Modeling Count Data

Cambridge University Press : 2014

Joseph M. Hilbe

hilbe@asu.edu or j.m.hilbe@gmail.com

18Dec2014: 15ep2016 update

* CHAPTER 1

P. 5

```
. use smoking
. regress sbp male smoker age, nohead
. predict mu
```

P. 6

```
. gen xb =  _b[male]*male + _b[smoker]*smoker + _b[age]*age + _b[_cons]
. l
. di _cons
1
. di _b[_cons]    /* intercept slope x 1  */
```

P 13

Table 1.2b Stata: *Code for Figure 1.2b*

```
=================================================================
clear
set obs 15
gen byte mu = 4
gen byte y = _n-1
gen yp = (exp(-mu)*mu^y)/exp(lngamma(y+1))
gen alpha = .5
gen amu = mu*alpha
gen ynb = exp(y*ln(amu/(1+amu)) - (1/alpha)*ln(1+amu) + lngamma(y +1/alpha) /*
          */ - lngamma(y+1) - lngamma(1/alpha))
PigPr y, m(2) mu(4) pr(ypig)
gen ygp = exp(ln((1-alpha)*mu) + (y-1)*ln((1-alpha)*mu+alpha*y) - /*
          */ (1-alpha)*mu - alpha*y - lngamma(y+1))
gen ynb1 = exp(lngamma(mu/alpha + y) - lngamma(y+1) - lngamma(mu/alpha) + /*
          */ (mu/alpha)*ln(1/(1+alpha)) + y*ln(1-1/(1+alpha)))
lab var yp "Poisson"
lab var ynb "NB2"
lab var ynb1 "NB1"
lab var ypig "PIG"
lab var ygp "generalized Poisson"
graph twoway connected ygp ypig ynb1 ynb yp y, ms(s D T d S) /*
*/ title("POI | NB | NB1 | GP |PIG distributions: MEAN = 4; a=0.5")
=================================================================
```

PigPr command for calc synthetic PIG

```
======================================================================
capture program drop PigPr
program define PigPr
    syntax varname [if] [in] [, m(real 0.1) mu(real 3.0) pr(name) REPLACE ]
        quietly {
          capture confirm variable `pr'
          if _rc == 0 & "`replace'" == "" {
```

```
            noi di as err "Variable `pr' already exists"
            exit 198
          }
          tempvar pp
          gen double `pp' = .
          summ `varlist' `if' `in'
          local n = r(max)
          local i  = 0
          local x  = 0

          while `x' <= `n' {
             local p0   = exp((1/`m')* (1-sqrt(1+2*`m'*`mu')))
             local pjm2 = `mu'/sqrt(1+2*`m'*`mu')*`p0'
             local pjm1 = 2*`m'*`mu'/(1+2*`m'*`mu')*.25*`pjm2' + ///
               `mu'^2/(1+2*`m' *`mu')*.50*`p0'
             local ty   = -lngamma(`x'+1) + ln(`p0')
             local ty   = `ty' + (`x'>0)*log((0+1)*`pjm2'/`p0')
             local ty   = `ty' + (`x'>1)*log((1+1)*`pjm1'/`pjm2')
             local ti   = (`x'==0)*1*`pjm2'/`p0'
             local ti   = (`x'==1)*2*`pjm1'/`pjm2'
             local tip1 = (`x'==0)*2*`pjm1'/`pjm2'

             local ym1 = `x'-1
             forvalues ii=2/`ym1' {
               local pj = 2*`m'*`mu'/(1+2*`m'*`mu')*(1-3/(2*(`ii'+1)))* ///
                 `pjm1' + `mu'^2/(1+2*`m'*`mu')*(1/(`ii'*(`ii'+1)))*`pjm2'
               local tip1 = (`x'==`ii'-1) * (`ii'+1)*`pj'/`pjm1'
               local ti   = (`x'==`ii')   * (`ii'+1)*`pj'/`pjm1'
               local ty   = `ty' + (`x'>`ii')*log((`ii'+1)*`pj'/`pjm1')
               replace `pp' = exp(`ty') if `varlist'==(`x')
               local pjm2 = `pjm1'
               local pjm1 = `pj'
          }
          replace `pp' = exp(`ty') if `varlist'==`x'
          local x = `x'+1
      }
      capture drop `pr'
      gen double `pr' = `pp' `if' `in'
 }
end
======================================================================
```

P 26

```
. di  exp(-2) * (2^0)/0
. di exp(-2)* (2^0)/exp(lnfactorial(0))
. di exp(-2)

. di exp(-2)* (2^1)/exp(lnfactorial(1))
. di exp(-2)* (2^2)/exp(lnfactorial(2))
. di exp(-2)* (2^3)/exp(lnfactorial(3))
. di exp(-2)* (2^4)/exp(lnfactorial(4))
```

P 27

```
. di poissonp(2,2)
. di poisson(2,2)
```

P 28

Figure 1.3

```
===============================================================================
clear
set obs 11
gen y = _n-1
gen mu = .
gen mu0_5 = (exp(-.5)* .5^y)/exp(lngamma(y+1))
forvalues i = 1(2)6 {
gen mu`i' = (exp(-`i')*`i'^y)/exp(lngamma(y+1))
}
graph twoway connected mu0_5 mu1 mu3 mu5 y, title("Poisson Distributions")
================================================================================
```

* CHAPTER 2

P. 42

```
. clear
. set obs 50000
. set seed 4590
. gen x1 = runiform()
. gen x2 = runiform()
. gen x3 = runiform()
. gen xb = 1 + 0.75*x1 - 1.25*x2 + .5*x3
. gen exb =  exp(xb)
. gen py = rpoisson(exb)
. tab py
```

P. 43

```
. sum py, detail
. glm py x1 x2 x3, fam(poi) nolog
. abic
```

P. 46

Table 2.5 Stata: *Monte Carlo Poisson Code*

```
=========================================================================
program define poix_sim, rclass /* poix_sim.ado */
version 10
drop _all
set obs 50000
gen x1 = runiform()
gen x2 = runiform()
gen x3= runiform()
gen py = rpoisson(exp(1 + 0.75*x1 - 1.25*x2 + .5*x3))
glm py x1 x2 x3, nolog fam(poi)
return scalar sx1 = _b[x1] /// x1
return scalar sx2 = _b[x2] /// x2
return scalar sx3 = _b[x3] /// x3
return scalar sc = _b[_cons] /// intercept
return scalar ddisp = e(dispers_s) /// deviance dispersion
return scalar pdisp = e(dispers_p) /// Pearson dispersion
```

```
end
=================================================================

. simulate mx1 = r(sx1) mx2 = r(sx2) mx3 = r(sx3) mcon = r(sc)
  mdd = r(ddisp) mpd = r(pdisp) , reps(500) : poix_sim
. sum
```

P 48

```
. use rwm1984
. tab docvis
```

P. 49

```
. sum docvis
. di exp(-3.162881)* (3.162881^0)/exp(lnfactorial(0))
. tab outwork
```

P. 50

```
. sum age
. distinct age
. center age, pre(c)
. glm docvis outwork cage, fam(poisson) nolog
```

p. 51

```
. abic
```

p. 53

```
. tab edlevel
. glm docvis outwork cage female married kids i.edlevel, fam(poi) nolog
. tab edlevel, gen(edlevel)  // create indicator vars for each level of edlevel
. glm docvis outwork cage female married kids edlevel3-edlevel4, fam(poi) nolog
```

p. 54

```
. abic
. gen elevel = edlevel
. recode elevel 2=1  3=2  4=3
. lab define elevel 1 "HS" 2 "Coll/Univ" 3 "Grad School"
. lab values elevel elevel
. tab elevel
```

p. 56

```
. glm docvis outwork age, fam(poi) nolog nohead
```

p. 57

```
. matrix list e(V)
. di sqrt(.00035512)
. di sqrt(7.018e-07)
. di sqrt(.00153519)
```

p. 58

```
. di 0.4079314 /0.0188447
. di normal(-0.033517 / 0.0391815)*2
. di .4079314 -  1.96* .0188447
. di .4079314 +  1.96* .0188447
```

p. 60

```
. di   exp( _b[outwork])
. di   exp( _b[age])
. di   exp( _b[outwork]) * _se[outwork]
. di   exp( _b[age]) * _se[age]
. glm docvis outwork age, fam(poi) nolog nohead eform
```

p. 65

```
. use fasttrakg,clear
. l, nolab
```

p. 66

```
. glm die anterior hcabg kk2-kk4, nolog fam(poi) eform exposure(cases)
. abic
```

p. 67

```
. glm docvis outwork age, nolog fam(poi) nohead
. di _b[outwork]*0 + _b[age]*50 - .033517
. di exp(1.0706929)
```

Predicted counts and 95% confidence intervals

```
=============================================================
poisson docvis outwork age
predict mu ; predict eta, xb
predict se_eta, stdp
gen low = eta - invnormal(0.975) * se_eta
gen up  = eta + invnormal(0.975) * se_eta
gen lci = exp(low)
gen uci = exp(up)
sort mu;
twoway (line lci mu uci eta, lpattern( dash 1 dash 1)),
      ytitle("Predicted Count and   95% CI"); #delimit cr
==============================================================
```

p. 69/72

```
. margins, dydx(age) atmeans
. margins, dydx(age)
. qui glm docvis i.outwork age,  fam(poi)
. margins, dydx(outwork) atmean
. margins, dydx(outwork)
```

*   CHAPTER 3

p. 76

```
. glm docvis outwork age, fam(poi) nolog
. scalar dev=e(deviance)
. scalar df=e(df)
. di " deviance GOF "" D="dev " df="df " p-value= " chiprob(df, dev)
```

p. 80/83

```
. qui poisson docvis outwork age
. estat gof
. glm py x1 x3, nolog fam(poi)
```

p. 86

```
. glm docvis outwork age,  fam(poi) nolog nohead
. predict mu
. gen double z=((docvis-mu)^2-docvis)/ (mu*sqrt(2))
. regress z
```

p. 87

Lagrange Multiplier Test

```
==================================================================
summ docvis, meanonly
scalar nybar = r(sum)
gen double musq = mu*mu
summ musq, meanonly
scalar mu2 = r(sum)
scalar chival = (mu2-nybar)^2/(2*mu2)
display "LM value = " chival _n "P-value = " chiprob(1,chival)
==================================================================
```

p. 89

Table 3.6 -- Observed versus Predicted Counts

```
==================================================================
use rwm1984, clear
qui glm docvis outwork age, fam(poisson)
predict mu
count
gen nobs = e(N)
local i 0
local newvar "pr`i'"
* Predicted probability at each day
while `i' <=25 {
local newvar "pr`i'"
3
qui gen `newvar' = poissonp(mu, `i')
local i = `i' + 1
}
quietly gen cnt = .
quietly gen observ = .
quietly gen expect = .
local i 0
*: Observed and expected docvis
while `i' <=25 {
local obs = `i' + 1
replace cnt = `i' in `obs'
tempvar obser
gen `obser' = `e(depvar)' ==`i' /* (docvis==`i') */
sum `obser'
replace observ = r(mean)* nobs in `obs'
sum pr`i'
replace expect = r(mean)* nobs in `obs'
local i = `i' + 1
}
*: Preparation for table
gen byte count = cnt
gen diff = observ - expect
drop cnt pr0-pr25 nobs mu
list count observ expect diff in 1/21
lab var expect "Expected days"
lab var observ "Observed days"
```

```
label var count "Number visits to Physician"
twoway scatter expect observ count, c(l l) ms(T d) ///
title(Observed vs Expected visits) ytitle(Count of patients)
================================================================
```

p. 92

```
. count
. count if docvis==0
. di 1611/3874
. di exp(-3.162881) * 3.162881^0 / exp(lnfactorial(0))
```

p. 93/102

```
. glm los hmo white type2 type3, fam(poi) nolog
. glm los hmo white type2 type3, fam(poi) nolog scale(x2)
. glm los hmo white type2 type3, nolog fam(poi) eform scale(x2) nohead
. glm los hmo white type2 type3, fam(poi) nolog disp(6.260391) irls
. di .023944/sqrt(6.260391)
. glm los hmo white i.type, fam(poi) vce(robust) nolog
. glm los hmo white i.type, fam(poi) cluster(provnum)nolog
. di normal(-1.96)*2
. glm los hmo white i.type, fam(poi)  nolog nohead
```

p. 104/105

```
. use titanic
. gen died = survived
. recode died 1=0 0=1
. tab died age
. di 765/1207
. di (765/1207) / (52/109)
. glm died age, fam(poi) nolog nohead vce(robust) ef
. bootstrap, reps(1000): glm los hmo white type2 type3, fam(poi)
```

* CHAPTER 4

p. 112/113

Likelihood ratio test

```
=============================================================
use rwm1984,clear
qui glm docvis outwork age, fam(poi) // full model
est store A
qui glm docvis outwork, fam(poi) // reduced model, drop age
est store B
lrtest A B
qui poisson docvis outwork age
lrdrop1
================================================================
```

p. 115

```
. di chi2tail(1,3.84)
. di chi2tail(1,2.705)/2
```

p. 221

```
. glm los hmo white type2 type3, fam(poi) nolog nohead
```

```
. abich
. estat, ic
```

* CHAPTER 5

p. 137/139

```
. sum docvis
. di exp(-3.162881)* (3.162881^0)/exp(lnfactorial(0))
. use rwm1984
. global xvar "outwork age female married edlevel2 edlevel3 edlevel4"
. glm docvis $xvar, fam(poi) vce(robust) nolog nohead
. countfit docvis $xvar, prm nbreg max(12)
     /* PRM=Poisson regresson model; NBRM = NB regression model */
```

p. 141/142

```
. glm docvis $xvar, fam(nb ml) vce(robust) nolog
. qui nbreg docvis $xvar,  vce(robust) nolog
. abich
. scalar llnb2 = e(ll)
```

p. 143

Table 5.3 *Observed versus Predicted Counts for docvis*

```
=====================================================================
use rwm1984
qui {
qui nbreg docvis outwork age married female edlevel2 edlevel3 edlevel4
predict mu
local alpha = e(alpha)
gen amu = mu* e(alpha)
local i 0
local newvar "pr`i'"
while `i' <=15 {
local newvar "pr`i'"
qui gen `newvar' = exp(`i'*ln(amu/(1+amu)) - (1/`alpha')*ln(1+ amu) + /*
  */ lngamma(`i' +   1/`alpha') - lngamma(`i'+1) - lngamma(1/`alpha'))
local i = `i' + 1
}
quietly gen cnt = .
quietly gen obpr = .
quietly gen prpr = .
local i 0
while `i' <=15 {
 local obs = `i' + 1
 replace cnt = `i' in `obs'
 tempvar obser
 gen `obser' = (`e(depvar)'==`i')
 sum `obser'
 replace obpr = r(mean) in `obs'
 sum pr`i'
 replace prpr = r(mean) in `obs'
 local i = `i' + 1
}
gen byte count = cnt
label var prpr "NB2 - Predicted"
label var obpr "NB2 - Observed"
label var count "Count"
}
```

```
twoway scatter prpr obpr count, c(l l) ms(T d) title("docvis Observed vs Predicted
Probabilities") sub("Negative Binomial") ytitle(Probability of Physician Visits) ||
lowess obpr count, bwidth(.3)
=========================================================================
```

p. 144

```
. rename prpr prnb2
. rename obpr obnb2
. drop mu*
. drop count
```

p. 145/146

```
. nbreg docvis $xvar, nolog vce(robust) disp(constant)
. abich
. scalar llnb1 = e(ll)
. nbreg docvis $xvar, nolog  vce(robust) irr
```

p. 148/149

```
. tab type
. glm los hmo white type2 type3, nolog fam(poi) vce(robust)
. abich
. tab los
```

p. 150/151

```
. glm los hmo white type2 type3, nolog fam(nb ml) vce(robust)
. abich
. linktest
. qui nbreg los hmo white type2 type3, nolog disp(const)
. abich
. linktest
```

p. 154/155

```
. use rwm1984
. global xvar "outwork age female married edlevel2 edlevel3 edlevel4"
. nbregp docvis $xvar, nolog vce(robust)
. abich
. scalar llnbp = e(ll)
. di -2*(llnb2 - llnbp)
. di chi2tail(1, 56.72852)
. di (1.363081 -2)/.1248427
. di ttail(1, -5.1017721)
```

p. 160

Table 5.9 Stata: *Squirrel Data*

```
================================================================
. use nuts
. center ntrees, prefix(s) standard
. center height, prefix(s) standard
. center cover, prefix(s) standard
. global xvars "sntrees sheight scover"
. glm cones $xvars if dbh<.6, fam(pois) nolog
. nbreg cones $xvars if dbh<.6, nolog
. gnbreg cones $xvars if dbh<.6, nolog lnalpha($xvars)
. gnbreg cones $xvars if dbh<.6, nolog lnalpha($xvars) vce(robust)
===========================================================================
```

* CHAPTER 6

p. 166/167

```
. use rwm1984
. glm docvis outwork age, nolog  vce(robust) fam(nb ml)
. abic
. linktest
. pigreg docvis outwork age, nolog vce(robust)
. abich
. predict mupig
. linktest
. pigreg, irr
```

p. 169

```
. pigreg los hmo white type2 type3, vce(robust) irr
. abich
. linktest
```

* CHAPTER 7

p. 172

```
. di exp(-3) * 3^0 / exp(lnfactorial(0))
```

p. 175/177

```
. di exp(-5) * 5^0 / exp(lnfactorial(0))
. di  %10.9f exp(-12) * 12^0 / exp(lnfactorial(0))
. sum los
. di  %10.9f exp(-9.854181) * 9.854181^0 / exp(lnfactorial(0))
. di  %10.9f exp(-9.854181) * 9.854181^0 / exp(lnfactorial(0)) * 1495
. glm los white hmo type2 type3, fam(poi) nolog nohead
. abic
. ztp los white hmo type2 type3, nolog
. abic
```

p. 178

```
. ztnb los white hmo type2 type3,  nolog
. abic
. linktest
```

p. 179/180

```
. glm los white hmo type2 type3, fam(nb ml) nolog nohead
. abic
----------------
. gen a = 1
. gen mu = 2
. gen y=0
. di exp(y*log((a*mu)/(1+a*mu))-(1/a)*log(1+a*mu)+lngamma(y+1/a)- /*
       /*  lngamma(y+1)-lngamma(1/a))
----------------
```

p. 181/183

```
. ztpig los hmo white type2 type3, nolog
. abich
. ztnbp los hmo white type2 type3,nolog
```

```
. abich
. ztpnm los hmo white type2 type, vuong
. abich
```

p. 186/187

```
. hplogit docvis outwork age, nolog
. abich
. gen visit=docvis>0
. logit visit outwork age, nolog
. abich
```

p. 188/189

```
. ztp docvis outwork age if docvis>0, nolog
. abich
. hnblogit docvis outwork age, nolog
. abich
```

p. 190/191

```
. hnblogit docvis outwork age, nolog irr vce(robust)
. hnblogit_p mu, eq(#2) irr
. l mu docvis outwork age in 1/5
. di exp([#2]_b[outwork]*0 + [#2]_b[age]*54 + [#2]_b[_cons])
. di exp([#2]_b[outwork]*1 + [#2]_b[age]*44 + [#2]_b[_cons])
```

p. 191/192

Section 7.2.1 Hurdle Models

```
================================================================
use rwm1984
ztnb docvis i.outwork age if docvis>0
margins, dydx(*) atmeans noatlegend
gen visit = docvis>0
logit visit i.outwork age
margins, dydx(*) atmeans noatlegend
margins, dydx(*) atmeans noatlegend predict(equation(#1))
margins, dydx(*) atmeans noatlegend predict(equation(#2))
tab outwork, gen(outwork1)
================================================================
```

p. 192/194

```
. ztpig docvis outwork age if docvis>0, nolog
. abich
. di 11403.847+5101.605
. ztpnm docvis outwork age if docvis>0, nolog
. abich
. di 11403.333+5101.605   /// AIC value
```

p. 195

```
. gen rcen=1
. replace rcen=-1 if docvis>=1
. cpoissone docvis outwork age, censor(rcen) cright(1) nolog
. di 5098.184+11403.847     /// AIC
```

p. 199/202

Zero Inflated Poisson

```
==========================================================================================
zip docvis outwork age, nolog inflate(outwork age) vuong
abic
zip, irr
predict prob /* predicted count */
gen xb_c = [docvis]_cons + [docvis]_b[outwork] *outwork + [docvis]_b[age]*age
gen xb_b = [inflate]_cons + [inflate]_b[outwork]*outwork + [inflate]_b[age]*age
gen pr0 = 1/(1+exp(-xb_b))
gen prcnt=exp(xb_c)*(1-pr0)
su docvis prob prcnt pr0
==========================================================================================
```

p. 202/204

Zero Inflated NB

```
==========================================================================================
zinb docvis outwork age, nolog inflate(outwork age) zip vuong irr
abic
predict prob /* predicted count */
gen xb_c = [docvis]_cons + [docvis]_b[outwork] *outwork + [docvis]_b[age]*age
gen xb_b = [inflate]_cons + [inflate]_b[outwork]*outwork + [inflate]_b[age]*age
gen pr0 = 1/(1+exp(-xb_b))
gen prcnt=exp(xb_c)*(1-pr0)
su docvis prob prcnt pr0
==========================================================================================
```

p. 206/207

Zero Inflated PIG

```
====================================================================
. zipig docvis outwork age, nolog inflate(outwork age) vuong zip irr
. abich
. predict prpc, n=
. predict prp0, pr
. su docvis prpc prp0
====================================================================
```

*   CHAPTER 8

p. 212/215

Generalized Poisson model and diagnostics

```
======================================================================
. use azprocedure, clear
. sum los
. hist los if los<40, title("LOS for full Heart Procedure Data") discrete xlab( 5 8.83
      "mean" 10 15 20 25 30 35) percent
. glm los procedure sex admit, nolog fam(poi) vce(robust) nohead nolog eform
. abich
. di e(dispers_p)
. di e(N)
. drop if los>8
. sum los
. glm los procedure sex admit, nolog fam(poi) vce(robust) eform
. abich
. gpoisson los procedure sex admit, nolog  vce(robust)
. abich
```

```
. zigp docvis outwork age, nolog inflate(outwork age) vuong zip eform
. abic
. gpois_p prgc, n
. predict prg0, pr
. sum docvis  pr*
==================================================================
```

* CHAPTER 9

p. 218/223

9.1 Small and Unbalanced Data -- Exact Poisson Regression

```
========================================================
. use azcabgptca, clear
. tab los
. tab procedure type
. sort procedure
. by procedure: sum los
. sum los
. glm los procedure type, fam(poi) eform nolog
. abic
. glm los procedure type ,  fam(poi) eform scale(x2) nolog nohead
. glm los procedure type,  fam(poi) eform vce(robust) nolog nohead
. ztp los procedure type ,  nolog irr vce(robust)
. abic
. linktest
. expoisson los procedure type, irr
======================================================
```

p. 227/232

9.2 Modeling Truncated and Censored Counts

```
==========================================================================
. use rwm1984, clear
. treg docvis outwork age if docvis>3, dist(poisson) ltrunc(3) nolog vce(robust)
. treg docvis outwork age if docvis>10, dist(pig)  rtrunc(10)  nolog vce(robust)
. treg docvis outwork age if docvis>0 & docvis<19, dist(nbp) ltrunc(0) rtrunc(19)
     vce(robust) eform
. gen cenvar=1
. replace cenvar if docvis<=3
. cpoissone docvis outwork age, censor(cenvar) cleft(3) nolog
. replace cenvar=1
. replace cenvar=-1 if docvis>=10
. cpoissone docvis outwork age, censor(cenvar) cright(10) nolog
. replace cenvar=1
. replace cenvar=-1 if docvis>=3
. epoissone docvis outwork, nolog cright(3)
. cpoissone docvis outwork age, censor(cenvar) cright(3) nolog
. gen visit = docvis>=3
. tab visit
. glm visit outwork age, fam(bin) nolog nohead
==========================================================================
```

p. 233/235

9.3 Counts with Multiple Components – Finite Mixture Models

```
====================================================================
. use fishing
. fmm totabund meandepth, exposure(sweptarea) components(2) mixtureof(negbin2)
```

```
. predict mean1, equation(component1)
. predict mean2, equation(component2)
. sum mean*
. di .5160794 * 150.9075 + .4839206 * 335.3274
. tab period
. nbreg totabund meandepth period, exposure(sweptarea) nolog
========================================================================
```

p. 238
9.5 When All Else Fails: Quantile Count Models

```
===========================================
. use rwm1984, clear
. qcount docvis outwork age, q(.5) rep(1000)
===========================================
```

p. 239/244
9.6 A Word about Longitudinal and Clustered Count Models

```
============================================================
. use medpar, clear
. encode provnum, gen(hospital)
. xtgee los hmo white age80 type2 type3, i(hospital) c(exch) vce(robust)fam(poi) eform
. xtmepoisson los hmo white type2 type3 || provnum:
. predict raneff, ref
. sum raneff
. tab year
. meglm docvis outwork age female married   || id: || year:, fam(poi)
============================================================
```

p. 246/248
9.7 Three-Parameter Count Models

```
=================================================
. use rwm1984, clear
. global xvar "outwork age female married"
. zinbregw docvis $xvar, vce(robust) nolog inflate($xvar)
=================================================
```

p. 248/252
9.8 Bayesian count model

When this book was written Stata had not developed any Bayesian capability. In 2014, after this book was published, Stata developed a new release with a *bayesmh* command. I have written some 65 separate Bayesian models using Stata. In the second edition of this book I will expand this section to a full chapter and discuss Bayesian models using R/JAGS and Stata in more detail.

Stata’s bayesmh command has a built-in Poisson likelihood, but no negative binomial. It is possible to program the negative binomial in various ways using bayesmh. I’ll show here the basic Poisson command and output, and code for the negative binomial. I am writing a fullk book on *Coding Bayesian Models: using Stata, R, R-JAGS, and R-INLA* for Cambridge University Press, which I expect to be published in 2017.

Stata Bayesian Models

POISSON

We'll use diffuse priors on all parameters. The results should be close to the results of a maximum likelihood (or GLM) model on the same data.

Poisson evaluator – call built-in poisson likelihood

```
==========================================
use rwm1984
egen sage = std(age)    /* creates a standardized age variable */
bayesmh docvis outwork sage, likelihood(poisson)  ///
           prior({docvis:}, flat) initial({docvis:} 0)
==========================================

Model summary
------------------------------------------------------------------------------
Likelihood:
  docvis ~ poissonreg(xb_docvis)

Prior:
  {docvis:outwork sage _cons} ~ 1 (flat)                                   (1)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_docvis.

Bayesian Poisson regression                      MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =      3,874
                                                 Acceptance rate  =      .2256
                                                 Efficiency:  min =     .06858
                                                              avg =     .08319
Log marginal likelihood = -15647.075                          max =      .1066

------------------------------------------------------------------------------
             |                                                Equal-tailed
      docvis |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
     outwork |  .4077047   .0193005   .000591   .4070702   .3693081   .4471157
        sage |   .247892   .0090987   .000347   .2478292    .230764   .2658936
       _cons |  .9379984   .0129237   .000474   .9381974    .912948   .9627528
------------------------------------------------------------------------------


. glm docvis outwork sage, family(poisson) nolog nohead

------------------------------------------------------------------------------
             |                 OIM
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     outwork |   .4079314   .0188447    21.65   0.000     .3709965    .4448663
        sage |   .2482285   .0094161    26.36   0.000     .2297733    .2666838
       _cons |   .9380965   .0127571    73.54   0.000      .913093    .9630999
------------------------------------------------------------------------------
```

NEGATIVE BINOMIAL

Again, I'll give the model diffuse or non-informative priors. The results should therefore be somewhat close to a maximum likelihood negative binomial model. Since Stata's *bayesmh* command does not have a built-in evaluator for negative binomial models, we have to create our own. The *bayesmh* command will use it to create the model. Gte evaluator is called *bnbll.ado*

bnbll: Bayesian negative binomial likelihood - llevaluator nbinomialp() 20Sep2015 JMH

```
==============================================================================
program define bnbll
   version 14
   args lnf xb lnalpha
   tempname m
   tempvar p mu lnfj
   scalar `m' = 1/exp(`lnalpha')
   quietly {
      gen double `mu' = exp(`xb') if $MH_touse
      gen double `p' = 1/(1 + exp(`lnalpha') * `mu') if $MH_touse
      gen double `lnfj' = ln(nbinomialp(`m',$MH_y, `p')) if $MH_touse
   }
   summarize `lnfj' if $MH_touse, meanonly
   if r(N) < $MH_n {
          scalar `lnf' = .
          exit
   }
   scalar `lnf' = r(sum)
end
==============================================================================
```

Code to call NB evaluator

```
-----------------------------------------------------------
use rwm1984, clear
egen sage = std(age)
bayesmh docvis outwork sage, ///
       llevaluator(bnbll, parameters({lnalpha})) ///
        prior({docvis:} {lnalpha}, flat)
-----------------------------------------------------------
```

```
Bayesian regression                              MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =      3,874
                                                 Acceptance rate  =      .2349
                                                 Efficiency:  min =     .04396
                                                              avg =       .063
Log marginal likelihood = -8342.7773                          max =     .08043

------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. Dev.     MCSE     Median  [95% Cred. Interval]
-------------+----------------------------------------------------------------
docvis       |
     outwork |  .4151371   .0533644   .001882   .4153228   .3127353   .5243428
        sage |  .2491212    .027545   .001314   .2487239   .1941934   .3023446
       _cons |  .9356769   .0334416   .001233   .9359876   .8704549   .9996466
-------------+----------------------------------------------------------------
     lnalpha |  .8329398   .0309982   .001333   .8331401   .7734565   .8958581
------------------------------------------------------------------------------
```

```
. nbreg docvis outwork sage, nolog

Negative binomial regression                    Number of obs     =      3,874
                                                LR chi2(2)        =     184.89
Dispersion     = mean                           Prob > chi2       =     0.0000
Log likelihood = -8332.7623                     Pseudo R2         =     0.0110

------------------------------------------------------------------------------
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     outwork |   .4146624   .0553158     7.50   0.000     .3062455    .5230793
        sage |   .2489207   .0266265     9.35   0.000     .1967337    .3011077
       _cons |   .9347726   .0334834    27.92   0.000     .8691464    1.000399
-------------+----------------------------------------------------------------
    /lnalpha |   .8316161   .0308983                      .7710565    .8921756
-------------+----------------------------------------------------------------
       alpha |   2.297028   .0709743                      2.162049    2.440433
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 1.5e+04              Prob >= chibar2 = 0.000
```