Modeling Count Data

Cambridge University Press: 2014

Joseph M. Hilbe

hilbe@asu.edu: 7 July, 2014

DESCRIPTION OF DATA FILES : R-Stata-SAS

AZCABGPTCA

azcabgptca.Rdata
azcabgptca.dta ; azcabgptca.sas7bdat

Random subset of the 1991 Arizona Medicare data for patients hospitalized subsequent to undergoing a CABG (DRGs 106, 107) or PTCA (DRG 112) cardiovascular procedure. Prepared by Hilbe (1992) from national Medpar files for use in workshops and for examples in publications

1959 observations on the following 6 variables. (R – data frame)

```
los         hospital length of stay
died        systolic blood pressure of subject
procedure   1=CABG; 0=PTCA
gender      1=male; 0=female
age         age of subject
type        1=emerg/urgent; 0=elective
```

Books in: (CUP = Cambridge University Press)
Hilbe, JM (2014), *Modeling Count Data*, CUP
Hilbe, JM (2011), *Negative Binomial Regression, 2nd ed,* CUP

SOURCE
1991 Arizona Medpar data, cardiovascular patient files, National Health Economics & Research Co.

EXAMPLES - R

```
library(COUNT); data(azcabgptca); attach(azcabgpca)
table(los); table(procedure, type); table(los, procedure)
summary(los)

summary(c91a <- glm(los ~ procedure+ type, family=poisson, data=azcabgptca))
modelfit(c91a)

summary(c91b <- glm(los ~ procedure+ type,
                   family=quasipoisson, data=azcabgptca))
modelfit(c91b)

library(sandwich)
sqrt(diag(vcovHC(c91a, type="HC0")))
```

EXAMPLES - Stata

```
use azcabgptca
bysort procedure: sum los

glm los procedure type, fam(poi) eform nolog
abic

glm los procedure type,  fam(poi) eform vce(robust) nolog nohead
abic

glm los procedure type, fam(nb ml) vce(robust) eform nolog

nbreg los procdure type

ztp los procedure type ,  nolog irr vce(robust)
abic

ztnb los procedure type ,  nolog irr vce(robust)
abic
```

EXAMPLES - SAS

```
/*To call the data from .csv (excel) file*/
proc import OUT= WORK.rwm5yr
            DATAFILE= "D:\sas code from R and stata\rwm5yr.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
run;

/*Select data for year = 1984*/
data rwm1984;
  set rwm5yr;
  where year = 1984;
run;

proc sort data = Medpar;
  by descending type;
run;

/*Robust SEs from page    */
proc genmod data = Medpar order = data;
   class  type Provider_number   ;
   model Length_of_Stay =  type HMO_readmit_ __White  / dist = poi link =
log;
   repeated subject =  Provider_number;
run;
```

AZPROCEDURE

azprocedure.Rdata
azprocedure.dta ; azprocedure.sas7bdat

Data come from the 1991 Arizona cardiovascular patient files. A subset of the fields was selected to model the differential length of stay for patients entering the hospital to receive one of two standard cardiovascular procedures: CABG and PTCA. CABG is the standard acronym for Coronary Artery Bypass Graft, where the flow of blood in a diseased or blocked coronary artery or vein has been grafted to bypass the diseased sections. PTCA, or Percutaneous Transluminal Coronary Angioplasty, is a method of placing a balloon in a blocked coronary artery to open it to blood flow. It is a much less severe method of treatment for those having coronary blockage, with a corresponding reduction in risk.

3589 observations on the following 6 variables. (R – data frame)

```
los          length of hospital stay
procedure    1=CABG;0=PTCA
sex          1=Male; 0=female
admit        1=Urgent/Emerg; 0=elective (type of admission)
age75        1= Age>75; 0=Age<=75
hospital     encrypted facility code (string)
```

Count models use *los* as response variable. 0 counts are structurally excluded

Books in: (CUP = Cambridge University Press)
Hilbe, JM (2014), *Modeling Count Data*, CUP
Hilbe, JM (2011), *Negative Binomial Regression, 2nd ed,* CUP
Hilbe, JM (2009), *Logistic Regression Models*, Chapman & Hall/CRC

SOURCE
1991 Arizona Medpar data, cardiovascular patient files, National Health Economics & Research Co.

EXAMPLES - R

```
library(MASS)
library(COUNT)
data(azprocedure)

summary(glmazp <- glm(los ~ procedure + sex + admit,
                      family=poisson, data=azprocedure))
exp(coef(glmp))

nb2 <- nbinomial(los ~ procedure + sex + admit, data=azprocedure)
summary(nb2)
exp(coef(nb2))

glmaznb <- glm.nb(los ~ procedure + sex + admit, data=azprocedure)
summary(glmaznb)
exp(coef(glmaznb))
```

EXAMPLES - Stata

```
use azprocedure, clear
glm los procedure sex admit, fam(poi) nolog
abic

glm los procedure sex admit, fam(poi) vce(robust) nolog nohead

glm los procedure sex admit, fam(nb ml) vce(robust) nolog eform
abic

nbreg los procedure sex admit, vce(robust) irr
abic

ztnb los procedure sex admit, vce(robust) irr
abic

pigreg los procedure sex admit, nolog  irr vce(robust)
abic
```

EXAMPLES - SAS

```
proc genmod data = azprocedure ;
  model los =  procedure sex admit / dist = poisson;
run;

proc genmod data = azprocedure ;
  model los =  procedure sex admit / dist = nb;
run;
```

FASTTRAKG

```
fasttrakg.Rdata
fasttrakg.dta; fasttrakg.sas7bdat
```

Data are from the Canadian National Cardiovascular Disease registry called, FASTRAK. Years covered at 1996-1998. They have been grouped by covariate patterns from individual observations.

15 observations on the following 9 variables. (R – data frame)

```
died        number died from MI
cases       number of cases with same covariate pattern
anterior    1=anterior site MI; 0=inferior site MI
hcabg       1=history of CABG; 0=no history of CABG
age75       1= Age>75; 0=Age<=75
killip      Killip level of cardiac event severity (1-4)
  kk1(1/0)  non-symptomatic; stress; tightness left shoulder; not MI
  kk2(1/0)  moderate severity cardiac event; angina
  kk3(1/0)  Severe cardiac event; severe chest pains
  kk4(1/0)  Severe cardiac event; death
```

Count models use *died* as response numerator and *cases* as the demominator

Books in: (CUP = Cambridge University Press)
Hilbe, JM (2014), *Modeling Count Data*, CUP
Hilbe, JM (2011), *Negative Binomial Regression, 2nd ed,* CUP
Hilbe, JM (2009), *Logistic Regression Models*, Chapman & Hall/CRC

SOURCE
1996-1998 FASTTRAK data, Hoffman-LaRoche Canada, National Health Economics & Research Co.

EXAMPLES - R

```
library(COUNT); data(fasttrakg)
summary(glmfp <- glm(die ~ anterior + factor(killip) +
offset(log(cases)), family=poisson, data=fasttrakg))
exp(coef(glmfp))

library(COUNT); data(fasttrakg)
lcases <- fasttrakg$cases
nb2 <- nbinomial(die ~ anterior + factor(killip),
                       offset=lcases, data=fasttrakg)
summary(nb2)

summary(glmfnb <- glm.nb(die ~ anterior + factor(killip) +
                       offset(log(cases)), data=fasttrakg))
exp(coef(glmfnb))
```

EXAMPLES – Stata

```
use fasttrakg, clear
glm die anterior i.killip, exposure(cases) fam(poi) eform vce(robust)
abic

nbreg die anterior i.killip, exposure(cases) irr vce(robust)
abic

gpoisson die anterior i.killip, exposure(cases) irr vce(robust)
abic
```

EXAMPLES – SAS

```
/*To create fasttrakg data as given on page    */
data fasttrakg;
  input die cases anterior hcabg killip kk1 kk2 kk3 kk4;
  datalines ;
  5  19   0 0 4 0 0 0 1
  10 83   0 0 3 0 0 1 0
  15 412  0 0 2 0 1 0 0
  28 1864 0 0 1 1 0 0 0
  1  1    0 1 4 0 0 0 1
  0  3    0 1 3 0 0 1 0
  1  18   0 1 2 0 1 0 0
  2  70   0 1 1 1 0 0 0
  10 28   1 0 4 0 0 0 1
  9  139  1 0 3 0 0 1 0
  39 443  1 0 2 0 1 0 0
  50 1374 1 0 1 1 0 0 0
  1  6    1 1 3 0 0 1 0
  3  16   1 1 2 0 1 0 0
  2  27   1 1 1 1 0 0 0
run;

/*data shown on page   */
proc print data= fasttrakg;
run;

/*Create offset variable*/
data fasttrakg;
  set fasttrakg;
  lncase = log(cases);
run;

/*To get the output similar to STATA on page   */
ods output parameterestimates = estimate;
proc genmod data = fasttrakg ;
  model die = anterior hcabg  kk2 kk3 kk4 / dist = poisson
                                            link = log
                                            offset = lncase;
run;
```

FISHING

fishing.Rdata
fishing.dta ; fishing. sas7bdat

The fishing data is adapted from Zuur, Hilbe and Ieno (2013) to determine whether the data appears to be generated from more than one generating mechanism. The data are originally adapted from Bailey et al. (2008) who were interested in how certain deep-sea fish populations were impacted when commercial fishing began in locations with deeper water than in previous years. Given that there are 147 sites that were researched, the model is of (1) the total number of fish counted per site (*totabund*); (2) on the mean water depth per site (*meandepth*); (3) adjusted by the area of the site (*sweptarea*); (4) the log of which is the model offset.

147 observations on the following 3 variables. (R - data frame)

All continuous variables

```
Totabund            total fish counted per site
Meandepth           mean water depth per site
sweptarea           adjusted area of site
density             folage density index
site                catch site
year                1977-2002
period              0=1977-1989; 1=2000+
```

Count models use *totabund* as response variable. Counts start at 2

Books in: (CUP = Cambridge University Press)
Hilbe, Joseph M (2014), *Modeling Count Data*, CUP
Zuur, Hilbe, Ieno (2013), *A Beginner's Guide to GLM and GLMM using R,* Highlands

SOURCE

Bailey M. et al (2008), "Longterm changes in deep-water fish populations in the North East Atlantic", *Proc Roy Soc B,* 275:1965-1969.

EXAMPLES - R

```
library(MASS)
library(COUNT)
library(flexmix)
data(fishing); attach(fishing)
fmm_pg <- flexmix(totabund~meandepth + offset(log(sweptarea)),
         data=rwm1984, k=2,
         model=list(FLXMRglm(totabund~., family="NB1"),
                    FLXMRglm(tpdocvis~., family="NB1")))
parameters(fmm_pg, component=1, model=1)
parameters(fmm_pg, component=2, model=1)
summary(fmm_pg)
```

EXAMPLES – Stata

```
use fishing
sum fishing
nbreg totabund meandepth, exposure(sweptarea) nolog

fmm totabund meandepth, exposure(sweptarea) components(2) mixtureof(negbin2)
```

Note: The Stata *fmm* command was authored by Partha Deb of Hunter College and City University New York (2007).

EXAMPLES - SAS

```
proc fmm data=fishing; /*proc fmm is available in SAS 9.2 or later version*/
  model totabund = meandepth /dist=negbin k = 2 cl;
run;
```

MEDPAR

```
medpar.Rdata
medpar.dta ; medpar.sas7bdat
```

The US national Medicare inpatient hospital database is referred to as the Medpar data, which is prepared yearly from hospital filing records. Medpar files for each state are also prepared. The full Medpar data consists of 115 variables. The national Medpar has some 14 million records, with one record for each hospitalization. The data in the *medpar* file comes from 1991 Medicare files for the state of Arizona. The data are limited to only one diagnostic group (DRG 112). Patient data have been randomly selected from the original data.

1495 observations on the following 10 variables. (R – data frame)

```
los      length of hospital stay
hmo      Patient belongs to Health Maintenance Organization, binary
white    Patient identifies themselves as Caucasian, binary
died     Patient died, binary
age80    Patient age 80 and over, binary
type     Type of admission, categorical
  type1  Elective admission, binary
  type2  Urgent admission,binary
  type3  Elective admission, binary
provnum  Provider ID
```

Count models use los as response variable. 0 counts are structurally excluded

Books in: (CUP = Cambridge University Press)
Hilbe, Joseph M (2014), *Modeling Count Data,* CUP
Hilbe, Joseph M (2007, 2011), *Negative Binomial Regression*, CUP
Hilbe, Joseph M (2009), *Logistic Regression Models*, Chapman & Hall/CRC
Hardin, JW & JM Hilbe (2001, 2007, 2012), *Generalized Linear Models & Extensions*, Stata Press

SOURCE
1991 National Medpar data, National Health Economics & Research Co.

EXAMPLES - R

```
library(COUNT); data(medpar)
summary(glmp <- glm(los ~ hmo + white + factor(type),
                          family=poisson, data=medpar))
exp(coef(glmp))

nb2 <- nbinomial(los ~ hmo + white + factor(type), data=medpar)
summary(nb2)
exp(coef(nb2))
```

EXAMPLES - Stata

```
use medpar, clear
sum los
tab los
glm los hmo white i.type, fam(poi) vce(robust) eform nolog
abic

glm los hmo white i.type, fam(nb ml) vce(robust) eform nolog

nbreg los hmo white i.type, ml) vce(robust) eform nolog
gnbreg(los hmo white type2 type3, lnalpha(hmo white type2 type3)
```

EXAMPLES - SAS

```
proc genmod data = medpar;
  model los =  white hmo / dist = poisson;
run;

proc genmod data = medpar ;
  model los =  white hmo / dist = nb;
run;
```

NUTS

nuts.Rdata
nuts.dta ; nuts.sas7bdat

Squirrel data set (nuts) from Zuur, Hilbe, and Ieno (2013). As originally reported by Flaherty et al (2012), researchers recorded information about squirrel behavior and forest attributes across various plots in Scotland's Abernathy Forest. The study focused on the following variables.

52 observations on the following 5 variables. (R – data frame)

```
cones     number of cones stripped by red squirrels per plot
ntrees    number of trees per plot
dbh       mean diameter of tree
height    mean tree height per plot
cover     percentage of canopy cover per plot
```

The stripped cone count was only taken when the mean diameter of trees was under 0.6m (dbh). ‘s’ prefix to a predictor indicates predictor has been standardized

Count models use *ntrees* as response variable. Counts start at 3

Books in: (CUP = Cambridge University Press)
Hilbe, Joseph M (2014), *Modeling Count Data*, CUP.
Zuur, Hilbe, Ieno (2013), *A Beginner's Guide to GLM and GLMM using R*, Highlands.

SOURCE

Flaherty, S et al (2012), "The impact of forest stand structure on red squirrels habitat use", *Forestry* 85:437-444.

EXAMPLES - R

```
library(COUNT) ; data(nuts)
nut <- subset(nuts, dbh<.6)
sntrees <- scale(nut$ntrees)
sheight <- scale(nut$height)
scover  <- scale(nut$cover)
summary(PO <- glm(cones ~ sntrees + sheight + scover,
                  family=quasipoisson, data=nut))

# heterogeneous negative binomial
nbh <- nbinomial(cones ~ sntrees + sheight + scover,
                  formula2 = sntrees + sheight + scover,
                  mean.link = ”log”,
                  scale.link = “log_s”,
                  data=nut)
summary(nbh)
```

EXAMPLES - Stata

```
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
abic
nbreg cones $xvars if dbh<.6, nolog
abic

gnbreg cones $xvars if dbh<.6, nolog lnalpha($xvars)
abic

gnbreg cones $xvars if dbh<.6, nolog lnalpha($xvars) vce(robust)
```

EXAMPLES - SAS

```
proc genmod data = nuts ;
  sntrees = scale(ntrees);
  sheight = scale(height);
  scover  = scale(cover);
  model cones =  sntrees sheight scover / dist = poisson;
run;

proc genmod data = nuts ;
  sntrees = scale(ntrees);
  sheight = scale(height);
  scover  = scale(cover);
  model cones =  sntrees sheight scover / dist = nb;
run;
```

RWM5YR

`rwm5yr.Rdata`
`rwm5yr.dta ; rwm5yr.sas7bdat`

German health registry for the years 1984-1988. Health information for years immediately prior to health reform.

19,609 observations on the following 17 variables. (R – data frame)

```
Id          patient ID  (1=7028)
Docvis      number of visits to doctor during year (0-121)
Hospvis     number of days in hospital during year (0-51)
Year        year; (categorical: 1984, 1985, 1986, 1987, 1988)
Age         age: 25-64
Outwork     out of work=1; 0=working
Female      female=1; 0=male
Married     married=1; 0=not married
Kids        have children=1; no children=0
Hhninc      household yearly income in marks (in Marks)
Self        self-employed=1; not self employed=0
educ        years of formal education (7-18)
edlevel     educational level (categorical: 1-4)
  edlevel1  (1/0) not high school graduate
  edlevel2  (1/0) high school graduate
  edlevel3  (1/0) university/college
  edlevel4  (1/0) graduate school
```

Count models typically use *docvis* as response variable. 0 counts are included

Books in: (CUP = Cambridge University Press)
Hilbe, Joseph M (2014), *Modeling Count Data*, CUP.
Hilbe, Joseph M (2011), *Negative Binomial Regression*, CUP.
Hilbe, J. and W. Greene (2008). Count Response Regression Models, in ed. C.R. Rao, J.P Miller, and D.C. Rao, *Epidemiology and Medical Statistics, Elsevier Handbook of Statistics* Series. London, UK: Elsevier.

SOURCE
German Health Reform Registry, years pre-reform 1984-1988, in Hilbe and Greene (2007)

EXAMPLES - R

```
library(COUNT); data(rwm5yr)
glmrp <- glm(docvis ~ outwork + female + age + factor(edlevel),
                      family=poisson, data=rwm5yr)
summary(glmrp); exp(coef(glmrp))

nb2 <- nbinomial(docvis ~ outwork + female + age + factor(edlevel),
                          data=rwm5yr)
summary(nb2)
exp(coef(nb2))
```

```
library(pscl); library(COUNT)
data(rwm5yr) ; rwm1984 <- subset(rwm5yr, year==1984)
poi <- glm(docvis ~ outwork + age, data=rwm1984, dist="poisson")
summary(zip <- zeroinfl(docvis ~ outwork + age | outwork + age,
               data=rwm1984, dist="poisson"))
print(vuong(zip,poi))
exp(coef(zinp))

nbh <- nbinomial(docvis ~ outwork + age, data=rwm5yr,
                            formula2 = outwork + age,
                            mean.link = "log", scale.link = "log_s")
summary(nbh)
exp(coef(nbh))

library(gee)
mygee <- gee(docvis ~ outwork + age + factor(edlevel), id=id,
                      corstr(exchangeable),
                      family=poisson,
                      data=rwn5yr)
summary(mygee)
exp(coef(mygee))
```

EXAMPLES - Stata

```
use rwm5yr, clear
keep docvis outwork female age edlevel
glm docvis outwork female age i.edlevel, fam(poi) vce(robust) eform
glm docvis outwork female age i.edlevel, fam(nb ml) vce(robust) eform
nbreg docvis outwork female age i.edlevel, nolog irr vce(robust)

tab edlevel, gen(ed)
global xvars "outwork female age ed2 ed3 ed4"
zinb docvis $xvars, inflate($xvars) vuong zip nolog
abic

zignbreg docvis $xvars, inflate($xvars) lnalpha($xvars) nolog
zignbreg, eform

zipig docvis $xvars, inflate($xvars) irr  vuong zip nolog
abic
```

EXAMPLE - SAS

```
proc genmod data = rwm5yr;
      model docvis = outwork age / dist=negbin;
      ods output ParameterEstimates=pe_nb;
run;
```

RWM1984

rwm1984.Rdata
rwm1984.dta ; rwm1984.sas7bdat

German health registry for the year 1984, the first year of health reform data collection.

3,874 observations on the following 17 variables. (R – data frame)

```
docvis      number of visits to doctor during year (0-121)
hospvis     number of days in hospital during year (0-51)
year        year; (categorical: 1984, 1985, 1986, 1987, 1988)
age         age: 25-64
outwork     out of work=1; 0=working
female      female=1; 0=male
married     married=1; 0=not married
kids        have children=1; no children=0
hhninc      household yearly income in marks (in Marks)
self        self-employed=1; not self employed=0
educ        years of formal education (7-18)
edlevel     educational level (categorical: 1-4)
  edlevel1  (1/0) not high school graduate
  edlevel2  (1/0) high school graduate
  edlevel3  (1/0) university/college
  edlevel4  (1/0) graduate school
```

Count models typically use *docvis* as response variable. 0 counts are included

Books in: (CUP = Cambridge University Press)
Hilbe, Joseph, M (2014), *Modeling Count Data*, CUP
Hilbe, Joseph M (2011), *Negative Binomial Regression*, 2nd ed., CUP
Hilbe, J. and W. Greene (2008). Count Response Regression Models, in ed. C.R. Rao, J.P Miller, and D.C. Rao, *Epidemiology and Medical Statistics*, *Elsevier Handbook of Statistics Series*. London, UK: Elsevier.

SOURCE
German Health Reform Registry, year=1984, in Hilbe and Greene (2007)

EXAMPLES - R

```
library(MASS); library(COUNT); data(rwm1984)

summary(glmrp <- glm(docvis ~ outwork + female + age + factor(edlevel),
                              family=poisson, data=rwm1984))
exp(coef(glmrp))

summary(nbh <- nbinomial(docvis ~ outwork + female + age +
                         factor(edlevel), data=rwm1984,
                         formula2 = outwork + female + age,
                         mean.link = ”log”, scale.link = “log_s”))
exp(coef(nbh))
```

```
summary(glmrnb <- glm.nb(docvis ~ outwork + female + age, data=rwm1984))

library(pscl); library(COUNT); data(rwm5yr)
rwm1984 <- subset(rwm5yr, year==1984)
poi <- glm(docvis ~ outwork + age, data=rwm1984, dist="poisson")
summary(zip <- zeroinfl(docvis ~ outwork + age | outwork + age,
                data=rwm1984, dist="poisson"))
print(vuong(zip,poi))
exp(coef(zinp))
```

EXAMPLES - Stata

```
use rwm5yr, clear
keep if year==1984

*or
use rwm1984, clear
keep docvis outwork female age edlevel
glm docvis outwork female age i.edlevel, fam(poi) vce(robust) eform
glm docvis outwork female age i.edlevel, fam(nb ml) vce(robust) eform
nbreg docvis outwork female age i.edlevel, nolog irr vce(robust)

tab edlevel, gen(ed)
global xvars "outwork female age ed2 ed3 ed4"
zinb docvis $xvars, inflate($xvars) irr vce(robust) vuong zip
abic
```

EXAMPLES - SAS

```
proc genmod data = rwm1984;
      model docvis = outwork age / dist=negbin;
      ods output ParameterEstimates=pe_nb;
run;
```

SMOKING

smoking.Rdata
smoking.dta ; smoking.sas7bdat

A simple artificial data set with only 6 observations.

6 observations on the following 4 variables. (R – data frame)

```
sbp       systolic blood pressure of subject
male      1=male; 0=female
smoker    1=hist of smoking; 0= no hist of smoking
age       age of subject
```

Books in: (CUP = Cambridge University Press)
Hilbe, Joseph M (2014), *Modeling Count Data*, CUP

SOURCE
none

EXAMPLE - R

```
sbp    <- c(131,132,122,119,123,115)
male   <- c(1,1,1,0,0,0)
smoker <- c(1,1,0,0,1,0)
age    <- c(34,36,30,32,26,23)
summary(reg1 <- lm(sbp~ male+smoker+age))
```

EXAMPLE - Stata

```
use smoking, clear
reg sbp male smoker age
glm sbp male smoker age, fam(gau)
predict mug

glm sbp male smoker age, fam(poi)
predict xb, xb
gen mu = exp(xb)
su mup
```

EXAMPLE - SAS

```
data smoking;
  input sbp male smoker age;
  datalines ;
  131  1  1  34
  132  1  1  36
  122  1  0  30
  119  0  0  32
  123  0  1  26
  115  0  0  23
run;
```

```
proc print data= smoking;
run;

ods output parameterestimates = estimate;
proc genmod data = smoking ;
  model sbp = male smoker age / dist = poisson
                                                link = log;
run;
```

TITANIC

```
titanic.Rdata
titanic.dta ; titanic.sas7bdat
```

The data is an observation-based version of the 1912 Titanic passenger survival log,

2201 observations on the following 4 variables. (R – data frame)

```
survive   number of passengers who survived
age       1=adult; 0=child
sex       1=Male; 0=female
class     ticket class 1= 1st class; 2= second class; 3= third class
```

Used to assess risk ratios

Books in: (CUP = Cambridge University Press)
Hilbe, Joseph M (2014), *Modeling Count Data*, CUP
Hilbe, Joseph M (2007, 2011), *Negative Binomial Regression*, CUP
Hilbe, Joseph M (2009), *Logistic Regression Models*, Chapman & Hall/CRC

SOURCE
Found in many other texts

EXAMPLES - R

```
library(COUNT); data(titanic)
summary(glmlr <- glm(survived ~ age + sex + factor(class),
                 family=poisson, data=titanic))
exp(coef(glmlr))

nb2 <- nbinomial(survived ~ age + sex + factor(class), data=titanic)
summary(nb2)
exp(coef(nb2))

?nbinomial   # for more information on use of nbinomial function
```

EXAMPLE - Stata

```
use titanic, clear
glm survived age, fam(poi) nolog nohead vce(robust) ef

gen byte died =survived==0
glm died age, fam(poi) nolog nohead vce(robust) ef
```

EXAMPLE - SAS

```
proc genmod data = titanic ;
  model survived = age sex class2 class3 / dist = poisson
                                           link = log;
run
```

TITANICGRP

titanicgrp.Rdata

titanicgrp.dta ; titanicgrp.sas7bdat

12 observations on the following 5 variables. (R – data frame)

```
survived   number of passengers who survived
cases      number of passengers with same pattern of covariates
age        1=adult; 0=child
sex        gender 1=Male; 0=female
class      ticket class 1= 1st class; 2= second class; 3= third class
```

Used to assess risk ratios

Books in: (CUP = Cambridge University Press)

Hilbe, Joseph M (2014), *Modeling Count Data*, CUP

Hilbe, Joseph M (2007, 2011), *Negative Binomial Regression*, CUP

Hilbe, Joseph M (2009), *Logistic Regression Models*, Chapman & Hall/CRC

SOURCE

Found in many other texts

EXAMPLES - R

```
library(MASS)
library(COUNT)
data(titanicgrp)
glmlr <- glm(survived ~ age + sex + factor(class) +
                          offset(log(cases)),
                          family=poisson,
                          data=titanicgrp)
summary(glmlr)
exp(coef(glmlr))

lcases <- titanicgrp$cases
nb2o <- nbinomial(survived ~ age + sex + factor(class),
                                         formula2 =~ age + sex,
                                         offset = lcases,
                                         mean.link="log",
                                         scale.link="log_s",
                                         data=titanicgrp)
summary(nb2o)
exp(coef(nb2o))
```

EXAMPLE - Stata

```
use titanicgrp, clear
glm survived age i.class, fam(nb ml) exposure(cases) nolog vce(robust) ef
abic
```