Step 3 Tutorial #2: Obtaining equations for scoring new cases in a basic example with main affects

DemoData = 'data3.sav' (or 'data3_copy.sav')

In this tutorial, we will show how to use the Step3 module in Latent GOLD to obtain equations for a scoring algorithm and related SPSS syntax for scoring new cases as a function of the indicators used to develop the latent classes. That is, the equations show how the indicators can be used to *predict* latent class membership.

Obtaining scoring equations from Step3 Module

Recall the output file, data3.sav, created in Latent GOLD Tutorial 1.

- ➢ Open data3.sav using "File→Open" ('data3_copy.sav' works with the LG demo version)
- From the menu bar, click 'Model' and select 'Step3'.

The following dialog box appears:

/ariables	Advanced	Model	Output	Technical		
burpose accuraci underst	/		Post	eriors>		
coopera du#1 du#2 du#3	t		Cova	riates>		
			Case	Neiaht>	frq	
				Covariates Dependent Scoring	Method Proportional Modal	Adjustment None BCH I ML
	al Order		Selec	tion ection Variab	le>	
		eset (

Figure 1. Variables Tab of Step3 Module

- For Type, select 'Scoring'
- Click on the 4 indicators and click 'Covariates' to move them to the Covariates box
 Right click on the 4 predictors and select 'Nominal'
- Click on clu#1, clu#2, clu#3 and click 'Posteriors' to move them to the Posteriors box
- Click 'Scan'

5	Step3 - data3.sa	v - Model2		×
Variables Advanced Model	Output Technical			
clu# .	<posteriors< td=""><td>du#1 du#2 du#3</td><td></td><td>33 33 33</td></posteriors<>	du#1 du#2 du#3		33 33 33
	Variables>	purpose accuracy understa cooperat	Nominal Nominal Nominal	3 2 2 3
	Case Weight>	frq	21	
	Select>			
	Analysis Covariates Dependent	Classification Proportional Modal	Adjustment ML BCH	
Lexical Order	Scoring		None	
Scan Reset				
	Close	Cancel Est	imate He	lp

Figure 2. Step 3 Variables Tab: Selecting the variables and the analysis type

Your screen should now look like Figure 2.

Click on the 'Technical Tab' and set all Bayes constants to '0'.

T	Fechnical	
	Bayes Constants	
	Latent Variables	0
	Categorical Variables	0
	Poisson Counts	0
	Error Variances	o

Click 'Estimate'

Confirm that the $L^2 = 0$ (Figure 3), which means that the posterior probabilities are reproduced perfectly as a function of the 4 predictors. The Profile and Probmeans output obtained here also reproduce the Profile and Probmeans output from the original 3-class model developed in Latent GOLD Tutorial 1. (Note that since we used the default Bayes=1 technical setting in the original 3-class model, the parameters output may differ in the 4th decimal place).

Q											
File	Edit	View	Mod	lel	Win	dow	Help)			
2	H	ХÞ	ß	9	C)		•	\?			
d	lata3.s	av				STEP:	3-Sco	ring 3-	Cluster Model		
	- Mo	del1 - L	$^{2} = 0.0$	0000							
	· · · · ·	Parame	ters						Number of cases	1202	
		Drofile					Nu	mber	of parameters (Npar)	14	
	±								Robustness Effect	1.0008	
	±	ProbIMe	eans						Random Seed	32116	
		Estimat	edVal	ues-l	Mc				Best Start Seed	32116	

Figure 3. Step 3 Model output

The equations are provided in the Parameters output.

Click Parameters to view the coefficients in the Parameters output (Figure 4)

Note: Ignore the other statistics as they are not relevant for the scoring procedure

٥							
File Edit View Model Wi	ndow Help						
🖻 🖬 🐇 🖿 🗃 🚭 🕿) 🕨 🛛 📢						
⊡. data3.sav	Model for Classes						
	Intercept	Cluster1	Cluster2	Cluster3	Wald	p-value	
Parameters		-2.0479	0.9315	1.1164	5.4249	0.066	
Drofile							
Profile ProbMeans EstimatedValues-Mc	Variables	Cluster1	Cluster2	Cluster3	Wald	p-value	
	purpose						
	good	0.6762	1.0709	-1.7471	212.3499	8.3e-45	
Model2	depends	-0.4683	0.1954	0.2729			
	waste	-0.2079	-1.2664	1.4742			
	accuracy						
	mostly true	0.5692	0.6510	-1.2202	78.6256	8.4e-18	
	not true	-0.5692	-0.6510	1.2202			
	understa						
	good	1.7524	-1.3387	-0.4138	126.0971	4.2e-28	
	fair/poor	-1.7524	1.3387	0.4138			
	cooperat						
	interested	1.9660	-0.8573	-1.1087	55.3125	2.8e-11	
	cooperative	0.7000	-0.2535	-0.4466			
	impatient/hostile	-2.6660	1.1108	1.5552			

Figure 4. Step 3 Parameters output

Each response pattern is scored on each cluster, and is assigned to the cluster with the highest score. For example, the response pattern:

Purpose = Good, Accuracy = Mostly True, Understanding = good, Cooperation = Interested is scored based on the coefficients highlighted above. This results in the following scores:

Cluster 1 score = 2.916, Cluster 2 score = 0.457, Cluster 3 score = -3.373.

Thus, this response pattern is assigned to Cluster 1. To obtain more meaningful scores, we can translate these logit scores to posterior membership probabilities using the formula provided below. This yields the following posterior membership probabilities associated with this response pattern:

Posterior 1 = 0.9196, Posterior 2 = 0.0787, Posterior 3 = 0.0017

Here is the formula to obtain the posterior probabilities:

Posterior(k) = $\exp[\text{score}(k)]/[\exp(\text{score}1) + \exp(\text{score}2) + \exp(\text{score}3)], k = 1, 2, 3.$

It is also possible to obtain SPSS syntax equations which compute the posteriors:

- Double click on Model 2
- Click on the Output tab
- At the bottom of the tab, check the box 'Scoring Syntax'.
 By default, the syntax type will be SPSS syntax (see Figure 5).

Variables Advanced Model Outpu	It Technical
Output Sections Parameters Profile ProbMeans Bivariate Residuals Frequencies / Residuals Classification - Posterior Classification - Model Estimated Values Set Profile Set ProbMeans Importance Iteration Detail Default	Standard Errors and Wald Standard (Hessian) Robust (Sandwich) Fast (Outer Product) None Prediction Type Posterior (EB) HB-like (EB) Model (Marginal) Coding Nominal Effect Dummy First Dummy Last
Scoring Syntax C: \LatentGOLD5. 1\DemoData\Score Variance/Covariance Matrix Restore to Defaults Save as De	Type SPSS V 23.sps Browse Browse efault Cancel Changes

Figure 5. Step 3 Output tab: Requesting SPSS Syntax

Using 1-Step Scoring to obtain Scoring Equations from Cluster Module

In addition to providing the scoring equations, the Step 3 module can also be used to relate the latent classes to other variables not used in the latent class model. A faster way to obtain these scoring equations is to use the Output option for LC Cluster models implemented in LG 5.1:

- ➢ Open data3.sav using "File→Open"
- From the menu bar, click 'Model' and select 'Cluster'. A dialog box will pop up (Figure 6)

Clus	ter Model - data3.sav - Model1	×
Variables Advanced Model	Residuals ClassPred Output Technical	
burpose	Indicators>	
	Covariates>	
	Clusters 1	
Lexical Order	Case Weight> frq Select>	·
	Close Cancel Estimate Hel	p

Figure 6. Variables tab from Cluster Module

- Move the 4 indicators to the Indicators box
- Right click on the 4 indicators and select 'Nominal'
- Change the number of Clusters from '1' to '3'
- Click 'Scan'

		Clust	er Model - data3	.sav - M	odel1		×
Variables	Advanced	Model	Residuals ClassPred	Output	Technical		
du#1 du#2 du#3 du#			<indicators< td=""><td>purpose accuracy understa cooperat</td><td></td><td>Nominal Nominal Nominal Nominal</td><td>3 2 2 3</td></indicators<>	purpose accuracy understa cooperat		Nominal Nominal Nominal Nominal	3 2 2 3
			Covariates>				
			Clusters 3				
Lexica Scan	al Order	set	Case Weight> Select>	frq			21
			Close	Cancel	Estim	ate	Help

Figure 7. Variables tab from Cluster Module

> Open the Output tab and select 'Scoring Equations' (see Figure 8)

Cluster Model	I - data3.sav - Model1 🛛 🗙
Variables Advanced Model Residuals	ClassPred Output Technical
Output Sections Parameters Profile ProbMeans Bivariate Residuals Classification - Posterior Classification - Model Estimated Values Set Profile Set ProbMeans Importance Iteration Detail	Standard Errors and Wald Standard (Hessian) Robust (Sandwich) Fast (Outer Product) None Prediction Type Prediction Type
Default	Scoring Equations
Scoring Syntax Typ	pe SPSS v Browse
Restore to Defaults Save as Default	Cancel Changes
Close	e Cancel Estimate Help

Figure 8. Output tab from Cluster Module

➢ Click Estimate

> Expand the Parameters output to display the Scoring Equations

ρ	LatentGOLD		-	
File Edit View Model Wind	low Help			
🖻 🖶 🐰 🖻 💼 🎒 🕄	► \ ?			
⊡- data3.sav		Cluster1	Cluster2	Cluster3
	Intercept			
- Parameters		-2.0479	0.9315	1.1164
Loadings	purpose			
Caurings	good	0.6762	1.0709	-1.7471
	depends	-0.4683	0.1954	0.2729
Paired Comparisons	waste	-0.2079	-1.2664	1.4742
👜 Profile	accuracy	•		
ProbMeans	mostly true	0.5692	0.6510	-1.2202
Bivariate Residuals	not true	-0.5692	-0.6510	1.2202
EstimatedValues-Model	understa	0		0
Model2	good	1.7524	-1.3387	-0.4138
WINGEL	fair/poor	-1.7524	1.3387	0.4138
	cooperat	••••••		0
	interested	1.9660	-0.8573	-1.1087
	cooperative	0.7000	-0.2535	-0.4466
	impatient/hostile	-2.6660	1.1108	1 5552

Figure 9. Scoring Equations

Confirm that these coefficients are identical to those obtained using the Step3 Module (recall Figure 4).