Interactions in Logistic Regression

Andy Grogan-Kaylor

10 Mar 2020 13:53:49

# 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

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

. summarize // descriptive statistics  
  
 Variable │ Obs Mean Std. Dev. Min Max  
─────────────┼─────────────────────────────────────────────────────────  
 x1 │ 10,000 .0001408 1.991975 -7.634315 7.043506  
 x2 │ 10,000 .493 .499976 0 1

# Story A: Main Effects Only

. 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 .5741862 .3129822 .0008352 .9996789

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

. tab yA // descriptive statistics  
  
 yA │ Freq. Percent Cum.  
────────────┼───────────────────────────────────  
 0 │ 4,237 42.37 42.37  
 1 │ 5,763 57.63 100.00  
────────────┼───────────────────────────────────  
 Total │ 10,000 100.00

. logit yA x1 x2 // does it recover the parameters?  
  
Iteration 0: log likelihood = -6814.5818   
Iteration 1: log likelihood = -4484.6166   
Iteration 2: log likelihood = -4433.6528   
Iteration 3: log likelihood = -4432.9688   
Iteration 4: log likelihood = -4432.9686   
  
Logistic regression Number of obs = 10,000  
 LR chi2(2) = 4763.23  
 Prob > chi2 = 0.0000  
Log likelihood = -4432.9686 Pseudo R2 = 0.3495  
  
─────────────┬────────────────────────────────────────────────────────────────  
 yA │ Coef. Std. Err. z P>|z| [95% Conf. Interval]  
─────────────┼────────────────────────────────────────────────────────────────  
 x1 │ 1.02339 .0211541 48.38 0.000 .9819283 1.064851  
 x2 │ .9979633 .0539552 18.50 0.000 .892213 1.103714  
 \_cons │ .0220674 .0364477 0.61 0.545 -.0493687 .0935036  
─────────────┴────────────────────────────────────────────────────────────────

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

# Story B: Main Effects + Interactions

. 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 .5517331 .3477357 4.29e-06 .9999984

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

. tab yB // descriptive statistics  
  
 yB │ Freq. Percent Cum.  
────────────┼───────────────────────────────────  
 0 │ 4,529 45.29 45.29  
 1 │ 5,471 54.71 100.00  
────────────┼───────────────────────────────────  
 Total │ 10,000 100.00

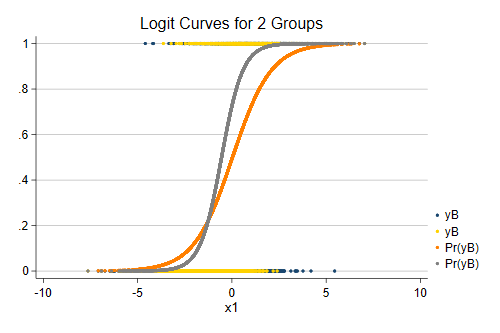
. logit yB c.x1##i.x2 // does it recover the parameters?  
  
Iteration 0: log likelihood = -6887.0378   
Iteration 1: log likelihood = -4019.1014   
Iteration 2: log likelihood = -3938.4941   
Iteration 3: log likelihood = -3934.2638   
Iteration 4: log likelihood = -3934.2528   
Iteration 5: log likelihood = -3934.2528   
  
Logistic regression Number of obs = 10,000  
 LR chi2(3) = 5905.57  
 Prob > chi2 = 0.0000  
Log likelihood = -3934.2528 Pseudo R2 = 0.4287  
  
─────────────┬────────────────────────────────────────────────────────────────  
 yB │ Coef. Std. Err. z P>|z| [95% Conf. Interval]  
─────────────┼────────────────────────────────────────────────────────────────  
 x1 │ .9756118 .0279305 34.93 0.000 .9208691 1.030355  
 1.x2 │ 1.001694 .0614978 16.29 0.000 .8811604 1.122227  
 │  
 x2#c.x1 │  
 1 │ .798564 .0592743 13.47 0.000 .6823885 .9147396  
 │  
 \_cons │ -.0330953 .035931 -0.92 0.357 -.1035187 .0373282  
─────────────┴────────────────────────────────────────────────────────────────

. 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)

. quietly graph export mygraph.png, width(500) replace



Logistic Regression With Interactions