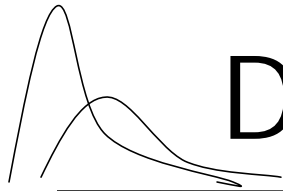


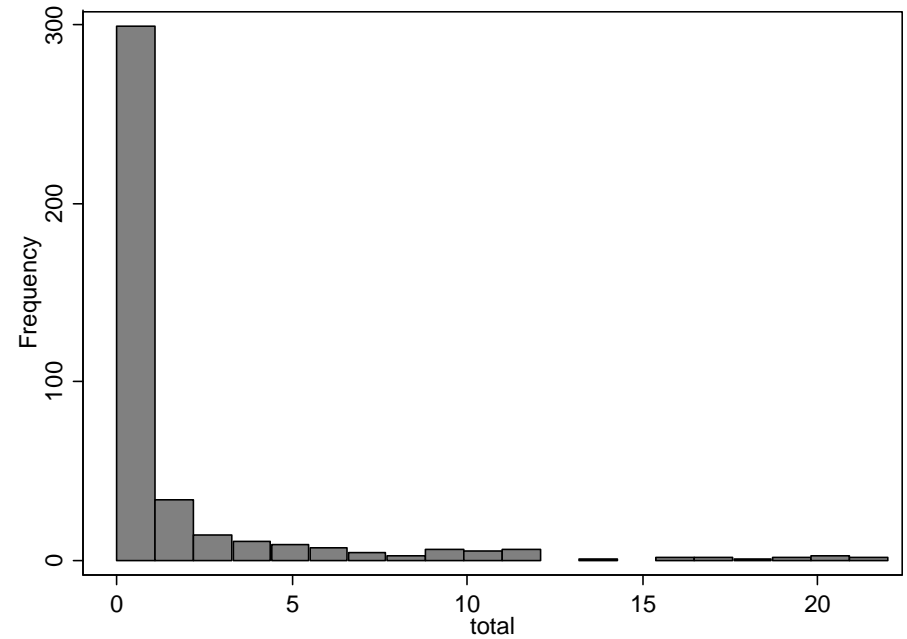
Longitudinal Models with Zero Inflation

Frauke Kreuter, UCLA Statistics



Data

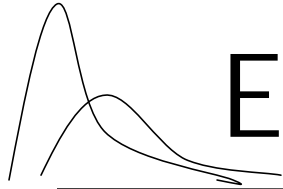
- Characteristics
 - Large number of zeros
 - Skewed
- Examples
 - Convictions
 - Drinking
 - Drug abuse





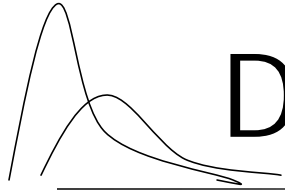
“Classic” examples

- Manufacturing applications
- Economics
- Medicine
- Public health
- Environmental science
- Education



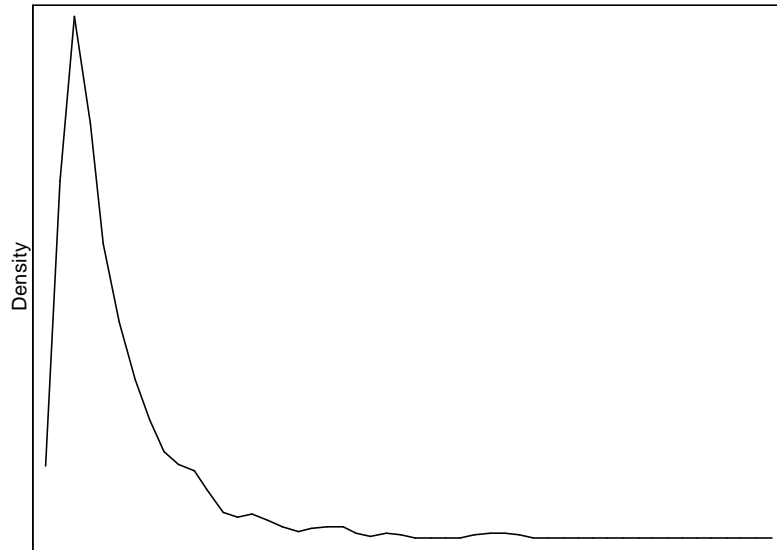
Example – Criminal behavior

- “Cambridge study” (Farrington/West 1990)
- N = 411 (403)
- Age 10 to 40
- Number of convictions each year
- 60 percent never convicted
- In any given year 98.5% to 88.8% zero
- Biannual 97.1% to 83.2% zero

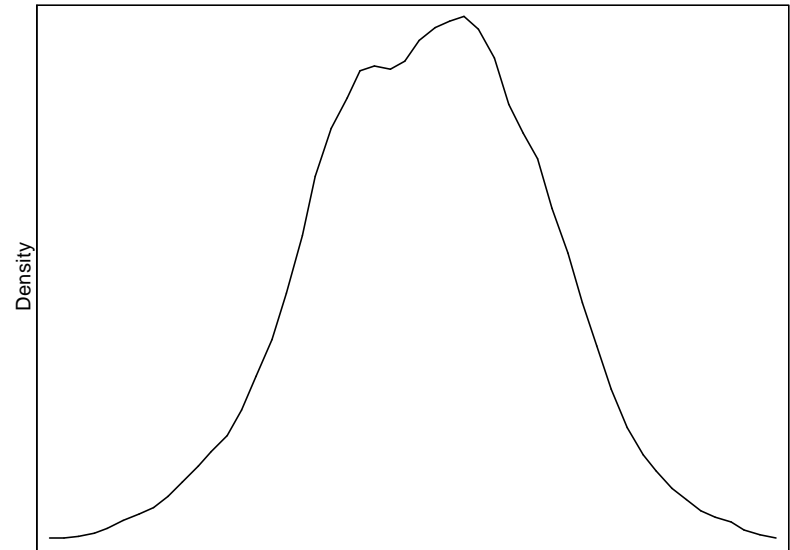


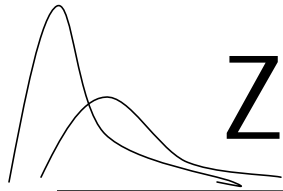
Distributions

- Skewed

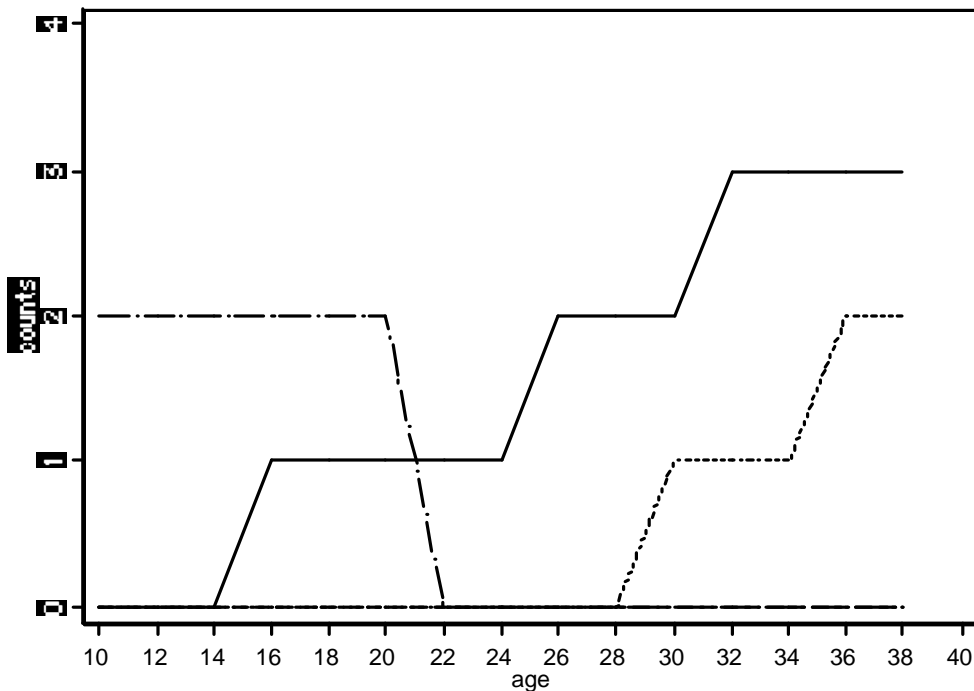


- Log transformed





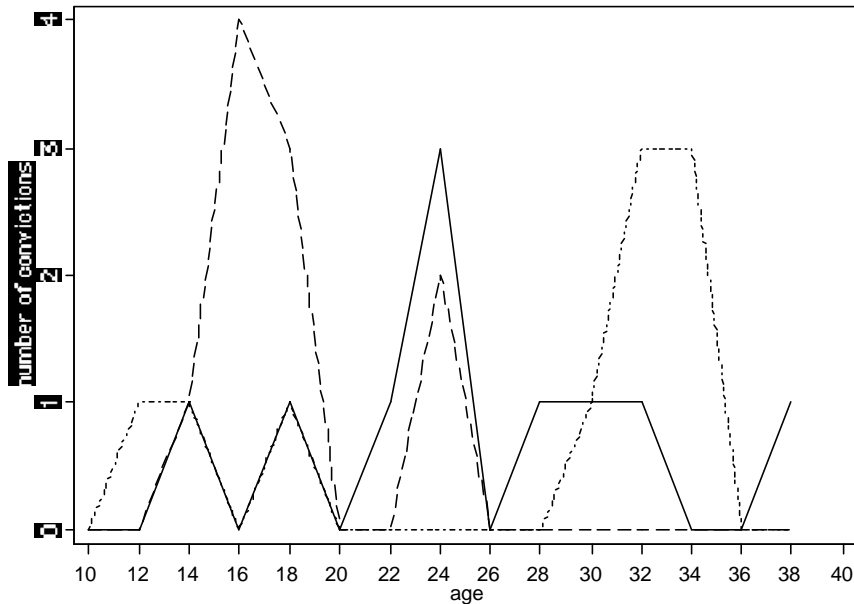
Zeros in longitudinal data



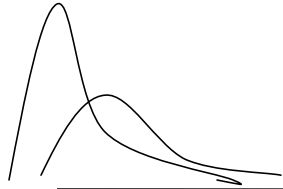
- Onset
 - Zero until onset
 - Once behavior shown we want to model counts
- “Offset”
 - Zero after a certain point
 - E.g. abstinence, “jail time”



“Types” of zeros



- “Careers”
 - In and out of zero (random zeros)
 - Measurement errors
 - Zero throughout (partly structural zeros)



One/Two-class models: Modeling alternatives

- One-class models
 - Censored normal
 - *Two-part modeling (Olsen & Schafer 2001)*
 - Onset followed by growth (Albert & Shih 2003)
- Two-class models
 - *Inflated models*
 - Zero inflated poisson
 - Censored inflated
 - Two-part modeling (mixture in 0-1 part)
 - *Two-class (Carlin et al. 2001)*
- Multi-class models



Modeling alternatives: Multi-class models

- Mixture censored
- Two-part models with mixture in 0-1 and growth part
 - LTA mover-stayer model
 - Two-class model for 0-1 part
- Onset followed by growth with mixture
- *LCGA with zero inflation (Roeder et al. 1999)*
- *GMM with zero class (Mplus V3)*



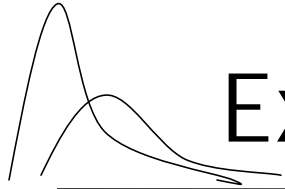
Substantive questions reg. zeros

- What is the process that generates zeros?
- Do we assume a mixture of zeros?
- Who should contribute to trajectory classes?

- Are covariates allowed to have different effects on zeros and positive outcomes?
- Are different covariates allowed for zeros and positive outcomes?

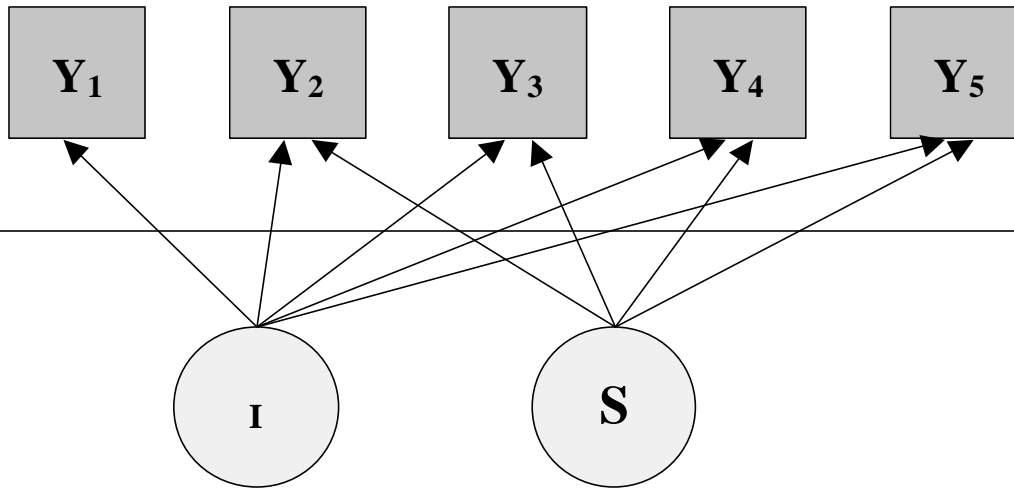


Separate process for zeros



Example – Olsen & Schafer (2001)

- **Data** $n=1961$
 - Panel of Adolescent Prevention Trial
 - Middle school and high school students
 - Grade 7 through 11
- **Variables**
 - Self reported recent alcohol use
 - Parental monitoring, rebelliousness, gender



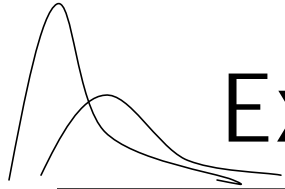
$$y_{it} = I_{0i} + S_{1i}a_{it} + \mathbf{e}_{it}$$

$$I_{0i} = \mathbf{a}_0 + \mathbf{V}_{0i}$$

$$S_{1i} = \mathbf{a}_1 + \mathbf{V}_{1i}$$

Assume, for simplicity ,

$$a_{it} = a_t = 0, t_1, t_2, \dots, T$$



Example – Olsen & Schafer (2001)

- Model: two-parts

- U-part

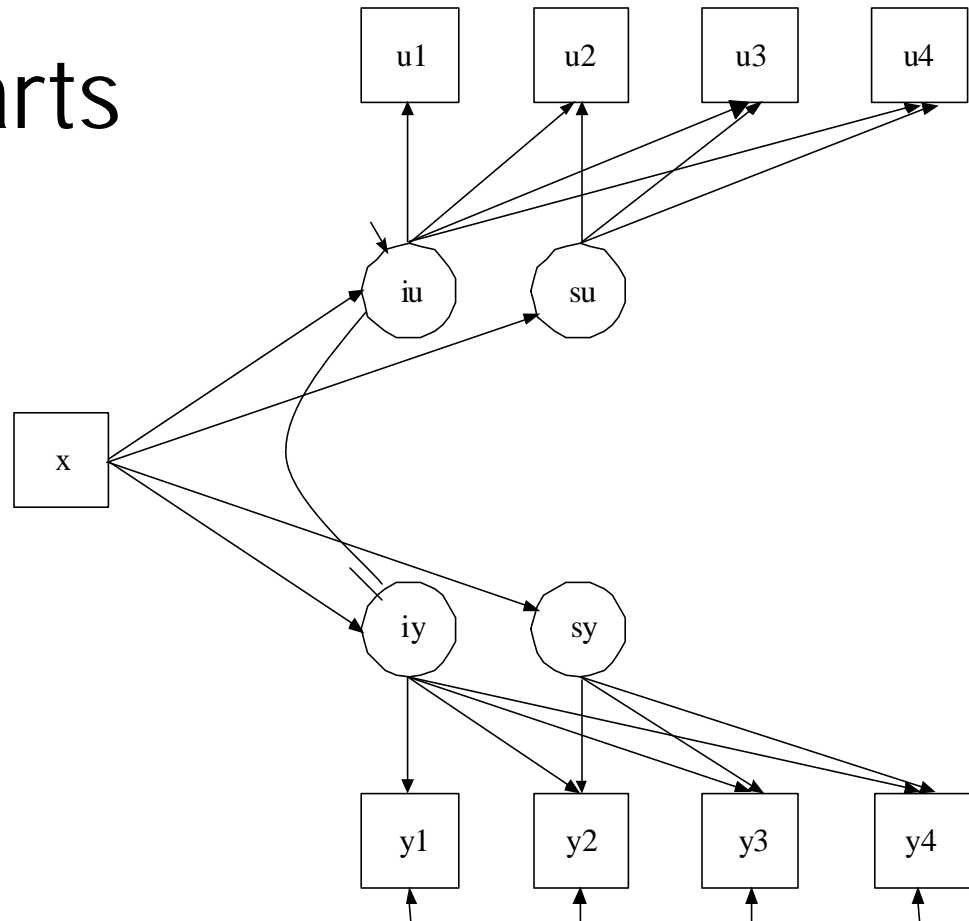
- Logit

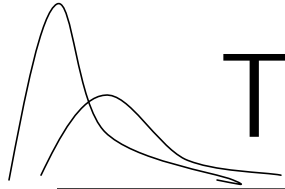
- use, no-use

- Y-part

- Log-normal

- $y > 0$



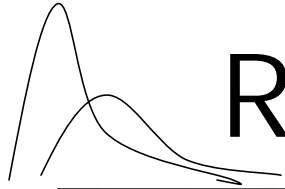


Two-part model

$$u_{it} = \begin{cases} 1 & \text{if } y_{it} > 0 \\ 0 & \text{if } y_{it} = 0 \end{cases}$$

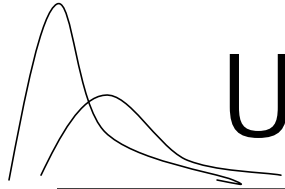
$$y_{it} = \begin{cases} m_{it} & \text{if } y_{it} > 0 \\ . & \text{if } y_{it} = 0 \end{cases}$$

$$m_{it} = \mathbf{h}_{0i} + \mathbf{h}_{1i} a_{it} + \mathbf{e}_{it}$$



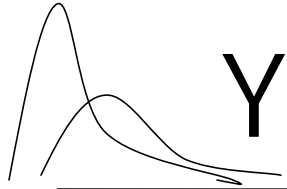
Raw data

i	<i>Grade</i> 7	<i>Grade</i> 8	<i>Grade</i> 9	<i>Grade</i> 10	<i>Grade</i> 11
1	0	0	0	0	0
2	0	0	1.7	2.3	3
3	0	1	0	1	1.7



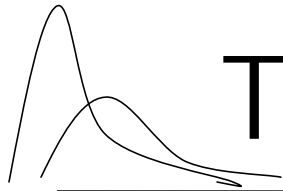
U-part

i	<i>Grade</i> 7	<i>Grade</i> 8	<i>Grade</i> 9	<i>Grade</i> 10	<i>Grade</i> 11
1	0	0	0	0	0
2	0	0	1	1	1
3	0	1	0	1	1



Y-part

i	<i>Grade</i> 7	<i>Grade</i> 8	<i>Grade</i> 9	<i>Grade</i> 10	<i>Grade</i> 11
1
2	.	.	1.7	2.3	3
3	.	1	.	1	1.7



Two part modeling - *Mplus*

VARIABLE:...

CATEGORICAL = u1-u4;

MODEL:

```
iu su | u1@0 u2@1 u3@2 u4@3;
```

```
iy sy | y1@0 y2@1 y3@2 y4@3;
```

```
iu-sy on x;
```

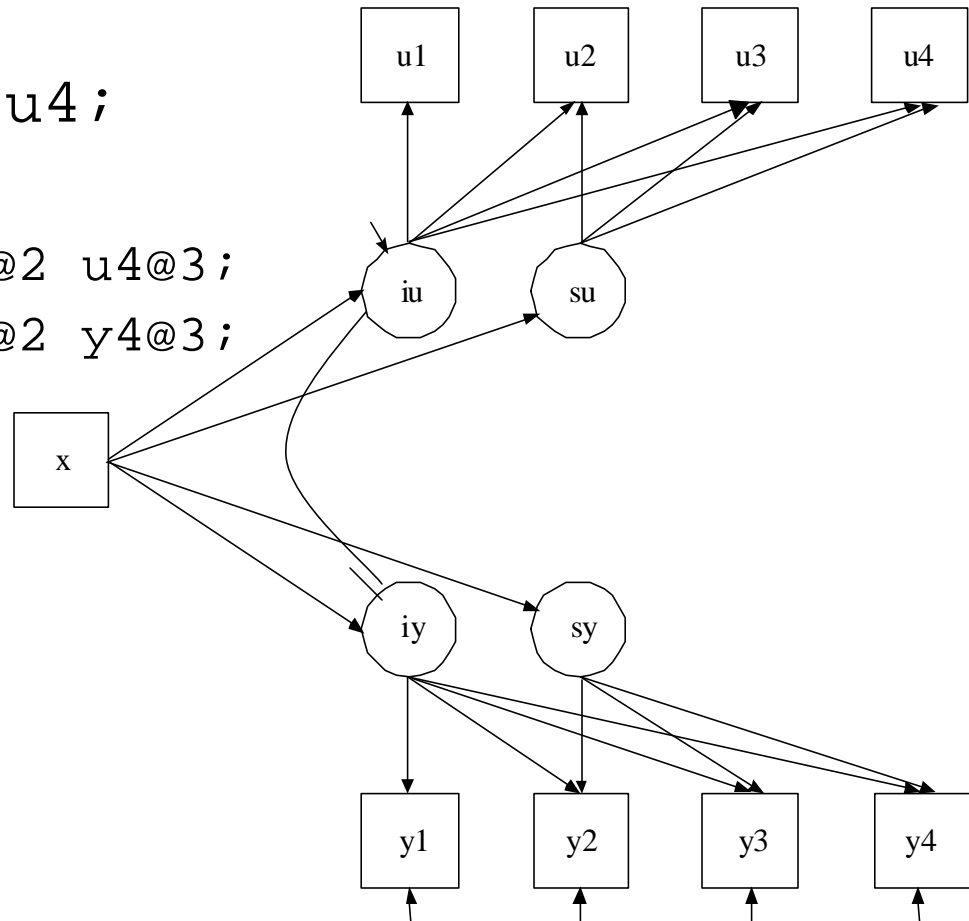
```
su@0; sy@0;
```

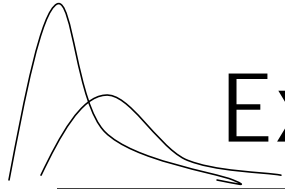
```
iu with su@0;
```

```
iy with sy@0;
```

```
iu with sy@0;
```

```
su with iy-sy@0;
```





Example – Olsen & Schafer (2001)

■ U-part

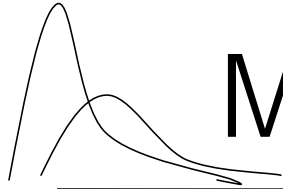
- In grade 7 unsupervised girls higher odds for drinking, effect diminishes over time
- Low monitoring in grade 7 no effect for boys, but unsupervised boys higher odds in grade 11

■ Y-part

- Reduced monitoring increase the amount of alcohol consumption in grade 7
- For girls effect increases over time, for boys it vanishes by grade 11



Separate sources for zeros



Modeling count variables (t=1)

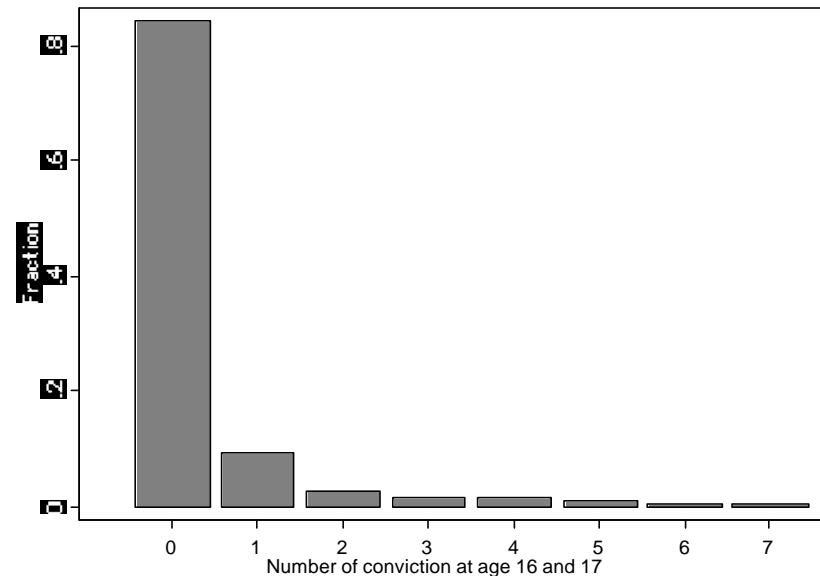
- Models used

- Poisson

- With parameter λ

- Use GLM to model $\log(\lambda)$

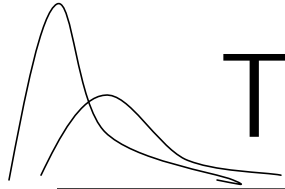
- Negative binomial



- Problem

- Zero inflation / overdispersion

- Model assumptions don't hold



The poisson model

$$\Pr(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \quad \text{Poisson model}$$

$$E(Y) = \lambda \text{ and } V(Y) = \lambda$$

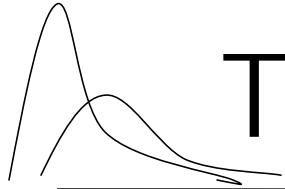
Example: Probability for count equal to 5, with lamda 4.5

$$\Pr(Y = 5) = \frac{e^{-4.5} 4.5^5}{5!} = .1708$$



Zero Inflated Poisson (ZIP)

- Zero outcome can arise from one of two sources, one where outcome is always zero, another where a poisson process is at work (Lambert 1992)
- The poisson process can produce zero or another outcome
- Covariates can predict group membership, and outcome of the poisson process



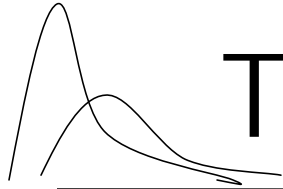
The model

$$\Pr(y_{it} = 0) = \Pr(\text{group1}) + \Pr(y_{it} = 0 | \text{group2}) * \Pr(\text{group2})$$

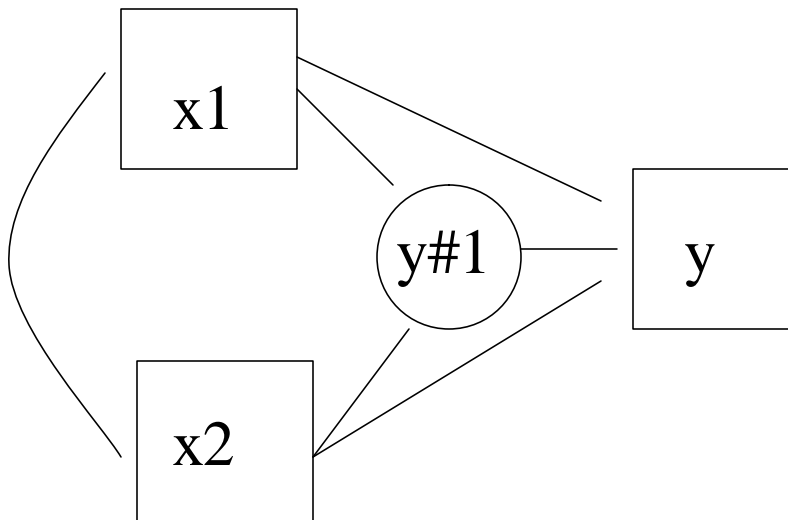
$$\Pr(y_{it} = 1) = \Pr(y_{it} = 1 | \text{group2}) * \Pr(\text{group2})$$

$$\Pr(\text{group1}) = \frac{e^L}{1 + e^L} \quad \text{Logistic regression model}$$

$$\Pr(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \quad \text{Poisson model}$$

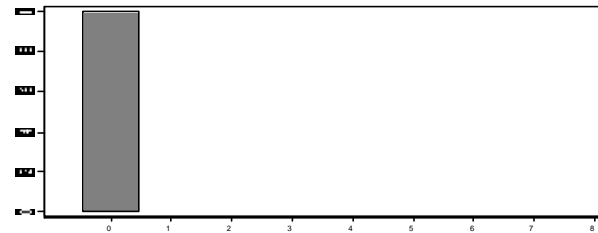


Two class models

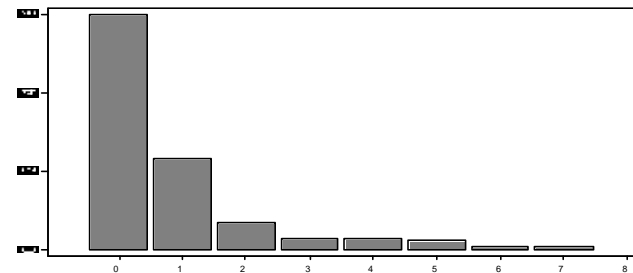


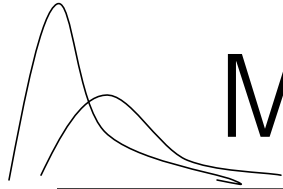
$y\#1$ is a two class variable

$y\#1 = 1$: “zero class”
 $P(y=0)=1$



$y\#1 = 2$: “convicted”
Poisson distribution





Mplus example - ZIP

```
!Input file
```

```
VARIABLE:
```

```
  NAMES ARE ...      ;  
  missing = .        ;  
  USEV      = ...     ;  
  COUNT     = total (i);
```

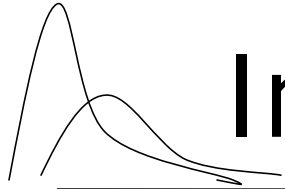
```
MODEL:
```

```
  total      ON x1 x2 ;  
  total#1    ON x1 x2 ;
```

```
MODEL RESULTS
```

```
Estimates
```

```
TOTAL      ON  
  norear    0.453  
  daring    0.434  
  
TOTAL#1     ON  
  norear    -0.260  
  daring    -0.952  
  
Intercepts  
  TOTAL#1   0.816  
  TOTAL     1.031
```



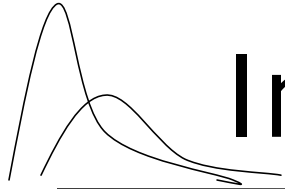
Interpreting the zip part

Odds of being in zero class: $e^{0.816} = 2.261$

Probability to be in zero class:

$$odds = \frac{\Pr(\text{zero_class})}{1 - \Pr(\text{zero_class})} = 2.26$$

$$\Pr(\text{zero_class}) = \frac{2.26}{1 + 2.261} = .693$$



Interpreting the count part

The average rate of conviction (average number of crimes) given a person is in the non-zero class, and both covariates are equal to zero:

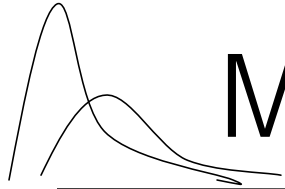
$$e^{1.031} = 2.804$$

The average rate of conviction for boys with $x_1=0$ and $x_2=0$

$$= 2.804 * (1 - \text{Pr}(\text{zero_class}))$$

$$= 2.804 * (1 - .693)$$

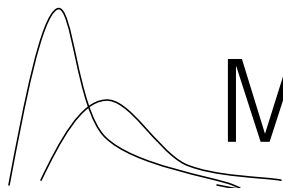
$$= .861$$



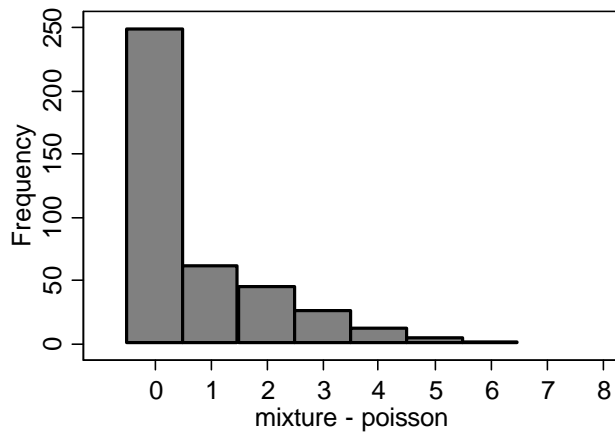
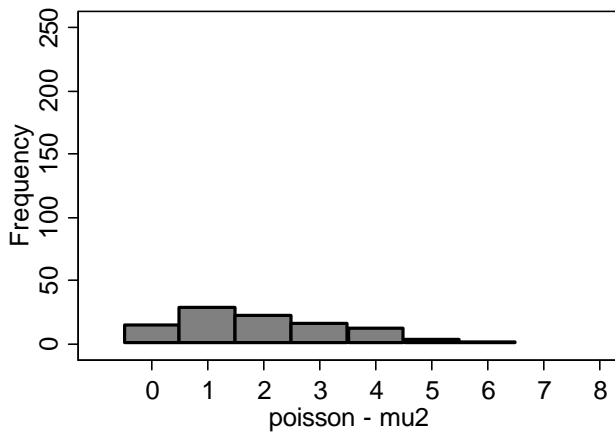
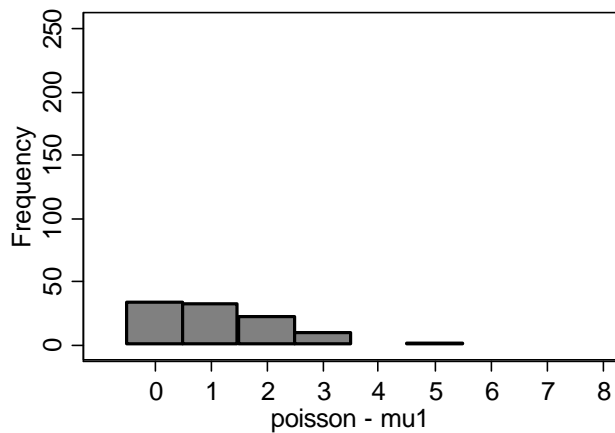
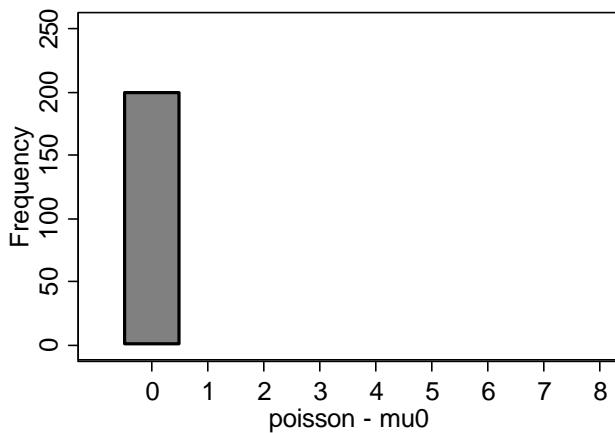
Mplus example – two classes

MODEL:

```
%overall%  
total ON norear daring;  
! class 1 has P(u=0)=1 ("zero class")  
! class 2 is regular Poisson  
c#1 on norear daring;  
%c#1%  
[total@-15];  
total on norear@0 daring@0;  
%c#2%
```

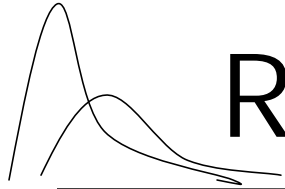


Multi-class model



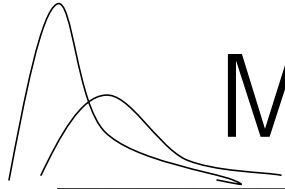


Growth models



Raw data

i	<i>Age</i> <i>10</i>	<i>Age</i> <i>12</i>	<i>Age</i> <i>14</i>	<i>Age</i> <i>16</i>	<i>Age</i> <i>18</i>
1	0	0	1	1	0
2	0	0	0	2	1
3	0	0	0	0	0



Mplus for (zero inflated) count

VARIABLE:

NAMES = ...

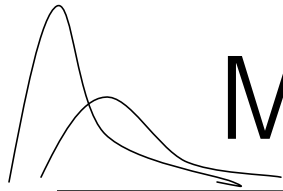
COUNT = cage10 - cage30(i);

MODEL:

i s q | cage10@0 ... cage30@10;

ii si qi | cage10#1@0 ... cage30#1@10;

s-qi@0;



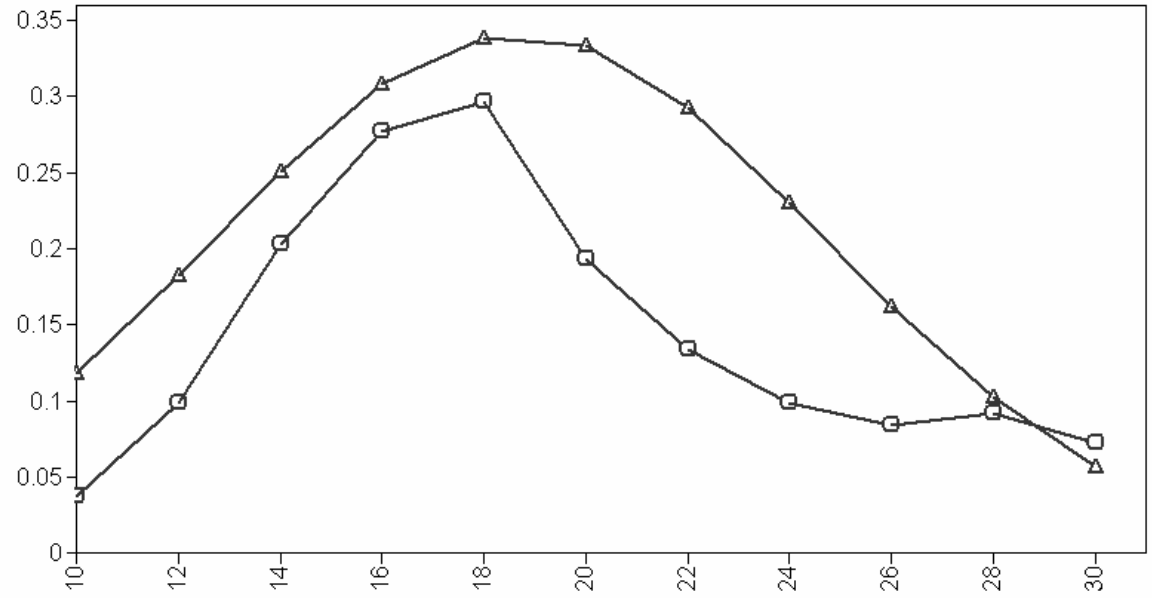
Mplus for (zero inflated) count

H0 Value	-1519.762
Free Param.	4
Bayesian	3063.530
est.	t
I	-4.265 -18.967
S	0.488 7.781
Q	-0.056 -8.023

H0 Value	-1496.937
Free Parameter	7
Bayesian (BIC)	3035.878
	est. t
I	-3.231 -10.461
S	0.130 1.272
Q	-0.027 -2.802
II	@0
SI	-1.776 -3.765
QI	0.132 4.866
CAGE10#1	1.532 2.861

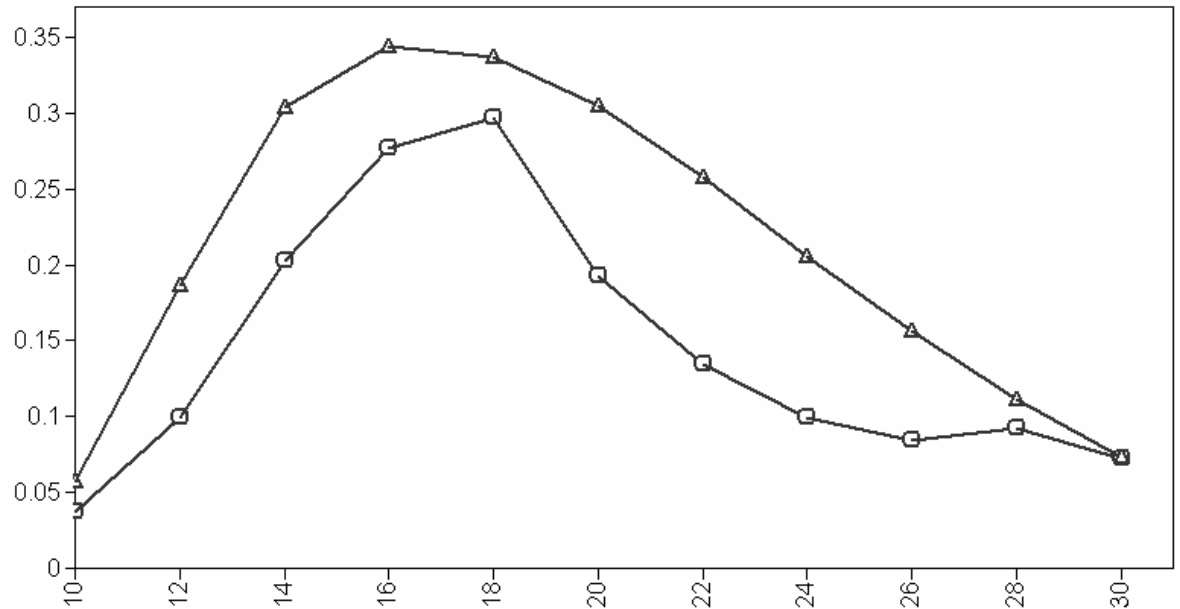
GMM

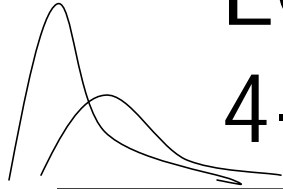
Predicted/ observed average conviction rate



GMM-ZIP

Predicted/ observed average conviction rate





LCGA with ZIP

4-classes

VARIABLE:

```
NAMES      = ... ;
missing    = . ;
USEV       = cage10-cage30;
COUNT     = cage10-cage30(i);
CLASSES    = c(4);
```

ANALYSIS:

```
TYPE       = MIXTURE;
ALGORITHM   = INTEGRATION;
```

MODEL:

```
%OVERALL%

i  s  q  |  cage10@0 ..  cage30@10;
ii si qi |  cage10#1@0 .. cage30#1@10;

i-qi@0;
```

Loglikelihood H0 Value	-1450.004
------------------------	-----------

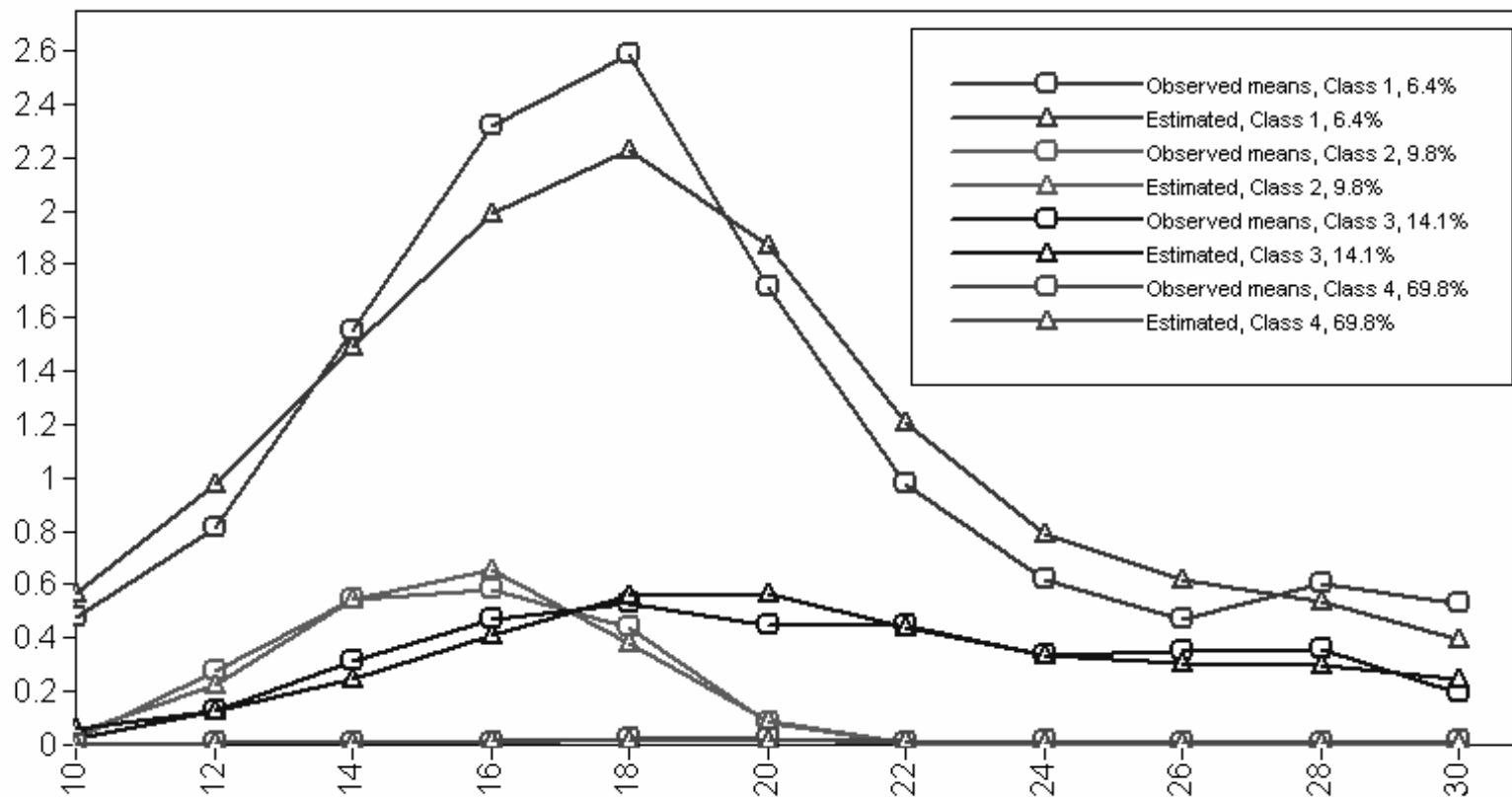
Information Criteria

Number of Free Parameters	18
Bayesian (BIC)	3008.033
Entropy	0.811

LO-MENDELL-RUBIN ADJUSTED LRT TEST

Value	26.463
P-Value	0.7244

Outlook Latent class growth model with 4 classes





GMM with random intercept

using Poisson, 3 classes of which one is structural zero

VARIABLE:

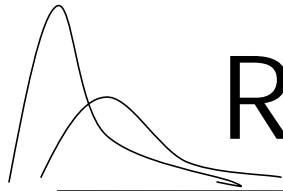
```
NAMES          = ... ; missing = . ;  
USEV           = cage10-cage30 ;  
COUNT        = cage10-cage30 ;  
CLASSES      = c(3) ;
```

ANALYSIS:

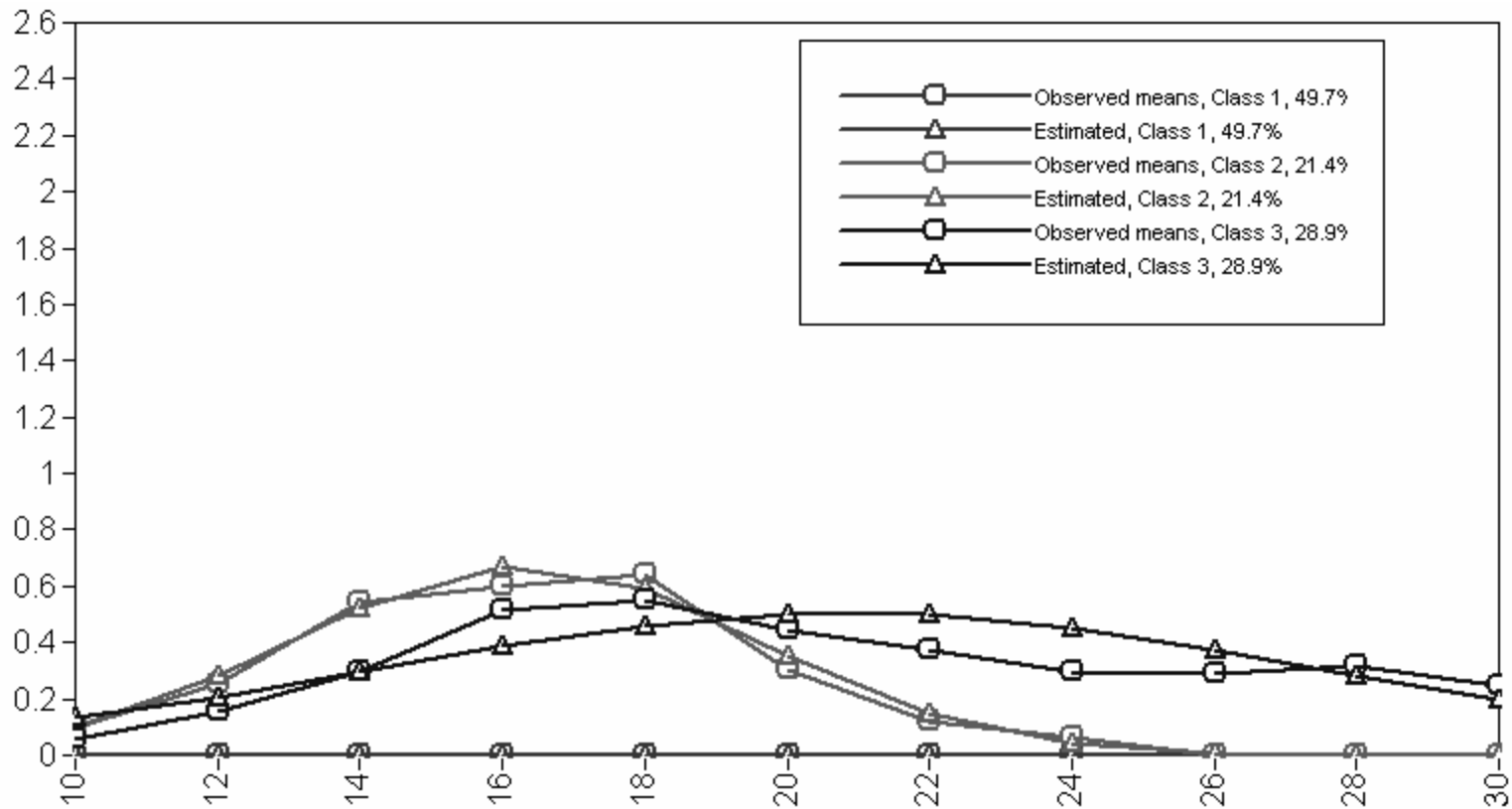
```
TYPE           = MIXTURE ;  
ALGORITHM      = INTEGRATION ;
```

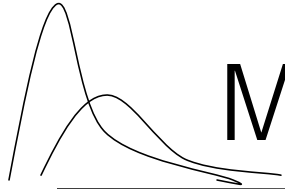
MODEL:

```
%OVERALL%  
i s q | cage10@0 .. cage30@10 ;  
s-q@0 ;  
%c#1% ! zero class throughout  
[i-q@0] ;  
[cage10-cage30@-15] ;
```



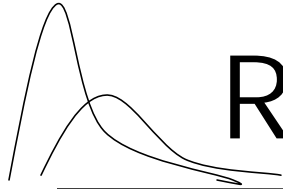
Results GMM zero class





Model comparison & evaluation

- Model fit
 - Loglikelihood
 - BIC
- Predictive power for distal outcomes
- Description
 - Entropy
 - Classification table
 - Comparing classifications across models



References (selection)

- Brown, R., Catalano, C., Fleming, C.B., Haggerty, K.P., Abbott, R.D. (2004): Adolescent Substance Use Outcomes in the Raising Healthy Children Project: A Two-Part Latent Growth Curve Analysis. Manuscript presented at SPR.
- Land, K.C., McCall, P.L., & Nagin, D.S. (1996): A Comparison of Poisson, Negative Binomial, and Semiparametric Mixed Poisson Regression Models. *Sociological Methods & Research*, 24, 387-442.
- Olsen, M.K. & Schafer, J.L. (2001): A Two-Part Random-Effects Model for Semicontinuous Longitudinal Data. *Journal of the American Statistical Association*, 96, 730-745.
- Roeder, K., Lynch, K.G. & Nagin, D.S. (1999): Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association*, 94, 766-776.