Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Smoking And Lung Cancer
[adapted from R.D. Cook (1980)]

In 1959, Fisher's [earlier] writings on the subject were reprinted in a booklet entitled "Smoking: The Cancer Controversy. Some Attempts to Assess the Evidence", together with a preface and an article on the inhaling issue. The preface bemoans the mathematical nature of statistical education without regard to the place of statistics in scientific research.

"If, indeed, the statistical departments ... were ... clarifying and confirming, in the future research workers who came within their influence, an understanding of the art of examining observational data, the fallacious conclusions drawn, from a simple association, could scarcely have been made the basis of a terrifying propaganda."

"There is nothing to stop those who greatly desire it from believing that lung cancer is caused by smoking cigarettes. They should also believe that inhaling cigarette smoke is a protection. To believe either is, however, to run the risk of failing to recognize and therefore failing to prevent other and more genuine causes..."

"No particular importance need be attached to the test of significance. It disproves at about the 1 percent level the hypothesis that inhalers and non-inhalers have the same cancer incidence. Even equality would be a fair knock-out for the theory that smoke in the lung causes cancer ..."

"Should not these workers have let the world know, not only that they have discovered the cause of lung cancer (cigarettes), but also that they had discovered the means of its prevention (inhaling cigarette smoke)? How had the M.R.C. the heart to withhold this information from the thousands who would otherwise die of lung cancer?"

"Those who refuse the jump from association to causation in the case of cigarette smoking will not be tempted to take it in the case of inhaling; but the M.R.C. and its Statistical Research Unit think the argument is valid in the first case. Can they refuse to admit it in the second?"

```
. use inhale.dta

. bysort male:table cont ninh cig

-------------------------------------------------------------------------------
-> male = 0

-------------------------------------------------------------------------
          |                    Cigarette Group and Inhaling
          | --- 1-4 --    -- 5-14 --    -- 15-24 -    -- 25-49 -    --- >49 --
   Cancer |    Y     N       Y     N       Y     N       Y     N       Y     N
----------+--------------------------------------------------------------------
     Case |    3     3       7     8       7     5       5     3
  Control |    2    10       2     7       6                         1
-------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> male = 1

-------------------------------------------------------------------------
          |                    Cigarette Group and Inhaling
          | --- 1-4 --    -- 5-14 --    -- 15-24 -    -- 25-49 -    --- >49 --
   Cancer |    Y     N       Y     N       Y     N       Y     N       Y     N
----------+--------------------------------------------------------------------
     Case |    7    17     141    67     133    63      96    78      21    24
  Control |   17    21     162    80     157    44      74    44      16     7
-------------------------------------------------------------------------
```

Fisher analyzes only the data for men.
We will use logistic regression to [essentially] reproduce Fisher's analysis.
Outcome : Cancer : Case Control
Exposure : Inhaling : Y N
Potential Confounder/Modifier : Cigarette Group : 5 levels

```
. logit case inh##i.cig if male==1

Logistic regression                             Number of obs     =      1,269
                                                LR chi2(9)        =      35.21
                                                Prob > chi2       =     0.0001
Log likelihood = -861.75455                     Pseudo R2         =     0.0200

------------------------------------------------------------------------------
        case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.inh |   -.675994   .5550885    -1.22   0.223    -1.763948    .4119594
             |
         cig |
        5-14 |   .0339751   .3658797     0.09   0.926    -.6831359    .7510861
       15-24 |   .5702542   .3808449     1.50   0.134    -.1761882    1.316697
       25-49 |   .7838283   .3768161     2.08   0.038     .0452823    1.522374
         >49 |   1.443447   .5394125     2.68   0.007      .386218    2.500676
             |
     inh#cig |
      1#5-14 |   .7144916   .5906045     1.21   0.226    -.4430719    1.872055
     1#15-24 |   .1511523   .6005096     0.25   0.801    -1.025825    1.328129
     1#25-49 |   .3637579   .6063013     0.60   0.549    -.8245707    1.552087
       1#>49 |  -.2842102   .7763798    -0.37   0.714    -1.805887    1.237466
             |
       _cons |  -.2113091   .3262554    -0.65   0.517    -.8507579    .4281397
------------------------------------------------------------------------------

. est stor A
```

```
. logit case inh i.cig if male==1

Logistic regression                               Number of obs     =      1,269
                                                  LR chi2(5)        =      29.41
                                                  Prob > chi2       =     0.0000
Log likelihood = -864.65245                       Pseudo R2         =     0.0167

------------------------------------------------------------------------------
        case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inh |  -.2971684   .1212979    -2.45   0.014     -.534908   -.0594289
             |
         cig |
       5-14  |   .3950898   .2803963     1.41   0.159    -.1544767    .9446564
      15-24  |   .5389397   .2834785     1.90   0.057    -.0166679    1.094547
      25-49  |   .9104857   .2887322     3.15   0.002     .3445809     1.47639
        >49  |   1.183409   .3673053     3.22   0.001     .4635039    1.903314
             |
       _cons |  -.3469659   .2652004    -1.31   0.191     -.866749    .1728173
------------------------------------------------------------------------------

. est stor B

. lrtest A B

Likelihood-ratio test                                 LR chi2(4)  =      5.80
(Assumption: B nested in A)                           Prob > chi2 =    0.2149


. logit case inh if male==1

Logistic regression                               Number of obs     =      1,269
                                                  LR chi2(1)        =       6.79
                                                  Prob > chi2       =     0.0092
Log likelihood = -875.96427                       Pseudo R2         =     0.0039

------------------------------------------------------------------------------
        case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inh |  -.3073255    .118229    -2.60   0.009      -.53905    -.075601
       _cons |   .2393382   .0954888     2.51   0.012     .0521836    .4264927
------------------------------------------------------------------------------

. est stor C

. lrtest B C

Likelihood-ratio test                                 LR chi2(4)  =     22.62
(Assumption: C nested in B)                           Prob > chi2 =    0.0002
```

There is no evidence that cigarette group modifies [p=0.2149]. There is no indication that cigarette group confounds [-0.2971684 compared with -0.3073255]. There is evidence that cigarette group is associated with cancer [p=0.0002]. So Model B looks fine [on the basis of this modeling!]. With either Model or Model C, we see that inhaling appears to be associated with cancer BUT the direction is crazy! See Fisher's remarks above.