlikely	3.2200	2.4427	2.1342	412.6102	8.1e-81	108.6332	7.3e-20	2.7469
Somewhat Unlikely	2.4563	1.6139	1.4135					1.9769
Neutral	0.2525	0.7284	0.2295					0.3711
Somewhat Likely	-1.9135	-1.3295	-0.6733					-1.4526
Very Likely	-4.0152	-3.4555	-3.1038					-3.6423

Predictors	Class1	Class2	Class3	Wald	p-value	Wald(=)	p-value	Mean
fashion								
	1.9400	1.1347	0.0000	472.1530	3.0e-103	472.1530	3.0e-103	1.2472
quality								
	0.0369	0.8709	2.1261	246.4190	3.9e-53	178.6510	1.6e-39	0.7743
price								
	-1.0031	-1.0031	-1.0031	495.6221	8.5e-110	0.0000	.	-1.0031

Model for Classes					
Intercept	Class1	Class2	Class3	Wald	p-value
	0.2742	-0.1472	-0.1270	3.8140	0.15
Covariates	Class1	Class2	Class3	Wald	p-value
sex					
Male	-0.5423	0.6943	-0.1520	25.9025	2.4e-6
Female	0.5423	-0.6943	0.1520		
age					
16-24	0.8341	-0.5994	-0.2347	53.4461	6.9e-11
25-39	-0.3087	0.5814	-0.2727		
40+	-0.5254	0.0180	0.5074		

First, note that the beta parameter estimates for the 3-class model with covariates (see Figure 7-69) are similar to those in the original 3-class model (see Figure 7-61).

The gamma parameters of the model for the latent distribution appear at the bottom of the Parameters output in Figure 7-69 under the heading 'Model for Classes'. The *p*-values associated with the Wald statistic shows that overall, both effects are significant. The betas associated with *SEX = Female* (0.5423, -0.6943, 0.1520) suggest that females are more likely than males of belonging to the "Fashion-Oriented Segment (segment 1), and much less likely to belong to segment 2. The *AGE* effects show that the youngest age group is more likely than other respondents to be in the "Fashion-Oriented Segment" while the oldest age group is more likely to be in the "Quality-Oriented Segment".

Classification Output

To obtain the Classification output, you need to specify it as an option in the Output Tab before estimating your model.

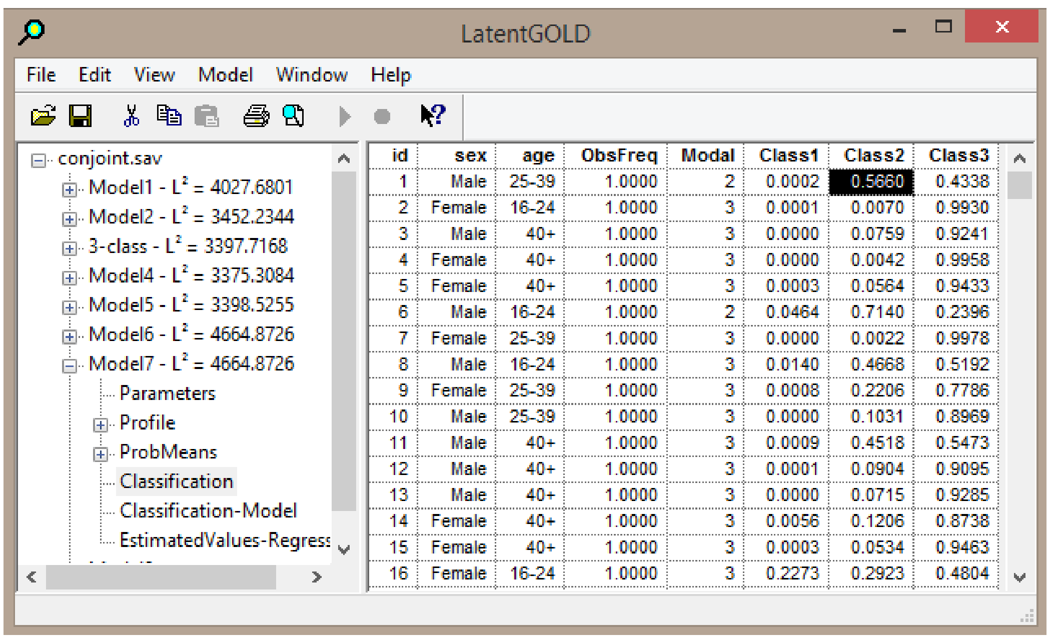
To view the standard classification output for Model 6,

- ➢ Double click on Model6 in the Outline pane to open the Analysis dialog box for this model.
- ➢ Click on the Output Tab.
- ➢ Click in the checkbox next to 'Classification - Posterior' and 'Classification - Model' to select this output.
- ➢ Click Estimate to re-estimate the model.

The new model estimated has been added to the bottom of the list as Model 7 (it is the same as Model6 except for the additional output file selected).

- ➢ In the Outline pane, for Model7, click Classification.

Figure 7-70: Classification Output for Model 7 (partial listing)



id	sex	age	ObsFreq	Modal	Class1	Class2	Class3
1	Male	25-39	1.0000	2	0.0002	0.5660	0.4338
2	Female	16-24	1.0000	3	0.0001	0.0070	0.9930
3	Male	40+	1.0000	3	0.0000	0.0759	0.9241
4	Female	40+	1.0000	3	0.0000	0.0042	0.9958
5	Female	40+	1.0000	3	0.0003	0.0564	0.9433
6	Male	16-24	1.0000	2	0.0464	0.7140	0.2396
7	Female	25-39	1.0000	3	0.0000	0.0022	0.9978
8	Male	16-24	1.0000	3	0.0140	0.4668	0.5192
9	Female	25-39	1.0000	3	0.0008	0.2206	0.7786
10	Male	25-39	1.0000	3	0.0000	0.1031	0.8969
11	Male	40+	1.0000	3	0.0009	0.4518	0.5473
12	Male	40+	1.0000	3	0.0001	0.0904	0.9095
13	Male	40+	1.0000	3	0.0000	0.0715	0.9285
14	Female	40+	1.0000	3	0.0056	0.1206	0.8738
15	Female	40+	1.0000	3	0.0003	0.0534	0.9463
16	Female	16-24	1.0000	3	0.2273	0.2923	0.4804

We can see that the first respondent *(ID=1)* would be classified into segment 2, since segment 2 has the highest membership probability (.5660) for that respondent (see Figure 7-70).

(For information on how to append classification scores to you original data file, see Step 9 in Chapter 5.)

Now, suppose that you wish to classify new data into the appropriate classes but you only had covariate information on these cases. You would use the Classification - Model (Covariate Classification) output for this purpose.

To view this output, click on Classification - Model.

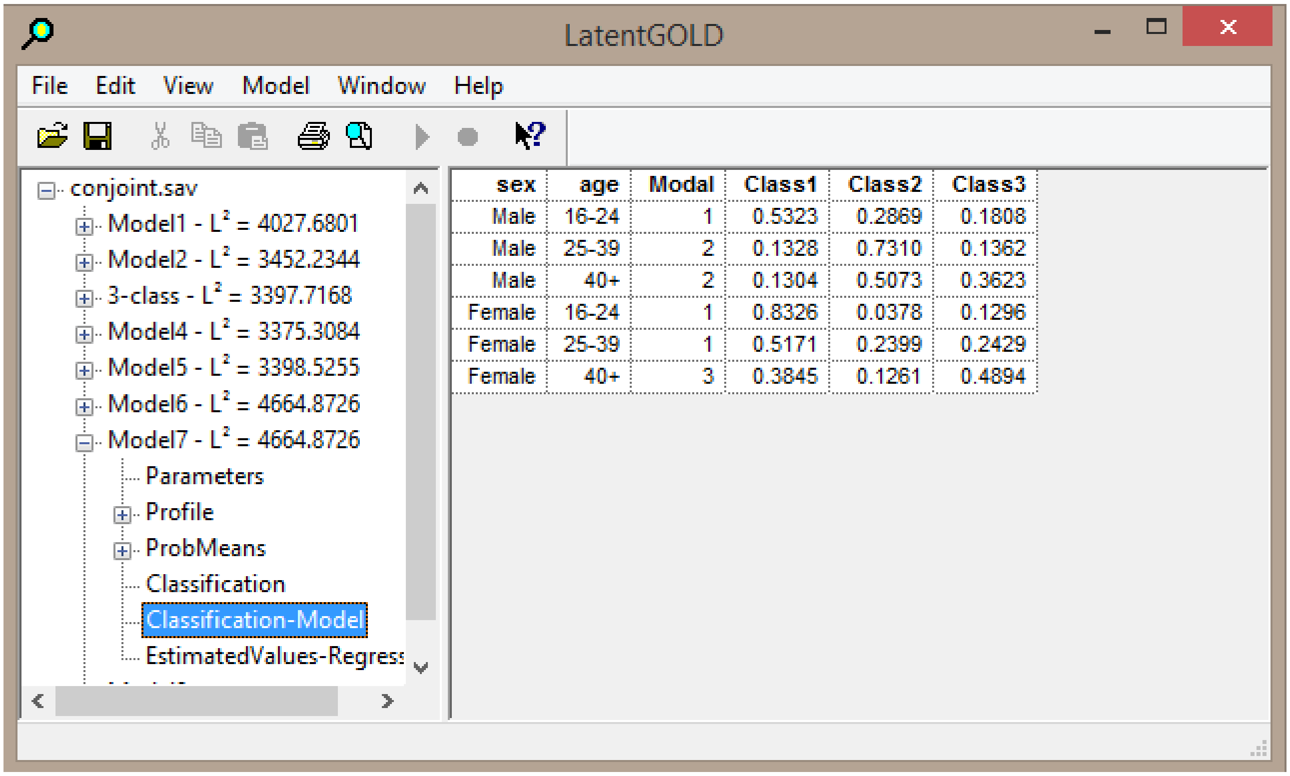


Figure 7-71. Covariate Classification Output for Model7