Interactions in Logistic Regression

Andy Grogan-Kaylor

19 Mar 2020 18:06:00

# The Math

Set to .

Then

# Simulate Some Data

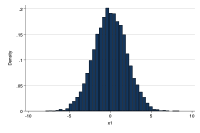
. clear all // empty data

. set obs 10000 // set observations  
number of observations (\_N) was 0, now 10,000

. generate x1 = rnormal(0, 2) // normally distributed

. histogram x1, scheme(michigan)  
(bin=40, start=-7.9101424, width=.40848519)

. graph export myx1graph.png, width(200) replace  
(file myx1graph.png written in PNG format)

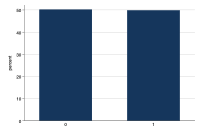


Histogram of x1

. generate x2 = rbinomial(1, .5) // categorical variable

. graph bar, over(x2) scheme(michigan)

. graph export myx2graph.png, width(200) replace  
(file myx2graph.png written in PNG format)



Bar Graph of x2

. summarize // descriptive statistics  
  
 Variable │ Obs Mean Std. Dev. Min Max  
─────────────┼─────────────────────────────────────────────────────────  
 x1 │ 10,000 -.0037079 2.006749 -7.910142 8.429265  
 x2 │ 10,000 .4984 .5000224 0 1

# Story A: Main Effects Only

## Set Up The Data

. generate zA = x1 + x2 // first z

. generate pA = exp(zA) / (1 + exp(zA)) // probabilities

. summarize pA // descriptive statistics  
  
 Variable │ Obs Mean Std. Dev. Min Max  
─────────────┼─────────────────────────────────────────────────────────  
 pA │ 10,000 .5726373 .314671 .0003669 .9999197

. generate yA = rbinomial(1, pA) // generate y with probability p

i.e.

. tab yA // descriptive statistics  
  
 yA │ Freq. Percent Cum.  
────────────┼───────────────────────────────────  
 0 │ 4,240 42.40 42.40  
 1 │ 5,760 57.60 100.00  
────────────┼───────────────────────────────────  
 Total │ 10,000 100.00

## Logistic Regression

. logit yA x1 x2 // does it recover the parameters?  
  
Iteration 0: log likelihood = -6815.5028   
Iteration 1: log likelihood = -4530.3951   
Iteration 2: log likelihood = -4472.0379   
Iteration 3: log likelihood = -4471.0229   
Iteration 4: log likelihood = -4471.0225   
  
Logistic regression Number of obs = 10,000  
 LR chi2(2) = 4688.96  
 Prob > chi2 = 0.0000  
Log likelihood = -4471.0225 Pseudo R2 = 0.3440  
  
─────────────┬────────────────────────────────────────────────────────────────  
 yA │ Coef. Std. Err. z P>|z| [95% Conf. Interval]  
─────────────┼────────────────────────────────────────────────────────────────  
 x1 │ .9982975 .0205714 48.53 0.000 .9579782 1.038617  
 x2 │ .9493676 .0537483 17.66 0.000 .8440228 1.054712  
 \_cons │ .0466175 .0361457 1.29 0.197 -.0242268 .1174618  
─────────────┴────────────────────────────────────────────────────────────────

. predict yhatA // predicted probabilities  
(option pr assumed; Pr(yA))

# Story B: Main Effects + Interactions

## Set Up The Data

. generate zB = x1 + x2 + (.75 \* x1 \* x2) // second z

. generate pB = exp(zB) / (1 + exp(zB)) // probabilities

. summarize pB // descriptive statistics  
  
 Variable │ Obs Mean Std. Dev. Min Max  
─────────────┼─────────────────────────────────────────────────────────  
 pB │ 10,000 .5492112 .3504014 6.87e-06 .9999999

. generate yB = rbinomial(1, pB) // generate y with probability p

i.e.

. tab yB // descriptive statistics  
  
 yB │ Freq. Percent Cum.  
────────────┼───────────────────────────────────  
 0 │ 4,501 45.01 45.01  
 1 │ 5,499 54.99 100.00  
────────────┼───────────────────────────────────  
 Total │ 10,000 100.00

## Logistic Regression

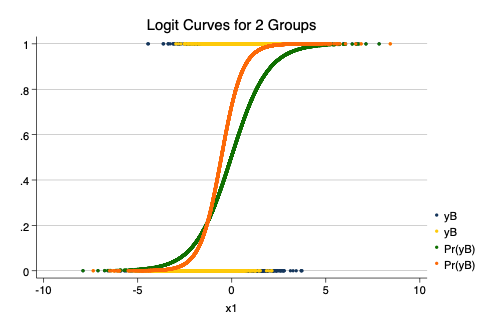
. logit yB c.x1##i.x2 // does it recover the parameters?  
  
Iteration 0: log likelihood = -6881.5886   
Iteration 1: log likelihood = -3889.0951   
Iteration 2: log likelihood = -3807.6537   
Iteration 3: log likelihood = -3803.8245   
Iteration 4: log likelihood = -3803.8112   
Iteration 5: log likelihood = -3803.8112   
  
Logistic regression Number of obs = 10,000  
 LR chi2(3) = 6155.55  
 Prob > chi2 = 0.0000  
Log likelihood = -3803.8112 Pseudo R2 = 0.4472  
  
─────────────┬────────────────────────────────────────────────────────────────  
 yB │ Coef. Std. Err. z P>|z| [95% Conf. Interval]  
─────────────┼────────────────────────────────────────────────────────────────  
 x1 │ 1.033507 .0296165 34.90 0.000 .9754593 1.091554  
 1.x2 │ .9603924 .0633321 15.16 0.000 .8362638 1.084521  
 │  
 x2#c.x1 │  
 1 │ .7527661 .0602815 12.49 0.000 .6346166 .8709155  
 │  
 \_cons │ .0276064 .0365199 0.76 0.450 -.0439713 .0991841  
─────────────┴────────────────────────────────────────────────────────────────

. predict yhatB // predicted probabilities  
(option pr assumed; Pr(yB))

# Inspect The Situation With A Graph

. twoway ///  
> (scatter yB x1 if x2 == 0, msize(tiny)) /// points  
> (scatter yB x1 if x2 == 1, msize(tiny)) /// points  
> (scatter yhatB x1 if x2 == 0, msize(tiny)) ///  
> (scatter yhatB x1 if x2 == 1, msize(tiny)), ///  
> title("Logit Curves for 2 Groups") ///  
> scheme(michigan)

. graph export mygraph.png, width(500) replace  
(file mygraph.png written in PNG format)



Logistic Regression With Interactions