Using Poisson Regression for Binary Outcomes

Andy Grogan-Kaylor

11 Oct 2021 08:27:18

Introduction

Logistic regression is the most frequently used model for binary outcomes. Logistic regression provides odds ratios, which while somewhat intuitive, may be misunderstood. Notably, odds ratios overstate the strength of the relationship that is implied by risk ratios (Viera, 2008).

Thus, a number of authors, including Zou (2004), have suggested that Poisson regression, which directly provides risk ratios, can be employed for binary outcomes. Zou (2004) indicates that the standard errors of the Poisson model will need to be adjusted.

This handout draws closely on the outline and presentation of ideas provided by Lindquist (n.d.) at IDRE, although the data source and variables are used are very different.

Get Data

We are using data from the U.S. Census Pulse Surveys

```
. clear all

. use "../data/Andy_June_5.10.21_1pc.dta"
```

Manage Data

```
. recode Anxious6 (0/1 = 0)(2/3 = 1)(. = .), generate(Anxiety6)
(383 differences between Anxious6 and Anxiety6)

. tabulate Anxiety6
```

RECODE of Anxious6 (ANXIOUS)	Freq.	Percent	Cum.
0	376	66.55	66.55
1	189	33.45	100.00
Total	565	100.00	

Logistic Regression

`logit` Command

```
. logit Anxiety6 Sex6 i.Race6 Age6 Income6, or // logistic regression with odds ratios
```

```
Iteration 0:   log likelihood = -339.85845
Iteration 1:   log likelihood = -327.09157
Iteration 2:   log likelihood = -326.88691
Iteration 3:   log likelihood = -326.88668
Iteration 4:   log likelihood = -326.88668

Logistic regression                                     Number of obs =    529
                                                        LR chi2(6)    =  25.94
                                                        Prob > chi2   = 0.0002
Log likelihood = -326.88668                             Pseudo R2     = 0.0382

------------------------------------------------------------------------------
    Anxiety6 | Odds ratio   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        Sex6 |   .6825173   .1366386    -1.91   0.056     .4610041    1.010468
             |
       Race6 |
 Black alone |   1.009843    .362237     0.03   0.978     .4999449    2.039789
 Asian alone |   .3294345   .1654222    -2.21   0.027     .1231252    .8814373
      Other  |   .4120474   .2162551    -1.69   0.091     .1473027    1.152614
             |
        Age6 |   .9891521   .0080552    -1.34   0.180     .9734895    1.005067
     Income6 |     .87352   .0402027    -2.94   0.003     .7981736     .955979
       _cons |   2.001622   .8446585     1.64   0.100     .8753591     4.57697
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.

. est store logit // store estimates
```

`glm` Command

```
. glm Anxiety6 Sex6 i.Race6 Age6 Income6, family(binomial) link(logit)

Iteration 0:   log likelihood = -327.29333
Iteration 1:   log likelihood = -326.88686
Iteration 2:   log likelihood = -326.88668
Iteration 3:   log likelihood = -326.88668

Generalized linear models                         Number of obs   =        529
Optimization     : ML                             Residual df     =        522
                                                  Scale parameter =          1
Deviance         =  653.7733688                   (1/df) Deviance =   1.252439
Pearson          =  529.2756818                   (1/df) Pearson  =   1.013938

Variance function: V(u) = u*(1-u/1)               [Binomial]
Link function    : g(u) = ln(u/(1-u))             [Logit]

                                                  AIC             =   1.262332
Log likelihood   = -326.8866844                   BIC             =  -2619.683

------------------------------------------------------------------------------
             |                 OIM
    Anxiety6 | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        Sex6 |  -.3819675    .200198    -1.91   0.056    -.7743484    .0104135
             |
       Race6 |
 Black alone |   .0097944   .3587065     0.03   0.978    -.6932573    .7128462
 Asian alone |  -1.110378     .50214    -2.21   0.027    -2.094554   -.1262014
      Other  |  -.8866169   .5248305    -1.69   0.091    -1.915266     .142032
             |
        Age6 |  -.0109071   .0081436    -1.34   0.180    -.0268682     .005054
     Income6 |  -.1352242   .0460238    -2.94   0.003    -.2254291   -.0450193
       _cons |   .6939581   .4219869     1.64   0.100    -.1331211    1.521037
------------------------------------------------------------------------------

. est store glm_logit // store estimates
```

Compare `logit` and `glm` Approaches

```
. est table logit glm_logit, b(%9.3f) star // nice table of estimates
```

Variable	logit	glm_logit
Sex6	-0.382	-0.382
Race6		
Black alone	0.010	0.010
Asian alone	-1.110*	-1.110*
Other	-0.887	-0.887
Age6	-0.011	-0.011
Income6	-0.135**	-0.135**
_cons	0.694	0.694

Legend: * p<0.05; ** p<0.01; *** p<0.001

Poisson Regression

poisson Command

```
. poisson Anxiety6 Sex6 i.Race6 Age6 Income6, irr vce(robust)
Iteration 0:   log pseudolikelihood = -366.52369
Iteration 1:   log pseudolikelihood = -366.52156
Iteration 2:   log pseudolikelihood = -366.52156
Poisson regression                                      Number of obs =    529
                                                        Wald chi2(6)  =  24.16
                                                        Prob > chi2   = 0.0005
Log pseudolikelihood = -366.52156                       Pseudo R2     = 0.0229
```

Anxiety6	IRR	Robust std. err.	z	P>|z|	[95% conf.	interval]
Sex6	.7797372	.104779	-1.85	0.064	.599192	1.014683
Race6						
Black alone	1.00453	.2111003	0.02	0.983	.6654021	1.516497
Asian alone	.4401884	.1834773	-1.97	0.049	.1944665	.996397
Other	.5482769	.2225452	-1.48	0.139	.2474559	1.214792
Age6	.9933699	.0048809	-1.35	0.176	.9838495	1.002982
Income6	.9192285	.0254323	-3.04	0.002	.8707096	.9704511
_cons	.7778068	.1849814	-1.06	0.291	.4880174	1.239676

```
Note: _cons estimates baseline incidence rate.

. est store poisson // store estimates
```

glm Command

```
. glm Anxiety6 Sex6 i.Race6 Age6 Income6, link(log) family(poisson) eform vce(robust)
Iteration 0:   log pseudolikelihood = -371.42226
Iteration 1:   log pseudolikelihood = -366.52249
Iteration 2:   log pseudolikelihood = -366.52156
Iteration 3:   log pseudolikelihood = -366.52156
Generalized linear models                         Number of obs   =        529
Optimization     : ML                             Residual df     =        522
                                                  Scale parameter =          1
Deviance         =  371.0431126                   (1/df) Deviance =   .7108106
Pearson          =  347.5824434                   (1/df) Pearson  =   .6658667
Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = ln(u)                   [Log]
```

```
                                                  AIC             =    1.41218
Log pseudolikelihood = -366.5215563               BIC             =  -2902.413
```

Anxiety6	IRR	Robust std. err.	z	P>|z|	[95% conf.	interval]
Sex6	.7797372	.104779	-1.85	0.064	.599192	1.014683
Race6						
Black alone	1.00453	.2111003	0.02	0.983	.6654021	1.516497
Asian alone	.4401884	.1834773	-1.97	0.049	.1944665	.996397
Other	.5482769	.2225452	-1.48	0.139	.2474559	1.214792
Age6	.9933699	.0048809	-1.35	0.176	.9838495	1.002982
Income6	.9192285	.0254323	-3.04	0.002	.8707096	.9704511
_cons	.7778068	.1849814	-1.06	0.291	.4880174	1.239676

```
Note: _cons estimates baseline incidence rate.

. est store glm_poisson // store estimates
```

Compare `poisson` and `glm` Approaches

```
. est table poisson glm_poisson, b(%9.3f) star // nice table of estimates
```

Variable	poisson	glm_poisson
Sex6	-0.249	-0.249
Race6		
Black alone	0.005	0.005
Asian alone	-0.821*	-0.821*
Other	-0.601	-0.601
Age6	-0.007	-0.007
Income6	-0.084**	-0.084**
_cons	-0.251	-0.251

Legend: * p<0.05; ** p<0.01; *** p<0.001

Compare Logistic Regression and Poisson Regression

```
. est table logit glm_logit poisson glm_poisson, b(%9.3f) star // nice table of estimates
```

Variable	logit	glm_logit	poisson	glm_poisson
Sex6	-0.382	-0.382	-0.249	-0.249
Race6				
Black alone	0.010	0.010	0.005	0.005
Asian alone	-1.110*	-1.110*	-0.821*	-0.821*
Other	-0.887	-0.887	-0.601	-0.601
Age6	-0.011	-0.011	-0.007	-0.007
Income6	-0.135**	-0.135**	-0.084**	-0.084**
_cons	0.694	0.694	-0.251	-0.251

Legend: * p<0.05; ** p<0.01; *** p<0.001

Get An Estimate of Risk Change From Logit Models

Re-Run the Logistic Regression Model

```
. logit Anxiety6 Sex6 i.Race6 Age6 Income6, or // re-run our logit model
Iteration 0:   log likelihood = -339.85845
Iteration 1:   log likelihood = -327.09157
Iteration 2:   log likelihood = -326.88691
Iteration 3:   log likelihood = -326.88668
Iteration 4:   log likelihood = -326.88668
Logistic regression                                     Number of obs =    529
                                                        LR chi2(6)    =  25.94
                                                        Prob > chi2   = 0.0002
Log likelihood = -326.88668                             Pseudo R2     = 0.0382
```

Anxiety6	Odds ratio	Std. err.	z	P>|z|	[95% conf.	interval]
Sex6	.6825173	.1366386	-1.91	0.056	.4610041	1.010468
Race6						
Black alone	1.009843	.362237	0.03	0.978	.4999449	2.039789
Asian alone	.3294345	.1654222	-2.21	0.027	.1231252	.8814373
Other	.4120474	.2162551	-1.69	0.091	.1473027	1.152614
Age6	.9891521	.0080552	-1.34	0.180	.9734895	1.005067
Income6	.87352	.0402027	-2.94	0.003	.7981736	.955979
_cons	2.001622	.8446585	1.64	0.100	.8753591	4.57697

```
Note: _cons estimates baseline odds.
```

Estimate Margins

We use the `eydx` option to get a *proportional* change in y for a unit change in x.

```
. margins, eydx(Income6) // proportional change in y for a change in x
Average marginal effects                                   Number of obs = 529
Model VCE: OIM
Expression: Pr(Anxiety6), predict()
ey/dx wrt:  Income6
```

	ey/dx	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
Income6	-.0889566	.0303987	-2.93	0.003	-.1485369	-.0293763

References

Lindquist, K. (n.d.). How Can I Estimate Relative Risk Using Glm For Common Outcomes In Cohort Studies? | Stata FAQ. Retrieved November 10, 2021, from https://stats.idre.ucla.edu/stata/faq/how-can-i-estimate-relative-risk-using-glm-for-common-outcomes-in-cohort-studies/

Viera, A. J. (2008). Odds ratios and risk ratios: What's the difference and why does it matter? *Southern Medical Journal*. https://doi.org/10.1097/SMJ.0b013e31817a7ee4

Zou, G. (2004). A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Journal of Epidemiology*, 159(7), 702–706. https://doi.org/10.1093/aje/kwh090