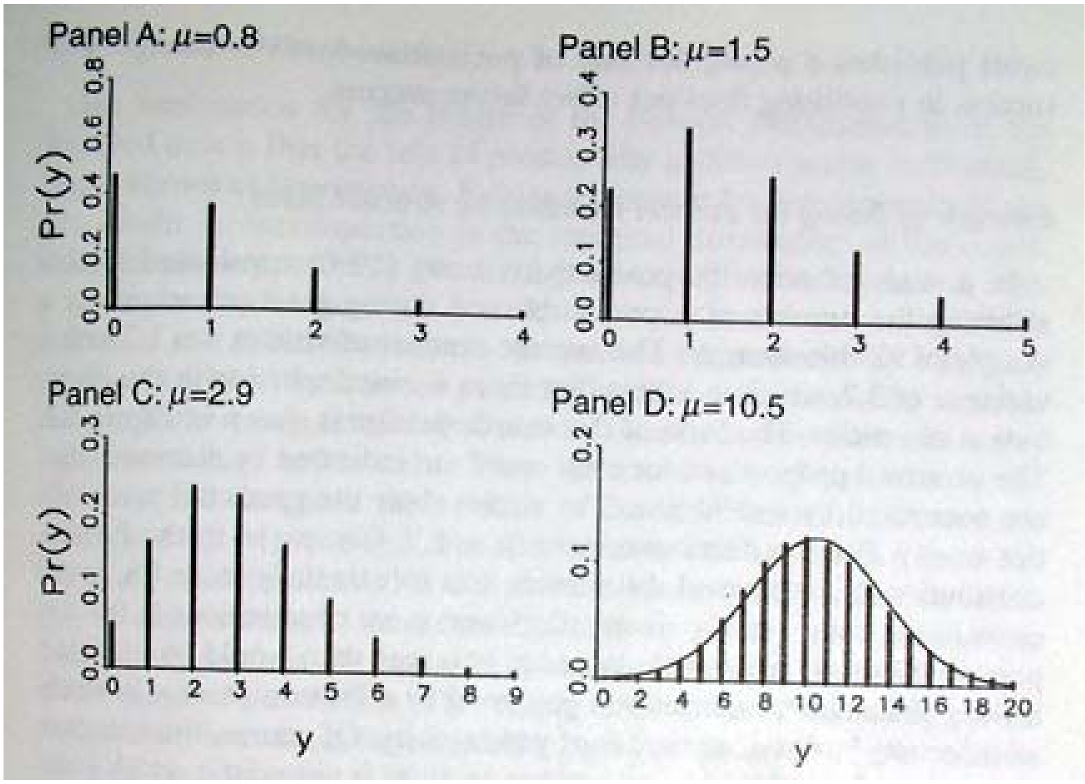
SC704: Topics in Multivariate Analysis
Instructor: Natasha Sarkisian

Poisson Regression

Count variables are often treated as though they are continuous, and OLS is used. OLS in this case can result in inefficient, inconsistent, and biased estimates. Need to use models that are developed specifically for count data. Poisson model is the most basic of them.

Poisson distributions:

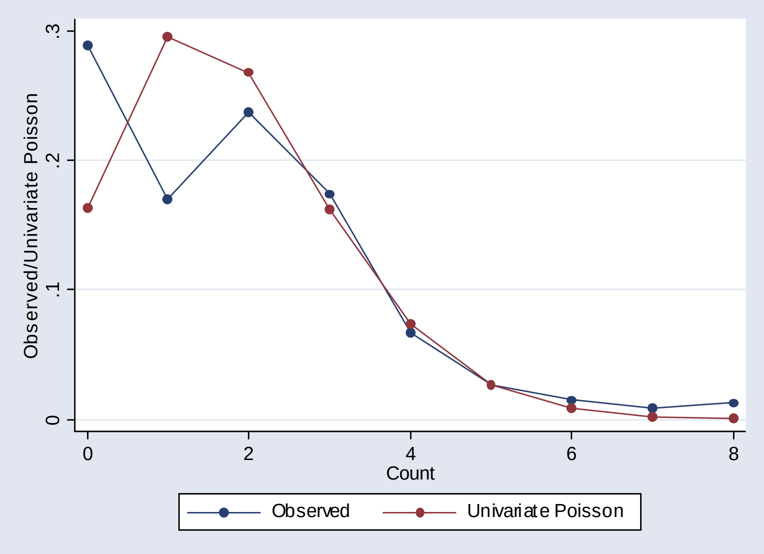


Characteristics of Poisson distribution:

1. E(y) = μ
2. The variance equals the mean: Var(y)=E(y)= μ -- equidispersion. In practice, the variance is often larger than μ: this is called overdispersion. The main reason for overdispersion is heterogeneity – if there are different groups within data that have different means and all of them are actually equal to their variances, when you put all of these groups together, the resulting combination will have variance larger than the mean. Therefore, we need to control for all those sources of heterogeneity. Thus, when using Poisson regression, we need to ensure that the conditional variance equals to the mean – that is Var(y|X)=E(y|X).
3. As μ increases, the probability of zeros decreases. But for many count variables, there are more observed zeros than would be predicted from Poisson distribution
4. As μ increases, the Poisson distribution approximates normal.
5. The assumption of independence of events – past outcomes don’t affect future outcomes.

We usually start by examining the raw distribution and comparing it with poisson:

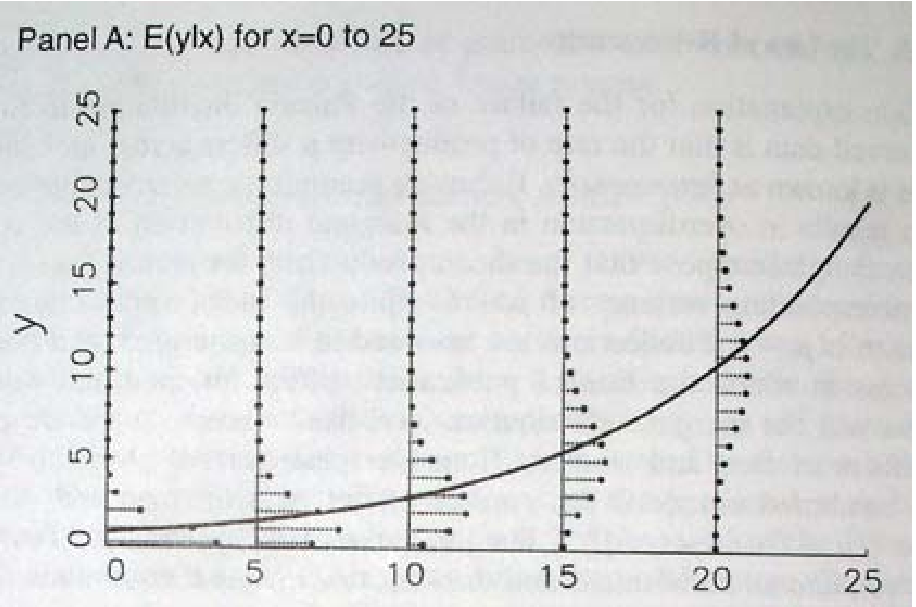
```
. tab childs
    number of |
     children |      Freq.     Percent        Cum.
--------------+-----------------------------------
         none |        799       28.95       28.95
          one |        469       16.99       45.94
          two |        657       23.80       69.75
        three |        481       17.43       87.17
         four |        185        6.70       93.88
         five |         73        2.64       96.52
          six |         40        1.45       97.97
        seven |         22        0.80       98.77
eight or more |         34        1.23      100.00
--------------+-----------------------------------
        Total |      2,760      100.00

. poisson childs

Iteration 0:   log likelihood = -5096.6865
Iteration 1:   log likelihood = -5096.6865

Poisson regression                                Number of obs   =       2760
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood = -5096.6865                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
      childs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5936071   .0141464    41.96   0.000     .5658807    .6213334
------------------------------------------------------------------------------
. prcounts pois, plot max(8)

. gr twoway connected poisobeq poispreq poisval
```



Overdispersion results in Poisson distribution underpredicting the outcomes in the two ends of the distribution – it underpredicts zeros and outcomes of 6 and larger. Fitting this kind of unconditional Poisson distribution does not take heterogeneity into account – the average number of children varies according to some characteristics of respondents. Next, we have to allow for that – need to incorporate the observed heterogeneity. A multivariate Poisson regression model does just that. It models the average count, μ:

$\mu=E(y|x)=\exp(Xb)$

We exponentiate to force the values to be positive–counts cannot be below 0. We get a nonlinear model that looks like this:



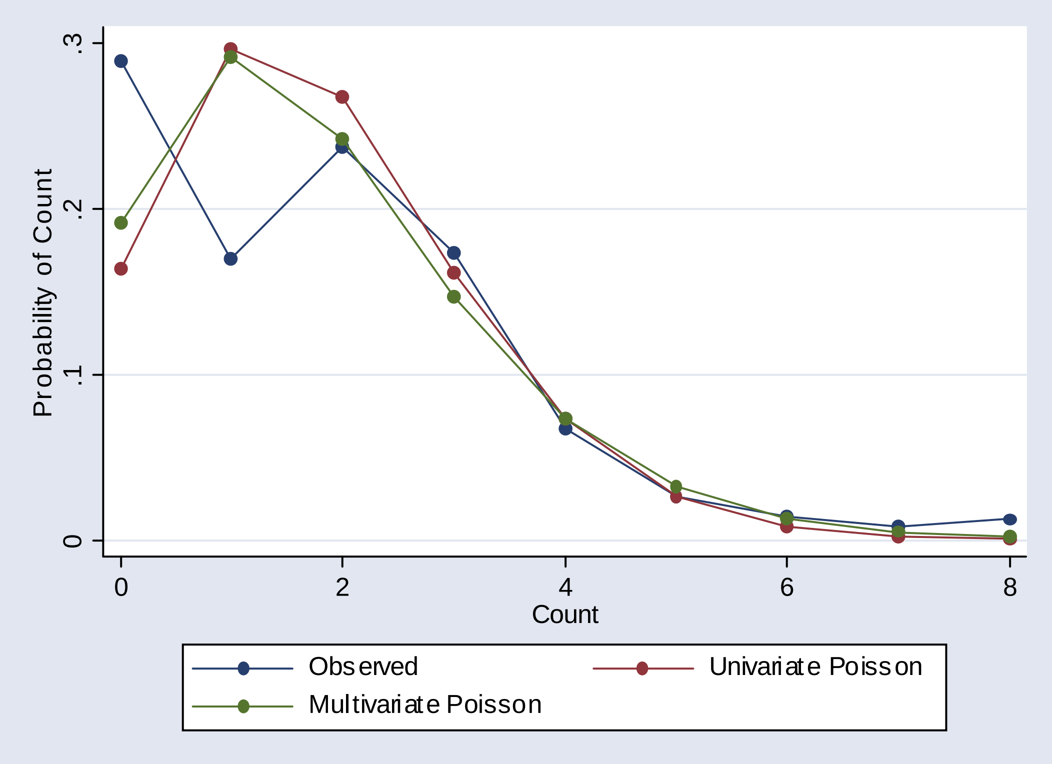
Let’s run a multivariate Poisson model:

```
. poisson childs sex married sibs  born educ

Iteration 0:   log likelihood = -4784.5123
Iteration 1:   log likelihood = -4784.5079
Iteration 2:   log likelihood = -4784.5079

Poisson regression                                Number of obs   =       2745
                                                  LR chi2(5)      =     572.66
                                                  Prob > chi2     =     0.0000
Log likelihood = -4784.5079                       Pseudo R2       =     0.0565

------------------------------------------------------------------------------
      childs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |    .195229   .0289993     6.73   0.000     .1383915    .2520665
     married |   .4486183   .0288777    15.54   0.000      .392019    .5052176
        sibs |   .0385556    .004219     9.14   0.000     .0302865    .0468246
        born |  -.2209195   .0522438    -4.23   0.000    -.3233154   -.1185235
        educ |   -.061697   .0048163   -12.81   0.000    -.0711369   -.0522572
       _cons |   .9547179   .1010692     9.45   0.000     .7566258     1.15281
------------------------------------------------------------------------------
```

Can interpret sign and significance – to interpret the size, we exponentiate the coefficients – generating so-called incidence-rate ratios (comparable to odds ratios). But we'll return to that later. First, let's consider model fit and diagnostics.

Once again, to assess how well our model predicts counts, we can graphically examine the predicted probabilities for different counts (these are probabilities for someone average on all characteristics) using prcounts command:

```
. prcounts prm, plot max(8)
(19 missing values generated)

. lab var poispreq "Univariate Poisson"

. lab var prmpreq "Multivariate Poisson"

. lab var prmobeq "Observed"

. gr twoway connected poisobeq poispreq prmpreq poisval, ylabel(0 (.1) .3)
ytitle("Probability of Count")
```



Multivariate Poisson offers a slight improvement over univariate Poisson – it explains some heterogeneity. But it still doesn't fit very well – underpredicts zeros, overpredicts ones, etc.

Just to clarify this, we can also obtain the probabilities presented in this graph using prvalue:

```
. prvalue

poisson: Predictions for childs
```

```
Predicted rate: 1.71     95% CI [1.66   ,   1.76]

Predicted probabilities:

  Pr(y=0|x):   0.1804  Pr(y=1|x):   0.3089
  Pr(y=2|x):   0.2646  Pr(y=3|x):   0.1510
  Pr(y=4|x):   0.0647  Pr(y=5|x):   0.0221
  Pr(y=6|x):   0.0063  Pr(y=7|x):   0.0015
  Pr(y=8|x):   0.0003  Pr(y=9|x):   0.0001

          sex    married       sibs       born       educ
x=  1.5555556  .45974499  3.6018215  1.0921676  13.358834
```

So we examined model fit graphically. We can also obtain a goodness-of-fit test (there are two versions of it, one based on deviance residuals, one is based on Pearson residuals; they usually produce similar results):

```
. estat gof
         Goodness-of-fit chi2  =  4279.437
         Prob > chi2(2739)     =    0.0000

. estat gof, pearson
         Goodness-of-fit chi2  =   3943.17
         Prob > chi2(2739)     =    0.0000
```

Since the probability is below .05, this suggests that predicted counts are significantly different from the observed ones, and therefore Poisson model doesn’t fit well.

Next, we turn to interpretation of Poisson models. First, as mentioned above, we can calculate incidence rate ratios:

```
. poisson childs sex married sibs  born educ, irr
Poisson regression                                Number of obs   =       2745
                                                  LR chi2(5)      =     572.66
                                                  Prob > chi2     =     0.0000
Log likelihood = -4784.5079                       Pseudo R2       =     0.0565
------------------------------------------------------------------------------
      childs |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   1.215589   .0352512     6.73   0.000     1.148425    1.286682
     married |   1.566147   .0452267    15.54   0.000     1.479966    1.657346
        sibs |   1.039308   .0043848     9.14   0.000      1.03075    1.047938
        born |   .8017812   .0418881    -4.23   0.000     .7237455    .8882309
        educ |   .9401677   .0045282   -12.81   0.000     .9313344    .9490847
------------------------------------------------------------------------------
```

So the number of children for women is 1.22 times (or 22%) higher than the number for men, the number of children for married is 1.57 times (or 57%) higher than for those not currently married, each additional sibling increases the number of children by almost 4%, being foreign born decreases the number of children by almost 10%, and each year of education reduces the number of children by 6%.

We can also obtain incidence rate ratios using listcoef – this will also allow us to see standardized ratios describing the change per one standard deviation of each variable.

```
. listcoef
poisson (N=2745): Factor Change in Expected Count
 Observed SD: 1.6887584
----------------------------------------------------------------------
      childs |      b         z     P>|z|    e^b    e^bStdX      SDofX
-------------+--------------------------------------------------------
         sex |   0.19523    6.732   0.000   1.2156   1.1019     0.4970
     married |   0.44862   15.535   0.000   1.5661   1.2506     0.4985
        sibs |   0.03856    9.139   0.000   1.0393   1.1227     3.0008
        born |  -0.22092   -4.229   0.000   0.8018   0.9381     0.2893
        educ |  -0.06170  -12.810   0.000   0.9402   0.8324     2.9741
----------------------------------------------------------------------
```

And we can get these as percents:

```
. listcoef, percent
poisson (N=2745): Percentage Change in Expected Count
 Observed SD: 1.6887584
----------------------------------------------------------------------
      childs |      b         z     P>|z|      %      %StdX      SDofX
-------------+--------------------------------------------------------
         sex |   0.19523    6.732   0.000     21.6     10.2     0.4970
     married |   0.44862   15.535   0.000     56.6     25.1     0.4985
        sibs |   0.03856    9.139   0.000      3.9     12.3     3.0008
        born |  -0.22092   -4.229   0.000    -19.8     -6.2     0.2893
        educ |  -0.06170  -12.810   0.000     -6.0    -16.8     2.9741
----------------------------------------------------------------------
```

Marriage and education seem to have the largest effects.

Listcoef with reverse option doesn't work after Poisson because we are now dealing with incidence rate ratios rather than odds ratios, so it doesn't make sense to report them. To compare the effect sizes between positive and negative effects, you can still calculate them, e.g., for education:

```
.di exp(.06170*2.9741)
1.2014173
```

So the effect of marriage is still stronger than that of education.

Next, to continue interpretation of this model, we can examine predicted rates for various groups. For example,

```
. prtab sex married, x(born=1)
poisson: Predicted rates for childs
-------------------------
responden |    married
ts sex    |      0       1
----------+--------------
     male | 1.2760  1.9983
   female | 1.5510  2.4292
-------------------------
          sex    married       sibs       born       educ
x=  1.5555556  .45974499  3.6018215          1  13.358834
```

We can see that for an average native-born woman, the average number of children she has if she is single is 1.55 and if she is married 2.43. An average native born man has 1.27 children on average if he is single and approximately 2 children if he is married.

We can also examine how such predicted rates change per change of each independent variable.

```
. prchange

poisson: Changes in Predicted Rate for childs
matsize too small
    You have attempted to create a matrix with more than 40 rows or columns or
to estimate a model with more than 40
    variables plus ancillary parameters.  You need to increase matsize using
the set matsize command; see help matsize.
r(908);
```

We are attempting a complex calculation, need to allow Stata to operate larger matrices.

```
. set matsize 800
```

Now we can do it:

```
. prchange
poisson: Changes in Predicted Rate for childs
         min->max      0->1     -+1/2    -+sd/2  MargEfct
    sex    0.3313    0.2725    0.3349    0.1662    0.3344
married    0.7889    0.7889    0.7748    0.3838    0.7683
   sibs    2.5712    0.0586    0.0660    0.1983    0.0660
   born   -0.3465   -0.4321   -0.3791   -0.1095   -0.3784
   educ   -2.7681   -0.2336   -0.1057   -0.3147   -0.1057

exp(xb):   1.7127
            sex  married     sibs     born     educ
    x=  1.55556  .459745  3.60182  1.09217  13.3588
sd(x)=  .496995  .498468  3.00084  .289315  2.97411
```

To make this more interpretable, let’s on rely on means for dummies:

```
. prchange, x(married=1 sex=2 born=1)

poisson: Changes in Predicted Rate for childs
         min->max      0->1     -+1/2    -+sd/2  MargEfct
    sex    0.4308    0.3544    0.4750    0.2358    0.4742
married    0.8781    0.8781    1.0989    0.5443    1.0898
   sibs    3.6468    0.0831    0.0937    0.2812    0.0937
   born   -0.4815   -0.6005   -0.5377   -0.1553   -0.5366
   educ   -3.9261   -0.3314   -0.1499   -0.4464   -0.1499

exp(xb):   2.4292
            sex  married     sibs     born     educ
    x=        2        1  3.60182        1  13.3588
sd(x)=  .496995  .498468  3.00084  .289315  2.97411
```

So for an average native born married women, each additional sibling increases her number of children by .09, and each additional year of education decreases it by .15. For an average native-born woman, marriage increases her number of kids by .88. For an average married woman, being foreign-born decreases the number of children by 0.48. And the difference between the number of children of an average native-born married woman and an average native-born married man is .43 of a child.

In addition to predicted rates, we can also obtain predicted probabilities for each count for specific combinations of independent variables. For a single native born man:

```
. prvalue, x(married=0 sex=1 born=1) save
poisson: Predictions for childs
Predicted rate: 1.22     95% CI [1.15   ,   1.29]
Predicted probabilities:
  Pr(y=0|x):   0.2953  Pr(y=1|x):   0.3602
  Pr(y=2|x):   0.2197  Pr(y=3|x):   0.0893
  Pr(y=4|x):   0.0272  Pr(y=5|x):   0.0066
  Pr(y=6|x):   0.0014  Pr(y=7|x):   0.0002
  Pr(y=8|x):   0.0000  Pr(y=9|x):   0.0000
          sex    married       sibs       born       educ        age
x=          1          0  3.6082663          1  13.354792  46.301756
```

For a single native-born woman, comparing the two groups side by side:

```
. prvalue, x(married=0 sex=2 born=1) dif
poisson: Change in Predictions for  childs
Predicted rate: 1.46     95% CI [1.39   ,   1.53]
         Saved: 1.22
    Difference: .242
Predicted probabilities:
                     Current       Saved  Difference
  Pr(y=0|x):          0.2317      0.2953     -0.0636
  Pr(y=1|x):          0.3388      0.3602     -0.0214
  Pr(y=2|x):          0.2477      0.2197      0.0280
  Pr(y=3|x):          0.1207      0.0893      0.0314
  Pr(y=4|x):          0.0441      0.0272      0.0169
  Pr(y=5|x):          0.0129      0.0066      0.0063
  Pr(y=6|x):          0.0031      0.0014      0.0018
  Pr(y=7|x):          0.0007      0.0002      0.0004
  Pr(y=8|x):          0.0001      0.0000      0.0001
  Pr(y=9|x):          0.0000      0.0000      0.0000
                sex    married       sibs       born       educ        age
Current=          2          0  3.6082663          1  13.354792  46.301756
  Saved=          1          0  3.6082663          1  13.354792  46.301756
   Diff=          1          0          0          0          0          0
```

Women have lower probabilities of having no kids or 1 child and higher probabilities of having 2, 3 and 4 children.

We can also graph predicted probabilities for each count.

```
. prgen educ, x(sex=1 married=1 born=1) from(1) to(20) gen(edprm) n(20)
poisson: Predicted values as educ varies from 1 to 20.
          sex    married       sibs       born       educ        age
x=          1          1  3.6082663          1  13.354792  46.301756

. prgen educ, x(sex=2 married=1 born=1) from(1) to(20) gen(edprf) n(20)
poisson: Predicted values as educ varies from 1 to 20.
          sex    married       sibs       born       educ        age
x=          2          1  3.6082663          1  13.354792  46.301756
```

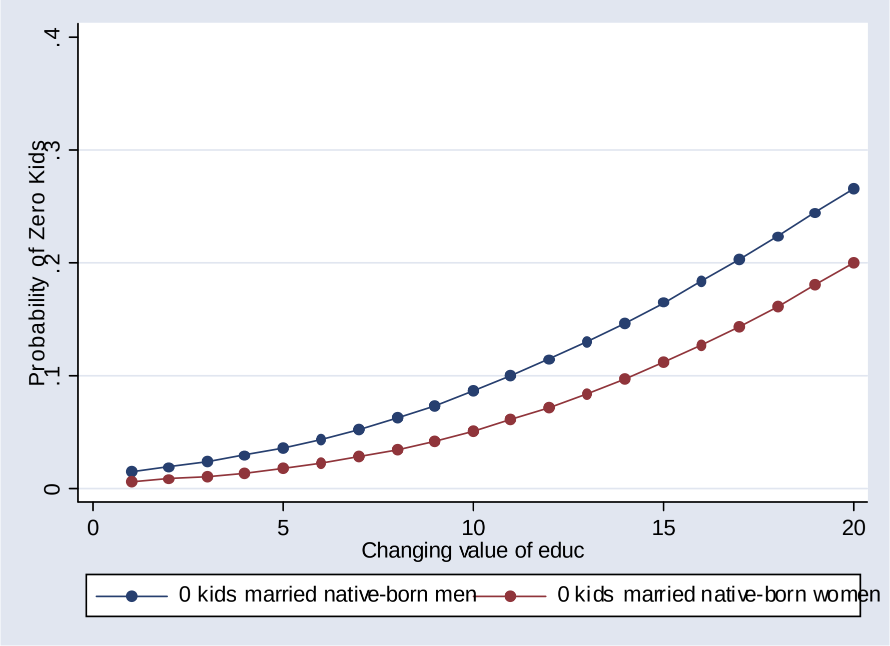
Note – it generated separate variables for each count. Let’s label the first four.

```
. lab var edprmp0 "0 kids married native-born men"
. lab var edprmp1 "1 kid married native-born men"
. lab var edprmp2 "2 kids married native-born men"
. lab var edprmp3 "3 kids married native-born men"
```

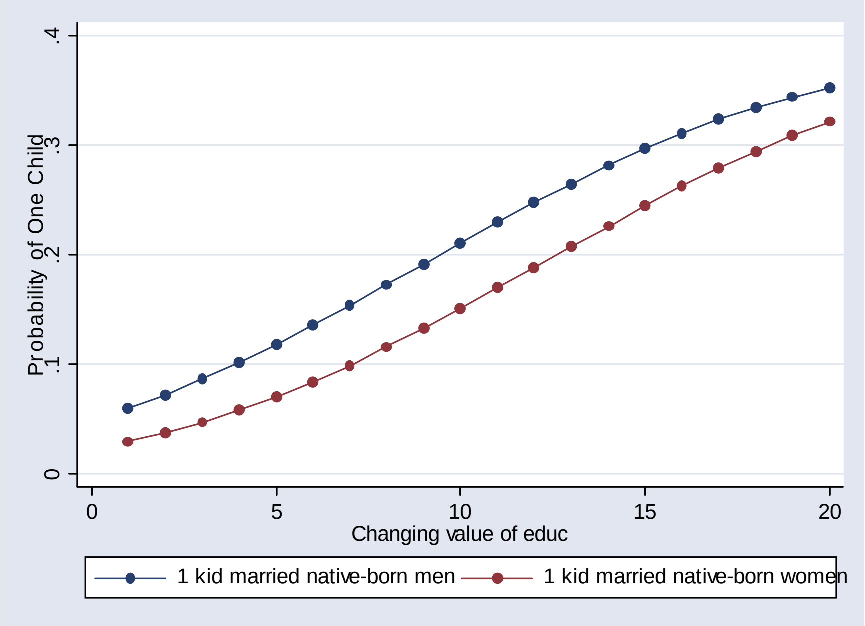
```
. lab var edprfp0 "0 kids married native-born women"
. lab var edprfp1 "1 kid married native-born women"
. lab var edprfp2 "2 kids married native-born women"
. lab var edprfp3 "3 kids married native-born women"
```

Now we are ready to graph them:

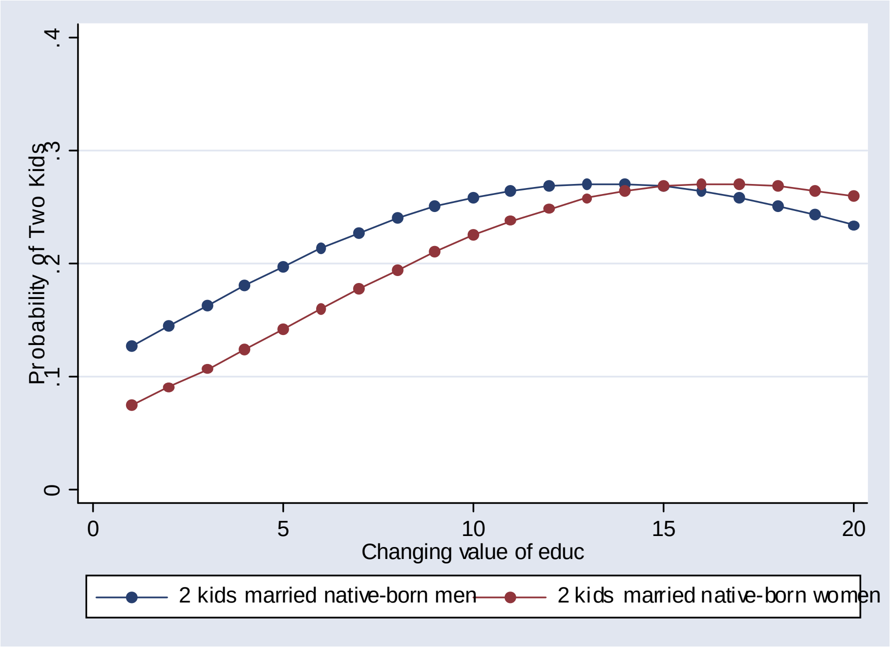
```
. gr twoway connected edprmp0 edprfp0 edprfx, ylabel(0(.1).4)
ytitle("Probability of Zero Kids")
```



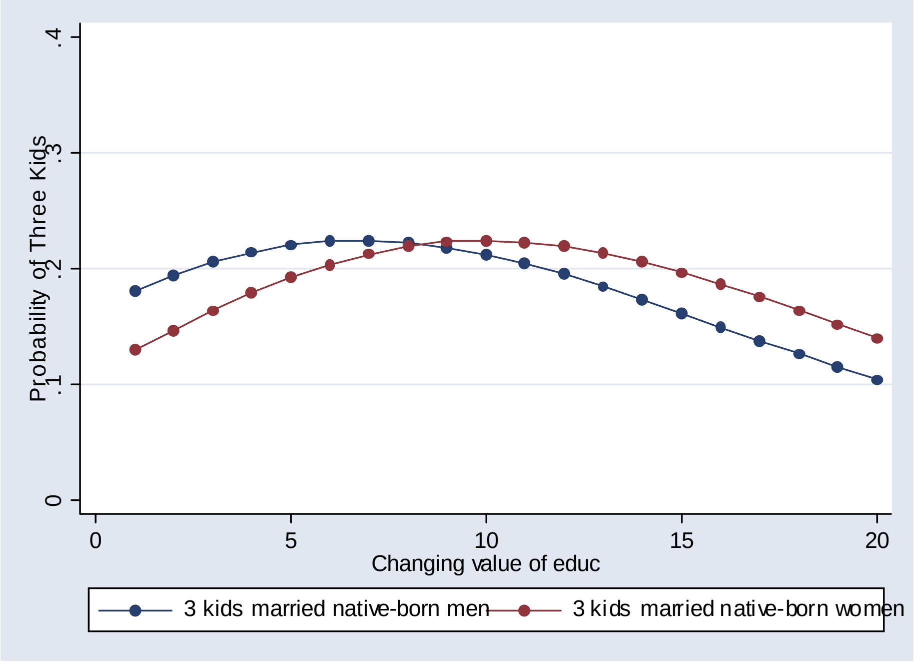
```
. gr twoway connected edprmp1 edprfp1 edprfx, ylabel(0(.1).4)
ytitle("Probability of One Child")
```



```
. gr twoway connected edprmp2 edprfp2 edprfx, ylabel(0(.1).4)
ytitle("Probability of Two Kids")
```

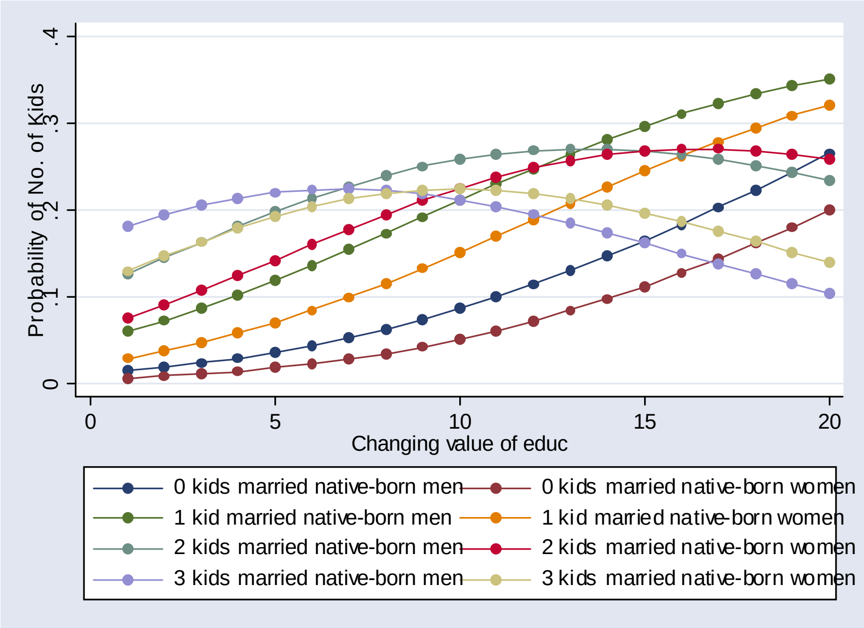


```
. gr twoway connected edprmp3 edprfp3 edprfx, ylabel(0(.1).4)
ytitle("Probability of Three Kids")
```



Can put these all together in one graph:

```
. gr twoway connected edprmp0 edprfp0 edprmp1 edprfp1 edprmp2 edprfp2 edprmp3
edprfp3 edprfx, ylabel(0(.1).4) ytitle("Probability of No. of Kids")
```



Diagnostics:

In terms of diagnostics, we can test for multicollinearity the same way we did with logistic models. To test for linearity and additivity, we can use Box-Tidwell test and mrunning and lowess using a log of the original count variable (add 1 to the count before logging it; otherwise zeros will become missing):

```
. gen countlg=log(childs+1)
```

We can also look at robust standard errors to compare them to the regular ones. We can also get residuals and leverage statistics to assess the outliers; however, to do that, we need to estimate the same model using generalized linear models command – GLM. Unfortunately, predict after Poisson is very limited, but after GLM version of Poisson we can get a range of statistics.

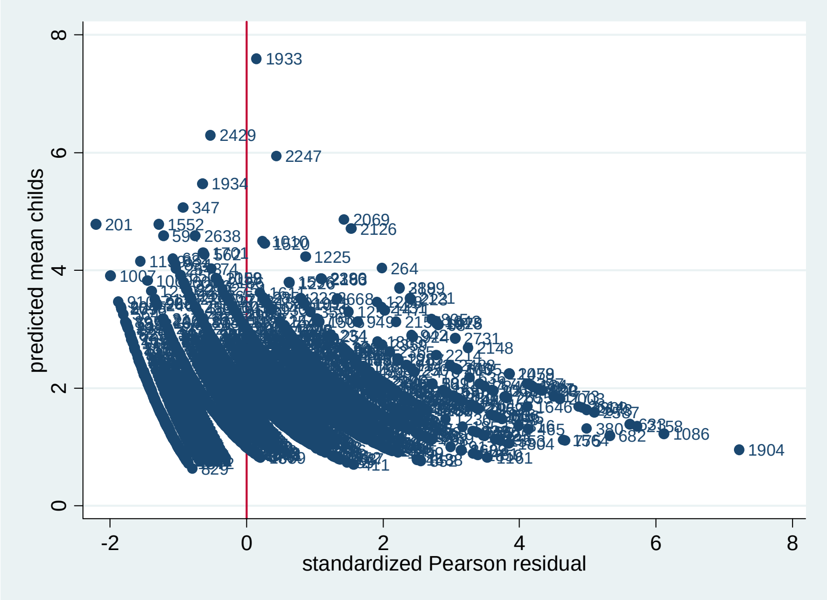
```
. glm childs sex married sibs  born educ, family(poisson)
Generalized linear models                          No. of obs      =      2745
Optimization     : ML                              Residual df     =      2739
                                                   Scale parameter =         1
Deviance         =  4279.437048                    (1/df) Deviance =  1.562409
Pearson          =  3943.169972                    (1/df) Pearson  =  1.439639
Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
                                                   AIC             =  3.490352
Log likelihood   =  -4784.50787                    BIC             =  -17406.7
------------------------------------------------------------------------------
             |                 OIM
      childs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |    .195229   .0289993     6.73   0.000     .1383915    .2520665
     married |   .4486183   .0288777    15.54   0.000      .392019    .5052176
        sibs |   .0385556    .004219     9.14   0.000     .0302865    .0468246
        born |  -.2209195   .0522438    -4.23   0.000    -.3233154   -.1185235
        educ |   -.061697   .0048163   -12.81   0.000    -.0711369   -.0522572
       _cons |   .9547179   .1010692     9.45   0.000     .7566258     1.15281
------------------------------------------------------------------------------
```

Here’s what we can obtain by using predict after this (among other statistics):

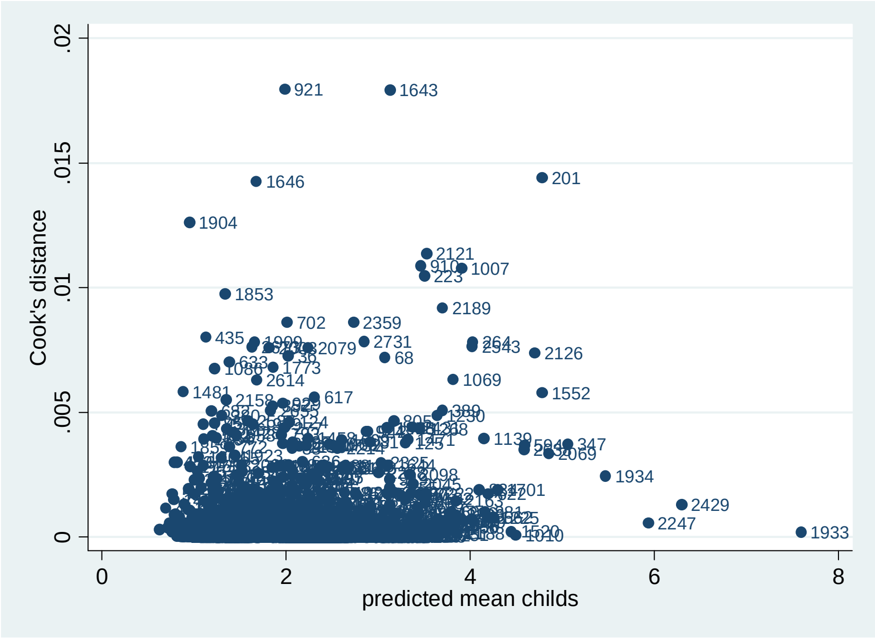
```
    cooksd calculates Cook's distance, which measures the aggregate change in
        the estimated coefficients when each observation is left out of the
        estimation.

    deviance calculates the deviance residuals.  Deviance residuals are
        recommended by McCullagh and Nelder and by others as having the best
        properties for examining the goodness of fit of a GLM.  They are
        approximately normally distributed if the model is correct.  They may
        be plotted against fitted values or against a covariate to inspect the
        model's fit.  Also see the pearson option below.

    hat calculates the diagonals of the "hat" matrix as an analog to simple
        linear regression.

    pearson calculates the Pearson residuals.  Be aware that Pearson residuals
        often have markedly skewed distributions for non-normal family
        distributions.  Also see the deviance option above.

    ----+ Options +----------------------------------------------------------

  standardized requests that the residual be multiplied by the factor
        (1-h)^[-1/2], where h is the diagonal of the hat matrix.  This is done
        to account for the correlation between depvar and its predicted value.

    studentized requests that the residual be multiplied by one over the
        square root of the estimated scale parameter.
```

We can use these the same way we have used them after logit, e.g.:

```
. predict p
(option mu assumed; predicted mean childs)
(19 missing values generated)
. predict rs, pearson standard
(20 missing values generated)
. predict cooksd, cooksd
(20 missing values generated)
. scatter p rs, xline(0) mlabel(id)
```



```
. scatter cooksd p, mlabel(id)
```



Would have a look at 1904, 921, 1643, 1646, 201.

Models Adjusted for Exposure:

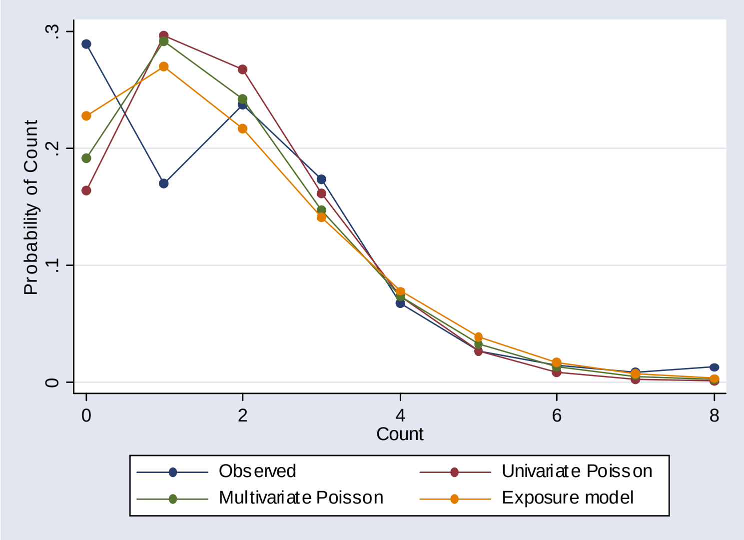
Models for count data also allow controlling for so-called exposure – that is usually a variable that indicates how long there has been an opportunity to accumulate counts. E.g. an 20 y.o. and a 40 y.o. had different time available to have kids, and that will likely be reflected in their number of children. So we can control for the duration of reproductive age – that's the amount of exposure one had. Let's assume reproductive age to start at 15 and end at 45 (these numbers of course will vary individually, and it would be best to get a variable with individual data on that, but this is our best approximation):

```
. gen reprage=age-15
(14 missing values generated)
. replace reprage=30 if age>45 & age~=.
(1312 real changes made)
. poisson childs sex married sibs  born educ, exposure(reprage)
Poisson regression                                Number of obs   =       2734
                                                  LR chi2(5)      =     365.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -4474.7807                       Pseudo R2       =     0.0392
------------------------------------------------------------------------------
      childs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .1829962   .0291302     6.28   0.000     .1259021    .2400902
     married |   .3223659   .0290622    11.09   0.000      .265405    .3793267
        sibs |   .0249154   .0042745     5.83   0.000     .0165375    .0332933
        born |  -.1354091   .0522745    -2.59   0.010    -.2378651    -.032953
        educ |  -.0575382    .004645   -12.39   0.000    -.0666423   -.0484341
       _cons |  -2.218856   .1006406   -22.05   0.000    -2.416108   -2.021604
     reprage |  (exposure)
------------------------------------------------------------------------------
```

What this actually does is: ln(reprage) is entered in the model, but its coefficient is constrained to 1. If we don't control for exposure, it's assumed that all cases have had the same exposure.

You can get the same result by using a log of exposure variable and specifying it using offset option:

```
. gen repragelog=log(reprage)
. poisson childs sex married sibs  born educ, offset(repragelog)
Poisson regression                                Number of obs   =       2734
                                                  LR chi2(5)      =     365.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -4474.7807                       Pseudo R2       =     0.0392
------------------------------------------------------------------------------
      childs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .1829962   .0291302     6.28   0.000     .1259021    .2400902
     married |   .3223659   .0290622    11.09   0.000      .265405    .3793267
        sibs |   .0249154   .0042745     5.83   0.000     .0165375    .0332933
        born |  -.1354091   .0522745    -2.59   0.010    -.2378651   -.032953
        educ |  -.0575382    .004645   -12.39   0.000    -.0666423   -.0484341
       _cons |  -2.218856   .1006406   -22.05   0.000    -2.416108   -2.021604
  repragelog |   (offset)
------------------------------------------------------------------------------.
. prcounts expo, plot max(8)
(31 missing values generated)
. lab var expopreq "Exposure model"
. gr twoway connected poisobeq poispreq prmpreq expopreq poisval, ylabel(0 (.1)
.3) ytitle("Probability of Count")
```



This model fits somewhat better but still has the same problems.
Further, when we think that our measure of exposure is not a perfect measure of how much time one had to accumulate counts, we may just enter log of exposure variable it into the model without constraining the coefficient to 1:

```
. poisson childs sex married sibs  born educ repragelog
Poisson regression                                Number of obs   =       2734
                                                  LR chi2(6)      =    1151.72
                                                  Prob > chi2     =     0.0000
Log likelihood = -4473.9245                       Pseudo R2       =     0.1140
------------------------------------------------------------------------------
      childs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .1835258   .0291282     6.30   0.000     .1264356     .240616
     married |   .3266819   .0292507    11.17   0.000     .2693516    .3840121
        sibs |   .0254587   .0042934     5.93   0.000     .0170438    .0338737
```

```
        born |  -.1396764   .0523749    -2.67   0.008    -.2423293   -.0370235
        educ |  -.0577855   .0046561   -12.41   0.000    -.0669113   -.0486597
  repragelog |   .9417878   .0441539    21.33   0.000     .8552478    1.028328
       _cons |  -2.028168   .1760539   -11.52   0.000    -2.373228   -1.683109
------------------------------------------------------------------------------
```

Here it has a coefficient not significantly different from 1 (the confidence interval includes 1), so reprage seems to be a good estimate of exposure time. If it would be significantly different from 1, and we would have substantive reasons to believe that our measure of exposure is imperfect, we might use this model instead of the one with exposure option or offset option.

When interpreting models with exposure, we cannot, unfortunately, directly use prtab, prgen, and prvalue as those show incidence rates (rates per unit of time) rather than total counts. We can, potentially, use them and multiply the values by average exposure values, but there is an easier way to get around that problem. As mentioned above, the way models with exposure are estimated is: the log of exposure variable is entered into the model, but its coefficient is constrained to 1. We could simulate that using constraints option – first, we specify that constraint #1 will mean repragelog coefficient should be 1, and then estimate the model adding repragelog and using constraint 1:

```
. constraint 1 repragelog=1
. poisson childs sex married sibs  born educ repragelog, constraints(1)
Poisson regression                                Number of obs   =       2734
                                                  Wald chi2(5)    =     371.72
Log likelihood = -4474.7807                       Prob > chi2     =     0.0000
 ( 1)  [childs]repragelog = 1
------------------------------------------------------------------------------
      childs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .1829962   .0291302     6.28   0.000     .1259021    .2400902
     married |   .3223659   .0290622    11.09   0.000      .265405    .3793267
        sibs |   .0249154   .0042745     5.83   0.000     .0165375    .0332933
        born |  -.1354091   .0522745    -2.59   0.010    -.2378651    -.032953
        educ |  -.0575382    .004645   -12.39   0.000    -.0666423   -.0484341
  repragelog |          1          .        .       .            .           .
       _cons |  -2.218856   .1006406   -22.05   0.000    -2.416108   -2.021604
------------------------------------------------------------------------------
```

Now prtab, prvalue, and prgen will work. E.g.:

```
. prtab married sex born
poisson: Predicted rates for childs
--------------------------------------------
         |  was r born in this country and
         |         respondents sex
         | ----- yes ----    ----- no -----
  married |   male  female      male  female
----------+---------------------------------
        0 | 1.2333  1.4810    1.0771  1.2934
        1 | 1.7025  2.0443    1.4869  1.7854
--------------------------------------------
           sex     married        sibs        born        educ  repragelog
x=   1.5548647   .46013168   3.6082663   1.0921726   13.354792   3.0594865
```

In terms of diagnostics and model fit for models with exposure, everything works the same except Box-Tidwell test which does not work with exposure or offset option, but does work with constraints:

```
. constraint 1 repragelog=1
. constraint 2 Irepr__1 =1
. boxtid poisson childs educ sex married sibs born repragelog, constraints(1 2)
Poisson regression                                Number of obs   =       2734
                                                  Wald chi2(8)    =     852.30
Log likelihood = -4472.2691                       Prob > chi2     =     0.0000
 ( 1)  [childs]Irepr__1 = 1
------------------------------------------------------------------------------
      childs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    Ieduc__1 |  -.5378193   .1694153    -3.17   0.002    -.8698671   -.2057714
    Ieduc_p1 |  -.0004982   .1461062    -0.00   0.997    -.2868611    .2858646
    Isibs__1 |   .3208799   .1250397     2.57   0.010     .0758066    .5659532
    Isibs_p1 |   .0009448     .12313     0.01   0.994    -.2403854    .2422751
    Irepr__1 |          1          .        .       .            .           .
    Irepr_p1 |   1.577093   .0834845    18.89   0.000     1.413467     1.74072
     Isex__1 |    .182057   .0292248     6.23   0.000     .1247774    .2393366
     married |   .3238347   .0292863    11.06   0.000     .2664346    .3812348
    Iborn__1 |  -.1413858   .0530498    -2.67   0.008    -.2453614   -.0374101
       _cons |   .2499269   .0315315     7.93   0.000     .1881263    .3117275
------------------------------------------------------------------------------
educ     |  -.0582693   .0046599    -12.504   Nonlin. dev. 0.069   (P = 0.793)
      p1 |   1.067851   .2711467      3.938
------------------------------------------------------------------------------
sibs     |   .0255532   .0042942      5.951   Nonlin. dev. 0.742   (P = 0.389)
      p1 |   .7165476   .3622967      1.978
------------------------------------------------------------------------------
repragelog|          1          0          .   Nonlin. dev. 4.167   (P = 0.041)
      p1 |   .2074807   .4305246      0.482
------------------------------------------------------------------------------
Deviance: 8944.406.
```

For those statistics that are obtained using predict after GLM, we need to use offset option with GLM (exposure option doesn’t work for that):

```
. glm childs sex married sibs  born educ, family(poisson) offset(repragelog)
Generalized linear models                          No. of obs      =       2734
Optimization     : ML                              Residual df     =       2728
                                                   Scale parameter =          1
Deviance         =  3675.111598                    (1/df) Deviance =   1.347182
Pearson          =  3353.513369                    (1/df) Pearson  =   1.229294
Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
                                                   AIC             =   3.277821
Log likelihood   = -4474.780694                    BIC             =  -17912.97
------------------------------------------------------------------------------
             |                 OIM
      childs |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .1829962   .0291302     6.28   0.000     .1259021    .2400902
     married |   .3223659   .0290622    11.09   0.000      .265405    .3793267
        sibs |   .0249154   .0042745     5.83   0.000     .0165375    .0332933
        born |  -.1354091   .0522745    -2.59   0.010    -.2378651    -.032953
        educ |  -.0575382    .004645   -12.39   0.000    -.0666423   -.0484341
       _cons |  -2.218856   .1006406   -22.05   0.000    -2.416108   -2.021604
  repragelog |   (offset)
------------------------------------------------------------------------------
```

After that, we can obtain residuals etc.