Poisson is the commonly encountered probability distribution after normal distribution. The latter assumes that true values are normally distributed around the mean and can take any non-imaginary value e.g. positive/ negative/ integer/ fractional, which makes it inappropriate for modeling count/discrete outcomes (e.g. number of days of hospital stay). As such, count measure are analysed using Poisson or negative binomial models.

The Poisson regression model assumes that the mean of the outcome measure is the same as its variance (equidispersion), non-compliance of which is dealt with negative binomial (for overdispersion cases) and generalized Poisson models (overdispersion or underdispersion cases). Some other methods are used for more rigorous handling of cases with excessive count of zeros e.g. zero-inflated Poisson, zero-altered Poisson and zero-inflated negative binomial models.

Here is a use case of poisson regression

Data

GSS 2016

Variables

Outcome

mntlhlth = days of poor mental health in past 30 days

Exposure

hhrace = race

relig= religion

hapmar = marital satisfaction

rhlthend = health rating

Glimpse of the variables

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    mntlhlth |      1,099    3.967243    7.517364          0         30
      hhrace |      2,855    1.557618    1.157873          1          5
       relig |      2,849    2.312039     1.93718          1         13
      hapmar |      1,204    1.436877     .570982          1          3
    rhlthend |      1,332    1.865616    .7561147          1          4
```

Output of the Poisson model

```
. poisson mntlhlth  i.hhrace i.relig i.hapmar i.rhlthend, irr
cformat(%9.2f)

          mntlhlth |        IRR   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------------+------------------------------------------------------
----------
            hhrace |
            white  |       1.00  (base)
            black  |       1.70       0.45     2.04   0.041         1.02
     2.84
```

```
  asiatic, oriental  |      12.58       4.57     6.97   0.000         6.17
    25.63
      other, mixed  |       0.17       0.18    -1.74   0.082         0.02
      1.25
                    |
             relig  |
       protestant  |       1.00  (base)
         catholic  |       3.43       0.57     7.43   0.000         2.48
      4.75
             none  |       1.35       0.26     1.56   0.119         0.93
      1.96
            other  |       0.00       0.00    -0.02   0.988         0.00
         .
                    |
           hapmar  |
       very happy  |       1.00  (base)
      pretty happy  |       2.39       0.33     6.33   0.000         1.83
      3.13
     not too happy  |       7.54       1.84     8.28   0.000         4.68
     12.16
                    |
          rhlthend  |
        excellent  |       1.00  (base)
             good  |       0.61       0.10    -3.09   0.002         0.45
      0.84
             fair  |       2.32       0.36     5.46   0.000         1.71
      3.13
             poor  |       0.38       0.18    -2.05   0.040         0.15
      0.96
                    |
            _cons  |       0.79       0.14    -1.29   0.198         0.56
      1.13
```

Reading example

Compared with participants who identify themselves as White, those who identify themselves as Black have 1.7 times and the Asian have 12.6 times the incident of depression i.e. Blacks and Asian race are at higher risk of depression compared with Whites (ceteris paribus).

Checking fitness of the model

```
. estat gof

         Deviance goodness-of-fit =  438.4644
         Prob > chi2(101)         =    0.0000

         Pearson goodness-of-fit  =  518.4127
         Prob > chi2(101)         =    0.0000
```

Poisson model does not fit the well, goodness-of-fit chi-squared test is statistically significant ($p<0.05$).

Running NB model as an alternative

```
 nbreg mntlhlth  i.hhrace i.relig i.hapmar i.rhlthend, irr cformat(%9.2f)
nolog
```

```
------------------------------------------------------------------------------
          mntlhlth |        IRR   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------------+----------------------------------------------------------
             hhrace |
             white  |       1.00  (base)
             black  |       1.21       0.94     0.24   0.808         0.26
     5.57
asiatic, oriental  |      15.06      24.79     1.65   0.099         0.60
   379.07
     other, mixed  |       0.15       0.20    -1.45   0.148         0.01
     1.97
                   |
              relig |
       protestant  |       1.00  (base)
         catholic  |       3.28       1.36     2.86   0.004         1.45
     7.41
              none  |       1.29       0.59     0.54   0.586         0.52
     3.18
             other  |       0.00       0.00    -0.00   0.999         0.00
         .
                   |
             hapmar |
       very happy  |       1.00  (base)
     pretty happy  |       3.29       1.23     3.19   0.001         1.59
     6.84
    not too happy  |       8.28       8.13     2.15   0.031         1.21
    56.72
                   |
           rhlthend |
        excellent  |       1.00  (base)
```

```
             good  |       0.78       0.31    -0.63   0.529         0.36
      1.69
             fair  |       2.31       1.16     1.68   0.093         0.87
      6.16
             poor  |       0.35       0.45    -0.82   0.410         0.03
      4.27
                   |
            _cons  |       0.66       0.27    -1.00   0.316         0.30
      1.48
```

Reading example: Those who reported being pretty happy and not too happy have higher incidence of depression compared with those who reported being very happy with their marriage i.e. higher marital satisfaction is inversely associated with the risk of depression (ceteris paribus).

Model comparison

```
Model:                        nbreg         poisson
N:                              113             113               0
Log-Lik Intercept Only:     -209.026        -435.259         226.233
Log-Lik Full Model:         -194.552        -298.508         103.956
D:                       389.104(96)     597.015(97)    -207.912(-1)
LR:                       28.949(11)     273.502(11)    -244.553(0)
Prob > LR:                     0.002           0.000           0.002
McFadden's R2:                 0.069           0.314          -0.245
McFadden's Adj R2:            -0.012           0.277          -0.290
```

```
Maximum Likelihood R2:          0.226          0.911         -0.685
Cragg & Uhler's R2:             0.232          0.912         -0.680
AIC:                            3.744          5.567         -1.822
AIC*n:                        423.104        629.015       -205.912
BIC:                          -64.726        138.459       -203.185
BIC':                          23.052       -221.501        244.553
```

Note: model with the lower AIC is a bit better i.e. nbreg is more suitable than Poisson for this data.