7.2. Tutorial #2: Using Latent GOLD to Estimate DFactor Models

DemoData = ‘gss82white.sav’

The Goal

In this tutorial, we re-examine the results obtained from tutorial #1 using discrete factor (DFactor) models instead of LC Cluster models. We show how a 2-DFactor model consisting of 2 dichotomous factors can be viewed as a restricted form of the 4-cluster model and use the L^2 difference statistic to test whether the unrestricted 4-class model provides an improvement. In addition, this tutorial illustrates:

The use of the Ordinal scale type
Estimating DFactor models
Factor Loadings Output
Restricting Factor Loadings to Zero
Joint Profile output
Classification Output
The Bi-plot

For these data the DFactor models provide additional insights into the different survey respondent types.

DFactor Analysis vs. Traditional Factor Analysis

In traditional factor analysis (FA), continuous observed variables are expressed as a linear function of 1 or more continuous latent factors (CFactors). DFactor analysis differ from FA in several respects:

The observed variables may be of mixed scale types including nominal, ordinal, continuous and count.
The latent variables are not continuous but discrete, containing 2 or more ordered categories (levels)
The model is not linear
Solutions need not be rotated to be interpretable (the indeterminacy issue of ‘rotation’ is unique to CFactors in a linear model).

In addition, a cross-tabulation of DFactors defines a set of clusters. For example, 2 dichotomous DFactors V and W, yields 4 latent classes (4 clusters).

Figure 7-26. The 4 latent classes

	W=1	W=2
V=1	X =1	X = 2
V=2	X =3	X = 4

As our starting point, we will re-estimate the 3- and 4-class Cluster models from tutorial #1.

Opening the data file

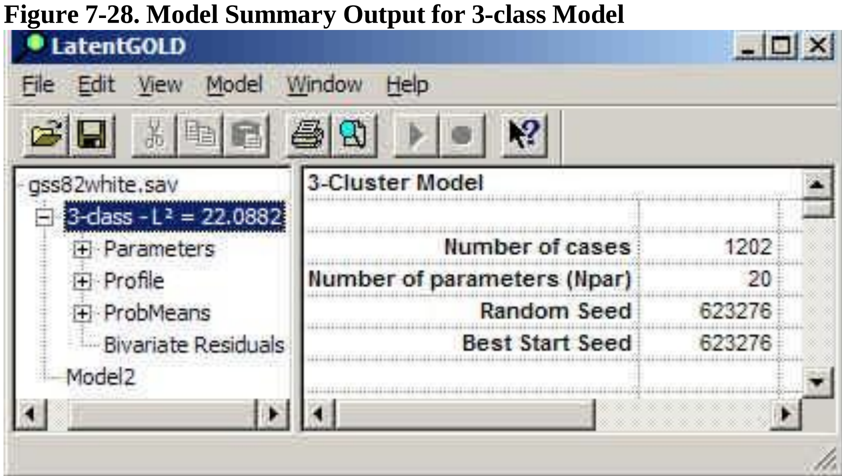
To retrieve the model setup for the 3-class model,

- Select File/Open ‘3-class.lgf’

Figure 7-27. Setup for '3-class' Model

- Double click on "3-class" to open the Analysis Dialog box
- Click Estimate to re-estimate this model

Figure 7-28. Model Summary Output for 3-class Model



To estimate the 4-class model,

- Double click on Model2 to re-open the Analysis Dialog box
- Change '3' to '4' in the Clusters box
- Click Estimate

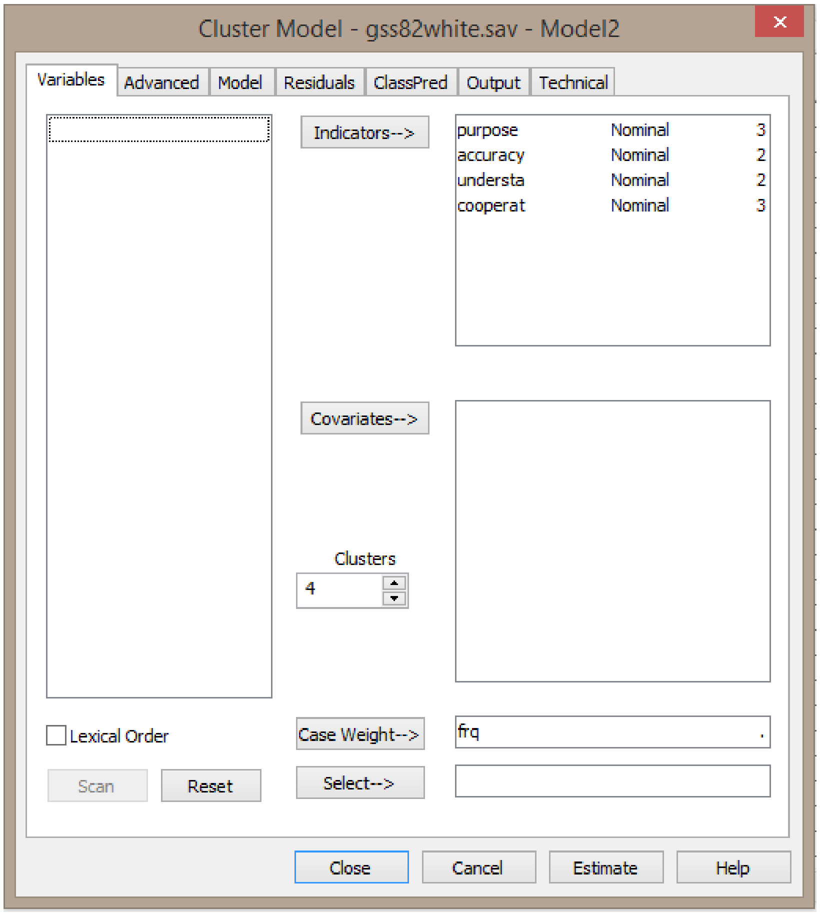
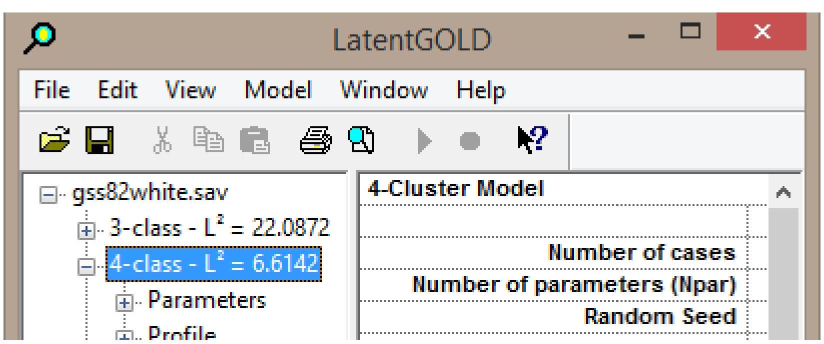


Figure 7-29. Estimating Model2

The 4-classmodel is named ‘Model2’. To change the name to ‘4-class’

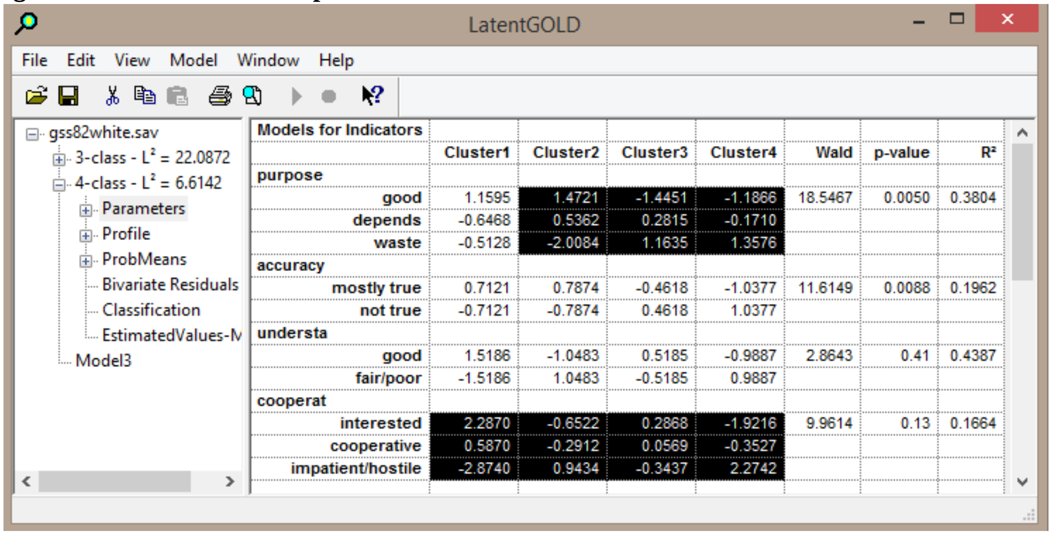
- Click once on Model2 to select it
- Click once again on it to enter Edit mode
- Type ‘4-class’

Figure 7-30. Editing the name of Model2



➢ Click Parameters to view the parameter estimates for the 4-class model

Figure 7-31. Parameters Output for 4-class model.



Notice that for the trichotomous variables PURPOSE and COOPERATE, the estimate for the middle level in each class is approximately midway between the estimates for the end categories (with the single exception of PURPOSE for cluster 1). This suggests that treating these variables as ordinal rather than nominal may be justified, using the default equidistant category scores.

To change the scale type to Ordinal,

- ➢ Double click on Model3 to re-open the Analysis Dialog box
- ➢ Ctrl-click on PURPOSE and COOPERAT to select these variables
- ➢ Right click to retrieve the scale type settings menu
- ➢ Select Ordinal

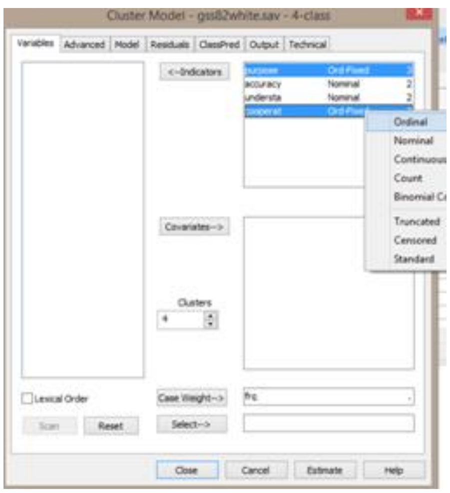


Figure 7-32. Changing PURPOSE and COOPERAT to ordinal variables

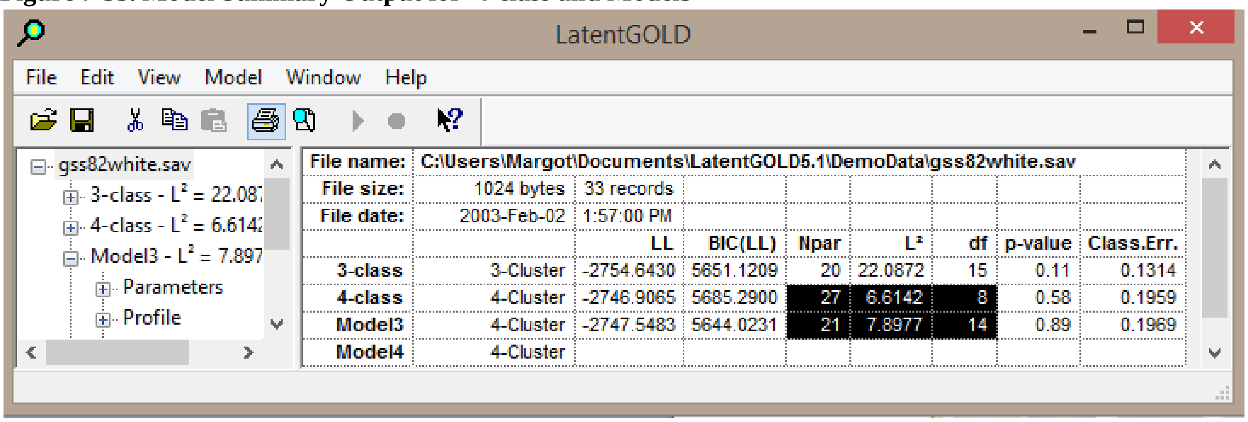
- ➢ Click Estimate to estimate the model

To compare the results from these models

- ➢ Click on the data file name in the Outline Pane

We see that imposing the ordinality restrictions increase L^2 from 6.6 to 7.9, a very small increase associated with the gain of 6 degrees of freedom. Thus, we choose Model3 over the unrestricted 4-class model, which results in
a 4-class model with 6 fewer parameters.

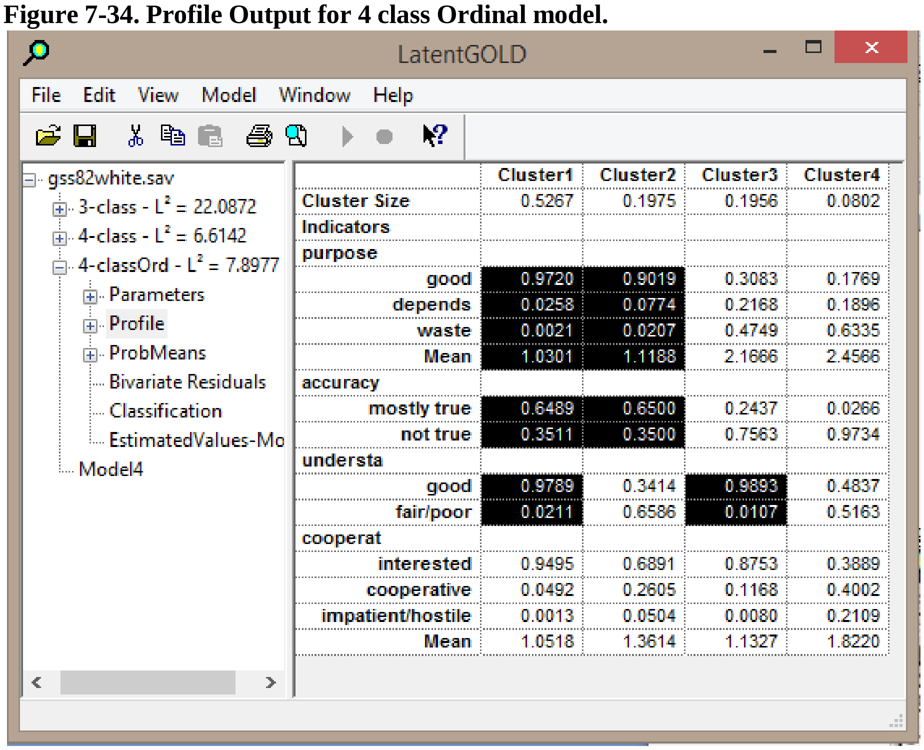
Figure 7-33. Model Summary Output for 4-class and Model3



File name:	C:\Users\Margot\Documents\LatentGOLD5.1\DemoData\gss82white.sav								
File size:	1024 bytes	33 records							
File date:	2003-Feb-02	1:57:00 PM							
		LL	BIC(LL)	Npar	L²	df	p-value	Class.Err.	
3-class	3-Cluster	-2754.6430	5651.1209	20	22.0872	15	0.11	0.1314	
4-class	4-Cluster	-2746.9065	5685.2900	27	6.6142	8	0.58	0.1959	
Model3	4-Cluster	-2747.5483	5644.0231	21	7.8977	14	0.89	0.1969	
Model4	4-Cluster								

- ➢ Change the name 'Model3' to '4-classOrd'
- ➢ Click Profile to view the Profile Output

Figure 7-34. Profile Output for 4 class Ordinal model.



	Cluster1	Cluster2	Cluster3	Cluster4
Cluster Size	0.5267	0.1975	0.1956	0.0802
Indicators				
purpose				
good	0.9720	0.9019	0.3083	0.1769
depends	0.0258	0.0774	0.2168	0.1896
waste	0.0021	0.0207	0.4749	0.6335
Mean	1.0301	1.1188	2.1666	2.4566
accuracy				
mostly true	0.6489	0.6500	0.2437	0.0266
not true	0.3511	0.3500	0.7563	0.9734
understa				
good	0.9789	0.3414	0.9893	0.4837
fair/poor	0.0211	0.6586	0.0107	0.5163
cooperat				
interested	0.9495	0.6891	0.8753	0.3889
cooperative	0.0492	0.2605	0.1168	0.4002
impatient/hostile	0.0013	0.0504	0.0080	0.2109
Mean	1.0518	1.3614	1.1327	1.8220

Notice that clusters 1 and 2 have similar response distributions associated with PURPOSE and ACCURACY, and the same is true for clusters 3 and 4. Also, notice that Clusters 1 and 3 are similar in their response distribution on UNDERSTAND, as is also true of Clusters 2 and 4. This pattern suggests that PURPOSE and ACCURACY may be associated with one DFactor, while UNDERSTAND may be associated with a second.

Estimating a 2-DFactor Model

➢ Right click on Model4 to open the Model Selection menu. ➢ Select DFactor to open the DFactor Analysis dialog box. ➢ Change the number of DFactors

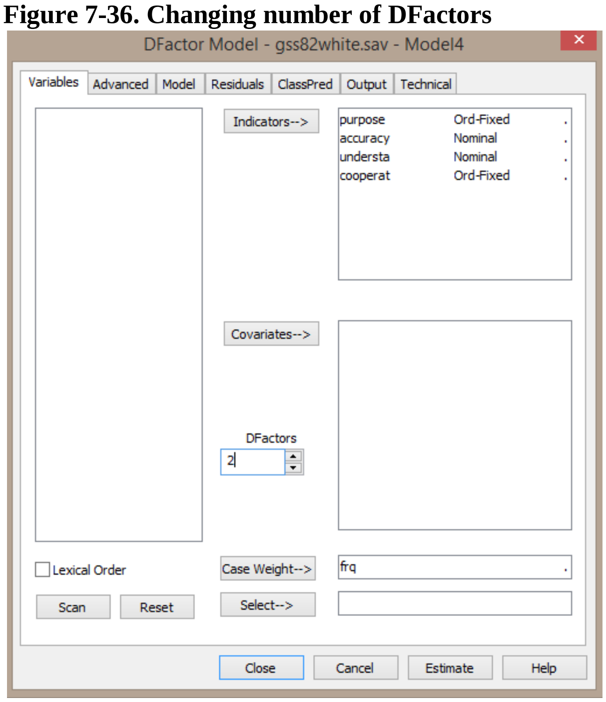
Figure 7-35. Selecting a DFactor Model

The DFactor Analysis Dialog Box opens, and the variable settings appear as before.

To estimate a 2-DFactor model.

➢ Set the number of DFactors in the DFactors Box to 2.

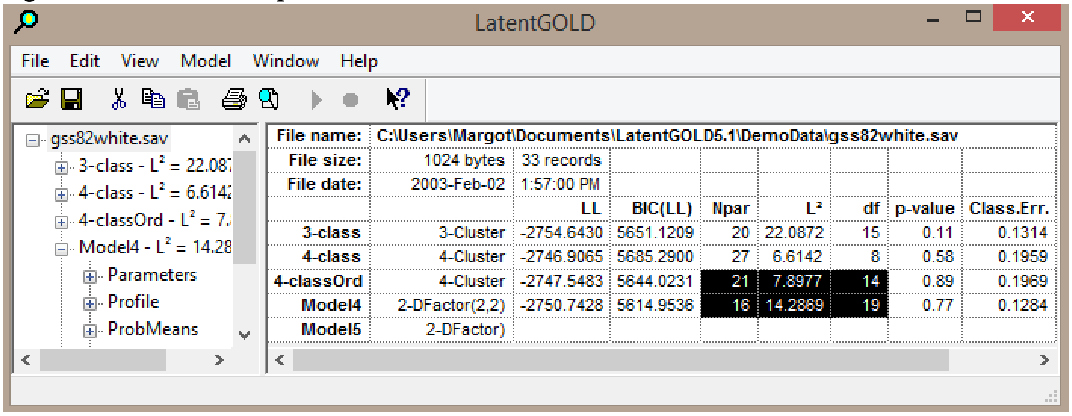
Figure 7-36. Changing number of DFactors



➢ Click Estimate

Now, highlight the file name gss82white.sav again to view the model comparisons

Figure 7-37. Model Comparison – 4-Cluster vs. 2-DFactor models



File name:	C:\Users\Margot\Documents\LatentGOLD5.1\DemoData\gss82white.sav							
File size:	1024 bytes	33 records						
File date:	2003-Feb-02	1:57:00 PM						
		LL	BIC(LL)	Npar	L²	df	p-value	Class.Err.
3-class	3-Cluster	-2754.6430	5651.1209	20	22.0872	15	0.11	0.1314
4-class	4-Cluster	-2746.9065	5685.2900	27	6.6142	8	0.58	0.1959
4-classOrd	4-Cluster	-2747.5483	5644.0231	21	7.8977	14	0.89	0.1969
Model4	2-DFactor(2,2)	-2750.7428	5614.9536	16	14.2869	19	0.77	0.1284
Model5	2-DFactor)							

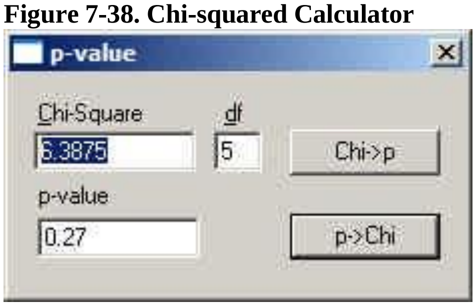
The 2-DFactor model applies further restrictions to the 4-class model, resulting in a model with 5 fewer parameters than model '4-classOrd' (resulting in a gain of 5 df). To test whether such restrictions are justified, we can test to see whether the increase in L^2 of 6.3875 (from 7.8978 to 14.2853) is statistically significant. We will test this in 2 ways.

First, we will use the chi-squared calculator which uses the chi-squared distribution to compute the p-value To open the chi-squared calculator

- Select View/ProbChi
- Enter 6.3875 in the Chi-Square box
- Enter 5 in the df box
- Click the Chi->p button

The p-value of .27 appears in the p-value box

Figure 7-38. Chi-squared Calculator



Since .27 > .05, we fail to reject the restrictions, and so we will accept the 2-DFactor model.

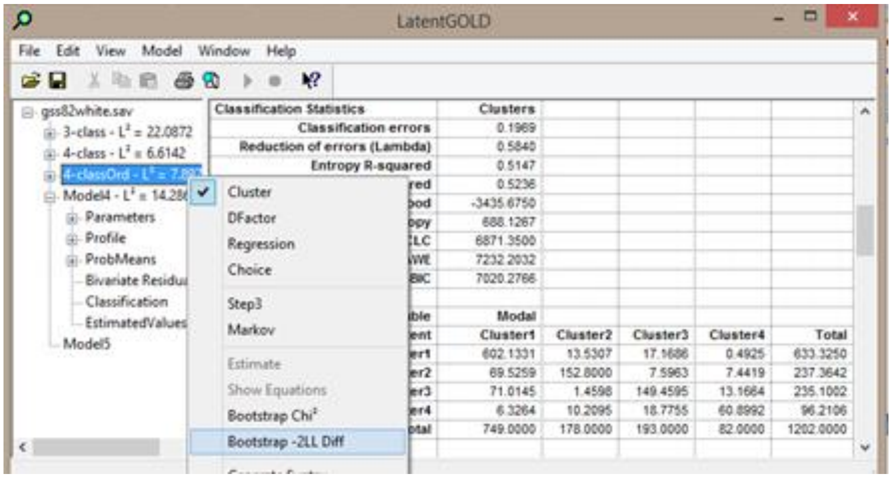
The second way to estimate the p-value utilizes the conditional bootstrap which does not rely on any specific distribution. To use this, we would select model '4-classOrd' as the source and test whether it represents a significant improvement over the reference model 'Model4'.

- Click on '4-classOrd'
- Select 'Bootstrap -2LL Diff' from the Model Menu.

Alternatively, you can

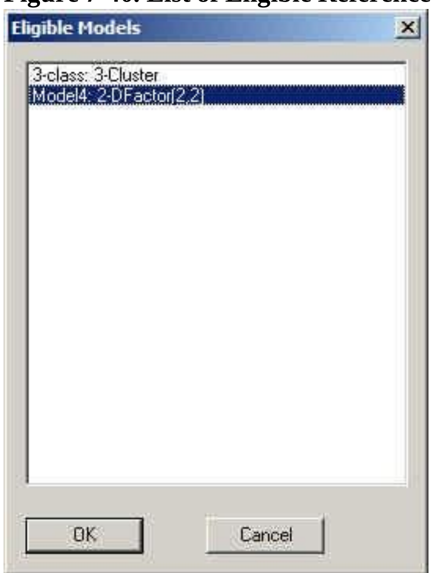
- Right-click on 'Model4'
- Select 'Bootstrap -2LL Diff' from the pop-up menu.

Figure 7-39. Estimating a Bootstrap –2LL Diff Model



Following this, a list of eligible *reference models* appears.

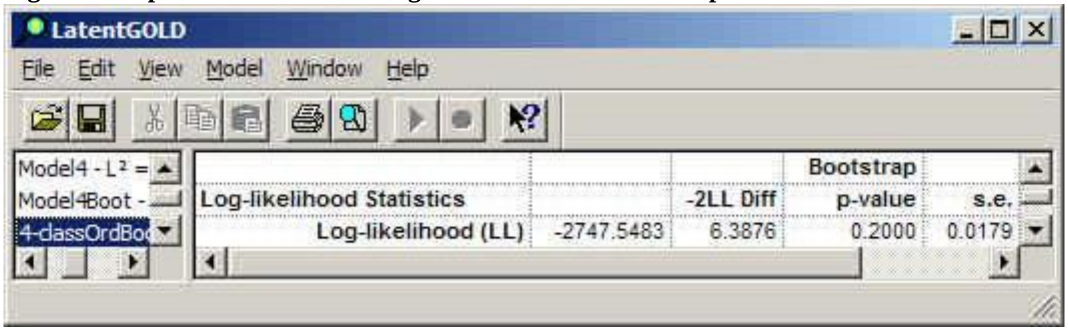
Figure 7-40. List of Eligible Reference Models for Conditional Bootstrap



- ➢ Select ‘Model4’, the 2-DFactor model
- ➢ Click OK

The conditional bootstrap procedure begins. Upon completion, 2 additional models named ‘Model4Boot’ and ‘4-ClassOrdBoot’ appear in the Outline Pane. The results from the conditional bootstrap appear in the Outline Pane associated with model ‘4-ClassOrdBoot’. You may need to scroll down to see these results.

Figure 7-41. **p-value Estimated using the Conditional Bootstrap**



The p-value is estimated to be .20 with a standard error of about .02. This result is similar to what we obtained using the chi-squared approach.

The conditional bootstrap also provides a bootstrap estimate of the p-value associated with the reference model. To view this,

➢ Click on 'Model4Boot'

The Contents Pane shows that the bootstrap estimate for the p-value associated with the 2-DFactor model ('Model4') is .866 with a standard error of .015. Again the results agree with the chi-squared based estimate of .77. (The assumptions underlying the use of the chi-squared based and the bootstrap estimates are both justified in this example.)

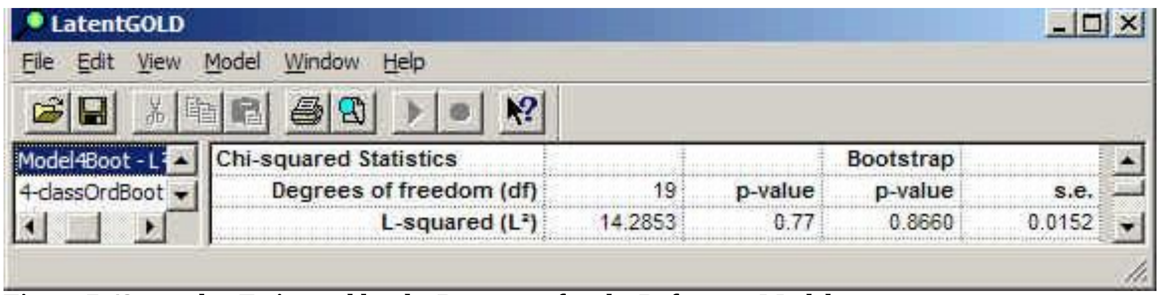
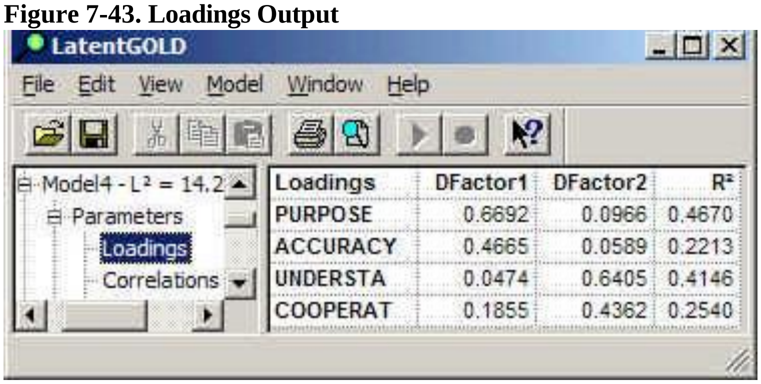


Figure 7-42. p-value Estimated by the Bootstrap for the Reference Model

Next, we will examine the output for the 2-DFactor model.

➢ Click on the expand/contract icon for Model4 to make the output listings visible
➢ Click on the expand/contract icon for Parameters to make the output subcategories visible
➢ Click 'Loadings' to view the DFactor loadings output

Figure 7-43. Loadings Output



This shows that PURPOSE and ACCURACY load primarily on DFactor1 (loadings of .67 and .47 on DFactor1 vs. loadings less than .1 on DFactor2), while UNDERSTAND loads primarily on DFactor2 (loading of .64 on DFactor2 vs. .05 on DFactor1).

Restricting Loadings to Zero

We can use the Model Tab to restrict some of these loadings to zero.

- ➢ Double click Model4 to re-open the DFactor Analysis Box for this model
- ➢ Click on Model to open the Model Tab

By default, DFactor1 is highlighted, indicating that the effects in the Included Effects box pertain to this DFacto
r

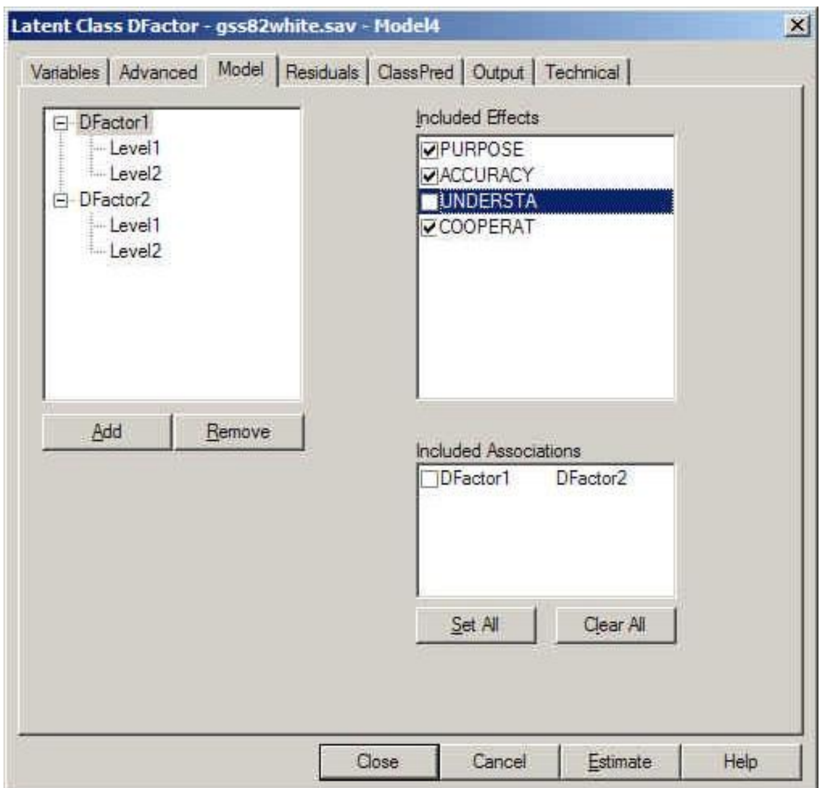


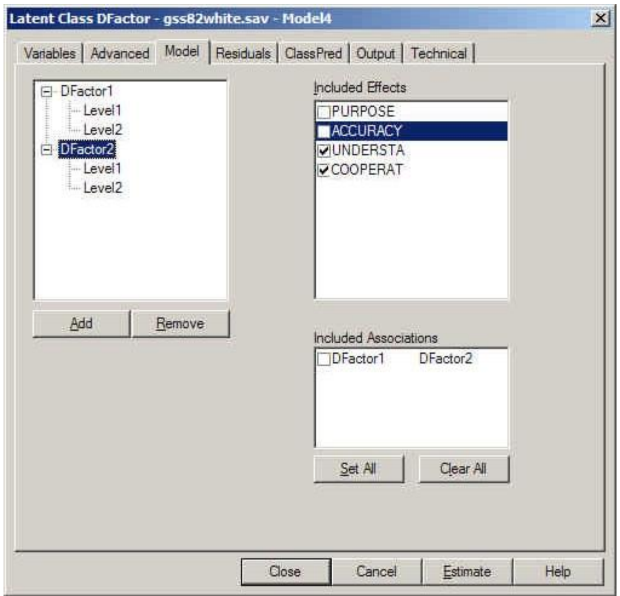
Figure 7-44. Included Effects Box

- ➢ Click the check-box preceding UNDERSTAND to set the loading on DFactor 1 to 0

To set loadings on DFactor2 to 0.

- ➢ Click DFactor2
- ➢ In the Included Effects Box, click to remove the checks for PURPOSE and ACCURACY

Figure 7-45. Restricting Effects to Zero

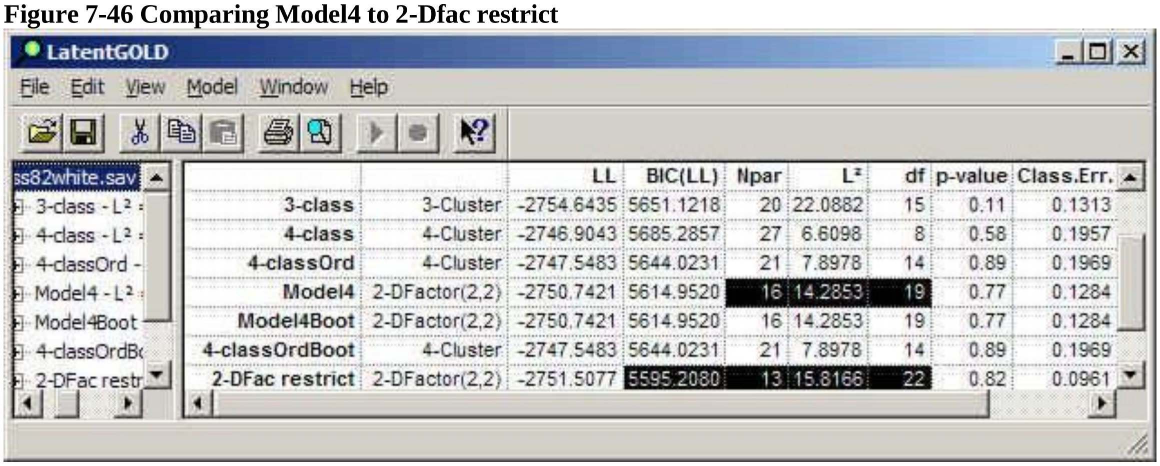


➢ Click Estimate 12

When the estimation is completed, rename the model to '2-DFac restrict'

➢ Click the data file name in the Outline Pane

Figure 7-46 Comparing Model4 to 2-Dfac restrict



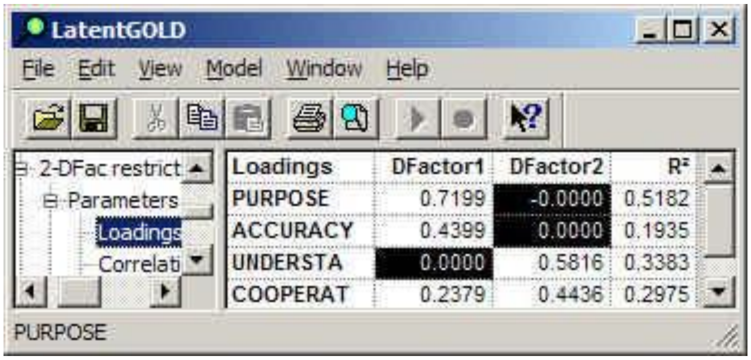
		LL	BIC(LL)	Npar	L²	df	p-value	Class.Err.
3-class	3-Cluster	-2754.6435	5651.1218	20	22.0882	15	0.11	0.1313
4-class	4-Cluster	-2746.9043	5685.2857	27	6.6098	8	0.58	0.1957
4-classOrd	4-Cluster	-2747.5483	5644.0231	21	7.8978	14	0.89	0.1969
Model4	2-DFactor(2,2)	-2750.7421	5614.9520	16	14.2853	19	0.77	0.1284
Model4Boot	2-DFactor(2,2)	-2750.7421	5614.9520	16	14.2853	19	0.77	0.1284
4-classOrdBoot	4-Cluster	-2747.5483	5644.0231	21	7.8978	14	0.89	0.1969
2-DFac restrict	2-DFactor(2,2)	-2751.5077	5595.2080	13	15.8166	22	0.82	0.0961

Comparing the restricted model with the unrestricted 2-DFactor model, we see that the number of parameters has been reduced by 3 due to the 3 parameters that we set to zero, and L^2 increased only slightly. The restricted model is also preferred according the BIC criteria (lowest BIC).

The parameters for this model may be viewed in several different forms. We will look at the factor loadings, and the associated (marginal) conditional probabilities. To view the loadings:

➢ Click on the expand/contract icon for Parameters to make the output subcategories visible
➢ Click 'Loadings'

Figure 7-47. Loadings Output for Model '2-DFac restrict'



Note that the factor loadings associated with the 3 parameter restrictions are zero.

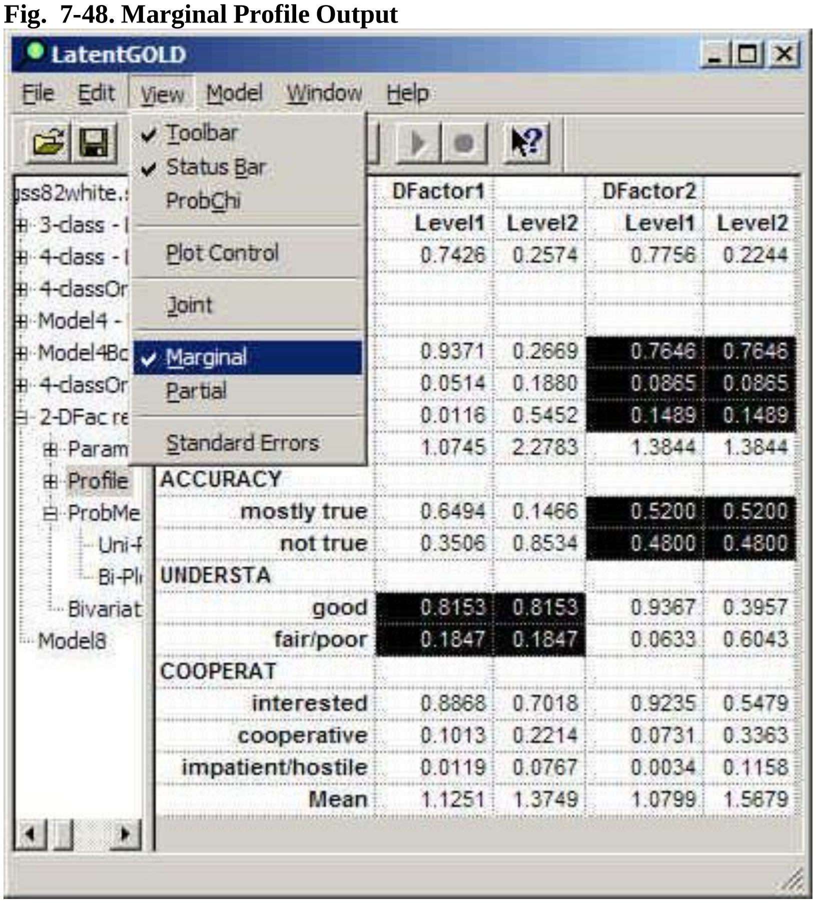
To view the model parameters as (marginal) conditional probabilities:

➢ Click on Profile

The parameters associated with each DFactor are shown in separate columns. Notice that for DFactor2, the conditional probabilities associated with PURPOSE and ACCURACY are identical for each factor level, indicating no effect. The same is true for DFactor1, regarding the effect of UNDERSTAND.

Note: This zero effect pattern would not be seen with *marginal* conditional probabilities if the DFactors were allowed to be correlated in the model. In the correlated situation, partial conditional probabilities would show this same pattern. (You may select Partial from the View menu to replace the marginal probabilities with partial probabilities. When the DFactors are restricted to be uncorrelated, both probability options show this zero-effect pattern).

Fig. 7-48. Marginal Profile Output

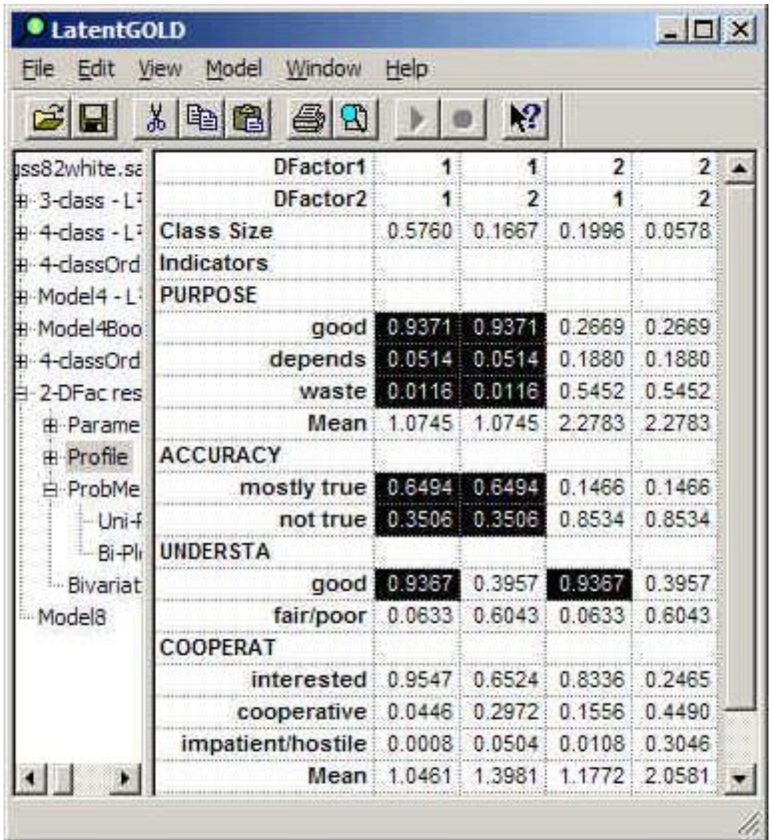


Viewing Joint Profile Output for the 2-DFactor Model

The Joint Profile View re-expresses the parameters in a form comparable to the corresponding cluster model. For this example, there are 4 joint categories formed by cross-tabulating the 2-DFactors, which correspond to 4 clusters.

➢ Select Joint from the View menu

The table now displays the Joint Profile Output



DFactor1	1	1	2	2
DFactor2	1	2	1	2
Class Size	0.5760	0.1667	0.1996	0.0578
Indicators				
PURPOSE				
good	0.9371	0.9371	0.2669	0.2669
depends	0.0514	0.0514	0.1880	0.1880
waste	0.0116	0.0116	0.5452	0.5452
Mean	1.0745	1.0745	2.2783	2.2783
ACCURACY				
mostly true	0.6494	0.6494	0.1466	0.1466
not true	0.3506	0.3506	0.8534	0.8534
UNDERSTA				
good	0.9367	0.3957	0.9367	0.3957
fair/poor	0.0633	0.6043	0.0633	0.6043
COOPERAT				
interested	0.9547	0.6524	0.8336	0.2465
cooperative	0.0446	0.2972	0.1556	0.4490
impatient/hostile	0.0008	0.0504	0.0108	0.3046
Mean	1.0461	1.3981	1.1772	2.0581

Figure 7-49. Joint Profile Output

Note the similarity between this Profile output view and the standard Profile output view obtained earlier for the '4-classOrd' cluster model (recall Fig. 2-9).

This information is plotted in the Joint View of the Profile Plot

➢ Click the expand icon (+) to the left of Profile and click on Prf-Plot.

The Profile Plot for the Joint Profile Output appears.

➢ Right-click on the Plot to view the Plot Control Dialog Box.

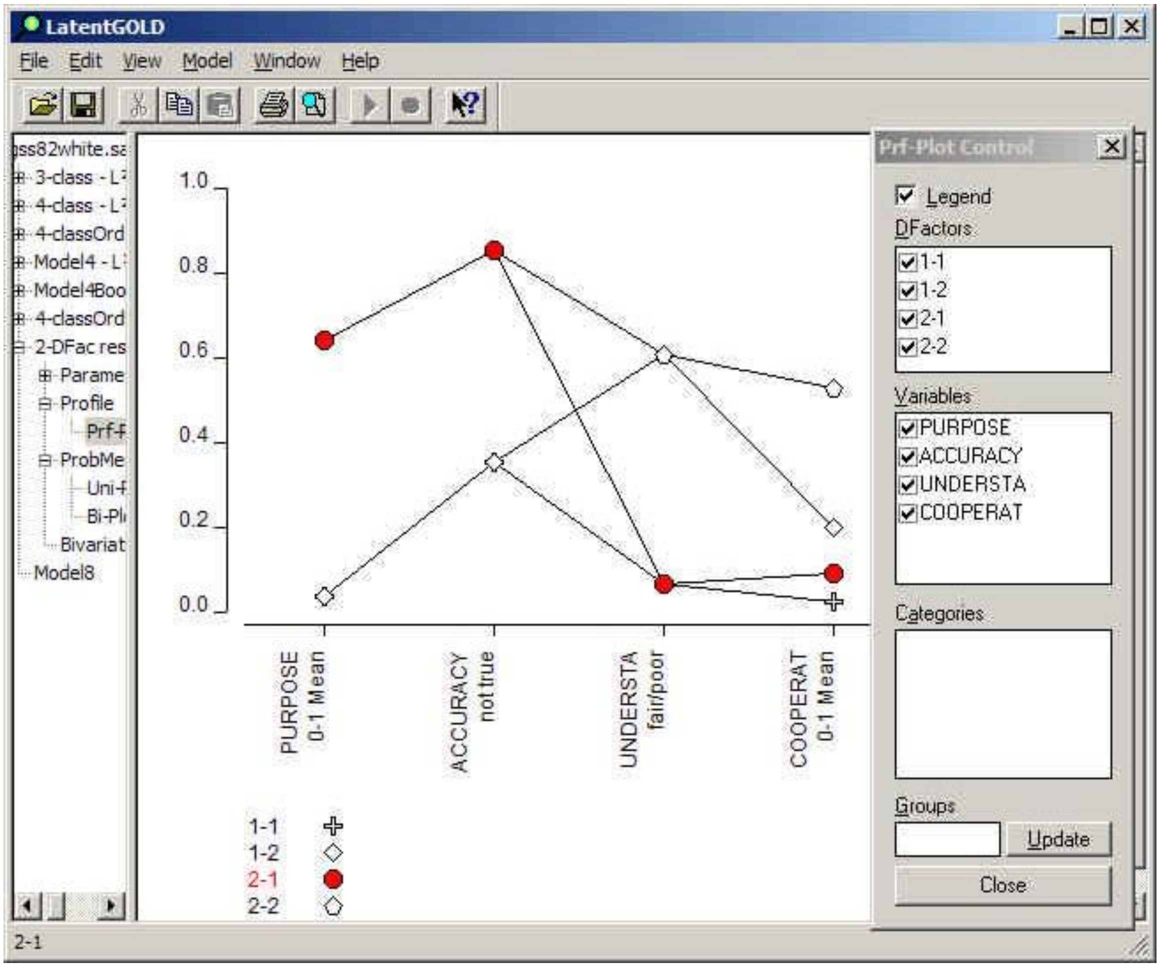
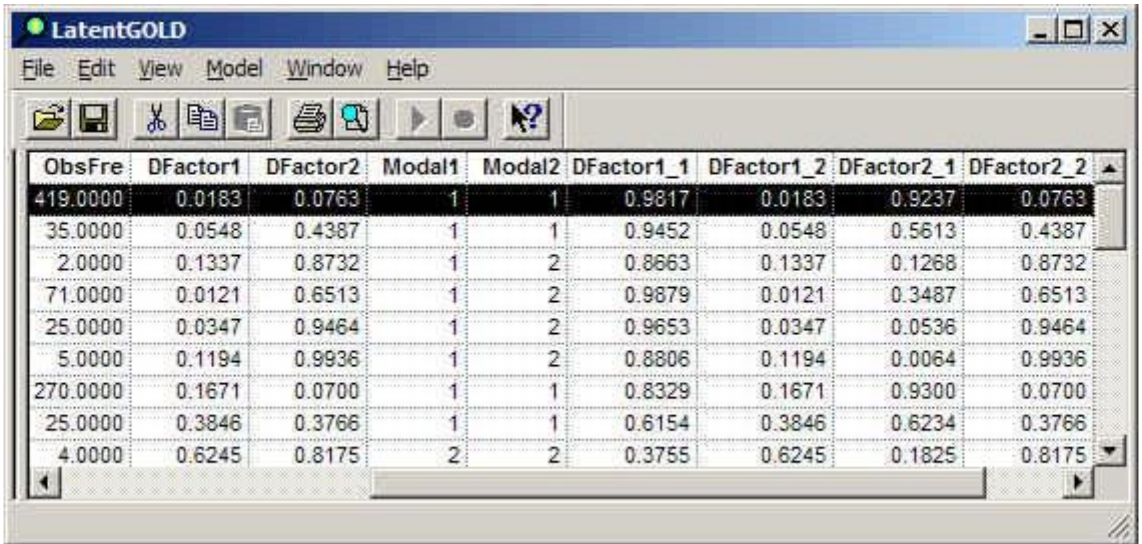


Figure 7-50. Profile Plot

You may use the Plot Control to select/ deselect the joint DFactor levels and variables to be shown on the plot.

Classifying Cases

Standard Classification Information may be requested from the Output Tab prior to estimating the model (Classification – Posterior). The classification information is presented in the right-most columns, as shown in Figure 7-51 below.



ObsFre	DFactor1	DFactor2	Modal1	Modal2	DFactor1_1	DFactor1_2	DFactor2_1	DFactor2_2
419.0000	0.0183	0.0763	1	1	0.9817	0.0183	0.9237	0.0763
35.0000	0.0548	0.4387	1	1	0.9452	0.0548	0.5613	0.4387
2.0000	0.1337	0.8732	1	2	0.8663	0.1337	0.1268	0.8732
71.0000	0.0121	0.6513	1	2	0.9879	0.0121	0.3487	0.6513
25.0000	0.0347	0.9464	1	2	0.9653	0.0347	0.0536	0.9464
5.0000	0.1194	0.9936	1	2	0.8806	0.1194	0.0064	0.9936
270.0000	0.1671	0.0700	1	1	0.8329	0.1671	0.9300	0.0700
25.0000	0.3846	0.3766	1	1	0.6154	0.3846	0.6234	0.3766
4.0000	0.6245	0.8175	2	2	0.3755	0.6245	0.1825	0.8175

Figure 7-51. Standard Classification Output

For each DFactor, this information includes the posterior probability of belonging to each level of that DFactor (e.g., for DFactor 1, DFactor1_1 =.98 and DFactor1_2 = .02), and the corresponding modal levels. For example, the first row contains 419 observations with the response pattern shown in the left-most columns (not visible in Figure7-51). Using the modal assignment rule, these cases would be classified into level 1 of DFactor1 (‘Modal1 = 1) and level 1 of DFactor2 (‘Modal2 = 1).

DFactor scores are also provided. Assigning 0 to the first level of the DFactor, and 1 to the last, the mean score can be computed, using the corresponding posterior probabilities as weights. Thus, for cases in the first row, their scores on DFactor1 and DFactor2 are .0183 and .0763 respectively, which correspond to the posterior probability of being in level 2 of each DFactor.

This classification information will also be appended to your data file if requested from the ClassPred Tab. Such output, which also contains posterior probabilities associated with the joint DFactor is illustrated in Figure 7-54.

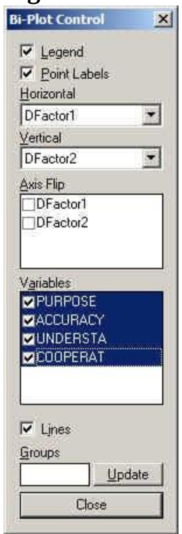
Viewing the Bi-Plot Display for the 2-DFactor Model

Note that the DFactor mean scores -- DFactor1 and DFactor2 -- can be plotted in a 2-dimensional space. While plotting respondents may not be of interest, plotting Indicator categories in a bi-plot display as in Correspondence Analysis may provide useful insights. Each such category can be positioned at a point whose coordinates are aggregated mean DFactor scores obtained for all cases responding in this category. Demographics and other covariate levels can also be appended to this plot. For the exact formula for producing the bi-plot coordinates, see Section 7.4 of Technical Guide.

The DFactor mean scores are summarized in the ProbMeans output. To view the bi-plot of this information:

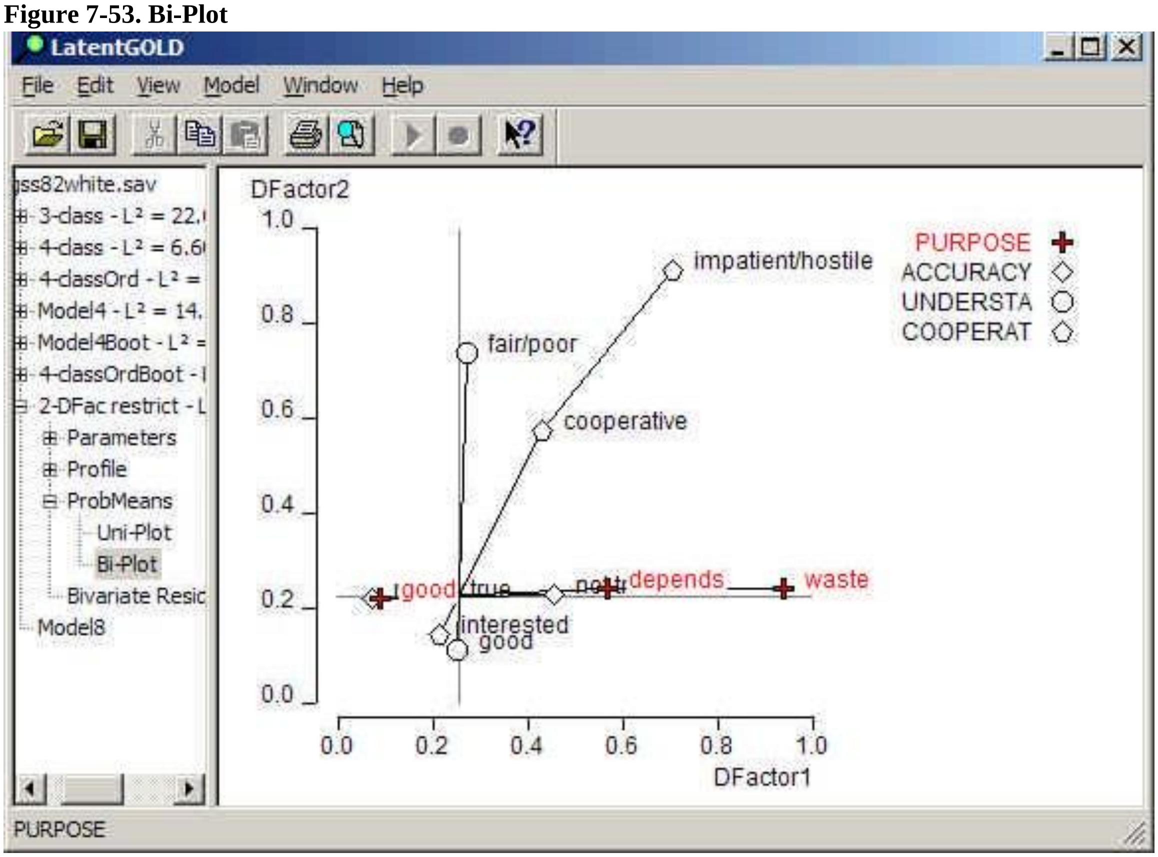
- Click on the expand/contract icon for ProbMeans to make the output subcategories visible
- Click Bi-Plot
- Right click on the bi-plot to retrieve the Plot Control
- Select all the variables and click in the Lines checkbox, to connect the categories of each variable with lines.

Figure 7-52. Plot Control Menu



➢ Click on the K symbol in the plot to highlight the categories of PURPOSE

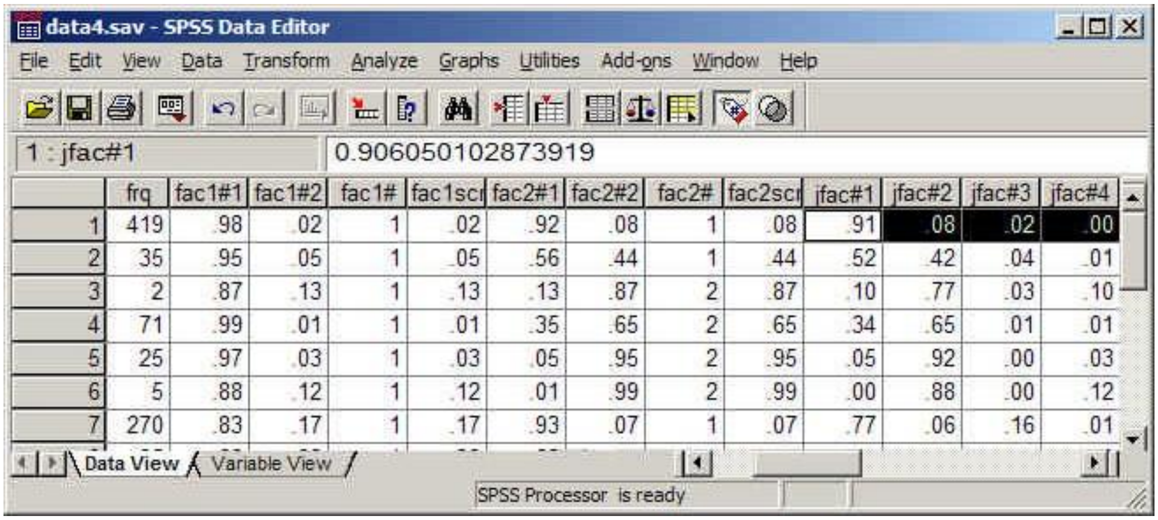
Figure 7-53. Bi-Plot



As can be seen, the categories of PURPOSE (and ACCURACY) vary along the horizontal axis associated with DFactor1 but not the vertical axis associated with DFactor2. Similarly, the categories for UNDERSTAND (denoted by the O symbol), vary only with respect to the DFactor2 axis.

The bi-plot can help you interpret the DFactors and can also serve a diagnostic function prior to restricting DFactor loadings to zero by plotting any 2 DFactors to help determine what restrictions to make.
Figure 7-54 shows the standard classification output as appended to an SPSS .sav file.

Figure 7-54. Standard Classification Output appended to an SPSS .sav file



	frq	fac1#1	fac1#2	fac1#	fac1scr	fac2#1	fac2#2	fac2#	fac2scr	jfac#1	jfac#2	jfac#3	jfac#4
1	419	.98	.02	1	.02	.92	.08	1	.08	.91	.08	.02	.00
2	35	.95	.05	1	.05	.56	.44	1	.44	.52	.42	.04	.01
3	2	.87	.13	1	.13	.13	.87	2	.87	.10	.77	.03	.10
4	71	.99	.01	1	.01	.35	.65	2	.65	.34	.65	.01	.01
5	25	.97	.03	1	.03	.05	.95	2	.95	.05	.92	.00	.03
6	5	.88	.12	1	.12	.01	.99	2	.99	.00	.88	.00	.12
7	270	.83	.17	1	.17	.93	.07	1	.07	.77	.06	.16	.01

This file contains the same information shown in Figure 7-51 plus posterior probabilities associated with all DFactors jointly. For example, for the 419 observations shown in the first row of the file, jfac#1, jfac#2, jfac#3 and jfac#4 contain the posterior probabilities associated with the joint DFactor (labeled within SPSS 'Joint DFactor 1 1', 'Joint DFactor 1 2', 'Joint DFactor 2 1', 'Joint DFactor 2 2' respectively). The most likely joint level for these cases is (1,1) since the probability of being in this level is .9817.