

CHAPTER 6. MODEL AND MODEL SUMMARY OUTPUT

Following estimation of a model, various output file listings and graphical displays are produced. The specific output produced is controlled by the check-boxes on the Output Tab (see Step 9: Set Output Options in Chapter 5). In addition, case level information such as modal class assignments, posterior membership probabilities, regression predictions, and individual coefficients can be requested to be output to a data file using check-boxes on the ClassPred Tab (described at the end of Step 9 in Chapter 5).

This chapter describes each of the various output listings and plots that may be obtained using the Output Tab, and how to view this output. With the Advanced version of Latent GOLD, the Advanced Tab can be used to request the inclusion of CFactors, GClasses, and or GCFactors in a model. When such advanced features are used, additional information is included in the output. This additional information is also described in this chapter in sections labeled 'Advanced'. For further technical details and formulae see Part I (chapters 1-7) of Technical Guide for the various parameters and statistics associated with Latent GOLD Basic, and Part II (chapters 8-12) of Technical Guide for the various parameters and statistics associated with Latent GOLD Advanced.

LATENT GOLD® 4.0 USER'S GUIDE

In general, *all* output listings and displays are organized in sections using a hierarchical structure.

- For each data file opened, the Data File Summary Output contains summary information for all models estimated on these data.
- Within each data file, for each model estimated, the Model Summary Output contains information such as model specifications and performance statistics for that model.
- Within each model, several specific kinds of model output (such as Parameters Output) are listed in separate sections in the Outline pane beneath the model name.
- Within certain model output section listings, subsections containing specialized tables or plots are available as separate files.

The model output files Parameters, Profile, and ProbMeans are produced by default, regardless of the type of model estimated (Cluster, DFactor or Regression), and the Bivariate Residuals are also produced by default in Cluster and DFactor. The Profile and ProbMeans Output re-express certain parameters in the Parameters file in terms of conditional probabilities and/or means, and contain subcategories for plots (e.g., Prf-Plot). Another element common to the output for all three-model types is that output associated with active covariates is always presented in the bottom portion of Parameters Output. For Cluster models, this section is labeled Model for Clusters; for DFactor models, Model for DFactors, and contains the results for each factor separately; for Regression models, it is labeled Model for Classes. If no active covariates are included in the model, results in this section are given for the intercept only which relates to the size of the latent classes.

To view a particular output file, click on a name in the Outline pane and the associated output will be opened and displayed in the Contents pane (see Figure 6-1). To make visible the names of the specific kinds of output produced for an estimated model you will need to click on the expand/contract icon (+) to expand a model name or model output name. For example, to make the Prf-Plot visible you will need to expand the Profile output. When expanded, the expand/contract icon changes from (+) to (-).

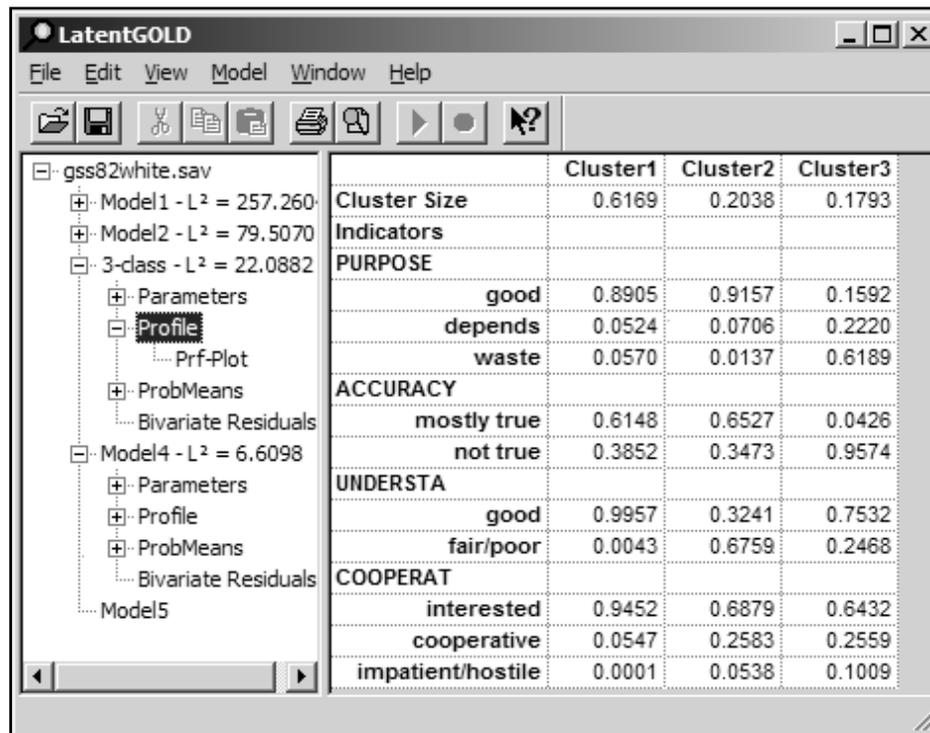


Figure 6-1. Expand/Contract Icon for 3-class Model changes to (-) which makes 'Prf-Plot' Visible. Expand/Contract Icon for 'Model4' remains as (+) so 'Prf-Plot' is Not Visible

Clicking on Prf-Plot causes this display to appear in the Contents Pane. For an example of this, see Figure 6-7.

The expanded output sub-categories may differ by type of model estimated. For example, for DFactor models, Parameters Output contains the subcategory listings Loadings, Correlations, and possibly also Error Correlations which re-express the parameters in a form similar to linear factor analysis, the latter subcategory available only when one or more indicators are continuous. For Cluster models, the Correlations subcategory does not appear and for Regression models, Parameters contains no subcategories. The use of the expand/contract icon is illustrated below.

ORDERING OF LATENT CLASSES IN OUTPUT

Towards the goal of assuring comparability in the output when a model is re-estimated, Latent GOLD applies rules to establish a unique ordering of the classes. In general, classes are ordered by class size, from largest to smallest. However, if parameter restrictions are applied, this rule is ignored to allow for better correspondence between a model when estimated with and without such restrictions. For example, in an unrestricted model if a parameter associated with the largest class (class #1) is restricted to 0, the restricted model will show a 0 for that parameter for class #1, even if class #1 is no longer largest after imposing the restriction.

ITERATION DETAIL

The model output file Iteration Detail produces similar kinds of information regardless of the type of model estimated (Cluster, DFactor or Regression). It is generated only if the Iteration Detail option is selected in the Output Tab before model estimation or if convergence problems or boundary solutions were encountered during the estimation.

This file contains technical information associated with the performance of the estimation algorithm, such as log-posterior and log-likelihood values at selected iterations:

- for each Random Set of starting values,
- for selected iterations of the EM algorithm,
- for selected iterations of the Newton algorithm.

When applicable, this file also contains warning messages concerning nonconvergence, unidentified parameters and boundary solutions, and additional iterations associated with an 'inactive sampling weight' when specified in the Advanced Tab.

Estimation Warnings

WARNING: negative number of degrees of freedom

This warning indicates that the model contains more parameters than cell counts. A necessary (but not sufficient) condition for identification of the parameters of a latent class model is that the number of degrees of freedom is nonnegative. This warning thus indicates that the model is not identified. The remedy is to use a model with fewer latent classes (or DFactors)

WARNING: # boundary or non-identified parameter(s)

This warning is derived from the rank of the information matrix (Hessian or its outer-product approximation). When there are non-identified parameters, the information matrix will not be full rank. The number reported is the rank deficiency, which gives an indication of the number of non-identified parameters.

Note that there are two problems associated with this identification check. The first is that boundary estimates also yield rank deficiencies. In other words, when there is a rank deficiency, we do not know whether it is caused by boundaries or non-identified parameters. The Latent GOLD Bayes Constants prevent boundaries from occurring, which solves the first problem related to this message. However, a second problem is that this identification check cannot always detect non-identification when Bayes Constants are used; that is, Bayes Constants can make an otherwise non-identified model appear to be identified.

A more reliable method to check the identification of a LC model (when estimated with a particular data set) is as follows. Set all Bayes Constants, Random Sets, and both EM and Newton-Raphson Iteration Limits to 0, and select the Standard Errors and Wald method Fast (Outer-Product) prior to estimating the specified model. If no warning message appears, the model is identified. However, non-identification detected in this manner may be related to "true" non-identification (model is non-identified with any data set) or to the fact that there is not enough data (too few non-zero cells in the observed frequency table). The latter will lead to boundary estimates. These two explanations can easily be distinguished by adding one or more fictitious cases to the data set (with response patterns that are not already in the data set). With enough data, if the warning message still appears, the model is non-identified, irrespective of the data set.

WARNING: # CFactor effect(s) should be excluded for identification

For CFactor models with continuous indicators -- standard and mixture factor analysis - there is a special identification check for the factor structure. This may yield a warning indicating how many constraints should be imposed to get an identified factor structure. Usually, identification will be achieved by excluding one or more CFactor-Indicator effects as suggested by the message. Sometimes, one may instead wish to use the "Equal Effects" option to achieve identification.

WARNING: maximum number of iterations reached without convergence

This warning is provided if the maximum specified EM and Newton-Raphson iterations are reached without meeting the tolerance criterion. If the (by default very strict) tolerance is almost reached, the solution is probably ok. Otherwise, the remedy is to reestimate the model with a sharper EM tolerance and/or more EM iterations, which makes sure that the switch from EM to Newton-Raphson occurs later. The default number of 50 Newton-Raphson iterations will generally be more than sufficient.

WARNING: estimation procedure did not converge (# gradients larger than 1.0e-3)

This message may be related to the previous message, in which case the same remedy may be used. If the previous message is not reported, this indicates that there is a more serious non-convergence problem. The algorithms may have gotten trapped in a very flat region of the parameters space (a saddle point). The best remedy is to reestimate the model with other starting sets, and possibly with a larger number of Start Sets and more Iterations per set. If the problem repeats, the likely explanation is that the data is not informative enough about the parameter values.

WARNING: # strata with only one PSU

A requirement for the complex survey estimator of the variance-covariance matrix of the model parameters is that each Stratum contains at least 2 PSUs. If a particular Stratum contains only one PSU, it is omitted from the computations of the variance-covariance matrix, which is clearly not the best solution. A better solution is for you to merge this stratum with another stratum, and re-estimate the model. The best remedy is, therefore, to perform the merging yourself.



For further technical details regarding the estimation, see sections 6.3 - 6.8 in Technical Guide.

The remainder of this chapter provides a description of each of the other output file listings and each output sub-category.

Data File Summary Output and Model Fit

If you highlight a data file name in the Outline pane, the Contents pane lists general summary information regarding the data file (file name, location, size, date) and also lists names of all models that have been estimated in the current Latent GOLD session for this data file. By default, this output file reports the log-likelihood (LL), BIC based on LL, number of parameters (Npar), and the proportion of classification errors (Class.Err.) for all models. In addition, if chi-squared statistics are available, the likelihood-ratio statistic (L^2), degrees of freedom (df), and the p-value are also reported.

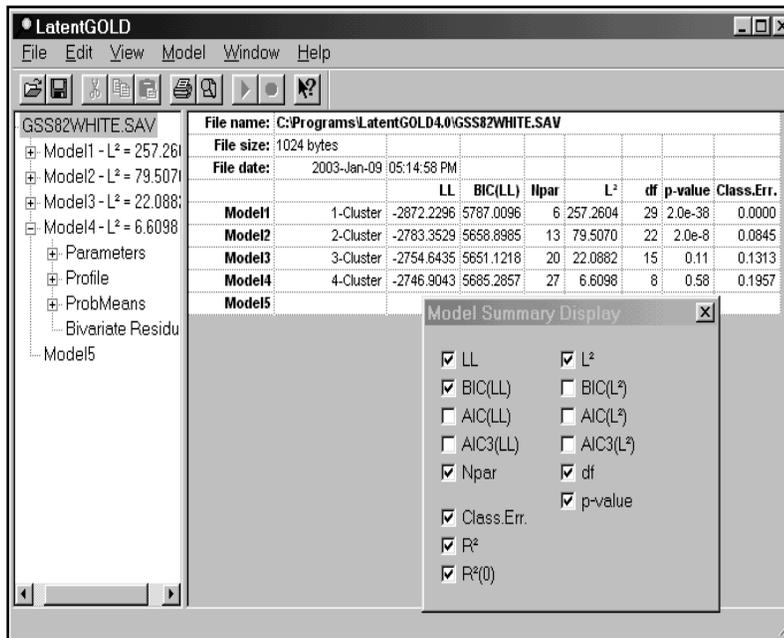


Figure 6-2. Data File Summary Output and associated Model Summary Display

The Model Fit likelihood ratio chi-squared statistic (L^2) is one of several statistics that can be used to assess how well the model fits the data (how similar model-based estimated frequencies are to observed frequencies). In the context of latent class analysis, L^2 can also be interpreted as indicating the amount of the observed relationship

between the variables that remains unexplained by a model; the larger the value, the poorer the model fits the data and the worse the observed relationships are described by the specified model. The associated p-value is a formal assessment of the extent to which the model fits the data (the null hypothesis of this test is that the specified model holds true in the population). It is obtained from a chi-squared table lookup with the reported number of degrees of freedom. Thus, $p < .05$ indicates a poor fit.

As a general rule of thumb, a good fit is provided by a model when the L^2 for that model is not substantially larger than the degrees of freedom which is the expected value for L^2 under the assumptions that 1) the model is true and 2) L^2 follows a chi-square distribution.

When dealing with a small sample size or sparse data, chi-square does not provide a good approximation to L^2 and hence the p-value reported is not valid. Thus, the program also provides optional alternative statistics to assess model fit using the bootstrap procedure. The Bootstrap estimate of the p-value can be obtained when chi-square statistics are available and the conditional Bootstrap (Bootstrap -2LL Diff) can be obtained for any model to assess the significance of the improvement in fit due to an additional latent class, DFactor or DFactor level (see chapter 7 for examples of the use of these Bootstrap procedures).

In addition, information criteria such as the BIC may be used when the table is not sparse as well as when it is sparse. When chi-squared statistics are available such information criteria can be based on L^2 , and when chi-squared statistics are not available, they can be based upon LL.

By selecting 'Summary Control' from the View Menu or with a right click from within the Contents Pane (right-hand) pane, the Model Summary Display menu becomes visible. This can be used to display the BIC and other items or to delete items to customize your summary table by clicking on the associated check-boxes in the Model Summary Display control panel.

Additional items that can be requested are AIC, AIC3, and BIC (based on L^2 or LL) and BIC based on L^2 . In addition to model fit, AIC, AIC3, and BIC take into account the parsimony of the model. They differ from one another according to how much weight is applied to penalize for each additional model parameter. When comparing models, the lower the value of the BIC (or AIC, AIC3), the better the model.



See section 7.1.3 of Technical Guide for specific formulae for each of these statistics.

Model Summary Output

Once estimation of a model has been completed or you have paused the estimation prior to completion (by selecting Stop from the Model Menu or by clicking the Stop button), a Model Summary Output file is generated for that model and a name such as 'Model1', 'Model1Paused' (if the model was Paused), or 'Model1Boot' (if a bootstrap option is used) is assigned by the program and appears in the Outline pane. In addition, if one or more Output listings were requested prior to model estimation using the Output Tab, these listings are also generated and an expand/contract icon (+/-) appears to the left of the Model name. Opening the expand icon (+) by clicking it, changes its appearance to (-) and causes the names associated with these Output listings to appear beneath the model name.

To view the Model Summary Output, click on the model name and the Contents pane lists the contents of this output. Similarly, the contents of any other Model Output file can be viewed by clicking on the desired Output Section name. At the top of the Model Summary Output information appears which describes the type of model estimated (for example, '2-Cluster Model'). The next line contains any warning messages, followed by general model summary information. Additional information follows in sections labeled Chi-squared Statistics, Log-likelihood Statistics, Log-likelihood Statistics, Classification Statistics, Covariate Classification Statistics, and Variable Detail. These are described further below. When the Bootstrap L^2 option is used, the Chi-squared Statistics section also contains the estimate for the associated bootstrap p-value. When the conditional bootstrap option is used, the Log-likelihood Statistics section contains the associated results. More specific information on the statistics including detailed formulae is provided in section 7 of Technical Guide.

Possible warning/alert messages include the following:

Estimation Warnings! This message appears if either boundary solutions, identification problems or convergence problems have been encountered during the estimation of the model. See **Iteration Detail** for more information on estimation warnings.

Model Paused. This message occurs whenever you cause the model to Pause prior to completion of the estimation. The output files in this case should be viewed as preliminary in this case. (To resume a paused model, you may simply right click on the model name of the paused model and select Resume from the pop-up menu.). See Step 10 in Chapter 5 for more information on pausing and resuming model estimation.

The specific contents of the Model Summary Output and other Model Output files depend upon the type of model estimated. In the remainder of this chapter we describe the various model output separately for each of the three modules - Cluster, DFactor and Regression.

Cluster

GENERAL INFORMATION

The Cluster Module is used to estimate LC Cluster models containing 1 or more latent classes (clusters). When 2 or more clusters are included in a model, as a general rule, the clusters are ordered in the output according to their size from the largest to the smallest. Exceptions may occur when parameter restrictions are applied. For example, when the Order Restricted Clusters option is used, Clusters are ordered according to the parameter values.

In addition to the Model Summary Output, additional specific model output file listings are generated following estimation of the model. The specific output sections that can be generated are those that appear as active in the Output Sections portion of the Output Tab. For Cluster models, these are the first 7 model Output Sections listed, plus the last one, Iteration Detail (see Figure 6-3 below).

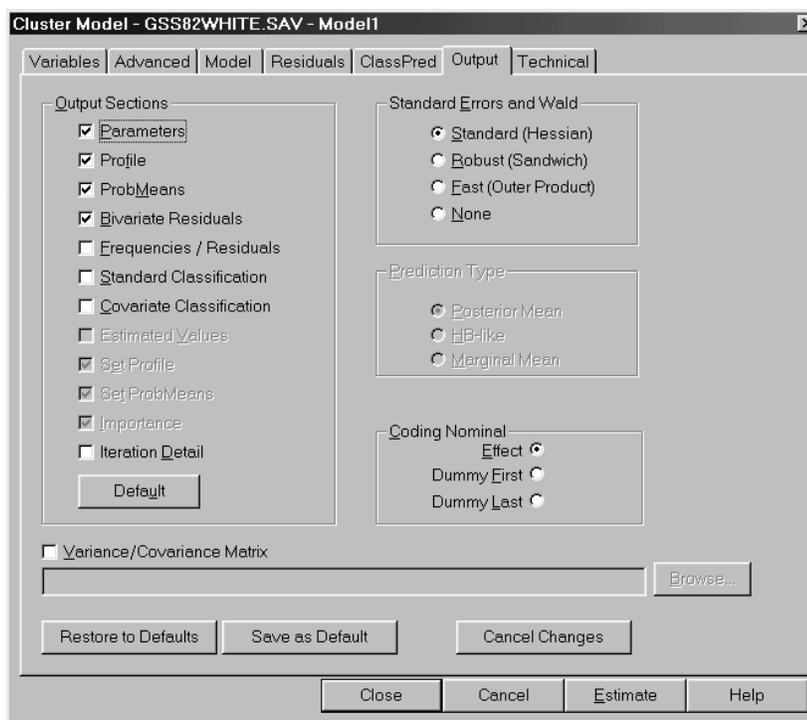


Figure 6-3. Output Tab for Cluster Models

Prior to estimating a Cluster model, click in the active check boxes to the left of the Output Sections to include or exclude that Section from being part of the model output files that are generated following the model estimation. A detailed description of each of the Model Summary Output as well as each specific Model Output Section file that can be output appears below.

LC CLUSTER MODEL SUMMARY OUTPUT

If you click in the Outline pane on the name of any estimated Cluster model (e.g., 'Model1'), the Contents pane lists general summary information associated with this model including the type of model, any warning messages, the value of the Random Seed used in starting sets that can be used to reproduce exactly the same run, and the value of the best seed that can be used to reproduce the reported model results. See section 6.6 of Technical Guide for further details regarding the random seeds. If order-restricted models have been estimated, following the number of parameters, the number of activated constraints is provided.

Specifically, the general section of the Model Summary Output contains the following items:

Type of Model. For example, '3-Cluster Model' indicates that a 3-class Cluster model has been estimated.

Warning messages (generally not present). Estimation Warning message(s) appear here to indicate that boundary solutions, identification or convergence problems were encountered during the estimation of the model.

Model Paused (if you paused the model). If you Paused the model prior to estimation being completed this message appears to alert you that the output files should be viewed as preliminary.

Number of cases. This is the number of cases used in model estimation. This number may be less than the original number of cases on the data file if missing cases have been excluded.

Advanced: Number of groups. In multilevel latent class models, the program reports the number of groups used in model estimation.

Number of parameters. This is the number of distinct parameters estimated.

Number of Activated Constraints (appears only if order-restricted clusters were requested). This is the number of parameters that needed to be restricted in order to obtain ordered clusters plus the number of error variances that were fixed to the smallest allowed value. This number plus the number of parameters estimated equals the total number of model parameters in a model obtained without the order-restriction.

Random Seed. The seed required to reproduce this model.

Best Start Seed. The single best seed that can reproduce this model more quickly using the number of starting sets =0. This is the seed that is automatically inserted in a saved definition (.lgf) file.

Advanced: Design Effect. When the Survey option is used, the program reports the generalized design effect.

3-Cluster Model			
Number of cases	1202		
Number of parameters (Npar)	20		
Random Seed	836991		
Best Start Seed	2129580		
Chi-squared Statistics			
Degrees of freedom (df)	15	p-value	
L-squared (L²)	22.0882		0.11
X-squared	23.5099		0.074
Cressie-Read	22.6617		0.091
BIC (based on L²)	-84.2880		
AIC (based on L²)	-7.9118		
AIC3 (based on L²)	-22.9118		
CAIC (based on L²)	-99.2880		
Dissimilarity Index	0.0272		
Log-likelihood Statistics			
Log-likelihood (LL)	-2754.6435		
Log-prior	-4.8632		
Log-posterior	-2759.5067		
BIC (based on LL)	5651.1218		
AIC (based on LL)	5549.2869		
AIC3 (based on LL)	5569.2869		
CAIC (based on LL)	5671.1218		
Classification Statistics			
Classification errors	0.1313	Clusters	
Reduction of errors (Lambda)	0.6570		
Entropy R-squared	0.5794		
Standard R-squared	0.6116		
Classification log-likelihood	-3224.8126		
AWE	6793.2949		
Classification Table			
	Model		
Probabilistic	Cluster1	Cluster2	Cluster3
Cluster1	704.1204	3.0920	34.5941
Cluster2	72.9409	163.7828	8.1632
Cluster3	27.9387	11.1252	176.2427
Total	805.0000	178.0000	219.0000
			1202.0000

Figure 6-4. Model Summary Output for Cluster Model

Chi-squared Statistics

This section lists various chi-square based statistics related to model fit.

If the scale type for one or more indicators in a model has been set to 'continuous', no chi-squared statistics are available and this section is not displayed. The same applies when the multilevel option (Advanced) is used.

The information reported:

- Degrees of freedom (df). The degrees of freedom for the current model.
- L-squared (L²). The likelihood-ratio goodness-of-fit value for the current model. If the bootstrap p-value for the L² statistic has been requested, the results will be displayed here.
- X-squared and Cressie-Read. These are alternatives to L² that should yield a similar p-value according to large sample theory if the model specified is valid and the data is not sparse.
- BIC, AIC and CAIC (based on L²). In addition to model fit, these statistics take into account the parsimony (df or Npar) of the model. When comparing models, the lower the BIC, AIC and CAIC value the better the model.

- **Dissimilarity Index.** A descriptive measure indicating how much the observed and estimated cell frequencies differ from one another. It indicates the proportion of the sample that needs to be moved to another cell to get a perfect fit.



For more detailed information about Chi-squared statistics including the formulae for each, see section 7.1.1 of Technical Guide.

Log-likelihood Statistics

This section contains additional statistics related to the model fit that are especially useful when L^2 and the other chi-squared statistics are not available. The statistics reported are:

- **Log-likelihood (LL).** If the conditional bootstrap (bootstrap -2LL Diff) has been requested, the results will be displayed here.
- **Log-prior** - this is the term in the function maximized in the parameter estimation that is associated with the Bayes constants. This term equals 0 if all Bayes constants are set to 0.
- **Log-posterior** - this is the function that is maximized in the parameter estimation. The value of the log-posterior function is obtained as the sum of the log-likelihood and log-prior values.

BIC, AIC, AIC3 and CAIC (based on LL) - these statistics (information criteria) weight fit and parsimony by adjusting the LL to account for the number of parameters in the model. The lower the value, the better the model. For example, according to the BIC values shown in Figure 6-4, the 3-class model is preferred over the 1-class, 2-class and 4-class models. Note: the same conclusions are obtained from information criteria based on LL or L^2 . The numbers differ only by a constant that depends on the data set at hand.



For more detailed information about Log-likelihood statistics including the formulae for each, see section 7.1.2 of Technical Guide

Classification Statistics

This information can be used to assess how well the model classifies cases into clusters. The statistics reported are:

- **Classification Errors.** When classification of cases is based on modal assignment (to the class having the highest membership probability), the proportion of cases that are estimated to be misclassified is reported by this statistic. The closer this value is to 0 the better.
- **Reduction of Errors (lambda), Entropy R-squared and Standard R-squared.** These pseudo R-squared statistics indicate how well one can predict class memberships based on the observed variables (indicators and covariates). The closer these values are to 1 the better the predictions.
- **Classification log-likelihood.** Log-likelihood value under the assumption that the true class membership is known.
- **AWE.** Similar to BIC, but also takes classification performance into account.
- **Classification Table.** The Classification Table cross-tabulates modal and probabilistic class

assignments.

- **A** **Standard R-squared** is reported for CFactors and GCFactors. For GClasses, as for Clusters, Classification Errors, Reduction of Errors, Entropy R-squared, and Standard R-squared values are reported.



For more detailed information about the classification statistics including the formulae for each, see section 7.1.3 of Technical Guide.

Covariate Classification Statistics

If one or more active covariates are included in the model, additional statistics are provided as above but these are now based only on the active covariates. The statistics reported are:



The same information is obtained for classification into GClasses based on group-level covariates - **Classification errors, Reduction of Errors (lambda), Entropy R-squared** and **Standard R-squared**



For more detailed information about the classification statistics including the formulae for each, see section 7.1.3 of Technical Guide.

Variable Detail

This section contains details about the variables entered into the model including variable names, scale types, number of categories and category labels and scores (if used).

PARAMETERS OUTPUT (OPTIONAL)

For any estimated model, a Parameters file listing is generated by default. It contains estimates of parameters in the Model for Indicators (betas, sigmas), and in the Model for Clusters (gammas), as well as measures of significance for these estimates and the model R^2 for each indicator.

The R^2 indicates how well an indicator is explained by the model. For **ordinal**, **continuous**, and **counts**, these are standard R^2 measures. For **nominal** variables, these are Goodman-Kruskal tau-b coefficients, representing a weighted average of separate R^2 measures for each category treated as a separate dichotomous response variable. The R^2 is similar to the explained variance in analysis of variance and to item communalities in factor analysis.



If CFactors have been included in the model, estimates and related information for additional parameters (lambdas) are included in the Models for Indicators output.



If GClasses and/or GCFactors have been included in the model, estimates and related information for additional parameters are included in the Models for Indicators and/or the Model for Clusters section of the Parameters output, and a separate section called Model for GClasses appears at the bottom of this output listing.

	Cluster1	Cluster2	Cluster3	Wald	p-value	R²
PURPOSE						
good	0.6766	1.0706	-1.7472	29.5602	6.0e-6	0.3440
depends	-0.4682	0.1950	0.2732			
waste	-0.2084	-1.2656	1.4740			
ACCURACY						
mostly true	0.5695	0.6512	-1.2207	8.3506	0.015	0.2003
not true	-0.5695	-0.6512	1.2207			
UNDERSTA						
good	1.7485	-1.3369	-0.4116	7.4225	0.024	0.4645
fair/poor	-1.7485	1.3369	0.4116			
COOPERAT						
interested	1.9589	-0.8539	-1.1050	18.9606	0.00080	0.1068
cooperative	0.6931	-0.2500	-0.4431			
impatient/hostile	-2.6520	1.1039	1.5481			

Figure 6-5. Parameters Output for Cluster Models

Betas. For Cluster models, the most important betas are that ones that indicate the strength of the effects of the clusters on the indicators. Other betas are the indicator intercepts, the direct associations between categorical (Nominal/Ordinal) indicators, and direct effects of covariates on indicators.

Gammas. Parameters of the multinomial logit model used to predict the clusters as a function of the covariates. Parameters include the intercept as well as the effects of each covariate.

Sigmas. Error variances and covariances for continuous indicators.

Viewing Wald Statistics and Standard Errors

By default, Wald statistics are provided in the output to assess the statistical significance of a set of parameter estimates. For example, for each indicator the Wald statistic tests the restriction that each estimate in the set of beta parameter estimates associated with that indicator equals zero. A non-significant p-value associated with this Wald statistic means that the indicator does not discriminate between the clusters in a statistically significant way. For variables specified as Nominal, the set includes parameters for each category of the variable.

To view standard errors or related statistics associated with the parameter estimates, simply right click on the Parameters Output in the Contents Pane to retrieve a pop-up menu, and select/deselect the items to appear in various columns of this output listing. Alternatively, you can select the desired items from the View Menu. By default, Wald statistics appear in Parameters Output. You can suppress these Wald statistics, or replace them with standard errors, a Z Statistic, or both standard errors and the associated Z Statistic.

Parameters Output Subcategories

Clicking on the + to the left of Parameters makes visible the Parameters Output Subcategories, which provide additional information regarding Parameters Output. The subcategories are Loadings and Error Correlations.

Loadings. The Loadings output contains the 2 columns labeled 'Clusters' and 'R²'. When only these 2 columns appear, the quantities reported under Clusters are simply correlations, representing the square root of the corresponding 'R²' or communality of the associated indicator. These numbers can be interpreted as standardized linear regression coefficients. The R² is the same as reported in the Parameters Output.



Additional columns appear immediately to the right of the 'Clusters' column when CFactors, GClasses and/or GCFactors are included in the model. The quantities reported in the columns to the left of the R² represent the decomposition of the R² into correlation components associated with each of these column effects. For further details, see section 12.2 of Technical Guide.

Error Correlations (available only when one or more indicators are continuous)-

The Error Correlations output provides the estimates of the within-cluster correlations between continuous indicators variables. Often, these correlations are easier to interpret than covariances, which are the actual model parameters. (Note that the off diagonal entries are all 0 unless direct effects between continuous indicators have been included in the model)



For more information, see section 7.2 in the Technical Guide.

PROFILE OUTPUT (OPTIONAL)

To view the profile table for a selected model, click Profile in the Outline pane. The Profile table contains probabilities or means associated with each Indicator or Dependent variable.

The screenshot shows the LatentGOLD software window with the 'Profile' option selected in the left-hand Outline pane. The main window displays a table of profile data for Model4 (L² = 6.6098). The table has five columns: Cluster1, Cluster2, Cluster3, and Cluster4. The rows are grouped by indicator variables: Cluster Size, PURPOSE (good, depends, waste), ACCURACY (mostly true, not true), UIIDERSTA (good, fair/poor), and COOPERAT (interested, cooperative, impatient/hostile).

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster Size	0.5524	0.2237	0.1544	0.0695
Indicators				
PURPOSE				
good	0.9294	0.9131	0.1973	0.2366
depends	0.0347	0.0811	0.2523	0.1475
waste	0.0359	0.0058	0.5504	0.6159
ACCURACY				
mostly true	0.6282	0.6627	0.1383	0.0486
not true	0.3718	0.3373	0.8617	0.9514
UIIDERSTA				
good	0.9916	0.4105	0.9405	0.4389
fair/poor	0.0084	0.5895	0.0595	0.5611
COOPERAT				
interested	0.9569	0.7081	0.8316	0.3340
cooperative	0.0430	0.2496	0.1628	0.3960
impatient/hostile	0.0001	0.0422	0.0056	0.2699

- For a Cluster model, the first row of numbers shows how large each cluster is. For example, Figure 6-6 shows that Cluster 1 contains about 55% of the respondents (.5524) , Cluster 2 contains 22%, Cluster 3 contains 15% and Cluster 4 contains the remaining 7% .
- The body of the table contains (marginal) conditional probabilities that show how the clusters are related to the Nominal, Ordinal or Binomial Count indicator variables. These probabilities sum to 1 within each cluster (column). For example, Figure 6-6 shows that respondents in Cluster 1 have a 92.94% chance of responding that surveys serve a 'good' purpose.
- For indicators specified as Continuous or Count, the body of the table contains means (rates) instead of probabilities. For indicators specified as Ordinal, means are displayed in addition to the conditional probabilities.
- For covariates the Profile output contains rescaled ProbMeans output rather than model probabilities or means. These are aggregate class membership probabilities which are rescaled to sum to one within clusters. For Covariates specified as Numeric, means are displayed in addition to the conditional probabilities, and if the Covariate contains more than 5 distinct values, the probabilities are displayed for 5 grouped ranges. The number of such grouped ranges may be changed (increased or decreased) using Groups option in the Plot Control.
- The probabilities and means that appear in the Profile Output, are displayed graphically in the Profile Plot (see below for details).

Viewing Standard Errors

To view the associated standard errors in a Profile table, from the View Menu select Standard Errors, or simply right click from within the Contents Pane to retrieve the pop-up menu and click Standard Errors. A column containing the standard errors appears to the right of each estimate. Standard Errors are not provided for Covariates.

Partial Profile

Latent GOLD reports not only marginal but also partial probabilities/means for indicators in the Profile Table Output. Except for zero-truncated and zero-censored response variables, these two differ only if direct effects between indicators or between a covariate and indicator are included in a model using the Residuals Tab (See Step 7, Chapter 5). As explained in section 7.3 of the Technical Guide, partial probabilities/means are obtained by conditioning on a certain value (mean) of the other variables involved in the model for the indicator concerned. *Marginal* probabilities/means, on the other hand, are obtained by collapsing over the categories of these "other" variables. Both have their advantages: the *Marginal Profile* output is somewhat easier to interpret and is displayed by default. *Partial Profile* reflects somewhat better the strength of the effects of the cluster on the indicators.

For zero-truncated and zero-censored response variables, the reported class-specific marginal means are expected values conditional on truncation or censoring at $y_{it}=0$. In contrast, the reported partial means are expected values unconditional on truncation or censoring.

The distinction between Marginal and Partial is not relevant for Covariates.

To change from displaying Marginal to Partial probabilities/means, from the View Menu select Partial, or right click from within the Contents Pane to retrieve the pop-up menu and click Partial.



For more detailed information about the Profile Output, see section 7.3 of Technical Guide.

PROFILE PLOT (FOR INDICATORS AND COVARIATES)

To view the Profile Plot, click the '+' icon to expand the Profile output and highlight Prf-Plot. The Profile Plot is constructed from the conditional probabilities for the nominal variables and means for the other indicators and covariates as displayed in the columns of the Profile table. Specific clusters (columns) are selected for display using the Plot Control pop-up menu. The quantities associated with the selected clusters are plotted and connected to form a line graph.

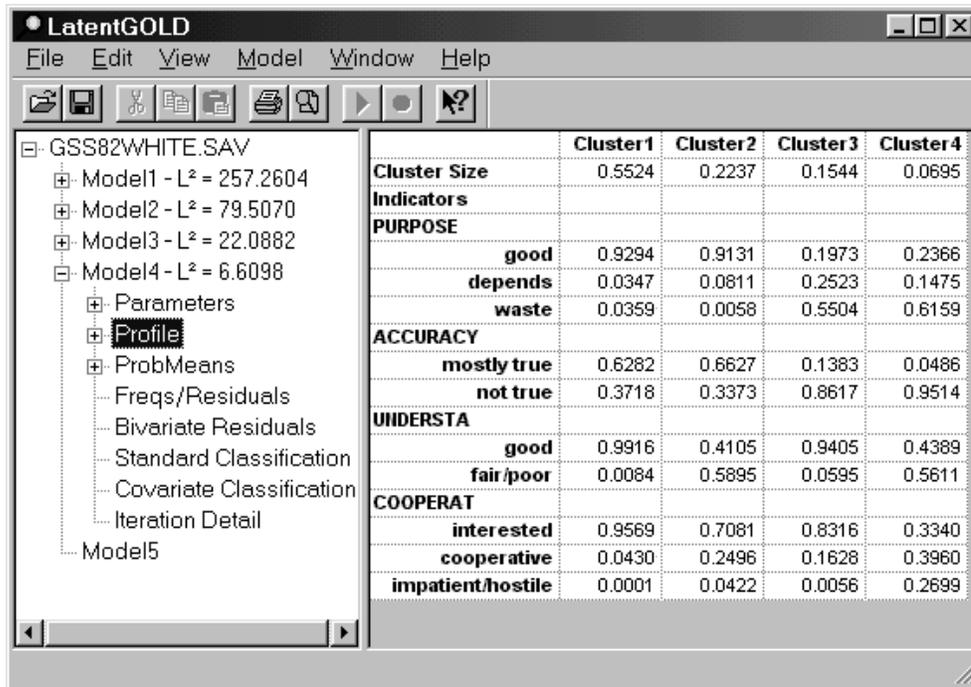


Figure 6-7. Profile Plot for Cluster Model

For the scale types **ordinal**, **continuous**, **count**, and **numeric covariate**, prior to plotting the class-specific means, they are re-scaled to always lie within the 0-1 range. Scaling of these "0-1 Means" is accomplished by subtracting the lowest observed value from the class-specific means and dividing the results by the range, which is simply the difference between the highest and the lowest observed value. The advantage of such scaling is that these numbers can be depicted on the same scale as the class-specific probabilities for **nominal** variables.

For **nominal** variables containing more than 2 categories, all categories are displayed simultaneously. For dichotomous variables specified as nominal, by default only the last category is displayed. The Plot Control (see Figure 6-8 below) can be used to display the first or both categories, as well as to specify the columns (Clusters), variables, and selected categories to appear in the display.

- A separate line is displayed for each cluster.
- Click on any variable symbol in the Profile Plot and the status bar describes it (variable name, cluster number, value). When 0-1 means are displayed, the status bar displays values for both the 0-1 mean and the original mean as shown in the Profile output.
- Click on any cluster name or symbol in the legend and Latent GOLD highlights all the symbols that refer to that cluster.

- When the contents of the Profile Table is changed from the default view to display Partial (instead of Marginal) probabilities, the points plotted change to reflect the current contents of the Profile Table.

By default, only the first 8 variables are displayed in the profile plot. Additional variables may be selected from the Plot Control (using a control-click or a shift click to select more than one at a time) and added or removed from the plot with a single click in the box to the left of the variable names.

To Change Settings for a Profile Plot

To change the settings for a Profile Plot, from the View Menu select Plot Control, or right click within the Contents pane when a Profile Plot is displayed to open the Plot Control dialog box. To change the font type/size for a plot, see Chapter 2.

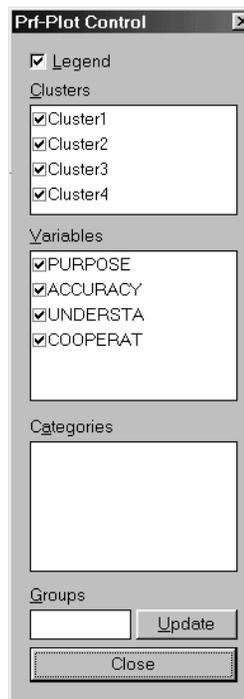


Figure 6-8: Plot Control for Profile Plots

The following plot settings are available for Profile Plots:

Legend. When this option is selected, a legend appears at the bottom of the Profile Plot.

Clusters. For Cluster models, a line will be drawn for each cluster selected. Those clusters with a checkmark in the checkbox are included in the plot.

Variables. Select which variables to include in the plot. Those with a checkmark in the checkbox are included in the plot.

Categories. For Nominal Indicators and Nominal Covariates, select which category of a variable to include in the plot. The category currently being plotted is listed in the plot beneath the variable name. To change the category that is plotted, highlight the variable name in the Variables box (the category currently being plotted will appear in the Category box), click the drop-down list to the right of the Categories box and select the category you wish to have plotted.

Groups: Select a variable for which range groups appear in the Profile Output, and the number of such

groups appears in the Groups Box. Replace this number with a value less than the total number of categories/ values for this variable and click Update. Certain adjacent categories will automatically be combined and the range groups for the selected variable change to the new number of groups in the Profile output and Profile Plot. Entering the value '0' causes each level to appear as a distinct value. The group option may be used with nominal/ordinal indicators and nominal/numeric covariates.

Note: A value of '1' appears in the Groups box for Continuous and Count Indicators when the Plot Control is opened from within the Profile Plot. This indicates that only one value is included in the Profile Output (the mean) and hence entering a different value in the Groups box will not alter the Output for such variables.

A GPROFILE OUTPUT (ADVANCED, OPTIONAL)

If GClasses and/or GCFactors are included in a model, this output file listing is generated. Similar to the Profile output, the top of this file contains the size of each GClass, followed by the probability of being in each cluster for each GClass. This section is followed by the Indicators section where means and (marginal) probabilities associated with each indicator are provided.

The View Menu options (which can be obtained by a right click on the GProfile output), can be used to switch between *Marginal* and *Partial* GProfile Output, or to obtain standard errors.

PROBMEANS OUTPUT (OPTIONAL)

This table contains aggregated cluster membership probabilities for (ranges of) values of indicators and covariates which are displayed in the Uni-Plot and Tri-Plot.

To view the Probability/Means table for a selected model, click ProbMeans in the Outline pane. To view a plot, click on the expand '+' icon to the left of ProbMeans to list the type of plots produced for a particular model. Highlight a plot type to view it in the Contents pane.

	Cluster1	Cluster2	Cluster3	Cluster4
Overall	0.5524	0.2237	0.1544	0.0695
Indicators				
PURPOSE				
good	0.6717	0.2672	0.0397	0.0214
depends	0.2212	0.2096	0.4506	0.1186
waste	0.1328	0.0084	0.5708	0.2879
ACCURACY				
mostly true	0.6676	0.2852	0.0409	0.0063
not true	0.4277	0.1571	0.2773	0.1380
UNDERSTA				
good	0.6719	0.1125	0.1782	0.0373
fair/poor	0.0250	0.7139	0.0496	0.2115
COOPERAT				
interested	0.6304	0.1889	0.1532	0.0276
cooperative	0.1794	0.4221	0.1900	0.2084
impatient/hostile	0.0011	0.3241	0.0295	0.6453

Figure 6-9. ProbMeans Output for Cluster Model

- The first row of the table contains the overall probability of being in a cluster (the size of each cluster), also reported in the first row of numbers in the Profile table.
- The body of the table contains conditional probabilities associated with each category of Nominal and Ordinal indicator variables (these probabilities sum to 100% across rows). For example, in Figure 6-9, for those respondents who responded that surveys serve a 'good' purpose, about 67% are classified as belonging in Cluster 1, 27% in Cluster 2, 04% in Cluster 3, and the remaining 2% in Cluster 4.
- At the bottom of the ProbMeans output, similar information is presented for each covariate.
- For any Indicator with scale type Continuous or Count, or for any numeric covariate, if more than 5 distinct values exists, the probabilities are displayed for 5 grouped ranges. The number of such grouped ranges may be changed (increased or decreased) using the Plot Control.



When the CFactor option is used, ProbMeans contains aggregated posterior CFactor means.

UNI-PLOT

For a Cluster model, the membership probabilities in the body of the ProbMeans output are plotted to form a Uni-Plot. To view the Uni-Plot, click on the expand/contract icon (+) and highlight Uni-Plot.

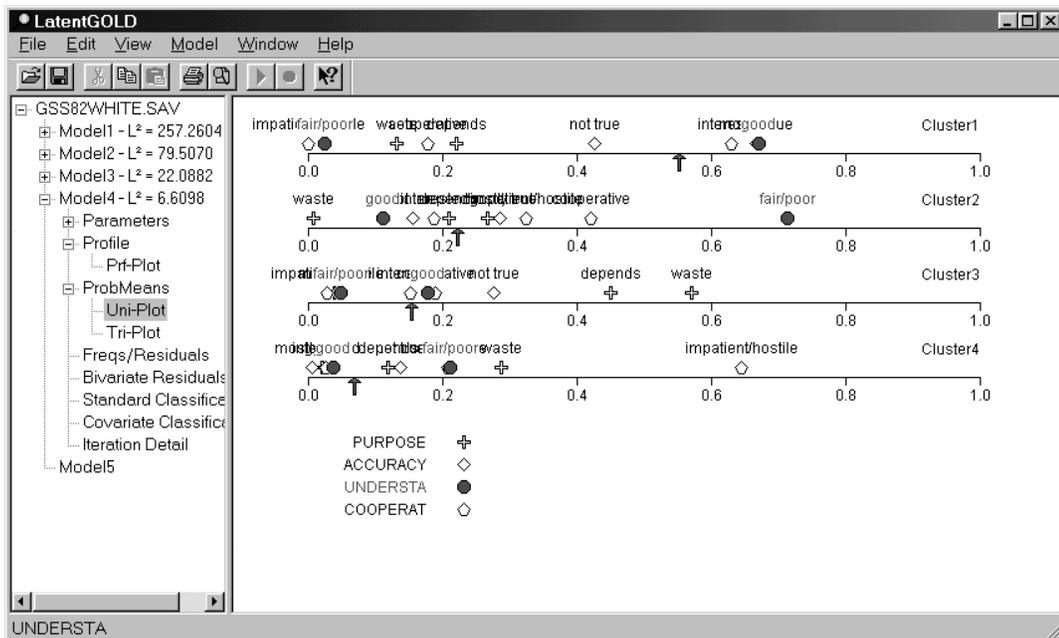


Figure 6-10. ProbMeans Uni-Plot for Cluster Model

- By default, a separate Uni-Plot is created for each cluster. Symbols appear in the plots for each value of each variable specified as an indicator or covariate in the model.
- The ↑ symbol marks the overall probability (the size) for a cluster.
- Click on any variable symbol in the Uni-Plot and the plot label will appear and the status bar will contain a description of the point (variable name and category value).
- Click on any variable name or symbol in the legend and Latent GOLD highlights points associated with that variable.

To Change Settings for a Uni-Plot

To change the settings for a Uni-Plot, right click (or select Plot Control from the View Menu) within the Contents pane when a Uni-Plot is displayed to open the Plot Control. To change the font for a plot, see Chapter 2.

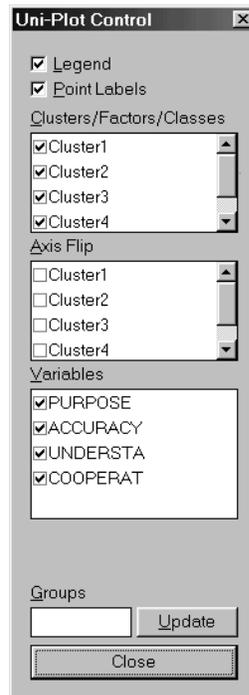


Figure 6-11: Plot Control for Uni-Plots

The following plot settings are available for Uni-Plots:

Legend. When this option is selected, a Legend appears at the bottom of the Uni-Plot.

Point Labels. When this option is selected, category labels for each variable are listed on the Uni-Plot next to the variable symbol.

Clusters/DFactors/Classes. Select which Clusters to include in the Uni-Plots. For each cluster selected (a checkmark in the checkbox) a separate Uni-Plot is displayed. By default, all clusters are selected.

Axis Flip. To flip (reverse) the axis for a Uni-Plot, select the corresponding cluster name. By default, the cluster probabilities range is from 0 to 1 (increasing). Selecting Axis flip for a cluster will reverse the axis to range from 1 to 0 (decreasing).

Variables. Select which indicators/covariates to include in the Uni-Plots. Selected variables are indicated by a checkmark in the checkbox. By default, the Uni-Plots contain the first 16 indicators/covariates included in the model. Additional variables may be selected from the Plot Control (using a control-click or a shift click to select more than one at a time) and added or removed from the plot with a single click in the box to the left of the variable names.

Groups. Select an Indicator with scale type Continuous or Count, or a numeric Covariate for which range groups appear in the ProbMeans Output, and the number of such groups appears in the Groups Box. Replace this number with an alternative number of groups and click Update. The range groups for the selected Covariate change to the new number of groups in the ProbMeans output, the UniPlot and Triplot. Entering the value '0' causes each level to appear as a distinct value.

Groups may be used more generally to produce range groups for the categories of any Nominal or Ordinal Indicator or any Covariate. Select any of these variables and enter a value less than the total number of categories/ values for this variable in the Groups box and click Update. Certain adjacent categories will automatically be combined and the output tables and plots are updated to present results for the associated grouped ranges.

TRI-PLOT

For Cluster models, the cluster membership probabilities in the body of the ProbMeans output table are plotted to form a Tri-Plot. To view the Tri-Plot, click on the expand/contract icon (+) and highlight Tri-Plot. (Note: No Tri-Plot is produced for a 1- Cluster model; for a 2-Cluster model, the Tri-Plot reduces to the Uni-Plot.)

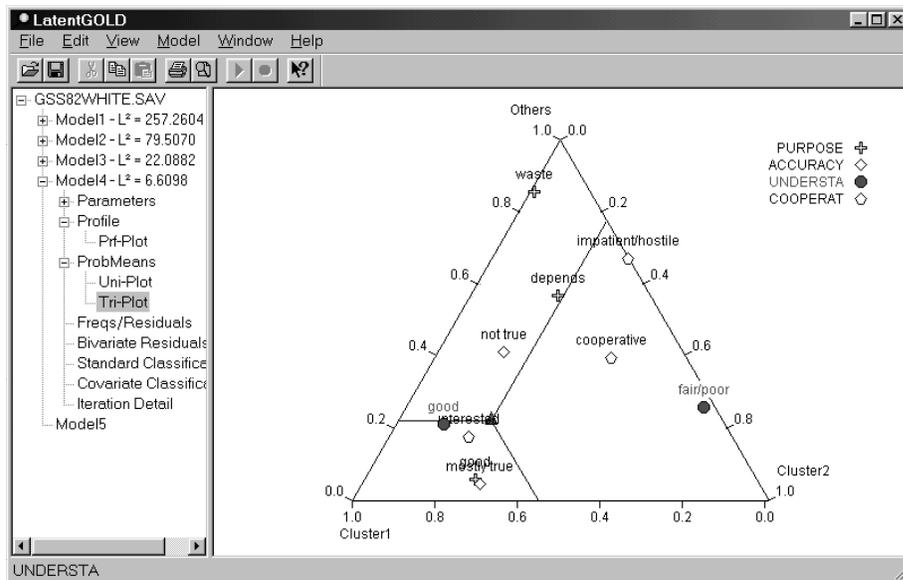


Figure 6-12. Tri-Plot for Cluster Model

- By default, Vertex A (left-most base vertex) is labeled 'Cluster 1', Vertex B (right-most base vertex) 'Cluster 2', and the third Vertex (the top point of the triangle) represents the aggregate of all other clusters. For a 3-Cluster model, by default, the third vertex will represent Cluster 3 and is labeled 'Cluster 3'. For a 4-or-more Cluster model, the third vertex is labeled 'Others'. For a 2-Cluster model, the cluster 3 membership probability is 0 and the Tri-Plot reduces to the Uni-Plot.
- The ▲ symbol marks the overall probabilities for the 3 clusters associated with the vertices. It represents the centroid of the triangle.
- Click on any variable symbol in the Tri-Plot and 1) the status bar contains a description of the point (variable name and category, cluster probabilities) 2) the category label appears next to that point on the plot and 3) lines emanate from that point to each side of the triangle, intersecting the side at the corresponding cluster probabilities value.
- Click on any variable symbol or name in the legend and all the symbols for that variable will be highlighted and their category labels listed in the Tri-Plot.

To Change Settings for a Tri-Plot

To change the settings for a Tri-Plot, right click (or select Plot Control from the View Menu) within the Contents pane when a Tri-Plot is displayed to open the Plot Control dialog box. To change the font for a plot, see Chapter 2.

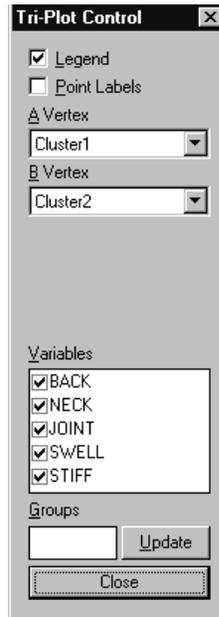


Figure 6-13: Plot Control for Tri-Plots

The following plot settings are available for Tri-Plots:

Legend. When this option is selected, a Legend appears to the right of the Tri-Plot.

Point Labels. When this option is selected, category labels for each selected variable are listed on the Tri-Plot next to the variable symbol.

Vertices. Latent GOLD allows you to select the base vertices in the Tri-Plot. The top vertex corresponds to the aggregate of the remaining clusters.

A vertex. The cluster currently used as the A vertex is listed in the drop down box. To select a different cluster, click on the down arrow to the right of the vertex box. A drop list containing all clusters will appear. Select the cluster to use as the A vertex.

B vertex. The cluster currently used as the B vertex is listed in the drop down box. To select a different cluster, click on the down arrow to the right of the vertex box. A drop list containing all clusters will appear. Select the cluster to use as the B vertex.

Variables. Select which variables to include in the Tri-Plot. Those with a checkmark in the checkbox are included in the plot. By default, the Tri-Plot contains the first 16 indicators/covariates that were input as part of the model. Additional variables may be selected from the Plot Control (using a control-click or a shift click to select more than one at a time) and added or removed from the plot with a single click in the box to the left of the variable names.

Groups. Select an Indicator with scale type Continuous or Count, or a numeric Covariate for which range groups appear in the ProbMeans Output, and the number of such groups appears in the Groups Box. Replace this number with an alternative number of groups and click Update. The range groups for the selected Covariate change to the new number of groups in the ProbMeans output, the UniPlot and Triplot. Entering the value '0' causes each level to appear as a distinct value.

Groups may be used more generally to produce range groups for the categories of any Nominal or Ordinal Indicator or any Covariate. Select any of these variables and enter a value less than the total number of categories/values for this variable in the Groups box and click Update. Certain adjacent categories will automatically be combined and the output tables and plots are updated to present results for the associated grouped ranges.



For more detailed information about the ProbMeans Output, see section 7.4 of Technical Guide.

FREQUENCIES AND RESIDUALS (OPTIONAL)

This output appears only if the Frequencies/ Residuals option was selected in the Output Tab before model estimation.

This output is not available if one or more indicators is continuous or when the multilevel option (Advanced) is used.

Click Freqs/Residuals to view a table containing the observed and expected frequencies along with the standardized residuals for a model.

	accuracy	cooperat	understa	purpose	Observed	Estimated	StdResid	Cook's D
Model 1 - L ² = 40	mostly true	interested	Good	GOOD PURPOSE	535.0000	523.2405	0.5141	0.0026
Parameters	mostly true	interested	Good	DEPENDS	29.0000	45.4186	-2.4362	0.0220
Profile	mostly true	interested	Good	WASTE OF TIME AND \$	32.0000	30.5503	0.2623	0.0657
ProbMeans	mostly true	interested	Fair/Poor	GOOD PURPOSE	105.0000	107.5559	-0.2464	0.0154
Freqs/Residuals	mostly true	interested	Fair/Poor	DEPENDS	9.0000	7.7654	0.4430	0.0743
	mostly true	interested	Fair/Poor	WASTE OF TIME AND \$	4.0000	4.6829	-0.3156	0.1796

Figure 6-14. Freqs/Residuals Output for Cluster.



For more detailed information about the Frequencies Output, see section 7.5 of Technical Guide.

BIVARIATE RESIDUALS (OPTIONAL)

This output only appears if the Bivariate Residuals option was selected in the Output Tab before model estimation

Click Bivariate Residuals to view a table containing the bivariate residuals (BVRs) for a model. A sorted view of this output is provided in the Residuals Tab for an estimated model. Double click on the name of an estimated

model to open the Analysis Dialog Box (or select an estimated model and from the Model Menu select Cluster) and open the Residuals Tab. The BVRs appear sorted from high to low. The Residuals Tab may be used to specify direct effect parameters in a model. Normally, a direct effect parameter associated with a large BVR would be included in a model to improve the fit of the model without the direct effect.

In addition to the L^2 criterion, finding no significant residuals is another indication that a model provides a good fit to the data. In general, BVRs larger than 3.84 identify correlations between the associated variable pairs that have not been adequately explained by the model. (For 1 degree of freedom effects, bivariate residuals larger than 3.84 indicate statistical significance at the .05 level.)

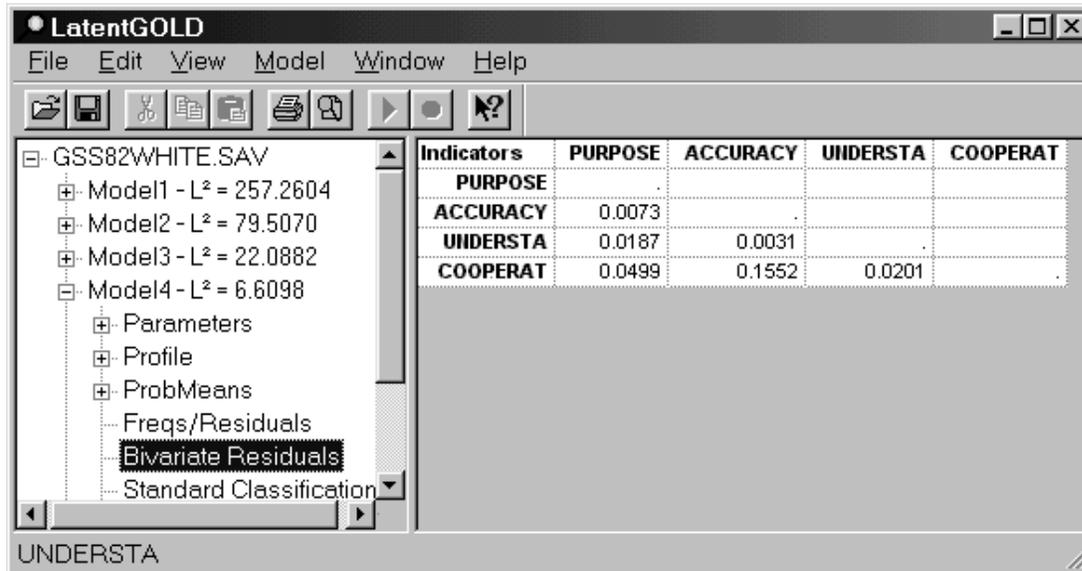


Figure 6-15. Bivariate Residuals Output in Cluster Model



For further details about Bivariate Residuals, see section 7.6 of Technical Guide.

STANDARD CLASSIFICATION (OPTIONAL)

This output only appears if Standard Classification option was selected in the Output Tab before model estimation

Click Standard Classification to view a table containing the posterior membership probabilities and other classification information for a model..

For each observed response pattern, the classification output contains the frequency of occurrence ("ObsFreq"), the cluster for which the posterior membership probability is highest ('Modal'), the associated posterior membership probabilities of belonging to each cluster ('Cluster 1', 'Cluster 2', ...)

PURPOSE	ACCURACY	UNDERSTA	COOPERAT	ObsFreq	Modal	Cluster1	Cluster2	Cluster3	Cluster4
good	mostly true	good	interested	419,0000	1	0.8774	0.1128	0.0095	0.0003
good	mostly true	good	cooperative	35,0000	2	0.4842	0.4882	0.0227	0.0049
good	mostly true	good	impatient/hostile	2,0000	2	0.0088	0.9442	0.0089	0.0381
good	mostly true	fair/poor	interested	71,0000	2	0.0438	0.9502	0.0035	0.0025
good	mostly true	fair/poor	cooperative	25,0000	2	0.0058	0.9834	0.0020	0.0088
good	mostly true	fair/poor	impatient/hostile	5,0000	2	0.0001	0.9649	0.0004	0.0346
good	not true	good	interested	270,0000	1	0.8087	0.0894	0.0917	0.0102
good	not true	good	cooperative	25,0000	1	0.3711	0.3217	0.1832	0.1239
good	not true	good	impatient/hostile	4,0000	4	0.0041	0.3734	0.0432	0.5794
good	not true	fair/poor	interested	42,0000	2	0.0446	0.8328	0.0376	0.0849
good	not true	fair/poor	cooperative	16,0000	2	0.0050	0.7273	0.0182	0.2494
good	not true	fair/poor	impatient/hostile	5,0000	4	0.0000	0.4190	0.0021	0.5788
depends	mostly true	good	interested	23,0000	1	0.5946	0.1820	0.2196	0.0038
depends	mostly true	good	cooperative	4,0000	2	0.1931	0.4637	0.3106	0.0326
depends	mostly true	good	impatient/hostile	1,0000	2	0.0028	0.7027	0.0957	0.1989
depends	mostly true	fair/poor	interested	8,0000	2	0.0177	0.9165	0.0487	0.0170
depends	mostly true	fair/poor	cooperative	2,0000	2	0.0023	0.9136	0.0270	0.0572
depends	not true	good	interested	43,0000	3	0.1865	0.0491	0.7249	0.0395
depends	not true	good	cooperative	9,0000	3	0.0391	0.0807	0.6618	0.2183
depends	not true	good	impatient/hostile	2,0000	4	0.0003	0.0737	0.1229	0.8031
depends	not true	fair/poor	interested	9,0000	2	0.0094	0.4187	0.2722	0.2997

Figure 6-16. Standard Classification for Cluster Model



If 1 or more CFactors have been included in the model, the associated CFactor scores ('Cfac1', 'Cfac2', 'Cfac3'), appear in the right-most portion of the standard classification output.

If 2 or more GClasses have been included in the model, posterior membership probabilities associated with these group classes are given in additional columns (GClass1, GClass2, ...) that appear in the right-most portion of the standard classification output.

If 1 or more GCFactors have been included in the model, the associated factor scores associated with these group level CFactors ('GCfac1', 'GCfac2', 'GCfac3'), appear in the right-most portion of the standard classification output.

COVARIATE CLASSIFICATION (OPTIONAL)

This output only appears if the Covariate Classification option was selected in the Output Tab before model estimation and 1 or more active covariates have been included in the model.

This table contains the estimated probabilities in the multinomial logistic regression model for the clusters, $P(x|z)$, as well as the modal cluster assignments based on these probabilities. The table has one row for each unique covariate pattern..



If 2 or more GClasses have been included in the model, estimated group-level class membership probabilities given group-level covariates are provided in additional columns (GClass1, GClass2, ...) that appear in the right-most portion of the covariate classification output.



For more detailed information about the Classification Output, see sections 7.8 and 12.7 of Technical Guide.

DFactor

GENERAL INFORMATION

The DFactor Module is used to estimate a type of restricted LC Cluster Model known as an DFactor model. DFactor models contain 1 or more discrete factors (DFactors), each of which may contain 2 or more ordered levels (ordered latent classes). A 1-DFactor model containing $K > 2$ levels differs from an order-restricted LC Cluster analysis with K ordered clusters in that the levels of a DFactor are assumed to be equally spaced. For further details see sections 3.7 and 4 of Technical Guide.

When 2 or more DFactors are included in a model, DFactors are ordered according to the R-squared (indicating how well the model predicts the DFactor score). By default, the DFactor having the highest R-squared is listed first (DFactor 1). The levels of the DFactors are ordered such that the lowest level always contains at least as many cases as the highest level.

Note: The ordering of DFactors may not be according to the standard R^2 (reliability) if any of the user options in the Model Tab are applied to increase the number of DFactor levels or restrict effects within one or more classes to zero.

In addition to the Model Summary Output file, additional model output files are generated following estimation of the model. The specific output sections that are available are those that appear as active in the Output Sections portion of the Output Tab. For DFactor models, these are the first 7 model output files listed, plus the last file (Iteration Detail).

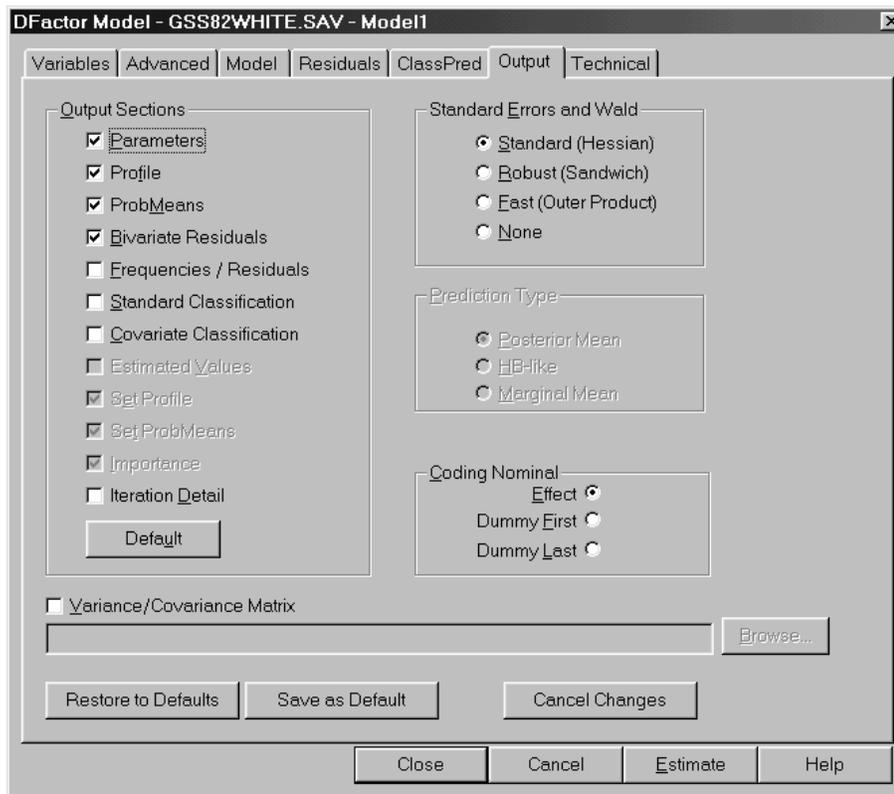


Figure 6-17. Output Tab for DFactor Models

Prior to estimating a DFactor model, click in the active check boxes to the left of the Output Sections to include or exclude that Section from being part of the model output files that are generated following the model estimation. A detailed description of each of the Model Summary Output as well as each specific Model Output Section file that can be output appears below.

DFACTOR MODEL SUMMARY OUTPUT

Following the estimation of a DFactor model, if you click on the Model name in the Outline pane, the Model Summary Output for this model will be displayed the Contents Pane. The information presented is similar to that reported for the LC Cluster Models.

The general section of the Model Summary Output contains the following items:

Type of Model. For example '2-DFactor Model' describes a DFactor model containing 2 discrete factors.

Warning messages (generally not present). Estimation Warning message(s) appear here to indicate that boundary solutions, identification or convergence problems were encountered during the estimation of the model.

Model Paused (if you paused the model). If you Paused the model prior to estimation being completed this message appears to alert you that the output files should be viewed as preliminary.

Number of cases. This is the number of cases used in model estimation. This number may be less than the original number of cases on the data file if missing cases have been excluded.

A **Number of groups:** In multilevel latent class models, the program reports the number of groups used in model estimation.

Number of parameters. This is the number of distinct parameters estimated.

Random Seed. The seed required to reproduce this model.

Best Start Seed. The single best seed that can reproduce this model more quickly using the number of starting sets =0. This is the seed that is automatically inserted in a saved definition (.lgf) file.

A **Design Effect.** When the Survey option is used, the program reports the generalized design effect.

2-Factor Model		
	Factor1(2)	Factor2(2)
Number of cases	1202	
Number of parameters (Npar)	16	
Random Seed	1147811	
Best Start Seed	1424465	
Chi-squared Statistics		
Degrees of freedom (df)	19	p-value
L-squared (L²)	14.2853	0.77
X-squared	11.5619	0.90
Cressie-Read	12.0806	0.88
BIC (based on L²)	-120.4578	
AIC (based on L²)	-23.7147	
AIC3 (based on L²)	-42.7147	
CAIC (based on L²)	-139.4578	
Log-likelihood Statistics		
Log-likelihood (LL)	-2750.7421	
Log-prior	-5.0898	
Log-posterior	-2755.8318	
BIC (based on LL)	5614.9520	
AIC (based on LL)	5533.4841	
AIC3 (based on LL)	5549.4841	
CAIC (based on LL)	5630.9520	
Classification Statistics		
	Factor1	Factor2
Classification errors	0.1284	0.1155
Reduction of errors (Lambda)	0.5715	0.5184
Entropy R-squared	0.4875	0.4272
Standard R-squared	0.5444	0.4980
Classification log-likelihood	-3500.1696	
AWE	7275.2750	

Figure 6-18. Model Summary Output for DFactor Model

Chi-squared Statistics

This section lists various chi-square based statistics related to model fit.

If the scale type for one or more indicators in a model has been set to 'continuous' or when the multilevel option (Advanced) is used, no chi-squared statistics are available and this section is not displayed.

The information reported:

Degrees of freedom (df). The degrees of freedom for the current model.

L-squared (L²). The likelihood-ratio goodness-of-fit value for the current model. If the bootstrap p-value for the L² statistic has been requested, the results will be displayed here.

X-squared and Cressie-Read. These are alternatives to L² that should yield a similar p-value according to large sample theory if the model specified is valid and the data is not sparse.

BIC, AIC and CAIC (based on L²). In addition to model fit, these statistics take into account the parsimony (df or Npar) of the model.

Dissimilarity Index. A descriptive measure indicating how much the observed and estimated cell

frequencies differ from one another. It indicates the proportion of the sample that needs to be moved to another cell to get a perfect fit.

Log-likelihood Statistics

Log-likelihood (LL). If the conditional bootstrap (bootstrap -2LL Diff) has been requested, the results will be displayed here.

Log-prior. This is the term in the function maximized in the parameter estimation that is associated with the Bayes constants. This term equals 0 if all Bayes constants are set to 0.

Log-posterior. This is the function that is maximized in the parameter estimation. The value of the log-posterior function is obtained as the sum of the log-likelihood and log-prior values.

BIC, AIC, AIC3 and CAIC (based on LL) - these statistics (information criteria) weight fit and parsimony by adjusting the LL to account for the number of parameters in the model. The lower the value, the better the model. For example, according to the BIC values shown in Figure 6-18, the 3-class model is preferred over the 1-class, 2-class and 4-class models.

Classification Statistics

This information can be used to assess how well the model classifies cases into clusters. The statistics reported are:

Classification Errors. When classification of cases is based on modal assignment (to the DFactor level having the highest membership probability), the proportion of cases that are expected to be misclassified is reported by this statistic. The closer this value is to 0 the better.

Reduction of Errors (lambda), Entropy R-squared and Standard R-squared. These statistics indicate how well the model predicts class memberships or DFactor scores. The closer these values are to 1 the better the predictions.

Classification log-likelihood. Log-likelihood value under the assumption that the true class membership is known.

AWE. Similar to BIC, but also takes classification performance into account.

Classification Table. The Classification Table cross-tabulates modal and probabilistic class assignments.

A **Standard R-squared** is reported for CFactors and GCFactors. For GClasses, as for Clusters, Classification Errors, Reduction of Errors, Entropy R-squared, and Standard R-squared values are reported.



For more detailed information about the classification statistics including the formulae for each, see section 7.1.3 of Technical Guide.

Covariate Classification Statistics

If one or more active covariates is included in the model, additional statistics are provided as above but these are now based only on the active covariates. The statistics reported are:

Classification errors, Reduction of Errors (lambda), Entropy R-squared and Standard R-squared

A **Standard R-squared** is reported for CFactors and GCFactors. For GClasses, as for Clusters, Classification Errors, Reduction of Errors, Entropy R-squared, and Standard R-squared values are reported.



For more detailed information about the classification statistics including the formulae for each, see section 7.1.4 of Technical Guide.

Variable Detail

This section contains details about the variables entered into the model including variable names, scale types, number of categories and category labels and scores (if used).

PARAMETERS OUTPUT (OPTIONAL)

For any estimated DFactor model, a Parameters file listing is generated by default. It contains estimates of parameters in the Model for Indicators (betas, sigmas), followed by the estimates in the Model for DFactors (gammas), as well as measures of significance for these estimates and the model R^2 for each indicator. The parameter estimates are arranged in separate columns corresponding to each DFactor included in the model.



If CFactors have been included in the model, estimates and related information for additional parameters (lambdas) are included in the Models for Indicators output.



If GClasses and/or GCFactors have been included in the model, estimates and related information for additional parameters are included in the Models for Indicators and/or Model for DFactors section of the output, and a separate section called Model for GClasses appears beneath this section.

Models for Indicators							
	Factor1	Wald	p-value	Factor2	Wald	p-value	R²
PURPOSE							
	2.5163	14.5193	0.00014	0.5760	0.9819	0.32	0.4670
ACCURACY							
mostly true	-1.1807	37.5215	9.0e-10	-0.1750	0.4814	0.49	0.2213
not true	1.1807			0.1750			
UNDERSTA							
good	0.2315	0.2448	0.62	-1.7988	2.0306	0.15	0.4146
fair/poor	-0.2315			1.7988			
COOPERAT							
	1.0344	5.8527	0.016	2.1478	4.7648	0.029	0.2540
Intercepts	Overall	Wald	p-value				
PURPOSE							
good	2.7286	48.0558	3.7e-11				
depends	-0.4982						
waste	-2.2305						
ACCURACY							
mostly true	0.4034	4.1121	0.043				
not true	-0.4034						

Figure 6-19. Parameters Output for DFactor Model

The Models for Indicators section contains:

Betas. For DFactor models, effects are organized separately for each DFactor. The most important betas are that ones that indicate the strength of the effects of the clusters on the indicators. Other betas are the indicator intercepts, the direct associations between categorical (Nominal/Ordinal) indicators, and direct effects of covariates on indicators.

Sigmas. Error variances and covariances for continuous indicators.

The Models for DFactors section contains:

Gammas. Parameters of the logit model used to predict each DFactor as a function of the covariates. These parameters include the intercepts, the associations between the DFactors, as well as the effects of each covariate on each DFactor.

Viewing Wald Statistics and Standard Errors

By default, Wald statistics are provided in the output to assess the statistical significance of a set of parameter estimates. For example, for each indicator the Wald statistic tests the restriction that each estimate in the set of beta parameter estimates associated with that indicator equals zero. A non-significant p-value associated with this Wald statistic means that the indicator does not discriminate between the clusters in a statistically significant way. For variables specified as Nominal, the set includes parameters for each category of the variable.

To view standard errors or related statistics associated with the parameter estimates, simply right click on the parameters output and select/deselect the items that you want to appear. Alternatively, you can click on the appropriate item in the View Menu. In Parameters, you can suppress the Wald statistics, or replace the Wald with standard errors, a Z Statistic, or both standard errors and the associated Z Statistic for each parameter estimate.

Parameters Output Subcategories

Clicking on the + to the left of Parameters makes visible the Parameters Output Subcategories, which provide additional information regarding Parameters Output. The subcategories are Loadings, Correlations, and Error Correlations.

Loadings. DFactor loadings for the indicators are given in separate columns for each DFactor, labeled 'DFactor1', 'DFactor2', ... followed by a column labeled 'R²' representing the communality of the indicator. The quantities reported in the DFactor columns are comparable to loadings (standardized regression coefficients) in traditional linear factor analysis. The R² is the same as reported in the Parameters Output.



Additional columns appear immediately to the right of the 'DFactors' columns when CFactors, GClasses and/or GCFactors are included in the model. The quantities reported in the columns to the left of the R² represent the decomposition of the R² into components associated with each of these column effects. For further details, see the Technical Guide.

Correlations. These are the correlation between the DFactors and the correlation between DFactors and indicators. Correlations between the DFactors are equal to 0 when neither DFactor associations nor Covariates are included in the model. In this case (and only in this case) DFactor-Indicator Correlations will be identical to the corresponding Loadings.

Error Correlations (available only when one or more indicators are continuous) - The Error Correlations output provides the estimates of the within-class correlations between continuous indicators. Often, these correlations are easier to interpret than covariances, which are the actual model parameters. (Note that the off diagonal entries are all 0 unless direct effects between continuous indicators have been included in the model)



For more detailed information about the Parameters Output, see section 7.2 of Technical Guide.

PROFILE OUTPUT (OPTIONAL)

To view the profile table for a selected model, click Profile in the Outline pane. The Profile table contains conditional response probabilities and/or means associated with each Indicator. Means are computed based on the category scores, using the conditional probabilities as weights. For Covariates the Profile table reports aggregated classification probabilities which are rescaled to sum to 1 within DFactor levels.

	Factor1		Factor2	
	Level1	Level2	Level1	Level2
Factor Level Size	0.7003	0.2997	0.7601	0.2399
Indicators				
PURPOSE				
good	0.9457	0.3414	0.7858	0.6976
depends	0.0443	0.1852	0.0839	0.0949
waste	0.0100	0.4734	0.1303	0.2076
Mean	1.0643	2.1320	1.3446	1.5100
ACCURACY				
mostly true	0.6724	0.1637	0.5365	0.4676
not true	0.3276	0.8363	0.4635	0.5324
UIIDERSTA				
good	0.8032	0.8434	0.9549	0.3729
fair/poor	0.1968	0.1566	0.0451	0.6271
COOPERAT				
interested	0.8801	0.7419	0.9257	0.5628
cooperative	0.1053	0.1952	0.0710	0.3262
impatient/hostile	0.0146	0.0629	0.0032	0.1109
Mean	1.1345	1.3210	1.0775	1.5481

Figure 6-20. Profile Output for DFactor Model

- By default, columns of the Profile table display information separately for each level of each DFactor
- At the top of this output table, the size of each level of each DFactor is indicated (for each DFactor these numbers sum to 100%). For example, regarding DFactor #1, 70% are in level 1, the remaining 30% in level 2 of this DFactor.
- The body of the table contains (marginal) conditional probabilities that show how the DFactor levels are related to the indicator variables and covariates. Within each level, these probabilities sum to 1. For ordinal indicators and numeric covariates, cluster-specific means are also reported beneath the probabilities.
- For indicators specified as Continuous or Count, the body of the table contains means instead of probabilities.
- Optionally, when 2 or more DFactors are included in a model, the columns may be changed to correspond to joint levels of the DFactors (see the section on Joint Profile below). The joint view makes clear that the DFactor model is an LC Cluster model. For example, Figure 6-21 below shows the 4 clusters corresponding to the joint DFactor levels (1,1), (1,2), (2,1), and (2,2).

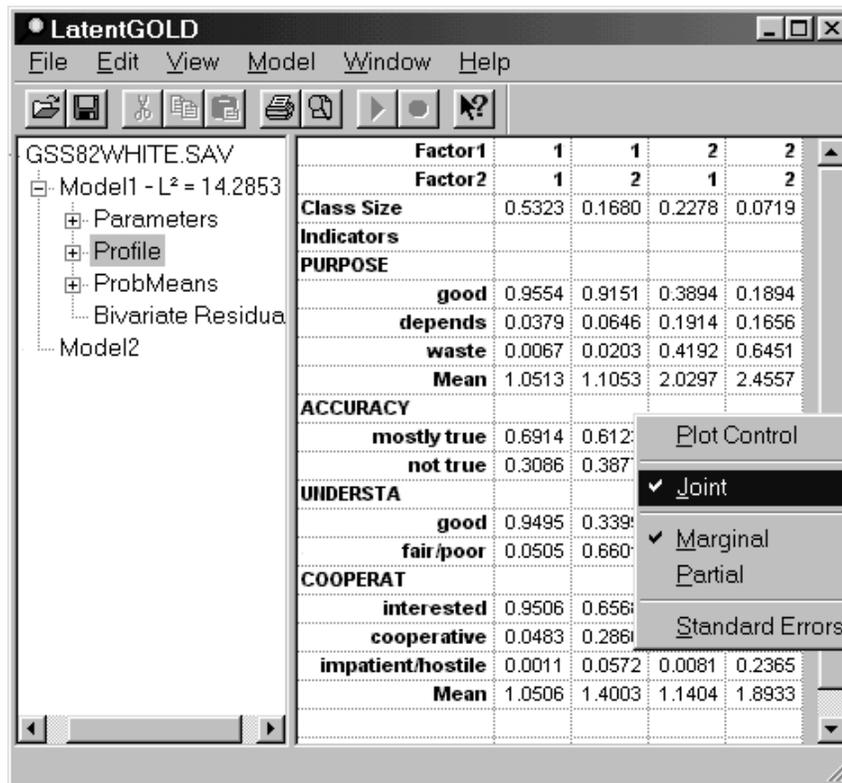
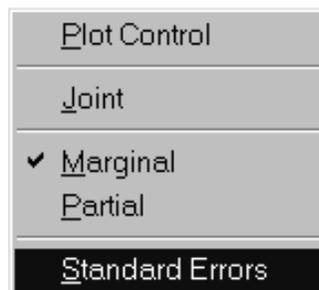


Figure 6-21. Joint Profile Output for DFactor Model



The View Menu options (which can be obtained by a right click on the Profile output), can be used to obtain the Joint Profile Output, to switch between Marginal and Partial Profile Output, or to obtain standard errors.

Partial Profile

Latent GOLD 4.0 reports not only marginal but also partial probabilities/means for indicators in the Profile Table Output. These two differ in models with more than a single DFactor or if local dependencies or effects of covariates on indicators are included using the Residuals Tab. As explained in section 7.3 of the Technical Guide, partial probabilities/means are obtained by conditioning on a certain value (mean) of the other variables involved in the model for the indicator concerned. The marginal probabilities/means, on the other hand, are obtained by collapsing over the categories of these "other" variables. Both have their specific advantages: the Marginal Profile output is somewhat easier to interpret and is displayed by default. Partial Profile reflects somewhat better the strength of the effects of the cluster variable on the indicators.

Joint Profile

A *Joint Profile* amounts to treating each DFactor combination as a cluster. For example, with 3 DFactors having 2 levels each, we have $2 \times 2 \times 2 = 8$ clusters, represented as (1,1,1), (1,1,2), (1,2,1), (1,2,2), (2,1,1), (2,1,2), (2,2,1), and (2,2,2).

The Joint Profile corresponds to the actual model probabilities if no direct effects are included in the model. Joint Profile probabilities can be Partial or Marginal (the default). As in Cluster, these only differ if direct effects are included in a model (direct effects are added to a model using the Residuals Tab - see Chapter 5 - Residuals Tab).

PROFILE PLOT (FOR INDICATORS AND COVARIATES)

To view the Profile Plot, click the '+' icon to expand the Profile output and highlight Prf-Plot. The Profile Plot is constructed from the conditional probabilities for the nominal variables and from the means for the other indicators and covariates as displayed in the columns of the Profile table. Specific columns (DFactor levels) are selected for display using the Plot Control pop-up menu. The quantities associated with the selected columns are plotted and connected to form a line graph.

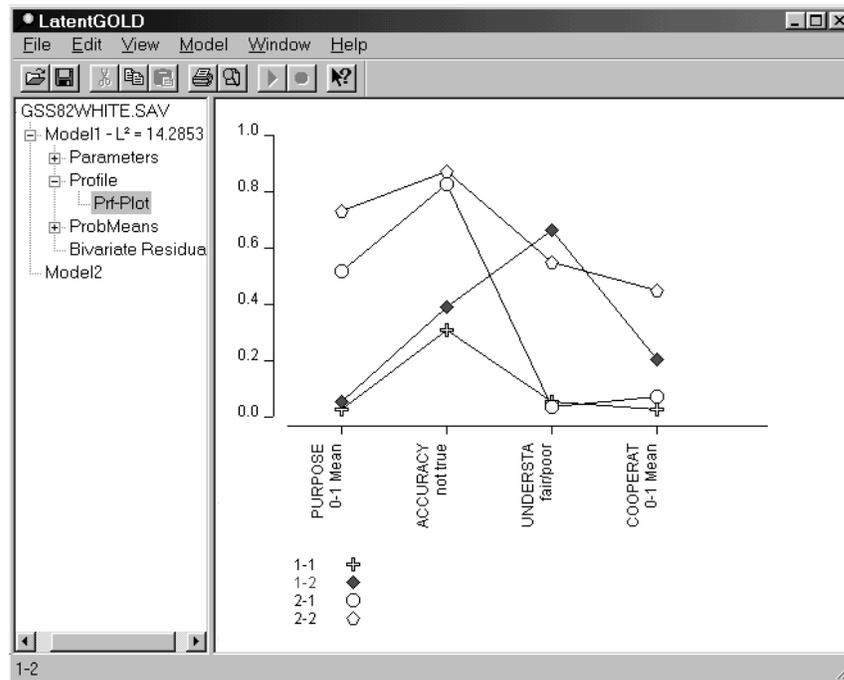


Figure 6-22. Profile Plot for DFactor Model

For the scale types **ordinal**, **continuous**, **count**, and **numeric** covariate, prior to plotting the class-specific means, they are re-scaled to always lie within the 0-1 range. Scaling of these "0-1 Means" is accomplished by subtracting the lowest observed value from the DFactor-level-specific means and dividing the results by the range, which is simply the difference between the highest and the lowest observed value. The advantage of such scaling is that these numbers can be depicted on the same scale as the corresponding probabilities for **nominal** variables.

For nominal variables containing more than 2 categories, all categories are displayed simultaneously. For dichotomous variables specified as nominal, by default only the last category is displayed. The Plot Control can

be used to display the first or both categories, as well as to specify the columns, variables, and selected categories to appear in the display.

By default, all levels associated with DFactor #1 are displayed, or when the Joint Profile table view has been selected, all joint factor levels are displayed.

Each column is associated with either a DFactor level, or a joint DFactor level . When the Profile Table has been changed to the Joint Profile view, the associated probabilities (means) within each selected joint DFactor level displayed graphically in the Profile Plot and connected to form a line graph. the profile information is displayed.

When the contents of the Profile Table is changed from the default view to display Partial (instead of Marginal) probabilities, or to the Joint Profile view, the points plotted change to reflect the current contents of the Profile Table.

- A line is displayed for each level of each selected DFactor (or each joint DFactor level if the Joint Profile has been selected).
- Click on any variable symbol in the Profile Plot and the status bar describes it (variable name, DFactor level number, and value).
- Click on any DFactor level name or symbol in the legend and Latent GOLD highlights all the symbols that refer to that DFactor level.

By default, only the first 8 variables are displayed in the profile plot. Additional variables may be selected from the Plot Control (using a control-click or a shift click to select more than one at a time) and added or removed from the plot with a single click in the box to the left of the variable names.

To Change Settings for a Profile Plot

To change the settings for a Profile Plot, right click (or select Plot Control from the View Menu) within the Contents pane when a Profile Plot is displayed to open the Plot Control dialog box. The appearance of the Profile Plot Control is somewhat different depending upon whether the Joint Profile view has been selected.

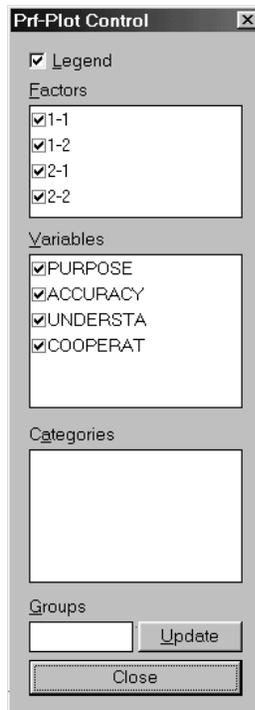


Figure 6-23: Plot Control for Profile Plots

The following plot settings are available for Profile Plots:

Legend. When this option is selected, a legend appears at the bottom of the Profile Plot.

DFactors. For DFactor models, a line will be drawn for each level of each DFactor selected. Those DFactors with a checkmark in the checkbox are included in the plot.

Variables. Select which variables to include in the plot. Those with a checkmark in the checkbox are included in the plot.

Categories. Select which category of a variable to include in the plot. The category currently being plotted is listed in the plot beneath the variable name. To change the category that is plotted, highlight the variable name in the Variables box (the category currently being plotted will appear in the Category box), click the drop-down list to the right of the Categories box and select the category you wish to have plotted.

Groups: Click Update once you have specified a new number of groups.



For more detailed information about the Profile Output, see section 7.3 of Technical Guide.



GPROFILE OUTPUT (ADVANCED, OPTIONAL)

If GClasses and/or GCFactors are included in a model, this output file listing is generated. Similar to the Profile output, the top of this file contains the size of each GClass, followed by the probability of being in each DFactor level for each GClass. This section is followed by the Indicators section where (marginal) means and (marginal) probabilities associated with each indicator are provided.

The View Menu options (which can be obtained by a right click on the *GProfile* output), can be used to switch between *Marginal* and *Partial* GProfile Output, or to obtain standard errors.

PROBMEANS OUTPUT (OPTIONAL)

For DFactor models, this table contains aggregated DFactor means for (ranges of) indicator and covariate values which are displayed in the Uni-Plot and Bi-Plot (when there are 2 or more DFactors).

To view the Probability/Means table for a selected model, click ProbMeans in the Outline pane. To view a plot (not available for Regression models), click on the expand '+' icon to the left of ProbMeans to list the type of plots produced for a particular model. Highlight a plot type to view it in the Contents pane.

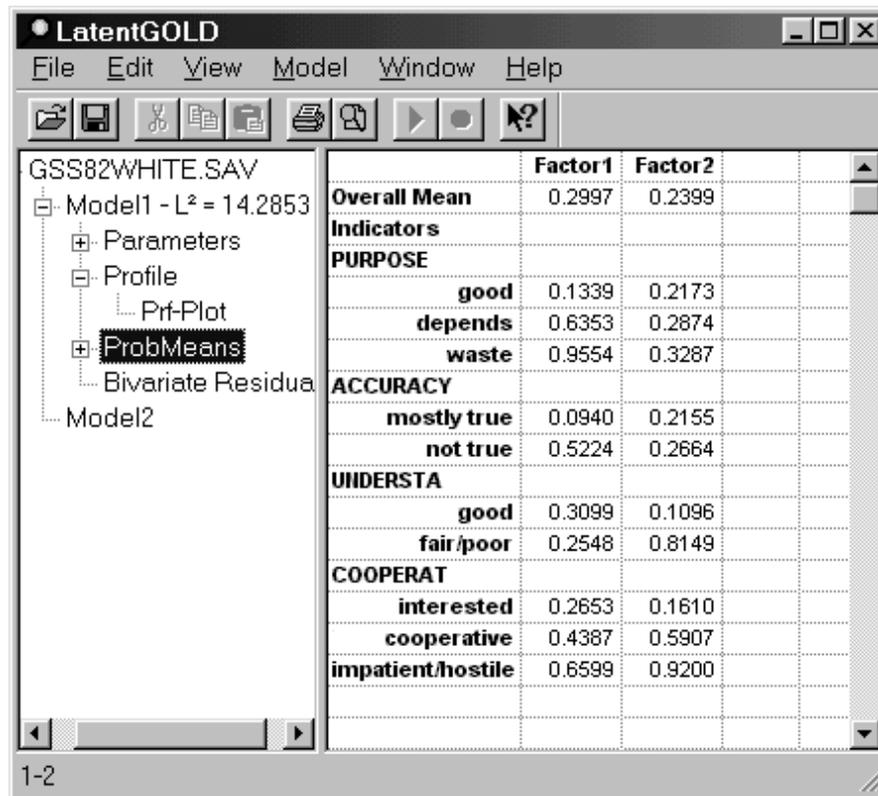


Figure 6-24. ProbMeans Output for DFactor Model

For a DFactor model, the first row contains the mean for each DFactor. For example, if DFactor 1 contained 3 levels, this value is the mean of those three levels computed using equidistant (uniform) scores between 0 and 1 (the first level of the DFactor is scored 0, the last level 1). In the case of a dichotomous DFactor the mean equals the probability of being at level 2. The body of the table contains (partial) DFactor means that show how the indicator variables are related to the DFactors.



When the CFactor option is used, ProbMeans contains aggregated posterior CFactor means.

UNI-PLOT

For a DFactor model, the membership probabilities in the body of the ProbMeans output are plotted to form a Uni-Plot. To view the Uni-Plot, click on the expand/contract icon (+/-) to list the ProbMeans plots and highlight Uni-Plot.

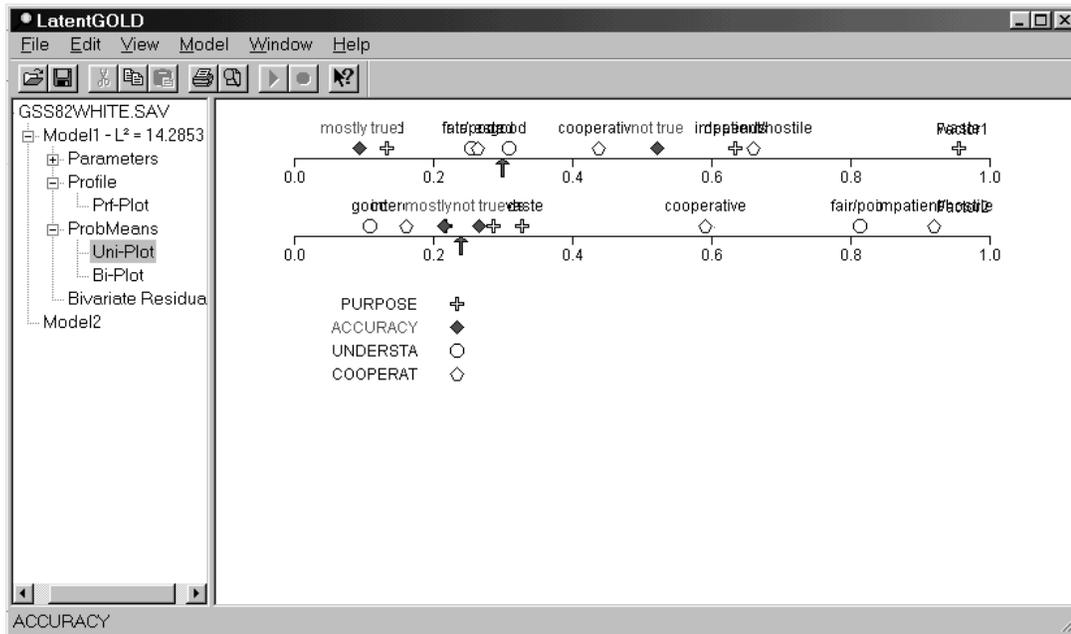


Figure 6-25. ProbMeans Uni-Plot for DFactor Model

The larger the distance (range) between points belonging to a particular variable, the stronger the variable is related to the latent variable.

- By default, a separate Uni-Plot is created for each DFactor. Symbols appear in the plots for each value of each variable specified as an indicator or covariate in the model.
- The ↑ symbol marks the overall mean for a DFactor (corresponding to the first row of the table).
- Click on any variable symbol in the Uni-Plot and the plot label will appear and the status bar will contain a description of the point (variable name and category value).
- Click on any variable name or symbol in the legend and Latent GOLD will highlight all the points that refer to that variable.

To Change Settings for a Uni-Plot

To change the settings for a Uni-Plot, right click (or select Plot Control from the View Menu) within the Contents pane when a Uni-Plot is displayed to open the Plot Control dialog box.

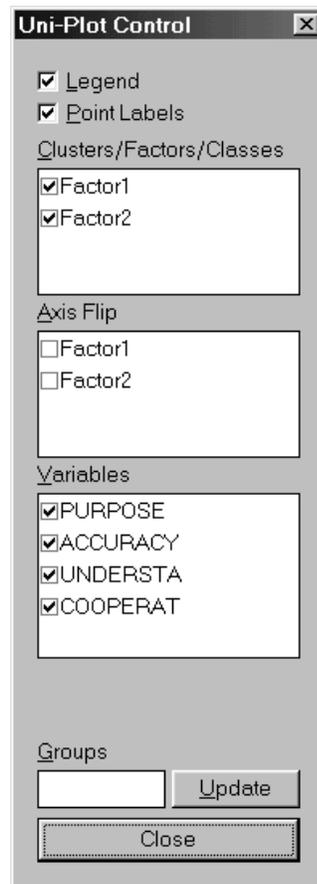


Figure 6-26: Plot Control for Uni-Plots

The following plot settings are available for Uni-Plots:

Legend. When this option is selected, a Legend appears at the bottom of the Uni-Plot.

Point Labels. When this option is selected, category labels for each variable are listed on the Uni-Plot next to the variable symbol.

DFactors. Select which DFactors to include in the Uni-Plots. For each DFactor selected (a checkmark in the checkbox) a Uni-Plot will be displayed. By default, all DFactors are selected.

Axis Flip. To flip (reverse) the axis for a Uni-Plot, select the corresponding DFactor name. By default, the DFactor mean range is from 0 to 1 (increasing). Selecting Axis flip for a DFactor will reverse the axis to range from 1 to 0 (decreasing).

Variables. Select which indicators/covariates to include in the Uni-Plots. Selected variables are indicated by a checkmark in the checkbox. By default, the Uni-Plots contain all the indicators/covariates included in the model.

Groups. Use the grouping option to reduce the number of categories for variable, click Update once you have specified a new number of groups (see Chapter 5, Step 4 for further details on the grouping option).

BI-PLOT

For a 2-or-more DFactor model, the DFactor means for any 2 selected DFactors are plotted to form a Bi-Plot. A symbol in the Bi-Plot is displayed for each category of indicators and covariates for the two DFactors selected. To view the Bi-Plot, click on the expand/contract (+/-) icon to list the ProbMeans plots and highlight Bi-Plot.

- By default, DFactor1 is used as the horizontal axis and DFactor2 as the vertical axis.
- For each DFactor, a reference line representing the overall mean is displayed. The intersection of the reference lines represents the origin of the plot.
- Click on any variable symbol in the Bi-Plot and the status bar will contain a description of the point (variable name and category, values of the DFactor means plotted).
- Click on any variable name or symbol in the legend and Latent GOLD will highlight all the symbols that refer to that variable.

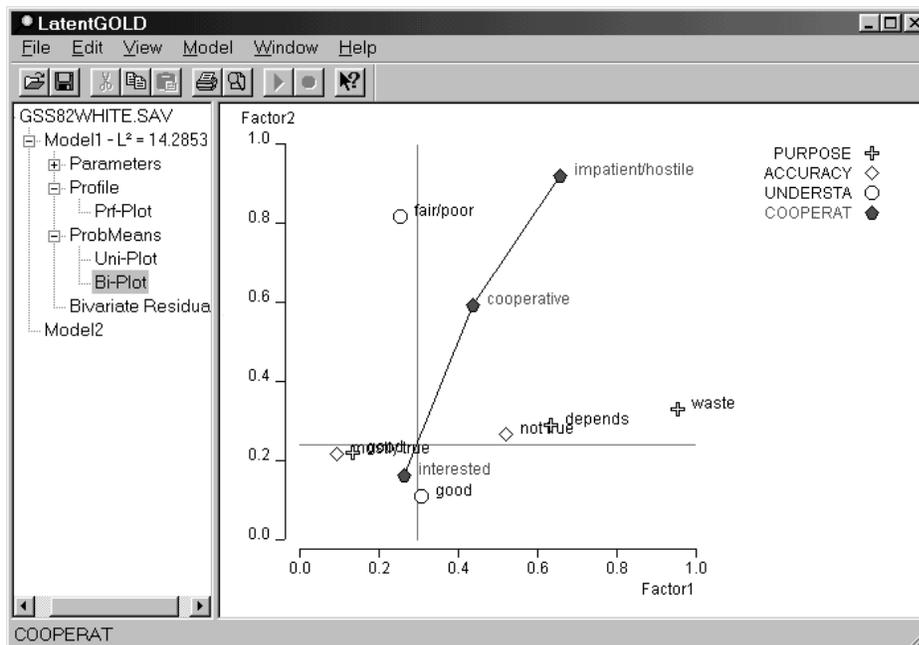


Figure 6-27: Bi-Plot with Lines Option

To Change Settings for a Bi-Plot

To change the settings for a Bi-Plot, right click (or select Plot Control from the View Menu) within the Contents pane when a Bi-Plot is displayed to open the Plot Control dialog box.

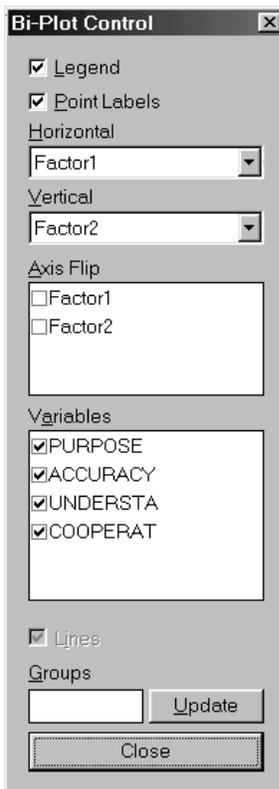


Figure 6-28: Plot Control for Bi-Plots

The following plot settings are available for Bi-Plots:

Legend. When this option is selected, a Legend appears to the right of the Bi-Plot.

Point Labels. When this option is selected, category labels for each variable are listed on the Bi-Plot next to the variable symbol.

Horizontal. Select the DFactor to be used as the horizontal axis. The DFactor currently used as the horizontal axis is listed in the drop-down box. To change the DFactor, click on the downward arrow to the right of the drop-down box. A drop list containing the possible DFactors will appear.

Vertical. Select the DFactor to be used as the vertical axis. The DFactor currently used as the vertical axis is listed in the drop-down box. To change the DFactor, click on the downward arrow to the right of the drop-down box. A drop list containing the possible DFactors will appear.

Axis Flip. To flip either or both axes for a Bi-Plot, select the corresponding DFactor associated with that axis. By default, the DFactor means range from 0 to 1 (increasing in the likelihood of being at a high DFactor level). Selecting Axis Flip for a DFactor (adding a checkmark in the checkbox), will reverse the axis to range from 1 to 0 (decreasing) for that DFactor.

Variables. Select which variables to include in the plot. Those with a checkmark in the checkbox are included in the plot. By default, the Bi-Plots contain all the indicators/covariates that were input as part of the model.

Groups. Click Update once you have specified a new number of groups.

Lines. Highlight the variable(s) and click Lines to suppress (or include, the default) lines that connect the category points for the selected variable(s).



For more detailed information about the ProbMeans Output, see section 7.4 of Technical Guide.

FREQUENCIES AND RESIDUALS (OPTIONAL)

This output appears only if the Frequencies/ Residuals option was selected in the Output Tab before model estimation.

Click Freqs/Residuals to view a table containing the observed and expected frequencies along with the standardized residuals for a model. In addition, for any model in which L^2 has not been calculated (models where the dependent variable has been specified as Continuous or for multilevel models), this output will not be produced even if selected.



For more detailed information about the Frequencies Output, see section 7.5 of Technical Guide.

BIVARIATE RESIDUALS (OPTIONAL)

This output only appears if the Bivariate Residuals option was selected in the Output Tab before model estimation

Click Bivariate Residuals to view a table containing the bivariate residuals for a model. As an advanced option, users may include parameters associated with the bivariate residuals in a model to improve the fit of an estimated model.

In addition to the L^2 criterion, finding no significant residuals is another indication that a model provides a good fit to the data. In general, those that are larger than 3.84 identify correlations between the associated variable pairs that have not been adequately explained by the model. (For 1 degree of freedom effects, bivariate residuals larger than 3.84 indicate statistical significance at the .05 level.



For more detailed information about the Bivariate Residuals Output, see section 7.6 of Technical Guide.

STANDARD CLASSIFICATION (OPTIONAL)

This output only appears if the Standard Classification option was selected in the Output Tab before model estimation

Click Standard Classification to view a table containing posterior DFactor means, as well as posterior DFactor-level membership probabilities and DFactor modal assignments for each DFactor.

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 6-29. Standard Classification Output for a DFactor Model

For DFactor models, each unique data pattern corresponds to a row in the output., Columns consist of the following information associated with the pattern:

- observed frequency (ObsFreq)
- factor (mean) scores (DFactor1, DFactor2, etc) -- assigning 0 to the first level of the DFactor, and 1 to the last, the mean score can be computed, using the probabilities of being in the DFactor levels as weights. For example, in Figure 6-29 above, for cases in the first row, their scores on Dfactor1 and DFactor2 are .0183 and .0763 respectively.
- the modal level for each DFactor (Modal1, Modal2) -- that DFactor level corresponding to the modal probability (the one that is most likely).
- the posterior membership probabilities corresponding to each level of each DFactor. For example, with a dichotomous DFactors, the posterior probabilities for level 2 (labeled 'DFactor1_2' and 'DFactor2_2' in Figure 6-29) are equal to the (mean) score of the corresponding DFactor.
- **A** If 1 or more CFactors have been included in the model, the associated CFactor scores ('Cfac1','Cfac2','Cfac3'), appear in the right-most portion of the standard classification output.
- **A** If 2 or more GClasses have been included in the model, posterior membership probabilities associated these group classes are given in additional columns (GClass1, GClass2, ...) that appear in the right-most portion of the standard classification output.
- **A** If 1 or more GCFactors have been included in the model, the associated factor scores associated with these group level CFactors ('GCfac1','GCfac2','GCfac3'), appear in the right-most portion of the standard classification output.

The classification information shown in the classification output can be appended to your data file using the Standard Classification Output option located in the ClassPred Tab. This option must be set before estimating your model.



For more information on this option, see Step 9 in Chapter 5.

COVARIATE CLASSIFICATION (OPTIONAL)

This output only appears if the Covariate Classification option was selected in the Output Tab before model estimation and 1 or more active covariates have been included in the model.

This table contains the estimated probabilities in the multinomial logistic regression model for each DFactor, $P(x|z)$, as well as the modal cluster assignments for the joint DFactor based on these probabilities. The table has one row for each unique covariate pattern.



If two or more GClasses have been included in the model, estimated group-level class membership probabilities given group-level covariates are provided in additional columns (GClass1, GClass2, ...) that appear in the right-most portion of the covariate classification output.



For more detailed information about the Classification Output, see sections 7.8 and 12.7 of Technical Guide.

Regression

GENERAL INFORMATION

The Regression Module is used to estimate LC Regression models containing 1 or more latent classes .

The appropriate Regression model is estimated according to the dependent variable scale type.

Continuous - Linear regression (with normally distributed residuals)

Dichotomous (specified as nominal, ordinal, or a binomial count) - Binary logistic regression

Nominal (with more than 2 levels) - Multinomial logistic regression

Ordinal (with more than 2 ordered levels) - Adjacent-category ordinal logistic regression

Count - Log-linear Poisson regression

Binomial Count - Binomial logistic regression model

The subtype *censored continuous* yields a tobit regression model. The subtypes *truncated continuous*, *truncated count*, and *truncated binomial count* yield truncated versions of the linear, log-linear Poisson, and binomial logistic regression model, respectively.

In addition to the Model Summary Output file, additional model output files are generated following estimation of the model. The specific output sections that are available are those that appear as active in the Output Sections portion of the Output Tab. For Regression models, these are the same as those available for Cluster and DFactor models with 2 exceptions: Bivariate Residuals output is not available for Regression, Estimated Values is available.

In addition, unlike the Cluster and DFactor models, the Prediction Type section of the Output Tab is also active for Regression models.

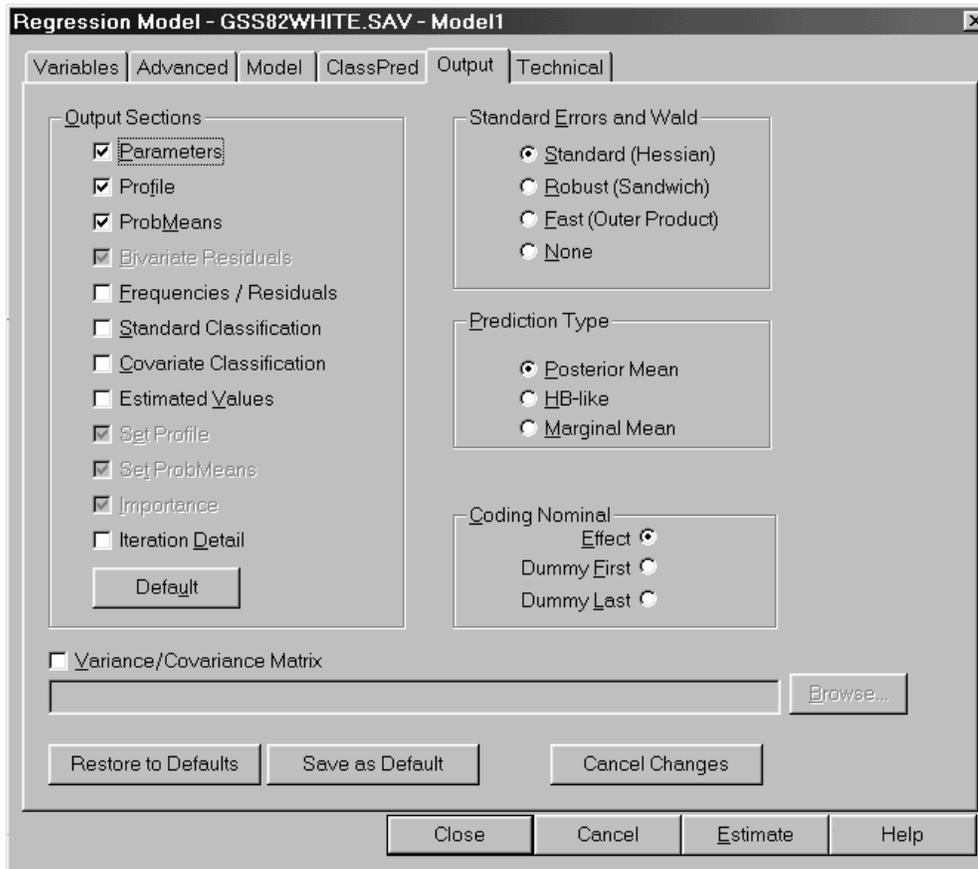


Figure 6-30. Output Tab for Regression Models

Prior to estimating a DFactor model, click in the active check boxes to the left of the Output Sections to include or exclude that Section from being part of the model output files that are generated following the model estimation. A detailed description of each of the Model Summary Output as well as each specific Model Output Section file that can be output appears below.

LC REGRESSION MODEL SUMMARY OUTPUT

By default, Latent GOLD 4.0 reports the log-likelihood (LL), number of parameters (Npar), BIC based on LL, and the proportion of classification errors (Class.Err.) for all models. In addition, if chi-squared statistics are available, we also report the likelihood-ratio statistic (L^2), degrees of freedom (df), and the p-value. The table also contains an overall R^2 based on mean squared error. For more information on these statistics, see the LC Cluster section above.

Specifically, the general section of the Model Summary Output contains the following items:

Type of Model. For example, '3-Cluster Model' indicates that a 3-class Cluster model has been estimated.

Warning messages (generally not present). Estimation Warning message(s) appear here to indicate that boundary solutions, identification or convergence problems were encountered during the estimation of the model.

Model Paused (if you paused the model). If you Paused the model prior to estimation being completed this message appears to alert you that the output files should be viewed as preliminary.

A **Number of groups:** In multilevel latent class models, the program reports the number of groups used in model estimation.

Number of cases. This is the number of cases used in model estimation. This number may be less than the original number of cases on the data file if missing cases have been excluded.

Number of replications. If an ID is used (for repeated measure Regression models) the number of replications (records) is provided.

Number of parameters. This is the number of distinct parameters estimated.

Random Seed. The seed required to reproduce this model.

Best Start Seed. The single best seed that can reproduce this model more quickly using the number of starting sets =0. This is the seed that is automatically inserted in a saved definition (.lgf) file.

A **Design Effect.** When the Survey option is used, the program reports the generalized design effect.

4-Class Regression Model			
Number of cases	400		
Number of replications	3200		
Number of parameters (llpar)	40		
Random Seed	1428079		
Best Start Seed	1428079		
Chi-squared Statistics			
Degrees of freedom (df)	2343704	p-value	
L-squared (L²)	4641.6663	1.00	
X-squared	813959.2758	1.00	
Cressie-Read	76010.1316	1.00	
BIC (based on L²)	-14037577.7		
AIC (based on L²)	-4682766.33		
AIC3 (based on L²)	-7026470.33		
CAIC (based on L²)	-16361281.7		
Dissimilarity Index	0.9878		
Log-likelihood Statistics			
Log-likelihood (LL)	-4025.1174		
Log-prior	-5.6032		
Log-posterior	-4030.7206		
BIC (based on LL)	8269.8934		
AIC (based on LL)	8130.2348		
AIC3 (based on LL)	8170.2348		
CAIC (based on LL)	8329.8934		
Classification Statistics			
	Classes		
Classification errors	0.0989		
Reduction of errors (Lambda)	0.8105		
Entropy R-squared	0.7769		
Standard R-squared	0.7751		
Classification log-likelihood	-4124.3050		
AWE	8847.9272		
Classification Table			
	Modal		
Probabilistic	Class1	Class2	Class3
Class1	180.8498	9.4228	0.28
Class2	13.3871	80.9720	7.37

Figure 6-31. Model Summary Output for Regression Model

Chi-squared Statistics

This section lists various chi-square based statistics related to model fit.

If the scale type for the dependent variable has been set to 'continuous', no chi-squared statistics are available and this section is not displayed.

The information reported:

Degrees of freedom (df). The degrees of freedom for the current model.

L-squared (L^2). The likelihood-ratio goodness-of-fit value for the current model. If the bootstrap p-value for the L^2 statistic has been requested, the results will be displayed here.

X-squared and Cressie-Read. These are alternatives to L^2 that should yield a similar p-value according to large sample theory if the model specified is valid and the data is not sparse.

BIC, AIC and CAIC (based on L^2). In addition to model fit, these statistics take into account the parsimony (df or Npar) of the model. When comparing models, the lower the BIC, AIC and CAIC value the better the model.

Dissimilarity Index. A descriptive measure indicating how much the observed and estimated cell frequencies differ from one another. It indicates the proportion of the sample that needs to be moved to another cell to get a perfect fit.



For more detailed information about Chi-squared statistics including the formulae for each, see section 7.1.1 of Technical Guide.

Log-likelihood Statistics

This section contains additional statistics related to the model fit that are especially useful when L^2 and the other chi-squared statistics are not available. The statistics reported are:

Log-likelihood (LL). If the conditional bootstrap (bootstrap -2LL Diff) has been requested, the results will be displayed here.

Log-prior -- this is the term in the function maximized in the parameter estimation that is associated with the Bayes constants. This term equals 0 if all Bayes constants are set to 0.

Log-posterior - this is the function that is maximized in the parameter estimation. The value of the log-posterior function is obtained as the sum of the log-likelihood and log-prior values.

BIC, AIC, AIC3 and CAIC (based on LL) - these statistics (information criteria) weight fit and parsimony by adjusting the LL to account for the number of parameters in the model. The lower the value, the better the model. For example, according to the BIC values shown in Figure 6-31, the 3-class model is preferred over the 1-class, 2-class and 4-class models.



For more detailed information about Log-likelihood statistics including the formulae for each, see section 7.1.2 of Technical Guide.

Classification Statistics

This information can be used to assess how well the model classifies cases into clusters. The statistics reported are:

Classification Errors. When classification of cases is based on modal assignment (to the class having the highest membership probability), the proportion of cases that are expected to be misclassified is reported by this statistic. The closer this value is to 0 the better.

Reduction of Errors (lambda), Entropy R-squared and Standard R-squared. These statistics indicate how well the model predicts class memberships or DFactor scores. The closer these values are to 1 the better the predictions.

Classification log-likelihood. Log-likelihood value under the assumption that the true class membership is known.

AWE. Similar to BIC, but also takes classification performance into account.

Classification Table. The Classification Table cross-tabulates modal and probabilistic class assignments.

A **Standard R-squared** is reported for CFactors and GCFactors. For GClasses, as for Clusters, Classification Errors, Reduction of Errors, Entropy R-squared, and Standard R-squared values are reported.

Covariate Classification Statistics

If one or more active covariates is included in the model, additional statistics are provided as above but these are now based only on the active covariates. The statistics reported are:

Classification errors, Reduction of Errors (lambda), Entropy R-squared and Standard R-squared

A **Standard R-squared** is reported for CFactors and GCFactors. For GClasses, as for Clusters, Classification Errors, Reduction of Errors, Entropy R-squared, and Standard R-squared values are reported.



For more detailed information about the classification statistics including the formulae for each, see section 7.1.3 of Technical Guide.

Prediction Statistics

The prediction statistics are based on the comparison between observed and predicted responses.

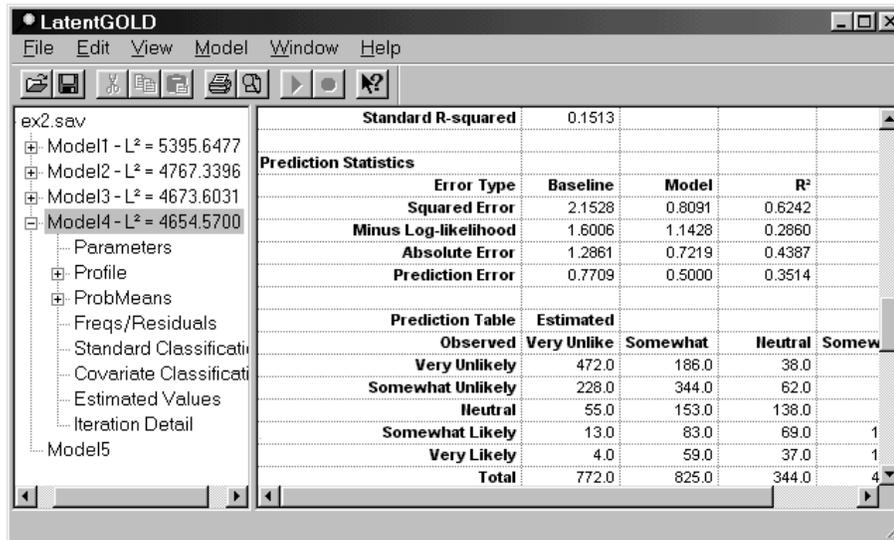


Figure 6-32. Prediction Statistics

This information can be used to assess prediction performance of the model. It is also possible to output predicted values for the **dependent** variable to an external file (*For more information on this option, see the section on the ClassPred Tab in Step 9 of Chapter 5.*)

Prediction Statistics, contains the following measures of Prediction error: **mean squared error** (MSE), **mean absolute error** (MAE), **minus mean log-likelihood** (-MLL), and for ordinal/nominal dependent variables, the proportion of predictions errors under modal prediction (PPE). For each error measure, we provide the prediction error of the baseline, or intercept-only model, the prediction error of the estimated model, and a R² value, which is the proportional reduction of errors in the estimated model compared to the baseline model. For nominal and ordinal dependent variables, a prediction table that cross-classifies observed and against estimated values is also provided.



For more technical information on Prediction Statistics, see section 7.1.4 in the Technical Guide.

Variable Detail This section contains details about the variables entered into the model including variable names, scale types, number of categories and category labels and scores (if used).

PARAMETERS OUTPUT (OPTIONAL)

For any estimated model, click Parameters and Latent GOLD displays a table containing class-specific parameter estimates (betas, gammas, sigmas), class-specific and overall R² values, and measures of significance for these estimates. In addition, the right-most columns contain means and standard deviations for the regression coefficients. In the case of a 1-class model, the mean is identical to the corresponding coefficient and the standard deviation is 0 which indicates that this is a fixed effect. For K>1 class models and unequal coefficients across classes, these are the means and the standard deviations for the (discrete) random effects. The size of the standard deviation indicates the extent of the between-class variation in the coefficient concerned.

Model for Dependent										
	Class1	Class2	Class3	Class4	Overall					
R ²	0.6121	0.6060	0.4699	0.6307	0.6242					
RATING										
	Class1	Class2	Class3	Class4	Wald	p-value	Wald(=)	p-value	Mean	Std.Dev.
Intercept										
Very Unlikely	9.6169	7.8809	8.6659	10.9167	936.7533	3.8e-189	112.0182	2.4e-18	9.0022	0.8446
Somewhat Unlikely	5.6726	4.2513	4.7563	6.7016					5.1332	0.7033
Neutral	0.2288	0.2276	0.7513	0.9928					0.3990	0.2645
Somewhat Likely	-5.1515	-3.5597	-4.4488	-5.4478					-4.5823	0.6756
Very Likely	-10.3667	-8.8001	-9.7247	-13.1633					-9.9521	0.9894
Predictors										
	Class1	Class2	Class3	Class4	Wald	p-value	Wald(=)	p-value	Mean	Std.Dev.
FASHION	1.9916	0.0473	1.1627	2.6712	419.5290	1.7e-89	184.7218	8.4e-40	1.3206	0.8481
QUALITY	-0.0340	2.0416	0.8621	0.9380	176.5660	4.1e-37	127.5144	1.9e-27	0.7754	0.8397
PRICE	1.1484	0.9277	1.0607	-0.0305	448.8721	7.6e-96	10.0969	0.018	1.0071	0.2615
Model for Classes										
	Class1	Class2	Class3	Class4	Wald	p-value				
Intercept	-0.1628	-1.3348	1.1720	0.3256	10.8725	0.012				
Covariates										
	Class1	Class2	Class3	Class4	Wald	p-value				
SEX	0.9493	0.1441	-1.4415	0.3482	23.6178	3.0e-5				
AGE	-0.1763	0.8663	0.7183	-1.4084	40.7094	7.5e-9				

Figure 6-33. Parameters Output for Regression Model

Betas. For Regression models, in addition to the class-specific regression intercept, betas indicate the class-specific effect of each predictor on the dependent variable.

Gammas. Parameters of the multinomial logit model used to predict the latent distribution as a function of the covariates. Parameters include the intercept as well as the effects of each covariate on the latent variable(s).

Sigmas. Error variances for a continuous dependent variable.

The Parameters output reports class-specific R² values and the overall R² value. These are reduction of error measures based on mean squared error. The overall R² indicates how well the dependent variable is overall predicted by the model (same measure as appearing in Prediction Statistics). For ordinal, continuous, and (binomial) counts, these are standard R² measures. For nominal dependent variables, these can be seen as weighted averages of separate R² measures for each category treated as a separate dichotomous response variable. For specific formulae for the overall and class-specific R², see section 7.2 of Technical Guide.



If CFactors have been included in the model, estimates and related information for additional parameters (lambdas) are included in the Model for Dependent output.



If GClasses and/or GCFactors have been included in the model, estimates and related information for additional parameters are included in the Model for Dependent and/or the Model for Classes output, and a separate section called Model for GClasses appears at the bottom of the listing.

Viewing Wald Statistics and Standard Errors

By default, Wald statistics are provided in the output to assess the statistical significance of the set of parameter estimates associated with a given variable. Specifically, for each variable, the Wald statistic tests the restriction that each of the parameter estimates in that set equals zero (for variables specified as Nominal, the set includes parameters for each category of the variable). For Regression models, by default, two Wald statistics (Wald, Wald(=)) are provided in the table when more than 1 class has been estimated. For each set of parameter estimates, the Wald(=) statistic considers the subset associated with each class and tests the restriction that each parameter in that subset equals the corresponding parameter in the subsets associated with each of the other classes. That is, the Wald(=) statistic tests the equality of each set of regression effects across classes.

To view standard errors or related statistics associated with parameter estimates, simply right click on the parameters output and select/deselect the items that you want to appear. Alternatively, you can click on the appropriate item in the View Menu. In Parameters, you can alter the output view to obtain standard errors, a Z Statistic, and can suppress the Wald statistics.



For more general information about the Parameters Output, see section 7.2 of Technical Guide.

A PARAMETERS OUTPUT SUBCATEGORY (ADVANCED)

If 1 or more CFactors have been included in the model, clicking on the + to the left of Parameters makes visible the Parameters Output Subcategory called Random Effects.

Random Effects. This table reports the CFactor effects in the first columns (labeled 'CFactor1', 'CFactor2', etc.) and provides the variance-covariance matrix for the continuous random effects in the columns to the right of these CFactor effects (labeled 'eff1', 'eff2', etc.). For further information see Section 9.2.4 of Technical Guide.

PROFILE OUTPUT (OPTIONAL)

To view the profile table for a selected model, click Profile in the Outline pane. The Profile table contains probabilities or means associated with the Dependent variable and covariates.

	Class1	Class2	Class3	Class4
Class Size	0.4405	0.2577	0.2487	0.0532
Dependent				
RATING				
Very Unlikely	0.2838	0.1796	0.1778	0.2553
Somewhat Unlikely	0.2821	0.1466	0.1818	0.2697
Neutral	0.1239	0.1240	0.2641	0.1942
Somewhat Likely	0.1117	0.2696	0.1859	0.1966
Very Likely	0.1986	0.2802	0.1904	0.0842
Mean	2.6592	3.3242	3.0292	2.5848
Covariates				
SEX				
Male	0.2463	0.4120	0.7762	0.3727
Female	0.7537	0.5880	0.2238	0.6273
Mean	1.7537	1.5880	1.2238	1.6273
AGE				
16-24	0.6417	0.2615	0.2617	0.8944
25-39	0.1750	0.1950	0.3216	0.0964
40+	0.1832	0.5435	0.4167	0.0092
Mean	1.5415	2.2819	2.1551	1.1147

Figure 6-34. Profile Output for Regression Model

- For a Regression model, the first row of numbers shows how large each class is. The body of the table contains the (marginal) conditional probabilities that show how the classes are related to the dependent variable, specified as Nominal, Ordinal or Binomial Count. The probabilities within each class (column) sum to 1. For dependent variables specified as Ordinal and Continuous, the body of the table contains means, and for Counts these are rates.
- For covariates the Profile output contains rescaled ProbMeans output rather than model probabilities or means. These are aggregate class membership probabilities which are rescaled to sum to one within clusters. For Covariates specified as Numeric, means are displayed in addition to the conditional probabilities, and if the Covariate contains more than 5 distinct values, the probabilities are displayed for 5 grouped ranges. The number of such grouped ranges may be changed (increased or decreased) using Groups option in the Plot Control.

Viewing Standard Errors

To view the associated standard errors in a Profile table, simply right click on the parameters output to retrieve the pop-up menu and click Standard Errors. Column containing the standard errors appear to the right of each estimate.

Partial Profile

Latent GOLD 4.0 reports not only marginal but also partial probabilities/means for the dependent variable in the Profile Table Output. These two differ if predictors have been included in the model. As explained in section 7.3 of the Technical Guide, partial probabilities/means are obtained by conditioning on a certain value (mean) of the

predictors. The marginal probabilities/means, on the other hand, are obtained by aggregating over the categories of these predictors. Both have their specific advantages: the Marginal Profile output is somewhat easier to interpret and is displayed by default.

By a right click on the Profile output, you can retrieve a pop-up menu that can be used to switch between Partial to Marginal and to obtain standard errors. (The partial/marginal switch has no effect on covariate information).

PROFILE PLOT (FOR DEPENDENT AND COVARIATES)

To view the Profile Plot, click the '+' icon to expand the Profile output and highlight Prf-Plot. The Profile Plot is constructed from the conditional probabilities for the nominal variables and means for the other indicators and covariates as displayed in the columns of the Profile Table. Specific classes (columns) are selected for display using the Plot Control pop-up menu. The quantities associated with the selected clusters are plotted and connected to form a line graph.

For the scale types ordinal, continuous, count, and numeric covariate, prior to plotting the class-specific means, they are re-scaled to always lie within the 0-1 range. Scaling of these "0-1 Means" is accomplished by subtracting the lowest observed value from the class-specific means and dividing the results by the range, which is simply the difference between the highest and the lowest observed value. The advantage of such scaling is that these numbers can be depicted on the same scale as the class-specific probabilities for nominal variables.

For nominal variables containing more than 2 categories, all categories are displayed simultaneously. For dichotomous variables specified as nominal, by default only the last category is displayed. The Plot Control can be used to display the first or both categories, as well as to specify the columns, variables, and selected categories to appear in the display.

When the contents of the Profile Table is changed from the default view to display Partial (instead of Marginal) probabilities, the points plotted change to reflect the current contents of the Profile Table.

- A separate line is displayed for each class.
- Click on any variable symbol in the Profile Plot and the status bar describes it (variable name, Class number, and value).
- Click on any Class name or symbol in the legend and Latent GOLD highlights all the symbols that refer to that Class.

By default, only the first 8 variables are displayed in the profile plot. Additional variables may be selected from the Plot Control (using a control-click or a shift click to select more than one at a time) and added or removed from the plot with a single click in the box to the left of the variable names.

To Change Settings for a Profile Plot

To change the settings for a Profile Plot, right click (or select Plot Control from the Model Menu) within the Contents pane when a Profile Plot is displayed to open the Plot Control dialog box. To change the font type/size for a plot, see Chapter 2.

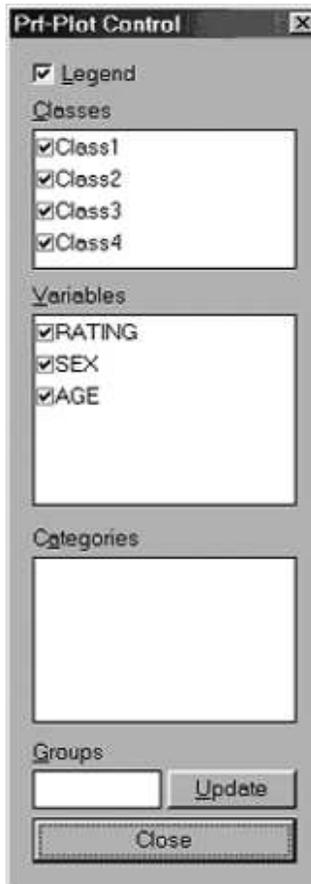


Figure 6-35: Plot Control for Profile Plots

The following plot settings are available for Profile Plots:

Legend. When this option is selected, a legend appears at the bottom of the Profile Plot.

Classes. For Regression models, a line will be drawn for each class selected. Those classes with a checkmark in the checkbox are included in the plot.

Variables. Select which variables to include in the plot. Those with a checkmark in the checkbox are included in the plot.

Categories. Select which category of a variable to include in the plot. The category currently being plotted is listed in the plot beneath the variable name. To change the category that is plotted, highlight the variable name in the Variables box (the category currently being plotted will appear in the Category box), click the drop-down list to the right of the Categories box and select the category you wish to have plotted.

Groups: Click Update once you have specified a new number of groups.



For more detailed information about the Profile Output, see section 7.3 of Technical Guide.

A GPROFILE OUTPUT (ADVANCED, OPTIONAL)

If GClasses and/or GCFactors are included in a model, this output file listing is generated. Similar to the Profile output, the top of this file contains the size of each GClass, followed by the probability of being in each Class for each GClass. This section is followed by the Dependent section where means and (marginal) probabilities associated with the dependent variable are provided.

The View Menu options (which can be obtained by a right click on the GProfile output), can be used to switch between Marginal and Partial GProfile Output, or to obtain standard errors.

PROBMEANS OUTPUT (OPTIONAL)

For Regression models, this table contains aggregated class membership probabilities for (ranges of) dependent and covariates values which are displayed in the Uni-Plot and Tri-Plot.

To view the Probability/Means table for a selected model, click ProbMeans in the Outline pane. To view a plot, click on the expand '+' icon to the left of ProbMeans to list the type of plots produced for a particular model. Highlight a plot type to view it in the Contents pane.

For Regression models, the first row of the table contains the overall probability of being in a class (the size of each class, as also reported as the first row of numbers in the Profile Table). The body of the table contains conditional class probabilities associated with each category of categorical dependent variables (these probabilities will sum to 100% across rows).

UNI-PLOT

For a Regression model, the membership probabilities in the body of the ProbMeans output are plotted to form a Uni-Plot. To view the Uni-Plot, click on the expand/contract icon (+/-) to list the ProbMeans plots and highlight Uni-Plot.

The larger the distance (range) between points belonging to a particular variable, the stronger the variable is related to the latent variable.

- By default, a separate Uni-Plot is created for each class. Symbols appear in the plots for each value of each variable specified as an indicator or covariate in the model.
- The ↑ symbol marks the overall probability (the size) for a class.
- Click on any variable symbol in the Uni-Plot and the plot label will appear and the status bar will contain a description of the point (variable name and category value).
- Click on any variable name or symbol in the legend and Latent GOLD will highlight all the points that refer to that variable.

To Change Settings for a Uni-Plot

To change the settings for a Uni-Plot, right click (or select Plot Control from the View Menu) within the Contents pane when a Uni-Plot is displayed to open the Plot Control dialog box. To change the font for a plot, see Chapter 2.

The following plot settings are available for Uni-Plots:

Legend. When this option is selected, a Legend appears at the bottom of the Uni-Plot.

Point Labels. When this option is selected, category labels for each variable are listed on the Uni-Plot next to the variable symbol.

Classes. Select which Classes to include in the Uni-Plots. For each class selected (a checkmark in the checkbox) a Uni-Plot will be displayed. By default, all classes are selected.

Axis Flip. To flip (reverse) the axis for a Uni-Plot, select the corresponding class name. By default, the class probabilities range is from 0 to 1 (increasing). Selecting Axis flip for a class will reverse the axis to range from 1 to 0 (decreasing).

Variables. Select which indicators/covariates to include in the Uni-Plots. Selected variables are indicated by a checkmark in the checkbox. By default, the Uni-Plots contains all the indicators/covariates included in the model.

Groups. Use the grouping option to reduce the number of categories for a variable, click Update once you have specified a new number of groups (see Chapter 5, Step 4 for further details on the grouping option).

TRI-PLOT

For Regression models, the class membership probabilities in the body of the ProbMeans output table are plotted to form a Tri-Plot. To view the Tri-Plot, click on the expand/contract icon (+/-) to list the ProbMeans plots and highlight Tri-Plot. (Note: No Tri-Plot is produced for a 1-class model; for a 2-class model, the Tri-Plot reduces to the Uni-Plot.)

- By default, Vertex A (left-most base vertex) is labeled 'Class 1', Vertex B (right-most base vertex) 'Class 2', and the third Vertex (the top point of the triangle) represents the aggregate of all other classes. For a 3-class Regression model, by default, the third vertex will represent Class 3 and is labeled 'Class 3'. For a 4-or-more class Regression model, the third vertex is labeled 'Others'. For a 2-class Regression model, the class 3 membership probability is 0 and the Tri-Plot reduces to the Uni-Plot.
- The ▲ symbol marks the overall probabilities for the 3 classes associated with the vertices. It represents the centroid of the triangle.
- Click on any variable symbol in the Tri-Plot and 1) the status bar will contain a description of the point (variable name and category, class probabilities) 2) the category label will appear next to that point on the plot and 3) lines emanate from that point to each side of the triangle, intersecting the side at the corresponding class probabilities value.
- Click on any variable symbol or name in the legend and all the symbols for that variable will be highlighted and their category labels listed in the Tri-Plot.

To Change Settings for a Tri-Plot

To change the settings for a Tri-Plot, right click (or select Plot Control from the View Menu) within the Contents pane when a Tri-Plot is displayed to open the Plot Control dialog box. To change the font for a plot, see Chapter 2.

The following plot settings are available for Tri-Plots:

Legend. When this option is selected, a Legend appears to the right of the Tri-Plot.

Point Labels. When this option is selected, category labels for each variable are listed on the Tri-Plot next to the variable symbol.

Vertices. Latent GOLD allows you to select the base vertices in the Tri-Plot. The top vertex corresponds to the aggregate of the remaining classes.

A vertex. The class currently used as the A vertex is listed in the drop down box. To select a different class, click on the down arrow to the right of the vertex box. A drop list containing all classes will appear. Select the class to use as the A vertex.

B vertex. The class currently used as the B vertex is listed in the drop down box. To select a different class, click on the down arrow to the right of the vertex box. A drop list containing all classes will appear. Select the class to use as the B vertex.

Variables. Select which variables to include in the Tri-Plot. Those with a checkmark in the checkbox are included in the plot. By default, the Tri-Plot contains all the indicators/covariates that were input as part of the model.

Groups. Click Update once you have specified a new number of groups.



For more detailed information about the ProbMeans Output, see section 7.4 of Technical Guide.

FREQUENCIES AND RESIDUALS (OPTIONAL)

This output appears only if the Frequencies/ Residuals option was selected in the Output Tab before model estimation.

Click Freqs/Residuals to view a table containing the observed and expected frequencies along with the standardized residuals for a model. In addition, for any model in which L^2 has not been calculated (models where a dependent variable has been specified as Continuous or in multilevel models), this output will not be produced even if selected.



For more detailed information about the Frequencies Output, see section 7.5 of Technical Guide.

STANDARD CLASSIFICATION (OPTIONAL)

This output only appears if the Standard Classification Information option was selected in the Output Tab before model estimation.

Click Standard Classification to view a table containing the posterior membership probabilities and other classification information for a model.

ID	SEX	AGE	ObsFreq	Modal	Class1	Class2	Class3	Class4
1	Male	25-39	1.0000	2	0.0002	0.7094	0.2905	0.0000
2	Female	16-24	1.0000	2	0.0000	0.9896	0.0103	0.0000
3	Male	40+	1.0000	2	0.0000	0.9188	0.0812	0.0000
4	Female	40+	1.0000	2	0.0000	0.9939	0.0061	0.0000
5	Female	40+	1.0000	2	0.0003	0.9241	0.0757	0.0000
6	Male	16-24	1.0000	3	0.0283	0.1938	0.7777	0.0002
7	Female	25-39	1.0000	2	0.0000	0.9994	0.0006	0.0000
8	Male	16-24	1.0000	3	0.0089	0.4206	0.5704	0.0000
9	Female	25-39	1.0000	2	0.0006	0.9029	0.0964	0.0000
10	Male	25-39	1.0000	2	0.0000	0.9753	0.0247	0.0000
11	Male	40+	1.0000	2	0.0006	0.5416	0.4578	0.0000
12	Male	40+	1.0000	2	0.0001	0.9062	0.0937	0.0000
13	Male	40+	1.0000	2	0.0000	0.9165	0.0835	0.0000
14	Female	40+	1.0000	2	0.0035	0.8557	0.1408	0.0000
15	Female	40+	1.0000	2	0.0002	0.9305	0.0693	0.0000
16	Female	16-24	1.0000	2	0.1079	0.5093	0.3559	0.0268
17	Male	40+	1.0000	2	0.0002	0.5323	0.4674	0.0000
18	Male	40+	1.0000	2	0.0000	0.9040	0.0960	0.0000
19	Female	40+	1.0000	2	0.0000	0.9822	0.0178	0.0000
20	Female	25-39	1.0000	2	0.0000	0.9697	0.0302	0.0000
21	Male	40+	1.0000	3	0.0003	0.3800	0.6197	0.0000
22	Female	25-39	1.0000	3	0.0105	0.4665	0.5230	0.0000
23	Female	40+	1.0000	2	0.0000	0.9994	0.0006	0.0000

Figure 6-36. Standard Classification Output for Regression Model

Each row contains grouped cases corresponding to responses to covariates, predictors, and dependent (in that order). If an ID variable is specified, it is included in the first column. If an ID is specified, a frequency count of 2 or more means that some cases have been combined (included in the same row of the classification output) because their data response patterns are identical. In this case, the ID of the first case only is displayed. (An output data file consisting of separate records for each case can be obtained using the ClassPred Tab. This option must be set before estimating your model. For more information on this option, see the section on the ClassPred Tab in Step 9 of Chapter 5).

For each data response pattern, the classification output contains separate columns for frequencies ('ObsFreq'), modal class ('Modal'), posterior probabilities of belonging to each class ('Class 1', 'Class 2', ...), and CFactor scores ('Cfac1', 'Cfac2', 'Cfac3'). The classification variable 'Modal' contains that class number associated with the modal class.



If 1 or more CFactors have been included in the model, the associated CFactor scores ('Cfac1','Cfac2','Cfac3'), appear in the right-most portion of the standard classification output.

A

If 2 or more GClasses have been included in the model, posterior membership probabilities associated these group classes are given in additional columns (GClass1, GClass2, ...) that appear in the right-most portion of the standard classification output.

A

If 1 or more GCFactors have been included in the model, the associated factor scores associated with these group level CFactors ('GCfac1','GCfac2','GCfac3'), appear in the right-most portion of the standard classification output.

COVARIATE CLASSIFICATION (OPTIONAL)

This output only appears if the Covariate Classification option was selected in the Output Tab before model estimation and if 1 or more active covariates have been included in the model.

SEX	AGE	Modal	Class1	Class2	Class3	Class4
Male	16-24	1	0.3993	0.1568	0.3398	0.1041
Male	25-39	3	0.2341	0.2608	0.4873	0.0178
Male	40+	3	0.1078	0.3407	0.5491	0.0024
Female	16-24	1	0.7162	0.1257	0.0558	0.1023
Female	25-39	1	0.5780	0.2878	0.1101	0.0241
Female	40+	2	0.3459	0.4886	0.1613	0.0042

Figure 6-37. Covariate Classification Output for Regression Model

This table contains the estimated probabilities in the multinomial logistic regression model for the classes, $P(x|z)$, as well as the modal cluster assignments based on these probabilities. The table has one row for each unique covariate pattern.

A

If 2 or more GClasses have been included in the model, estimated group-level class membership probabilities given group-level covariates are provided in additional columns (GClass1, GClass2, ...) that appear in the right-most portion of the covariate classification output.



For more detailed information about the Classification Output, see sections 7.8 and 12.7 of Technical Guide.

ESTIMATED VALUES (OPTIONAL)

This output only appears if the Estimated Values option was selected in the Output Tab before model estimation.

The Estimated Values output section reports the estimated class-specific, the estimated overall, and the observed means (probabilities) for each unique predictor pattern.

FASHION	QUALITY	PRICE	RATIOG	Class1	Class2	Class3	Class4	Overall	Observed
Traditional	Low	Lower	Very Unlikely	0.6761	0.5431	0.6109	0.6267	0.6230	0.6050
Traditional	Low	Lower	Somewhat Unlikely	0.2924	0.2942	0.2679	0.3316	0.2889	0.2975
Traditional	Low	Lower	Neutral	0.0282	0.1075	0.1069	0.0394	0.0688	0.0725
Traditional	Low	Lower	Somewhat Likely	0.0029	0.0497	0.0129	0.0023	0.0174	0.0200
Traditional	Low	Lower	Very Likely	0.0004	0.0054	0.0014	0.0000	0.0019	0.0050
Traditional	Low	Lower	Mean	1.3590	1.6800	1.5261	1.4174	1.4863	1.5225
Traditional	Low	Higher	Very Unlikely	0.3373	0.1811	0.2273	0.6346	0.2855	0.2625
Traditional	Low	Higher	Somewhat Unlikely	0.4599	0.2481	0.2880	0.3257	0.3555	0.3750
Traditional	Low	Higher	Neutral	0.1400	0.2292	0.3317	0.0375	0.2052	0.2225
Traditional	Low	Higher	Somewhat Likely	0.0454	0.2682	0.1156	0.0021	0.1180	0.0975
Traditional	Low	Higher	Very Likely	0.0174	0.0734	0.0374	0.0000	0.0359	0.0425
Traditional	Low	Higher	Mean	1.9457	2.8046	2.4477	1.4072	2.2632	2.2825
Traditional	High	Lower	Very Unlikely	0.6843	0.0107	0.2963	0.3540	0.3967	0.4225
Traditional	High	Lower	Somewhat Unlikely	0.2860	0.0446	0.3078	0.4786	0.2395	0.2250
Traditional	High	Lower	Neutral	0.0267	0.1254	0.2907	0.1453	0.1241	0.1175
Traditional	High	Lower	Somewhat Likely	0.0027	0.4469	0.0831	0.0212	0.1381	0.1300
Traditional	High	Lower	Very Likely	0.0003	0.3725	0.0220	0.0009	0.1016	0.1050
Traditional	High	Lower	Mean	1.3486	4.1260	2.2266	1.8363	2.3086	2.2700
Traditional	High	Higher	Very Unlikely	0.3482	0.0005	0.0415	0.3631	0.1831	0.1950
Traditional	High	Higher	Somewhat Unlikely	0.4589	0.0048	0.1244	0.4761	0.2596	0.2475
Traditional	High	Higher	Neutral	0.1350	0.0343	0.3394	0.1402	0.1602	0.1650
Traditional	High	Higher	Somewhat Likely	0.0423	0.3090	0.2802	0.0199	0.1690	0.1650
Traditional	High	Higher	Very Likely	0.0157	0.6514	0.2144	0.0008	0.2281	0.2275
Traditional	High	Higher	Mean	1.9185	4.6061	3.5017	1.8192	2.9995	2.9825

Figure 6-38. Estimated Values Output for Regression Model



For further information about Estimated Values, see section 7.7 of Technical Guide.