Zero-Inflated Growth and Mixture Models Using Mplus

Alan C. Acock
alan.acock@oregonstate.edu
Department of HDFS
322 Milam Hall
Oregon State University
Corvallis, OR 97331
7/2008

This document and selected references, data, and programs can be downloaded from

http://oregonstate.edu/~acock/growth

*Alan C. Acock, Department of Human Development and Family Sciences, Oregon State University, Corvallis OR 97331 (alan.acock@oregonstate.edu) . This was supported in part by 1R01DA13474, The Positive Action Program: Outcomes and Mediators, A Randomized Trial in Hawaii and R305L030072 CFDA U.S. Department of Education; Positive Action for Social and Character Development. Randomized trial in Chicago, Brian Flay, PI. ; and R215S020218 CFDA, Uintah Character Education Randomized Trial, U.S. Department of Education. An earlier version of this was presented at Presented at the APA Workshop on New Methods for the Analysis of Family and Dyadic Processes, Center for Research on Families, University of Massachusetts, Amherst, October 15, 2006.

Introduction

Expectations that researchers utilize longitudinal data are becoming standard. Often, these data involve variables that are measured as counts or frequencies. Much of these data involves counts of events that are relatively rare and that include many observations with a count of zero. Proper analysis of these data involves understanding **two interdependent outcomes**.

- First, the **binary outcome** is whether the event occurs at all or not (physical conflict, social approval, criticism, abortion, birth, clinical depression) and what covariates increase or decrease the trajectory of the likelihood (probability) of this occurrence vs. nonoccurrence. This is referred to as growth in the onset of the outcome.
- The second outcome is the **count** of how often the event occurs (rate of occurrence). The covariates that predict the binary trajectory may be the same or different from those that predict the trajectory of the count. One of these outcomes may be normative and the other problematic. For example, the presence of nonviolent conflict in a relationship (binary outcome) may be a positive relationship characteristic, but a high frequency of nonviolent conflict in a relationship (count outcome) may be a negative relationship characteristic. Interventions may seek to modify one or both of these components and may have varying success with each of these goals.

Interest in these two types of outcomes without longitudinal data has a long history and there are excellent methods for their analysis (see, for example, King, 1989; Long & Freese, 2006). Major structural equation modeling programs (EQS, LISREL, & Mplus) have extended these methodologies to the analysis of growth trajectories (Bollen & Curran, 2006; Duncan, Duncan, & Stryker, 2006; Rabe-Hesketh & Skrondal, 2005). Rather than just predicting the likelihood of a binary event or the rate of occurrence of the event, with longitudinal data we predict their trajectories. Does the probability of a binary outcome occurring increase or decrease? Does the frequency of an outcome increase or decrease? What factors influence the growth trajectories of the binary and frequency components of the outcome? What are the consequences of these trajectories for distal outcomes?

- The occurrence (binary) and count outcomes are related processes.
- We estimate the rate of change in both aspects of the outcome variable.
- We empirically isolate subpopulations (latent classes) that have different growth trajectories. Our sample is a mixture from multiple populations in which each class or group has a different growth trajectory.
- We identify time invariant and time varying covariates that predict the initial level and rate of growth of both the binary and the count components.
- Distal outcomes can be influenced directly by the initial level and the rate of growth of both growth trajectories and these growth factors may mediate the effect of other predictors.

It may be easiest to appreciate this redirection in our research questions by considering a hypothetical example. This example is presented as illustrative and not as a developed research study. Researchers have been interested in conflict conduct of men in dating relationships.

- Let’s assume we identify adolescent couples near the start of a romantic relationship and interview them once a week for the next five weeks of their relationship.
- We are only interested in the response of the man and his reports of conflict conduct in the relationship in this hypothetical example.

The Binary Component

- The first question is—Does the behavior occur or not?
    - Our survey might ask whether he engaged in any conflict conduct in the relationship during the last week (binary outcome) and, if so, next ask him how often this happened.
    - Alternatively, we might ask how many days last week there was conflict and zero days would be part of our 0 – 7 set of response options.
- Conflict conduct for some men is what we call **always zero or structurally zero** meaning that they simply do not engage in conflict conduct.
    - These men would respond *no* to our first question.
    - These men would select the 0 days response option to the alternative question.
    - They would give the same answer no matter what week was in the question.
- There are other men who may answer *no* to our question simply because conflict did not happen in that particular week even though such conduct is part of their normal behavior repertoire.
- It is important to recognize the difference between always zero and not always zero responses when selecting an analytic strategy. The not always zero responses that just happening to be zero may conform to a random process
    - A Poisson distribution where we expect a substantial number of zeros just by chance.
    - A negative binomial distribution where we would expect somewhat more zeros than under a Poisson distribution.
- The **covariates that explain which men are *structurally zero* may be different from the factors that explain which men just happen to be zero** for the measured period.
    - Core values may explain being structurally zero.
    - Lack of recent stressful events may explain why a man just happened to be zero last week. With recent stressful events he might have had a high count (days last week) for conflict.
- A similar example might be asking adolescents if they had sexual intercourse in the last month.
    - Those who are abstinent will say no because this is a structural response.
    - Others will say no because they did not have the opportunity in the last month.

What happens to the binary component over time?

- Does the likelihood of conflict conduct increase, decrease, or follow a more complex trajectory?
- Does a bigger proportion of our sample become structural zeros over time?
- Does the proportion of our sample that are structural zeros get smaller over time?
- When studying the growth trajectory of the occurrence of conflict conduct, there are two parameters being estimated.
    - We need to identify the occurrence of conflict at the start of the relationship, α.

- This parameter is the initial level (with conventional coding) or the latent intercept growth factor of the growth process.
- Second, we need to identify the rate of change or the slope of the growth process of the likelihood of occurrence, β (not to be confused with a standardized Beta weight).
- This parameter is the latent slope growth factor.

- Different covariates may influence the latent intercept growth factor and the latent slope growth factor.
  - Adolescent males may have a higher level of acts of physical violence at the start of an intervention (initial level) than adolescent females, but
  - An intervention may have just as strong an influence for adolescent males on changes in this rate of behavior (slope) as it has on females.
  - Hence, gender will strongly influence the latent intercept growth factor, but
  - Gender will not have a significant effect on the latent slope growth factor.
- Beyond this, some personality characteristics might not have an influence on the initial level of the behavior, but might serve as major barriers to program effectiveness (slope).

Two types of predictors can be used when considering a growth trajectory.

- Some predictors are what we call **time invariant covariates** because they are the same across the duration of the study.
  - Examples include core psychological traits, ethnicity, prior conflict between his parents.
  - These time invariant variables can influence both his initial level of the likelihood and his rate of change in the likelihood.
- The second type of predictor is what we call **time varying covariates** because these may vary from one time to the next.
  - In our example of violent conflict a time varying covariate might be peer or parental approval of the relationship as measured each week.
  - The level of implementation of the program (possibly a continuous predictor) could be different at each wave.

Another research question involves distal outcomes. In our study of conflict in adolescent romantic relationships we might think our results have long term implication for distal outcomes.

- For example, the age when these adolescents will marry or how stable their marriage will be are examples of distal outcomes.
- A distal outcome may be influenced by
  - Time invariant variables,
  - Time varying variables,
  - The initial level of conflict conduct (intercept), and
  - The growth trajectory of conflict conduct (slope).
  - The growth trajectory of conflict conduct might also mediate the effects of time invariant predictors on the distal outcome.

The Count Component. In addition to the likelihood of conflict occurring, for those men who are not always zero, we are interested in how often the conflict occurs. A relationship that involves a high frequency of conflict poses special problems.

- The occurrence vs. non-occurrence of conflict may not have an influence on distal outcomes, but the frequency of the occurrence of conflict might.
- A goal in an intervention program may be to reduce the frequency of conflict even though the occurrence of some conflict persists.
- The adolescent who experiments a couple times with shop lifting is a very different problem than an adolescent who engages in such behavior several times each week.

The following is a list of variables we would need to consider for our hypothetical study of conflict in intimate relationships:

Growth Variable

- Adolescent-peer relationship conflict, measured each of the first 5 weeks of a relationship.
    - Binary Component
        - Coded as 1 if any and
        - 0 if none.
    - Count Component
        - A count of how often
        - Coded as 0 to *k*

Time invariant covariates such as:

- Gender of the child,
- Mother's education,
- Prior parent-adolescent level of conflict
- Prior parental conflict between mother and father

Time varying covariates might include

- Conflict between parents and adolescent the week prior to each interview.
- Conflict between parents with each other the week prior to each interview,

Potential Distal Outcomes

- Does the adolescent graduate from High School at "normal" age?
- Does adolescent have job related difficulties?
- Likelihood first marriage will end in a divorce within 5 years.
- Partner conflict when adolescent marries or cohabits.

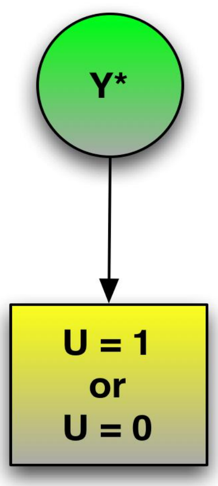
The Binary Component

The observed indicator is whether the man did or did not engaged in conflict conduct in the last week. This is a discrete, binary response of yes or no. The latent variables, both the latent intercept growth factor and the latent slope growth factor are his propensity to engage in this behavior. The latent variable is continuous having an expected value and variance. Consider Figure 1 in which:

- *U* is a binary response variable for whether there was any conflict or not.
- *Y** is a continuous latent response variable representing the propensity for peer conflict.
- τ is a *threshold* parameter determining that $Y = 1$ when $Y^* > \tau$ and $Y = 0$ when $Y^* \leq \tau$. τ is the minimum value of the propensity toward conflict that must be reached for the adolescent male to report conflict. This minimum value is the threshold, τ.

Figure 1
A Continuous Latent Factor and a Binary Response Variable and Threshold

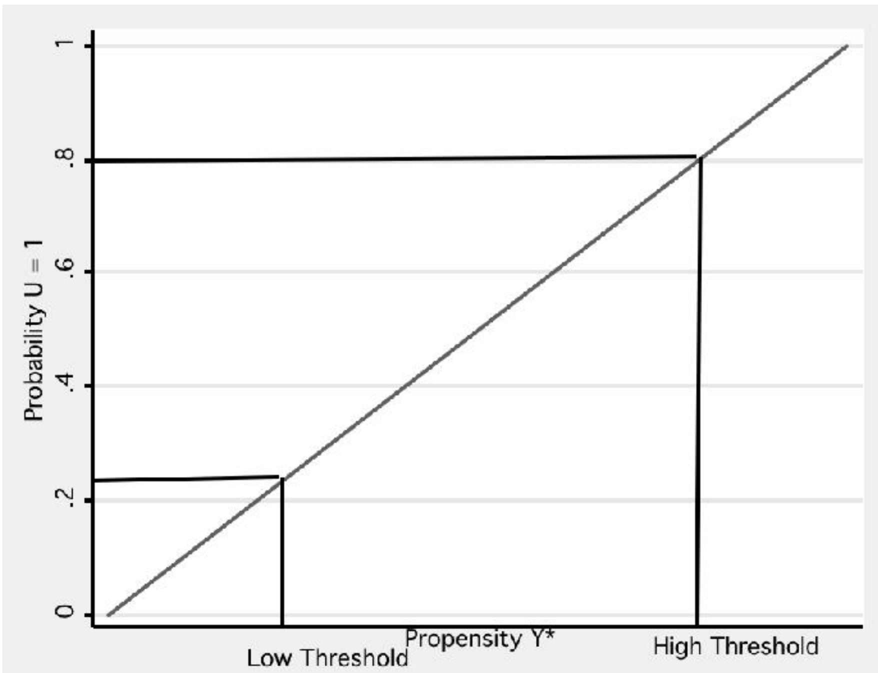
Rule: τ *is the threshold,*
where
$U = 1$ if $Y^* > \tau$,
$U = 0$ if $Y^* \leq \tau$



Rule: τ *is the threshold,*
where
$U = 1$ if $Y^* > \tau$,
$U = 0$ if $Y^* \leq \tau$

Another way of looking at this is as a simple bivariate plot. Remembering that the Latent variable is the predictor, we can say that if

- The threshold is low, then the probability of saying yes to U will be low and
- The threshold is high, then the probability of say yes to U will also be high.



An Empirical Example

We can estimate a growth model for negative feelings among a group of young children. We are using data from a major study of Positive Action in grades 1 to 4. This longitudinal study included 20 schools and over 100 classrooms (Flay, et al, 2006). For our purposes we will focus on those children who participated in the Positive Action Program.

- We are specifically interested in the number of these items for which they have a negative response, i.e., they report feeling bad when they engage in the positive behavior.
- We have their gender as a time invariant covariate because boys may start with a higher probability of some negative responses to positive behaviors and a higher count for negative responses
- It is not clear how this will change over the duration of the project. Is the trajectory different for boys than it is for girls?

We will show how to estimate the binary component by itself, the count component by itself, and then how to merge this using the zero-inflated models.

The Binary Component

We could estimate a growth curve for the binary component by itself (did they have any negative responses (coded 1), or not (coded 0). Estimating the growth model for the binary component involves a logistic model (Mplus also can do this using a probit model). Because of the difficulty interpreting the parameters or even odds ratios, Mplus lets us express the growth curve in terms of the probability that a person will engage in the behavior or not. The probability of a child reporting any negative response can be computed as:

$$\Pr(U=1)=\frac{e^B}{1+e^B}$$

since $e^B =$ the odds ratio

we can express

$$\Pr(U=1)=\frac{\text{odds ratio}}{1+\text{odds ratio}}$$

Wide Data vs. Long Data

Longitudinal models are often done with what is known as long data. Mplus can work with wide data as we are doing here, or with long data using its multilevel capabilities. Wide data simply means there is a row for each observation and a column for each variable. The variables include the score the person has on the outcome at each wave and any covariates that are used in the analysis.

Here is some of our data for the negative feelings data. Observation 213 reported

- 2 negative feelings at wave 1,
- 1 negative feeling at wave 2,
- 1 negative feeling at wave 3, and
- 1 negative feeling at wave 4.

Not all children were available for all years, e.g., observations 320 has no data for wave 1. Although not apparent in this small subset of data there is a consistent decrease in the count of negative feelings being expressed about positive action behavior.

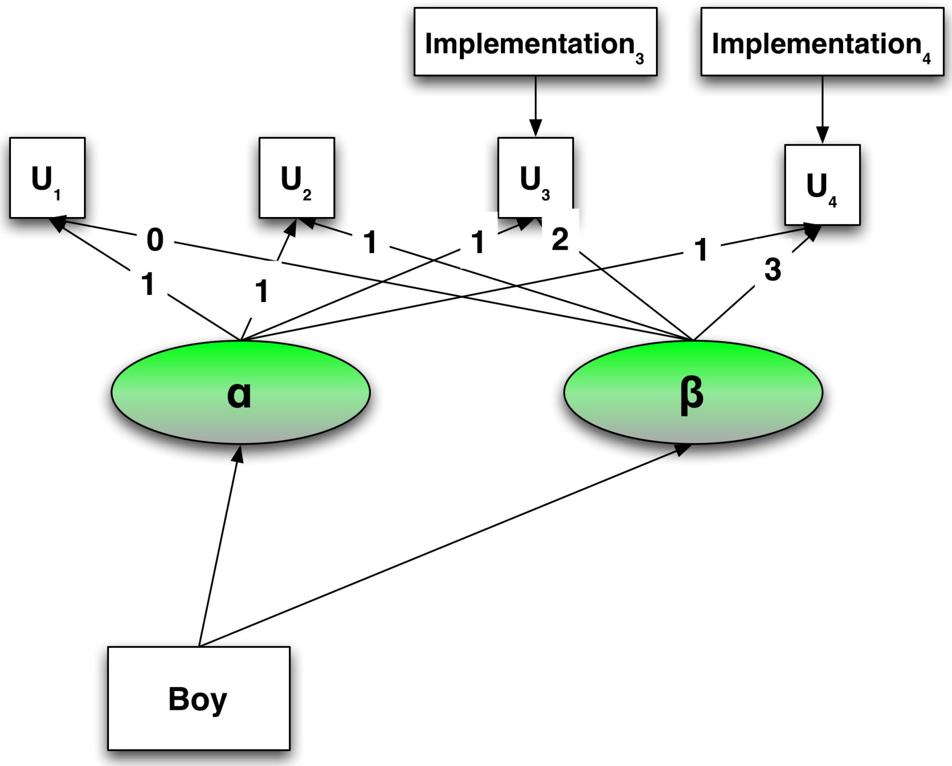
	s1flbadc	s2flbadc	s3flbadc	s4flbadc
203.	3	0	1	0
204.	.	1	1	0
207.	0	2	0	0
208.	.	.	1	0
211.	0	1	1	0
213.	2	1	1	1
320.	.	0	2	2

We can make the data binary by recoding them so a 1 indicates a positive count and a 0 indicates a count of zero. We still have missing values where the adolescent did not participate at a particular wave. For the binary growth model we use the following rule where U_{it} is a 0 or 1 for child *i* at wave *t*.

$$U_{it} = \begin{Bmatrix} 1 \text{ if } Y_{it} > 0 \\ 0 \text{ if } Y_{it} = 0 \end{Bmatrix}$$

	s1flbadd	s2flbadd	s3flbadd	s4flbadd
203.	1	0	1	0
204.	.	1	1	0
	[illegible]	[illegible]	[illegible]	[illegible]

Binary Growth Curve Model



- The U_i are the binary indicators (1's and 0's in our data).
- The α is the latent intercept growth factor or initial level in the probability (propensity) of expressing any bad feelings about positive behaviors.
- The β is the latent slope growth factor, or linear rate of change in the probability (propensity).

We used Latent Profile Analysis to generate two classes for waves 3 and 4 to represent consistently high implementation or consistently low implementation. These are time varying covariates to explain why the U_i*'s* are higher or lower than the overall trajectory would predict.

Boy represents a time invariant covariate. We expect boys may have a positive effect on the intercept (boys are going to have a higher initial count). We will also test whether boys have a different trajectory.

```
Title:          workshop binary growth.inp
                Stata2Mplus conversion for workshop_growth.dta
Data:           File is workshop_growth.dat ;
Variable:       Names are
                idnum s1flbadc s2flbadc s3flbadc s4flbadc male s1flbadd
                s2flbadd s3flbadd s4flbadd s1flbadm s2flbadm s3flbadm
                s4flbadm c3 c4 s3techer room ;
                Usevariables are male s1flbadd s2flbadd s3flbadd
                s4flbadd c3 c4 ;
                Categorical are s1flbadd s2flbadd s3flbadd s4flbadd ;
                Missing are all (-9999) ;
Analysis:       Estimator = ML ;
Model:          alpha beta | s1flbadd@0 s2flbadd@1 s3flbadd@2 s4flbadd@3 ;
                alpha on male ;
                beta on male ;
                s3flbadd on c3 ;
                s4flbadd on c4 ;
Output:  Patterns sampstat standardized tech1 tech8;
```

A workshop on growth curve modeling explaining how Mplus sets up the program is available at http://www.oregonstate.edu/~acock/growth. Here we will just explain key features and those that are different from the procedures for estimating a growth curve for a continuous variable. Under the `Variable:` section of the program we have the line `Categorical are s1 slbadd s2flbadd s3flbadd s4flbadd ;`.

- **When we specify maximum likelihood estimation, `Estimator = ML ;` in the `Analysis:` section, Mplus will determine if the dependent categorical variables are dichotomous and then use logistic regression modeling.**
- The latent growth curve factors, α and β are continuous, so logistic regression is not used when these are the dependent variables (e.g., `alpha on male` and `beta on male`).

Under the `Model:` section we have a single line that sets up the growth curve, `alpha beta | s1flbadd@0 s2flbadd@1 s3flbadd@2 s4flbadd@3 ;`

- The `alpha` is the arbitrary name we selected for the intercept and `beta` is the arbitrary name we selected for the slope.
- The intercept and slope both depend `on` the time invariant covariate, `male`, and the score at waves 3 and 4 on `s3flbadd` and `s4flbadd` are influenced by the level of implementation.
- This model does not have a distal outcome, but that would be easy to add, if it were appropriate to our model.

Selected output

Here is the observed probability of being in each categorical outcome variable (endogenous variables).

- Notice that 66.9% of the children expressed some negative feelings at wave 1 and
- This drops steadily to 62.8% at wave 2,
- 45.3% at wave 3, and
- Just 25.6% at wave 4.

This is clear evidence that any negative feelings about positive actions are getting less and less likely each year of the program. Examining these proportions, we can see that we might try other models such as a quadratic. The program gets off to a slow start reducing the proportion with negative feelings from 66.9% to 62.8% after the first year. After that the program's cumulative effect starts to show with larger drops in the subsequent years.

```
SUMMARY OF CATEGORICAL DATA PROPORTIONS
    S1FLBADD
      Category 1    0.331
      Category 2    0.669
    S2FLBADD
      Category 1    0.372
      Category 2    0.628
    S3FLBADD
      Category 1    0.547
      Category 2    0.453
    S4FLBADD
      Category 1    0.744
      Category 2    0.256
```

The following section gives information for evaluating the model. There is far less of this than with simple SEM models. We have no Chi-Square test of our overall model. We do not have the usual measures of goodness of fit such as the comparative fit index. These would have been available if we had not included observations with missing values. In this case *N* = 530, Mplus would have used a Weighted Least Squares Estimator with adjustments for means and variances (WLMV), and would have provide the various additional fit indicators. However, we still have an `AIC, BIC`, and `Adjusted BIC`. These, along with the degrees of freedom are sometimes useful for comparing competing models.

```
TESTS OF MODEL FIT
Loglikelihood
          H0 Value                          -2049.169
Information Criteria
```

```
Number of Free Parameters                    9
Akaike (AIC)                          4116.338
Bayesian (BIC)                        4160.390
Sample-Size Adjusted BIC              4131.806
```

Even though the indicators are binary variables, the latent variables are always continuous. At a threshold value on the latent variable you switch from predicting the child will not report negative feelings to predicting that the child will report negative feelings. Mplus, by default assumes these threshold values are constant for each wave. All that changes is the number of children above or below the threshold value.

The unstandardized effect of being `male` on the intercept, α, is .548 and this is highly significant with a $z = 2.98$, $p < .01$. It would be helpful to have a more substantive interpretation.

- The standardized Beta weight makes little sense when you have a binary predictor such as gender because its interpretation depends upon being one standard deviation higher or lower on gender and this is not conceptually sensible.
- Mplus output includes an additional standardized measure labeled `STDY standardization`. This is a partially standardized value (Stavig & Acock, 1981) for which only the latent growth factor variables are standardized.
    - Since the intercept and slopes are continuous variables, we can say that the partially standardized coefficient for the effect of gender on alpha, .464, indicates that boys have an intercept that is .464 standard deviations higher than girls.
    - Thus, boys have a propensity to respond negatively to actions that are positive that is nearly half a standard deviation higher than the propensity of girls.

Although boys are initially more likely to be negative than girls, the path of the slope, `beta on Male` is not significant, .03, $z = .37$, $\beta = .04$, *p ns* and the partially standardized coefficient is .077. This is an important finding because it says that

- However effective the program is at reducing negative feelings, it is approximately equally effective for both boys and girls.
- There is no interaction between gender and the intervention.

The time varying covariates are both significant. The higher the implementation level at wave 3 the lower the likelihood of reporting negative feelings. At wave 3 the $B = -.23$, $z = -2.71$, $p < .05$; β = -.06. At wave 4 the $B = -.64$, $z = -4.48$, $p < .001$, β = -.147. Because we are predicting binary outcomes, `s3flbadd` and `s4flbadd`, we need to be careful in how we interpret these parameter estimates.

- We are predicting a logit and this makes interpreting the *B's* more difficult and interpreting the β weights largely meaningless.

- We can compute the odds ratios, $e^{-.23} = .79$.
    - Thus, the odds of a child reporting any negative feelings is reduced by a factor of .79 when the child is in the high implementation category.
    - Alternatively, we could report that a child who is in the high implementation group is 21% (1-.79 × 100) less likely to report any negative feelings than a child in the low implementation group.
    - Similarly, at wave 4 the odds of a child in the high implementation group reporting any negative feelings is reduced by a factor of $e^{-.642} = .53$, a reduction of 47%.
    - The odds ratios where a binary outcome is predicted appear at the end of this section of output. It is clear that the implementation category is an important factor in what the program accomplishes.

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ALPHA    |
    S1FLBADD           1.000      0.000    999.000    999.000
    S2FLBADD           1.000      0.000    999.000    999.000
    S3FLBADD           1.000      0.000    999.000    999.000
    S4FLBADD           1.000      0.000    999.000    999.000

 BETA     |
    S1FLBADD           0.000      0.000    999.000    999.000
    S2FLBADD           1.000      0.000    999.000    999.000
    S3FLBADD           2.000      0.000    999.000    999.000
    S4FLBADD           3.000      0.000    999.000    999.000

 ALPHA      ON
    MALE               0.548      0.184      2.980      0.003

 BETA       ON
    MALE               0.033      0.088      0.371      0.711

 S3FLBADD   ON
    C3                -0.231      0.085     -2.714      0.007

 S4FLBADD   ON
```

```
   C4                 -0.642      0.144     -4.476      0.000

 BETA     WITH
   ALPHA              -0.344      0.217     -1.581      0.114

 Intercepts
   ALPHA               0.000      0.000    999.000    999.000
   BETA               -0.475      0.094     -5.078      0.000

 Thresholds
   S1FLBADD$1         -0.714      0.139     -5.137      0.000
   S2FLBADD$1         -0.714      0.139     -5.137      0.000
   S3FLBADD$1         -0.714      0.139     -5.137      0.000
   S4FLBADD$1         -0.714      0.139     -5.137      0.000

 Residual Variances
   ALPHA               1.321      0.529      2.497      0.013
   BETA                0.180      0.113      1.589      0.112
```

LOGISTIC REGRESSION ODDS RATIO RESULTS

```
 S3FLBADD   ON
   C3                  0.794

 S4FLBADD   ON
   C4                  0.526


STANDARDIZED MODEL RESULTS


STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ALPHA    |
   S1FLBADD            0.546      0.034     16.260      0.000
```

```
   S2FLBADD           0.577      0.049     11.759      0.000
   S3FLBADD           0.586      0.064      9.090      0.000
   S4FLBADD           0.566      0.077      7.336      0.000

BETA     |
   S1FLBADD           0.000      0.000    999.000    999.000
   S2FLBADD           0.207      0.046      4.509      0.000
   S3FLBADD           0.421      0.089      4.730      0.000
   S4FLBADD           0.609      0.114      5.332      0.000

ALPHA      ON
   MALE               0.232      0.073      3.187      0.001

BETA       ON
   MALE               0.038      0.103      0.372      0.710

S3FLBADD   ON
   C3                -0.055      0.020     -2.720      0.007

S4FLBADD   ON
   C4                -0.147      0.031     -4.824      0.000

BETA     WITH
   ALPHA             -0.705      0.187     -3.767      0.000

Intercepts
   ALPHA              0.000      0.000    999.000    999.000
   BETA              -1.120      0.308     -3.641      0.000

Thresholds
   S1FLBADD$1        -0.330      0.064     -5.191      0.000
   S2FLBADD$1        -0.349      0.069     -5.091      0.000
   S3FLBADD$1        -0.354      0.072     -4.922      0.000
   S4FLBADD$1        -0.342      0.073     -4.657      0.000

Residual Variances
   ALPHA              0.946      0.034     28.062      0.000
   BETA               0.999      0.008    125.435      0.000
```

```
STDY Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ALPHA    |
    S1FLBADD           0.546      0.034     16.260      0.000
    S2FLBADD           0.577      0.049     11.759      0.000
    S3FLBADD           0.586      0.064      9.090      0.000
    S4FLBADD           0.566      0.077      7.336      0.000

 BETA     |
    S1FLBADD           0.000      0.000    999.000    999.000
    S2FLBADD           0.207      0.046      4.509      0.000
    S3FLBADD           0.421      0.089      4.730      0.000
    S4FLBADD           0.609      0.114      5.332      0.000

 ALPHA      ON
    MALE               0.464      0.145      3.195      0.001

 BETA       ON
    MALE               0.077      0.207      0.372      0.710

 S3FLBADD   ON
    C3                -0.114      0.042     -2.725      0.006

 S4FLBADD   ON
    C4                -0.308      0.063     -4.851      0.000

 BETA     WITH
    ALPHA             -0.705      0.187     -3.767      0.000

 Intercepts
    ALPHA              0.000      0.000    999.000    999.000
    BETA              -1.120      0.308     -3.641      0.000

 Thresholds
    S1FLBADD$1        -0.330      0.064     -5.191      0.000
```

```
    S2FLBADD$1        -0.349      0.069     -5.091      0.000
    S3FLBADD$1        -0.354      0.072     -4.922      0.000
    S4FLBADD$1        -0.342      0.073     -4.657      0.000

 Residual Variances
    ALPHA              0.946      0.034     28.062      0.000
    BETA               0.999      0.008    125.435      0.000
```

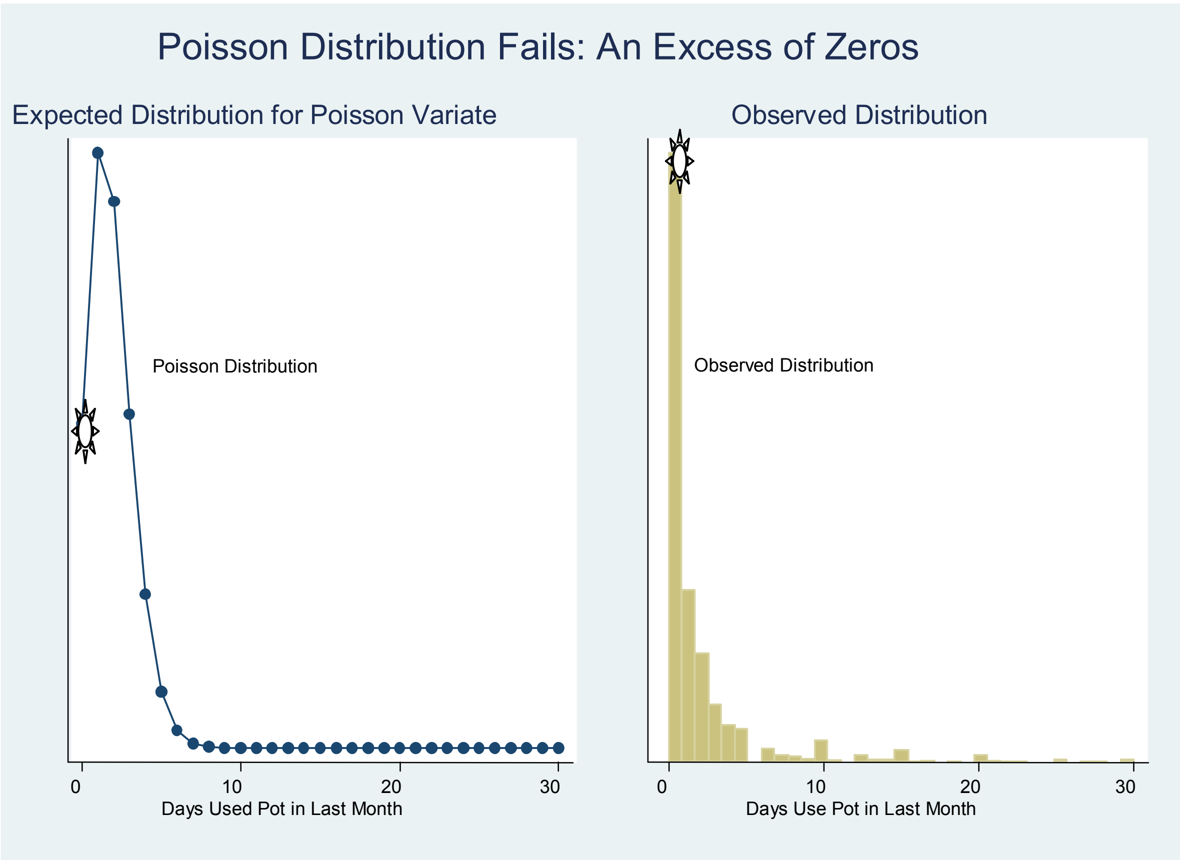
Graphs

Mplus does not make a graph of the estimated probabilities for each wave of data when there are covariates because the variances depend on the level of the covariates. If you want a series of graphs (e.g., boy/low intervention both wave 3 and wave 4, boy/high intervention both waves, girl/high intervention both waves, and girl/low intervention both waves) you would need to estimate the model treating each of these combinations as a separate group. Doing this, each group would not have any covariates and the `Plot:` section as described in the previous example would produce the four plots. An example, `workshop binary growth no covariates.inp`, is in the Appendix.

The Count Component

The Poisson distribution is not adequate to represent the data when there is an excess of zeros. At the risk of being confusing, we will skip back to the NLSY data temporally because it provides an excellent illustration. If we are considering the number of days adolescents used marijuana in the last month we are likely to have a large number of adolescents who report zero days, especially in the first year of our study when the adolescents are younger. A Poisson distribution would randomly generate a considerable number of zeros, but we are likely to have an excess beyond what would be generated by a random process assuming a Poisson distribution. Using the NLSY97 data, here is the actual observed distribution compared to the Poisson distribution for the first year of that dataset:

Figure 5
Poisson Variate Compared to Observed Distribution



The Poisson distribution looks great, except for underestimating the number of zeros—far more adolescents report no use of marijuana than the Poisson distribution would estimate. The Poisson distribution can be used to estimate proportion of observations with each count value. That is, what is the estimated proportion of participants who have a count of 0? A count of 1? A count of 2? Etc. In this example, the mean number of days is $M = 1.8345$. We can use this single parameter estimate to estimate the proportion for each possible number of days:

$$\Pr(Y = k) = \frac{e^{-\lambda}\lambda^{k}}{k!} \qquad (5)$$

where k = count

$$\lambda = E(Y) = Var(Y)$$

Example--What is Probability of exactly 0, 1, or 2 successes, if λ = 1.8345

$$\Pr(Y = 0) = \frac{e^{-1.8345} \times 1.8345^{0}}{0!} = e^{-1.8345} = .16$$

$$\Pr(Y = 1) = \frac{e^{-1.8345} \times 1.8345^{1}}{1!} = .29$$

$$\Pr(Y = 2) = \frac{e^{-1.8345} \times 1.8345^{2}}{2!} = .27$$

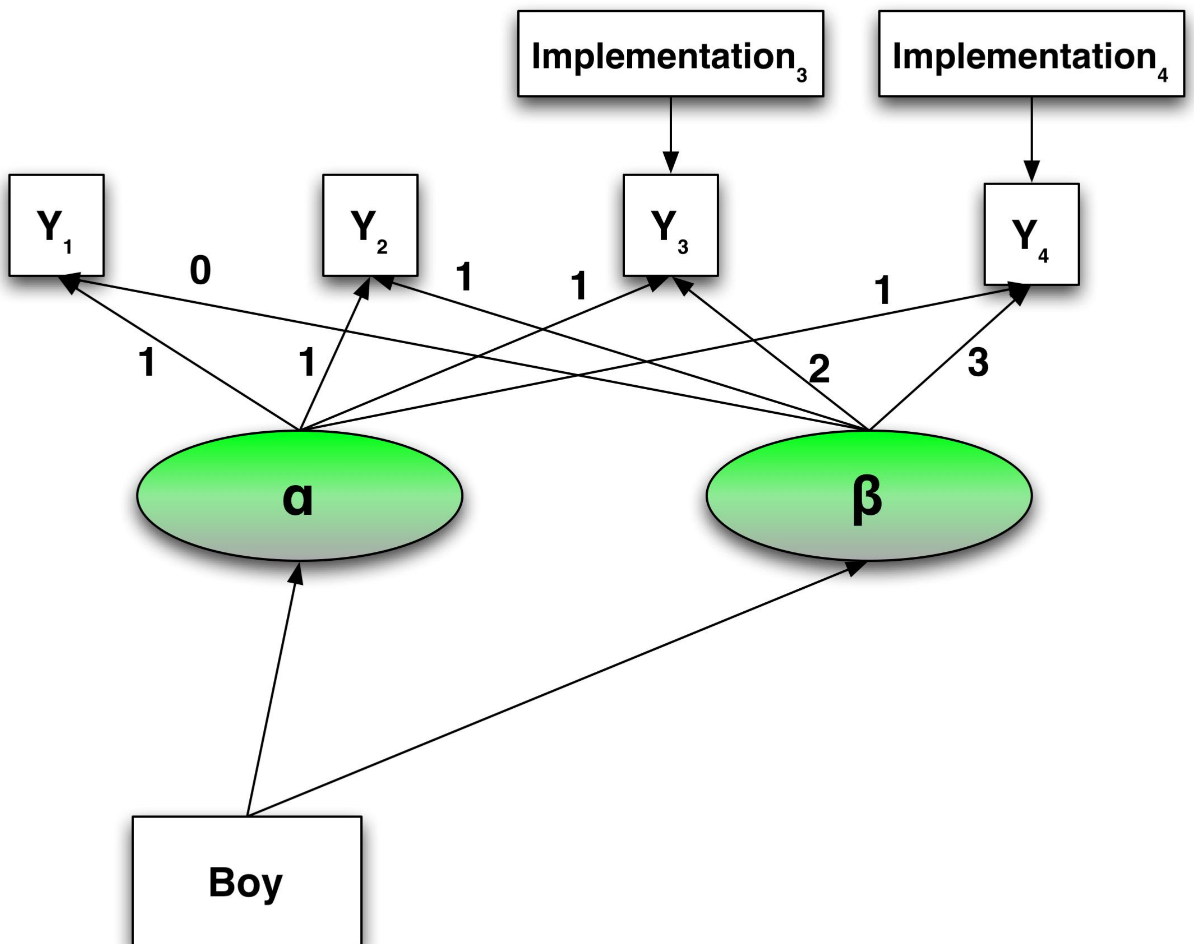
- The estimated proportion that would have a count of zero days, according to the Poisson model would be .16,
- The estimated proportion that would have a count of one day is .29.
- The estimated proportion that would have a count of two days is .27.

Our actual observed distribution has far more adolescents with a count of zero (.55) than would be estimated using a Poisson model (.16).

Although the Poisson may not provide an ideal fit the data when there is an excess of zeros, it has been widely used and can be applied with growth curves. The problem with the Poisson distribution is that it assumes there is only one parameter needed to describe the distribution and this is the mean. The variance is assumed to be equal to the mean. When there are an excess of zeros, the variances is typically greater than the mean. Mplus can also use a negative binomial distribution as the model for a count variable. The negative binomial distribution has a residual term added to the equation to represent the amount the variance is greater than the mean. With this added parameter, the negative binomial distribution will estimate a greater number of zero values (and a few more at the high end of the distribution as well. If this parameter is significant, the negative binomial distribution might be preferable to the more widely used Poisson distribution. Although the negative binomial model may be a better fit for a count variable than a Poisson model it is more complicated with the extra parameter and it doesn't explain the over dispersion. When we estimate the binary and count models simultaneously using a zero inflated model, we have the advantage of explaining the excess of zeros. For example, what covariates predict who will be in the structurally zero category. The zero inflated model is available in Mplus for both the Poisson model and the negative binomial model.

The model for our empirical example is the same as before except we are now dealing with indicators that are the count of the number of negative responses to positive behaviors. Because of this we use Y_i to represent the count indicators instead of U_i.

Figure 6
Poisson Growth Curve With Covariates (Fixed Effects)



First we will estimate the model without the covariates (always start simple!). We can do this with α and β as fixed effects or as random effects. Here is the fixed effects program:

```
Title:    workshop_growth_count_fixed.inp
Data:     File is workshop_growth.dat ;
Variable:
          Names are
               idnum s1flbadc s2flbadc s3flbadc s4flbadc
               male s1flbadd s2flbadd s3flbadd s4flbadd
               s1flbadm s2flbadm s3flbadm s4flbadm c3 c4
               s3techer room ;
          Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc ;
          Missing are all (-9999) ;
          Count are s1flbadc s2flbadc s3flbadc s4flbadc ;
Analysis: estimator = MLR ; !This is default since clearly not normal
          processors = 2 ;
Model:
          alpha beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
          alpha@0 ; !fixes intercept variance at 0--fixed effect
          beta@0 ; !fixes slope variance at 0--fixed effect for slope
Output:
          residual tech1 tech4 tech8;
Plot:
          Type = Plot3 ;
          Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
```

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 Means
    ALPHA              0.559      0.026     21.410      0.000
    BETA              -0.644      0.018    -34.867      0.000

 Intercepts
    S1FLBADC           0.000      0.000    999.000    999.000
    S2FLBADC           0.000      0.000    999.000    999.000
    S3FLBADC           0.000      0.000    999.000    999.000
    S4FLBADC           0.000      0.000    999.000    999.000

 Variances
```

```
    ALPHA                  0.000       0.000     999.000    999.000
    BETA                   0.000       0.000     999.000    999.000
```

Interpretation

- Intercept. We exponentiate the positive $\alpha = .559$ to get the **expected count** at the start of the program, $e^{.559} = 1.748$. We expect 1.748 negative responses to positive actions at the start of the program. Not surprisingly, this is statistically significant, $z = 21.410$, $p < .001$. The observed mean count at wave 1 is 1.526 so the estimate using the linear growth curve is not too far off the observed mean.

```
  Model Estimated Means
                S1FLBADC      S2FLBADC      S3FLBADC      S4FLBADC
                ________      ________      ________      ________
                   1.748         0.918         0.482         0.253
```

- Slope. A Poisson model of a count variable is estimating the log of the expected count, $\log(\lambda_t)$, where t is the wave, 0,1,2,3. Thus the model may be expressed as

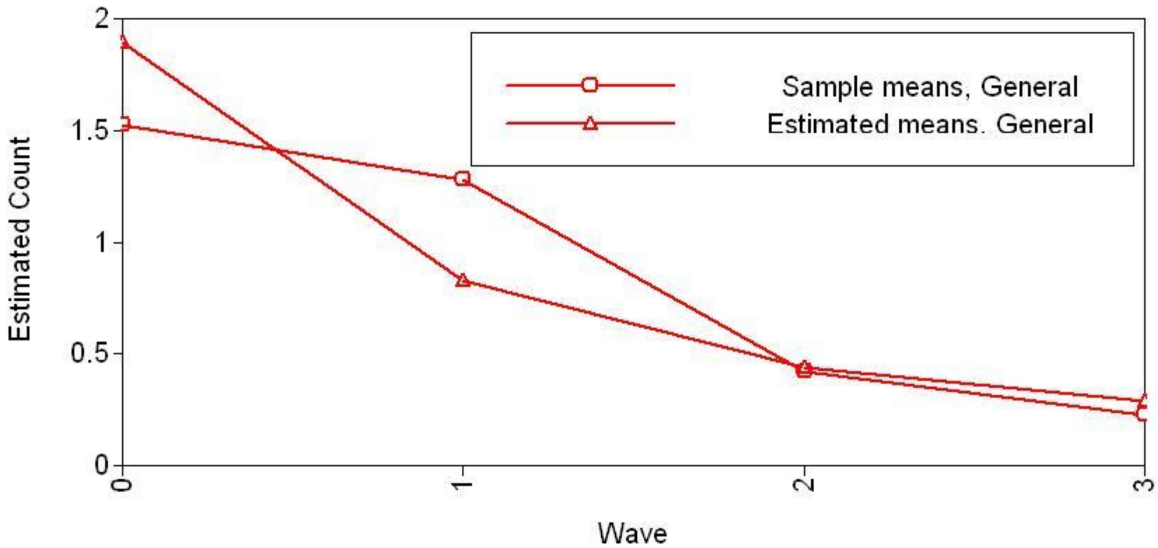
$$\log(\lambda_t) = \alpha + \beta X_t$$

where X_t is coded as 0,1,2,3 for the four waves. The negative $B = -.539$ for the slope indicates that the expected count of negative responses goes down each wave. To get the expected count for each wave we can use

$$\text{Expected count (initial)} = e^{\alpha}\times e^{\beta*0} = e^{\alpha} = 1.748$$
$$\text{Expected count (wave1)} = e^{\alpha}\times e^{\beta*1} = .918$$
$$\text{Expected count (wave2)} = e^{\alpha}\times e^{\beta*2} = .482$$
$$\text{Expected count (wave3)} = e^{\alpha}\times e^{\beta*3} = .253$$

Although we have a linear growth curve for $\log(\lambda_t)$, it is important to remember that the rate of change for the estimated count, λ_t, is nonlinear. This is illustrated in the following graph of these expected counts.

Figure 7
Plot of Sample and Estimated Count



In the program we asked for `residual` under the section labeled `Output:` and these expected counts are printed there. We obtain `Model Estimated Means` that match our computations:

```
Model Estimated Means
              S1FLBADC      S2FLBADC      S3FLBADC      S4FLBADC
              ________      ________      ________      ________
                1.748         0.918         0.482         0.253
```

Poisson Growth Curve Without Covariates (Random Effects)

We can also estimate the model allowing for the α and β to have positive variances, thus making it a random effects model. The program simply drops the commands `alpha@0` and `beta@0` so the variances of α and β are now free. When we do this we can also test if there is a significant covariance between α and β. If we were considering a random intercept only model we would leave the `beta@0` in the program.

Note that the z-tests for the variance of alpha and beta are problematic. This is because a variance has a lower limit of 0 and cannot be negative.

The program listing for `workshop count growth random effects.inp` appears in Appendix A.

```
Title:        workshop_count_growth_random effects.inp
Data:         File is workshop_growth.dat ;
Variable:     Names are
                  idnum s1flbadc s2flbadc s3flbadc s4flbadc
```

```
                 male s1flbadd s2flbadd s3flbadd s4flbadd
                 s1flbadm s2flbadm s3flbadm s4flbadm c3 c4
                 s3techer room ;
             Usevariables are s1flbadc s2flbadc s3flbadc
                 s4flbadc ;
             Missing are all (-9999) ;
             Count are s1flbadc s2flbadc s3flbadc s4flbadc ;
Analysis:    ALGORITHM = integration ; !Numerical integration
! can be very CPU intense. It is needed for random effects models
             processors = 2 ;
Model:       alpha beta | s1flbadc@0 s2flbadc@1 s3flbadc@2
             s4flbadc@3 ;
Output:
             residual tech1 tech4 tech8;
Plot:
             Type = Plot3 ;
             Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
```

The following are selected results:

```
TESTS OF MODEL FIT

Loglikelihood

          H0 Value                       -9378.619

Information Criteria

          Number of Free Parameters              5
          Akaike (AIC)                   18767.237
          Bayesian (BIC)                 18797.146
          Sample-Size Adjusted BIC       18781.259
            (n* = (n + 2) / 24)
MODEL RESULTS

                   Estimates     S.E.  Est./S.E.    Std      StdYX

 Beta WITH
```

```
    Alpha              -0.086    0.021     -4.056   -0.276   -0.276
 Means
    Alpha               0.396    0.028     14.194
    Beta               -0.842    0.021    -40.115
Variances
    Alpha               0.483    0.040     12.064
    Beta                0.199    0.016     12.077
```

We do not have a direct test of the difference between these two models.

- This model is more complex with 5 free parameters [α, β, Var(α), Var(β), and Cov(α,β)] compared to the fixed effects model that has only two free parameters (α and β).
- Normally, we would do a model with a random intercept and fixed slope between these two runs.
- The sample size adjusted BIC is 9,954.67 for the random effects model which is smaller than the 10,679.59 value for the fixed effects model. It is difficult to say which is better given the differences in free parameters. This is a big difference.
- The random effects solution has a highly significant variance for both α and β and this indicates making them fixed would not be a good idea. These z-tests are problematic because the lower limit on the variance is zero so a z-test should be viewed with caution.
- There is also a marginally significant covariance between the intercept and slope of -.040, $z = -1.690$, $p = .097$. Those that start with a higher initial count, have a more negative slope than those who start with a lower initial count.

The expected counts are not computed by the simple formula, $\log(\lambda_t) = \alpha + \beta X_t$, we used for the fixed effect model. With a random effects model estimating the expected count requires numerical integration over the growth factors. However, we can get these estimated counts from the graph which is comparable to the graph for the fixed effects model and these expected counts are reported in the `Model Estimated Means` when we ask for a `residual` analysis:

```
  Model Estimated Means
                 S1FLBADC      S2FLBADC      S3FLBADC      S4FLBADC
                 ________      ________      ________      ________
                   1.716         0.924         0.550         0.362
```

The Poisson Growth Curve without covariates and with random effects would have a similar interpretation for the α and β parameters to what we had for the fixed effects model.

Poisson Growth Curve With Covariates (Random Effects)

Analyzing the growth curves for the counts without including the covariates does not address our key questions about the effects of the covariates. To do this we will use our random effects model and simply add the covariates the `Usevariables are` command:

```
Title:        workshop_count_growth_random_effects_cov.inp
Data:         File is workshop_growth.dat ;
Variable:     Names are
                  idnum s1flbadc s2flbadc s3flbadc s4flbadc
                  male s1flbadd s2flbadd s3flbadd s4flbadd
                  s1flbadm s2flbadm s3flbadm s4flbadm c3 c4
                  s3techer room ;
              Usevariables are s1flbadc s2flbadc s3flbadc
                  s4flbadc male c3 c4;
              Missing are all (-9999) ;
              Count are s1flbadc s2flbadc s3flbadc s4flbadc ;
Analysis:     ALGORITHM = integration ;
              Integration = 20 ; !default is 7 integration points
! can run with even fewer and might for developing models and then
! use integration = 50 ; for the final run (and enjoy lunch)
              processors = 2 ;
              starts 40 2 ;
Model:        alpha beta | s1flbadc@0 s2flbadc@1 s3flbadc@2
                  s4flbadc@3 ;
              alpha on male ;
              beta on male ;
              s3flbadc on c3 ;
              s4flbadc on c4 ;
Output:       residual tech1 tech4 tech8;
Plot:
              Type = Plot3 ;
              Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
```

Here are selected results:

```
TESTS OF MODEL FIT

Loglikelihood
```

```
          H0 Value                         -4085.484
          H0 Scaling Correction Factor         1.247
            for MLR

Information Criteria

          Number of Free Parameters                9
          Akaike (AIC)                      8188.967
          Bayesian (BIC)                    8233.019
          Sample-Size Adjusted BIC          8204.435
            (n* = (n + 2) / 24)



MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ALPHA      ON
    MALE               0.278      0.091      3.055      0.002

 BETA       ON
    MALE               0.095      0.054      1.760      0.078

 S3FLBADC   ON
    C3                -0.295      0.063     -4.665      0.000

 S4FLBADC   ON
    C4                -0.827      0.118     -6.996      0.000

 BETA     WITH
    ALPHA             -0.253      0.054     -4.732      0.000

 Intercepts
    ALPHA             -0.098      0.073     -1.339      0.181
    BETA              -0.180      0.062     -2.898      0.004
    S1FLBADC           0.000      0.000    999.000    999.000
    S2FLBADC           0.000      0.000    999.000    999.000
    S3FLBADC           0.000      0.000    999.000    999.000
```

```
    S4FLBADC           0.000      0.000    999.000    999.000

 Residual Variances
    ALPHA              0.657      0.096      6.831      0.000
    BETA               0.238      0.037      6.357      0.000
```

As in the binary model,

- Gender has a significant effect on the intercept, α, with $B = .278$, $z = 3.055$, $p < .001$. Thus, boys have more negative responses to positive behaviors in the first wave.
- By contrast, gender does not have a significant effect on the slope, β with $B = .095$, $z = 1.760$, $p = .078$. Thus boys and girls do not differ significantly in the decline of negative responses.
- This is an important finding for the Positive Action Program that has a program goal of working as well for boys as it works for girls.
- The slope, β, is significant and negative, $B = -.181$, $z = 2.909$, $p < .01$. This supports the program's goal of reducing negative responses to positive actions.

The two time varying covariates, the level of implementation at waves 3 and wave 4, are both highly significant. For year 3, $B = -.295$, $z = -4.665$, $p < .001$. For Year 4, $B = -.827$, z = -6.996, $p < .001$. Thus the classrooms we classified as having high implementation of the program had fewer negative responses to positive actions than the classrooms we classified as having low implementation.

The output also reports the covariance between α and β, -.251, $z = -4.68$, $p < .001$. Those students who started very high on negative responses have a more negative slope than those who started out low on the initial count. This makes sense since students who start out very low face a floor effect where there is little or no room for improvement. Mplus allows for Poisson Growth Models that include censoring from below, but we will not present that here.

Putting it Together: Growth Models of Counts with an Excess of Zeros

We are now ready to put things together and estimate a model that contains both the growth in the binary outcome and the growth in the count outcome. Mplus offers several solutions to this goal. One approach is a **two-part model**. This estimates two models simultaneously:

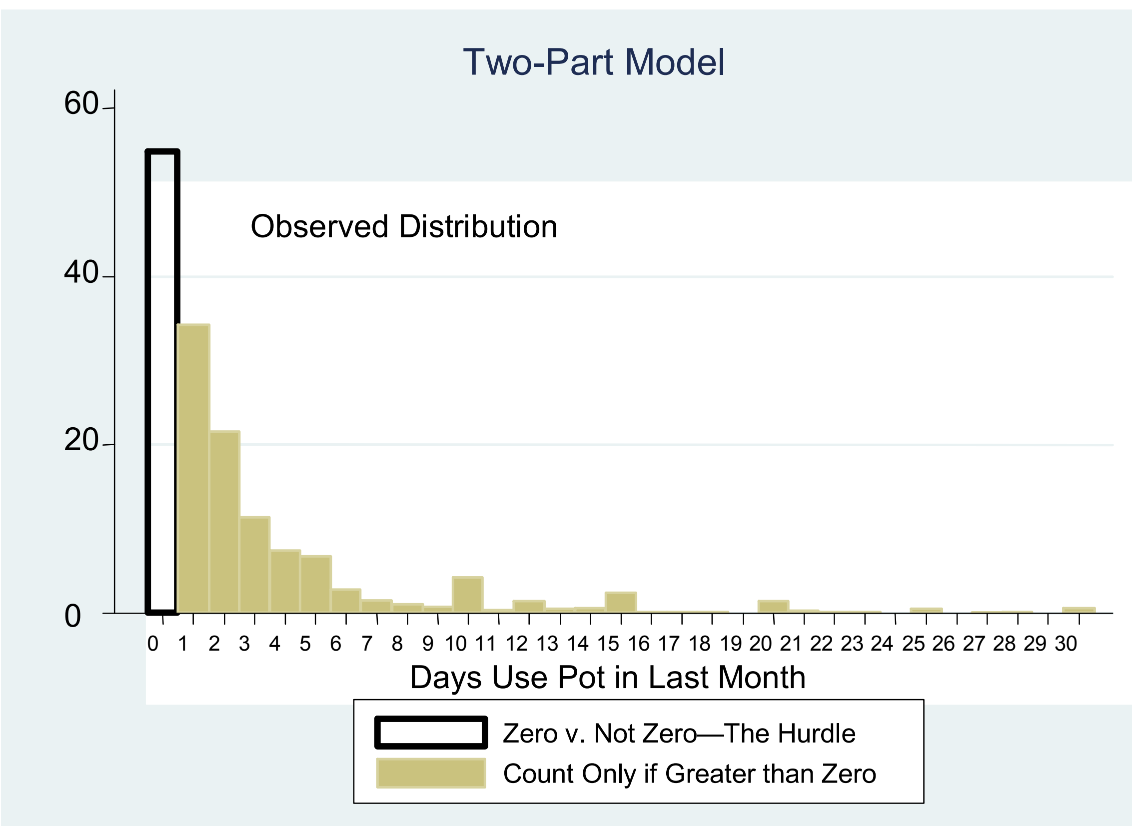
1. Part 1 models binary outcome of zero versus not zero days, asking what factors predict the presence or absence of the zeros. This uses a logistic model. We want to know what factors predict who will overcome whatever hurtle there is to engaging in the behavior.
2. Part 2 models the semi-continuous part for the observations that are greater than zero (zeros are treated as missing values in a two-part solution) using a Poisson model. What factors

predict the count for observations that are greater than zero. We use it to identify factors that predict the frequency of the behavior for those people who have overcome the hurdle.

An example of this approach is provided by King (1989). Returning to our marijuana example, this two-part model may be clarified by seeing a graph of the two distributions for one month.

- The bar that has no fill (the white bar) is the count of people who have a score of zero—they did not overcome the hurdle that month.
- These people are contrasted to those that have a count greater than zero—they overcame the hurdle. The bars that are filled show the distribution of the observations that have overcome the hurdle by reporting that they used marijuana in the last month at least once.
- Fully 55% of the people are zero and 45% of the people did use marijuana at least once in the last month. The bars that are filled show the distribution of the observations that have overcome the hurdle by reporting that they used marijuana in the last month at least once. Notice, the second part begins at 1, meaning that 55% of all observations with zeros are excluded from this analysis.

Figure 9
The Two-Part Model



The second approach is to distinguish between the zeros that are generated by a random process under the assumption of a Poisson distribution and the excess of zeros. This is a zero-inflated approach.

1. Some zeros are always zero. They are *structurally zero* if the outcome never occurs. The number of days you used marijuana is not a relevant question to an adolescent who never uses marijuana.
2. Some zeros are zero the last month just by chance, but might have been another value the previous month.

The zero-inflated Model estimates the two parts simultaneously using all of the data for both parts:

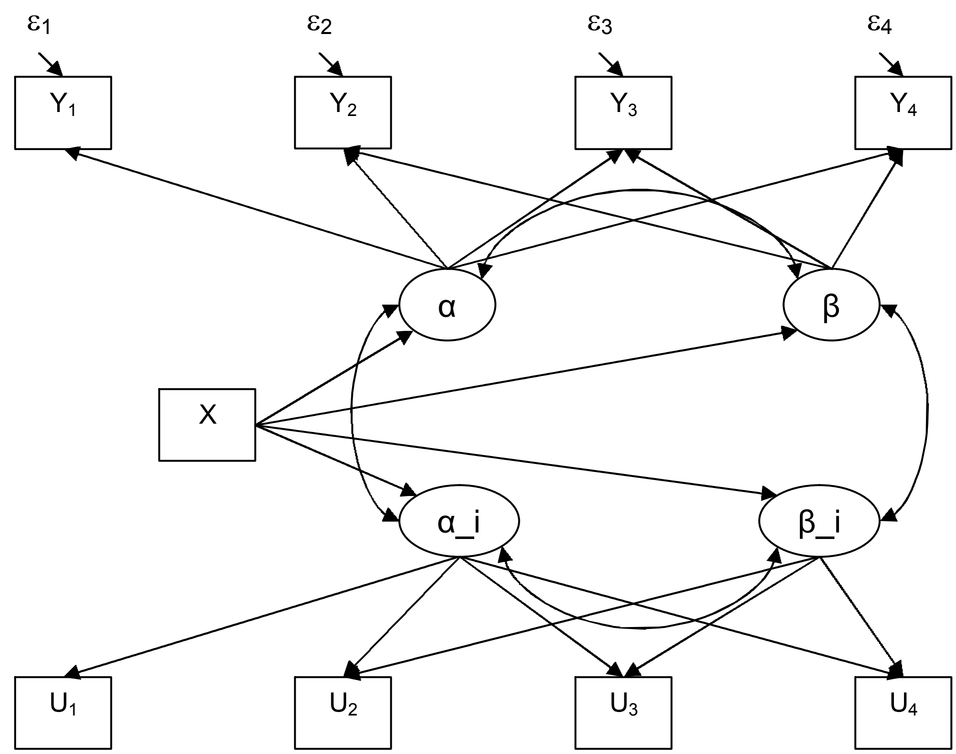
1. Part 1 models the zero-inflated part asking what factors predict the absence (called inflation) of the outcome using a logistic model.
2. Part 2 models all of the data including the zeros that are explained by a Poisson process, asking what factors predict the count even when the count is zero.

With both approaches it is important that both the onset and the count are explained in the same analysis and different covariates may be relevant to the two parts. In other words, factors that explain the onset of an outcome may be very different than factors that explain the frequency of an outcome. The same figure can be used to represent both the two-part approach and the zero-inflated approach

The two-part model is a reasonable approach. Some researchers like the way it clearly separates the two questions of occurrence vs. count. Others do not like the way it dismisses the people with a count of zero from the second analysis and ignores the fact that many of these people with a count of zero for the given time period, are not structural zeros, just chance zeros.

Figure 10
Model for Zero-Inflated Growth Model and Two-Part Model

Conceptual Model for Both Two-Part and Zero-Inflated Growth Curves



Zero-inflated Growth Curve Approach

We will focus on the zero-inflated growth curve approach. The two-part approach assumes that everybody who has a zero is the same and thereby fails to distinguish between those observations who are structural zeros and those who are zeros just by chance. The zero-inflated growth model seeks to make this distinction. A critique of the zero-inflated approach is that the distinction between structural zeros and chance zeros are based on assumptions about the appropriate distribution, Poisson or negative binomial.

Zero-Inflated Poisson Model without Covariates

Mplus offers several ways of estimating a zero-inflated model. The first one we will consider estimates the model directly. This is a traditional approach allowing for the growth parameters to have individual variation (random effects). This is computationally intensive and may take a very long time to converge even with fairly simple models. However, setting up the program involves only a few modifications to what was done for the count model.

- Under the `Variable:` section we have the command `Count are s1flbadc s2flbadc s3flbadc s4flbadc (i) ;`. This looks familiar except for the `(i)` on the end. This addition means that we are doing a zero-inflated count model using a Poisson model. For a negative binomial model we would use (nbi).
- When Mplus sees this, it looks for parallel growth models under the `Model:` section. The first growth model is unchanged; we have `alpha beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;`.
- The second growth model is for the binary outcome, we have `alpha_i beta_i | s1flbadc#1@0 s2flbadc#1@1 s3flbadc#1@2 s4flbadc#1@3 ;`
  - `alpha_i` is the intercept for the inflation part of the model and `beta_i` is the slope for the inflation part of the model. The choice of names is arbitrary.
  - Each variable name has `#1` (number sign 1) appended to it. Mplus uses `#1` to distinguish the inflation part of the model from the count part of the model.
  - No recoding is required since Mplus handles this when it sees `#1` at the end of a variable name.

The following assumptions are defaults for the inflation part of the model. The intercepts of the outcome indicators, `s1flbadc` to `s4flbadc`, are constrained to be equal representing a consistent threshold for linking the latent growth factors to the indicators.

- The mean of the intercept growth factor, α_i, is fixed at zero, but
- The mean of the growth factor, β_i is free,
- The variances of both α_i and β_i are free as is the covariance of α_i and β_i).

The following assumptions are defaults for the count part of the model. The intercepts of the outcome indicators, `s1flbadc to s4flbadc` are fixed at zero. The means, variances, and covariance of the growth factors, α and β, are free. Covariances of α_i with α and β and the covariances of β_i with α and β are also free.

The model is estimated using maximum likelihood with robust standard errors using numerical integration. This is a very CPU intensive program. To reduce the amount of time it took to estimate the model, I changed the default number for the integration. The default is 15 and I used 5. For a final publication it would be desirable to estimate the model with the largest number our computer can support. However, for anything but a very simple model this will quickly exceed the capabilities of most personal computers. There is a 64 bit version of Mplus, but I'm using the 32 bit version so I can run out of memory.

Here is the program:

```
Title:        workshop_zip_poisson_model.inp
              Linear growth model for a count outcome using
              The zero-inflated Poisson (ZIP) model
Data:         File is workshop_growth.dat ;
Variable:     Names are
               idnum s1flbadc s2flbadc s3flbadc s4flbadc
               male s1flbadd s2flbadd s3flbadd s4flbadd
               s1flbadm s2flbadm s3flbadm s4flbadm c3 c4
               s3techer room ;
              Usevariables are s1flbadc s2flbadc s3flbadc
               s4flbadc ;
              Missing are all (-9999) ;
              Count are s1flbadc s2flbadc s3flbadc
               s4flbadc (i);
  ! the (i) tells Mplus to set this up for an inflated model
  ! we need class(1) and type mixture to get the graphs we
  ! will use
Analysis:     processors = 2 ;
              Integration = 5 ;
              Starts 40 2 ;
Model:        Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2
               s4flbadc@3 ;
              Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1
               s3flbadc#1@2 s4flbadc#1@3 ;
  !Alpha is intercept for count part and Beta is slope
  !for count part
  !Alpha_i is intercept for inflated part and Beta_i is
  !the slope for inflated part
Output: Patterns residual tech1 tech8 ;
Plot:
         Type = Plot3 ;
         Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
```

The output is shown below. We now have parallel growth curves, one for the inflation part of the model and one for the count part of the model.

We will first consider the inflation part of the model. Here we are estimating the probability of a child being unable to assume any value except zero. That is that the child will have a structural zero score meaning that he or she has no negative response repertoire to positive behaviors. This can change over the waves of data. We would hope that the probability of a child being structurally zero will increase across the four years of this study. We obtain this using the `residual` option in the Output: command. We fine the results in the output file under the label `Residual Output` under the heading: `Model Estimated Inflation Probability.`

- Notice that the probability of being structurally zero increases dramatically.
- Although the first wave has many children who have a score of zero, Mplus estimates that only .020 of them are structurally zero,
- The second wave has a probability of .065 of being structurally zero,
- The third wave has a probability of .335, and
- The fourth wave has a probability of .516.

As shown in Figure 11, the Positive Action Program inoculates children from negative responses and by the fourth year the majority of the children are in the always zero class. This is also reflected in the estimated mean for β_i under the `Model Results`, $B = 2.353$, $z = 7.149$, $p < .001$. This is a large, positive value. The threshold for the zero-inflated part of the model is shown under the label of Intercepts. For each wave the threshold is -6.756, z = 3.45, $p < .05$. This large negative value will be confusing, unless we remember that the outcome for the inflated part of the model is predicting always zero rather than predicting a positive count. The more negative the threshold value the smaller the likelihood of being in the always zero class at the start. (display exp(-6.756) → .001.) Logistic regression usually is predicting the presence of an outcome, but now we are predicting its absence.

The results for the count growth curve has a marginally significant mean intercept of $\alpha = .267$, $z = 1.721$, $p = .09$ and a mean slope of $\beta = -.258$, $z = -1.032$, $p = .234$. The actual and estimated count for each wave are shown in Figure 12 that appears below. When we separate out the inflation part of this model, the count portion does not have a statistically significant intercept or growth factor (see Figure 12). Thus, it appears that the intervention is effective in inoculating an increasing proportion of the children from any negative responses (zero-inflation component), but not having a statistically significant effect on the count component. That is the program is not significantly reducing the frequency of this outcome except for significantly increasing the number of children who are structurally zero. Those who do not move into the structurally zero class are not significantly benefited. The output labeled `Model Results` does show Those who do not move into the structurally zero class are not significantly benefited.

Here is selected output:

```
Estimator                                                   MLR

TESTS OF MODEL FIT
```

```
Loglikelihood

          H0 Value                        -8958.486
          H0 Scaling Correction Factor        0.918
            for MLR

Information Criteria

          Number of Free Parameters              14
          Akaike (AIC)                    17944.972
          Bayesian (BIC)                  18028.716
          Sample-Size Adjusted BIC        17984.233
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit for the Count Outcomes**

          Pearson Chi-Square

          Value                            2505.791
          Degrees of Freedom                   9959
          P-Value                            1.0000

          Likelihood Ratio Chi-Square

          Value                             784.573
          Degrees of Freedom                   9959
          P-Value                            1.0000

Chi-Square Test for MCAR under the Unrestricted Latent Class Indicator
Model
for the Count Outcomes

          Pearson Chi-Square

          Value                             552.234
          Degrees of Freedom                   2097
          P-Value                            1.0000
```

```
          Likelihood Ratio Chi-Square

          Value                              658.342
          Degrees of Freedom                    2097
          P-Value                             1.0000
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
BETA     WITH
    ALPHA             -0.348      0.035     -9.864      0.000

 ALPHA_I  WITH
    ALPHA              0.648      0.448      1.447      0.148
    BETA              -0.968      0.383     -2.529      0.011

 BETA_I   WITH
    ALPHA             -1.110      0.237     -4.682      0.000
    BETA               2.596      0.397      6.536      0.000
    ALPHA_I           -9.785      3.050     -3.208      0.001

 Means
    ALPHA              0.269      0.036      7.396      0.000
    BETA              -0.301      0.041     -7.276      0.000
    ALPHA_I            0.000      0.000    999.000    999.000
    BETA_I             2.049      0.334      6.137      0.000

 Intercepts
    S1FLBADC#1        -6.796      0.886     -7.666      0.000
    S1FLBADC           0.000      0.000    999.000    999.000
    S2FLBADC#1        -6.796      0.886     -7.666      0.000
    S2FLBADC           0.000      0.000    999.000    999.000
    S3FLBADC#1        -6.796      0.886     -7.666      0.000
    S3FLBADC           0.000      0.000    999.000    999.000
    S4FLBADC#1        -6.796      0.886     -7.666      0.000
    S4FLBADC           0.000      0.000    999.000    999.000

 Variances
    ALPHA              0.485      0.045     10.676      0.000
```

```
    BETA                 0.797      0.077     10.306      0.000
    ALPHA_I             10.592      4.164      2.544      0.011
    BETA_I              13.226      3.386      3.906      0.000

RESIDUAL OUTPUT
     ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED)
           Model Estimated Inflation Probability
              S1FLBADC      S2FLBADC      S3FLBADC      S4FLBADC
              ________      ________      ________      ________
 1              0.034         0.042         0.304         0.470
           Model Estimated Means
              S1FLBADC      S2FLBADC      S3FLBADC      S4FLBADC
              ________      ________      ________      ________
 1              1.585         1.176         0.478         0.201
           Residuals for Means
              S1FLBADC      S2FLBADC      S3FLBADC      S4FLBADC
              ________      ________      ________      ________
 1             -0.059         0.103        -0.057         0.026
```

Figure 11
Graph of Binary Part of Zero-Inflated Model with No Covariates

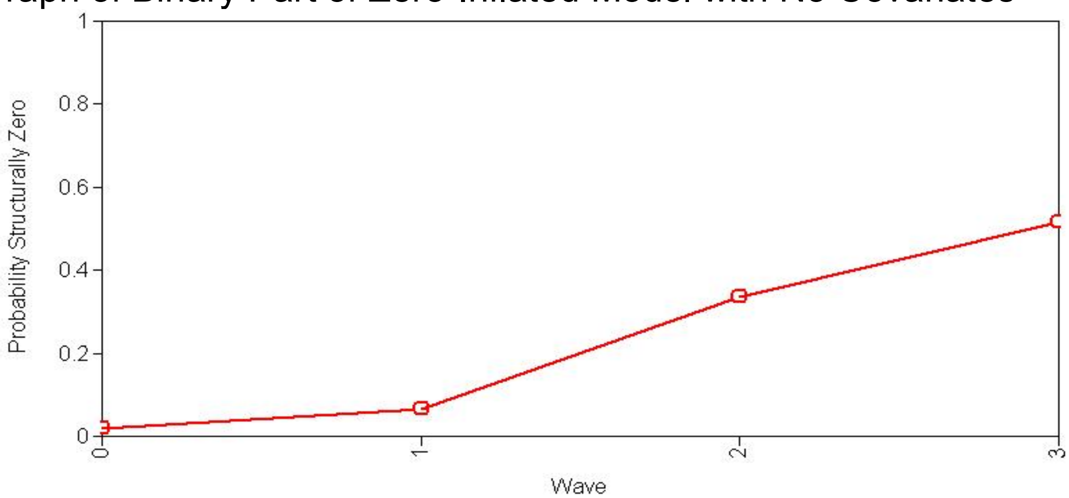
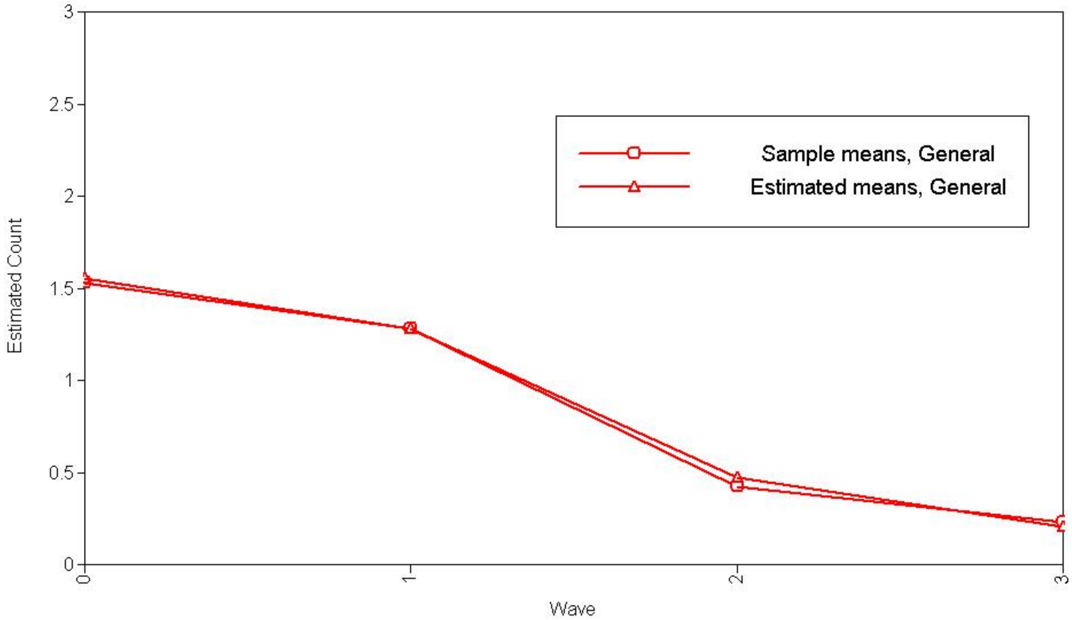


Figure 12
Graph of Count Part of Zero-Inflated Model with No Covariates



Zero-Inflated Poisson Model with Covariates

The next model we estimated uses the same approach but adds gender as a time invariant covariate and level of implementation at waves 3 and 4 as time varying covariates. The program is identical to the previous one except for additions to the section labeled `Model:`. When we add covariates we no longer get residual output because their values vary as a function of the covariate values. For the same reason, we do not get the graphic output. This analysis can be require a great amount of computer memory and is computationally intense. We limited the integration number.

Here we highlight in red the code for the covariates that we added to the section labeled `Model:`
.

```
Title:      workshop_zip_poisson_model_cov.inp
            Linear growth model for a count outcome using the
            zero-inflated Poisson (ZIP) model
Data:       File is workshop_growth.dat ;
Variable:   Names are
              idnum s1flbadc s2flbadc s3flbadc s4flbadc male s1flbadd
              s2flbadd s3flbadd s4flbadd s1flbadm s2flbadm s3flbadm
              s4flbadm c3 c4 s3techer room ;
            Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc
              male c3 c4;
            Missing are all (-9999) ;
            Count are s1flbadc s2flbadc s3flbadc s4flbadc (i);
```

```
  ! the (i) tells Mplus to set this up for an inflated model
Analysis:    Integration = 5 ;
             Starts 40 2 ;
             processors = 2
Model:       Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
             Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1 s3flbadc#1@2
               s4flbadc#1@3 ;
  !Alpha is intercept for count part and Beta is slope for count part
  !Alpha_i is intercept for inflated part and Beta_i is slope for
  !inflated part
             Alpha on male ;
             Beta on male ;
             Alpha_i on male ;
             Beta_i on male ;
             s3flbadc on c3 ;
             s4flbadc on c4 ;
Output:      Patterns sampstat tech1 tech8 stand;
```

Selected results appear below. We do not get the `Estimated Inflation Probability` or the `Model Estimated Means` (counts) when we have covariates because the values vary as a function of the covariate values. This makes it harder to provide a simple interpretation of the results. This time Mplus puts the means for α, β, α_i, and β_i as well as the thresholds for the inflation part of the model under the section labeled Intercepts. The inflation part of the model results shows that the thresholds are a large negative number, -7.316, $z = -3.876$, $p < .001$ and we now know that this means the initial probability of being in the structurally zero class is extremely small, $e^{-7.316} = .0007$. If the initial likelihood of being structurally zero approached a probability of zero, Mplus would fix the threshold at -15 since exp(-15) = 3.059e-07 which for practical purposes is zero and fixing the threshold becomes necessary for model identification. The slope for the inflation part of the model, β_i has a mean of 1.733 and this is statistically significant, z = 2.200, $p < .05$. This is a positive slope indicating an increase in the likelihood of a child being structurally zero.

For the count part of the model the intercept is -.083, $z = -1.087$, $p = .277$ and the slope $B = -.138$, $z = -2.020$, $p < .05$, thus the expected count decreases each wave.

The coefficients for the trajectory are smaller than when we did not have the covariates included meaning that the covariates may be important. When we examine the effect of gender on the intercepts and slopes we find that gender is only having a significant effect on the initial count, .294,

z = 2.876, $p < .01$. Thus, boys have a higher initial count, but do not have more who are structurally zero and the program is not significantly more effective for girls than it is for boys.
However, the time varying covariates, namely, the level of implementation of the program are both statistically significant. At wave 3, $B = -.286$, $z = -4.933$, $p < .001$ for an incidence risk ratio of, $e^{-.286} = .751$. At wave 4, $B = -.548$, $z = -4.357$, $p < .001$ for an incidence risk ratio of $e^{-.548} = .578$. Thus, the level of implementation directly reduces the expected negative responses by 24.9% at wave 3 and by 42.4% at wave 4. Again, these results suggest that the level of implementation of the Positive Action program is the pivotal variable.

The final selected output shown includes the R^2'a for each endogenous variable, i.e., the intercept and slope for both the inflation part and the count part of the model. All of these are small and insignificant. This is not unusual and my own experience has been that these are often dismal values. Clearly, they are worth noting because they demonstrate that there must be important variables that we have not included. I suspect that many researchers will not report these. The substantive interpretation of the parameter estimates are important evidence that something is happening.

The program does not provide the information needed to obtain the nice graphs we had when there were no covariates. To get graphs like this we would need to physically select subsamples such as

- Boys with low implementation for both wave 3 and wave 4 for one analysis and
- Boys with high implement for both waves for a separate analysis.

By substituting physical control for covariates we would be able to obtain graphs. Of course, our sample would become unacceptably small if there were many covariates treated this way.

```
THE MODEL ESTIMATION TERMINATED NORMALLY
TESTS OF MODEL FIT
Loglikelihood
          H0 Value                         -4020.962
          H0 Scaling Correction Factor         0.894
            for MLR

Information Criteria
          Number of Free Parameters               20
          Akaike (AIC)                      8081.924
          Bayesian (BIC)                    8179.818
          Sample-Size Adjusted BIC          8116.297
            (n* = (n + 2) / 24)
MODEL RESULTS
                                                     Two-Tailed
```

```
                       Estimate       S.E.  Est./S.E.    P-Value
 ALPHA      ON
    MALE                  0.294      0.102      2.876      0.004
 BETA       ON
    MALE                  0.034      0.076      0.451      0.652
 ALPHA_I    ON
    MALE                 -0.327      1.178     -0.278      0.781
 BETA_I     ON
    MALE                 -0.150      0.521     -0.288      0.774
 S3FLBADC   ON
    C3                   -0.286      0.058     -4.933      0.000
 S4FLBADC   ON
    C4                   -0.548      0.126     -4.357      0.000
 BETA     WITH
    ALPHA                -0.168      0.054     -3.139      0.002
 ALPHA_I  WITH
    ALPHA                 1.670      0.462      3.616      0.000
    BETA                  0.043      0.438      0.098      0.922
 BETA_I   WITH
    ALPHA                -0.524      0.245     -2.141      0.032
    BETA                  0.261      0.238      1.099      0.272
    ALPHA_I              -3.035      2.279     -1.332      0.183
 Intercepts
    ALPHA                -0.083      0.077     -1.087      0.277
    BETA                 -0.138      0.069     -2.020      0.043
    ALPHA_I               0.000      0.000    999.000    999.000
    BETA_I                1.733      0.788      2.200      0.028
    S1FLBADC#1           -7.316      1.888     -3.876      0.000
    S1FLBADC              0.000      0.000    999.000    999.000
    S2FLBADC#1           -7.316      1.888     -3.876      0.000
    S2FLBADC              0.000      0.000    999.000    999.000
    S3FLBADC#1           -7.316      1.888     -3.876      0.000
    S3FLBADC              0.000      0.000    999.000    999.000
    S4FLBADC#1           -7.316      1.888     -3.876      0.000
    S4FLBADC              0.000      0.000    999.000    999.000
 Residual Variances
    ALPHA                 0.614      0.101      6.070      0.000
    BETA                  0.244      0.041      5.980      0.000
    ALPHA_I               9.056      5.669      1.597      0.110
```

```
    BETA_I               1.636      1.164      1.405      0.160

R-SQUARE
    Latent                                              Two-Tailed
    Variable          Estimate       S.E.  Est./S.E.    P-Value

    ALPHA                0.034      0.023      1.445      0.148
    BETA                 0.001      0.005      0.226      0.821
    ALPHA_I              0.003      0.021      0.140      0.889
    BETA_I               0.003      0.024      0.144      0.885
```

Mixture Model Analysis using a Zero-Inflated Poisson Model (LCGA Poisson)

The second approach to zero-inflated Poisson model estimation involves the use of mixture models in order to find latent classes of participants. This approach can be done with the number of classes fixed at one. In this case it estimates the model for the overall sample and serves as a baseline for evaluate other solutions that have more than one class. The other solutions utilize the concept of mixture models by evaluating whether it is useful to think of a sample as being a mixture of two or more distinct subgroups. For example, our empirical example might have one class of children who start out with little likelihood of responding negatively to any positive action behaviors and they stay there. Such a class might be thought of has having a floor effect where the program is of little benefit simply because the response is not part of their initial behavioral repertoire. A second class of students might consist of children who start out at a higher likelihood of responding negatively, but the program works for them and they have a higher likelihood of being structural zero and a lower expected count each year. There might also be a third class of children for whom the program is ineffective and they maintain a consistent level of negative responses. Although it would not be expected, there might be some children in a fourth class that start out with a low level of negative responses but this level increases with each passing year of the program. Distinguishing such classes of children as a part of a zero-inflated Poisson model is an important extension. For example, there would be no need to include students who have a floor effect (first class above) in an evaluation of a program other than to acknowledge the number of children in this class who do not benefit from the additional attention. We might compare the classes by examining differences in other possible covariates. Are children in the second class, those that show marked improvement, more likely to be boys? Are they more likely to come from economically advantaged or disadvantaged families? Do they have greater parental support for education? Do they have greater parental involvement in the Positive Action Program? Such differences offer great opportunities to evaluate program effectiveness so that the program can be strengthened in subsequent modifications.

Mixture Model Using a Zero-Inflated Poisson Model with No Covariates

A One Class Solution. The first step is to estimate a LCGA Zero-Inflated Poisson Model using a single class solution and excluding any covariates. This model is in the same class as the growth models for Zero-Inflated Poisson Models and it serves as a baseline when we want to show that there is more than a single class. This solution will also provide a useful graph for both the zero-inflated component and the count component of the parallel growth model. The program utilizes a few new features. First, as with LCA and LPA, we add a `class = c(1)` to the section labeled `Variable:`. Subsequently we will change this command so we can have 2 or more classes. The first part of the `Model:` section adds a new line labeled `%Overall%` that defines the growth model for the overall sample and since this solution has only a single class, this is all we do. Here α, β, α_i, and β_i are defined as before.

The intercept for the inflation part of the model is automatically fixed at zero and the threshold values are used for estimating the odds and probability of inflation. For the count part of the model the means, variances, and covariance of α and β are estimated by default. For the inflation part of the model the means are computed for α_I, and β_i, but the variances of the inflation growth factors and their covariances are fixed at zero. This is because we conceptualize these as discrete classes. With these restrictions we can have different means for α, β, α_I, and β_i in each class we estimate a model that allows for more than one class. The default estimator is maximum likelihood with robust standard errors.

Here is the program.

```
Title:       workshop_mixture_poisson_model_no_cov_c1.inp
             Mixture Model Growth Analysis for a count outcome
             using a ZIP Model with no covariates and with just
             one class
Data:        File is workshop_growth.dat ;
Variable:    Names are
                idnum s1flbadc s2flbadc s3flbadc s4flbadc male
                s1flbadd s2flbadd s3flbadd s4flbadd s1flbadm
                s2flbadm s3flbadm s4flbadm c3 c4 s3techer room ;
             Usevariables are s1flbadc s2flbadc s3flbadc
                s4flbadc  ;
             Missing are all (-9999) ;
             Count are s1flbadc s2flbadc s3flbadc s4flbadc (i);
! the (i) tells Mplus to set this up for an inflated model
! we need class(1) and type mixture to get the graphs we will
! use
             Classes = c(1) ;
! this says there is a single class so it is not really looking
! for a mixture model. This just serves as a baseline
```

```
Analysis:    Type = Mixture ;
             stiterations = 20 ;
Model:       %Overall%
                Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2
                s4flbadc@3 ;
                Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1
                s3flbadc#1@2 s4flbadc#1@3 ;
Output:      sampstat residual tech1 tech8 ;
Plot:        Type = Plot3 ;
! Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
             Series = s1flbadc#1 s2flbadc#1 s3flbadc#1
                s4flbadc#1(*) ;
! We estimate the model twice, once with the count series
! commented out and once with the inflation series commented out
```

Here is selected output. We will not provide a detailed interpretation of the model estimates but note that the threshold value, α, β, α_i, and β_i are all highly significant and in the expected directions consistent with the zero-inflated Poisson Model with No Covariates that we estimated previously.

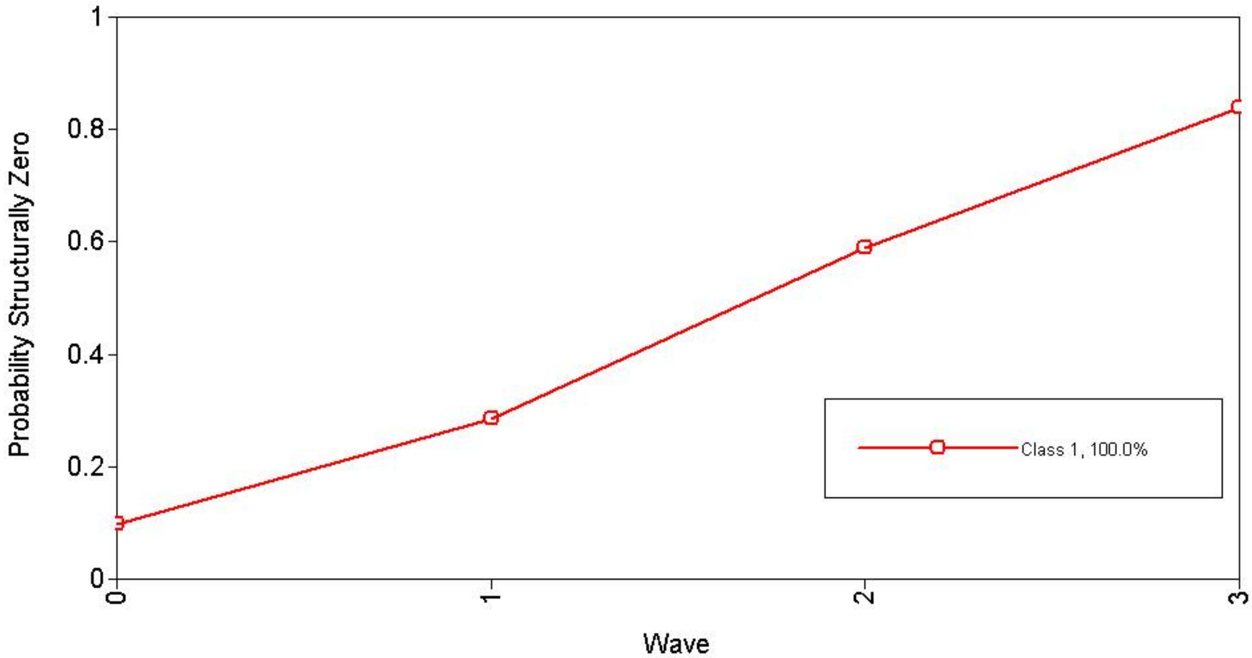
Figure 13 shows a graph of the growth curve for the inflation part of the model. Notice this curve increases more quickly than what we obtained with the zero-inflated Poisson model. These two solutions made different assumptions about missing data and this may account for part of the difference in estimates. However the basic trajectories have similar patterns.

```
THE MODEL ESTIMATION TERMINATED NORMALLY



TESTS OF MODEL FIT

Loglikelihood

          H0 Value                        -9581.115
          H0 Scaling Correction Factor        1.946
            for MLR

Information Criteria

          Number of Free Parameters               4
          Akaike (AIC)                    19170.230
          Bayesian (BIC)                  19194.157
          Sample-Size Adjusted BIC        19181.447
```

```
            (n* = (n + 2) / 24)

CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY LATENT CLASS
MEMBERSHIP
Class Counts and Proportions
    Latent
   Classes
       1             2927          1.00000
MODEL RESULTS
                   Estimates     S.E.  Est./S.E.
Latent Class 1
```

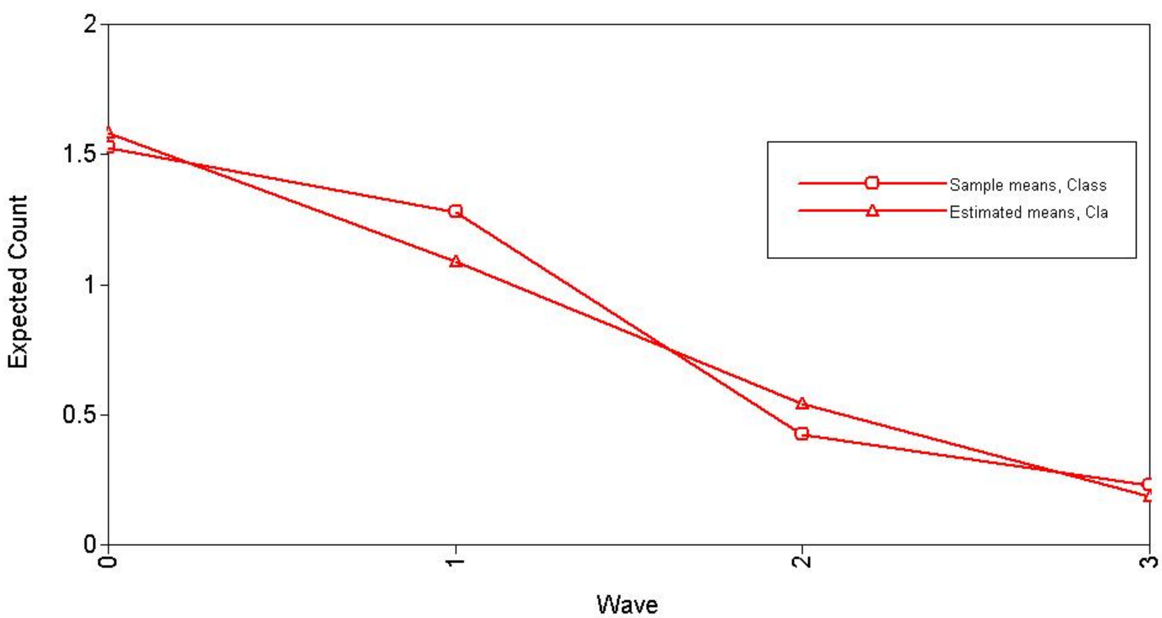
```
Latent Class 1
 Intercepts
    S1FLBADC#1        -2.216      0.116    -19.092      0.000
    S1FLBADC           0.000      0.000    999.000    999.000
    S2FLBADC#1        -2.216      0.116    -19.092      0.000
    S2FLBADC           0.000      0.000    999.000    999.000
    S3FLBADC#1        -2.216      0.116    -19.092      0.000
    S3FLBADC           0.000      0.000    999.000    999.000
    S4FLBADC#1        -2.216      0.116    -19.092      0.000
    S4FLBADC           0.000      0.000    999.000    999.000
 Means
    ALPHA              0.562      0.035     16.162      0.000
    BETA              -0.143      0.028     -5.082      0.000
    ALPHA_I            0.000      0.000    999.000    999.000
    BETA_I             1.290      0.055     23.511      0.000
RESIDUAL OUTPUT
     ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED) FOR CLASS 1
           Model Estimated Inflation Probability
              S1FLBADC      S2FLBADC      S3FLBADC      S4FLBADC
              ________      ________      ________      ________
 1              0.098         0.284         0.590         0.839
           Model Estimated Means
              S1FLBADC      S2FLBADC      S3FLBADC      S4FLBADC
              ________      ________      ________      ________
 1              1.582         1.089         0.541         0.184
           Residuals for Means
              S1FLBADC      S2FLBADC      S3FLBADC      S4FLBADC
              ________      ________      ________      ________
 1             -0.056         0.190        -0.120         0.044
```

Figure 13
Growth in Likelihood of a Structural Zero



The count part of the model has a significant intercept as well as a significant slope, α = .562, z = 16.162, $p < .001$ and β = -.143, z = -5.082, $p < .001$. This indicates a significant decline in negative responses to positive actions. Because there are no covariates we can show this graphically (see Figure 14).

Figure 14
Growth in Number of Negative Responses to Positive Actions



Extension of Mixture Model, Zero-Inflated Poisson Model to Two Classes

This program can be extended to have 2 or more classes. The program would simply add `classes = c(2)` or `classes = c(3)` for two and three class solutions, respectively. As with the mixture model presented in the first section of this paper, the AIC, BIC, Sample Adjusted BIC, and Lo-Mendell-Rubin Adjusted Likelihood Ratio Test can be used for model comparisons.

When multiple classes are identified the defaults for Mplus allow differences in the count part of the model, but impose equality constraints on the inflation part of the model. Getting around these defaults is a bit problematic and can lead to convergence problems. These issues will be explained in the section where we have covariates.

For now we will simply present the plots of the two classes for the growth component. As Figure 15 shows, the default does not differentiate between the two classes in the inflation part of the model. As Figure 16 shows there is a normative group that shows a steady decline in the number of negative responses they have to positive actions. It appears that the normative group is dropping to such a low rate of negative responses that there may be a censoring effect on the ability to further reduce the rate. There is a much smaller group that starts with a much higher frequency of negative responses. An important feature of this smaller group is that the program has an even more dramatic effect on them and although they do not fall to the level of the normative group by wave 4 they have come much closer than they were at wave 1.

Figure 15
Estimated Inflation Probabilities for Process

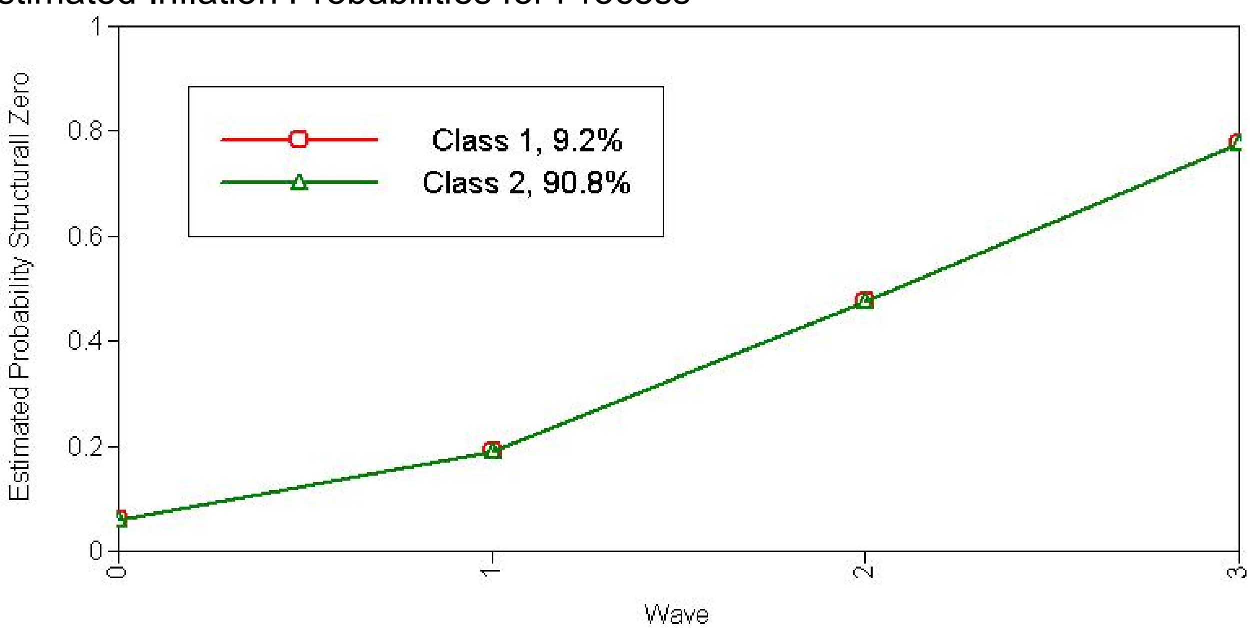
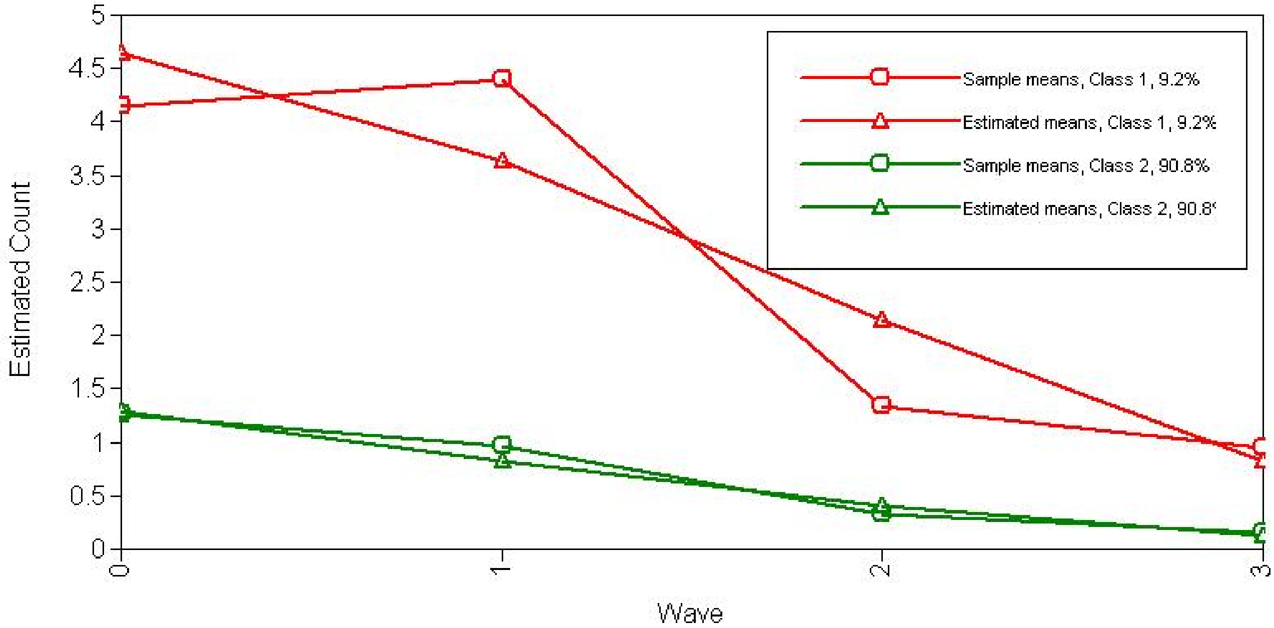


Figure 16

Growth in Number of Negative Responses to Positive Behaviors Using a Two Class Latent Class Analysis Zero-Inflated Poisson Model



Mixture Model, Zero-Inflated Poisson Models with Covariates

We can extend our model to include covariates. We will select a model of the number of classes and subsequently we will remove restrictions on the differentiation between classes on the inflation part of the model. Repeating the single class solution our program has only one change. We add the covariate effects under the section labeled `Model:` (see `workshop LCA zip poisson model with covariates c1.inp` in the Appendix). The first two lines below add the time invariant covariate for gender and the last two add the time varying covariates, `c3` and `c4`, which represent the level of implementation at waves 3 and 4.

```
Alpha on male ;
Beta on male ;
s3flbadc on c3 ;
s4flbadc on c4 ;
```

We estimated this model with from one to four classes. The following table summarizes the comparison of these models.

- The number of free parameters being estimated varies from 8 for the single class solution to 17 for the 4 class solution.
- The AIC, BIC, and Sample Adjusted BIC get smaller with more classes.
- The entropy is reasonable for all four solutions.

- The Lo, Mendell, Rubin test shows significant improvement moving from one to two classes and from two to three classes, but adding a fourth class has a barely significant improvement.
- The class sizes are an important consideration. What is an ideal number of classes depends on our research purposes.
    - Both the two and three class solutions have a clear normative class and either one or two "deviant" classes.
    - The smaller classes are quite small relative to the normative class and for some purposes we may not want to consider them. For example, with a sample of 987 children we might be hesitant to suggest substantial program modifications for the group of 41 children in the second class of the three class solution.
    - The four class solution has one class with just 4 children.
- Not shown is a warning message for the 4 class solution. Even with 200 random starts Mplus still doesn't converge on a solution and warns the results may not be trustworthy.

	1 Class	*2 Classes*	*3 Classes*	4 *Classes*
Free Parameters	*8*	11	14	17
AIC	*8647.18*	8232.32	8126.24	8118.85
BIC	*8686.33*	8286.16	8194.71	8192.06
Sample Adjusted BIC	*8660.93*	8251.23	8150.30	8138.06
Entropy		.83	.82	.83
Lo, Mendell, Rubin	*na*	2 *v* 1 Value =401.45 *P* < .001	3 *v* 2 Value = 106.91 *p* < .01	4 *v* 3 Value = 22.31 *p* = .047
***N* for each class**	*C1 = 987*	C1=85 C2=902	C1=891 C2=55 C3=41	C1=64 C2=38 C3=881 C4=4

Mplus cannot provide graphs for solutions when we have covariates. We might be tempted to drop the covariates and report graphs for the one to four class solutions. Because Mplus is using all available information we would find people distributed differently across the classes when we did not have covariates. If we did not have good reasons for including the covariates relying on the solution that has no covariates might have advantages in terms of simplicity of interpretation. We could then do a profile analysis of the members of each class to see how they differed. This would require using the `Savedata:` command as discussed in our LCA (see `workshop LPA.inp` in the Appendix).

For our purposes we will examine the two class solution with covariates. This two class solution does significantly better than the one class solution, Lo-Mendell-Rubin test statistic = 401.45, $p <$ 001 and there is a substantial reduction in all of the fit indexes. An argument could be made favoring the three class solution because it does significantly better than the two class solution and reduces the fit indexes. The choice would center on whether the researcher could interpret the three classes in a way that was useful. Here are selected results for the **two class solution with covariates:**

```
TESTS OF MODEL FIT

Loglikelihood

          H0 Value                        -4105.161
          H0 Scaling Correction Factor        1.891
            for MLR

Information Criteria

          Number of Free Parameters              11
          Akaike (AIC)                     8232.321
          Bayesian (BIC)                   8286.163
          Sample-Size Adjusted BIC         8251.226

MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Latent Class 1

 ALPHA    ON
    MALE               0.199      0.106      1.871      0.061
 BETA     ON
    MALE               0.057      0.067      0.858      0.391
 S3FLBADC ON
    C3                -0.266      0.060     -4.427      0.000
 S4FLBADC ON
    C4                -0.385      0.122     -3.165      0.002
 Intercepts
    S1FLBADC#1        -3.160      0.502     -6.293      0.000
    S1FLBADC           0.000      0.000    999.000    999.000
    S2FLBADC#1        -3.160      0.502     -6.293      0.000
```

```
    S2FLBADC           0.000      0.000    999.000    999.000
    S3FLBADC#1        -3.160      0.502     -6.293      0.000
    S3FLBADC           0.000      0.000    999.000    999.000
    S4FLBADC#1        -3.160      0.502     -6.293      0.000
    S4FLBADC           0.000      0.000    999.000    999.000
 Means
    ALPHA              1.260      0.208      6.055      0.000
    BETA               0.050      0.121      0.413      0.680
    ALPHA_I            0.000      0.000    999.000    999.000
    BETA_I             0.989      0.199      4.962      0.000

Latent Class 2
ALPHA    ON
    MALE               0.199      0.106      1.871      0.061
 BETA     ON
    MALE               0.057      0.067      0.858      0.391
 S3FLBADC ON
    C3                -0.266      0.060     -4.427      0.000
 S4FLBADC ON
    C4                -0.385      0.122     -3.165      0.002
 Intercepts
    S1FLBADC#1        -3.160      0.502     -6.293      0.000
    S1FLBADC           0.000      0.000    999.000    999.000
    S2FLBADC#1        -3.160      0.502     -6.293      0.000
    S2FLBADC           0.000      0.000    999.000    999.000
    S3FLBADC#1        -3.160      0.502     -6.293      0.000
    S3FLBADC           0.000      0.000    999.000    999.000
    S4FLBADC#1        -3.160      0.502     -6.293      0.000
    S4FLBADC           0.000      0.000    999.000    999.000
 Means
    ALPHA              0.022      0.100      0.224      0.823
    BETA              -0.095      0.063     -1.517      0.129
    ALPHA_I            0.000      0.000    999.000    999.000
    BETA_I             0.989      0.199      4.962      0.000
Categorical Latent Variables
 Means
    C#1               -2.113      0.223     -9.481      0.000
```

The program defaults do not provide for many possible differences between the classes.

- The table below shows that the effects of gender and level of implementation are held constant across classes.
- The default does not allow for possible differences in the inflation part of the model.
- All that vary by default are the intercept growth factor, α, and the slope growth factor, β for the count part of the model.

When we had no covariates we had all of these parameter estimates significant but with the greatest drop in negative responses in the small class. Here the results appear disappointing. The classes are distinguished by the smaller class starting with a higher rate of negative responses, Mean α = 1.26 (remember we are predicting a log of the rate) so the actual estimate rate would be much higher, $e^{1.26} = 3.53$. The Normative class also has a trend of a decline in the count that is marginally significant.

Parameter Estimate	*Class 1 (N = 85)*	*Class 2 (N=902)*	
Mean α	**1.26*****	**.022^{ns}**	
Mean β	**.050^{ns}**	**-.095†**	
Mean α_i (fixed)	**.000**	**.000**	
Mean β_i	**.989*****	**.989*****	
Threshold	**-3.160*****	**-3.160*****	All of these
Gender→α	**.199†**	**.199†**	Are equal
Gender→β	**.057^{ns}**	**.057^{ns}**	Across groups
Implement(3)→s3count	**-.266*****	**-.266*****	By default
Implement(4)→s4count	**-.385*****	**-.385*****	

Freeing Additional Constraints

The next step is to free additional constraints. As noted above, the classes are constrained to have equal parameter estimates for the inflation part of the model and for the effects of the covariates. Freeing additional constraints can create problems with convergence (Muthén & Muthén, 2006). However, this may be very important conceptually if you want to explore differences in the inflation part of the model or differences in the effects of covariates across classes. We estimated a model that allowed the inflation part of the model to vary across classes, but still constrained the effects of the covariates. This was done in the `Model:` section of the program.

The `%c#2%` tells Mplus that class number 2 will have some differences in the specification. The line, `[s1flbadc#1 s2flbadc#1 s3flbadc#1 s4flbadc#1](1) ;` Puts an equality constraint on the thresholds for items but allows the value to be different than it was for class 1. The (1) is an index that the constraint applies to this set of estimates. The `[Beta_i] ;` indicates that this intercept parameter is free to differ for the second class.

```
Title:       workshop_mixture_zip_model_class_diff2.inp
Data:        File is workshop_growth.dat ;
Variable:    Names are
                idnum s1flbadc s2flbadc s3flbadc s4flbadc male
                s1flbadd s2flbadd s3flbadd s4flbadd s1flbadm
                s2flbadm s3flbadm s4flbadm c3 c4 s3techer room ;
             Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc male c3 c4 ;
             Missing are all (-9999) ;
             Count are s1flbadc s2flbadc s3flbadc s4flbadc (i);
             Classes = c(2) ;
Analysis:    Type = Mixture ;
             Starts = 50 2 ;
             Stiterations = 20 ;
Model:       %Overall%
                Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
                Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1 s3flbadc#1@2
                s4flbadc#1@3 ;
                Alpha on male ;
                Beta on male  ;
                S3flbadc on c3 ;
                S4flbadc on c4 ;
             %c#2%
                [s1flbadc#1 s2flbadc#1 s3flbadc#1 s4flbadc#1](1) ;
                [Beta_i] ;
             Output:     sampstat residual tech1 tech8 ;
```

The key results of this model appear below. The normative class (class 1) has a threshold of -15.00 estimating that the initial probability of being structurally zero is virtually zero, e^{-15} = .0000003. (Note, as a likelihood approaches zero Mplus fixes the parameter at approximately zero by making the threshold -15.000 since and tells you it did this.)

Notice the previous solution missed an enormous difference between the normative and deviant classes in the inflation part of the model.

- Although the normative class starts with approximately zero likelihood of being structurally zero, the mean β_i = 4.853, $p < .001$ indicates a rapid increase in the likelihood of being structurally zero.
- We saw a similar result much earlier in that the program appears to be inoculating children from having negative responses to positive actions as part of their behavioral repertoire.

By contrast, there is a smaller group, $N = 101$, of children who may start with a greater likelihood of being structurally zero (threshold is -1.505, $e^{-1.505}$ = .222, but who nonetheless make much less progress, Mean β_i = .552 vs. Mean β_i = 4.853 for the normative group.

The count part of the model shows that the normative class ($N = 886$) starts with a much lower initial level than the deviant class as we saw in Figure 16 when we had no covariates, here α = -.028.

However, when we have the covariates it is interesting that the normative class declines below an already low initial level, β = -.159, $p < .05$, but the deviant class does not change significantly in the rate of negative responses. The deviant class starts with a high initial count α = 1.271, $p < .001$ and has an insignificant slope β = .030, *ns*. This is quite different from Figure 16 where the deviant class gained the most benefit (larger negative mean β). Thus, controlling for the covariates results in two classes. A small class, *N* = 101, that starts with a higher level of negative responses and doesn't benefit from the intervention and a large class, *N* = 886, that starts with a fairly lower level of negative behavior and drops even lower while increasing the likelihood of the children being inoculated into a structural zero rate of negative responses.

Categorical Latent
Inflation free

Parameter Estimate	*Class 1 (N = (101)*	*Class 2 (N=886)*
Mean α	1.271**	-.028ns
Mean β	$.003^{ns}$	-.159*
Mean α_i (fixed)	.000	.000 fixed
Mean β_i	.552***	4.853***
Threshold	-1.505***	-15.000 fixed by Mplus
Gender→α	$.192^{\dagger}$	.227*
Gender→β	$.065^{ns}$	$.079^{ns}$
Implement(3)→s3count	-.292***	$-.191^{\dagger}$
Implement(4)→s4count	-.335**	$-.237^{\dagger}$

We do not show the results here but we can also allow the covariates to operate differently in the two classes. This is accomplished by modifying the `Model:` section of the program. The `Model:` section becomes:

```
Model:    %Overall%
             Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
             Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1
             s3flbadc#1@2 s4flbadc#1@3 ;
             Alpha on male ;
             Beta on male  ;
             S3flbadc on c3 ;
             S4flbadc on c4 ;
          %c#2%
             [s1flbadc#1 s2flbadc#1 s3flbadc#1 s4flbadc#1](1) ;
```

```
[Beta_i] ;
Alpha on male (1);
Beta on male  (2);
S3flbadc on c3 (3);
S4flbadc on c4 (4);
```

Where the last four lines are the additional code required.

Summary

Many outcomes are best studied using longitudinal data to identify growth trajectories. Growth trajectories for a binary component can be distinguished from growth trajectories for a count component. Different time invariant and time variant covariates may influence both of these trajectories. Although not illustrated with an empirical example, these ideas can be extended to have distal outcomes. These distal outcomes may vary based on the time invariant covariates, the time varying covariates, the initial levels of the binary trajectory, the slope of the binary trajectory, the initial level of the count component, and the slope of the count component. It is also possible to estimate the size and significance of indirect effects and mediation.

Mplus offers many features that are especially useful for longitudinal studies of individuals and families

- Conventional growth curve models can be extended to zero-inflated models
- Estimation can assume random variation of parameters or fixed variance of parameters
- Estimation can use a mixture model identify subgroups with different trajectories assuming all variance in parameters is explained by class differences
- Estimation can utilize mixture modeling to allow for random variances of parameters within classes

References

Bollen, K. A., & Curran, P. K., (2006). *Latent Curve Models: A Structural Equation Perspective*. New York: Wiley-Interscience.

Bauer, D. J., & Curran, P. J. (2002). Distributional assumptions of growth mixture models: Implications of over-extraction of latent trajectory classes. *Psychological Methods*.

Duncan, T.E., Duncan, S.C., & Stryker, L. (2006). *An Introduction to Latent Variable Growth Curve Modeling* (2nd ed.). Mahwah NJ: Lawrence Erlbaum.

Flay, B., Acock, A., Vuchinich, S., & Beets, M. (2006). Progress report of the randomized trial of Positive Action in Hawaii: End of third year of intervention (Spring, 20065). Available from Positive Action, Inc. 264 4th Avenue South, Twin Falls, ID 83301.

Goodman, L. A. Exploratory latent structure analysis using both identifiable and unidentifiable models, *Biometrika*, *61,* 215-231.

Jeffries, N.O. (2003). A note on 'testing the number of components in a normal mixture'. *Biometrika, 90*, 991-994.

Khoo, S. T., & Muthén, B. O. (2000). Longitudinal Data on Families: Growth Modeling Alternatives. In J. S. rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate Applications in Substance Use Research: New Methods for New Questions* (pp. 43-78). Mahwah, NJ: Lawrence Erlbaum Associates.

King, G. (1989). Event Count Models for International Relations: Generalizations and Applications. *International Studies Quarterly 33*, 123-147.

Kreuter, F., & Muthen, B. O. (2006). Analyzing criminal trajectory profiles: Bridging multilevel and grrop-based approaches using growth mixture modeling. Available at www.statmodel.com .

Lazarfeld, P. F. & Henry, Neil W. (1968). *Latent Structure Analysis.*

Long, J. S., & Freese, J. (2006). Regression Models for Categorical Dependent Variables Using Stata. College Station TX : Stata Press.

Mitchell, M. N. (no date). stata2mplus. UCLA Academic Technology Services.

Muthén, B. O. (1996). Growth modeling with binary responses. In A. V. Eye, & C. Clogg (Eds.), Categorical Variables in Developmental Research: Methods of Analysis (pp. 37-54). San Diego CA: Academic Press

Muthén, B. O. (2002). Statistical and Substantive Checking in Growth Mixture Modeling. UCLA

Muthen, B. O. (2006). The potential of growth mixture modeling. *Infant and Child Development*, *15*.

Muthén, B. O., & Muthén, L. K. (2000a). The Development of Heaving Drinking and Alcohol-Related Problems from Ages 18-37 in a U.S. National Sample. *Journal of Studies on Alcohol*.

Muthén, B. & Muthén, L. (2000b). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. Alcoholism: Clinical and Experimental Research, 24, 882-891.

Nylunk, K., Asparauhov, T., Muthén, O. (2006). Deciding on the number of classes in Latent Classs analysis and growth mixture modeling: A Monte-Carolo simulation study. Available at www.statmodel.com .

Rabe-Hesketh, S. & Skrondal, A. (2005). *Multilevel and Longitudinal Modeling Using Stata*. College Station TX: Stata Press.
Raudenbush, S. W. (2005). How do we study "What Happens Next"? *Annals of the American Academy of Political and Social Sciences, 602*, 131-144.
Stavig, G. R. & Acock, A C. (1981). Applying the Semi-Standardized Regression Coefficient to Factor, Canonical, and Path Analysis. *Multivariate Behavioral Research* 2: 255-258.

Appendix

Programs used in this presentation. Some of these programs appeared in the text. The name of the program appears in the first line.

Converting Files

Currently, Mplus cannot import files directly from other programs. SPSS and SAS users should write a program to save a plain ASCII file that is either space or comma delimited. This output file can be read by Mplus.

Stata users can take advantage of a user written command, `stata2mplus`. Install this command by typing `findit stata2mplus` and following the screen directions. This command does two things to a Stata File.

1. Creates a dataset that Mplus can read. You should delete irrelevant variables in your Stata dataset before running this.
2. Writes an Mplus program (*.inp) to analyze the data using Mplus. This Mplus program sets up missing values and includes value labels.

An example of the command is:

```
save "e:\flash\workshop\Workshop.dta", replace
```

A Simple growth model for a binary outcome

```
Title:       workshop binary growth.inp
Data:        File is workshop_growth no covariates.dat ;
Variable:    Names are
               idnum s1flbadc s2flbadc s3flbadc s4flbadc male
               s1flbadd s2flbadd s3flbadd s4flbadd s1flbadm
               s2flbadm s3flbadm s4flbadm c3 c4 s3techer room ;
             Usevariables are s1flbadd s2flbadd s3flbadd s4flbadd ;
             Categorical are s1flbadd s2flbadd s3flbadd s4flbadd ;
             Missing are all (-9999) ;
Analysis:    Type = Missing ;
             Estimator = ML ;
Model:       alpha beta | s1flbadd@0 s2flbadd@1 s3flbadd@2 s4flbadd@3 ;
Output:      residual Patterns sampstat standardized tech8;
Plot:        Type = Plot3 ;
               Series = s1flbadd s2flbadd s3flbadd s4flbadd(*) ;
```

A simple growth model for a count variable: Fixed Effects

```
Title:       workshop_growth_count_fixed.inp
Data: File is workshop_growth.dat ;
Variable:
             Names are
             idnum s1flbadc s2flbadc s3flbadc s4flbadc
                   male s1flbadd s2flbadd s3flbadd s4flbadd
                   s1flbadm s2flbadm s3flbadm s4flbadm c3 c4
                   s3techer room ;
             Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc ;
             Missing are all (-9999) ;
             Count are s1flbadc s2flbadc s3flbadc s4flbadc ;
Analysis:    estimator = MLR ; !This is default since clearly not normal
             processors = 2 ;
Model:
             alpha beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
             alpha@0 ; !fixes intercept variance at 0--fixed effect
             beta@0 ;    !fixes slope variance at 0--fixed effect for slope
Output:
             residual tech1 tech4 tech8;
Plot:
             Type = Plot3 ;
             Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
```

A Simple Growth model for a count variable: Random Effects

```
Title:             workshop_count_growth_random effects.inp
Data:              File is workshop_growth.dat ;
Variable:    Names are
                         idnum s1flbadc s2flbadc s3flbadc s4flbadc
                         male s1flbadd s2flbadd s3flbadd s4flbadd
                         s1flbadm s2flbadm s3flbadm s4flbadm c3 c4
                         s3techer room ;
                   Usevariables are s1flbadc s2flbadc s3flbadc
                         s4flbadc ;
                   Missing are all (-9999) ;
                   Count are s1flbadc s2flbadc s3flbadc s4flbadc ;
Analysis:    ALGORITHM = integration ; !Numerical integration
! can be very CPU intense. It is needed for random effects models
                   processors = 2 ;
Model:             alpha beta | s1flbadc@0 s2flbadc@1 s3flbadc@2
                   s4flbadc@3 ;
Output:
                   residual tech1 tech4 tech8;
Plot:
                   Type = Plot3 ;
                   Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
```

Latent Class Analysis

```
Title: workshop lca2.inp
  Latent Class Analysis of implementation for year 3
  with 2 latent classes
Data:
  File is
    lcalpa34.dat ;
Variable:
  Names are
    idnum s3ptp1 s3ptp2 s3ptp3 s3ptp4 s3ptp5 s3ptp6 s3ptp7 s3ptp8 s3ptp9
    s3ptp12 s4ptp1 s4ptp2 s4ptp3 s4ptp4 s4ptp5 s4ptp6 s4ptp13 s4ptp7 s4ptp14
    s4ptp8 s4ptp9 s4ptp10 s4ptp11 s4ptp12 s3ptp1b s3ptp2b s3ptp3b s3ptp4b
    s3ptp5b s3ptp6b s3ptp7b s3ptp8b s3ptp9b s3ptp12b s4ptp1b s4ptp2b s4ptp3b
    s4ptp4b s4ptp5b s4ptp6b s4ptp7b s4ptp8b s4ptp9b s4ptp10b s4ptp11b
    s4ptp12b s4ptp13b s4ptp14b s3techer room ;    Missing are all (-9999) ;
  Usevariables are
    s3ptp1b s3ptp2b s3ptp3b s3ptp4b s3ptp5b s3ptp6b s3ptp7b s3ptp8b s3ptp9b
    s3ptp12b ;
  Categorical are
    s3ptp1b s3ptp2b s3ptp3b s3ptp4b s3ptp5b s3ptp6b s3ptp7b s3ptp8b s3ptp9b
    s3ptp12b ;
  Classes = c(2) ;
Analysis:
  Type = Mixture Missing ;
  Starts = 20 2;
Output:
  Stand Tech11 ;
Plot:
    Type = Plot3 ;
    series = s3ptp1b(1) s3ptp2b(2) s3ptp3b(3) s3ptp4b(4) s3ptp5b(5)
             s3ptp6b(6) s3ptp7b(7) s3ptp8b(8) s3ptp9b(9) s3ptp12b(10) ;
```

Growth Curve Model for Count Variables: Random Effects

```
Title:        workshop_count_growth_random_effects_cov.inp
Data:         File is workshop_growth.dat ;
Variable:     Names are
                          idnum s1flbadc s2flbadc s3flbadc s4flbadc
                          male s1flbadd s2flbadd s3flbadd s4flbadd
                          s1flbadm s2flbadm s3flbadm s4flbadm c3 c4
                          s3techer room ;
                    Usevariables are s1flbadc s2flbadc s3flbadc
                          s4flbadc male c3 c4;
                    Missing are all (-9999) ;
                    Count are s1flbadc s2flbadc s3flbadc s4flbadc ;
Analysis:           ALGORITHM = integration ;
                    Integration = 20 ; !default is 7 integration points
! can run with even fewer and might for developing models and then
! use integration = 50 ; for the final run (and enjoy lunch)
                    processors = 2 ;
                    starts 40 2 ;
Model:              alpha beta | s1flbadc@0 s2flbadc@1 s3flbadc@2
                          s4flbadc@3 ;
                    alpha on male ;
```

```
                    beta on male ;
                    s3flbadc on c3 ;
                    s4flbadc on c4 ;
Output:      residual tech1 tech4 tech8;
Plot:
                    Type = Plot3 ;
                    Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
```

A Zero Inflated Poisson Model

```
Title:             workshop_zip_poisson_model.inp
                   Linear growth model for a count outcome using
                   The zero-inflated Poisson (ZIP) model
Data:              File is workshop_growth.dat ;
Variable:      Names are
                idnum s1flbadc s2flbadc s3flbadc s4flbadc
                      male s1flbadd s2flbadd s3flbadd s4flbadd
                      s1flbadm s2flbadm s3flbadm s4flbadm c3 c4
                      s3techer room ;
                   Usevariables are s1flbadc s2flbadc s3flbadc
                    s4flbadc ;
                   Missing are all (-9999) ;
                   Count are s1flbadc s2flbadc s3flbadc
                    s4flbadc (i);
  ! the (i) tells Mplus to set this up for an inflated model
  ! we need class(1) and type mixture to get the graphs we
  ! will use
Analysis:      processors = 2 ;
                   Integration = 5 ;
                   Starts 40 2 ;
Model:             Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2
                    s4flbadc@3 ;
                   Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1
                    s3flbadc#1@2 s4flbadc#1@3 ;
  !Alpha is intercept for count part and Beta is slope
  !for count part
  !Alpha_i is intercept for inflated part and Beta_i is
  !the slope for inflated part
Output: Patterns residual tech1 tech8 ;
Plot:
             Type = Plot3 ;
             Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
```

A Zero Inflated Poisson Model with Covariates

```
Title:      workshop_zip_poisson_model_cov.inp
            Linear growth model for a count outcome using the
            zero-inflated Poisson (ZIP) model
Data:       File is workshop_growth.dat ;
Variable:   Names are
              idnum s1flbadc s2flbadc s3flbadc s4flbadc male s1flbadd
              s2flbadd s3flbadd s4flbadd s1flbadm s2flbadm s3flbadm
```

```
              s4flbadm c3 c4 s3techer room ;
            Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc
              male c3 c4;
            Missing are all (-9999) ;
            Count are s1flbadc s2flbadc s3flbadc s4flbadc (i);
  ! the (i) tells Mplus to set this up for an inflated model
Analysis:   Integration = 5 ;
            Starts 40 2 ;
            processors = 2
Model:      Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
            Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1 s3flbadc#1@2
              s4flbadc#1@3 ;
  !Alpha is intercept for count part and Beta is slope for count part
  !Alpha_i is intercept for inflated part and Beta_i is slope for
  !inflated part
            Alpha on male ;
            Beta on male ;
            Alpha_i on male ;
            Beta_i on male ;
            s3flbadc on c3 ;
            s4flbadc on c4 ;
Output:     Patterns sampstat tech1 tech8 stand;
```

Latent Profile Analysis

```
Title:  latent profile analysis of implementation for year 3
Data:
  File is lpa3or4.dat ;
Variable:
  Names are
     idnum childid cond1234 data0405 diffgend doe_id doe_id2 file1_id file2_da
     form_ab grade id mast0405 rec_form recs4gen s3ans s3ans2 s3apa1 s3apa2
     s3apa3 s3apa4 s3apa5 s3apa6 s3apa7 s3apa8 s3apa9 s3apa10 s3apa11 s3apa12
     s3apa13 s3apa14 s3apa15 s3apa16 s3cond s3esll s3eth2 s3gend s3gend2
     s3grad s3grade s3island s3lowinc s3mth s3ptp1 s3ptp2 s3ptp3 s3ptp4
     s3ptp5 s3ptp6 s3ptp7 s3ptp8 s3ptp9 s3ptp10 s3ptp11 s3ptp12 s3rdg s3sccd
     s3sccdx s3sped s4apa1 s4apa2 s4apa3 s4apa4a s4apa4b s4apa5 s4apa15
     s4apa16 s4apa17 s4cond s4gend s4grad s4grade s4ptp1 s4ptp2 s4ptp3
     s4ptp4 s4ptp5 s4ptp6 s4ptp7 s4ptp8 s4ptp9 s4ptp10 s4ptp11 s4ptp12
     s4ptp13 s4ptp14 s4race s4racehl s4sccd school unique v3 room s3techer
     s4room ;
  Missing are
     all (-9999) ;
  Usevariables are
     s3ptp1 s3ptp2 s3ptp3 s3ptp4 s3ptp5
     s3ptp6 s3ptp7 s3ptp8 s3ptp9 s3ptp12 ;
  Classes = c(2) ;
  Cluster = s3techer ;
  Idvariable = idnum ;
Analysis:
```

```
  Type = Mixture Missing Complex ;
  Starts = 40 2;
Output:
  samp Stand Tech11 ;
Plot:
  Type = Plot3 ;
  series = s3ptp1(1) s3ptp2(2) s3ptp3(3) s3ptp4(4) s3ptp5(5)
           s3ptp6(6) s3ptp7(7) s3ptp8(8) s3ptp9(9) s3ptp12(10) ;
Savedata:
  File is wave3.dat ;
  Save =  Cprobabilities ;
  Format is F6.0 ;
```

Zero-inflated Poisson Model without covariates

```
Title: workshop zip poisson model.inp
  Stata2Mplus conversion for workshop_growth.dta
  Linear growth model for a count outcome using the
  zero-inflated Poisson (ZIP) model
Data:
  File is workshop_growth.dat ;
Variable:
  Names are
     idnum s1flbadc s2flbadc s3flbadc s4flbadc male s1flbadd s2flbadd s3flbadd
     s4flbadd s1flbadm s2flbadm s3flbadm s4flbadm c3 c4 s3techer room ;
  Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc ;
  Missing are all (-9999) ;
  Count are s1flbadc s2flbadc s3flbadc s4flbadc (i);
  ! the (i) tells Mplus to set this up for an inflated model
  ! we need class(1) and type mixture to get the graphs we will use
Analysis:
  Type = Missing ;
  Integration = 5 ;
  ! Starts 40 2 ;
Model:
  Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
  Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1 s3flbadc#1@2 s4flbadc#1@3 ;
  !Alpha is intercept for count part and Beta is slope for count part
  !Alpha_i is intercept for inflated part and Beta_i is slope for inflated part
Output:
  Patterns sampstat residual tech1 tech8 ;
Plot:
  Type = Plot3 ;
  Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
```

Zero-Inflated Poisson Model with covariates

```
Title: workshop zip poisson model cov.inp
  Stata2Mplus conversion for workshop_growth.dta
  Linear growth model for a count outcome using the
  zero-inflated Poisson (ZIP) model
Data:
```

```
  File is workshop_growth.dat ;
Variable:
  Names are
     idnum s1flbadc s2flbadc s3flbadc s4flbadc male s1flbadd s2flbadd s3flbadd
     s4flbadd s1flbadm s2flbadm s3flbadm s4flbadm c3 c4 s3techer room ;
  Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc male c3 c4;
  Missing are all (-9999) ;
  Count are s1flbadc s2flbadc s3flbadc s4flbadc (i);
  ! the (i) tells Mplus to set this up for an inflated model
  ! we need class(1) and type mixture to get the graphs we will use
Analysis:
  Type = Missing ;
  Integration = 5 ;
  ! Starts 40 2 ;
Model:
  Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
  Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1 s3flbadc#1@2 s4flbadc#1@3 ;
  !Alpha is intercept for count part and Beta is slope for count part
  !Alpha_i is intercept for inflated part and Beta_i is slope for inflated part
  Alpha on male ;
  Beta on male ;
  Alpha_i on male ;
  Beta_i on male ;
  s3flbadc on c3 ;
  s4flbadc on c4 ;
Output:
  Patterns sampstat tech1 tech8 stand;
```

Mixture Analysis Using Zero-Inflated Poisson with a single class

```
Title: workshop LCA zip poisson model NO covariates c1.inp
  Latent Class Growth Analysis for a count outcome
  using a ZIP Model
Data:
  File is workshop_growth.dat ;
Variable:
  Names are
     idnum s1flbadc s2flbadc s3flbadc s4flbadc male s1flbadd s2flbadd s3flbadd
     s4flbadd s1flbadm s2flbadm s3flbadm s4flbadm c3 c4 s3techer room ;
  Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc  ;
  Missing are all (-9999) ;
  Count are s1flbadc s2flbadc s3flbadc s4flbadc (i);
  ! the (i) tells Mplus to set this up for an inflated model
  ! we need class(1) and type mixture to get the graphs we will use
  Classes = c(1) ;
Analysis:
  Type = Mixture missing ;
  ! Stiterations = 20 ;
Model:
  %Overall%
  Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
  Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1 s3flbadc#1@2 s4flbadc#1@3 ;
  !Alpha is intercept for count part and Beta is slope for count part
  !Alpha_i is intercept for inflated part and Beta_i is slope for inflated part
```

```
Output:
  sampstat residual tech1 tech8 ;
Plot:
  Type = Plot3 ;
  ! Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
  Series = s1flbadc#1 s2flbadc#1 s3flbadc#1 s4flbadc#1(*) ;
```

Mixture Zero-Inflated Growth Models with no Covariates—two classes.

```
Title: Mixture Zero-Inflated Growth Model with No Covariates
  using a ZIP Model
Data:
  File is workshop_growth.dat ;
Variable:
  Names are
     idnum s1flbadc s2flbadc s3flbadc s4flbadc male s1flbadd s2flbadd s3flbadd
     s4flbadd s1flbadm s2flbadm s3flbadm s4flbadm c3 c4 s3techer room ;
  Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc  ;
  Missing are all (-9999) ;
  Count are s1flbadc s2flbadc s3flbadc s4flbadc (i);
  ! the (i) tells Mplus to set this up for an inflated model
  ! we need class(1) and type mixture to get the graphs we will use
  Classes = c(2) ;
Analysis:
  Type = Mixture missing ;
  ! Stiterations = 20 ;
Model:
  %Overall%
  Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
  Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1 s3flbadc#1@2 s4flbadc#1@3 ;
  !Alpha is intercept for count part and Beta is slope for count part
  !Alpha_i is intercept for inflated part and Beta_i is slope for inflated part
Output:
  sampstat residual  tech8 tech11;
Plot:
  Type = Plot3 ;
  Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
  ! Series = s1flbadc#1 s2flbadc#1 s3flbadc#1 s4flbadc#1(*) ;
```

Mixture ZIP Growth Models with no Covariates—two classes inflation parameters different

```
Title: workshop mixture zip poisson model no covariates c2 diff.inp
    Latent Class Growth Analysis for a zero-inflated model with no
    covariates allowing for class differences in inflation parameters
    using a ZIP Model
  Data:
    File is workshop_growth.dat ;
  Variable:
    Names are
       idnum s1flbadc s2flbadc s3flbadc s4flbadc male s1flbadd s2flbadd s3flbadd
       s4flbadd s1flbadm s2flbadm s3flbadm s4flbadm c3 c4 s3techer room ;
    Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc  ;
    Missing are all (-9999) ;
    Count are s1flbadc s2flbadc s3flbadc s4flbadc (i);
  ! the (i) tells Mplus to set this up for an inflated model
  ! we need class(1) and type mixture to get the graphs we will use
    Classes = c(2) ;
  Analysis:
    Type = Mixture Missing ;
    Starts = 50 2 ;
    Stiterations = 20 ;
  Model:
    %Overall%
    Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
    Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1 s3flbadc#1@2 s4flbadc#1@3 ;
  !Alpha is intercept for count part and Beta is slope for count part
  !Alpha_i is intercept for inflated part and Beta_i is slope for inflated part
    %c#2%
    [s1flbadc#1 s2flbadc#1 s3flbadc#1 s4flbadc#1](1) ;
    [Beta_i] ;
  Output:
    sampstat residual tech1 tech8 ;
  Plot:
    Type = Plot3 ;
    Series = s1flbadc s2flbadc s3flbadc s4flbadc(*) ;
  ! Series = s1flbadc#1 s2flbadc#1 s3flbadc#1 s4flbadc#1(*) ;
  ! I only do one plot at a time.
```

MIXTURE ZIP Growth Models with Covariates—two classes

```
Title:      workshop_mixture_zip_model_class_diff2.inp
Data:       File is workshop_growth.dat ;
Variable:   Names are
               idnum s1flbadc s2flbadc s3flbadc s4flbadc male
               s1flbadd s2flbadd s3flbadd s4flbadd s1flbadm
               s2flbadm s3flbadm s4flbadm c3 c4 s3techer room ;
            Usevariables are s1flbadc s2flbadc s3flbadc s4flbadc male c3 c4 ;
            Missing are all (-9999) ;
            Count are s1flbadc s2flbadc s3flbadc s4flbadc (i);
            Classes = c(2) ;
Analysis:   Type = Mixture ;
```

```
            Starts = 50 2 ;
            Stiterations = 20 ;
Model:      %Overall%
               Alpha Beta | s1flbadc@0 s2flbadc@1 s3flbadc@2 s4flbadc@3 ;
               Alpha_i Beta_i | s1flbadc#1@0 s2flbadc#1@1 s3flbadc#1@2
               s4flbadc#1@3 ;
               Alpha on male ;
               Beta on male  ;
               S3flbadc on c3 ;
               S4flbadc on c4 ;
            %c#2%
               [s1flbadc#1 s2flbadc#1 s3flbadc#1 s4flbadc#1](1) ;
               [Beta_i] ;
            Alpha on male ;
            Beta on male  ;
            S3flbadc on c3 ;
            S4flbadc on c4 ;
Output:     sampstat residual tech1 tech8 ;
```