MLE: Categorical and Limited Dependent Variables

Unit 2: Ordered, Multinomial, Count, and Limited Dependent Variables

3a. Poisson Regression

PS2730-2021

Week 7-8

Professor Steven Finkel

Limited Dependent Variables: Counts

- So far have talked about models for different kinds of categorical data – 2 choice, ordered 3+ choice, non-ordered 3+ choice with individual and choice-specific independent variables.
- Next segment of course will be devoted to situations where DV is another kind of non-continuous variable, what is called *limited*, in that can only take on some values and not others.
- One kind of limited DV is very common in social science, and that is a variable that represents a *count* of something, how many times something occurred, or how many things a person knows or does, etc. Example: how many acts of political participation a person engages in, how many wars a country is involved in, how many terrorist attacks experienced, how many presidential vetoes in a legislative session. Limited by 0 on the left, and must be integer value on the right.
- Other limited DVs: *censored*, *sample-selected* which we'll get to later

- Count and other limited variables often analyzed using OLS or regression models as if the DV was really continuous. But leads to nonsense predictions about negative numbers of wars, negative participations, or, in censoring case, erroneous conclusions because the clump of values at 60 can lead the regression line to be very far off its true value. So we move to other methods to handle these situations.
- Begin with count data
- Many count models available, depending on how the data were generated, and the distribution of the dependent variable
- **Poisson, Negative Binomial, Zero Inflated, Hurdle Models**: differ primarily due to theoretical considerations about how distribution of counts came about, how the 0s versus non-zero values may have been generated, and whether the models can empirically accurately account for the number of zeros in the data, a common problem in the estimation of count models
- All build on Poisson Regression as foundation

```
tab polpart
```

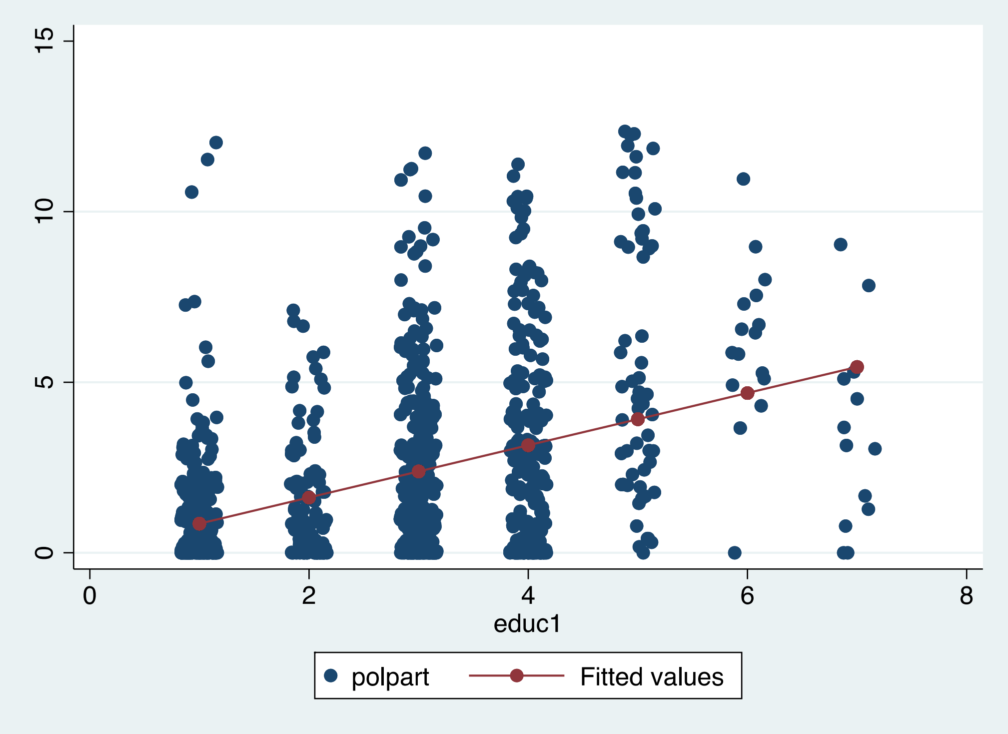
polpart	Freq.	Percent	Cum.
0	346	36.81	36.81
1	145	15.43	52.23
2	125	13.30	65.53
3	93	9.89	75.43
4	63	6.70	82.13
5	44	4.68	86.81
6	34	3.62	90.43
7	23	2.45	92.87
8	17	1.81	94.68
9	19	2.02	96.70
10	12	1.28	97.98
11	10	1.06	99.04
12	9	0.96	100.00
Total	940	100.00	

- Example in South African data: political participation, runs from 0-12
- How many acts of participation engaged in over the past 2 years? Some people do nothing (37%), some people a few acts, some people do 10 or more (3%)
- Two features of count data: 1) consist solely of non-negative integers; and 2) often very skewed in terms of the distribution

- What happens if we use OLS on this DV?

```
regress polpart educ1

      Source |       SS           df       MS      Number of obs   =       940
-------------+----------------------------------   F(1, 938)       =    152.93
       Model |  1027.39097         1  1027.39097   Prob > F        =    0.0000
    Residual |  6301.57073       938  6.71809246   R-squared       =    0.1402
-------------+----------------------------------   Adj R-squared   =    0.1393
       Total |   7328.9617       939  7.80507104   Root MSE        =    2.5919

------------------------------------------------------------------------------
     polpart |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       educ1 |   .7654651   .0618986    12.37   0.000     .6439894    .8869408
       _cons |   .0892404   .1972857     0.45   0.651     -.297932    .4764128
------------------------------------------------------------------------------
```

You can start seeing some of the problems: Negative predictions, heteroskedastic and non-normal error variance; possible non-linearity in relationships



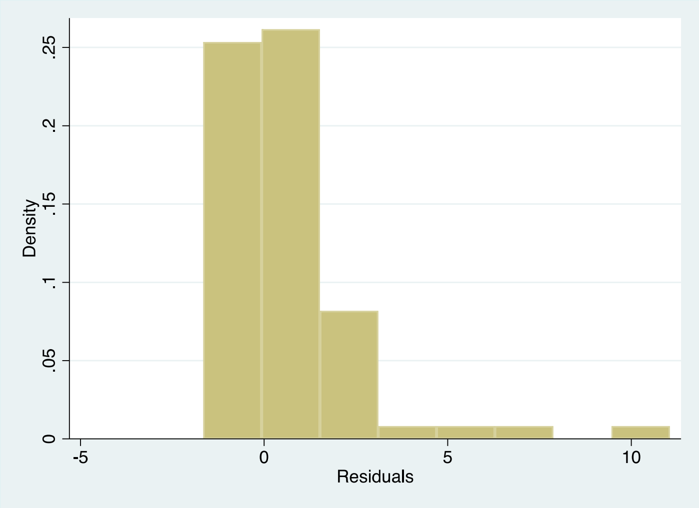
- Multivariate example: Polpart as function of education, civic ed exposure, political interest, group memberships

```
summarize olsprd if olsprd<0
```

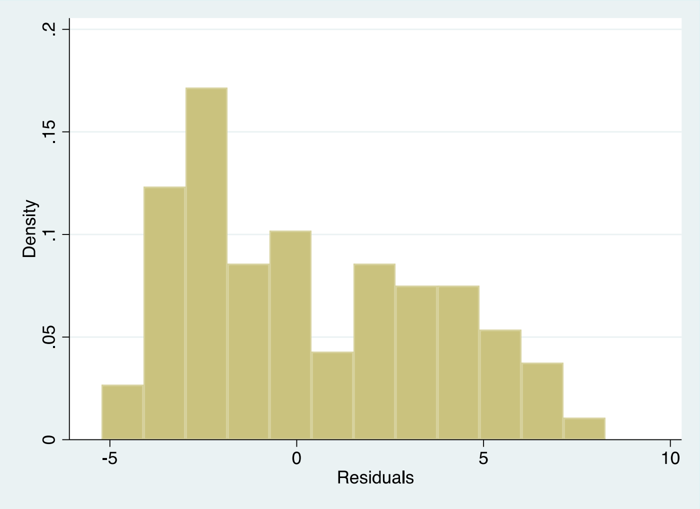
Variable	Obs	Mean	Std. Dev.	Min	Max
olsprd	65	-.4877836	.4593379	-1.982731	-.0028756

- 65 cases with negative participation predictions, average of nearly -.5 and one -1.98 prediction!
- So
- OLS PROBLEM #1: POTENTIAL NEGATIVE PREDICTIONS

• OLS PROBLEM #2: SKEWED (NON-NORMAL) RESIDUALS

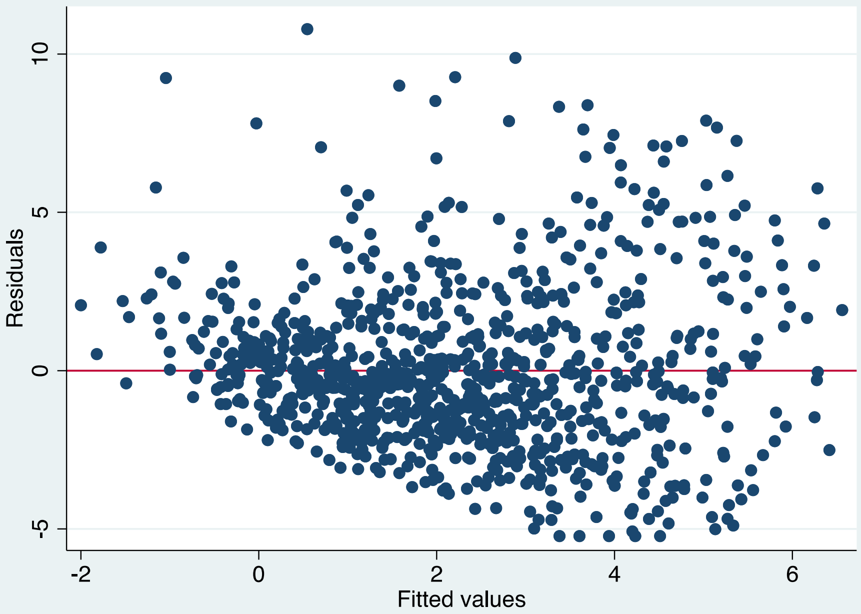


OLS Residuals when GROUPS=0



OLS Residuals when GROUPS=5

- OLS PROBLEM #3: HETEROSKEDASTIC ERROR VARIANCE



Error Variance increases as predicted Y increases

```
summarize olsres if groups==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      olsres |        227   -.1420849    1.632169  -2.975147    10.3427

summarize olsres if groups==3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      olsres |        164    -.158865    2.254063  -4.579563   7.821376

summarize olsres if groups==5

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      olsres |        166    .3419108    3.303307  -5.223773   8.260745
```

All this suggests that count data is likely not to follow a normal distribution at all, and not one that has a constant variance that can be modeled with OLS or normal regression techniques, plus relationship is possibly non-linear in the first place

- What to do? Initial Model: Assume that data follows a "Poisson distribution"
- Poisson: a theoretical distribution that is often used to model rare events, things that happen relatively infrequently.
- Poisson distribution:
  - non-negative,
  - often highly skewed
  - generates intrinsic non-linearities between the conditional means and the IVs
  - has the characteristic that the mean and variance are the same
- This means that as the mean increases, the variance increases, so it is a promising distribution for the kind of heteroskedasticity that is often seen in count data.
- As we will see, sometimes in count distributions the variance is even *greater* than the mean, and large numbers of zeros are also possible problems for Poisson. So other methods developed to compensate for these deficiencies

- When model the conditional mean as a function of Xs, will see how the relationship is non-linear, which is also another advantage of Poisson.
- Can look at Poisson regression as modeling a non-linear relationship that always predicts positive outcomes while accommodating skewed distributions on Y
- Form of the Poisson distribution: governed by a single term μ, the mean of the distribution, which is also equal to its variance

$$\Pr(y \mid \mu) = \frac{\exp^{-\mu}(\mu^{y})}{y!}$$

- This gives the probability of observing a 0, 1, 2, 3 ….. etc, given a value of μ, the mean of the distribution or what is termed the “rate” of occurrence, or the **expected** number of times something will happen in the given time interval
- So given an “expected” or “average frequency”, with what probability do we observe particular frequencies 0, 1, 2, etc.?

- Example: μ =.85

$$\Pr(y\,|\,.85)=\frac{\exp^{-.85}(.85^{y})}{y!}$$

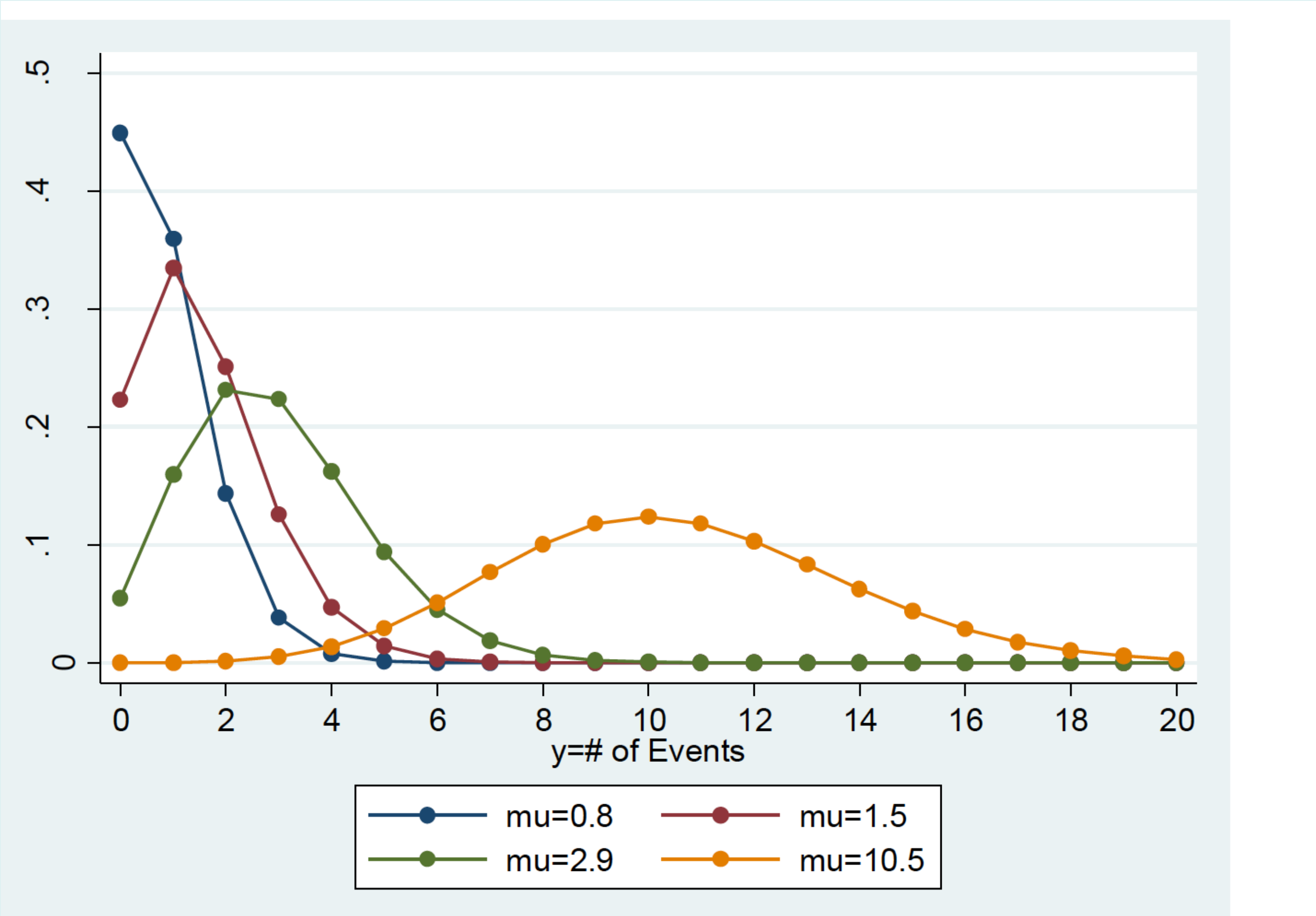
$$\Pr(0\,|\,.85)=\frac{\exp^{-.85}(.85^{0})}{0!}=\exp^{-.85}=.43$$

$$\Pr(1\,|\,.85)=\frac{\exp^{-.85}(.85^{1})}{1!}=\exp^{-.85}(.85)=.36$$

$$\Pr(2\,|\,.85)=\frac{\exp^{-.85}(.85^{2})}{2!}=\frac{\exp^{-.85}(.85^{2})}{2*1}=.15$$

$$\Pr(6\,|\,.85)=\frac{\exp^{-.85}(.85^{6})}{6!}=\frac{\exp^{-.85}(.85^{6})}{6*5*4*3*2*1}=.0002$$

- Can plot these values for different μ



Can view the distribution as governed by the latent "rate" parameter μ, which generates different probability sets of 0/1/2/3/ etc., depending on the value of the rate or the "average" of the distribution

- Demonstrates several features of the Poisson distribution
  - No negative values in the Poisson distribution. Can be seen in the denominator of the distribution, as the factorial of a negative number is undefined.
  - μ is the mean of the distribution. As the mean increases the bulk of the distribution shifts to the right
  - The variance equals the mean (also known as "equidispersion"): $Var(Y) = E(Y) = \mu$. As as μ increases, can see that the variance also increases to match it
  - As μ increases, the probability of 0s decreases
  - As μ increases, the Poisson distribution approximates the normal distribution

- Important assumptions for using Poisson distribution:

1. The observations are independent
   - One act of participation doesn't influence the probability of another one; in a given time period one terrorist incident doesn't influence another in a given time period, etc. This would be violated, e.g., if there are *contagion* effects
2. There is no over-dispersion (or under-dispersion).
   - The variance equals the mean, or, when we expand the model to include independent X variables, the *conditional* variance, given the Xs, equals the *conditiona*l mean, given the Xs
3. There are no more 0s than would be predicted by the Poisson distribution

- If these assumptions are violated, we either modify our model or move to alternative models for count data

- Begin modeling political participation by seeing how it looks compared to a Poisson distributed variable with the same mean. POLPART has a mean of 2.29; what does a Poisson variable with that mean look like in terms of the distribution of counts?

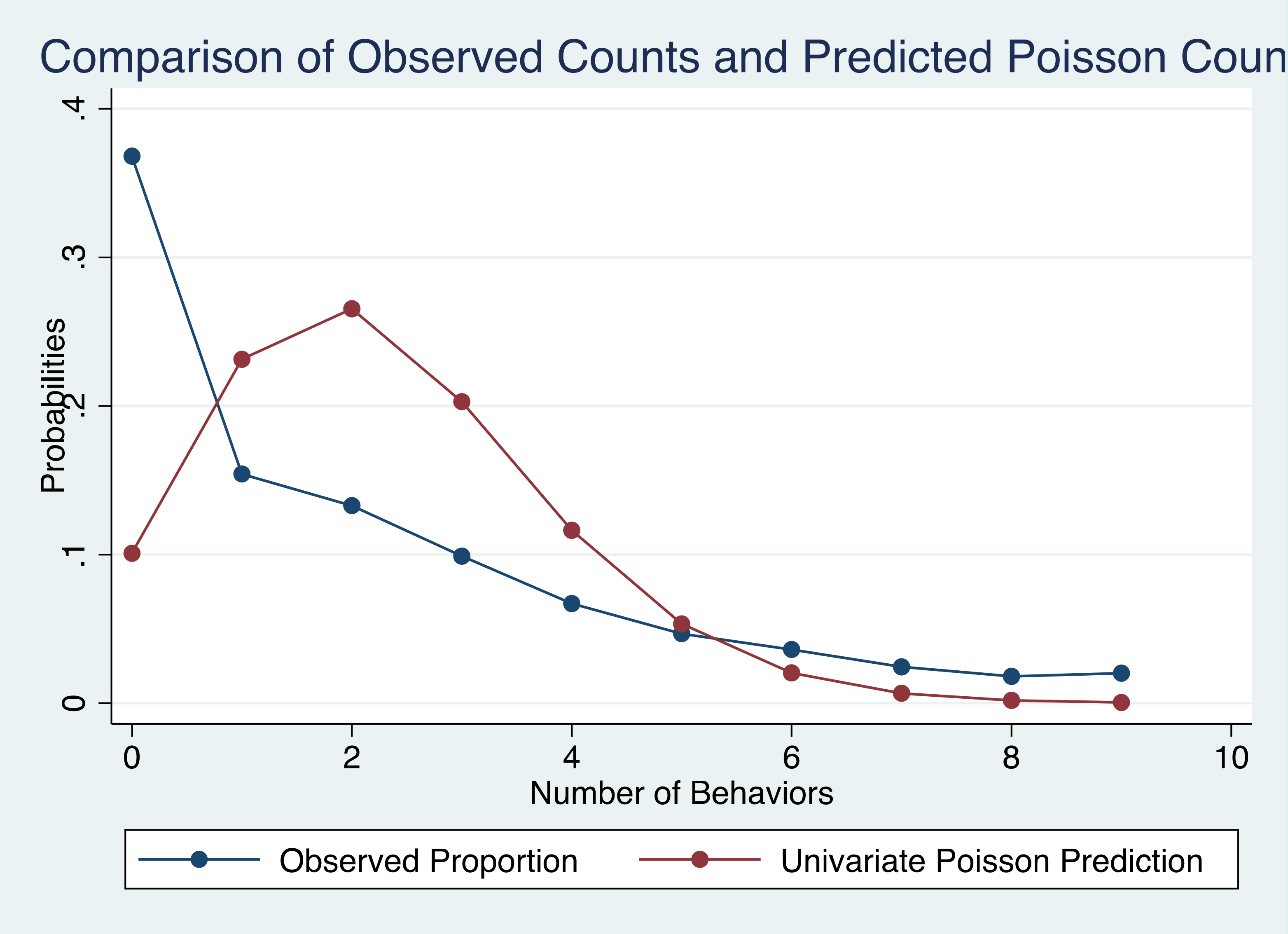
$$\Pr(0 \mid 2.29) = \frac{\exp^{-2.29}(2.29^0)}{0!} = \exp^{-2.29} = .10$$

$$\Pr(1 \mid 2.29) = \frac{\exp^{-.2.29}(2.29^1)}{1!} = \exp^{-2.29}(2.29) = .23$$

$$\Pr(2 \mid 2.29) = \frac{\exp^{-.2.29}(2.29^2)}{2*1} = .27$$

$$\Pr(3 \mid 2.29) = \frac{\exp^{-.2.29}(2.29^3)}{3*2*1} = .20$$

$$\Pr(6 \mid 2.29) = \frac{\exp^{-.2.29}(2.29^6)}{6*5*4*3*2*1} = .02$$



Comparison of Observed Counts and Predicted Poisson Coun
Probabilities
0
.1
.2
.3
.4
0
2
4
6
8
10
Number of Behaviors
Observed Proportion
Univariate Poisson Prediction

- Differences?
  - Poisson predicts far too few 0s than in actual distribution
  - Poisson predicts far too many low counts, too few high counts
  - This indicates *overdispersion* in the actual data; we can confirm this by looking at the variance of POLPART, which is 7.81 compared to the mean of 2.29!
- What accounts for overdispersion?
  - One possible reason: **individual heterogeneity**
  - Until now we are assuming ***one*** rate parameter which is constant across all units. This is unrealistic! Should not the latent rate of participation be different for highly educated, poorly educated, etc.?
  - If we ignore this heterogeneity, we'll typically see such overdispersion since more variance in the rate parameter will translate into more variance in the observed counts
  - If we include additional variables to predict the latent rate, we may achieve *conditional* equidispersion, such that $Var(y|x) = \mu_i|x$

- We want the rate to always be positive, so we can predict the rate as a very simple exponentiated function of the independent variables:

$$\mu = E(y \mid X) = \exp(XB)$$

- This is Poisson regression: the X variables predict the latent rate of occurrence of some (assumed poisson-distributed) outcome, which then generates predictions of different distributions of counts for all observations with similar latent rates

$$\Pr(y \mid X) = \frac{\exp(-\mu)(\mu^{y})}{y!} = \frac{\exp(-\exp(XB))(\exp(XB))^{y}}{y!}$$

where $\mu = \exp(XB)$

- And once we estimate μ_i conditional on the Xs, we can then assess whether the conditional distribution of counts is equidispersed as one measure of the fit of the model to the data

- This makes for a very straightforward GLM version of the Poisson model:

$$\mu = \exp^{XB}$$

$$\text{and } \ln\mu = \eta = XB$$

- So a one-unit change in the independent variables generates a linear β change in the "log-rate" parameter
- So the Poisson model is "linear in the log-rate" and non-linear in the rate
- Remember in GLM: g(μ) is the "linearizing link" of a non-linear response function, and the "mean function" g^{-1} (η) gets you back to μ from the linear fuction

$$g(\mu) = \ln(\mu) = \eta = XB$$

$$g^{-1}(\eta) = \exp^{\mu} = \exp^{XB}$$

ML Estimation of Poisson Regression

- Steps:

- Assume a probability distribution for Y – Poisson in this case
- Express the joint probability of the data (i.e., all of the Y) using the assumed probability distribution
- Calculate the joint probability of the data given the parameters–the "likelihood function" (taking the log of the likelihood to simplify)
- Maximize this function with respect to the unknown parameters (e.g., the Bs in the regression function)

$$\Pr(y \mid X) = \prod_{i=1}^{N} \frac{\exp(-\mu_i)(\mu_i^{y_i})}{y_i!}$$

$$L(\beta \mid y, X) = \prod_{i=1}^{N} \frac{\exp(-\exp(XB)(\exp(XB)^{y_i})}{y_i!}$$

$$\ln L(\beta \mid y, X) = -n\exp^{XB} + \sum_{i=1}^{N} y_i X\beta - \sum_{i=1}^{N} \ln(y_i!)$$

- Which is maximized wrt the β
- ML estimates are those that generated the highest predicted probability of observing the count for each unit that was observed, given the predicted rate and the assumed Poisson distribution of the outcomes

```
. poisson polpart

Iteration 0:   log likelihood = -2409.8308
Iteration 1:   log likelihood = -2409.8308

Poisson regression                              Number of obs     =        940
                                                LR chi2(0)        =       0.00
                                                Prob > chi2       =          .
Log likelihood = -2409.8308                     Pseudo R2         =     0.0000

------------------------------------------------------------------------------
     polpart |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .8301301   .0215365    38.55   0.000     .7879192    .8723409
------------------------------------------------------------------------------
```

- The univariate Poisson regression (i.e., no independent variables)
- Estimated constant=.83
- So predicted μ=exp(.83) = 2.29
- Mean of POLPART=2.29!!!
- Every unit as the same predicted rate, which translates into the distribution of predicted outcomes seen on the graph on slide 16
- Log-likelihood maximized at -2409.83, with no further iterations since the mean of POLPART was just plugged in and maximum was achieved

```
. poisson polpart groups

Iteration 0:   log likelihood =  -2066.933
Iteration 1:   log likelihood =  -2066.933

Poisson regression                              Number of obs     =        940
                                                LR chi2(1)        =     685.80
                                                Prob > chi2       =     0.0000
Log likelihood =  -2066.933                     Pseudo R2         =     0.1423

------------------------------------------------------------------------------
     polpart |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      groups |   .3582984   .0140516    25.50   0.000     .3307578    .3858389
       _cons |  -.2296391   .0524266    -4.38   0.000    -.3323933    -.126885
------------------------------------------------------------------------------
```

- One independent variable: Log-rate=-.23+.358*Groups
- Each additional group increases the predicted log-rate by .358
- Each additional group increases the rate by a constant factor exp(.358)=1.43
- Log-likelihood maximized at -2066.93
- Model chi-square= G^2 = 2 ln L (Full Model) – 2 LnL(Reduced Model)
- G^2 = 2 (-2066.93)- 2 (-2409.83) = 685.8
- Pseudo R-squared=(-2409.83 –(-2066.93))/ (-2409.83)=.142

Interpretation: Impact of X on μ

- Exponentiate the β to get the factor change in the rate for an additional unit (or standard unit) change in X
- Increasing groups by 1 leads to a 1.43 factor change in the rate
- Predicted rate for 0 groups: exp(-.23)=.79
- 1 group: exp(-.23+.358)=1.137 (which is .79*1.43)
- 2 group: 1.63 (which is 1.137*1.43), etc.
- Percent change in the rate is the (factor change*-1)*100
- So every unit change in X changes the predicted rate by 43%!!

```
. listcoef, help

poisson (N=940): Factor change in expected count

  Observed SD:  2.7938
```

	b	z	P>|z|	e^b	e^bStdX	SDofX
groups	0.3583	25.499	0.000	1.431	1.761	1.579
constant	-0.2296	-4.380	0.000	.	.	.

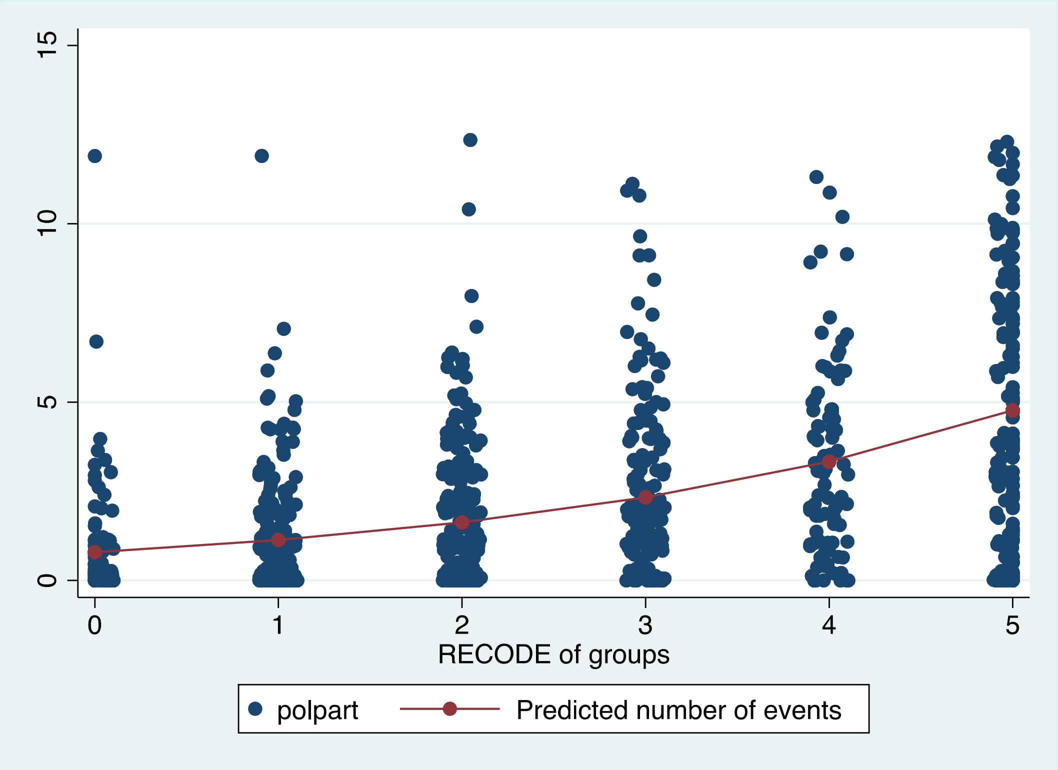
```
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in expected count for unit increase in X
 e^bStdX = exp(b*SD of X) = change in expected count for SD increase in X
   SDofX = standard deviation of X
```

```
. listcoef, percent help

poisson (N=940): Percentage change in expected count

  Observed SD:  2.7938
```

	b	z	P>|z|	%	%StdX	SDofX
groups	0.3583	25.499	0.000	43.1	76.1	1.579
constant	-0.2296	-4.380	0.000	.	.	.

```
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
       % = percent change in expected count for unit increase in X
   %StdX = percent change in expected count for SD increase in X
   SDofX = standard deviation of X
```

- Increasing by 2 groups leads to a factor change of exp(.358*2)=2.05 etc.
- This suggests a non-linear effect of X on the rate, since a one-unit change in X leads to a 1.43 factor change in the rate, while a 2 unit change leads to a 2.05 factor change in the rate, etc.



- So Poisson is a non-linear model of the effect of the independent variables on the rates or expected counts!!

```
. poisson polpart educ1 civiced groups interest

Iteration 0:   log likelihood = -1892.9049
Iteration 1:   log likelihood = -1892.9024
Iteration 2:   log likelihood = -1892.9024

Poisson regression                              Number of obs     =        940
                                                LR chi2(4)        =    1033.86
                                                Prob > chi2       =     0.0000
Log likelihood = -1892.9024                     Pseudo R2         =     0.2145
```

polpart	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
educ1	.1747755	.0164016	10.66	0.000	.142629	.206922
civiced	.1905193	.0460366	4.14	0.000	.1002892	.2807495
groups	.2354591	.015671	15.03	0.000	.2047445	.2661737
interest	.4559433	.0370617	12.30	0.000	.3833037	.528583
_cons	-2.009032	.125183	-16.05	0.000	-2.254387	-1.763678

```
. listcoef, percent

poisson (N=940): Percentage change in expected count

  Observed SD:  2.7938
```

	b	z	P>|z|	%	%StdX	SDofX
educ1	0.1748	10.656	0.000	19.1	27.0	1.367
civiced	0.1905	4.138	0.000	21.0	10.0	0.500
groups	0.2355	15.025	0.000	26.5	45.0	1.579
interest	0.4559	12.302	0.000	57.8	37.8	0.703
constant	-2.0090	-16.049	0.000	.	.	.

Interpretation: Impact of Marginal and Discrete Change in X on μ

- Can also examine the effect of X on μ from discrete or marginal change in X
- Marginal effect (slope of tangent to curve for very small change in X):

$$\frac{\partial E(y \mid X)}{\partial X_k} = E(y \mid X)\beta_k$$

- Marginal effects depends on both the regression coefficient and the predicted rate; when β is positive, the bigger the rate, the larger the marginal effect; when β is negative, the smaller
- Can compute with other variables at their observed values (default in Stata) or setting them at their mean

- Discrete change, for centered/uncentered unit/standard unit:

$$\frac{\Delta E(y \mid x)}{\Delta x_k (x_k^{start} \rightarrow x_k^{end})} = E(y \mid x, x_k^{end}) - E(y \mid x, x_k^{start})$$

```
. mchange

poisson: Changes in mu | Number of obs = 940

Expression: Predicted number of polpart, predict()

             |    Change    p-value
-------------+---------------------
educ1        |
          +1 |     0.438      0.000
         +SD |     0.619      0.000
    Marginal |     0.401      0.000
civiced      |
          +1 |     0.481      0.000
         +SD |     0.229      0.000
    Marginal |     0.437      0.000
groups       |
          +1 |     0.609      0.000
         +SD |     1.033      0.000
    Marginal |     0.540      0.000
interest     |
          +1 |     1.325      0.000
         +SD |     0.867      0.000
    Marginal |     1.046      0.000

Average prediction

    2.294
```

Note: AMEs by default

All variables held at their observed sample values; use "atmeans" option for alternative MEM

Interpretation: Impact of X on P(y_i)

- Can also see how changing IVs impacts the probability of a count being at a certain value or values, and can graph this

$$\Pr(y = k \mid X) == \frac{\exp(-\exp(x\hat{B}))(\exp(x\hat{B}))^k}{k!}$$

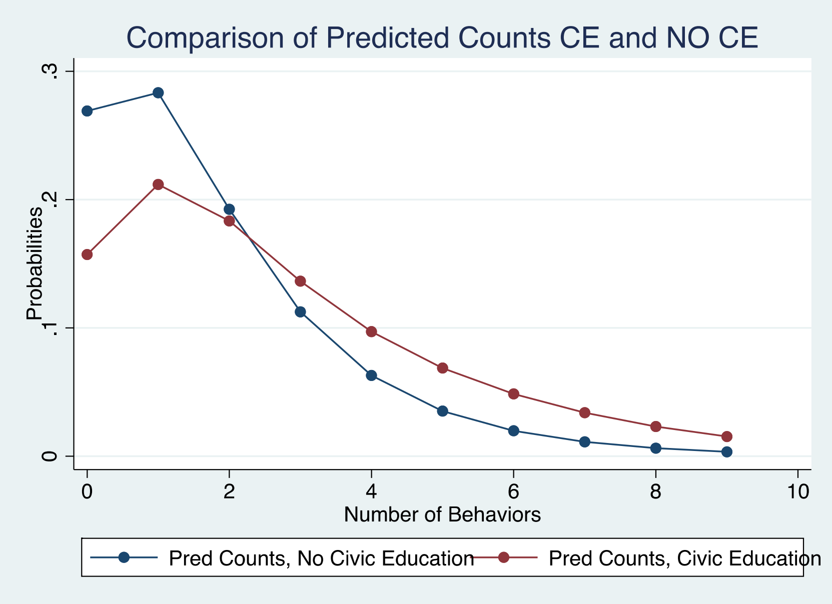
poisson: Change in Predictions for **polpart**

Confidence intervals by delta method

	Current	Saved	Change	95% CI for Change	
Rate:	1.9743	1.6318	.34248	[0.1798,	0.5052]
Pr(y=0|x):	0.1389	0.1956	-0.0567	[-0.0837,	-0.0297]
Pr(y=1|x):	0.2742	0.3191	-0.0450	[-0.0662,	-0.0238]
Pr(y=2|x):	0.2706	0.2604	0.0102	[0.0035,	0.0170]
Pr(y=3|x):	0.1781	0.1416	0.0365	[0.0194,	0.0535]
Pr(y=4|x):	0.0879	0.0578	0.0301	[0.0159,	0.0444]
Pr(y=5|x):	0.0347	0.0189	0.0159	[0.0081,	0.0236]
Pr(y=6|x):	0.0114	0.0051	0.0063	[0.0030,	0.0095]
Pr(y=7|x):	0.0032	0.0012	0.0020	[0.0009,	0.0031]
Pr(y=8|x):	0.0008	0.0002	0.0006	[0.0002,	0.0009]
Pr(y=9|x):	0.0002	0.0000	0.0001	[0.0000,	0.0002]

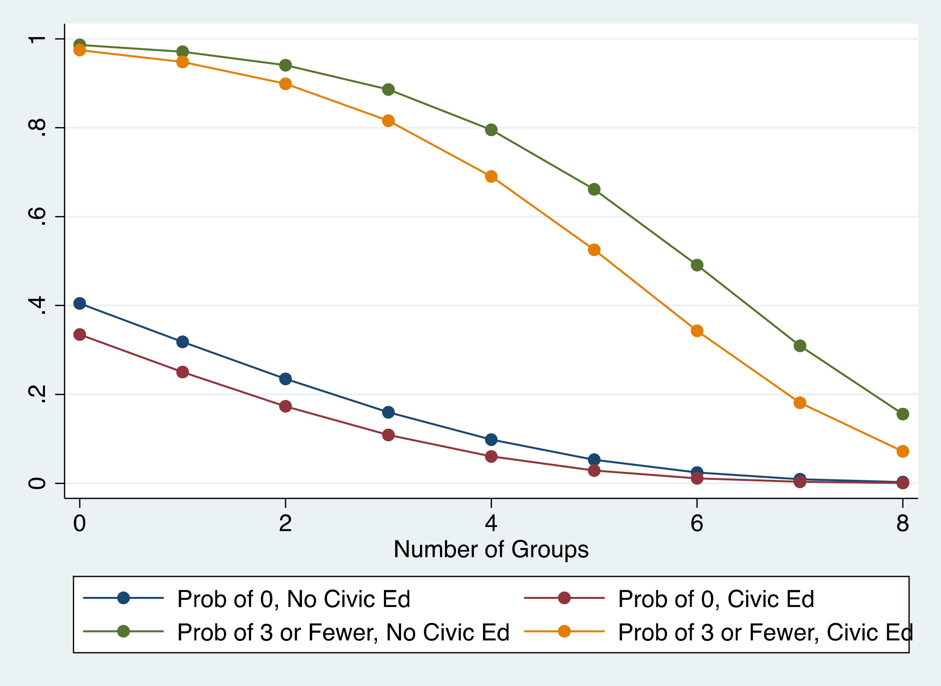
	educ1	civiced	groups	interest
Current=	2.8797872	1	2.5053191	3.0826241
Saved=	2.8797872	0	2.5053191	3.0826241
Diff=	0	1	0	0

These are differences in count probabilities for individuals who were exposed to Civic Ed and individuals who were not



LEFT: Predicted distribution of counts, no civic education versus civic education

RIGHT: Predicted probability of POLPART=0 and POLPART=3 for different levels of group memberships and civic education exposure



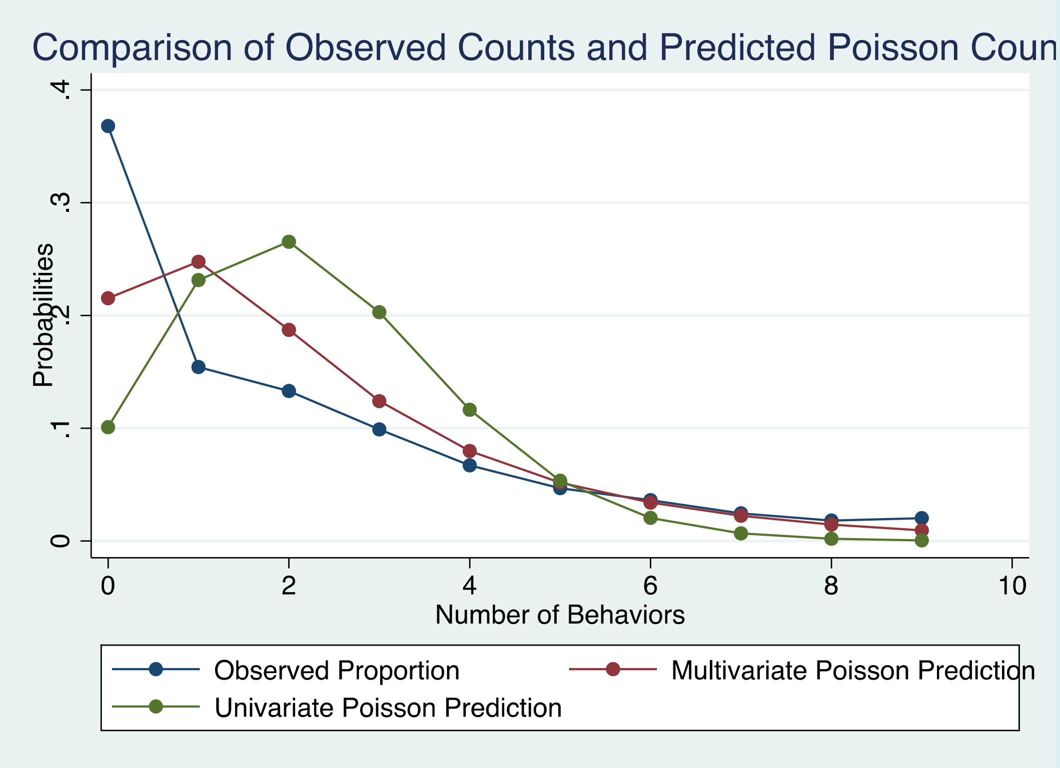
Assessing Model Fit

```
. fitstat
```

	poisson
Log-likelihood	
Model	**-1892.902**
Intercept-only	**-2409.831**
Chi-square	
Deviance(df=935)	**3785.805**
LR(df=4)	**1033.857**
p-value	**0.000**
R2	
McFadden	**0.215**
McFadden(adjusted)	**0.212**
Cox-Snell/ML	**0.667**
Cragg-Uhler/Nagelkerke	**0.671**
IC	
AIC	**3795.805**
AIC divided by N	**4.038**
BIC(df=5)	**3820.034**

Fit indices via "fitstat": Model chi-square, McFadden R-squareds, and Deviance-based Statistics (AIC and BIC)

- Finally, can see adequacy of the PRM model as a whole by comparing average probability of each count with observed data



- Observed proportion of 0 still much higher than predicted!
- Still overpredict 1/2/3 and underpredict 8/9/10
- So still have overdispersion: why?
- Contagion and/or **unobserved** heterogeneity! Move to alternative models