Interactions in Logistic Regression

Andy Grogan-Kaylor

10 Mar 2020 09:56:27

# 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 .0006275 1.989384 -8.936604 8.655005  
 x2 │ 10,000 .496 .500009 0 1

# Story A: Main Effects Only

. generate zA = x1 + x2 // first z

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

. summarize pA // descriptive statistics  
  
 Variable │ Obs Mean Std. Dev. Min Max  
─────────────┼─────────────────────────────────────────────────────────  
 pA │ 10,000 .5741806 .3136176 .0001315 .9999359

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

. tab yA // descriptive statistics  
  
 yA │ Freq. Percent Cum.  
────────────┼───────────────────────────────────  
 0 │ 4,228 42.28 42.28  
 1 │ 5,772 57.72 100.00  
────────────┼───────────────────────────────────  
 Total │ 10,000 100.00

. logit yA x1 x2 // does it recover the parameters?  
  
Iteration 0: log likelihood = -6811.7968   
Iteration 1: log likelihood = -4535.0406   
Iteration 2: log likelihood = -4477.7577   
Iteration 3: log likelihood = -4476.7392   
Iteration 4: log likelihood = -4476.7387   
  
Logistic regression Number of obs = 10,000  
 LR chi2(2) = 4670.12  
 Prob > chi2 = 0.0000  
Log likelihood = -4476.7387 Pseudo R2 = 0.3428  
  
─────────────┬────────────────────────────────────────────────────────────────  
 yA │ Coef. Std. Err. z P>|z| [95% Conf. Interval]  
─────────────┼────────────────────────────────────────────────────────────────  
 x1 │ 1.002792 .0206627 48.53 0.000 .962294 1.04329  
 x2 │ .9381053 .0536428 17.49 0.000 .8329674 1.043243  
 \_cons │ .0508602 .0361359 1.41 0.159 -.0199649 .1216853  
─────────────┴────────────────────────────────────────────────────────────────

. 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 .5514415 .3490153 .0000345 .9999999

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

. tab yB // descriptive statistics  
  
 yB │ Freq. Percent Cum.  
────────────┼───────────────────────────────────  
 0 │ 4,499 44.99 44.99  
 1 │ 5,501 55.01 100.00  
────────────┼───────────────────────────────────  
 Total │ 10,000 100.00

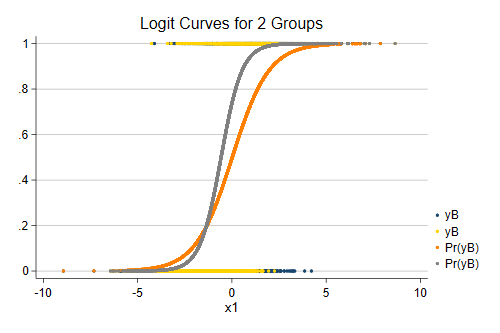
. logit yB c.x1##i.x2 // does it recover the parameters?  
  
Iteration 0: log likelihood = -6881.1873   
Iteration 1: log likelihood = -3908.2789   
Iteration 2: log likelihood = -3827.5789   
Iteration 3: log likelihood = -3823.7445   
Iteration 4: log likelihood = -3823.732   
Iteration 5: log likelihood = -3823.732   
  
Logistic regression Number of obs = 10,000  
 LR chi2(3) = 6114.91  
 Prob > chi2 = 0.0000  
Log likelihood = -3823.732 Pseudo R2 = 0.4443  
  
─────────────┬────────────────────────────────────────────────────────────────  
 yB │ Coef. Std. Err. z P>|z| [95% Conf. Interval]  
─────────────┼────────────────────────────────────────────────────────────────  
 x1 │ 1.028367 .0293532 35.03 0.000 .9708354 1.085898  
 1.x2 │ 1.047891 .0632684 16.56 0.000 .9238873 1.171895  
 │  
 x2#c.x1 │  
 1 │ .7680259 .0605722 12.68 0.000 .6493066 .8867453  
 │  
 \_cons │ -.0214365 .0364055 -0.59 0.556 -.09279 .0499171  
─────────────┴────────────────────────────────────────────────────────────────

. 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