Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Gender Bias Study

Data were collected as part of a University of California at Berkeley study to assess whether men were being given preferential treatment over women in admission to graduate programs (Bickel & O'Connell, 1975, Freedman et al., 1991, pp. 16 - 19). Assuming that the men and women who applied for admission to the graduate programs were equally well-qualified, one would expect equal acceptance rates by gender. However, it initially appeared as if men were being admitted in greater proportions than women. The data are in bias.dta

```
. table nacc female major

----------------------------------------------------------------------------------------------
          |                              major and female
          | ---- 1 ---     ---- 2 ---     ---- 3 ---     ---- 4 ---     ---- 5 ---     ---- 6 ---
     nacc |    0     1        0     1        0     1        0     1        0     1        0     1
----------+-----------------------------------------------------------------------------------
        0 |  512    89      353    17      120   202      138   131       53    94       22    24
        1 |  313    19      207     8      205   391      279   244      138   299      351   317
----------------------------------------------------------------------------------------------

. cc accept male,exact
                                                         Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |      1198         557  |       1755       0.6826
        Controls |      1493        1278  |       2771       0.5388
-----------------+------------------------+------------------------
           Total |      2691        1835  |       4526       0.5946
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |          1.84108       |    1.621141    2.090932 (exact)
 Attr. frac. ex. |         .4568406       |    .3831507    .5217445 (exact)
 Attr. frac. pop |          .311849       |
                 +-------------------------------------------------
                                  1-sided Fisher's exact P = 0.0000
                                  2-sided Fisher's exact P = 0.0000
```

On the basis of this 2x2 table, one might be tempted to conclude that the odds of acceptance for males is estimated to be more than 1.8 times the odds of acceptance for females.

```
. cs accept male,exact

                 | male                   |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+------------
           Cases |      1198         557  |       1755
        Noncases |      1493        1278  |       2771
-----------------+------------------------+------------
           Total |      2691        1835  |       4526
                 |                        |
            Risk |  .4451877    .3035422  |   .3877596
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
 Risk difference |         .1416454       |     .113447    .1698439
      Risk ratio |         1.466642       |     1.35235    1.590592
```

```
                 +-----------------------------------------------
                                  1-sided Fisher's exact P = 0.0000
                                  2-sided Fisher's exact P = 0.0000
```

Notice that Stata gives the 'risk of acceptance'. We would say 'success rate' rather than 'risk'. It would be better to reverse code accept and gender.

```
. cc accept male,by(major)

           major |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |    .349212      .1971148   .5920123     29.85745 (exact)
               2 |   .8025007      .2944774   2.005611     6.015385 (exact)
               3 |    1.13306      .8452664   1.516368     45.10893 (exact)
               4 |   .9212838      .6789529   1.250452     46.14773 (exact)
               5 |   1.221631      .8067614   1.838862     22.21233 (exact)
               6 |   .8278727       .433305   1.575535     11.79832 (exact)
-----------------+-------------------------------------------------
           Crude |    1.84108      1.621141   2.090932              (exact)
    M-H combined |   .9046968      .7719074    1.06033
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(5) =    18.00  Pr>chi2 = 0.0029

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =      1.52
                                                Pr>chi2 =    0.2169
```

The relationship between the odds of acceptance and gender is modified by major.

```
. cs accept male,by(major)

           major |       RR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |     .753095     .6799474   .8341117     78.69775
               2 |    .9269958     .7032172   1.221985      16.2735
               3 |     1.08393     .9045434   1.298892     71.51416
               4 |    .9473337     .7801407   1.150358     68.97348
               5 |    1.160131     .8690445   1.548718     30.74315
               6 |     .838025     .4788977   1.466463     12.53782
-----------------+-------------------------------------------------
           Crude |    1.466642      1.35235   1.590592
    M-H combined |     .944905     .8664522   1.030461
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(5) =   23.288  Pr>chi2 = 0.0003

. cc accept male if major!=1,by(major)

           major |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               2 |   .8025007      .2944774   2.005611     6.015385 (exact)
               3 |    1.13306      .8452664   1.516368     45.10893 (exact)
               4 |   .9212838      .6789529   1.250452     46.14773 (exact)
               5 |   1.221631      .8067614   1.838862     22.21233 (exact)
               6 |   .8278727       .433305   1.575535     11.79832 (exact)
-----------------+-------------------------------------------------
           Crude |   1.563947      1.353682   1.806933              (exact)
    M-H combined |    1.03103      .8701288   1.221685
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(4) =     2.55  Pr>chi2 = 0.6350

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =      0.12
                                                Pr>chi2 =    0.7237
```

The modification appears to be attributable to major==1 compared with the other majors.
We might then argue that, apart from this modification, the relationship is confounded by major.

```
. cs accept male if major!=1,by(major)

           major |       RR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               2 |    .9269958     .7032172   1.221985      16.2735
               3 |     1.08393     .9045434   1.298892     71.51416
               4 |    .9473337     .7801407   1.150358     68.97348
               5 |    1.160131     .8690445   1.548718     30.74315
               6 |     .838025     .4788977   1.466463     12.53782
-----------------+-------------------------------------------------
           Crude |    1.356622     1.230477     1.4957
    M-H combined |    1.020364     .9131851   1.140123
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(4) =    2.688  Pr>chi2 = 0.6113

. cc accept male,by(major1)

          major1 |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |    .349212      .1971148   .5920123     29.85745 (exact)
               2 |   1.563947      1.353682   1.806933     153.6989 (exact)
-----------------+-------------------------------------------------
           Crude |    1.84108      1.621141   2.090932              (exact)
    M-H combined |   1.366357      1.195366   1.561807
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    30.45  Pr>chi2 = 0.0000

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     20.79
                                                Pr>chi2 =    0.0000

. cs accept male,by(major1)

          major1 |       RR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |     .753095     .6799474   .8341117     78.69775
               2 |    1.356622     1.230477     1.4957     243.0526
-----------------+-------------------------------------------------
           Crude |    1.466642      1.35235   1.590592
    M-H combined |    1.209004     1.115439   1.310417
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =   87.801  Pr>chi2 = 0.0000

. cs nacc male,by(major)

           major |       RR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |    2.156555     1.420635   3.273698     16.80064
               2 |    1.155134     .6457439   2.066352      7.65812
               3 |    .9566398      .864462   1.058646     138.4259
               4 |    1.028276     .9301607    1.13674     128.4697
               5 |    .9496577     .8559325   1.053646     97.78938
               6 |    1.012263     .9738279   1.052215     165.6036
-----------------+-------------------------------------------------
           Crude |    .7966202     .7612901     .83359
    M-H combined |    1.027683     .9829126   1.074493
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(5) =   16.993  Pr>chi2 = 0.0045
```

Now we will see if some modelling can enhance this analysis. We will now use non-acceptance and female.

We will have p=pr(not accepted), F=female, M=major, and then the model with major=1 as baseline :

$$\log\left(\frac{p}{1-p}\right)=\beta_0+\beta_1 F+\beta_2 M_2+\beta_3 M_3+\beta_4 M_4+\beta_5 M_5+\beta_6 M_6+\beta_7 FM_2+\beta_8 FM_3+\beta_9 FM_4+\beta_{10} FM_5+\beta_{11} FM_6$$

and so we have

$$\log(\mathrm{OR})=\beta_1+\beta_7 M_2+\beta_8 M_3+\beta_9 M_4+\beta_{10} M_5+\beta_{11} M_6$$

```
. logit nacc i.female##i.major

Logistic regression                             Number of obs     =      4,526
                                                LR chi2(11)       =     877.06
                                                Prob > chi2       =     0.0000
Log likelihood = -2583.6421                     Pseudo R2         =     0.1451

------------------------------------------------------------------------------
        nacc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |  -1.052076   .2627081    -4.00   0.000    -1.566974   -.5371775
             |
       major |
          2  |  -.0416278   .1131892    -0.37   0.713    -.2634746    .1802189
          3  |    1.02764   .1354968     7.58   0.000     .7620707    1.293209
          4  |    1.19608   .1264066     9.46   0.000     .9483272    1.443832
          5  |   1.449083   .1768115     8.20   0.000     1.102539    1.795627
          6  |   3.261867   .2311961    14.11   0.000     2.808731    3.715003
             |
female#major |
        1 2  |   .8320534   .5103948     1.63   0.103     -.168302    1.832409
        1 3  |   1.176998    .299558     3.93   0.000     .5898748     1.76412
        1 4  |   .9700888   .3026187     3.21   0.001     .3769669    1.563211
        1 5  |   1.252263    .330322     3.79   0.000     .6048437    1.899682
        1 6  |   .8631784   .4026666     2.14   0.032     .0739662     1.65239
             |
       _cons |  -.4921214   .0717497    -6.86   0.000    -.6327482   -.3514947
------------------------------------------------------------------------------

. logit nacc i.female##i.major if major!=1

Logistic regression                             Number of obs     =      3,593
Log likelihood = -1985.7974                     Pseudo R2         =     0.1196

------------------------------------------------------------------------------
        nacc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |  -.2200224   .4375926    -0.50   0.615    -1.077688    .6376434
             |
       major |
          3  |   1.069267   .1444824     7.40   0.000     .7860872    1.352448
          4  |   1.237707   .1359941     9.10   0.000     .9711639    1.504251
          5  |   1.490711   .1837881     8.11   0.000     1.130493    1.850929
          6  |   3.303488   .2365737    13.96   0.000     2.839812    3.767164
             |
female#major |
        1 3  |    .344944    .460659     0.75   0.454     -.557931    1.247819
        1 4  |   .1380352   .4626552     0.30   0.765    -.7687524    1.044823
        1 5  |   .4202094   .4812321     0.87   0.383    -.5229883    1.363407
        1 6  |   .0311313   .5334903     0.06   0.953     -1.01449    1.076753
             |
       _cons |  -.5337492    .087543    -6.10   0.000    -.7053304   -.3621681
------------------------------------------------------------------------------
. tab major major1
```

```
           |        major1
     major |         1          2 |     Total
-----------+----------------------+----------
         1 |       933          0 |       933 
         2 |         0        585 |       585 
         3 |         0        918 |       918 
         4 |         0        792 |       792 
         5 |         0        584 |       584 
         6 |         0        714 |       714 
-----------+----------------------+----------
     Total |       933      3,593 |     4,526 


. logit nacc i.female##i.major1

Logistic regression                             Number of obs     =      4,526
                                                LR chi2(3)        =     376.21
                                                Prob > chi2       =     0.0000
Log likelihood = -2834.0656                     Pseudo R2         =     0.0622

--------------------------------------------------------------------------------
          nacc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
      1.female |  -1.052078   .2627083    -4.00   0.000    -1.566977   -.5371797
      2.major1 |   1.034514   .0863319    11.98   0.000     .8653061    1.203721
               |
female#major1 |
          1 2  |   1.499291    .272492     5.50   0.000     .9652167    2.033366
               |
         _cons |  -.4921214   .0717497    -6.86   0.000    -.6327482   -.3514947
--------------------------------------------------------------------------------

. counterplot nacc female major1, measure(or)
```