Count outcomes - Poisson regression (Chapter 6)

- Exponential family
- Poisson distribution
- Examples of count data as outcomes of interest
- Poisson regression
- Variable follow-up times - Varying number “at risk” - offset
- Overdispersion - pseudo likelihood
- Using Poisson regression with robust standard errors in place of binomial log models

The Exponential Family

- Assume Y has a distribution for which the density function has the following form:

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

  for some specific function $a(\cdot)$, $b(\cdot)$, and $c(\cdot, \cdot)$.

  θ: canonical (natural) parameter – parameter of interest

  ϕ: scale parameter – nuisance parameter

- The above density define an exponential family if ϕ is known; if ϕ unknown, it may or may not define a two-parameter exponential family, depending on the form of $c(y, \phi)$.

- Examples: Normal
  Binomial
  Poisson
  Negative Binomial
  Gamma

Properties of Exponential Family and Generalized Linear Models

- If ϕ is known in the previous density function, then:

$$\mu \equiv E[Y] = b'(\theta)$$

$$Var(Y) = b''(\theta)a(\phi)$$

- Generalized linear models (GLM):
  - We assume the observation are independent with non-constant variance.
  - We extend the linear model by:
    - Replacing the linear model for μ with a linear model for $g(\mu)$.
    - Replacing the constant variance assumption with mean-variance relationship.
    - Replacing the normal distribution with the exponential family.
- Linear predictor: $\eta = X^T\beta$ (systematic component)

  Link function to link η and μ : $\eta = g(\mu)$ (when $\eta = \theta$, the corresponding link function is called the canonical link function)

Poisson distribution

- The Poisson distribution, Y ~ Poisson(μ), $\Pr(Y = y) = \frac{e^{-\mu}\mu^{y}}{y!}$, $\mu > 0$, is the most widely-used distribution for counts.
- The Poisson distribution assigns a positive probability to every nonnegative integer 0, 1, 2, . . ., so that every nonnegative integer becomes a mathematical possibility (albeit practically zero possibility for most count values)
- The Poisson is different than the binomial, Bin(n, π), which takes on numbers only up to some n, and leads to a proportion (out of n).
- But the Poisson is similar to the binomial in that it can be show that the Poisson is the limiting distribution of a Binomial for large n and small π. Furthermore, because of the simple form of the Poisson distribution, it is often computationally preferred over the Binomial.

Examples of count data

- Number of visits to emergency room during last year. A study looks at the effectiveness of a new treatment compared to standard care on reducing emergency room visits controlling for demographics and alcohol and drug use of individuals. (from VGSM)
- Number of damage reports on ships out to sea in the 1960-80. Look for systematic variables influencing the likelihood of damage occurring to the ship. (from McCullagh and Nelder 1989)
- Length of stay (in days) of hospital admissions. Look for systematic variables (i.e. insurance type, type of admission, demographics) related to the average length of stay (from Hardin and Hilbe 2007)
- Number of homicides within each census tract throughout the Twin Cities area. Look at whether there are relationships between homicide rate and density of alcohol outlets (Jones-Webb R and Wall MM. Neighborhood Racial/Ethnic Concentration, Social Disadvantage, and Homicide Risk: An Ecological Analysis of 10 Cities. *Journal of Urban Health*, 2008.

Examples of count data

- Number of injuries that resulted in lost work time during the construction of the Denver Airport. Look at characteristic of construction contracts and see if there are things that are related to higher injury rates (form Lowery et al *Am Journal of Industrial Medicine* 1998)
- Deaths from coronary heart disease after 10 years in a population of British male doctors. Look at how smoking is related to the risk of death. Have person time at risk (from Breslow and Day 1987).
- Count of number of abstainers of alcohol and how this is related to treatment. We can use Poisson regression (with robust standard errors) to estimate common risks in places where we might have computational problems using binomial regression.

The Poisson Regression model

Let Y_i be the observed count for experimental unit *i*

$$Y_i|X_i \sim \text{Poi}(\mu_i)$$

$$\log(\mu_i) = X_i\beta$$

The log link is the most commonly used, indicating we think that the covariates influence the mean of the counts (μ) in a multiplicative way, i.e. as a covariate increases by 1 unit, the log of the mean increases by β units and this implies the mean increases by a "fold-change" of or "scale factor" of exp(β).

* The log link is the canonical link in GLM for Poisson distribution.

Poisson regression for modeling rates

Often we are modeling the count of events within a particular time period, or within a particular region, or within a particular risk group of people. In each of these cases what is of interest is to model the RATE.

So given, for example, a specific time period t, we want to model the events occurring in the time period t. Thus, the Poisson mean μ is better described as $\mu = \lambda * t$ where λ is the RATE of events.

$$
\begin{aligned}
Y_i | X_i &\sim Poi(\lambda_i * t_i) \\
log(\lambda_i) &= \mathbf{X}_i \beta \\
log(\mu_i / t_i) &= log(\mu_i) - log(t_i) = \mathbf{X}_i \beta \\
log(\mu_i) &= log(t_i) + \mathbf{X}_i \beta
\end{aligned}
$$

The term $\log(t_i)$ is known as the offset and it provides the adjustment for the variable risk sets (e.g. varying time periods followed for each person, or variable numbers of people at risk). It can be thought of as a predictor but it does not have a parameter in front of it to be estimated, so it must be treated different from other predictors in the software.

Poisson regression produces relative rates

Let Y_i be the count of events within a risk set t_i, and X_i predictors of interest. Consider,

$$
\begin{aligned}
Y_i | X_i &\sim Poi(\lambda_i * t_i) \\
log(\lambda_i) &= \mathbf{X}_i \beta \\
log(\mu_i) &= log(t_i) + \mathbf{X}_i \beta
\end{aligned}
$$

Now, a change of one unit in a predictor variable relates to β unit change in the log RATE (i.e. $\log(\lambda_i)$), so if we exponentiate this we have a Relative rate (or Rate ratio).

Over and Under dispersion

- Recall that the if Y ~ Poi(μ) this means that E(Y) = μ AND Var(Y) = μ. It is quite common for the equality of the mean and variance to be incorrect for count data. In other words, Poisson distributional assumption is often not strictly correct.
- A common cause of overdispersion is there are other variable causing variability in the outcome which are not being included in the model, unexplained random variation. Underdispersion does not have an obvious explanation.
- A common solution is to assume that the variance is proportional to the mean, i.e. Var(Y) = φμ, and estimate the proportionality factor φ, which is called the scale parameter, from the data.
- Use the Goodness of fit tests (Pearson or the Residual Deviance) to estimate the φ. $\hat{\varphi} = X^2/(n - p)$ or $\hat{\varphi}$ = Residual Deviance/(n − p).

Adjusting for over/under-dispersion

- If $\hat{\varphi}$ >> 1 we say the data exhibits overdispersion and if it is << 1 it is called underdispersion. Note when φ = 1 this means E(Y) = Var(Y) which is supportive of the Poisson assumption.
- There is no clear cut decision rules when to decide that there is definitely over/under dispersion. Common rule of thumb is > 2.
- When controlling for over/under dispersion, basically, the parameter estimates do not change but the standard errors do. All standard errors are multiplied by sqrt($\hat{\varphi}$), hence they get wider in the case of overdispersion and smaller with underdispersion.
- In SAS simply add /scale = deviance OR /scale = pearson to the model statement.
- In Stata add scale(x2) or scale(dev) in the glm function.

Using Poisson regression for incidence rates

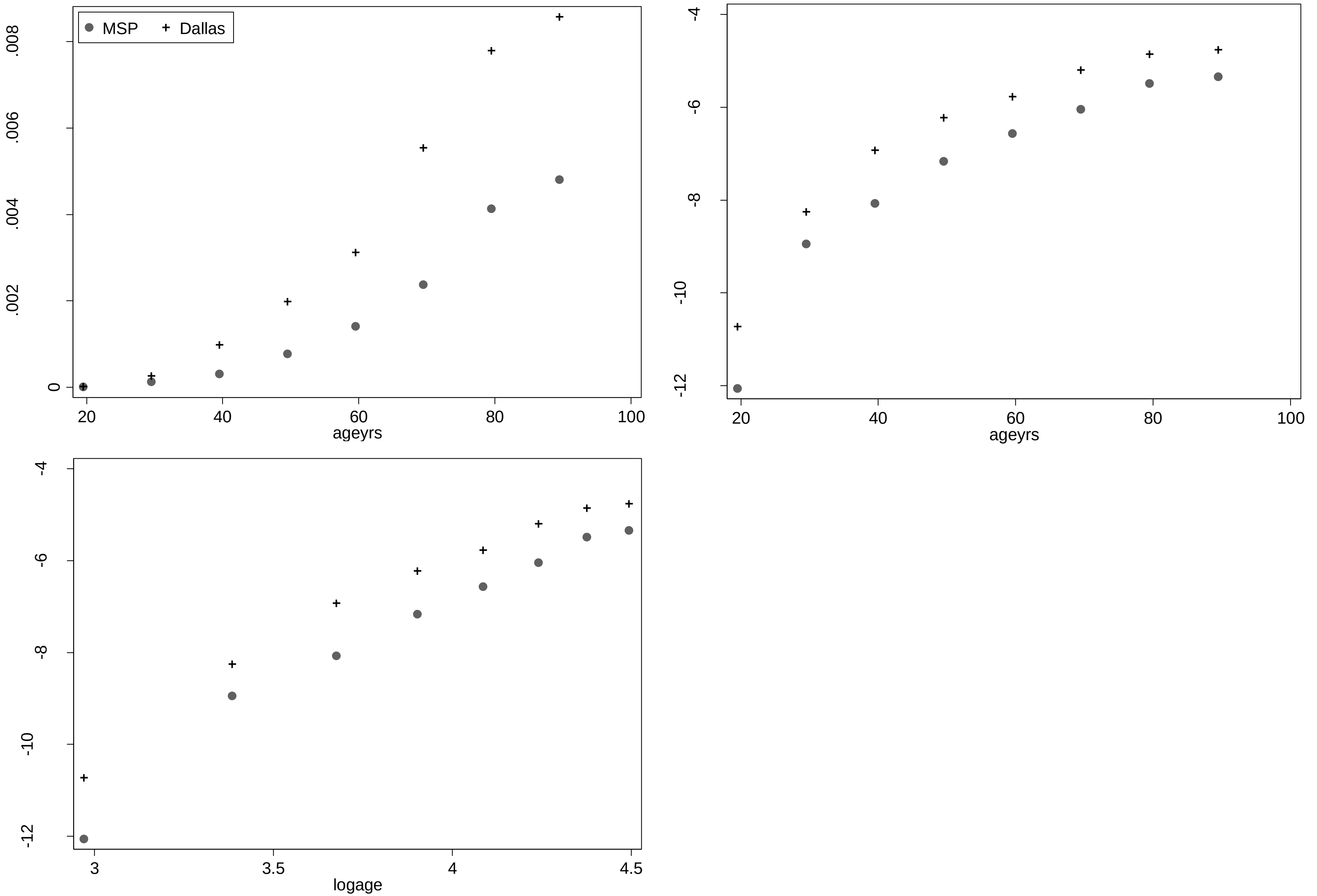
The data show the incidence of nonmelanoma skin cancer among women in Minneapolis-St Paul, Minnesota, and Dallas-Fort Worth, Texas in 1970. One would expect sun exposure to be greater in Texas than in Minnesota.

--Data from Kleinbaum, D., Kupper, L., and Muller, K. (1989). Applied regression analysis and other multivariate methods. PWS-Kent, Boston, Massachusetts. And adapted from Scotto, Kopf and Ruvbach (1974)

```
input case town str5 age ageyrs pop
1 0 15-24 19.5 172675
16 0 25-34 29.5 123065
30 0 35-44 39.5 96216
71 0 45-54 49.5 92051
102 0 55-64 59.5 72159
130 0 65-74 69.5 54722
133 0 75-84 79.5 32185
40 0 85+ 89.5 8328
4 1 15-24 19.5 181343
38 1 25-34 29.5 146207
119 1 35-44 39.5 121374
221 1 45-54 49.5 111353
259 1 55-64 59.5 83004
310 1 65-74 69.5 55932
226 1 75-84 79.5 29007
65 1 85+ 89.5 7583
end
```

* town: 1 = Dallas, 0 = MSP

Skin Cancer data



Using Poisson regression for incidence rates: SAS

```
data skin1; set skin;
logpop = log(pop); ***by default log is natural log in SAS;
logage = log(ageyrs);
run;
****Using Poisson regression*********************************************;
***categorical age;
proc genmod data = skin1;
class ageyrs;
model case = ageyrs town/ dist = poisson link = log offset = logpop;
estimate 'age adjusted RR of skincancer in Dallas vs MSP' town 1;
run;
***Continuous age;
proc genmod data = skin1;
model case = ageyrs town/ dist = poisson link = log offset = logpop;
estimate 'age adjusted RR of skincancer in Dallas vs MSP' town 1;
run;
***Log transformed continuous age;
proc genmod data = skin1;
model case = logage town/ dist = poisson link = log offset = logpop;
estimate 'age adjusted RR of skincancer in Dallas vs MSP' town 1;
run;
*************************************************************************;
Using Binomial regression with log link - practically same answer;
**************************************************************************;
proc genmod data = skin1;
class ageyrs;
model case/pop = ageyrs town / dist = binomial link = log;
estimate 'age adjusted RR of skincancer in Dallas vs MSP' town 1;
run;
```

Treating age as categorical

```
. glm case i.age_cat town, family(poisson) link(log) offset(logpop)
Generalized linear models                          No. of obs      =         16
Optimization     : ML                              Residual df     =          7
                                                   Scale parameter =          1
Deviance         =  8.258494053                    (1/df) Deviance =   1.179785
Pearson          =  8.127296469                    (1/df) Pearson  =   1.161042
Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
                                                   AIC             =   7.531475
Log likelihood   = -51.25179981                    BIC             =  -11.14963
------------------------------------------------------------------------------
             |                 OIM
        case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     age_cat |
          2  |   2.630202   .4674615     5.63   0.000     1.713994     3.54641
          3  |   3.847367   .4546594     8.46   0.000     2.956251    4.738483
          4  |   4.595197   .4510287    10.19   0.000     3.711197    5.479197
          5  |   5.087289   .4503011    11.30   0.000     4.204715    5.969863
          6  |   5.645412   .4497475    12.55   0.000     4.763924    6.526901
          7  |   6.058534   .4503204    13.45   0.000     5.175923    6.941146
          8  |    6.17419   .4577405    13.49   0.000     5.277035    7.071345
             |
        town |   .8038983   .0522048    15.40   0.000     .7015788    .9062177
       _cons |  -11.65761   .4487091   -25.98   0.000    -12.53706   -10.77816
      logpop |          1  (offset)
------------------------------------------------------------------------------
```

Treating age as categorical: Risk (Rate) Ratio

```
. glm, eform
Generalized linear models                          No. of obs      =        16
Optimization     : ML                              Residual df     =         7
                                                   Scale parameter =         1
Deviance         =  8.258494053                    (1/df) Deviance =  1.179785
Pearson          =  8.127296469                    (1/df) Pearson  =  1.161042
Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
                                                   AIC             =  7.531475
Log likelihood   = -51.25179981                    BIC             = -11.14963
------------------------------------------------------------------------------
             |                 OIM
        case |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     age_cat |
          2  |   13.87657   6.486763     5.63   0.000      5.55109    34.68855
          3  |   46.86951   21.30966     8.46   0.000     19.22577    114.2608
          4  |   99.00766    44.6553    10.19   0.000     40.90274    239.6543
          5  |   161.9502   72.92636    11.30   0.000     67.00151     391.452
          6  |   282.9902   127.2742    12.55   0.000     117.2049    683.2777
          7  |   427.7481   192.6237    13.45   0.000     176.9598    1033.955
          8  |   480.1938   219.8042    13.49   0.000     195.7885    1177.731
             |
        town |   2.234234   .1166376    15.40   0.000     2.016935    2.474944
       _cons |   8.65e-06   3.88e-06   -25.98   0.000     3.59e-06    .0000208
      logpop |          1  (offset)
------------------------------------------------------------------------------
```

Treating age as continuous

```
. glm case ageyrs town, family(poisson) link(log) offset(logpop)
Generalized linear models                          No. of obs      =        16
Optimization     : ML                              Residual df     =        13
                                                   Scale parameter =         1
Deviance         =   190.561268                    (1/df) Deviance =  14.65856  <--***
Pearson          =  148.1964085                    (1/df) Pearson  =  11.39972
Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
                                                   AIC             =   18.1754  <--***
Log likelihood   = -142.4031868                    BIC             =  154.5176
------------------------------------------------------------------------------
             |                 OIM
        case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ageyrs |   .0600283   .0013123    45.74   0.000     .0574562    .0626003
        town |   .8192832   .0521801    15.70   0.000     .7170121    .9215543
       _cons |  -10.31721   .0955464  -107.98   0.000    -10.50448   -10.12994
      logpop |          1  (offset)
------------------------------------------------------------------------------

. lincom town, rrr
 ( 1)  [case]town = 0
------------------------------------------------------------------------------
        case |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   2.268873     .11839    15.70   0.000     2.048304    2.513194
------------------------------------------------------------------------------
```

Treating log(age) as continuous

```
. glm case logage town, family(poisson) link(log) offset(logpop)
Generalized linear models                          No. of obs      =         16
Optimization     : ML                              Residual df     =         13
                                                   Scale parameter =          1
Deviance         =  40.20069588                    (1/df) Deviance =   3.092361  <--***
Pearson          =  34.36515837                    (1/df) Pearson  =   2.643474
Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
                                                   AIC             =   8.777863  <--***
Log likelihood   = -67.22290072                    BIC             =   4.157042
------------------------------------------------------------------------------
             |                 OIM
        case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      logage |   3.313291   .0820583    40.38   0.000      3.15246    3.474122
        town |   .8095186   .0521754    15.52   0.000     .7072567    .9117806
       _cons |  -20.10271   .3423286   -58.72   0.000    -20.77367   -19.43176
      logpop |          1  (offset)
------------------------------------------------------------------------------

. lincom town, rrr
 ( 1)  [case]town = 0
------------------------------------------------------------------------------
        case |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   2.246826   .1172291    15.52   0.000     2.028419     2.48875
------------------------------------------------------------------------------
```

Binomial regression

```
. binreg case i.age_cat town, rr n(pop)
Generalized linear models                          No. of obs      =        16
Optimization     : MQL Fisher scoring              Residual df     =         7
                   (IRLS EIM)                      Scale parameter =         1
Deviance         =  8.281933795                    (1/df) Deviance =  1.183133
Pearson          =  8.150373408                    (1/df) Pearson  =  1.164339
Variance function: V(u) = u*(1-u/pop)              [Binomial]
Link function    : g(u) = ln(u/pop)                [Log]
                                                   BIC             = -11.12619
------------------------------------------------------------------------------
             |                 EIM
        case | Risk Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     age_cat |
          2  |   13.87658   6.485948     5.63   0.000     5.551734    34.68456
          3  |   46.87123   21.30752     8.46   0.000     19.22882     114.251
          4  |   99.01094   44.65057    10.19   0.000     40.90913    239.6327
          5  |     161.95   72.91578    11.30   0.000      67.0099    391.4018
          6  |   282.9974   127.2581    12.55   0.000     117.2235    683.2037
          7  |   427.6687    192.555    13.46   0.000     176.9537    1033.607
          8  |    480.063   219.6807    13.49   0.000     195.7859    1177.104
             |
        town |   2.233754   .1164364    15.42   0.000     2.016814     2.47403
       _cons |   8.65e-06   3.88e-06   -25.98   0.000     3.59e-06    .0000209
------------------------------------------------------------------------------
```

*. Similar results as Poisson regression

Comparing models

- Fitzmaurice,G.M (1997). Model selection with overdispersed data, *The Statistician*, 46(1):81-91. recommends using Adjusted Information criterion to choose model in cases with overdispersion.
- The general form of information criterion is = -2logL + penalty factor

AIC = -2logL + 2*p

SC = Schwarz's Criterion = BIC = - 2logL + 2*p*log(n)

AICC (Corrected AIC) = -2logL + 2*p*(n/n-p)

- Recall that if $\widehat{\mu_i}$ is small (less than 5) then the Deviance and Pearson are not good measures of the goodness-of-fit and using them as a measure of overdispersion is not recommended.
- This is the same rule of thumb that tells us not to use the chi-square test when expected cell sizes in a table are < 5.

Note: In Stata, the AIC and BIC in -glm- output use different formula from -estat ic- command. The latter is closer to the SAS output (if not the same).

Ship damage example

- From McCullagh and Nelder (1989), Sec 6.3.2.
- Each row represents the AGGREGATED number of months of service and number of damage incidents to all ships in the fleet; particular type built in the particular year (60-64, 65-69, 70-74, 75-79) and operating during the particular time period (60-74, 75-79).
- Notice there are some structural zeros (0) under service because it was not possible for a ship built between 75-79 to operate between 1960-1974.

index	shiptype	year	period	months	damage
1	A	60	60	127	0
2	A	60	75	63	0
3	A	65	60	1095	3
4	A	65	75	1095	4
5	A	70	60	1512	6
6	A	70	75	3353	18
7	A	75	60	0	0
8	A	75	75	2244	11
9	B	60	60	44882	39
10	B	60	75	17176	29

Ship damage example: Stata output

```
. glm damage i.type i.year i.period, family(poisson) link(log) offset(logmonths)
Generalized linear models                         No. of obs      =         34
Optimization     : ML                             Residual df     =         25
                                                  Scale parameter =          1
Deviance         =  38.69504856                   (1/df) Deviance =   1.547802
Pearson          =   42.2752462                   (1/df) Pearson  =    1.69101
Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = ln(u)                   [Log]
                                                  AIC             =   4.545928
Log likelihood   = -68.28076994                   BIC             =  -49.46396
------------------------------------------------------------------------------
             |                 OIM
      damage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        type |
          2  |  -.5433442   .1775899    -3.06   0.002    -.8914141   -.1952744
          3  |  -.6874015   .3290472    -2.09   0.037    -1.332322   -.0424808
          4  |  -.0759614   .2905787    -0.26   0.794    -.6454851    .4935623
          5  |   .3255796   .2358794     1.38   0.168    -.1367355    .7878948
             |
        year |
         65  |   .6971404   .1496414     4.66   0.000     .4038486    .9904321
         70  |   .8184267   .1697737     4.82   0.000     .4856764    1.151177
         75  |   .4534269   .2331705     1.94   0.052    -.0035789    .9104327
             |
   75.period |   .3844667   .1182722     3.25   0.001     .1526575    .6162759
       _cons |  -6.405901   .2174441   -29.46   0.000    -6.832084   -5.979719
   logmonths |          1  (offset)
------------------------------------------------------------------------------
```

Ship damage example: Stata output

```
// relative rate for B vs. C type
. lincom 2.type-3.type, eform
 ( 1)  [damage]2.type - [damage]3.type = 0
------------------------------------------------------------------------------
      damage |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |    1.15495   .3449144     0.48   0.630     .6432213    2.073796
------------------------------------------------------------------------------

// relative rate for 1975- vs. 1960-64
. lincom 75.year, eform
 ( 1)  [damage]75.year = 0
------------------------------------------------------------------------------
      damage |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.573696   .3669394     1.94   0.052     .9964275    2.485398
------------------------------------------------------------------------------
```

Note: The Pearson X2 was 1.69 and Deviance 1.55 indicating moderate overdispersion. In SAS, this can be addressed in Proc Genmod by using the scale option in the model statement: scale = p uses the Pearson, while scale = d uses the deviance to adjust the standard error estimates. In Stata, use scale(x2) or scale(dev) option.

Ship damage example: correct for overdispersion

```
. glm damage i.type i.year i.period, family(poisson) link(log) offset(logmonths) scale(x2)
...
------------------------------------------------------------------------------
             |                 OIM
      damage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        type |
          2  |  -.5433442   .2309359    -2.35   0.019    -.9959702   -.0907183
          3  |  -.6874015   .4278892    -1.61   0.108    -1.526049    .1512459
          4  |  -.0759614   .3778651    -0.20   0.841    -.8165634    .6646406
          5  |   .3255796   .3067348     1.06   0.288    -.2756096    .9267688
             |
        year |
         65  |   .6971404   .1945919     3.58   0.000     .3157472    1.078534
         70  |   .8184267   .2207717     3.71   0.000     .3857221    1.251131
         75  |   .4534269   .3032122     1.50   0.135    -.1408581    1.047712
             |
   75.period |   .3844667   .1537997     2.50   0.012     .0830247    .6859086
       _cons |  -6.405901   .2827618   -22.65   0.000    -6.960104   -5.851699
   logmonths |          1  (offset)
------------------------------------------------------------------------------
(Standard errors scaled using square root of Pearson X2-based dispersion.)
```

Ship damage example: Stata output

```
// relative rate for B vs. C type
. lincom 2.type-3.type, eform
 ( 1)  [damage]2.type - [damage]3.type = 0
------------------------------------------------------------------------------
      damage |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |    1.15495   .4485226     0.37   0.711     .5395116     2.47244
------------------------------------------------------------------------------

// relative rate for 1975- vs. 1960-64
. lincom 75.year, eform
 ( 1)  [damage]75.year = 0
------------------------------------------------------------------------------
      damage |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.573696   .4771638     1.50   0.135     .8686126     2.85112
------------------------------------------------------------------------------
```

Note: the coefficient estimates are the same as previous output, but the standard errors and test p-values are now more conservative.

Poisson regression with common risks

Previously we saw that OR and RR were quite different when the prevalence of the risk was not small, bigger than 10%. In order to estimate an adjusted RR we could use the Binomial model with a log link. But sometimes there are computational difficulties with this model. An alternative is to use the Poisson model AND to utilize the ROBUST STANDARD ERRORS. Regular standard errors for Poisson are known to be too conservative in these cases of approximating binomial data with common risk.

```
*********************Binomial Regression;
proc genmod data = birthwgt2 descending;
     class c_baseline_bmi (ref = "2") /param = ref;
     model hibwt = totalweightgain c_baseline_bmi/dist=binomial link=log type3;
run;
*********************Poisson Regression;
proc genmod data = birthwgt2 descending;
     class c_baseline_bmi (ref = "2") id /param = ref;  <- need to generate an ‘id’ variable for
                                                             every sample;
     model hibwt = totalweightgain c_baseline_bmi/dist=poisson link=log type3;
     repeated sub = id / type=ind;   <- This is the statement to obtain robust std errors;
run;
```

Birthweight example: log link

```
. glm hibwt totalweightgain ib2.c_baseline_bmi, fam(bin) link(log) eform iter(100)
Iteration 100: log likelihood = -1039.0036  (not concave)
convergence not achieved
                                                    AIC             =  1.042004
Log likelihood   = -1039.003635                     BIC             = -13100.99
---------------------------------------------------------------------------------
                |                 OIM
          hibwt | Risk Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
totalweightgain |   1.024364   8.78e-10  2.8e+07   0.000     1.024364    1.024364
                |
 c_baseline_bmi |
             1  |   .0617007   .0352668    -4.87   0.000     .0201262    .1891555
             3  |   .6152626   .0750599    -3.98   0.000     .4844141    .7814554
             4  |   1.855169          .        .       .            .           .
                |
          _cons |   .1368256   7.65e-09 -3.6e+07   0.000     .1368255    .1368256
---------------------------------------------------------------------------------
Warning: parameter estimates produce inadmissible mean estimates in one or
         more observations.
Warning: convergence not achieved
```

Birthweight example: Poisson regression with regular S.E.

```
. glm hibwt totalweightgain ib2.c_baseline_bmi, fam(poisson) link(log) eform
Generalized linear models                          No. of obs      =      2000
Optimization     : ML                              Residual df     =      1995
                                                   Scale parameter =         1
Deviance         =   963.011489                    (1/df) Deviance =  .4827125
Pearson          =  1643.646513                    (1/df) Pearson  =   .823883
Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]
                                                   AIC             =  .7465057
Log likelihood   = -741.5057445                    BIC             = -14200.79

---------------------------------------------------------------------------------
                |                 OIM
          hibwt |        IRR   Std. Err.       z    P>|z|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
totalweightgain |   1.032877   .0046171     7.24   0.000     1.023868    1.041966
                |
 c_baseline_bmi |
             1  |    .187722   .1100813    -2.85   0.004     .0594793    .5924677
             3  |   1.875006   .3199926     3.68   0.000     1.341945    2.619814
             4  |   2.387442   .3428306     6.06   0.000     1.801783    3.163466
                |
          _cons |   .0334007   .0064157   -17.70   0.000      .022922    .0486695
---------------------------------------------------------------------------------
```

Birthweight example: Poisson regression with robust S.E.

```
. glm hibwt totalweightgain ib2.c_baseline_bmi, fam(poisson) link(log) eform robust
Generalized linear models                           No. of obs      =       2000
Optimization     : ML                               Residual df     =       1995
                                                    Scale parameter =          1
Deviance         =   963.011489                     (1/df) Deviance =   .4827125
Pearson          =  1643.646513                     (1/df) Pearson  =    .823883
Variance function: V(u) = u                         [Poisson]
Link function    : g(u) = ln(u)                     [Log]
                                                    AIC             =   .7465057
Log pseudolikelihood = -741.5057445                 BIC             =  -14200.79

---------------------------------------------------------------------------------
                |               Robust
          hibwt |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
totalweightgain |   1.032877   .0038243     8.74   0.000     1.025409      1.0404
                |
 c_baseline_bmi |
             1  |    .187722   .1077046    -2.92   0.004     .0609737    .5779468
             3  |   1.875006   .2884014     4.09   0.000     1.386999    2.534714
             4  |   2.387442   .3093241     6.72   0.000     1.852032    3.077634
                |
          _cons |   .0334007    .005491   -20.68   0.000     .0242002    .0460989
---------------------------------------------------------------------------------
```

Other count models

- Zero-truncated Poisson

  Example: A study by the county traffic court on the number of tickets received by teenagers as predicted by school performance, amount of driver training and gender. Only individuals who have received at least one citation are in the traffic court files.

  SAS: `proc nlmixed ...;`

  (http://www.ats.ucla.edu/stat/sas/faq/zt_nlmixed.htm)

  Stata: `tpoisson y x1 x2, inflate(x1)`

  (http://www.ats.ucla.edu/stat/sas/faq/zt_nlmixed.htm)

Other count models

- Zero-inflated Poisson

  Example: The state wildlife biologists want to model how many fish are being caught by fishermen at a state park. Visitors are asked whether or not they have a camper, how many people were in the group, were there children in the group and how many fish were caught. Some visitors do not fish, but there is no data on whether a person fished or not. Some visitors who did fish did not catch any fish so there are excess zeros in the data because of the people that did not fish.

  SAS:
  ```
  proc genmod data = T; model y = x1 x2 / dist=zip;
       zeromodel x1 / link = logit;
  ```

  (http://www.ats.ucla.edu/stat/sas/dae/zipreg.htm)

  Stata: `zip y x1 x2, inflate(x1)`

  (http://www.ats.ucla.edu/stat/stata/dae/zip.htm)

Other count models

- Negative binomial model:
  – Have one more parameter than the Poisson, the second parameter can be used to adjust the variance independently of the mean.
  – Alternative solution to overdispersed Poisson

SAS: `proc genmod data = T; model y = x1 x2 / dist=negbin;`

Stata: `glm y x1 x2, family(nb)`

- Zero-inflated negative binomial model
- Zero-truncated negative binomial model

(*. Check the UCLA website for examples)

Estimation and Goodness of fit for GLM

- Estimation:
  - MLE: maximize the log-likelihood
  - Iterated Weighted Least Squares (IWLS): analogous to weighted least squares, with the weights updated iteratively

  GLM fits using IWLS is more general than MLE (see e.g. quasi-likelihood if you want to know more details)

- Goodness of fit:
  - Pearson's X^2: (does not depend on distribution assumption)

$$X^2 = \sum_{i=1}^{n} \frac{\left(Y_i - \hat{\mu}_i\right)^2}{V\left(\hat{\mu}_i\right)/m_i}$$

    where m_i = 1 for Normal and Poisson, n_i for Binomial.

  - Deviance: (does not depend on ϕ, but depends on log-likelihood)

$$D\left(\underline{Y}, \underline{\hat{\mu}}\right) = 2\phi\left\{l\left(\underline{Y}, \phi; \underline{Y}\right) - l\left(\underline{\hat{\mu}}, \phi; \underline{Y}\right)\right\}$$

$$= 2\sum_{i=1}^{n} m_i \left\{Y_i\left(\tilde{\theta}_i - \hat{\theta}_i\right) - \left[b\left(\tilde{\theta}_i\right) - b\left(\hat{\theta}_i\right)\right]\right\}$$

    where $b'\left(\tilde{\theta}_i\right) = Y_i$ (natural parameter for the saturated model)

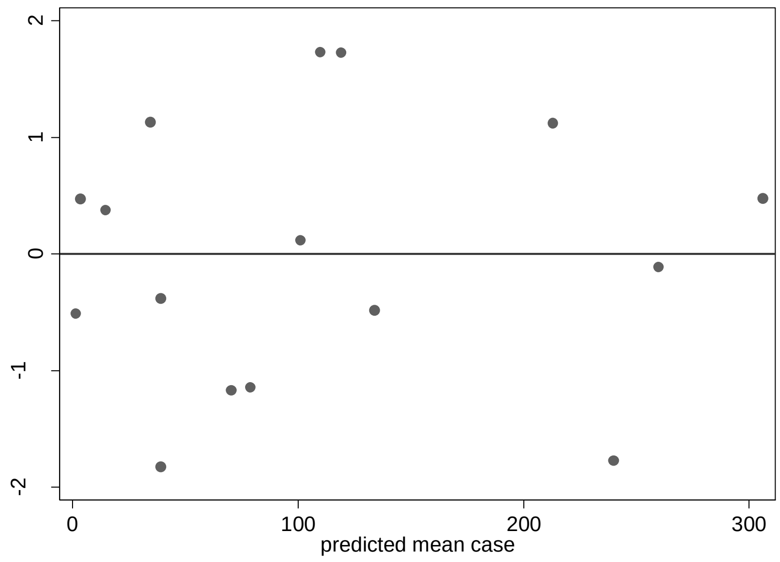
Residuals

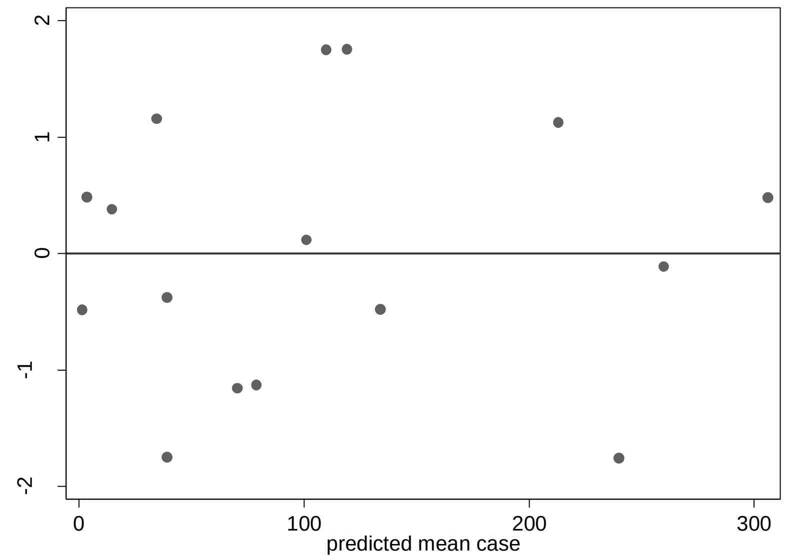
- Types of Residuals:
  - Pearson residual: $$r_i^P = \frac{Y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$
  - Deviance residual: $$r_i^D = sign(Y_i - \hat{\mu}_i)\sqrt{d_i}$$
  - Response residual: $$r_i^R = Y_i - \hat{\mu}_i$$
  - Working residual: $$r_i^W = (Y_i - \hat{\mu}_i)\frac{\partial \hat{\eta}_i}{\partial \hat{\mu}_i}$$
  - Partial residual: $$r_{ki}^T = (Y_i - \hat{\mu}_i)\frac{\partial \hat{\eta}_i}{\partial \hat{\mu}_i} + x_{ik}\hat{\beta}_k$$

  Note: residual diagnostics non-informative for binary data

- Overdispersion: Var(Y) > $a(\phi)V(\mu)$ for any GLM
  check if $X^2/(n\text{-}p)$ or $D/(n\text{-}p) >> 1$

Checking outliers

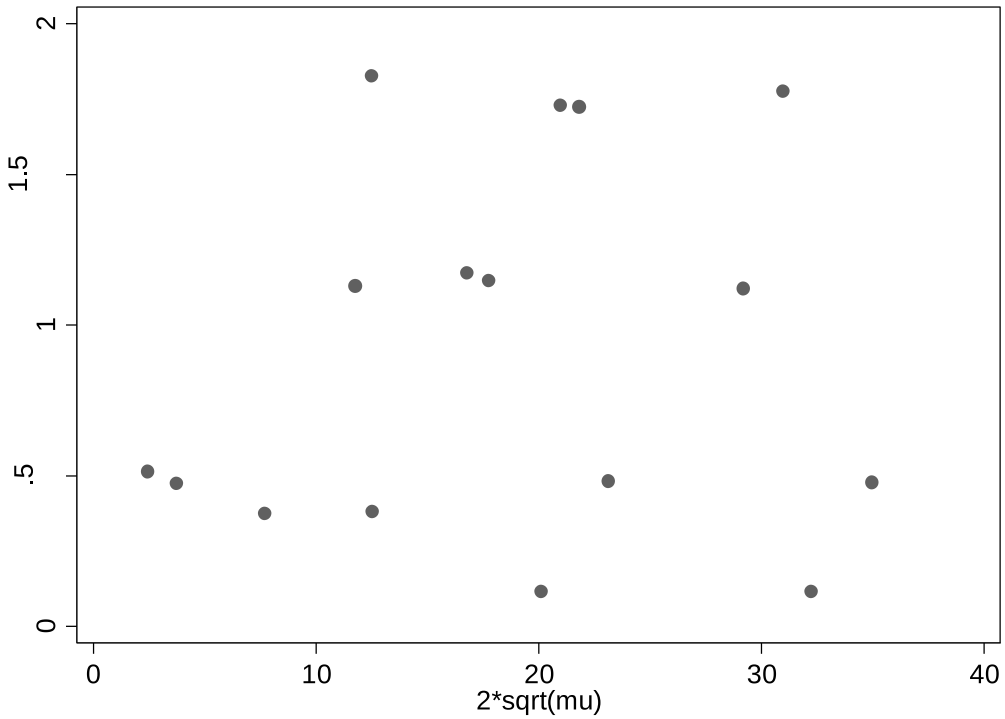




Check for observations with large (standardized) Pearson or deviance residuals.

Deviance residual is generally preferred to the Pearson residual, since its distributional properties are closer to the residuals in linear regression.

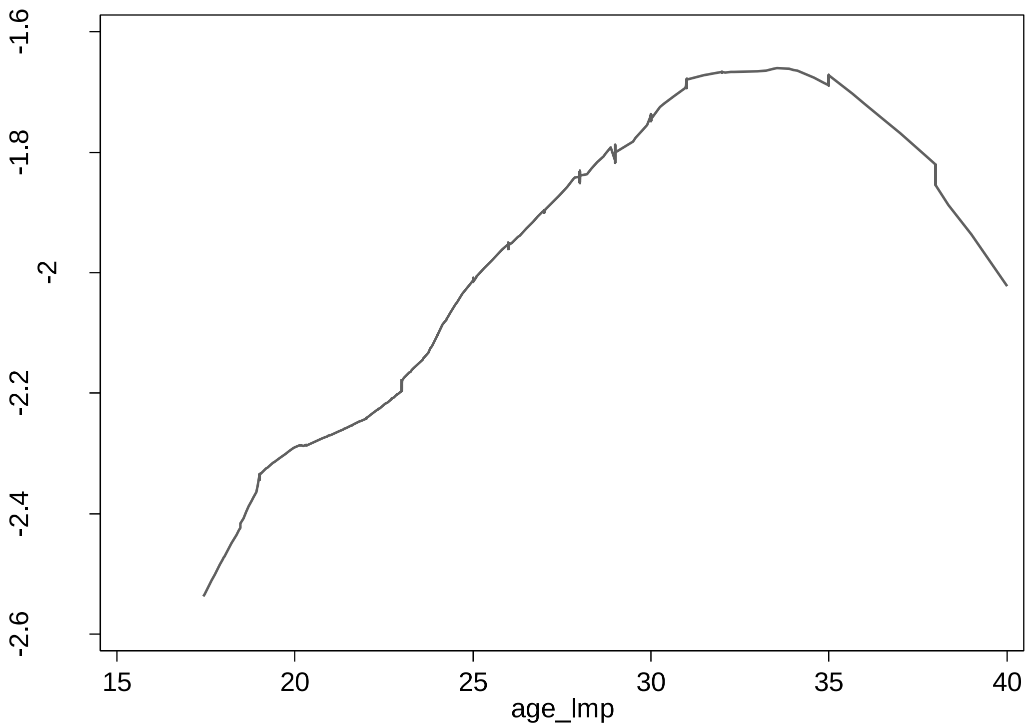
Checking the variance function



The x-axis is “constant-information scale” for Poisson errors. See McCullagh and Nelder, “Generalized Linear Models” (Chapter 12.6.1).

Positive trend indicates the current variance function increases too slowly with the mean.

Checking for linearity: lowess plot



Other approach to check departure from linearity:

- Add polynomial terms (e.g., age^2) and test for their statistical significance;
- Group continuous predictor into categories (e.g., age_cat).

Other Diagnostics

- Checking the link function: the simplest method is to include $\hat{\eta}^2$ as covariate and assessing the fall in deviance. (Hinkley, 1985)
- Partial residual is helpful to assess the form of covariates, i.e., plot partial residual against the corresponding predictor.
- Checking influential points: modified Cook's distance (> 4/(*n*-*p*) to be problematic)