

# Tutorial #6A: Comparing Segments obtained from LC Cluster and DFactor Models in a Consumer Preference Study

DemoData = 'crackers0.sav'

This tutorial is based on a consumer preference study commissioned by the Kellogg Company. For more information, see

Applications of latent class models to food product development: a case study

Popper, Richard, Kröll, Jeff and Magidson, Jay. Sawtooth Software Proceedings, 2004.

## The Data

In this case study, consumers (N=157) rated their liking of 15 crackers on a nine-point liking scale that ranged from “Dislike Extremely” to “Like Extremely.” Consumers tasted the crackers over the course of three sessions, conducted on separate days. The serving order of the crackers was balanced to account for the effects of day, serving position, and carry-over.

	ID	R#117	R#138	R#231	R#342	R#376	R#410	R#495	R#548	R#603	R#682	R#755	R#812	R#821	R#951	R#967	AvgRtg
1	1101	6	7	6	6	6	8	9	9	7	8	6	9	9	8	8	7.47
2	1102	8	7	6	6	9	7	9	9	4	9	3	6	7	9	7	7.07
3	1103	8	3	5	6	7	6	3	9	7	8	5	8	2	7	2	5.73
4	1104	4	2	3	2	8	6	7	5	2	7	4	7	6	7	6	5.07
5	1105	2	2	8	2	7	4	9	8	5	5	3	9	7	7	7	5.67
6	1106	3	7	2	2	3	6	6	7	8	8	1	7	4	6	6	5.07
7	1107	1	1	1	2	5	9	1	8	5	9	1	9	8	9	5	4.93
8	1108	2	2	2	7	9	9	9	6	8	7	8	7	9	6	8	6.60
9	1109	8	8	7	3	8	8	9	8	7	9	7	9	8	9	9	7.80
10	1110	6	4	4	2	8	7	9	8	7	8	5	7	8	8	5	6.40

Figure 1: Data Layout for the LC Cluster and LC Factor Models

The variable ID uniquely identifies each case. The “R#...” variables contain ratings for each of the 15 cracker products. The variable ‘AvgRtg’ contains the average rating for each case across all 15 cracker products.

## The Goal

The goal of this research was to determine if consumers can be segmented in a meaningful way on the basis of their liking ratings of the crackers. In this tutorial 6A we will estimate and compare a number of LC Cluster and DFactor models and interpret the resulting segments.

In Tutorial 6B we will use Regression Models to obtain segments.

In this tutorial, you will:

- Estimate LC Cluster and DFactor Models
- Examine various output including the Loadings output from DFactor Models

- Use the DFactor module to obtain 'clusters' after 'factoring out' a nuisance factor
- > Use the 'Equal Effects' option to obtain a general factor

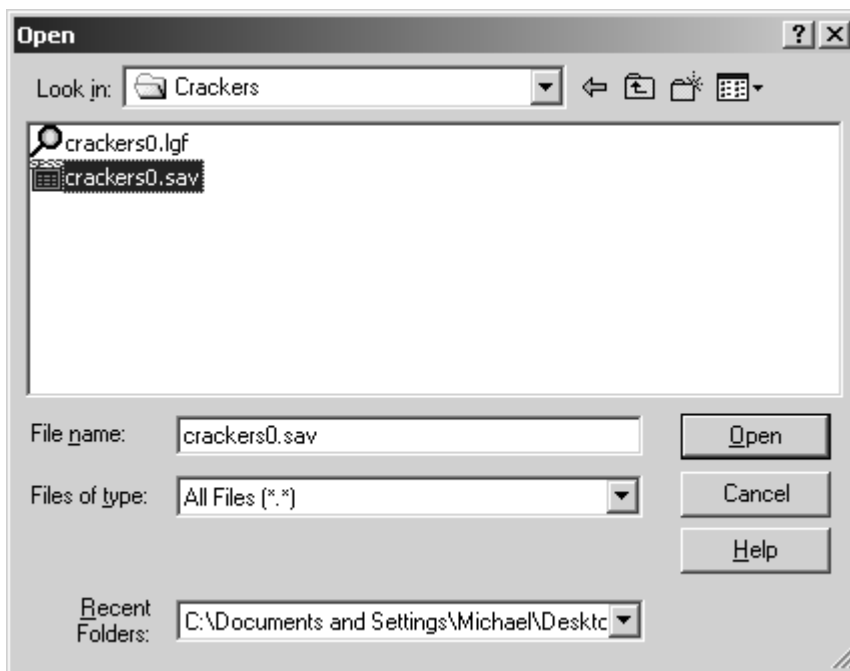
### **Opening the Data File**

For this example, the data file is in SPSS system file format.

- To open the file, from the menus choose:

File  
Open

- Open crackers0.sav



**Figure 2. Opening crackers0.sav**

The filename crackers0.sav appears in the left-hand pane of Latent GOLD (the Outline pane).

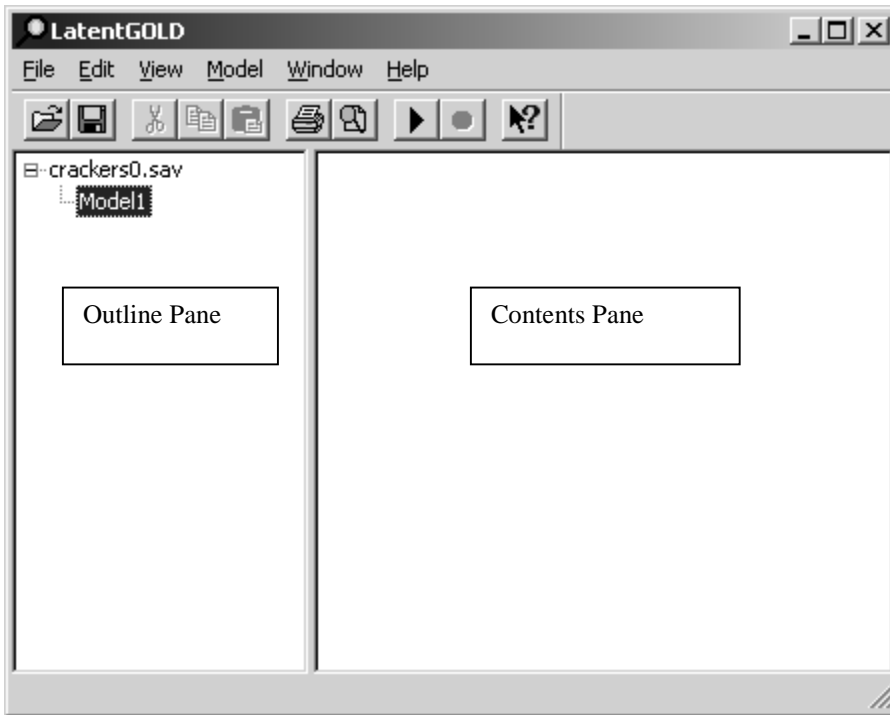
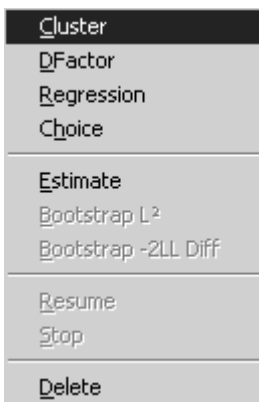


Figure 3. Data file in the Outline Pane

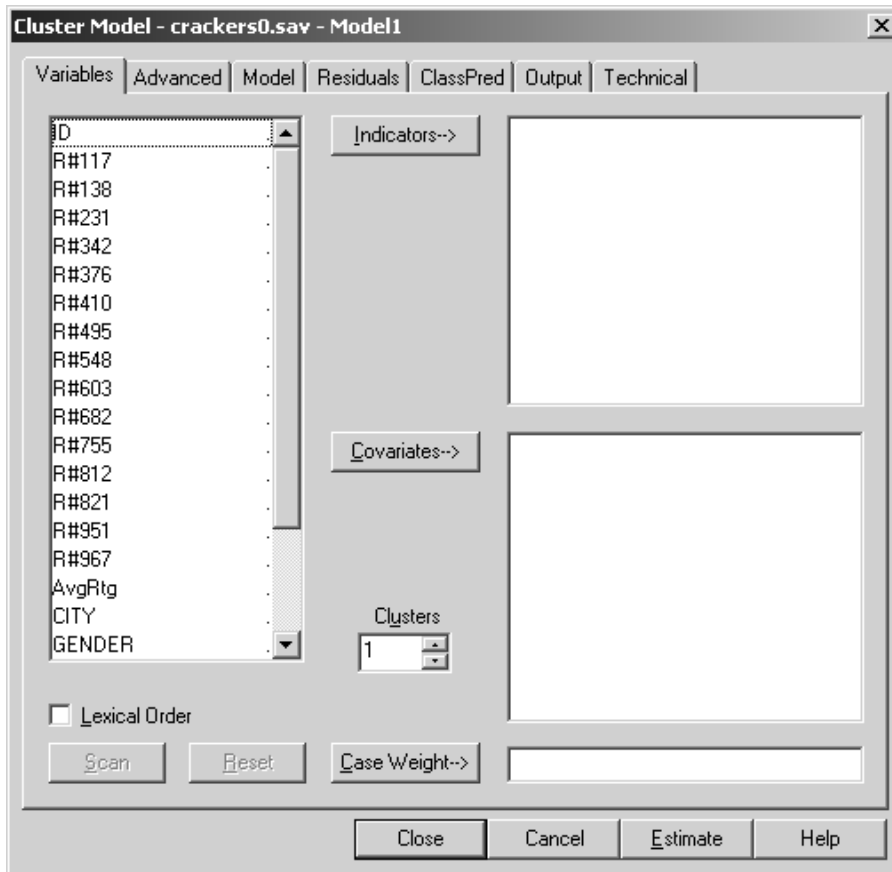
### ***Estimating a LC Cluster Model***

We will begin by estimating a LC Cluster Model. This is the traditional latent class model, which imposes no special structure to distinguish between variation due to differences in overall response level and those due to relative differences in the liking of crackers. That is, the latent classes simply represent unordered levels (i.e., a nominal factor).

- Right-click on Model1 and select Cluster from the menu:



The Analysis Dialog Box for the Cluster Model opens:



**Figure 4. Cluster Model Analysis Dialog Box**

- Select all 15 ratings variables
- Click on “Indicators” to move these variables to the Indicator Box

We will begin by estimating 1 to 3 Clusters. Thus,

- In the Clusters Box, type in “1-3”

Your final model analysis dialog box should look like this:

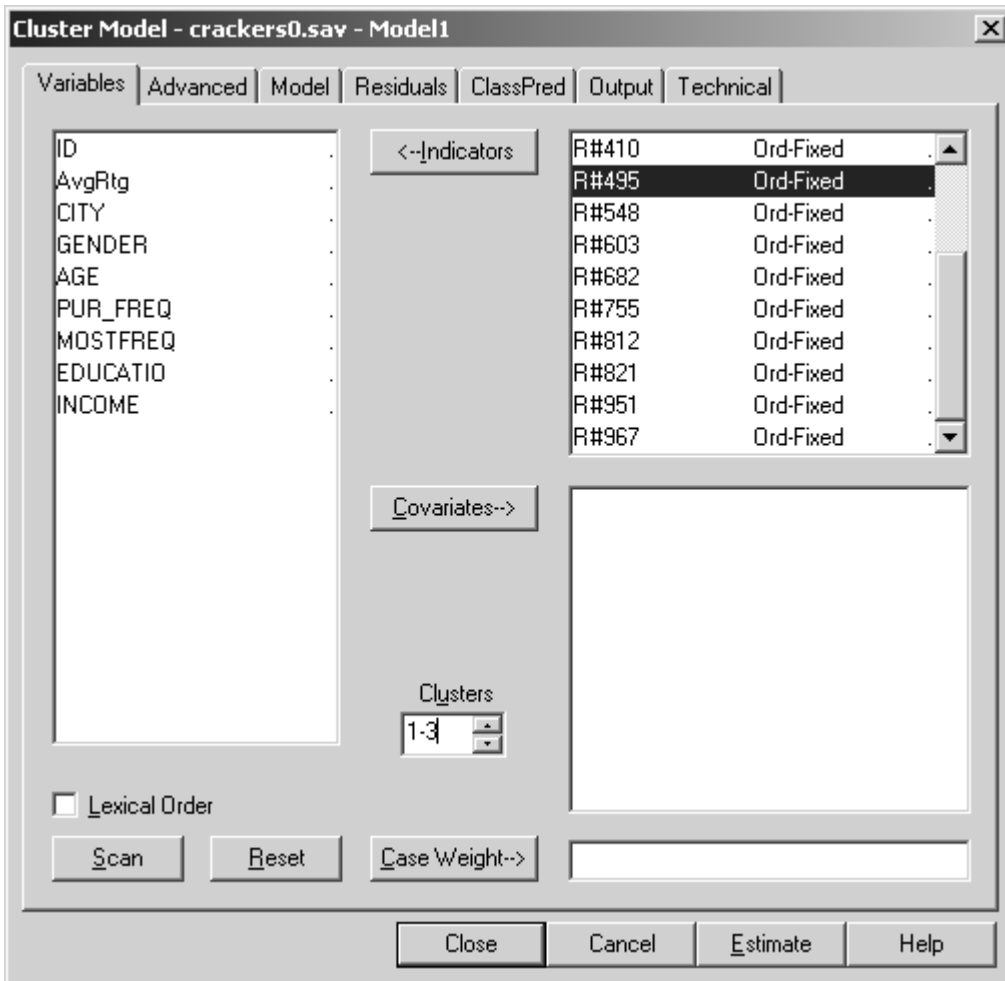


Figure 5. Cluster Model analysis dialog box prior to estimation

- Click Estimate

After the model has estimated, click on the data file name “crackers0.sav” in the Outline (left-hand) Pane. Observing the output, we can see that the 2-cluster model has the lowest BIC value, and, thus, is the best fit:

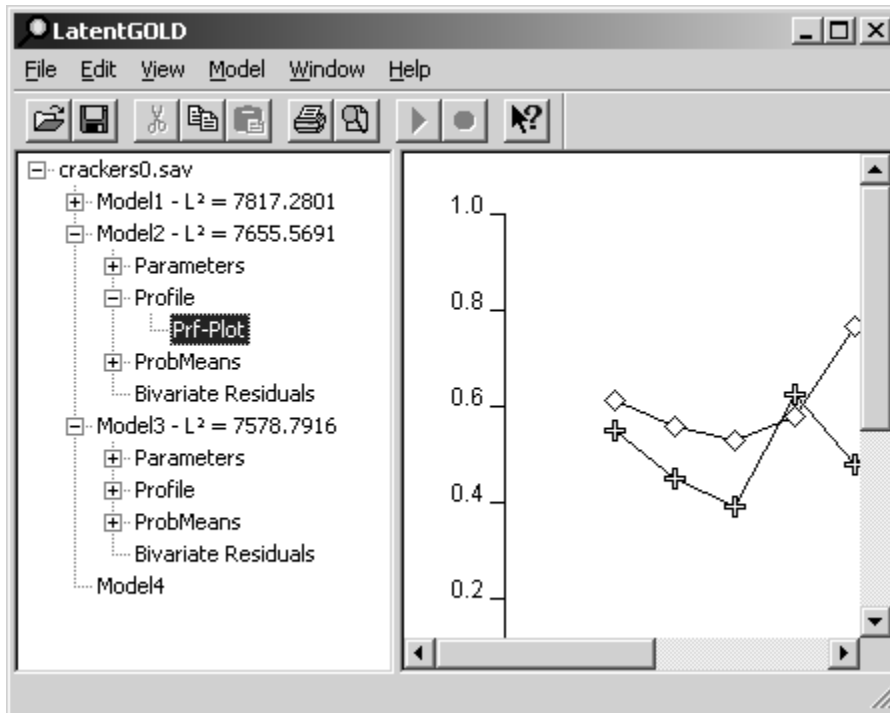
			LL	BIC(LL)	Npar
Model1	1-Cluster		-4702.4706	10006.6345	119
Model2	2-Cluster		-4621.6151	<b>9925.8235</b>	135
Model3	3-Cluster		-4583.2264	9929.9459	151

Figure 6. 2-Cluster Model showing the best fit

According to the BIC, a two-cluster solution was a better fit to the data than either a one-cluster or three-cluster solution. The two clusters (segments) were approximately equal in size (53% and 47%).

Now, in the Outline Pane,

- Click on the '+' symbol to the left of Model2 to expand it.
- Click on the '+' next to Profile and click on Prf-Plot:



**Figure 7. Prf-Plot for the 2-Cluster Model.**

The Profile Plot is now in the Contents Pane. By default, it is showing only the first 8 variables, but we can set it to show all 15. To do that:

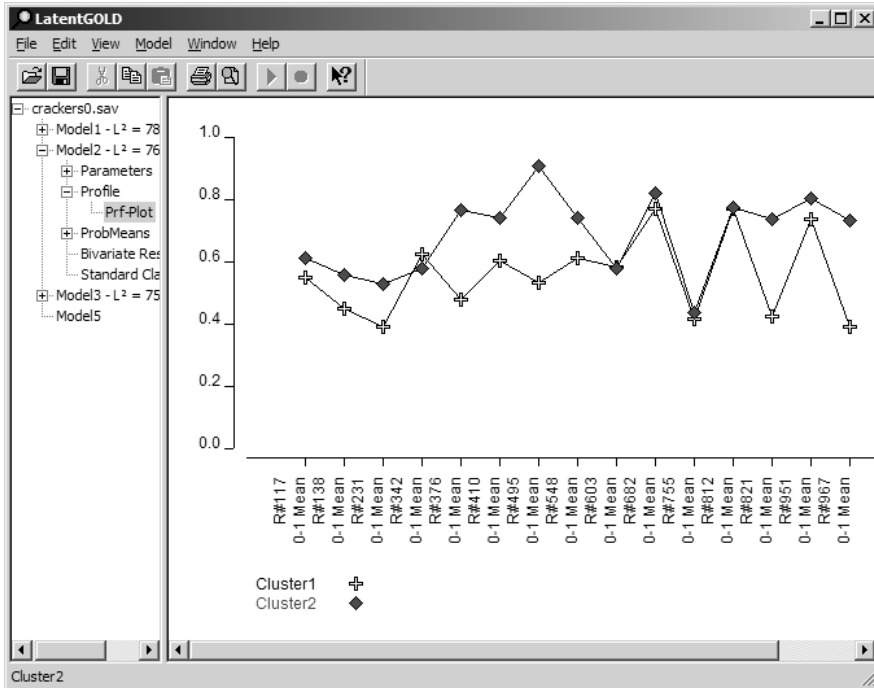
- Right-click on the plot to bring up the Profile Plot Control Menu.
- In the Variables Tab, scroll down and check all of the variables that are not checked:



**Figure 8. Profile Plot Control**

The Profile Plot will now show all the variables.

- At the bottom of the Profile Plot, click on the  $\diamond$  symbol next to Cluster 2 to highlight Cluster 2 on the Plot:



**Figure 9. Profile Plot with Cluster 2 highlighted**

The Profile Plot illustrates that Cluster 2 tends to rate almost all the crackers higher than Cluster 1.

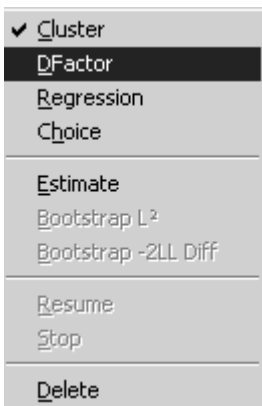
Figure 8 shows that the two segments are clearly and almost exclusively differentiated by their overall average liking of the crackers. This result is not unexpected, since no attempt was made in the analysis to adjust the data for differences in response level..

Note: For comparability for variables that have different range of values, the Profile plot transforms the values to between 0 and 1 (labeled '0-1 Mean'). If you click on a particular symbol, both the 0-1 mean and the original means are displayed in the status bar.

### ***Estimating a DFactor Model***

Let us now apply a factor-based version of the latent class model in order to try to “factor out” response level effects.

- In the Outline Pane, right-click on Model4 and select DFactor from the menu



The DFactor Model Analysis Dialog Box appears.

- Change the number of Factors from 1 to 2 by typing 2 in the Factors box

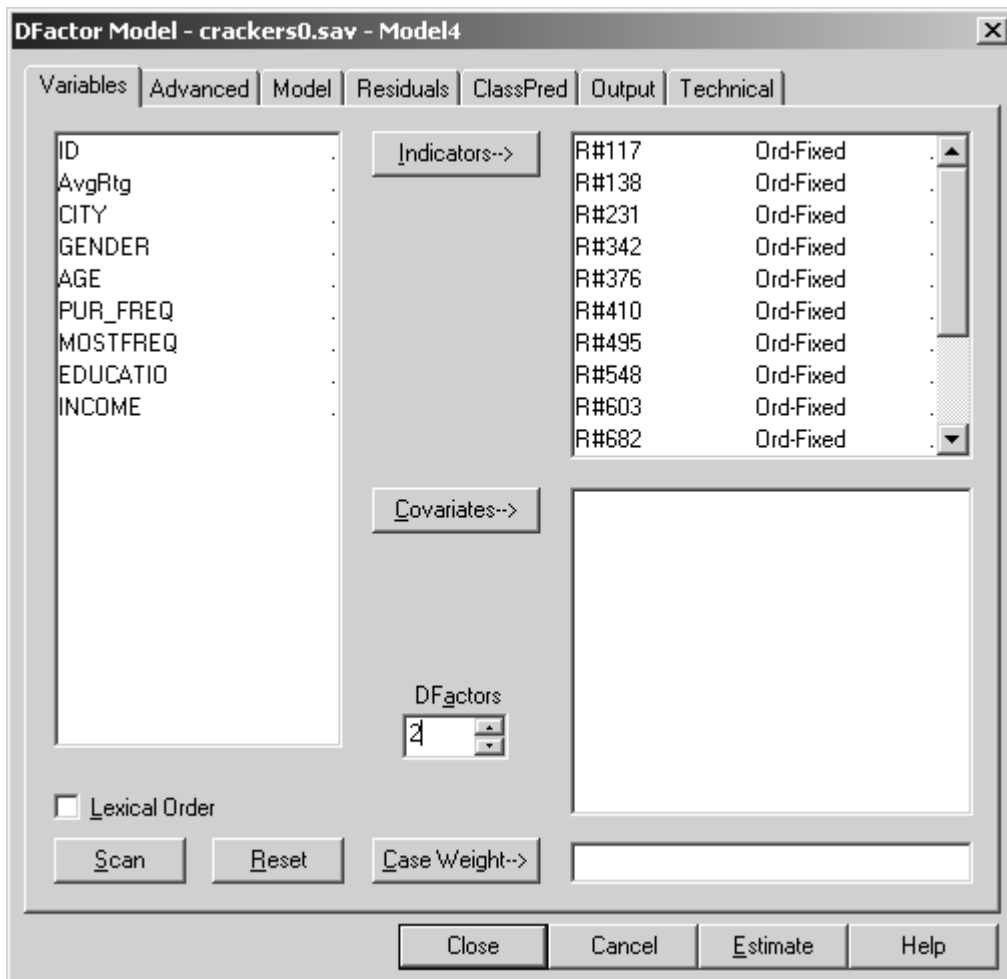


Figure 10. DFactor Model Analysis Dialog Box prior to estimation

- Click Estimate
- Click on crackers0.sav to go to the new model summary

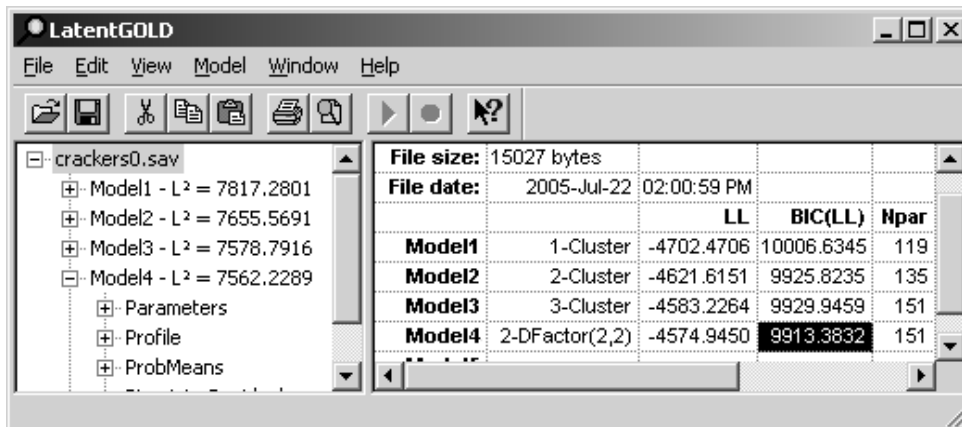


Figure 11. Model Summary Output with 2 –DFactor Model

The 2 DFactor model appears in the model summary as '2-DFactor(2,2)' to denote that each DFactor has 2 levels, the default specification for DFactor models. Note that it has the same number of parameters as a 3-class model, but is preferred over both the 3 and 2-class (Cluster) models according to the BIC.

We want to interpret one of the DFactors (say DFactor1) as a general factor which measures the overall 'rating level' effect. Persons at one end of the scale might tend to like crackers better (or otherwise rate crackers higher) than those at the other end of the scale, regardless of the particular type of cracker. We will then compare the 2 levels associated with DFactor2 with the 2 clusters we obtained in the 2-class Cluster model.

To increase the number of levels for DFactor1,

- Double-click on Model4 in the Outline Pane.

The Model Analysis Dialog Box appears.

- Click on Model to go to the Model Tab
- In the Model tab, click on one of the levels ('Level1' or 'Level2') and click the Add button. The label 'Level3' appears. Click the Add again and 'Level4' appears.

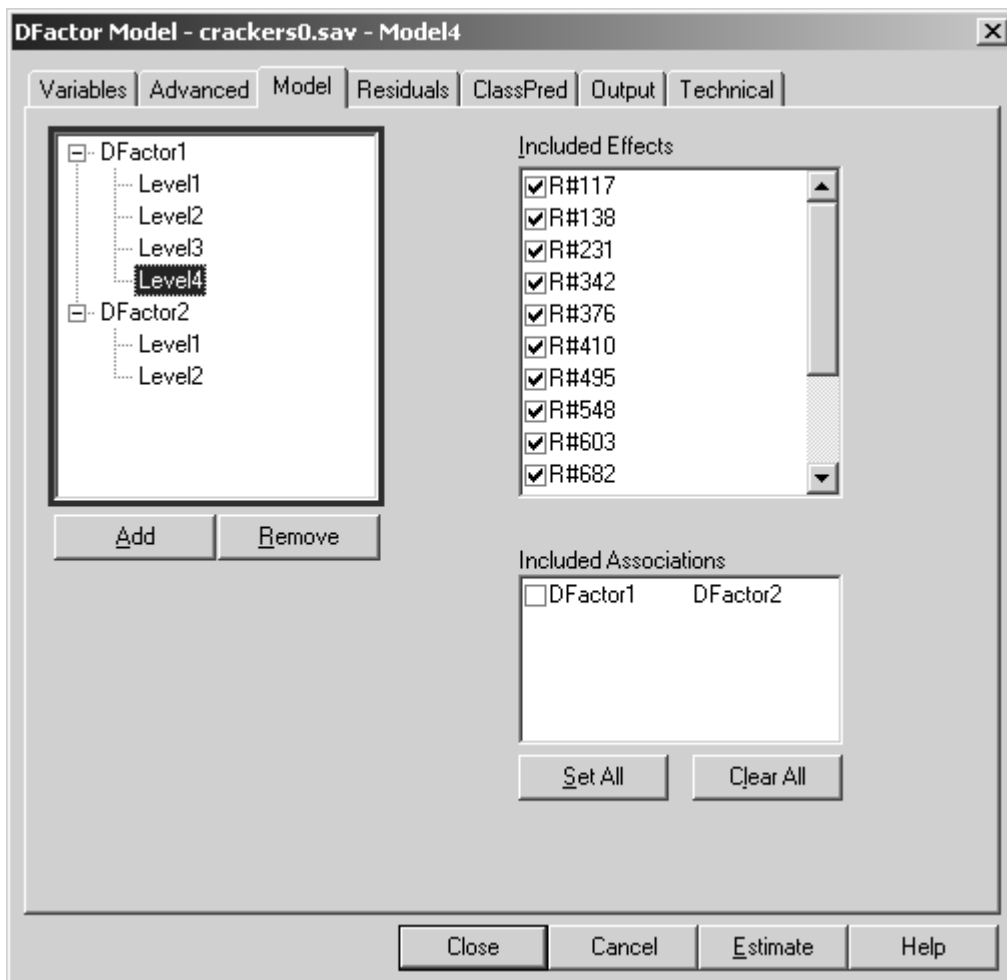


Figure 12. Adding levels to DFactor1

- Click Estimate to estimate the 2-DFactor(4,2) model.
- Click on crackers0.sav to view the new model summary.

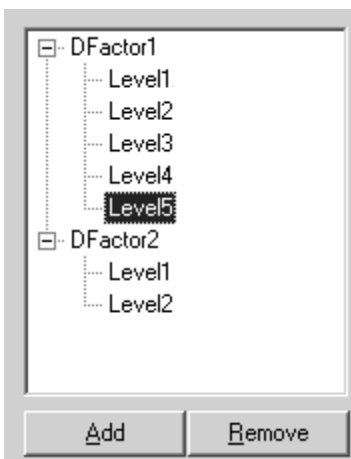
The BIC suggests that increasing the number of levels for DFactor1 to 4 provides an improvement:

Model	Structure	LL	BIC(LL)	Npar
Model1	1-Cluster	-4702.4706	10006.6345	119
Model2	2-Cluster	-4621.6151	9925.8235	135
Model3	3-Cluster	-4583.2264	9929.9459	151
Model4	2-DFactor(2,2)	-4574.9450	9913.3832	151
Model5	2-DFactor(4,2)	-4556.8514	9887.3085	153

**Figure 13. Model Summary Display with DFactor (4,2) Model**

Now, let's try various other level combinations:

- In the Outline Pane, Double-click on Model6
- Click on the Model Tab
- Add another level to DFactor1



Thus, this Model is now DFactor (5,2)

- Click Estimate
- Double-click on Model7
- Click on the Model Tab

- Add another level to DFactor1 (Level 6)
- Click Estimate again

Going back to the Model Summary Display, you can see that the (5,2) combination is best according to the BIC:

			LL	BIC(LL)	Npar
Model1	1-Cluster		-4702.4706	10006.6345	119
Model2	2-Cluster		-4621.6151	9925.8235	135
Model3	3-Cluster		-4583.2264	9929.9459	151
Model4	2-DFactor(2,2)		-4574.9450	9913.3832	151
Model5	2-DFactor(4,2)		-4556.8514	9887.3085	153
Model6	2-DFactor(5,2)		-4552.7475	9884.1568	154
Model7	2-DFactor(6,2)		-4557.0383	9897.7948	155

Figure 14. Model Summary Display – 2DFactor (5,2) is the best fit

Now, rename the 2-DFactor(5,2) model:

- In the Outline Pane, Click on Model6 to select it and click on the name again
- A cursor appears. Type in 2Dfac(5,2)

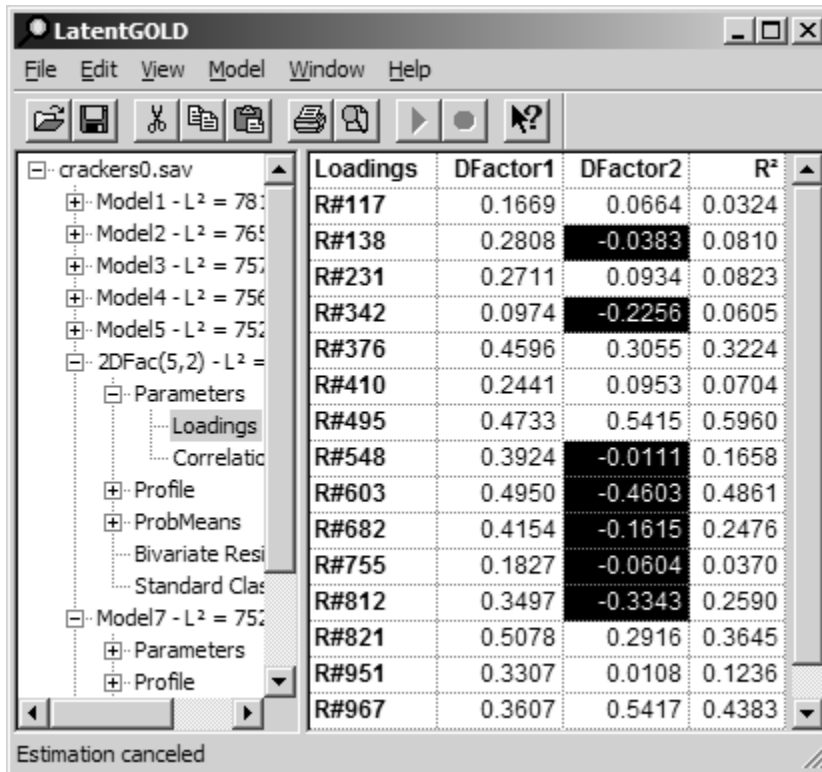
Model	L <sup>2</sup>
Model1	7817.2801
Model2	7655.5691
Model3	7578.7916
Model4	7562.2289
Model5	7526.0417
2DFac (5,2)	
Model7	7517.7010
Parameters	
Profile	
ProbMeans	
Bivariate Residuals	
Model8	

Figure 15. Changing the Model name to 2-Dfac (5,2)

## Loadings Output

For this model, let's view the Loadings Output to view the factor loadings

- Click on the '+' symbol to the left of 2Dfac(5,2) to expand it.
- Click on the '+' next to Parameters and click on Loadings:

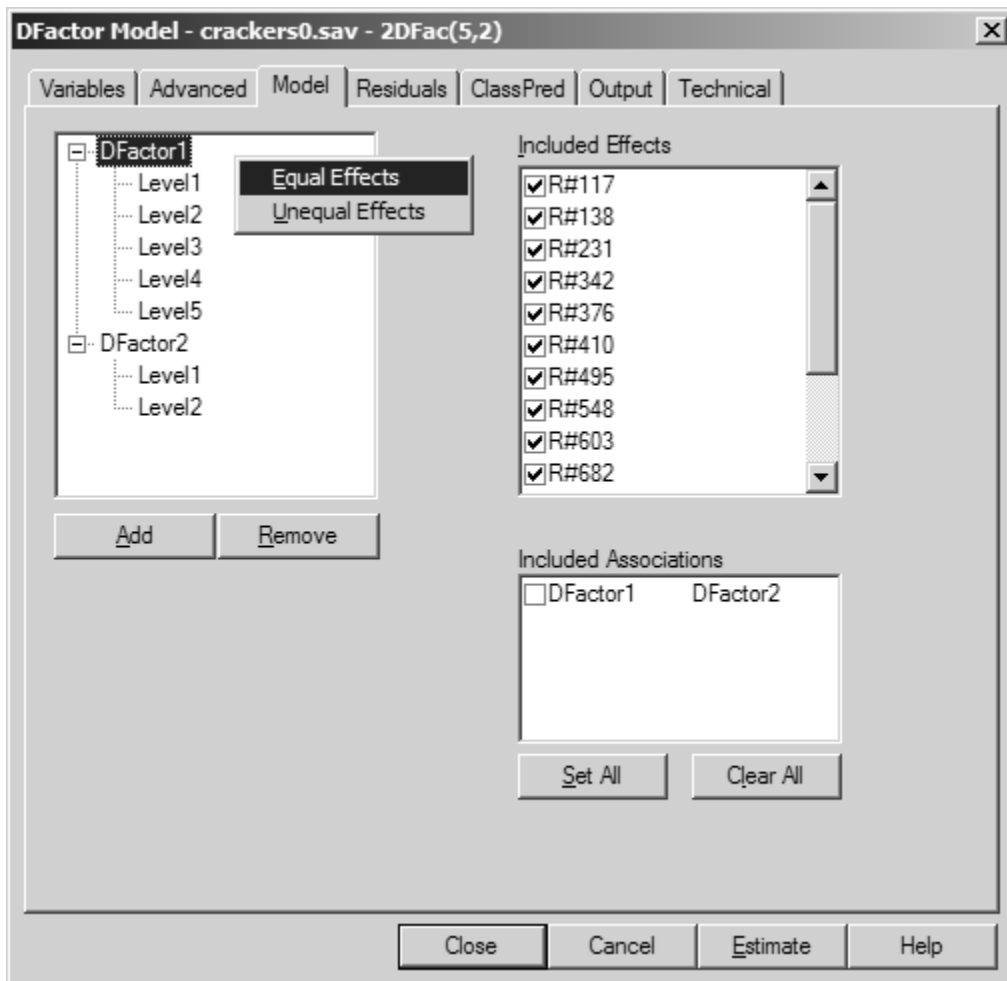


Loadings	DFactor1	DFactor2	R <sup>2</sup>
R#117	0.1669	0.0664	0.0324
R#138	0.2808	-0.0383	0.0810
R#231	0.2711	0.0934	0.0823
R#342	0.0974	-0.2256	0.0605
R#376	0.4596	0.3055	0.3224
R#410	0.2441	0.0953	0.0704
R#495	0.4733	0.5415	0.5960
R#548	0.3924	-0.0111	0.1658
R#603	0.4950	-0.4603	0.4861
R#682	0.4154	-0.1615	0.2476
R#755	0.1827	-0.0604	0.0370
R#812	0.3497	-0.3343	0.2590
R#821	0.5078	0.2916	0.3645
R#951	0.3307	0.0108	0.1236
R#967	0.3607	0.5417	0.4383

Figure 16. Loadings Output for the 2Dfac (5,2) Model

Notice that all the loadings on DFactor1 are positive while DFactor2 contains both positive as well as negative loadings. To make it especially clear that DFactor1 is a 'general' factor:

- In the Outline Pane, double click on model name 'Dfac(5,2)' and click on the Model tab
- Right click on 'DFactor1' to retrieve the pop-up menu
- Select 'Equal effects'



**Figure 17. Selecting Equal Effects for DFactor1**

The symbol '<E>' now appears next to DFactor1

- Click Estimate.

By setting all 15 of the (log-linear) effects equal, we reduce the number of parameters by 14. The resulting model is preferred according to the BIC:

Model	Factor	LL	BIC(LL)	Npar
Model1	1-Cluster	-4702.4706	10006.6345	119
Model2	2-Cluster	-4621.6151	9925.8235	135
Model3	3-Cluster	-4583.2264	9929.9459	151
Model4	2-DFactor(2,2)	-4574.9450	9913.3832	151
Model5	2-DFactor(4,2)	-4556.8514	9887.3085	153
2DFac(5,2)	2-DFactor(5,2)	-4552.7475	9884.1568	154
Model7	2-DFactor(6,2)	-4557.0383	9897.7948	155
Model8	2-DFactor(5,2)	-4576.1962	9860.2668	140

Figure 18. New Model Summary Display

Notice that the (linear) loadings for DFactor 1 are closer in value to reflect the equality of the log-linear counterparts. The fact that the loadings are now all negative is due to the arbitrary ordering of the levels (i.e., one can always multiply all the loadings by minus 1 and reverse the order of the 5 levels so that level 5 becomes the new level 1, level 4 becomes new level 2, etc. without changing the model).

- Click on the '+' symbol to the left of Model8 to expand it.
- Click on the '+' next to Parameters and click on Loadings:

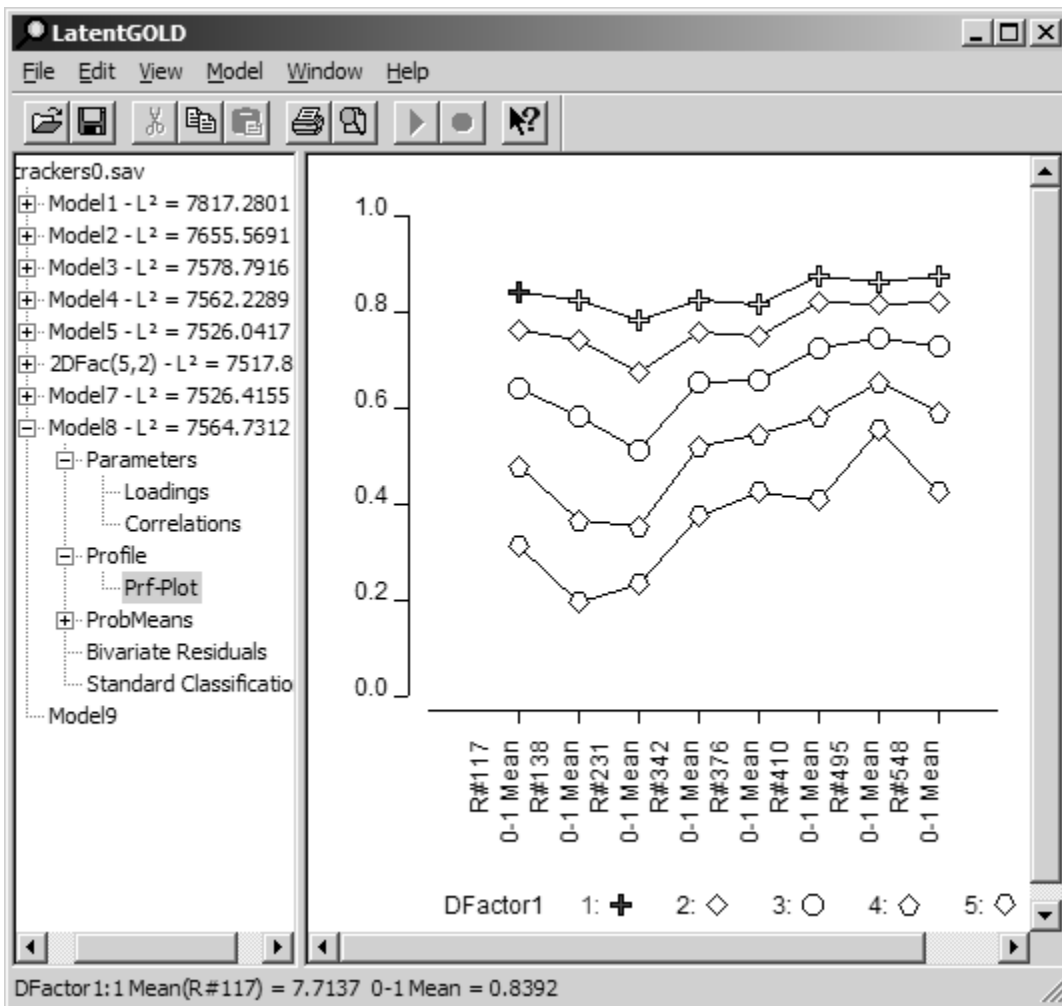
Loadings	DFactor1	DFactor2	R <sup>2</sup>
R#117	-0.3726	0.0029	0.1432
R#138	-0.4081	-0.0458	0.1752
R#231	-0.3805	0.1356	0.1645
R#342	-0.3311	-0.2865	0.1968
R#376	-0.2891	0.4457	0.2888
R#410	-0.3464	0.0796	0.1336
R#495	-0.2250	0.6198	0.4541
R#548	-0.3389	0.1316	0.1393
R#603	-0.3533	-0.1772	0.1615
R#682	-0.2584	0.0211	0.0755
R#755	-0.3668	-0.0935	0.1438
R#812	-0.2670	-0.1222	0.0921
R#821	-0.3312	0.4012	0.2767
R#951	-0.2939	0.1242	0.1122
R#967	-0.2383	0.6694	0.5130

Figure 19. Loadings for Model8

Under this model, level 1 of DFactor1 contains those persons with the highest ratings. To see this, we can view the Profile plot for DFactor1.

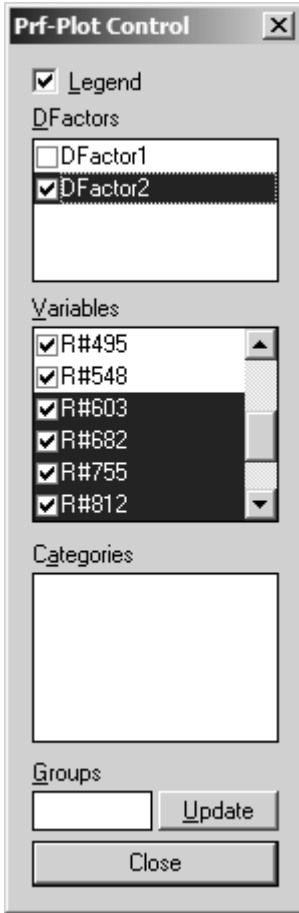
- Click on the '+' next to Profile and click on Prf-Plot
- The Profile Plot appears

Click the symbol for level 1 associated with Cracker product R#117, and compare the average rating with that obtained for level 5. The status bar shows that the average rating on this cracker for persons in level 1 is 7.7 (a 0-1 mean of .84). This compares to an average rating of 3.5 for those in level 5.



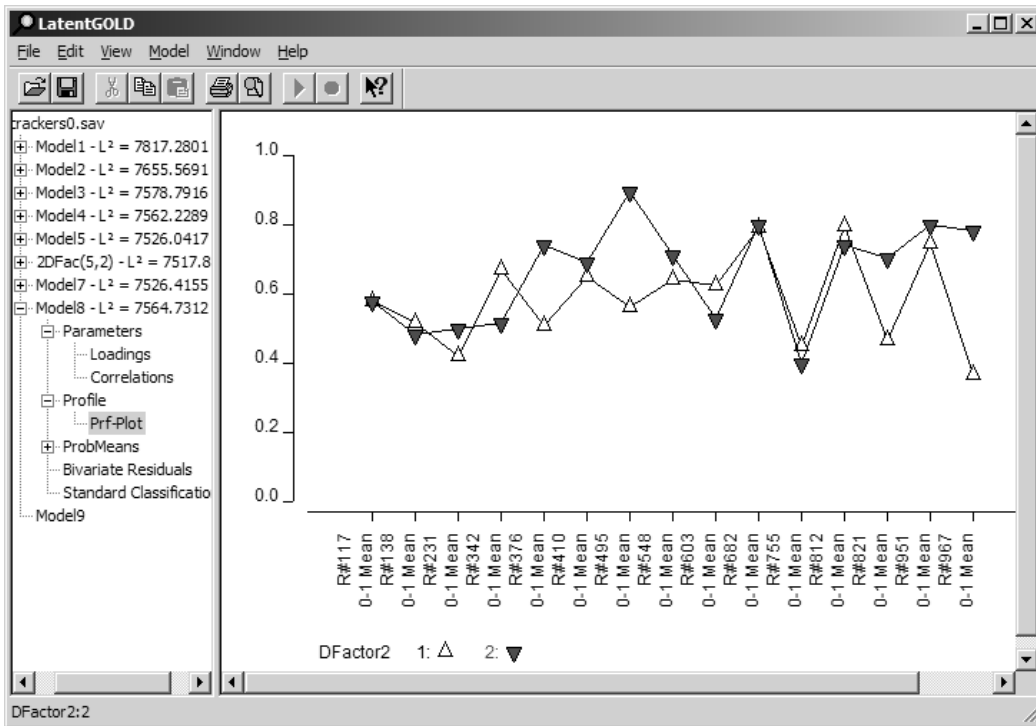
**Figure 20. Profile Plot for Model8**

- Right click on the plot to retrieve the Profile Plot Control.
- In the Dfactors box, uncheck DFactor1 and check DFactor2 to change the setting to display DFactor2 instead of DFactor1.
- In the Variables box, scroll down and check all the variables to display all the crackers on the plot.



**Figure 21. Profile Plot**

- Click on the ▾ symbol for level 2 of DFactor2 to highlight the average ratings for this level.



**Figure 22. Profile Plot for DFactor2**

Notice that controlling for the overall level effect (DFactor1), persons in level 2 of DFactor2 provide significantly higher ratings for cracker products #376, #495, #821 and #967 and significantly lower ratings for #342, and #603.

To see that these differences are statistically significant, examine the log-linear parameters (and associated p-values) associated with DFactor2 in the Parameters output. We can also confirm that the parameter estimates associated with DFactor1 are equal for all 15 rating variables (the estimate is  $-1.1047$ ).

- Click on 'Parameters' to view the Parameters output:

	DFactor1	Wald	p-value	DFactor2	Wald	p-value	R <sup>2</sup>
R#117	-1.1047	77.8304	1.1e-18	0.0030	0.0011	0.97	0.1432
R#138	-1.1047	77.8304	1.1e-18	-0.0425	0.2736	0.60	0.1752
R#231	-1.1047	77.8304	1.1e-18	0.1398	2.3854	0.12	0.1645
R#342	-1.1047	77.8304	1.1e-18	-0.3239	8.6774	0.0032	0.1968
R#376	-1.1047	77.8304	1.1e-18	0.5956	16.0722	6.1e-5	0.2888
R#410	-1.1047	77.8304	1.1e-18	0.0876	0.7571	0.38	0.1336
R#495	-1.1047	77.8304	1.1e-18	1.3411	18.4496	1.7e-5	0.4541
R#548	-1.1047	77.8304	1.1e-18	0.1485	2.2128	0.14	0.1393
R#603	-1.1047	77.8304	1.1e-18	-0.1896	3.8132	0.051	0.1615
R#682	-1.1047	77.8304	1.1e-18	0.0311	0.0598	0.81	0.0755
R#755	-1.1047	77.8304	1.1e-18	-0.1040	1.1109	0.29	0.1438
R#812	-1.1047	77.8304	1.1e-18	-0.1726	1.7617	0.18	0.0921
R#821	-1.1047	77.8304	1.1e-18	0.4564	16.1442	5.9e-5	0.2767
R#951	-1.1047	77.8304	1.1e-18	0.1625	1.9517	0.16	0.1122
R#967	-1.1047	77.8304	1.1e-18	1.0514	18.9698	1.3e-5	0.5130

Figure 23. Parameters Output for Model8

### Comparison of LC Cluster and DFactor Models

The most important difference between the two models is that the LC Cluster model confounded relative differences in liking with average response level (one segment rated almost all products higher than the other). The DFactor model, on the other hand, was able to separate out a response level effect (D-Factor # 1) from an effect that reflected the relative differences in liking (D-Factor #2). Also, the DFactor model was preferred over the LC Cluster models according to the BIC (BIC = 9,887 for the DFactor model, compared to 9,926 for the two-class cluster model and 9,930 for the three-class cluster model.)

The same products that differentiated the segments in the LC Cluster model differentiated the segments on D-Factor #2 (#495, 376, 821, and 967). But D-Factor #2 further differentiated the segments in their response to Products #812 and #603, whereas the LC Cluster model showed no difference between these products (see Figure 6). In the LC Cluster solution, these differences were masked by the differences in response level between the clusters. In addition to these differences between the two models, the two models also differed in the relative sizes of the two segments.