Logistic Regression - More Complicated Margins Plots

Some Ideas for Plotting More Complicated Margins Plots

Andy Grogan-Kaylor

2024-04-29

Table of contents

> 💡 Tip
>
> This tutorial builds off of some of the ideas first presented in this tutorial on interactions.

1 Get Data

```
library(Statamarkdown)
```

We start by obtaining *simulated data* from StataCorp.

```
clear all

use http://www.stata-press.com/data/r15/margex, clear
```

(Artificial data for margins)

2 describe The Data

The variables are as follows:

```
describe
```

Running C:\Users\agrogan\Desktop\GitHub\newstuff\categorical\logistic-plotting-
> margins\profile.do .

Contains data from http://www.stata-press.com/data/r15/margex.dta
 Observations: 3,000   Artificial data for margins
    Variables: 11   27 Nov 2016 14:27

Variable   Storage   Display   Value
    name   type   format   label   Variable label
----
y   float   %6.1f
outcome   byte   %2.0f
sex   byte   %6.0f   sexlbl
group   byte   %2.0f
age   float   %3.0f
distance   float   %6.2f
ycn   float   %6.1f
yc   float   %6.1f
treatment   byte   %2.0f
agegroup   byte   %8.0g   agelab
arm   byte   %8.0g

Sorted by: group

```

# 3 Estimate `logit`

We then run a logistic regression model in which `outcome` is the dependent variable. `sex`, `age` and `group` are the independent variables. We estimate an interaction of `sex` and `age`.

> We note that the regression coefficient for the interaction term is not statistically significant.

```
logit outcome sex##c.age i.group
```

```
Running C:\Users\agrogan\Desktop\GitHub\newstuff\categorical\logistic-plotting-
> margins\profile.do .

Iteration 0:  Log likelihood = -1366.0718
Iteration 1:  Log likelihood =  -1118.129
Iteration 2:  Log likelihood = -1070.8227
Iteration 3:  Log likelihood = -1068.0102
Iteration 4:  Log likelihood =   -1067.99
Iteration 5:  Log likelihood =   -1067.99

Logistic regression                              Number of obs =  3,000
                                                 LR chi2(5)    = 596.16
                                                 Prob > chi2   = 0.0000
Log likelihood = -1067.99                        Pseudo R2     = 0.2182

------------------------------------------------------------------------------
     outcome | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         sex |
      female |    .5565025   .6488407     0.86   0.391    -.7152019    1.828207
         age |    .0910807   .0113215     8.04   0.000     .0688909    .1132704
             |
   sex#c.age |
      female |    -.001211   .0134012    -0.09   0.928    -.0274769     .025055
             |
       group |
          2  |   -.5854237   .1349791    -4.34   0.000    -.8499779   -.3208696
          3  |   -1.355227   .2965301    -4.57   0.000    -1.936416   -.7740391
             |
       _cons |   -5.592272   .5583131   -10.02   0.000    -6.686545   -4.497998
------------------------------------------------------------------------------
```

4 margins

I use the `margins` command to estimate predicted probabilities at different values of `sex` and `age`.

I use the `pwcompare` option to get pairwise comparisons. This gives us **a lot** of output, which I have made *scrollable* to improve readability.

I'm not going to want to graph all of this output, so I'm saving these margins to a data file from which I can make a customized graph.

```
margins sex, at(age = (20 30 40 50 60)) pwcompare saving(marginsdemo.dta, replace)
```

```
Running C:\Users\agrogan\Desktop\GitHub\newstuff\categorical\logistic-plotting-
> margins\profile.do . margins sex, at(age = (20 30 40 50 60)) pwcompare saving(marginsdemo.c
> lace)

Pairwise comparisons of predictive margins                Number of obs = 3,000
Model VCE: OIM

Expression: Pr(outcome), predict()
1._at: age = 20
2._at: age = 30
3._at: age = 40
4._at: age = 50
5._at: age = 60
```

```
------------------------------------------------------------------------------
                         |            Delta-method        Unadjusted
                         |  Contrast    std. err.      [95% conf. interval]
-------------------------+----------------------------------------------------
                 _at#sex |
    (1#female) vs (1#male) |  .0102685    .0072777     -.0039956    .0245325
      (2#male) vs (1#male) |  .0214202    .0029968      .0155466    .0272938
    (2#female) vs (1#male) |   .044561    .0099352      .0250884    .0640335
      (3#male) vs (1#male) |  .0702044    .0066437      .0571829    .0832258
    (3#female) vs (1#male) |  .1179267    .0122683      .0938813    .1419721
      (4#male) vs (1#male) |  .1698721    .0172084      .1361443    .2035999
    (4#female) vs (1#male) |  .2527094    .0173073      .2187878     .286631
      (5#male) vs (1#male) |  .3367733    .0439099      .2507114    .4228352
    (5#female) vs (1#male) |  .4463801    .0328339      .3820268    .5107333
    (2#male) vs (1#female) |  .0111518    .0093316     -.0071379    .0294415
  (2#female) vs (1#female) |  .0342925    .0032639      .0278954    .0406896
```

```
      (3#male) vs (1#female) |   .0599359   .0112873    .0378132    .0820587
    (3#female) vs (1#female) |   .1076582   .0068839    .0941661    .1211504
      (4#male) vs (1#female) |   .1596037   .0172049    .1258827    .1933246
    (4#female) vs (1#female) |    .242441   .0161021    .2108814    .2740005
      (5#male) vs (1#female) |   .3265048    .041177    .2457994    .4072102
    (5#female) vs (1#female) |   .4361116   .0345021    .3684887    .5037345
      (2#female) vs (2#male) |   .0231407   .0115908    .0004231    .0458583
        (3#male) vs (2#male) |   .0487841   .0044858    .0399922    .0575761
      (3#female) vs (2#male) |   .0965065   .0138589    .0693435    .1236694
        (4#male) vs (2#male) |   .1484519    .017062     .115011    .1818928
      (4#female) vs (2#male) |   .2312892   .0188626    .1943192    .2682592
        (5#male) vs (2#male) |    .315353   .0448691    .2274111    .4032949
      (5#female) vs (2#male) |   .4249598   .0339939     .358333    .4915867
      (3#male) vs (2#female) |   .0256434    .013238   -.0003027    .0515895
    (3#female) vs (2#female) |   .0733657   .0045497    .0644484    .0822831
      (4#male) vs (2#female) |   .1253111   .0184164    .0892157    .1614066
    (4#female) vs (2#female) |   .2081485    .015823     .177136    .2391609
      (5#male) vs (2#female) |   .2922123   .0415145    .2108454    .3735792
    (5#female) vs (2#female) |   .4018191   .0353013    .3326299    .4710083
    (3#female) vs (3#male)   |   .0477223   .0153717    .0175943    .0778504
      (4#male) vs (3#male)   |   .0996677   .0137035    .0728094    .1265261
    (4#female) vs (3#male)   |    .182505   .0202228     .142869     .222141
      (5#male) vs (3#male)   |   .2665689   .0424277    .1834121    .3497257
    (5#female) vs (3#male)   |   .3761757   .0349486    .3076778    .4446736
      (4#male) vs (3#female) |   .0519454   .0196499    .0134322    .0904586
    (4#female) vs (3#female) |   .1347827   .0123002    .1106747    .1588908
      (5#male) vs (3#female) |   .2188466   .0414735    .1375599    .3001332
    (5#female) vs (3#female) |   .3284534   .0325423    .2646715    .3922352
      (4#female) vs (4#male) |   .0828373   .0229441    .0378678    .1278069
        (5#male) vs (4#male) |   .1669011    .029177    .1097153     .224087
      (5#female) vs (4#male) |    .276508   .0359297    .2060871    .3469288
      (5#male) vs (4#female) |   .0840638   .0417247    .0022849    .1658428
    (5#female) vs (4#female) |   .1936707   .0205646    .1533647    .2339766
      (5#female) vs (5#male) |   .1096068   .0482732     .014993    .2042206
-------------------------------------------------------------------------------
```

5 `use` Data File With `margins`

Now I'm going to use the data file with margins. It's worth taking a look at.

```
use marginsdemo.dta, clear

describe
```

Running C:\Users\agrogan\Desktop\GitHub\newstuff\categorical\logistic-plotting-
> margins\profile.do .
(Created by command margins; also see char list)

Contains data from marginsdemo.dta
 Observations: 45   Created by command margins;
                                                          also see char list
     Variables: 16   29 Apr 2024 14:35

Variable   Storage   Display   Value
    name   type   format   label   Variable label

_deriv   byte   %9.0g   Derivatives w.r.t.
_term   byte   %9.0g   _term   Margin term index
_predict   byte   %9.0g   predict() option index
_at   byte   %9.0g   Covariates fixed values index
_atopt   byte   %22.0g   _atopt   at() option index
_margin   float   %9.0g   Pr(outcome), predict()
_se_margin   float   %9.0g   Standard error
_statistic   float   %9.0g   z-statistic
_pvalue   float   %9.0g   P>|z|
_ci_lb   float   %9.0g   95% Confidence interval, LB
_ci_ub   float   %9.0g   95% Confidence interval, UB
_m1   byte   %6.0f   sexlbl   sex
_at1   byte   %6.0f   sexlbl   sex
_at2   byte   %3.0f   age
_at3   byte   %2.0f   group
_pw   byte   %14.0g   _pw

Sorted by:

6 Make The Graph!

I'm going to want to graph the `_margin` against values of the `_at2` variable. I also want to graph the confidence interval: `_ci_lb` and `_ci_ub`.

Because there is so much output, I only want to do this for specific values of the **_pw** variable.

```
label list _pw
```

Running C:\Users\agrogan\Desktop\GitHub\newstuff\categorical\logistic-plotting-
> margins\profile.do . label list _pw
_pw:
            1 (1 1) vs (1 0)
            2 (2 0) vs (1 0)
            3 (2 1) vs (1 0)
            4 (3 0) vs (1 0)
            5 (3 1) vs (1 0)
            6 (4 0) vs (1 0)
            7 (4 1) vs (1 0)
            8 (5 0) vs (1 0)
            9 (5 1) vs (1 0)
           10 (2 0) vs (1 1)
           11 (2 1) vs (1 1)
           12 (3 0) vs (1 1)
           13 (3 1) vs (1 1)
           14 (4 0) vs (1 1)
           15 (4 1) vs (1 1)
           16 (5 0) vs (1 1)
           17 (5 1) vs (1 1)
           18 (2 1) vs (2 0)
           19 (3 0) vs (2 0)
           20 (3 1) vs (2 0)
           21 (4 0) vs (2 0)
           22 (4 1) vs (2 0)
           23 (5 0) vs (2 0)
           24 (5 1) vs (2 0)
           25 (3 0) vs (2 1)
           26 (3 1) vs (2 1)
           27 (4 0) vs (2 1)
           28 (4 1) vs (2 1)
           29 (5 0) vs (2 1)
           30 (5 1) vs (2 1)
           31 (3 1) vs (3 0)
           32 (4 0) vs (3 0)
           33 (4 1) vs (3 0)
           34 (5 0) vs (3 0)
           35 (5 1) vs (3 0)
```

```
36  (4 0)  vs  (3 1)
37  (4 1)  vs  (3 1)
38  (5 0)  vs  (3 1)
39  (5 1)  vs  (3 1)
40  (4 1)  vs  (4 0)
41  (5 0)  vs  (4 0)
42  (5 1)  vs  (4 0)
43  (5 0)  vs  (4 1)
44  (5 1)  vs  (4 1)
45  (5 1)  vs  (5 0)
```

> ⚠️ Warning
>
> Remember that `pwcompare` gives us *pairwise* comparisons, i.e the *difference* between the predicted probabilities for the two groups.

```stata
twoway (line _margin _at) /// line graph for margins
(rcap _ci_lb _ci_ub _at, legend(off)) /// range plot w capped spikes for CIs
if _pw == 1 | _pw == 18 | /// long complicated if statement
_pw ==  31 | _pw == 40 | _pw == 45, /// broken into several lines
title("Difference in Predicted Probabilities Between Male and Female") ///
xtitle("age") ///
ytitle("predicted probability")

graph export mypwmarginsplot.png, width(1000) replace
```

```
Running C:\Users\agrogan\Desktop\GitHub\newstuff\categorical\logistic-plotting-
> margins\profile.do .

file mypwmarginsplot.png saved as PNG format
```
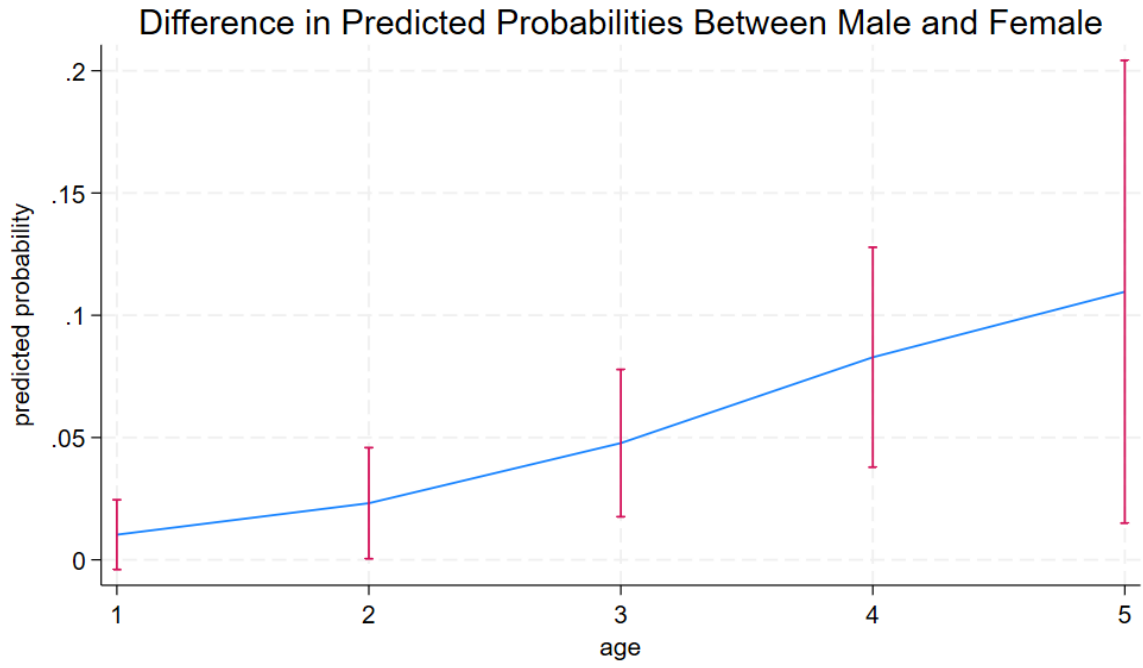
Figure 1: marginsplot of pairwise comparisons