PennState The Methodology Center
advancing methods, improving health

LCA_Distal_LTB SAS Macro Users' Guide (Version 1.1)

John J. Dziak
Jingyun Yang
Xianming Tan
Bethany C. Bray
Aaron T. Wagner
Stephanie T. Lanza
Penn State

NOTE: the functionality of this macro is identical to the (no longer distributed) LCA Distal SAS Macro (v. 3.0.2). For estimating LCA with a distal outcome, we recommend using the %LCA_Distal_BCH macro. See section 1 for more information.

Please send questions and comments to *MChelpdesk@psu.edu*.
The development of the SAS %LCA_Distal_LTB macro was supported by National Institute on Drug Abuse Grant P50-DA10075 and P50-DA039838. The authors would like to thank Bethany Bray and Amanda Applegate for their helpful comments.

Thank you for citing this users' guide when you use this macro. Suggested citation:

Dziak, J. J., Yang, J., Tan, X., Bray, B. C., Wagner, A. T., & Lanza, S. T. (2017). *LCA distal SAS macro users' guide* (Version 1.1). University Park: The Methodology Center, Penn State. Retrieved from http://methodology.psu.edu

Contents

1 About the %LCA_Distal_LTB Macro

The %LCA_Distal_LTB SAS macro estimates the association between a latent class variable and a distal outcome using a model-based approach. The %LCA_Distal_LTB macro is designed to work with SAS Version 9.1 or higher and PROC LCA version 1.3.2 or higher. This macro is based on an earlier macro, also by the Methodology Center, LCA_Distal version 3.0.2. The name of the macro has been changed to distinguish it from another macro, LCA_Distal_BCH.

Note: Currently, the recommended procedure for LCA with a distal outcome for any metric (e.g., binary, count, continuous) is a classification error adjusted approach, colloquially known as the BCH approach. The BCH approach can be implemented using the %LCA_Distal_BCH SAS macro, which is also available for free download at methodology.psu.edu. The %LCA_Distal_LTB SAS macro is provided primarily for methodological research, such as comparing results using the BCH approach and the LTB approach.

The %LCA_Distal_LTB macro

- uses simple, minimal syntax;
- estimates class-specific probabilities for binary and categorical distal outcomes;
- estimates class-specific means and modes for count and continuous distal outcomes;
- provides output including class-specific conditional densities for continuous distal outcomes; and
- can accommodate distal outcomes for multiple groups.

This guide assumes the user has a working knowledge of latent class analysis and PROC LCA. The book, *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (Collins & Lanza, 2010), provides a comprehensive introduction to the use of latent class analysis in applied research. PROC LCA & PROC LTA and the accompanying users' guide can be downloaded from http://methodology.psu.edu/downloads. A detailed explanation of the model-based estimation approach used in the %LCA_Distal_LTB macro appears in the article, Lanza, Tan, and Bray (2013). In that article, the macro is identified as the %LCA_Distal macro.

2 System Requirements

The %LCA_Distal_LTB macro requires

- SAS Version 9.1 or higher (Windows version),
- PROC LCA & PROC LTA Version 1.3.2 or higher (to fit LCA models),
- SAS/STAT (PROC KDE for non-parametric density estimation), and
- %create_group macro, included in the download of this macro (needed for categorical distal outcome with groups only).

Note for SAS version 9.1 or 9.2 users: Newer versions of Java Runtime Environment are incompatible with certain features of SAS. See Appendix A for details. For SAS 9.3 or higher, no action is required.

Note: SAS/STAT is sold separately from the base SAS package, but most university licenses include it. If you can run PROC LCA, you can run this macro.

3 The Model-Based Approach to LCA With Distal Outcomes

3.1 Recommended Method for Estimating LCA With Distal Outcomes

As previously stated, the recommended procedure for LCA with a distal outcome for any metric (e.g., binary, count, continuous) is a classification error adjusted approach, colloquially known as the BCH approach. The BCH approach can be implemented using the %LCA_Distal_BCH SAS macro, which is also available for free download at methodology.psu.edu. The %LCA_Distal_LTB SAS macro is provided primarily for methodological research, such as comparing results using the BCH approach and the LTB approach.

3.2 LTB Approach to LCA With a Distal Outcome for a Single Group

Researchers are often interested in the relationship between a latent class variable, *C,* and a distal outcome, *Z*. The %LCA_Distal_LTB macro provides a model-based approach to estimating the association between *C* and *Z*; that is, the conditional distribution of *Z* given C, *f*{*Z*|*C*}.

In order to estimate *f*{*Z*|*C*}, we assume conditional independence between an observed item in the model, *X*, and *Z* given *C*. Stated mathematically, we assume that *f*{*X,Z*|*C*} = *f*{*X*|*C*} *f*{Z|*C*}. This conditional independence assumption is similar to the local independence assumption essential to most LCA models (Collins & Lanza, 2010).

In LCA with a distal outcome, interest lies in the conditional distribution *f*{*Z*|*C*}. The distribution of *Z*|*C* can be determined by applying Bayes' theorem:

$$f\{Z = z \mid C = c\} = \frac{f\{Z = z\} \times f\{C = c \mid Z = z\}}{f\{C = c\}} \quad . \tag{1}$$

To obtain each element in the model,

- *f*{*C* = *c*} is derived from the LCA model,
- *f*{*C* = *c*|*Z* = *z*} is derived from the LCA model with *Z* included as a covariate, and
- *f*{*Z* = *z*} is the marginal distribution of *Z*, and can be empirically estimated.

When the distal outcome is binary, count, or categorical, we include *Z* in the LCA model as a covariate, and assume conditional independence between *X* and *Z* given *C*. The conditional distribution of $f\{Z = z|C = c\}$ for binary, count, or categorical *Z* can be calculated by re-expressing (1) as follows:

$$\Pr\{Z = z \mid C = c\} = \frac{\Pr\{Z = z\}e^{\beta_{0c}+\beta_{1c}z}}{\Pr\{C = c\}(1+\sum_{c'=1}^{K-1} e^{\beta_{0c'}+\beta_{1c'}z})}. \quad (2)$$

Specifically, $\Pr\{Z = z\}$ is estimated from the empirical distribution of *Z* (i.e., from the proportions in the observed data); the estimates for $\{\beta_{0c}, \beta_{1c}; c = 1,2,\ldots, K - 1\}$ are provided by the LCA with covariates model; and the marginal distribution $\Pr\{C = c\}$ can be obtained by multiplying $\Pr\{C = c|Z = z\}$ by the marginal empirical distribution $\Pr\{Z = z\}$. Thus, we can obtain $\Pr\{Z = z|C = c\}$ given these estimates for $\Pr\{Z = z\}$, $\Pr\{C = c\}$, and $\{\beta_{0c}, \beta_{1c}; c = 1,2,\ldots, K - 1\}$.

We follow a very similar procedure in the case of a continuous distal outcome, except that we use a smoothed empirical estimate of the distribution. Specifically, for a continuous distal outcome we estimate the distribution of *Z* using SAS PROC KDE, a kernel density estimation procedure that approximates the unknown population distribution using a smoothed version of the observed data histogram.

Currently, the macro assumes that if a distal outcome is a continuous variable, then it is homoscedastic; that is, the variance of the variable is the same across all latent classes. Heteroscedasticity in a covariate causes the estimation to be less accurate. If you have a continuous distal outcome variable, we recommend using the %LCA_Distal_BCH SAS macro.

The macro now generates standard errors in the binary case; see section 5.1 for more information. Standard errors are not yet available from the macro for other kinds of distal outcome variables.

An overview of the LCA model, a thorough explanation of the model-based approach to LCA with a distal outcome, and results from a Monte Carlo study for evaluating this approach can be found in a technical report by Lanza, Tan, and Bray (2011), which can be downloaded from http://methodology.psu.edu.

3.3 LTB Approach to LCA With a Distal Outcome for Multiple Groups

In many cases, there are multiple groups *G* (e.g., male and female) in the data, and researchers are interested in conditional distribution of *Z* given *C* and *G*. Similarly, within each group, we assume conditional independence between any given observed item in the model and the outcome *Z* given *C*.

We are interested in the conditional distribution of $f(Z=z \mid C=c, G=g)$. Using Bayes' theorem, this can be expressed as

$$f(Z=z \mid C=c,G=g) = \frac{f(C=c \mid Z=z,G=g)\,f(Z=z \mid G=g)\,f(G=g)}{f(G=g)\,f(C=c \mid G=g)} = \frac{f(Z=z \mid G=g)\,f(C=c \mid Z=z,G=g)}{f(C=c \mid G=g)}. \quad (3)$$

To obtain each element in the model,

- $f(C=c \mid G=g)$ is derived from the LCA model conditional on *G,*
- $f(C=c \mid Z=z,G=g)$ is derived from the LCA model with *Z* included as a covariate conditional on G, and
- $f(Z=z \mid G=g)$, the marginal distribution of *Z* conditional on *G*, is estimated empirically.

The macro can accommodate distal outcomes that are specified as **binary variables**, **continuous variables, count variables, or categorical variables**. In the case of a continuous distal outcome, a smoothed estimate of *f*{*Z = z*|*G=g*} is empirically estimated using kernel density estimation. **Currently, the macro assumes that if a distal outcome is a continuous variable, then it is homoscedastic; that is, the variance of the variable is the same across all latent classes.**

4 Using the %LCA_Distal_LTB Macro

Table 1. Argument Definitions for the %LCA_Distal_LTB Macro.

Argument	Required	Description
`input_data`	Y	Input data set. The distal outcome must be included as a covariate.
`param`	Y	Name of the data set generated by PROC LCA as the OUTPARAM output. The data set contains estimates of the beta parameters.
`distal`	Y	Distal outcome variable.
`group`	N	Variable for multiple groups. (If `metric=4`, use the new variable created by %create_group. See section 4.2.2 for details.) If no group argument is supplied, the macro assumes there is only one group.
`metric`	Y	Numeric value indicating whether the distal outcome is a binary (`metric = 1`), continuous (`metric = 2`), count (`metric = 3`) or categorical (`metric = 4`) variable.
`covariance_beta`	N	Name of the data set generated by PROC LCA as the OUTCOVB output. The data set contains the betas' covariance matrix.
`output_dataset_name`	Y	Name of the data set to contain output from the %LCA_Distal_LTB macro.
`out_density`	N	Name of the data set for conditional density of the continuous distal outcome for each latent class. Default is "cond_density."
`plot_density`	N	For multiple groups, whether to plot conditional density for each group (`plot_density=1`) or not (`plot_density=0`, default).
`weight`	N	Name of the variable specifying survey weight. This option only works in the binary outcome case. It assumes that `weight` has also been used in the previous call to PROC LCA.

4.1 Preparation

A SAS macro is a special block of SAS commands. The block is first defined and then called when needed. Four steps need to be completed before you run the macro.

1. If you haven't already done so, download and save the macro to a designated path (e.g., `S:\myfolder\`).

2. Direct SAS to read the macro code from the path, using a SAS %INCLUDE statement such as

```
%INCLUDE "S:\myfolder\LCA_Distal_LTB_macro_v11.sas";
```

3. Direct SAS to the input data file. We assume the data set is a permanent file saved to a designated directory. If so, we recommend using a "libname" statement. The statement should give the libname command, name the library, and then identify the path to the data. For example,
   ```
   libname sasf "s:\myfolder\";
   ```

4. Ensure that the distal outcome is coded as follows:
   - Binary: 0, 1
   - Continuous: original coding or standardized variable
   - Count: original coding (0, 1, 2, …)
   - Categorical: 1, 2, …

Note: Missingness in the distal outcome variable should be imputed (e.g., multiple imputation; Schafer, 1997). Otherwise, cases with missing values in the distal outcome variable must be removed from the analysis.

4.2 Estimation of the Latent Class Model in PROC LCA

4.2.1 Estimation in PROC LCA for a Single Group

Use PROC LCA to generate the output needed for use by the %LCA_Distal_LTB macro. First, you must select the LCA model. This process is described in chapter 5 of the *PROC LCA & PROC LTA Users' Guide* (Lanza, Dziak, Huang, Xu, & Collins, 2011).

Once model selection is complete, generate a file containing the parameter estimates to be used in the macro by estimating the latent class model with the distal outcome included as a covariate. This file can be generated using the OUTPARAM option in PROC LCA. (See section 5.3 of the *PROC LCA & PROC LTA Users' Guide* for more information.) If you are not using multiple groups, and if the distal outcome is binary, count, or continuous, the PROC LCA syntax will be similar to the following:

```
PROC LCA DATA = sasf.simdata_binary OUTPARAM = Binary_param OUTCOVB =
    Binary_covb; /* the input data set, the file to be generated
    containing the parameter estimates, and the file to be generated
    containing the covariance matrix (only needed for generating
    standard errors in the binary case) */
    NCLASS 5; /* the number of latent classes */
    ITEMS item001 item002 item003 item004 item005 item006 item007
    item008; /* indicator variables used to measure the latent class
    variable */
    CATEGORIES 2 2 2 2 2 2 2 2; /* number of response categories for
    each indicator variable (in this case, all binary) */
```

```
    COVARIATES z; /* the distal outcome variable*/
    SEED 54327; /* a random number. */
RUN;
```

If the distal outcome is categorical and you are not using multiple groups, the distal outcome is included using a `groups` statement instead of a `covariates` statement. The `measurement` statement will also be needed to restrict estimation so that item-response probabilities for each item-latent class combination are equal across all groups. In such cases, the PROC LCA syntax will be similar to the following:

```
PROC LCA DATA = sasf.simdata_categ OUTPARAM = categ_param;
           NCLASS 5;
           ITEMS item001-item008;
           CATEGORIES 2 2 2 2 2 2 2 2;
           GROUPS z; /* z= 1/2/3 */
           MEASUREMENT group;
 SEED 10;
RUN;
```

Other arguments available in PROC LCA, such as `id, rho prior, maxiter`, and `criterion` may also be necessary for estimation of the latent class model. Refer to the *PROC LCA & PROC LTA Users' Guide* for more information.

4.2.2 Estimation in PROC LCA for Multiple Groups

When there are multiple groups, if the distal outcome is binary, count, or continuous, the PROC LCA syntax will be similar to the following:

```
PROC LCA DATA = sasf.simdata_binary_group OUTPARAM = Binary_param
    OUTCOVB = Binary_covb; /* the input data set, the file to be
    generated containing the parameter estimates, and the file to be
    generated containing the covariance matrix (outcovb only needed for
    generating standard errors in the binary case) */
    NCLASS 5; /* the number of latent classes */
    ITEMS item001 item002 item003 item004 item005 item006 item007
    item008; /* indicator variables used to measure the latent class
    variable */
    CATEGORIES 2 2 2 2 2 2 2 2; /* number of response categories for
    each indicator variable (in this case, all binary) */
    COVARIATES z; /* the distal outcome variable*/
    GROUPS educ; /*the variable for multiple groups*/
    SEED 54327; /* a random number. */
RUN;
```

If the distal outcome is categorical, a new grouping variable needs to be created in order to include the distal outcome using the `group` statement in PROC LCA. This can be done using the provided %create_group SAS macro. Direct SAS to read the macro code from the path, using a SAS %INCLUDE statement such as

```
%INCLUDE "S:\myfolder\create_group.sas";
```

The new data set to be used in PROC LCA can be created by calling the macro using a percent sign, its name, and user-defined arguments in parentheses. The macro parameters, (i.e., the information in parentheses provided to the macro), are shown below.

```
%create_group(
input_data=sasf.simdata_categ_group,
group=educ,
distal=z,
output_data=newg
);
```

Table 2: Argument Definitions for the %create_group macro.

Argument	Required	Description
`input_data`	Y	Input data set.
`group`	Y	The variable for multiple groups.
`distal`	Y	Distal outcome variable.
`output_data`	Y	Name of the output data set to be used in PROC LCA.

Assuming that the variable for multiple groups is `educ`, the %create_group will create a new data set named "newg" with a new grouping variable named "`new_educ`".

This new data set should be used in PROC LCA with "`new_educ`" included in the `group` statement. The `measurement` statement will also be needed to restrict estimation so that item-response probabilities for each item-latent class combination are equal across all groups. In such cases, the PROC LCA syntax will be similar to the following:

```
PROC LCA DATA = newg OUTPARAM = categ_param;
           NCLASS 5;
           ITEMS item001-item008;
           CATEGORIES 2 2 2 2 2 2 2 2;
           GROUPS new_educ; /* the new group variable*/
           MEASUREMENT group;
           SEED 10;
RUN;
```

Other arguments available in PROC LCA, such as `id, rho prior, maxiter`, and `criterion` may also be necessary for estimation of the latent class model. Refer to the *PROC LCA& PROC LTA Users' Guide* for more information.

4.3 Macro Syntax and Input

Call the macro using a percent sign, its name, and user-defined arguments in parentheses. The macro parameters are shown below.

```
      %LCA_Distal_LTB (
                input_data = data set name,
                param = name of OUTPARAM data set created by PROC LCA,
                distal = variable,
                group = variable,
                metric = number,
                output_dataset_name = data set name
                out_density = data set name,
                plot_density = number,
                covariance_beta = data set name (used in binary case
only) ,
                weight = survey weighting variable name (optional and
used in binary case only)
      ) ;
```

4.4 Output

The output for this macro depends on the metric specified for the distal outcome. If there is a single group, the output will always contain a set of estimates representing the association between membership in each latent class and the distal outcome. For a binary outcome (`metric` = 1) or a categorical outcome (`metric` = 4), the output file will include a column for each latent class. The probability associated with each level of the distal outcome will be listed for each class. For a continuous outcome (`metric` = 2) or a count outcome (`metric` = 3), the output file will include a column for each latent class. The mean value of the outcome and the mode of the outcome for each latent class will be listed. For continuous outcomes *only*, a separate file will be generated that includes the conditional density of the distal outcome for each class. If the user decides to plot conditional densities (`plot_density` = 1), the macro will also generate a conditional density plot. If there are multiple groups, the output will include the above information for each group.

The macro will also display results on screen. For count or categorical outcomes, estimated means and modes within each latent class are shown. For binary or categorical outcomes, the results include the estimated probability of each outcome within each latent class.

For the binary case, the results displayed onscreen also include

- standard errors for the class-specific probabilities,
- confidence intervals for the class-specific probabilities, and
- a Wald Chi-squared test for every pair of classes to test whether their class-specific probabilities are significantly different. (This is expressed as the Wald Chi-squared test statistic and as the p-value associated with the test statistic.)

The additional output created in the binary case is also saved in three new data sets. (These will be generated in the “Work” library.) The new data sets are

- pi_by_class: the probability of the distal outcome being = 1 given each class, with 95% confidence bounds,
- Wald_teststats: the test statistics of the class-conditional probabilities
- Wald_pvalue: the p-values for the class conditional probabilities.

Some users who are running SAS 9.1 or 9.2 will not be able to generate the conditional density plot because of incompatibility between SAS and certain versions of Java Runtime Environment. See Appendix A for instructions on how to fix this issue.

5 Empirical Demonstrations of the %LCA_Distal_LTB Macro

In this section, we first describe the structure of the data sets and the variables to be analyzed. Then, we illustrate how to estimate the distribution of the distal outcome within each latent class using the %LCA_Distal_LTB macro and describe the output of the macro. Section 5.1 describes use of the macro with a binary distal outcome. Continuous, count, and categorical outcomes are discussed in sections 5.2, 5.3, and 5.4, respectively.

For demonstrations of the macro with multiple groups, see chapter 6.

5.1 Estimating a Binary Distal Outcome

Before attempting to complete the following example, please download the file *%LCA_Distal Examples* from the %LCA_Distal_LTB macro download page at http://methodology.psu.edu. Also, verify that you are running PROC LCA v.1.3.2 or higher.

5.1.1 Example Data

Below are the first 10 observations from the SAS data set **simdata_binary.sas7bdat**, which is contained in the *%LCA_Distal Examples* file.

ID	Item001	Item002	Item003	Item004	Item005	Item006	Item007	Item008	Z
1	2	2	1	2	2	2	2	2	1
2	1	1	2	2	2	2	2	2	0
3	2	1	2	1	1	1	1	1	0
4	2	2	2	2	2	2	2	2	1
5	2	2	2	2	2	2	2	2	1
6	1	1	1	2	2	2	2	2	1
7	2	2	1	2	2	2	2	2	1
8	2	2	2	2	2	2	2	2	1
9	2	2	2	2	2	2	2	2	1
10	2	2	2	2	1	2	2	2	1

ID= subject's identification variable,

Item001,…, Item008= 8 items used to measure the latent class variable

Z= the distal outcome (Note: binary distal outcome should be coded using 0s and 1s.)

5.1.2 Example Syntax

Include a “libname” statement prior to running the macro to direct SAS to the data file.

```
libname sasf "S:\myfolder\";
```

Note: We suppose that the SAS data set exists in the folder S:\myfolder\. This path represents any user-specified folder.

Once the LCA model has been identified, we estimate the LCA model including the distal outcome *Z* as a covariate using PROC LCA.

```
PROC LCA DATA = sasf.SimData_Binary OUTPARAM = Binary_param OUTCOVB =
                  binary_covb;
                  ID id;
                  NCLASS 5;
                  ITEMS item001-item008;
                  CATEGORIES 2 2 2 2 2 2 2 2;
                  COVARIATES z;
                  RHO PRIOR = 1;
                  SEED 10;
                  MAXITER 5000;
                  CRITERION 0.000001;
RUN;
```

The output is described in the *PROC LCA & PROC LTA Users' Guide*.

Output will also include the files BINARY_PARAM and BINARY_COVB in the work directory.

VIEWTABLE: Work.Binary_param

	Parameter Type	Group Number	Variable Name	Response Category	Estimate - Latent Class 1	Estimate - Latent Class 2	Estimate - Latent Class 3	Estimate - Latent Class 4	Estimate - Latent Class 5
1	GAMMA	1		.	0.323545	0.280305	0.099871	0.133263	0.163016
2	BETA	1		.	.	0.882074	0.784210	0.738405	0.701830
3	BETA	1	z	.	.	-1.15509	-2.49450	-1.95009	-1.61560
4	RHO	1	Item001	1	0.056834	0.906451	0.933252	0.977816	0.134035
5	RHO	1	Item002	1	0.066115	0.881024	0.871387	0.829746	0.146703
6	RHO	1	Item003	1	0.085492	0.850127	0.851169	0.961339	0.127452
7	RHO	1	Item004	1	0.078072	0.883143	0.949816	0.878651	0.055761
8	RHO	1	Item005	1	0.044679	0.087768	0.934019	0.686866	0.771810
9	RHO	1	Item006	1	0.047961	0.003221	0.856566	0.970499	0.713803
10	RHO	1	Item007	1	0.116454	0.062794	0.951015	0.002227	0.106036
11	RHO	1	Item008	1	0.115358	0.153006	0.771621	0.103327	0.103264
12	RHO	1	Item001	2	0.943166	0.093549	0.066748	0.022184	0.865965
13	RHO	1	Item002	2	0.933885	0.118976	0.128613	0.170254	0.853297
14	RHO	1	Item003	2	0.914508	0.149873	0.148831	0.038661	0.872548
15	RHO	1	Item004	2	0.921928	0.116857	0.050184	0.121349	0.944239
16	RHO	1	Item005	2	0.955321	0.912232	0.065981	0.313134	0.228190
17	RHO	1	Item006	2	0.952039	0.996779	0.143434	0.029501	0.286197
18	RHO	1	Item007	2	0.883546	0.937206	0.048985	0.997773	0.893964
19	RHO	1	Item008	2	0.884642	0.846994	0.228379	0.896673	0.896736

Binary_param

VIEWTABLE: Work.Binary_covb

	Col_1	Col_2	Col_3	Col_4	Col_5	Col_6	Col_7	Col_8
1	0.2414	-0.2073	0.1922	-0.1629	0.1922	-0.1632	0.2380	-0.1267
2	-0.2073	0.2080	-0.1600	0.1546	-0.1607	0.1530	-0.1908	0.1470
3	0.1922	-0.1600	0.2424	-0.2033	0.1920	-0.1639	0.2376	-0.1258
4	-0.1629	0.1546	-0.2033	0.2448	-0.1639	0.1488	-0.1918	0.1464
5	0.1922	-0.1607	0.1920	-0.1639	0.2520	-0.2110	0.2399	-0.1245
6	-0.1632	0.1530	-0.1639	0.1488	-0.2110	0.2303	-0.1924	0.1431
7	0.2380	-0.1908	0.2376	-0.1918	0.2399	-0.1924	0.3564	-0.1919
8	-0.1267	0.1470	-0.1258	0.1464	-0.1245	0.1431	-0.1919	0.2622

Binary_covb

Then, the BINARY_PARAM file is used to estimate the distribution of the distal outcome within each latent class. The BINARY_COVB file is used to estimate the confidence intervals. Include the macro and enter the proper syntax in SAS.

```
%LCA_Distal_LTB (input_data = sasf.simdata_binary,
      param = Binary_param,
      distal = z,
      metric = 1,
      covariance_beta = Binary_covb,
      output_data set_name = Binary_res);
```

The `input_data` argument identifies the data file. The `param` argument directs the macro to the parameters in the OUTPARAM file generated by PROC LCA. The `distal` argument identifies the distal outcome variable in the data set. The `metric` argument indicates that the distal outcome is binary. The `covariance_beta` argument indicates the data set generated by PROC LCA that contains the covariance matrix needed to generate confidence intervals and `output_data set_name` names the macro's output.

5.1.3 Example Output

Below is the onscreen output. It includes the class-specific distribution estimates for the distal outcome, the estimated class-conditional probabilities, the Wald test statistic on class-conditional probabilities and the p-value on class-conditional probabilities.

Results Viewer - SAS Output

Class 5	0.2547	0.7453

```
P(Z=k) in the table above is the probability that z equals k.
```

The SAS System

```
Estimated Class-Conditional Probabilities for P(z=1|Class=c)
                                        Confidence Interval
Class      Probability                  For Probability
   1            0.9364            (     0.6892,    0.9899 )
   2            0.8226            (     0.7404,    0.8829 )
   3            0.5485            (     0.3864,    0.7009 )
   4            0.6768            (     0.5629,    0.7729 )
   5            0.7453            (     0.6105,    0.8452 )
 Wald Test on Class-Conditional Probabilities
   Classes            Wald    df  p-value
  1  vs.   2        1.0160     1  0.3135
  1  vs.   3        5.3061     1  0.0213
  1  vs.   4        3.2849     1  0.0699
  1  vs.   5        2.4524     1  0.1173
  2  vs.   3       10.3763     1  0.0013
  2  vs.   4        5.1812     1  0.0228
  2  vs.   5        1.0841     1  0.2978
  3  vs.   4        1.4320     1  0.2314
  3  vs.   5        3.2440     1  0.0717
  4  vs.   5        0.5937     1  0.4410
```

The WORK directory will also contain four data sets: the user defined output, Wald_pvalue, Wald_teststats, and Pi_by_class.

VIEWTABLE: Work.Binary_res

	Z	prob_z	Z_C01	Z_C02	Z_C03	Z_C04	Z_C05
1	0	0.2	0.0636180804	0.1774039271	0.4514974759	0.3232147687	0.2547325748
2	1	0.8	0.9363819196	0.8225960729	0.5485025241	0.6767852313	0.7452674252

The output in the user-named file conveys the association between membership in each latent class and the distal outcome. The first row represents the probability of $z = 0$ conditional on latent class membership and the second row indicates the probability of $z = 1$ conditional on latent class membership. The column prob_Z indicates the marginal probability of each value of Z for the entire sample. The column Z_C01 represents the probability of the distal outcome equaling 0 or 1, given membership in latent class 1. Columns Z_C02 through Z_C05 represent the same thing for classes 2 through 5.

The Wald_pvalue, Wald_teststats, and Pi_by_class data sets are for technical use. They contain the data summarized in the onscreen output.

VIEWTABLE: Work.Pi_by_class

	PI_GIVEN_CLASS_BEST	PI_GIVEN_CLASS_UPPER	PI_GIVEN_CLASS_LOWER	CLASS
1	0.9363819196	0.9898661828	0.6892380933	1
2	0.8225960729	0.8828623347	0.7404392357	2
3	0.5485025241	0.7009336049	0.3863921246	3
4	0.6767852313	0.7729397092	0.5629365302	4
5	0.7452674252	0.8452190305	0.6105139055	5

Pi_by_class

VIEWTABLE: Work.Wald_pvalues

	COL1	COL2	COL3	COL4	COL5
1	.	0.3134795345	0.0212507683	0.0699194991	0.117342961
2	0.3134795345	.	0.0012764506	0.0228319817	0.2977784264
3	0.0212507683	0.0012764506	.	0.2314338364	0.071684598
4	0.0699194991	0.0228319817	0.2314338364	.	0.4409764815
5	0.117342961	0.2977784264	0.071684598	0.4409764815	.

Wald_pvalues

VIEWTABLE: Work.Wald_teststats

	COL1	COL2	COL3	COL4	COL5
1	.	1.0159587171	5.3061088762	3.2849091207	2.4524316726
2	1.0159587171	.	10.376273643	5.1812432194	1.0841167797
3	5.3061088762	10.376273643	.	1.4320274869	3.244018854
4	3.2849091207	5.1812432194	1.4320274869	.	0.5937400928
5	2.4524316726	1.0841167797	3.244018854	0.5937400928	.

Wald_teststats

5.2 Estimating a Continuous Distal Outcome

NOTE: Currently, the macro assumes that if a distal outcome is a continuous variable, then it is homoscedastic; that is, the variance of the variable is the same across all latent classes.

Before attempting to complete the following example, please download the file *%LCA_Distal Examples* from the %LCA_Distal_LTB macro download page.

5.2.1 Example Data

In **simdata_conti.sas7bdat**, the data structure is similar to the data set in section 5.1 of this document. However, instead of binary values for z, the values are continuous.

ID	Item001	Item002	Item003	Item004	Item005	Item006	Item007	Item008	Z
1	2	1	2	2	2	2	2	2	-1.8513098
2	1	1	2	1	2	2	2	2	-0.5950087
3	2	2	2	1	1	1	2	2	1.55437269
4	2	1	2	2	2	2	2	2	0.89742276
5	1	1	1	1	2	2	2	2	-0.3121734
6	1	1	1	1	2	2	2	2	-1.5068341
7	2	2	1	2	2	2	2	2	0.73713821
8	1	1	1	2	1	1	2	2	1.8747736
9	1	1	1	1	1	1	2	1	-0.0463611
10	1	1	1	1	1	1	1	1	-0.1706686

ID= subject's identification variable

Item001,…, Item008= 8 items used to measure the latent class variable

Z= the distal outcome (in this case a CONTINUOUS distal outcome)

5.2.2 Example Syntax

Include a "libname" statement prior to running the macro to direct SAS to the data file.

```
libname sasf "S:\myfolder\";
```

Note: we suppose that the SAS data set exists in the folder `S:\myfolder\`. This path represents any user-specified folder.

Once the LCA model has been identified, we estimate the LCA model including the distal outcome *Z* as a covariate using PROC LCA.

```
PROC LCA DATA = sasf.SimData_Conti OUTPARAM = Conti_param;
     ID id;
     NCLASS 5;
     ITEMS item001-item008;
     CATEGORIES 2 2 2 2 2 2 2 2;
     COVARIATES z;
     RHO PRIOR = 1;
     SEED 10;
     MAXITER 5000;
     CRITERION 0.000001;
RUN;
```

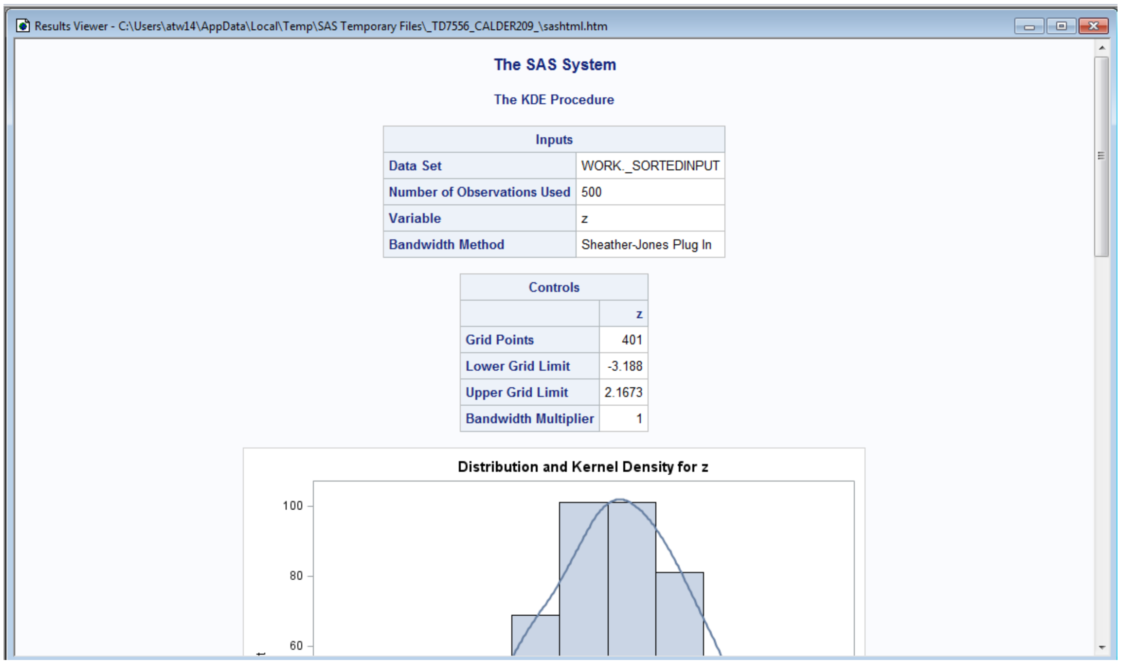
Then, the Conti_param file is used to estimate the distribution of the distal outcome within each latent class. Include the macro and enter the proper syntax in SAS.

```
%LCA_Distal_LTB (input_data = sasf.simdata_conti,
      param = conti_param,
      distal = z,
      metric = 2,
      output_dataset_name = Conti_res);
```

The `input_data` argument identifies the data file. The `param` argument directs the macro to the parameters generated in the OUTPARAM file generated by PROC LCA. The `distal` argument identifies the distal outcome. The `metric` argument indicates that the distal outcome is continuous, and `output_dataset_name` names the macro's output.

5.2.3 Example Output

The onscreen output contains the class-specific distribution estimates for the distal outcome. It also contains the kernel density for the distal outcome.



Results Viewer - C:\Users\atw14\AppData\Local\Temp\SAS Temporary Files_TD7556_CALDER209_\sashtml.htm

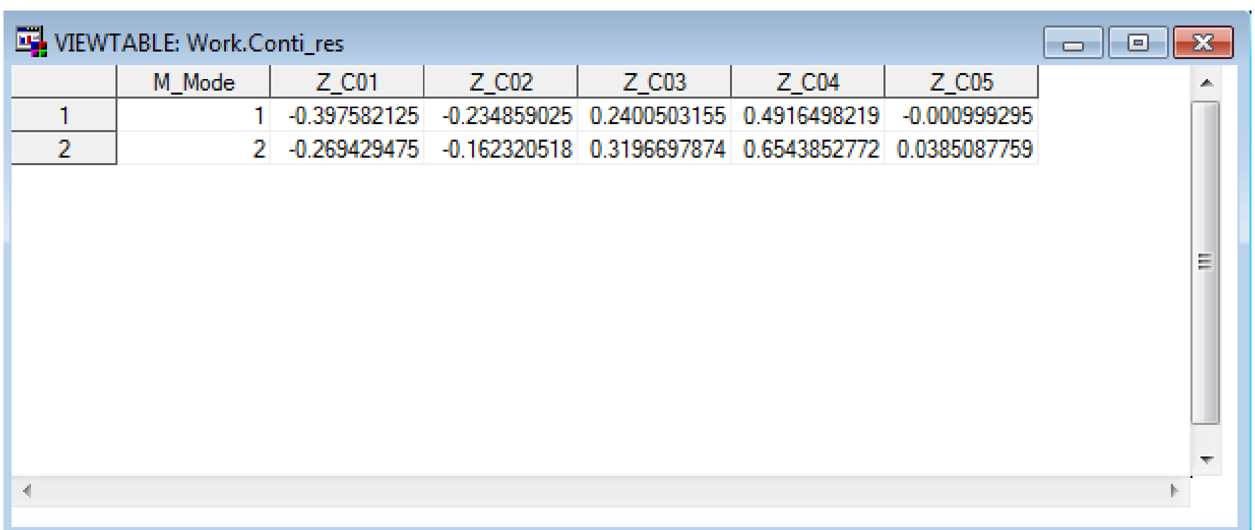
The SAS System

The KDE Procedure

Inputs	
Data Set	WORK._SORTEDINPUT
Number of Observations Used	500
Variable	z
Bandwidth Method	Sheather-Jones Plug In

Controls	
	z
Grid Points	401
Lower Grid Limit	-3.188
Upper Grid Limit	2.1673
Bandwidth Multiplier	1

The WORK directory will also contain the user-named output data set and the conditional density data set.



VIEWTABLE: Work.Conti_res

	M_Mode	Z_C01	Z_C02	Z_C03	Z_C04	Z_C05
1	1	-0.397582125	-0.234859025	0.2400503155	0.4916498219	-0.000999295
2	2	-0.269429475	-0.162320518	0.3196697874	0.6543852772	0.0385087759

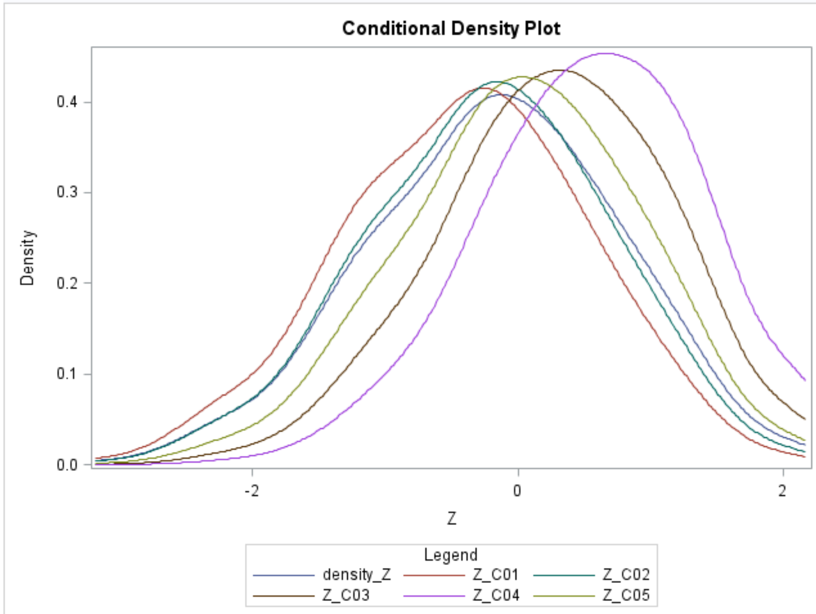
The output conveys the association between membership in each latent class and the distal outcome. The first row (M_Mode=1) represents the mean response for each class, and the second row (M_Mode=2) represents the modal response for each latent class. The column Z_C01 represents the mean and mode of the distal outcome for class 1. Columns Z_C02 through Z_C05 represent the same thing for classes 2 though 5.

The macro will also generate a data set for the conditional density; the name of the data set is specified by the user in the `out_density` statement. It includes the conditional density of the distal outcome for each latent class. It can be used to generate a plot of the conditional densities in SAS, Microsoft Excel, or other software.

VIEWTABLE: Work.Cond_density

	Z	Density_Z	Z_C01	Z_C02	Z_C03	Z_C04	Z_C05
1	-3.188148546	0.0046700935	0.0073009098	0.0043014108	0.0006802138	0.0002026224	0.0018449439
2	-3.174759926	0.0048324718	0.0075458683	0.0044566589	0.0007102311	0.0002125334	0.0019185895
3	-3.161371307	0.0050003098	0.0077987232	0.0046173194	0.0007415424	0.0002229199	0.0019950924
4	-3.147982687	0.005174152	0.0080602902	0.0047839136	0.0007742571	0.000233821	0.0020747074
5	-3.134594068	0.0053545838	0.0083314447	0.0049570032	0.0008084941	0.0002452792	0.0021577104
6	-3.121205448	0.0055422267	0.0086131136	0.0051371861	0.0008443815	0.0002573405	0.002244397
7	-3.107816828	0.0057377489	0.0089062911	0.0053251059	0.0008820586	0.0002700551	0.0023350869
8	-3.094428209	0.0059418451	0.0092120079	0.0055214342	0.000921673	0.0002834766	0.0024301167
9	-3.081039589	0.006155242	0.0095313386	0.0057268754	0.0009633819	0.0002976626	0.002529842
10	-3.06765097	0.0063787043	0.0098654107	0.0059421723	0.0010073532	0.000312675	0.0026346403
11	-3.05426235	0.0066130175	0.0102153779	0.0061680906	0.0010537633	0.000328579	0.0027449046
12	-3.04087373	0.0068589858	0.0105824168	0.006405417	0.0011027968	0.0003454441	0.0028610428
13	-3.027485111	0.0071174304	0.0109677236	0.0066549574	0.0011546468	0.0003633431	0.0029834767
14	-3.014096491	0.0073891911	0.0113725166	0.0069175388	0.001209515	0.000382353	0.0031126436
15	-3.000707872	0.0076751134	0.0117980162	0.0071939969	0.0012676101	0.0004025543	0.0032489907
16	-2.987319252	0.0079760455	0.0122454403	0.007485174	0.0013291473	0.0004240307	0.0033929742
17	-2.973930632	0.0082928336	0.0127159965	0.0077919137	0.0013943478	0.0004468696	0.0035450569
18	-2.960542013	0.0086263155	0.013210873	0.0081150554	0.0014634377	0.0004711611	0.0037057063
19	-2.947153393	0.0089773185	0.0137312349	0.0084554321	0.0015366478	0.0004969985	0.0038753929
20	-2.933764774	0.0093466415	0.0142781971	0.0088138533	0.0016142104	0.0005244769	0.0040545828
21	-2.920376154	0.0097350622	0.0148528356	0.009191112	0.0016963605	0.000553694	0.0042437407
22	-2.906987534	0.0101433328	0.0154561808	0.0095879798	0.0017833352	0.0005847499	0.0044433275
23	-2.893598915	0.0105721557	0.0160891815	0.0100051847	0.0018753691	0.0006177452	0.0046537896
24	-2.880210295	0.0110221975	0.0167527254	0.0104434229	0.0019726966	0.0006527824	0.0048755648
25	-2.866821676	0.0114940748	0.017447619	0.0109033462	0.0020755491	0.0006899643	0.0051090752

If the user decides to plot conditional densities (`plot_density = 1`), the macro will also generate a conditional density plot. For some users, this will be displayed in the SAS Results Viewer. For others, it will be saved as a .png file. (If this happens, it will be recorded in the SAS log.) This differs because of different installations of Java on each individual user's machine. Some users who are running SAS 9.1 or 9.2 will not be able to generate the conditional density plot because of incompatibility between SAS and certain versions of Java Runtime Environment. See Appendix A for instructions on how to fix this issue.



Conditional Density Plot
Density
0.0
0.1
0.2
0.3
0.4
-2
0
2
Z
Legend
density_Z
Z_C01
Z_C02
Z_C03
Z_C04
Z_C05

5.3 Estimating a Count Distal Outcome

Before attempting to complete the following example, please download the file *%LCA_Distal Examples* from the %LCA_Distal_LTB macro download page.

5.3.1 Example Data

In **simdata_count.sas7bdat**, the data structure is similar to the data set in section 5.1 of this document. However, the item *z* contains count responses with values from 0 to 4.

ID	Item001	Item002	Item003	Item004	Item005	Item006	Item007	Item008	Z
1	2	2	2	2	1	1	2	1	2
2	2	2	2	2	2	2	2	2	0
3	2	1	1	1	2	1	2	2	0
4	2	2	2	2	2	2	2	1	0
5	2	2	1	2	2	2	2	2	0
6	1	1	1	1	2	1	2	2	1
7	2	2	2	2	1	1	2	2	0
8	1	1	1	1	2	2	2	1	0
9	2	2	1	1	2	2	2	2	1
10	2	2	2	2	2	2	2	2	1

ID= subject's identification variable

Item001,…, Item008= 8 items used to measure the latent class variable

Z= the distal outcome (in this case a COUNT distal outcome)

5.3.2 Example Syntax

Include a "libname" statement prior to running the macro to direct SAS to the data file.

```
libname sasf "S:\myfolder\";
```

Note: we suppose that the SAS data set exists in the folder S:\myfolder\. This path represents any user-specified folder.

We estimate the LCA model including the distal outcome Z as a covariate using PROC LCA.

```
PROC LCA DATA = sasf.SimData_Count OUTPARAM = Count_param;
    ID id;
    NCLASS 5;
    ITEMS item001-item008;
    CATEGORIES 2 2 2 2 2 2 2 2;
    COVARIATES z;
    RHO PRIOR = 1;
    SEED 10;
```

```
        MAXITER 5000;
        CRITERION 0.000001;
RUN;
```

Then, the COUNT_PARAM file is used to estimate the distribution of the distal outcome within each latent class. Include the macro and enter the proper syntax in SAS.

```
%LCA_Distal_LTB (
    input_data = sasf.simdata_count,
    param = Count_param,
    distal = z,
    metric = 3,
    output_dataset_name = Count_res);
```

The `input_data` argument identifies the data file. The `param` argument directs the macro to the parameters generated in the OUTPARAM file generated by PROC LCA. The `distal` argument identifies the distal outcome. The `metric` argument indicates that the distal outcome is continuous, and `output_dataset_name` names the macro's output.

5.3.3 Example Output

The onscreen output contains the class-specific distribution estimates for the distal outcome.

The SAS System

Class-Specific Distribution Estimates for z

Class	Mean	Mode
1	0.3953	0
2	0.7607	0
3	1.2503	1
4	0.8759	0
5	0.7722	0

The Work directory will also contain the user-named output data set.

VIEWTABLE: Work.Count_res

	M_Mode	Z_C01	Z_C02	Z_C03	Z_C04	Z_C05
1	1	0.395293623	0.760684391	1.250341336	0.875912386	0.7721643944
2	2	0	0	1	0	0

The output conveys the association between membership in each latent class and the distal outcome. The first row (M_Mode=1) represents the mean count for each latent class, and the second row (M_Mode=2) represents the modal count for each latent class. The column Z_C01 represents the mean and mode of the distal outcome for members of class 1. Columns Z_C02 through Z_C05 represent the same thing for classes 2 though 5.

5.4 Estimating a Categorical Distal Outcome

Before attempting to complete the following example, please download the file *%LCA_Distal Examples* from the %LCA_Distal_LTB macro download page.

5.4.1 Example Data

First, we will examine the structure of the database and the variables to be analyzed. Below are the first 10 observations from the SAS data set **simdata_categ.sas7bdat**, which is contained in the *%LCA_Distal Examples* file available at http://methodology.psu.edu

ID	Item001	Item002	Item003	Item004	Item005	Item006	Item007	Item008	Z
1	2	2	2	2	1	2	2	2	2
2	2	1	2	2	2	2	2	2	3
3	1	1	2	1	1	2	2	1	2
4	2	1	1	1	2	1	2	2	3
5	2	2	2	1	2	2	2	2	1
6	2	2	2	2	1	2	2	2	1
7	1	1	1	1	1	2	2	2	2
8	2	2	2	2	2	2	2	1	3
9	2	2	2	2	2	2	1	2	2
10	2	2	2	1	1	1	2	1	1

ID= subject's identification variable

Item001,…, Item008= 8 items used to measure the latent class variable

Z= the distal outcome (Note: The categorical distal outcome should be coded using 1, 2, 3, …, g, where g = the number of categories.)

5.4.2 Example Syntax

Include a "libname" statement prior to running the macro to direct SAS to the data file.

```
libname sasf "S:\myfolder\";
```

Note: we suppose that the SAS data set exists in the folder `S:\myfolder\`. This path represents any user-specified folder.

Once the LCA model has been identified, we estimate the LCA model including the distal outcome *Z* as a grouping variable using PROC LCA. Note that measurement invariance across groups should be imposed using the `measurement` statement.

```
PROC LCA DATA = sasf.SimData_Categ OUTPARAM = Categ_param;
       ID id;
       NCLASS 5;
       ITEMS item001-item008;
       CATEGORIES 2 2 2 2 2 2 2 2;
       GROUPS z;
       MEASUREMENT group;
       RHO PRIOR = 1;
       SEED 10;
       MAXITER 5000;
       CRITERION 0.000001;
RUN;
```

The output is described in the *PROC LCA & PROC LTA Users' Guide.*

The output will also include the file Categ_param in the WORK directory. Then, the Categ_param file is used to estimate the effect of the latent class variable on the distal outcome. Include the macro and enter the proper syntax in SAS.

```
%LCA_Distal_LTB (input_data = sasf.simdata_categ,
      param = Categ_param,
      distal = z,
      metric = 4,
      output_dataset_name = Categ_res);
```

The `input_data` argument identifies the data file. The `param` argument directs the macro to the parameters in the OUTPARAM file generated by PROC LCA. The `distal` argument identifies the distal outcome variable in the data set. The `metric` argument indicates that the distal outcome is categorical, and `output_dataset_name` names the macro's output.

5.4.3 Example Output

The onscreen output contains the class-specific distribution estimates for the distal outcome.

The SAS System

Class-Specific Distribution Estimates for z

Estimated Probabilities			
	P(Z=1)	P(Z=2)	P(Z=3)
Overall	0.3360	0.3480	0.3160
Class 1	0.3285	0.2760	0.3955
Class 2	0.4333	0.2844	0.2823
Class 3	0.4201	0.0717	0.5082
Class 4	0.2137	0.7189	0.0674
Class 5	0.2800	0.4034	0.3166

P(Z=k) in the table above is the probability that z equals k.

The Work directory will also contain the user-named output data set.

VIEWTABLE: Work.Categ_res

	Z	Density_Z	Z_C01	Z_C02	Z_C03	Z_C04	Z_C05
1	1	0.336	0.3285090978	0.4332610414	0.4201022034	0.2136970797	0.2799539667
2	2	0.348	0.2759568269	0.2844404218	0.0717381916	0.7188570681	0.4034437767
3	3	0.316	0.3955340753	0.2822985368	0.5081596051	0.0674458522	0.3166022565

The output conveys the association between membership in each latent class and the distal outcome. The first row represents the probability of z = 1 conditional on latent class membership, the second row indicates the probability of z = 2 conditional on latent class membership, the third row indicates the probability of z = 3 conditional on latent class membership, and so on. The column prob_Z indicates the marginal probability of each value of Z for the entire sample. The column Z_C01 represents the probability of each level of the distal outcome, given membership in latent class 1. Columns Z_C02 through Z_C05 represent the same thing for classes 2 through 5.

6 Empirical Demonstrations of the %LCA_Distal_LTB Macro for Multiple Groups

In this section, we first describe the structure of the data sets and the variables to be analyzed when there are multiple groups. Then, we illustrate how to estimate the distribution of the distal outcome within each latent class using the %LCA_Distal_LTB macro and describe the output of the macro. Section 6.1 describes use of the macro with a binary distal outcome. Continuous, count, and categorical outcomes are discussed in sections 6.2, 6.3, and 6.4, respectively.

6.1 Estimating a Binary Distal Outcome for Multiple Groups

Before attempting to complete the following example, please download the file *%LCA_Distal Examples* from the %LCA_Distal_LTB macro download page. Also, verify that you are running PROC LCA v.1.3.2 or higher.

6.1.1 Example Data

Below are 10 putative observations from the SAS data set **simdata_binary_group.sas7bdat**, which is contained in the *%LCA_Distal Examples* file available at http://methodology.psu.edu.

ID	Item001	Item002	Item003	Item004	Item005	Item006	Item007	Item008	Z	Educ
1	2	2	1	2	2	2	2	2	1	1
2	1	1	2	2	2	2	2	2	0	1
3	2	1	2	1	1	1	1	1	0	1
4	2	2	2	2	2	2	2	2	1	2
5	2	2	2	2	2	2	2	2	1	2
6	1	1	1	2	2	2	2	2	1	2
7	2	2	1	2	2	2	2	2	1	3
8	2	2	2	2	2	2	2	2	1	3
9	2	2	2	2	2	2	2	2	1	3
10	2	2	2	2	1	2	2	2	1	3

ID= subject's identification variable,

Item001,…, Item008= 8 items used to measure the latent class variable,

Z= the distal outcome (Note: binary distal outcome should be coded using 0s and 1s.)

Educ=the variable for multiple groups.

6.1.2 Example Syntax

Include a “libname” statement prior to running the macro to direct SAS to the data file.

```
libname sasf "S:\myfolder\";
```

Note: we suppose that the SAS data set exists in the folder `S:\myfolder\`. This path represents any user-specified folder.

Once the LCA model has been identified, we estimate the LCA model including the distal outcome *Z* as a covariate and Educ as the grouping variable using PROC LCA.

```
PROC LCA DATA = sasf.SimData_Binary_group OUTPARAM =
Binary_param OUTCOVB = Binary_covb;
ID id;
NCLASS 5;
ITEMS item001-item008;
CATEGORIES 2 2 2 2 2 2 2 2;
COVARIATES z;
GROUP educ;
RHO PRIOR = 1;
SEED 10;
MAXITER 5000;
CRITERION 0.000001;
RUN;
```

The output is described in the *PROC LCA & PROC LTA Users’ Guide.*

The output will also include the files Binary_param and Binary_covb in the WORK directory.

VIEWTABLE: Work.Binary_param

	Parameter Type	Group Number	Variable Name	Response Category	Estimate - Latent Class 1	Estimate - Latent Class 2	Estimate - Latent Class 3	Estimate - Latent Class 4	Estimate - Latent Class 5
1	GAMMA	1		.	0.323545	0.280305	0.099871	0.133263	0.163016
2	GAMMA	2		.	0.319432	0.279297	0.102542	0.132176	0.166553
3	GAMMA	3		.	0.385522	0.281460	0.104810	0.128846	0.099362
4	BETA	1		.	.	0.882074	0.784210	0.738405	0.701830
5	BETA	1	z	.	.	-1.15509	-2.49450	-1.95009	-1.61560
6	BETA	2		.	.	0.122223	-1.05199	-.749811	-.497958
7	BETA	2	z	.	.	-.534813	-.160154	-.257956	-.301342
8	BETA	3		.	.	-.442129	-1.27610	-1.30794	-1.03786
9	BETA	3	z	.	.	0.261012	-.058713	0.417214	-.909474
10	RHO	1	Item001	1	0.056834	0.906451	0.933252	0.977816	0.134035
11	RHO	1	Item002	1	0.066115	0.881024	0.871387	0.829746	0.146703
12	RHO	1	Item003	1	0.085492	0.850127	0.851169	0.961339	0.127452
13	RHO	1	Item004	1	0.078072	0.883143	0.949816	0.878651	0.055761
14	RHO	1	Item005	1	0.044679	0.087768	0.934019	0.686866	0.771810
15	RHO	1	Item006	1	0.047961	0.003221	0.856566	0.970499	0.713803
16	RHO	1	Item007	1	0.116454	0.062794	0.951015	0.002227	0.106036
17	RHO	1	Item008	1	0.115358	0.153006	0.771621	0.103327	0.103264
18	RHO	1	Item001	2	0.943166	0.093549	0.066748	0.022184	0.865965
19	RHO	1	Item002	2	0.933885	0.118976	0.128613	0.170254	0.853297
20	RHO	1	Item003	2	0.914508	0.149873	0.148831	0.038661	0.872548
21	RHO	1	Item004	2	0.921928	0.116857	0.050184	0.121349	0.944239
22	RHO	1	Item005	2	0.955321	0.912232	0.065981	0.313134	0.228190
23	RHO	1	Item006	2	0.952039	0.996779	0.143434	0.029501	0.286197
24	RHO	1	Item007	2	0.883546	0.937206	0.048985	0.997773	0.893964
25	RHO	1	Item008	2	0.884642	0.846994	0.228379	0.896673	0.896736
26	RHO	2	Item001	1	0.059517	0.905435	0.932155	0.977287	0.127166
27	RHO	2	Item002	1	0.069464	0.880979	0.871984	0.828643	0.136172

Binary_param

VIEWTABLE: Work.Binary_covb

	Col_1	Col_2	Col_3	Col_4	Col_5	Col_6	Col_7	Col_8
1	0.2414	-0.2073	0.1922	-0.1629	0.1922	-0.1632	0.2380	-0.1267
2	-0.2073	0.2080	-0.1600	0.1546	-0.1607	0.1530	-0.1908	0.1470
3	0.1922	-0.1600	0.2424	-0.2033	0.1920	-0.1639	0.2376	-0.1258
4	-0.1629	0.1546	-0.2033	0.2448	-0.1639	0.1488	-0.1918	0.1464
5	0.1922	-0.1607	0.1920	-0.1639	0.2520	-0.2110	0.2399	-0.1245
6	-0.1632	0.1530	-0.1639	0.1488	-0.2110	0.2303	-0.1924	0.1431
7	0.2380	-0.1908	0.2376	-0.1918	0.2399	-0.1924	0.3564	-0.1919
8	-0.1267	0.1470	-0.1258	0.1464	-0.1245	0.1431	-0.1919	0.2622
9	0.0519	-0.0340	0.0361	-0.0209	0.0352	-0.0216	0.0765	-0.0256
10	-0.0340	0.0628	-0.0216	0.0299	-0.0228	0.0290	-0.0344	0.0364
11	0.0361	-0.0216	0.1048	-0.0756	0.0230	-0.0135	0.0771	-0.0251
12	-0.0209	0.0299	-0.0756	0.1197	-0.0128	0.0239	-0.0335	0.0351

Binary_covb

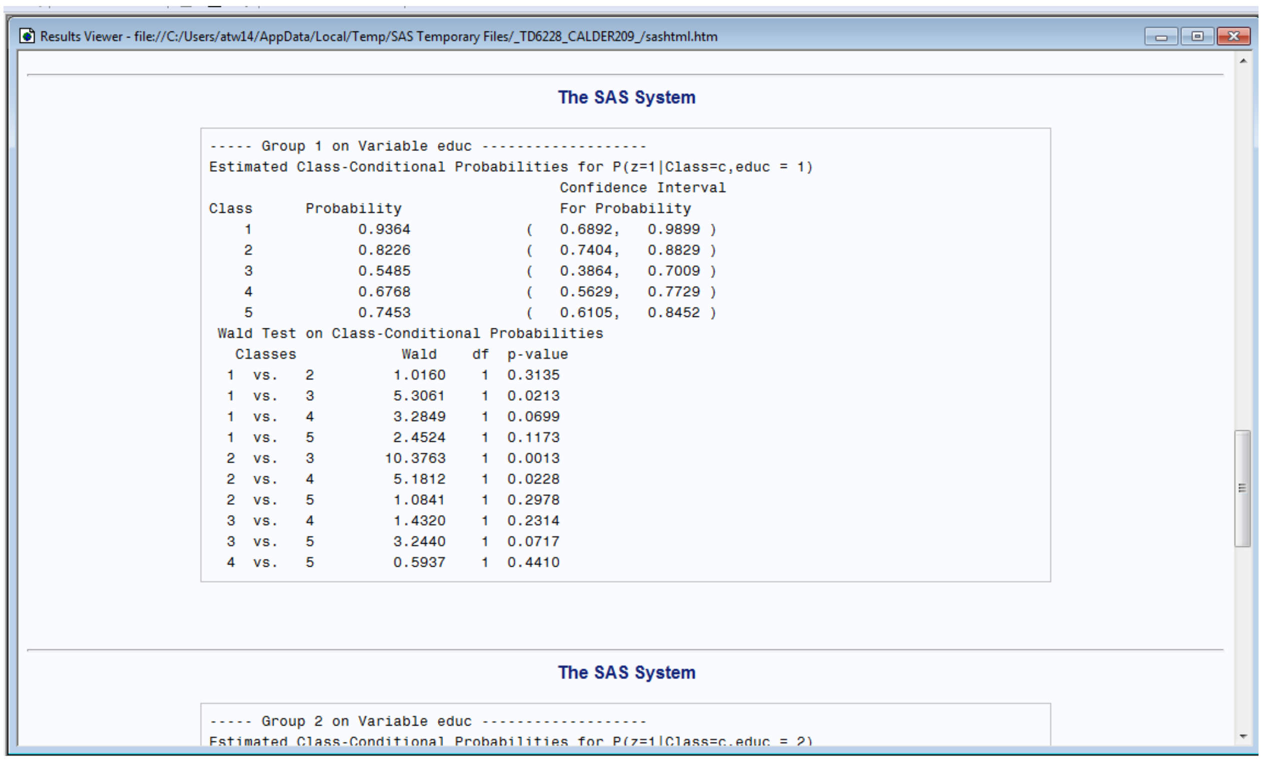
Then, the Binary_param file is used to estimate the distribution of the distal outcome within each latent class. The Binary_covb file is used to estimate the confidence intervals. Include the macro and enter the proper syntax in SAS.

```
%LCA_Distal_LTB (input_data = sasf.simdata_binary_group,
      param = Binary_param,
      distal = z,
      group = educ,
      metric = 1,
      covariance_beta = Binary_covb,
      output_dataset_name = Binary_res);
```

The `input_data` argument identifies the data file. The `param` argument directs the macro to the parameters in the OUTPARAM file generated by PROC LCA. The `distal` argument identifies the distal outcome variable in the data set. The `group` argument identifies the variable for multiple groups. The `metric` argument indicates that the distal outcome is binary. The `covariance_beta` argument indicates the data set generated by PROC LCA that contains the covariance matrix needed to generate confidence intervals, and `output_dataset_name` names the macro's output.

6.1.3 Example Output

Below is the onscreen output. It includes, for each group, the class-specific distribution estimates for the distal outcome, the estimated class-conditional probabilities, the Wald Test statistic on class-conditional probabilities and the p-values comparing class-conditional probabilities.



Results Viewer - file://C:/Users/atw14/AppData/Local/Temp/SAS Temporary Files/_TD6228_CALDER209_/sashtml.htm

The SAS System

```
----- Group 1 on Variable educ ------------------
Estimated Class-Conditional Probabilities for P(z=1|Class=c,educ = 1)
                                         Confidence Interval
Class      Probability                   For Probability
    1            0.9364             (    0.6892,    0.9899 )
    2            0.8226             (    0.7404,    0.8829 )
    3            0.5485             (    0.3864,    0.7009 )
    4            0.6768             (    0.5629,    0.7729 )
    5            0.7453             (    0.6105,    0.8452 )
 Wald Test on Class-Conditional Probabilities
   Classes             Wald    df   p-value
  1   vs.   2         1.0160    1   0.3135
  1   vs.   3         5.3061    1   0.0213
  1   vs.   4         3.2849    1   0.0699
  1   vs.   5         2.4524    1   0.1173
  2   vs.   3        10.3763    1   0.0013
  2   vs.   4         5.1812    1   0.0228
  2   vs.   5         1.0841    1   0.2978
  3   vs.   4         1.4320    1   0.2314
  3   vs.   5         3.2440    1   0.0717
  4   vs.   5         0.5937    1   0.4410
```

The SAS System

```
----- Group 2 on Variable educ ------------------
Estimated Class-Conditional Probabilities for P(z=1|Class=c,educ = 2)
```

The output in the user-named file conveys, for each group, the association between membership in each latent class and the distal outcome. The columns educ and Z indicate which values of the group and of distal outcome are indicated by each row. For each group, the first row represents the probability of z = 0 conditional on latent class membership, and the second row indicates the probability of z = 1 conditional on latent class membership. The column density_Z indicates the marginal probability of each value of Z for the entire sample. The column Z_C01 represents the probability of the distal outcome equaling 0 or 1, given membership in latent class 1. Columns Z_C02 through Z_C05 represent the same thing for classes 2 through 5.

VIEWTABLE: Work.Res

	educ	Z	Density_Z	Z_C01	Z_C02	Z_C03	Z_C04	Z_C05
1	1	0	0.2	0.0636180804	0.1774039271	0.4514974759	0.3232147687	0.2547325748
2	1	1	0.8	0.9363819196	0.8225960729	0.5485025241	0.6767852313	0.7452674252
3	2	0	0.516	0.4538220262	0.5865140526	0.4937288114	0.5181710051	0.5289930836
4	2	1	0.484	0.5461779738	0.4134859474	0.5062711886	0.4818289949	0.4710069164
5	3	0	0.534	0.543963037	0.4788364423	0.558486096	0.4400629878	0.7475864768
6	3	1	0.466	0.456036963	0.5211635577	0.441513904	0.5599370122	0.2524135232

The WORK directory will also contain four data sets, the user defined output plus Wald_pvalues, Wald_teststats, and Pi_by_class.

VIEWTABLE: Work.Wald_pvalues

	COL1	COL2	COL3	COL4	COL5	educ
1	.	0.3134795345	0.0212507683	0.0699194991	0.117342961	1
2	0.3134795345	.	0.0012764506	0.0228319817	0.2977784264	1
3	0.0212507683	0.0012764506	.	0.2314338364	0.071684598	1
4	0.0699194991	0.0228319817	0.2314338364	.	0.4409764815	1
5	0.117342961	0.2977784264	0.071684598	0.4409764815	.	1
6	.	0.0573159554	0.6628991044	0.4400935782	0.3731870146	2
7	0.0573159554	.	0.2951052334	0.4003859429	0.4536685338	2
8	0.6628991044	0.2951052334	.	0.8143185724	0.7191318162	2
9	0.4400935782	0.4003859429	0.8143185724	.	0.9054301436	2
10	0.3731870146	0.4536685338	0.7191318162	0.9054301436	.	2
11	.	0.2961879005	0.8650462043	0.2215381891	0.2951854647	3
12	0.2961879005	.	0.3558752995	0.652526948	0.1760022447	3
13	0.8650462043	0.3558752995	.	0.2637048968	0.3501220598	3
14	0.2215381891	0.652526948	0.2637048968	.	0.1387891611	3
15	0.2951854647	0.1760022447	0.3501220598	0.1387891611	.	3

Wald_pvalues

VIEWTABLE: Work.Wald_teststats

	COL1	COL2	COL3	COL4	COL5	educ
1	.	1.0159587171	5.3061088762	3.2849091207	2.4524316726	1
2	1.0159587171	.	10.376273643	5.1812432194	1.0841167797	1
3	5.3061088762	10.376273643	.	1.4320274869	3.244018854	1
4	3.2849091207	5.1812432194	1.4320274869	.	0.5937400928	1
5	2.4524316726	1.0841167797	3.244018854	0.5937400928	.	1
6	.	3.6133963371	0.1900210428	0.5960383539	0.79302494	2
7	3.6133963371	.	1.0961833246	0.7071668908	0.5614700725	2
8	0.1900210428	1.0961833246	.	0.0551585081	0.129326072	2
9	0.5960383539	0.7071668908	0.0551585081	.	0.0141145392	2
10	0.79302494	0.5614700725	0.129326072	0.0141145392	.	2
11	.	1.0912794383	0.0288844091	1.494379551	1.0958191304	3
12	1.0912794383	.	0.8523963685	0.2027291084	1.8310612722	3
13	0.0288844091	0.8523963685	.	1.249202256	0.8730147082	3
14	1.494379551	0.2027291084	1.249202256	.	2.1913329538	3
15	1.0958191304	1.8310612722	0.8730147082	2.1913329538	.	3

Wald_teststats

VIEWTABLE: Work.Pi_by_class

	PI_GIVEN_CLASS_BEST	PI_GIVEN_CLASS_UPPER	PI_GIVEN_CLASS_LOWER	CLASS	educ
1	0.936382	0.9898662	0.6892381	1	1
2	0.822596	0.8828623	0.7404392	2	1
3	0.548503	0.7009336	0.3863921	3	1
4	0.676785	0.7729397	0.5629365	4	1
5	0.745267	0.845219	0.6105139	5	1
6	0.546178	0.6323636	0.4571328	1	2
7	0.413486	0.4920067	0.3391308	2	2
8	0.506271	0.643797	0.3677898	3	2
9	0.481829	0.6019147	0.3638063	4	2
10	0.471007	0.5830731	0.3617892	5	2
11	0.456037	0.5305417	0.3834491	1	3
12	0.521164	0.5950711	0.4463177	2	3
13	0.441514	0.5802538	0.3113413	3	3
14	0.559937	0.6865098	0.4250589	4	3
15	0.252414	0.6252366	0.0639602	5	3

Pi_by_class

6.2 Estimating a Continuous Distal Outcome for Multiple Groups

NOTE: Currently, the macro assumes that if a distal outcome is a continuous variable, then it is homoscedastic; that is, the variance of the variable is the same across all latent classes.

Before attempting to complete the following example, please download the file *%LCA_Distal Examples* from the %LCA_Distal_LTB macro download page.

6.2.1 Example Data

In **simdata_conti_group.sas7bdat**, the data structure is similar to the data set in section 6.1 of this document. However, instead of binary values for z, the values are continuous.

ID	Item001	Item002	Item003	Item004	Item005	Item006	Item007	Item008	Z	Educ
1	2	1	2	2	2	2	2	2	-1.8513098	1
2	1	1	2	1	2	2	2	2	-0.5950087	1
3	2	2	2	1	1	1	2	2	1.55437269	1
4	2	1	2	2	2	2	2	2	0.89742276	2
5	1	1	1	1	2	2	2	2	-0.3121734	2
6	1	1	1	1	2	2	2	2	-1.5068341	2
7	2	2	1	2	2	2	2	2	0.73713821	3
8	1	1	1	2	1	1	2	2	1.8747736	3
9	1	1	1	1	1	1	2	1	-0.0463611	3
10	1	1	1	1	1	1	1	1	-0.1706686	3

where

ID= subject's identification variable,

Item001,…, Item008= 8 items used to measure the latent class variable,

Z= the distal outcome (in this case a continuous distal outcome) and

Educ= the variable for multiple groups.

6.2.2 Example Syntax

Include a "libname" statement prior to running the macro to direct SAS to the data file.

```
libname sasf "S:\myfolder\";
```

Note: we suppose that the SAS data set exists in the folder `S:\myfolder\`. This path represents any user-specified folder.

Once the LCA model has been identified, we estimate the LCA model including the distal outcome *Z* as a covariate using PROC LCA.

```
PROC LCA DATA = sasf.SimData_Conti_group OUTPARAM = Conti_param;
          ID id;
          NCLASS 5;
          ITEMS item001-item008;
          CATEGORIES 2 2 2 2 2 2 2 2;
          COVARIATES z;
          GROUPS educ;
          RHO PRIOR = 1;
          SEED 10;
          MAXITER 5000;
          CRITERION 0.000001;
RUN;
```

Then, the Conti_param file is used to estimate the distribution of the distal outcome within each latent class. Include the macro and enter the proper syntax in SAS.

```
%LCA_Distal_LTB (input_data = sasf.simdata_conti_group,
      param = conti_param,
      distal = z,
      group = educ,
      metric = 2,
      output_dataset_name = Conti_res,
      out_density=try_density,
      plot_density=1
);
```

The `input_data` argument identifies the data file. The `param` argument directs the macro to the parameters generated in the OUTPARAM file generated by PROC LCA. The `distal` argument identifies the distal outcome. The `group` argument identifies the variable for multiple groups. The `metric` argument indicates that the distal outcome is continuous. The `output_dataset_name` names the macro's output. The `out_density` argument names the data set for conditional densities, and the `plot_density` argument enables the users to choose whether to plot the conditional densities or not.

6.2.3 Example Output

The onscreen output contains estimates of the mean and mode of the distribution of the distal outcome within each class and group, based on kernel density estimation and Bayes' theorem. This information is also contained in the user-named output data set.

Results Viewer - file://C:/Users/atw14/AppData/Local/Temp/SAS Temporary Files/_TD6228_CALDER209_/sashtml.htm

density_Z Z_C01 Z_C02
Z_C03 Z_C04 Z_C05

The SAS System

```
Class-Specific Distribution Estimates for z
Within group        1 on z:
```

Class	Mean	Mode
1	-0.3976	-0.2694
2	-0.2349	-0.1623
3	0.2400	0.3197
4	0.4917	0.6544
5	-0.0010	0.0385

```
Within group        2 on z:
```

Class	Mean	Mode
1	0.0246	-0.0160
2	-0.0248	-0.1276
3	-0.1043	-0.2392
4	0.2354	0.4465
5	0.0273	-0.0160

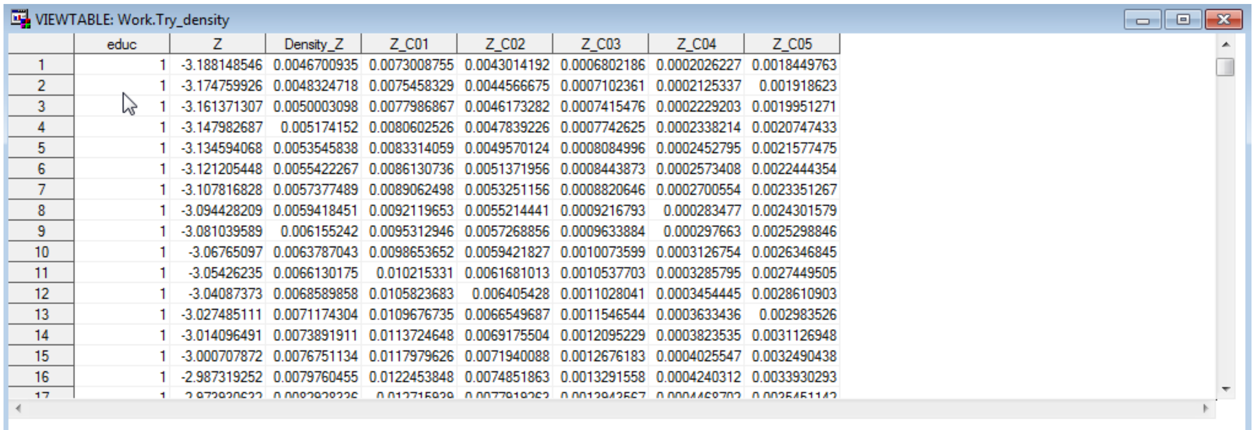
The WORK directory will also contain the user-named output data set and the conditional density data set.

VIEWTABLE: Work.Conti_res

	educ	M_Mode	Z_C01	Z_C02	Z_C03	Z_C04	Z_C05
1	1	1	-0.397579846	-0.234858795	0.2400495922	0.4916503915	-0.001003001
2	1	2	-0.269429475	-0.162320518	0.3196697874	0.6543852772	0.0385087759
3	2	1	0.0246288693	-0.02479146	-0.104316005	0.2353965621	0.0273430522
4	2	2	-0.015953022	-0.127580014	-0.239207006	0.4465016595	-0.015953022
5	3	1	-0.084562497	-0.007245186	0.355559102	-0.359267762	0.0061699482
6	3	2	-0.15975591	-0.102856412	0.2669903277	-0.401578778	-0.088631537

The output conveys, for each group, the association between membership in each latent class and the distal outcome. The columns educ and *Z* indicate which values of the group and of the distal outcome are indicated by each row. For each group, the first row (M_Mode=1) represents the mean response for each latent class, and the second row (M_Mode=2) represents the modal response for each latent class. The column Z_C01 represents the mean and mode of the distal outcome for class 1. Columns Z_C02 through Z_C05 represent the same thing for classes 2 though 5.

The macro will also generate a data set for the conditional density; the name of the data set is specified by the user in the `out_density` statement. It includes the conditional density of the distal outcome for each latent class in each group. It can be used to generate a plot of the conditional densities in SAS, Microsoft Excel, or other software.

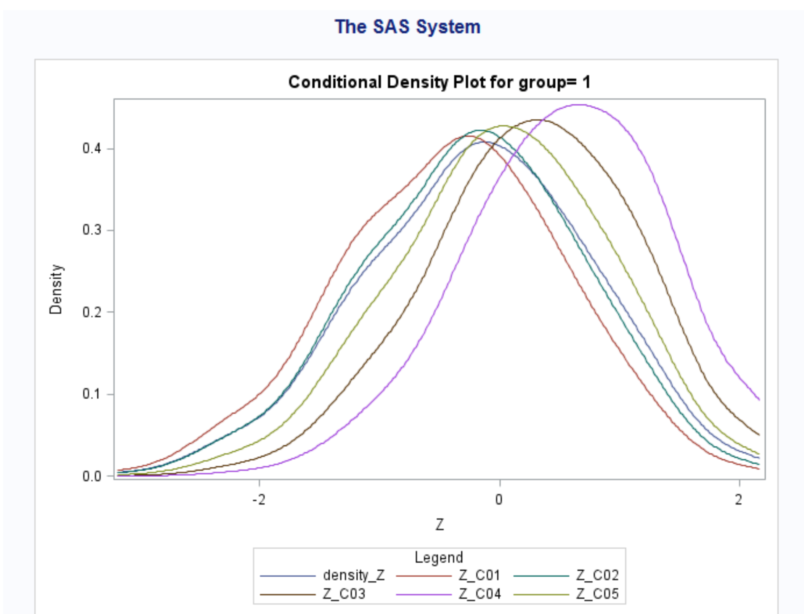


VIEWTABLE: Work.Try_density

	educ	Z	Density_Z	Z_C01	Z_C02	Z_C03	Z_C04	Z_C05
1	1	-3.188148546	0.0046700935	0.0073008755	0.0043014192	0.0006802186	0.0002026227	0.0018449763
2	1	-3.174759926	0.0048324718	0.0075458329	0.0044566675	0.0007102361	0.0002125337	0.001918623
3	1	-3.161371307	0.0050003098	0.0077986867	0.0046173282	0.0007415476	0.0002229203	0.0019951271
4	1	-3.147982687	0.005174152	0.0080602526	0.0047839226	0.0007742625	0.0002338214	0.0020747433
5	1	-3.134594068	0.0053545838	0.0083314059	0.0049570124	0.0008084996	0.0002452795	0.0021577475
6	1	-3.121205448	0.0055422267	0.0086130736	0.0051371956	0.0008443873	0.0002573408	0.0022444354
7	1	-3.107816828	0.0057377489	0.0089062498	0.0053251156	0.0008820646	0.0002700554	0.0023351267
8	1	-3.094428209	0.0059418451	0.0092119653	0.0055214441	0.0009216793	0.000283477	0.0024301579
9	1	-3.081039589	0.006155242	0.0095312946	0.0057268856	0.0009633884	0.000297663	0.0025298846
10	1	-3.06765097	0.0063787043	0.0098653652	0.0059421827	0.0010073599	0.0003126754	0.0026346845
11	1	-3.05426235	0.0066130175	0.010215331	0.0061681013	0.0010537703	0.0003285795	0.0027449505
12	1	-3.04087373	0.0068589858	0.0105823683	0.006405428	0.0011028041	0.0003454445	0.0028610903
13	1	-3.027485111	0.0071174304	0.0109676735	0.0066549687	0.0011546544	0.0003633436	0.002983526
14	1	-3.014096491	0.0073891911	0.0113724648	0.0069175504	0.0012095229	0.0003823535	0.0031126948
15	1	-3.000707872	0.0076751134	0.0117979626	0.0071940088	0.0012676183	0.0004025547	0.0032490438
16	1	-2.987319252	0.0079760455	0.0122453848	0.0074851863	0.0013291558	0.0004240312	0.0033930293
17	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

If the user decides to plot conditional densities in SAS (plot_density = 1), the macro will also generate a conditional density plot for each group. For some users, this will be displayed in the SAS Results Viewer. For others, it will be saved as a .png file. (If this happens, it will be recorded in the SAS log.) This differs because of different installations of Java on each individual user's machine.

Some users who are running SAS 9.1 or 9.2 will not be able to generate the conditional density plot because of incompatibility between SAS and certain versions of Java Runtime Environment. See Appendix A for instructions on how to fix this issue.



6.3 Estimating a Count Distal Outcome for Multiple Groups

Before attempting to complete the following example, please download the file *%LCA_Distal Examples* from the %LCA_Distal_LTB macro download page.

6.3.1 Example Data

In **simdata_count_group.sas7bdat**, the data structure is similar to the data set in section 6.1 of this document. However, the item *z* contains count responses with values from 0 to 5.

ID	Item001	Item002	Item003	Item004	Item005	Item006	Item007	Item008	Z	Educ
1	2	2	2	2	1	1	2	1	2	1
2	2	2	2	2	2	2	2	2	0	1
3	2	1	1	1	2	1	2	2	0	1
4	2	2	2	2	2	2	2	1	0	2
5	2	2	1	2	2	2	2	2	0	2
6	1	1	1	1	2	1	2	2	1	2
7	2	2	2	2	1	1	2	2	0	3
8	1	1	1	1	2	2	2	1	0	3
9	2	2	1	1	2	2	2	2	1	3
10	2	2	2	2	2	2	2	2	1	3

ID= subject's identification variable,

Item001,…, Item008= 8 items used to measure the latent class variable,

Z= the distal outcome (in this case a COUNT distal outcome) and

Educ= the variable for multiple groups.

6.3.2 Example Syntax

Include a "libname" statement prior to running the macro to direct SAS to the data file.

```
libname sasf "S:\myfolder\";
```

Note: we suppose that the SAS data set exists in the folder `S:\myfolder\`. This path represents any user-specified folder.

We estimate the LCA model including the distal outcome *Z* as a covariate using PROC LCA and specifying `educ` as the group variable.

```
PROC LCA DATA = sasf.SimData_Count_group OUTPARAM = Count_param;
         ID id;
```

```
                NCLASS 5;
                ITEMS item001-item008;
                CATEGORIES 2 2 2 2 2 2 2 2;
                COVARIATES z;
                GROUP educ;
                RHO PRIOR = 1;
                SEED 10;
                MAXITER 5000;
                CRITERION 0.000001;
RUN;
```

Then, the Count_param file will be used to estimate the distribution of the distal outcome within each latent class. Include the macro and enter the proper syntax in SAS.

```
%LCA_Distal_LTB (input_data = sasf.simdata_count_group,
     param = Count_param,
     distal = z,
     group = educ,
     metric = 3,
     output_dataset_name = Count_res);
```

The `input_data` argument identifies the data file. The `param` argument directs the macro to the parameters generated in the OUTPARAM file generated by PROC LCA. The `distal` argument identifies the distal outcome. The `group` argument identifies the variable for multiple groups. The `metric` argument indicates that the distal outcome is continuous, and `output_dataset_name` names the macro's output.

6.3.3 Example Output

The onscreen output contains the class-specific mean and mode estimates for the distal outcome, which are also provided in the user-named output data set.


Results Viewer - SAS Output

The SAS System

Class-Specific Distribution Estimates for z
Within group 1 on z:

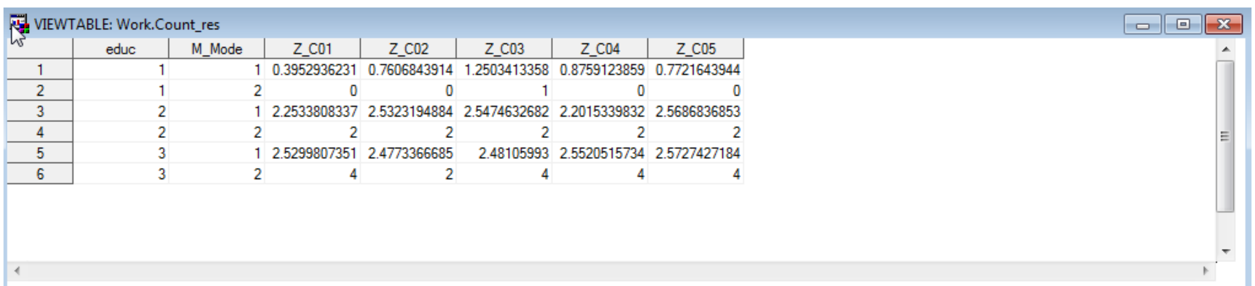
Class	Mean	Mode
1	0.3953	0
2	0.7607	0
3	1.2503	1
4	0.8759	0
5	0.7722	0

Within group 2 on z:

Class	Mean	Mode
1	2.2534	2
2	2.5323	2
3	2.5475	2
4	2.2015	2
5	2.5687	2

Within group 3 on z:

The WORK directory will also contain the user-named output data set.


VIEWTABLE: Work.Count_res

	educ	M_Mode	Z_C01	Z_C02	Z_C03	Z_C04	Z_C05
1	1	1	0.3952936231	0.7606843914	1.2503413358	0.8759123859	0.7721643944
2	1	2	0	0	1	0	0
3	2	1	2.2533808337	2.5323194884	2.5474632682	2.2015339832	2.5686836853
4	2	2	2	2	2	2	2
5	3	1	2.5299807351	2.4773366685	2.48105993	2.5520515734	2.5727427184
6	3	2	4	2	4	4	4

The output conveys, for each group, the association between membership in each latent class and the distal outcome. The columns educ and Z indicate which values of the group and of distal outcome are indicated by each row. For each group, the first row (M_Mode=1) represents the mean count for each latent class, and the second row (M_Mode=2) represents the modal count for each latent class. The column Z_C01 represents the mean and mode of the distal outcome for members of class 1. Columns Z_C02 through Z_C05 represent the same thing for classes 2 though 5.

6.4 Estimating a Categorical Distal Outcome for Multiple Groups

Before attempting to complete the following example, please download the file *%LCA_Distal Examples* from the %LCA_Distal_LTB macro download page.

6.4.1 Example Data

Below are the 10 putative observations from the SAS data set **simdata_categ_group.sas7bdat**, which is contained in the *%LCA_Distal Examples* file available at http://methodology.psu.edu.

ID	Item001	Item002	Item003	Item004	Item005	Item006	Item007	Item008	Z	Educ
1	2	2	2	2	1	2	2	2	2	1
2	2	1	2	2	2	2	2	2	3	1
3	1	1	2	1	1	2	2	1	2	1
4	2	1	1	1	2	1	2	2	3	2
5	2	2	2	1	2	2	2	2	1	2
6	2	2	2	2	1	2	2	2	1	2
7	1	1	1	1	1	2	2	2	2	3
8	2	2	2	2	2	2	2	1	3	3
9	2	2	2	2	2	2	1	2	2	3
10	2	2	2	1	1	1	2	1	1	3

where

ID= subject's identification variable,

Item001,…, Item008= 8 items used to measure the latent class variable,

Z= the distal outcome (Note: The categorical distal outcome should be coded using 1, 2, 3, …, g, where g = the number of categories.) and

Educ= the variable for multiple groups.

6.4.2 Example Syntax

Include a "libname" statement prior to running the macro to direct SAS to the data file.

```
libname sasf "S:\myfolder\";
```

Note: we suppose that the SAS data set exists in the folder `S:\myfolder\`. This path represents any user-specified folder.

First include and invoke the %create_group macro to create a new grouping variable in order to include the distal outcome in the `groups` statement in PROC LCA. (The macro was included in the download of the %LCA Distal macro.)

Direct SAS to read the macro code from the path, using a SAS %INCLUDE statement such as

```
%INCLUDE “S:\myfolder\create_group.sas”;
```

Then, invoke the macro.

```
%create_group(
input_data=sasf.simdata_categ_group,
group = educ,
distal=z,
output_data=newg
);
```

This will create a new data set named ‘newg’, which has an additional variable ‘new_educ’ that contains the re-defined grouping information. We estimate the LCA model using PROC LCA by specifying `new_educ` as the group variable using the new data set ‘newg’. Note that measurement invariance across groups should be imposed using the `measurement` statement.

```
PROC LCA DATA = newg OUTPARAM = Categ_param;
       ID id;
       NCLASS 5;
       ITEMS item001-item008;
       CATEGORIES 2 2 2 2 2 2 2 2;
       GROUPS new_educ;
       MEASUREMENT group;
       RHO PRIOR = 1;
       SEED 10;
       MAXITER 5000;
       CRITERION 0.000001;
RUN;
```

The output is described in the *PROC LCA & PROC LTA Users' Guide*.

The output will include the file Categ_param in the work directory. Then, the Categ_param file will be used to estimate the effect of the latent class variable on the distal outcome. Include the macro and enter the proper syntax in SAS.

```
%LCA_Distal_LTB (input_data = newg,
      param = Categ_param,
      distal = z,
```

```
group = new_educ,
metric = 4,
output_dataset_name = Categ_res);
```

The `input_data` argument identifies the data file (the newly created data set). The `param` argument directs the macro to the parameters in the OUTPARAM file generated by PROC LCA. The `distal` argument identifies the distal outcome variable in the data set. The `group` argument identifies the variable for multiple groups (the newly created grouping variable).The `metric` argument indicates that the distal outcome is categorical, and `output_dataset_name` names the macro's output.

6.4.3 Example Output

The onscreen output contains the class-specific distribution estimates for the distal outcome.

Results Viewer - SAS Output

The SAS System

Class-Specific Distribution Estimates for z

Estimated Probabilities						
	Overall	Class 1	Class 2	Class 3	Class 4	Class 5
Group 1, Outcome 1	0.3360	0.3163	0.4266	0.4134	0.2217	0.3173
Group 1, Outcome 2	0.3480	0.2556	0.2971	0.1180	0.6655	0.3632
Group 1, Outcome 3	0.3160	0.4281	0.2763	0.4687	0.1128	0.3195
Group 2, Outcome 1	0.2760	0.2879	0.2513	0.1910	0.3568	0.2722
Group 2, Outcome 2	0.4580	0.4176	0.4842	0.4792	0.4503	0.4650
Group 2, Outcome 3	0.2660	0.2944	0.2644	0.3298	0.1929	0.2628
Group 3, Outcome 1	0.3020	0.2746	0.2547	0.3754	0.3165	0.3232
Group 3, Outcome 2	0.4780	0.5379	0.4241	0.4306	0.5204	0.4680
Group 3, Outcome 3	0.2200	0.1875	0.3212	0.1940	0.1631	0.2088

The Work directory will also contain the user-named output data set.

VIEWTABLE: Work.Categ_res

	educ	Z	Density_Z	Z_C01	Z_C02	Z_C03	Z_C04	Z_C05
1	1	1	0.336	0.3148332943	0.4252595364	0.4123061332	0.230987974	0.3183394207
2	1	2	0.348	0.2556103415	0.2969202053	0.1125473187	0.6527378139	0.3611472557
3	1	3	0.316	0.4295563642	0.2778202583	0.4751465481	0.1162742121	0.3205133237
4	2	1	0.276	0.2885601645	0.2526183478	0.1888228991	0.3511774507	0.2721878637
5	2	2	0.458	0.417549576	0.4827419164	0.4773030951	0.4546093397	0.4639778679
6	2	3	0.266	0.2938902595	0.2646397358	0.3338740058	0.1942132096	0.2638342684
7	3	1	0.302	0.275585349	0.2534692904	0.3746611032	0.3178917813	0.3214766608
8	3	2	0.478	0.537149766	0.4241250948	0.4304259203	0.5164009053	0.470458297
9	3	3	0.22	0.187264885	0.3224056148	0.1949129765	0.1657073135	0.2080650422

The output conveys, for each group, the association between membership in each latent class and the distal outcome. The columns educ and Z indicate which values of the group and of the distal outcome are indicated by each row. For each group, the first row represents the probability of z = 1 conditional on latent class membership, the second row indicates the probability of z = 2 conditional on latent class membership, and the third row indicates the probability of z = 3 conditional on latent class membership, and so on. The column Density_Z indicates the marginal probability of each value of Z for the entire sample. The column Z_C01 represents the probability of each level of the distal outcome, given membership in latent class 1. Columns Z_C02 through Z_C05 represent the same thing for classes 2 through 5.

7 References

Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling: A Multidisciplinary Journal, 20*, 1-20.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. New York, NY: Wiley.

Lanza, S. T., Dziak, J. J., Huang, L., Wagner, A. T., & Collins, L. M. (2015). *Proc LCA & Proc LTA users'guide* (Version 1.3.2). University Park: The Methodology Center, Penn State. Available from methodology.psu.edu

Lanza, S. T., Tan, X., & Bray, B. C. (2011) *A technical introduction: A model-based approach to latent class analysis with distal outcomes*. (Report No. 11-116). University Park: The Methodology Center, Penn State. Retrieved from methodology.psu.edu

Schafer, J. L. (1997) *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman and Hall/CRC.

Appendix A. Troubleshooting: The %LCA_Distal_LTB Macro Does Not Generate the Conditional Densities Plot

1. Verify that you are running SAS version 9.1 or 9.2. In these versions there is a compatibility problem between SAS and Java. (This is a known problem listed on the SAS website. See http://support.sas.com/kb/34/512.html if you want the details.)

NOTE: If you are running SAS version 9.3 or higher, do not follow these instructions. They will not help you. Instead, email mchelpdesk@psu.edu for assistance.

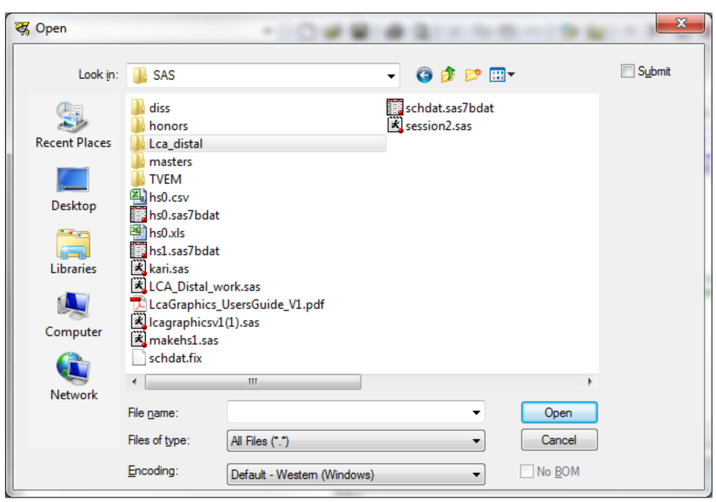
2. Login to your computer as an administrator.
3. Download and install Java Runtime Environment 5.0 Update 12 (This corresponds to JRE 1.5). The file is *jre-1_5_0_12-windows-i586-p.exe*, which can be downloaded free of charge from http://www.oracle.com/technetwork/java/javasebusiness/downloads/java-archive-downloads-javase5-419410.html#jre-1.5.0_12-oth-JPR
    - To download the file, you will need a (free) Oracle account.
4. Right-click the SAS icon and select "Run as administrator."
5. In the SAS Editor window, run the following code:

```
proc options option=config;
   run;
```

6. In the log, SAS will identify the location of the configuration file. It will look something like this:

```
CONFIG=C:\Program Files\SAS\SASFoundation\9.2\nls\en\sasv9.cfg
```

7. Copy the directory path and file name listed in the log (highlighted above).
8. Select the **File → Open Program** menu option in SAS.
9. In the window that opens, change the "Files of type" dropdown to "All Files."



10. Paste the file *name and path* you copied in step 7 into the “File name” field and click **Open**.
    - The configuration file will open in the Editor window.
    - For safety, save a copy of the existing configuration file before proceeding. (A damaged file can prevent SAS from functioning.)
11. Use “find” to search for “JRE.” Something similar to this will be displayed in the editor window.

```
-JREOPTIONS=(

        -DPFS_TEMPLATE=!SASROOT\tkjava\sasmisc\qrpfstpt.xml
        -Djava.class.path=C:\PROGRA~1\
         SASHome\SASVER~1\eclipse\plugins\SASLAU~1.JAR
        -Djava.security.auth.login.config=!SASROOT\tkjava\sasmisc\
         sas.login.config
        -Djava.security.policy=!SASROOT\tkjava\sasmisc\sas.policy
        -Djava.system.class.loader=com.sas.app.AppClassLoader
        -Dlog4j.configuration=file:/C:/Program%20Files/
         SASHome/SASFoundation/9.3/tkjava/sasmisc/sas.log4j.properties
        -Dsas.app.class.path=C:\PROGRA~1\SASHome\SASVER~1\eclipse\
         plugins\tkjava.jar
        -Dsas.ext.config=!SASROOT\tkjava\sasmisc\sas.java.ext.config
        -Dsas.jre.libjvm=C:\PROGRA~1\Java\JRE15~1.0_1\bin\client\jvm.dll
        -Dtkj.app.launch.config=C:\PROGRA~1\SASHome\SASVER~1\picklist
        -Xms128m
        -Xmx128m
        )
```

12. In the section that is highlighted in yellow, be sure that the text reads ”JRE15~1.0_1”. If it reads ANYTHING else, change it to JRE15~1.0_1.
    - If you are using 64-bit Windows, the section highlighted in cyan PROGRA~1 will need to be changed to PROGRA~2.
13. Save the file (**File → Save**).
14. Exit SAS. The next time you open SAS and run %LCA_Distal_LTB, you should be able to plot conditional densities.