Count modeling with different length of exposure uses an offset, that is, a term in the regression which has its coefficient fixed at 1. This can also be used for modeling proportions in that a count is the sum of all the binary events that lead to the proportion variable. The number of different events is N which may vary over observational units, so the offset log(N) should be used.

The Mplus Discussion thread below discusses modeling with offsets.

bmuthen posted on Tuesday, August 23, 2005 - 11:06 am

The following question appeared on SEMNET Aug 19, 2005.

Hi,
I'm using MPLUS to fit a zero-inflated poisson LGM following the example in the manual. However, I do not see how to include an OFFSET variable. My experience using SAS for such a model was to use a defined OFFSET variable that represented a denominator for the counts so a relative rate can be obtained. MlWin also has an option for the offset. Has anyone experience with this in MlWin (I think it should say Mplus)?

bmuthen posted on Tuesday, August 23, 2005 - 11:33 am

Poisson regression with an offset is useful with grouped data. It can be done in Mplus by adding an offset variable with a coefficient fixed at one. An example of doing this in SAS GENMOD is shown at:

http://v8doc.sas.com/sashtml/stat/chap29/sect6.htm

The same results are obtained by the Mplus input below, where the offset variable is specified in the Define command:

title: Poisson Offset

data: file is pooff.dat;

analysis: estimator=ml;

variable:
names are n c car1 car2 car3 age1 age2;
usevar are c car2 car3 age1 offset;
count=c;

define: offset=log(n);

model: c on offset@1 car2 car3 age1;

The data set for the Mplus run is:

500 42 1 0 0 1 0
1200 37 0 1 0 1 0
100 1 0 0 1 1 0
400 101 1 0 0 0 1
500 73 0 1 0 0 1
300 14 0 0 1 0 1

Not sure how one would do zero-inflated Poisson with an offset - is there literature on that?

Michael J. Zyphur posted on Tuesday, August 23, 2005 - 6:43 pm

I think this article:

http://www.stat.uga.edu/~dhall/pub/ZIMixed.pdf

by Hall in Biometrics talks about it. But I may be wrong.

Tihomir Asparouhov posted on Wednesday, August 24, 2005 - 9:24 am

The treatment of offset in this paper is a bit unorthodox in my point of view. The offset log(n) is being used simply as a covariate and its beta coefficient is not fixed to 1 but it is estimated.

It is possible to use log(n) as a covariate in ZIP for both or either, the mean part or the inflation part, however its interpretation is not very clear. Because sums of ZIPs is not a ZIP one cannot simply use procedures for estimating ZIP to estimate sums of ZIPs.

Susan E. Collins posted on Wednesday, January 21, 2009 - 10:54 pm

Hi there,

just as a follow-up to this thread: is it possible to similarly include an offset with the new negative binomial commands? Also, was the problem with treatment of the offset in a zip/zinb model resolved? or is that still not doable?

Thanks!
Susan

[Tihomir Asparouhov](#) posted on Thursday, January 22, 2009 - 11:15 am

It is always possible to use an offset variable and even estimate a slope coefficient for that offset. Technically speaking however including $\log(N)$ with a coefficient 1 is generally used for the Poisson model alone. This inclusion reflects a model where the dependent variable is not the same Poisson(mu) but it is a sum of N Poisson(mu). This modeling depends on the assumption that sum of independent Poisson variables is Poisson and this assumption does not work for zero-inflated distributions. Nevertheless, you can use $\log(N)$ as a covariate and estimate the slope. The interpretation of this model obviously cannot be that it is the sum of N independent zero-inflated distributions.

In general the Negative Binomial distribution is a sum of independent geometric distributions however the situation here is more complicated because not only the mean is affected by N but also the dispersion parameter and that is not reflected in a model with an offset variable only. You can off course still improve the model by including an offset variable.

If N is mostly a large number and the dependent variable is also large then the best choice might be to simply use a normal approximation model where both the mean and the variance of the normal dependent variable are the appropriate functions of N.

[Susan E. Collins](#) posted on Thursday, January 22, 2009 - 12:35 pm

Thanks, Tihomir! this was really helpful.

I didn't totally get the last paragraph, though. Just to clarify, we you say "N," I assume you mean that to be the exposure time (for example)? Also, the "normal approximation model" would be just treating the dv as if it were normal, using the Satorra-Bentler robust statistics and add in the N as a covariate?

Thanks again,
Susan

Tihomir Asparouhov posted on Thursday, January 22, 2009 - 2:06 pm

Yes, N is the total exposure time and indeed the robust ML will safeguard against the heteroscedasticity in the normal approximation model (here the residual variance will also be proportional to N). You can also model the varying residual variance - see web note #3
http://statmodel.com/download/webnotes/mc3.pdf
or Example 5.23 in the user's guide for how to use model constraints to build in mean and residual variance proportional to N.

Jason Payne posted on Wednesday, February 25, 2009 - 6:11 pm

How does one go about adjusting for exposure time when using a mixture model for longitudinal count data and where the exposure varies between individuals and over time?

For example, I have yearly counts of criminal convictions for 1000 individuals over 40 years (conv1-conv40) AND the number of days each year than each individual spent in custody (pt1-pt40). I want to model the latent class trajectories as in Kreuter, F. & Muthen, B. (2008) but with an offset for the number of exposure days in each each year.

Is it as simple as running the offset as a time-varying covariate?

Tihomir Asparouhov posted on Thursday, February 26, 2009 - 11:11 am

The general modeling approach is described here
http://en.wikipedia.org/wiki/Poisson_regression

You can implement this in Mplus as follows. I will do this just for 4 variables to make it short.

variable:
names = pt1-pt4 conv1-conv4;
usevar = conv1-conv4 exposure1-exposure4;

define:
exposure1=log((365-pt1)/365);
exposure2=log((365-pt2)/365);
exposure3=log((365-pt3)/365);
exposure4=log((365-pt4)/365);

model:

conv1-conv4 PON exposure1-exposure4@1;

etc ...

Thanks Tihomir. Is it possible to have Mplus generate graphics for the estimated latent class sample means when using exposure? Without exposure, PLOT3 generates all the right graphics, but when the exposure is included in the model no such graphics are generated? Any suggestions on what might be going on?

Also, any ideas where I might find a sample inp and out file for the the analysis conducted by Kreuter, F. & Muthen, B. (2008)? I've seen reference to it elsewhere on the discussion board... I'd like to cross check mine with that analysis to make sure im interpreting my own results correctly!

Thanks!

In some cases, we do not give these plots. There is no way to request them if they are not given automatically.

Email Frauke Kreuter for the input and output.

Thanks Linda. I have emailed Frauke and he was kind enough to oblige.

I am, however, having some issues running the negative binomial LCGA with exposure. I keep getting the following error:

WARNING: THE SAMPLE COVARIANCE OF THE INDEPENDENT VARIABLES IN CLASS 1 IS SINGULAR.

THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE

TRUSTWORTHY FOR SOME PARAMETERS DUE TO A NON-POSITIVE DEFINITE FIRST-ORDER DERIVATIVE PRODUCT MATRIX. THIS MAY BE DUE TO THE STARTING VALUES BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION. THE CONDITION NUMBER IS 0.314D-10. PROBLEM INVOLVING PARAMETER 1.

For your information, the model includes 20 observed dependents (ch11-ch30). These dependents are regressed on the defined exposure (ln((365-prison_time)/365)) as isuggested by Tihomir, but this only occurs for 15 of the 20 dependents (ch16-ch20) because the exposure variable is invariant for the remaining five (ch11-ch15). I wondered whether it was the uneven exposure that was causing the problem?

Thanks in advance.

Jason

Jason Payne posted on Sunday, April 05, 2009 - 2:10 am

Hi Linda,

Just to follow-up from my previous post. I note the discussion above about the validity of including an offset in a negative binomial model. I would just like to say that I get the same error message when estimating the model as a standard Poisson.

Jason

Linda K. Muthen posted on Sunday, April 05, 2009 - 12:42 pm

Please send your input, data, output, and license number to support@statmodel.com.

Kim Runyon posted on Friday, February 17, 2012 - 1:31 pm

I'm trying to use Mplus to analyze data using a Poisson LGM. My situation is unique because I'm trying to model a DV that is a proportion. More specifically, my unit of analysis is school district and I'm trying to model changes in the proportion of English Language Learner test-takers over four consecutive time points. I have referred to the Mplus discussion board for advice on modeling variables that are rates/proportions and attempted to run the syntax below,

where t02 through t05 are the number of test-takers in the school district each year and ell02c though ell05c are the number of English Language Learner test-takers in the school district each year. I cannot get the model to converge. Is there an error in my syntax?
VARIABLE: NAMES ARE id t02 t03
t04 t05 ell02c ell03c ell04c ell05c;
USEVARIABLES ARE ell02c ell03c ell04c ell05c
exp02 exp03 exp04 exp05;
COUNT = ell02c - ell05c;
MISSING = .;
DEFINE: exp02=log(t02);
exp03=log(t03);
exp04=log(t04);
exp05=log(t05);
MODEL: i s | ell02c@0 ell03c@1 ell04c@2 ell05c@3;
ell02c-ell05c on exp02-exp05@1;

Linda K. Muthen posted on Friday, February 17, 2012 - 1:58 pm

You should not log transform a count variable. Also, run the growth model first before you add the ON statement.

Dena Pastor posted on Monday, February 20, 2012 - 2:02 pm

I'm working with Kim and believe that clarification might be needed for her 2/17/2012 post. We were modeling our syntax after that provided by Asparouhov in an earlier post on Thursday, February 26, 2009 (in response to a post by Payne on 2/25/2009). The exp02 through exp05 variables we are creating are the offset variables. We believe exp02 through exp05 are created correclty, but are unsure how to include these offsets in the growth model.

Linda K. Muthen posted on Tuesday, February 21, 2012 - 8:48 am

The ON statement includes the offsets but it should be:

ell02c ON exp02@1;
ell03c ON exp03@1;
ell04c ON exp04@1;
ell05c ON exp05@1;

The way it is specified above it crosses the left- and right-hand sides.

Boris Forthmann posted on Wednesday, February 13, 2013 - 4:28 am

Hello,

I'm trying to fit a Poisson regression model with a constant off-set term, which means that the variance of the off-set term is zero and MPlus prints an error.

Here is an example in code:

```
VARIABLE:
NAMES ARE a1-a3 b1-b3 off1-off6 off_a1-off_a3 off_b1-off_b3;
COUNT ARE a1-b3;
ANALYSIS:
ESTIMATOR IS ML;
DEFINE:
off_a1 = log(off1);
off_a2 = log(off2);
off_a3 = log(off3);
off_b1 = log(off4);
off_b2 = log(off5);
off_b3 = log(off6);
MODEL:
!incorporating offsets:
a1 on off_a1@1;
a2 on off_a2@1;
a3 on off_a3@1;
b1 on off_b1@1;
b2 on off_b2@1;
b3 on off_b3@1;
!latent model part:
a by a1-a3@1;
a@1;
b by b1-b3@1;
b@1;
!end of code.
```

My goal is to get a model like this:

log(count) = intercept + 1*fscore + 1*log(offset)

I am interested in the intercepts, while taking into account the constant offset terms. Do you have any suggestions how to run this model with a constant offset?

Thank you very much!

Boris

Boris Forthmann posted on Wednesday, February 13, 2013 - 7:52 am

Okay, I tried to fit the model described above by integrating a model constraint.

Here is the code:

```
VARIABLE:
NAMES ARE a1-a3 b1-b3;
COUNT ARE a1-b3;
ANALYSIS:
ESTIMATOR IS ML;
MODEL:
!label intercepts:
[a1] (int_a1);
[a2] (int_a2);
[a3] (int_a3);
[b1] (int_b1);
[b2] (int_b2);
[b3] (int_b3);
!latent model part:
a by a1-a3@1;
a@1;
b by b1-b3@1;
b@1;
MODEL CONSTRAINT:
NEW(newint_a1 newint_a2 newint_a3 newint_b1 newint_b2 newint_b3);
int_a1 = newint_a1 + log(1.56);
int_a2 = newint_a2 + log(2.5);
int_a3 = newint_a3 + log(0.83);
int_b1 = newint_b1 + log(2.21);
int_b2 = newint_b2 + log(0.5);
int_b3 = newint_b3 + log(0.67);
```

! the second term is the known constant
! (the offset) in each line.
!end of code.

This code produced the desired output and coefficients were comparable to those estimated with R. Do you think this is the right way to implement the above model in MPlus?

Thank you very much!

Boris

Bengt O. Muthen posted on Wednesday, February 13, 2013 - 2:27 pm

Your Model Constraint approach is the way to go when your offset does not vary across subjects.

Boris Forthmann posted on Thursday, February 14, 2013 - 4:27 am

Hello Mr. Muthen,

thank you very much for your quick response. Your advices on this webpage are very helpful.

Boris