Green-comments **black-commands** **blue-output**

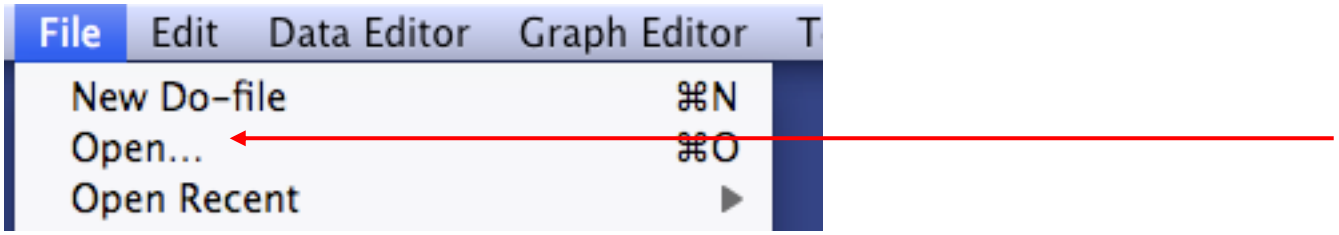
Stata/SE 11.1 File Edit Data Editor Graph Editor Tools Data Graphics Statistics User Window Help

From top menu bar - -
Create a log of your session by clicking on **FILE > LOG > BEGIN**
Format the log file as a "stata log". At File Format box: Choose the option **STATA LOG**

```
. *
. * Turn of the pausing of output
. set more off

. *
. * In this illustration we will using existing stata data sets.  Use FILE > OPEN
```

From top menu bar –
Read an existing stata data set by clicking on **FILE > OPEN**
You'll have to browse the folders on ***your*** computer to access the data set.

Stata/SE 11.1 File Edit Data Editor Graph Editor Tools Data Graphics Statistics User Window Help




```
. *
. * Input into memory the Ille-et-Villaine data set
. * This a study of the relationship between alcohol use and esophageal cancer.
. use "/Users/carolbigelow/Desktop/esophageal_cancer.dta"


. *
. * Look at data
. * Note that data are in tabular form with frequency counts stored in TALLY.
. codebook, compact

Variable   Obs Unique  Mean  Min  Max  Label
------------------------------------------------------------------------------------------
alcohol      8      4   2.5    1    4  Alcohol g per day
case         8      2    .5    0    1
tally       10      7  23.4    2   92
```

Green-comments **black-commands** **blue-output**

```
. *
. * Descriptives using TAB.  NOTE - Since data are tabular, must use [FWEIGHT = ].
. tab case alcohol [fweight=tally]

           |              Alcohol g per day
      case |     0-39g      40-79    80-119g       120+ |     Total
-----------+--------------------------------------------+----------
   control |        47         31          9          5 |        92
      case |         2          9          9          5 |        25
-----------+--------------------------------------------+----------
     Total |        49         40         18         10 |       117

. *
. * TAB with option COLUMN yields column percentages.
. tab case alcohol [fweight=tally], column

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

           |              Alcohol g per day
      case |     0-39g      40-79    80-119g       120+ |     Total
-----------+--------------------------------------------+----------
   control |        47         31          9          5 |        92
           |     95.92      77.50      50.00      50.00 |     78.63
-----------+--------------------------------------------+----------
      case |         2          9          9          5 |        25
           |      4.08      22.50      50.00      50.00 |     21.37
-----------+--------------------------------------------+----------
     Total |        49         40         18         10 |       117
           |    100.00     100.00     100.00     100.00 |    100.00
```

Green-comments **black-commands** **blue-output**

```
. *
. * TAB with option EXACT or CHI2 yields simple tests of association.
. tab case alcohol [fweight=tally], exact

Enumerating sample-space combinations:
stage 4:  enumerations = 1
stage 3:  enumerations = 9
stage 2:  enumerations = 97
stage 1:  enumerations = 0

           |              Alcohol g per day
      case |     0-39g      40-79    80-119g       120+ |     Total
-----------+--------------------------------------------+----------
   control |        47         31          9          5 |        92
      case |         2          9          9          5 |        25
-----------+--------------------------------------------+----------
     Total |        49         40         18         10 |       117

           Fisher's exact =                 0.000

. tab case alcohol [fweight=tally], chi2

           |              Alcohol g per day
      case |     0-39g      40-79    80-119g       120+ |     Total
-----------+--------------------------------------------+----------
   control |        47         31          9          5 |        92
      case |         2          9          9          5 |        25
-----------+--------------------------------------------+----------
     Total |        49         40         18         10 |       117

          Pearson chi2(3) =  22.4068   Pr = 0.000
```

Green-comments **black-commands** **blue-output**

```
. *-----------------------------------------------------*
. *  One Sample Proportion
. *-----------------------------------------------------*
. * Illustration using subset for whom alcohol=2
. keep if alcohol==2
(8 observations deleted)

. tab case [fweight=tally]

       case |      Freq.     Percent        Cum.
------------+-----------------------------------
    control |         31       77.50       77.50
       case |          9       22.50      100.00
------------+-----------------------------------
      Total |         40      100.00

. *
. * Exact Confidence Interval for Single Proportion
. * CCI command is an example of an "immediate" command – you supply the counts
. cii 40 9, level(90)

                                                    -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [90% Conf. Interval]
-------------+---------------------------------------------------------------
             |         40        .225    .0660256        .1227117    .3597924


. *
. * Exact Binomial Test for Single Proportion
. * BITESTI totaln eventcount nullp
. * Eg – n=40, x=9 null pi = .25
. bitesti 40 9 .25

        N   Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
       40          9           10       0.25000      0.22500

  Pr(k >= 9)            = 0.700168  (one-sided test)
  Pr(k <= 9)            = 0.439540  (one-sided test)
  Pr(k <= 9 or k >= 11) = 0.855636  (two-sided test)
```

Green-comments **black-commands** **blue-output**

```
. *
. * Normal Approximation Test for Single Proportion
. * BE CAREFUL!!  Here you must supply observed proportion instead of count
. * PRTESTI totaln observedp nullp
. * Eg - n=40, xbar=.225  null pi = .25
. prtesti 40 .225 .25

One-sample test of proportion                      x: Number of obs =       40
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.                     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |       .225   .0660256                      .0955923    .3544077
------------------------------------------------------------------------------
    p = proportion(x)                                           z =  -0.3651
Ho: p = 0.25

    Ha: p < 0.25                 Ha: p != 0.25                 Ha: p > 0.25
 Pr(Z < z) = 0.3575         Pr(|Z| > |z|) = 0.7150          Pr(Z > z) = 0.6425


. *-----------------------------------------------------*
. *  Comparison of Two Proportions
. *-----------------------------------------------------*
. * Illustration using 2 groups defined by alcohol=1 or alcohol=2
. clear
. use "/Users/carolbigelow/Desktop/esophageal_cancer.dta"
. keep if alcohol==1 | alcohol==2
(6 observations deleted)

. tab case alcohol [fweight=tally], column
+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

           |   Alcohol g per day
      case |     0-39g      40-79 |     Total
-----------+----------------------+----------
   control |        47         31 |        78
           |     95.92      77.50 |     87.64
-----------+----------------------+----------
      case |         2          9 |        11
           |      4.08      22.50 |     12.36
-----------+----------------------+----------
     Total |        49         40 |        89
           |    100.00     100.00 |    100.00
```

Green-comments **black-commands** **blue-output**

```
. * Normal Approximation Test of Equality of Two Proportions ╦ Using counts
. * PRTESTI N1 #EVENTS1 N2 #EVENTS2, COUNT
. prtesti 49 2 40 9, count

Two-sample test of proportion                     x: Number of obs =       49
                                                  y: Number of obs =       40
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .0408163   .0282664                     -.0145847    .0962174
           y |       .225   .0660256                      .0955923    .3544077
-------------+----------------------------------------------------------------
        diff |  -.1841837   .0718217                     -.3249517   -.0434157
             |  under Ho:   .0701327    -2.63   0.009
------------------------------------------------------------------------------
        diff = prop(x) - prop(y)                                      z =  -2.6262
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.0043         Pr(|Z| < |z|) = 0.0086          Pr(Z > z) = 0.9957

* Normal Approximation Test of Equality of Two Proportions ╦ Using Proportions
. * PRTESTI N1 P1 N2 P2
. prtesti 49 .041 40 .225

Two-sample test of proportion                     x: Number of obs =       49
                                                  y: Number of obs =       40
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |       .041   .0283272                     -.0145202    .0965202
           y |       .225   .0660256                      .0955923    .3544077
-------------+----------------------------------------------------------------
        diff |      -.184   .0718457                     -.324815    -.043185
             |  under Ho:   .0701573    -2.62   0.009
------------------------------------------------------------------------------
        diff = prop(x) - prop(y)                                      z =  -2.6227
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.0044         Pr(|Z| < |z|) = 0.0087          Pr(Z > z) = 0.9956
```

Green-comments **black-commands** **blue-output**

```
. *--------------------------------------------------
. * Single 2x2 Table
. * COHORT Study Design
. *--------------------------------------------------

. * PRELIMINARY Need to create 0/1 exposure variable for high alcohol consumption
. generate alcohol_high=(alcohol==2)
. label define highf 0 "0-39" 1 "40-79"
. label values alcohol_high highf
. tab alcohol_high [fweight=tally]

alcohol_hig |
          h |      Freq.     Percent        Cum.
------------+-----------------------------------
       0-39 |         49       55.06       55.06
      40-79 |         40       44.94      100.00
------------+-----------------------------------
      Total |         89      100.00

. * Estimation of Relative Risk Statistics and Fisher's Exact Test
. * CS command with option EXACT
. cs case alcohol_high [fweight=tally], exact

                | alcohol_high           |
                |   Exposed   Unexposed  |      Total
----------------+------------------------+-----------
          Cases |         9           2  |         11
       Noncases |        31          47  |         78
----------------+------------------------+-----------
          Total |        40          49  |         89
                |                        |
           Risk |      .225    .0408163  |   .1235955
                |                        |
                |      Point estimate    |    [95% Conf. Interval]
                |------------------------+------------------------
 Risk difference |         .1841837       |    .0434157    .3249517
      Risk ratio |           5.5125       |    1.262212    24.07491
 Attr. frac. ex. |         .8185941       |    .2077403     .958463
 Attr. frac. pop |         .6697588       |
                +-------------------------------------------------
                                  1-sided Fisher's exact P = 0.0100
                                  2-sided Fisher's exact P = 0.0108
```

Green-comments **black-commands** **blue-output**

```
. * Estimation of Relative Risk Statistics and Fisher's Exact Test: Immediate version
. * CSI casesexposed casesnot NONcasesexposed NONcasesnot
. csi 9 2 31 47, exact

                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+-----------
           Cases |         9           2  |         11
        Noncases |        31          47  |         78
-----------------+------------------------+-----------
           Total |        40          49  |         89
                 |                        |
            Risk |      .225    .0408163  |   .1235955
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
 Risk difference |         .1841837       |    .0434157    .3249517
      Risk ratio |           5.5125       |    1.262212    24.07491
 Attr. frac. ex. |         .8185941       |    .2077403     .958463
 Attr. frac. pop |         .6697588       |
                 +-------------------------------------------------
                                  1-sided Fisher's exact P = 0.0100
                                  2-sided Fisher's exact P = 0.0108
. *-------------------------------------------------------
. * Single 2x2 Table
. * CASE-CONTROL Study Design
. *-------------------------------------------------------
. * NOTE:  We already have 0/1 exposure variable for high alcohol consumption
. * Estimation of Odds Ratio Statistics and Fisher's Exact Test
. * CC command with option EXACT
. cc case alcohol_high [fweight=tally], exact
                                                        Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |         9           2  |         11      0.8182
        Controls |        31          47  |         78      0.3974
-----------------+------------------------+------------------------
           Total |        40          49  |         89      0.4494
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         6.822581       |    1.263628    67.65054 (exact)
 Attr. frac. ex. |         .8534279       |    .2086276    .9852182 (exact)
 Attr. frac. pop |         .6982592       |
                 +-------------------------------------------------
                                  1-sided Fisher's exact P = 0.0100
                                  2-sided Fisher's exact P = 0.0108
```

Green-comments **black-commands** **blue-output**

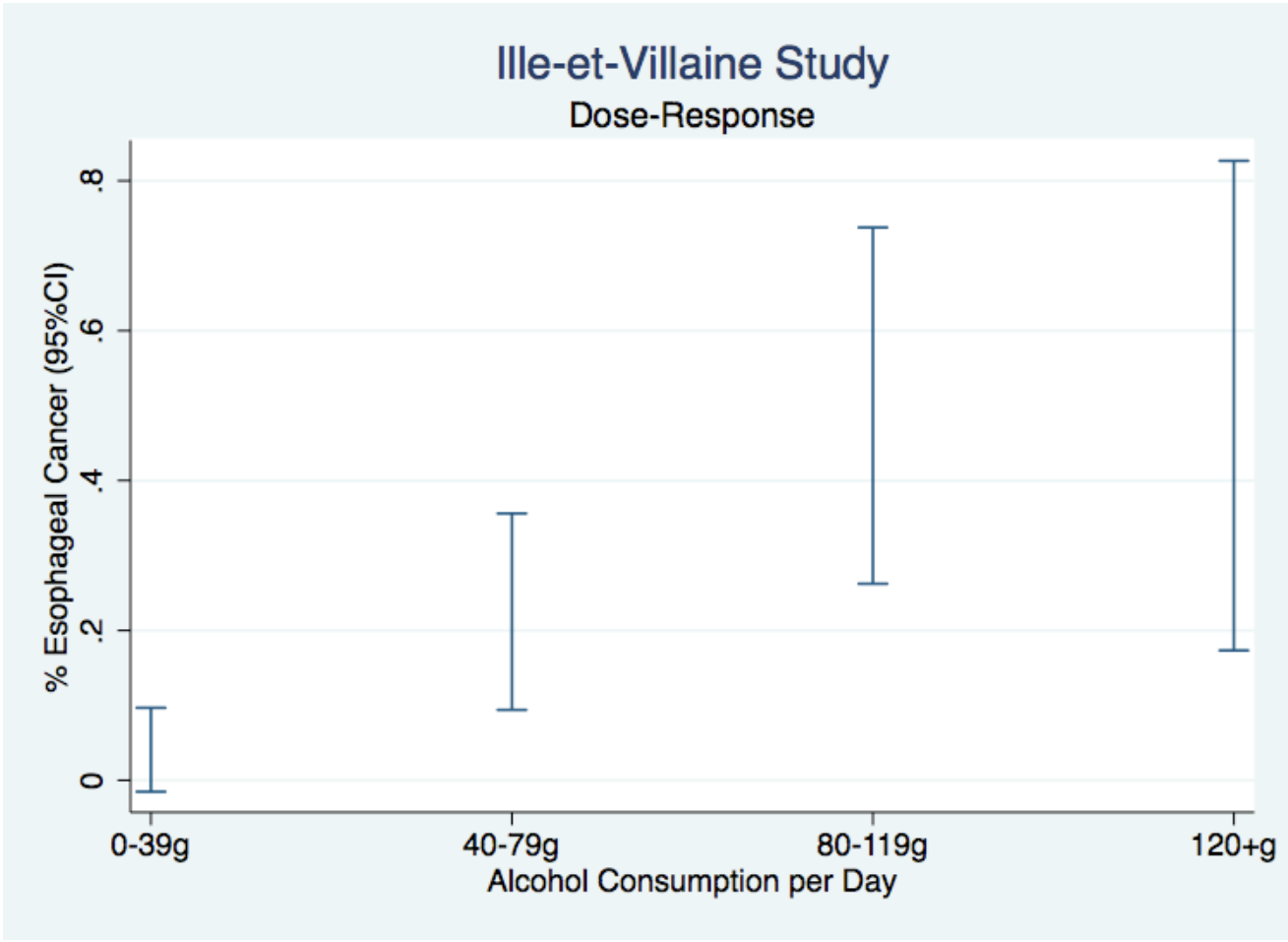
```
. * Estimation of Odds Ratio Statistics and Fisher's Exact Test: Immediate version
. * CCI casesexposed casesnot controlsesexposed controlssnot
. cci 9 2 31 47, exact
                                                        Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |         9           2  |         11      0.8182
        Controls |        31          47  |         78      0.3974
-----------------+------------------------+------------------------
           Total |        40          49  |         89      0.4494
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         6.822581       |    1.263628    67.65054 (exact)
 Attr. frac. ex. |         .8534279       |    .2086276    .9852182 (exact)
 Attr. frac. pop |         .6982592       |
                 +-----------------------------------------------
                                 1-sided Fisher's exact P = 0.0100
                                 2-sided Fisher's exact P = 0.0108

. *-------------------------------------------------------
. * R x C Table - Test of Trend
. *-------------------------------------------------
. * Illustration uses entire data set
. clear
. use "/Users/carolbigelow/Desktop/esophageal_cancer.dta"


. * TAB with option COLUMN to obtain descriptives
. tab case alcohol [fweight=tally], column

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

           |              Alcohol g per day
      case |     0-39g      40-79    80-119g       120+ |     Total
-----------+--------------------------------------------+----------
   control |        47         31          9          5 |        92
           |     95.92      77.50      50.00      50.00 |     78.63
-----------+--------------------------------------------+----------
      case |         2          9          9          5 |        25
           |      4.08      22.50      50.00      50.00 |     21.37
-----------+--------------------------------------------+----------
     Total |        49         40         18         10 |       117
           |    100.00     100.00     100.00     100.00 |    100.00
```

Green-comments **black-commands** **blue-output**

```
. * Test of Linear Trend in Event Probability Associated w Increasing Dose
. * TABODDS outcomevariable dosevariable
. tabodds case alcohol [fweight=tally]

------------------------------------------------------------------------------
   alcohol |      cases     controls       odds      [95% Conf. Interval]
-----------+------------------------------------------------------------------
      0-39g |          2           47    0.04255        0.01034   0.17518
      40-79 |          9           31    0.29032        0.13822   0.60979
    80-119g |          9            9    1.00000        0.39695   2.51919
       120+ |          5            5    1.00000        0.28950   3.45420
------------------------------------------------------------------------------
Test of homogeneity (equal odds): chi2(3)  =    22.22
                                  Pr>chi2  =   0.0001

Score test for trend of odds:     chi2(1)  =    20.85
                                  Pr>chi2  =   0.0000



. * Graph of % with event and 95% CI
. * Preliminary - Use COLLAPSE to create data set with means and se values*
. clear
. use "/Users/carolbigelow/Desktop/esophageal_cancer.dta"

. * COLLAPSE to obtain MEAN and SEM of the variable CASE separately for each DOSE
. collapse (mean) case (sem) semcase=case [fweight=tally], by(alcohol)
. list
     +------------------------------+
     | alcohol       case    semcase |
     |------------------------------|
  1. |   0-39g   .0408163   .0285593 |
  2. |   40-79       .225   .0668667 |
  3. | 80-119g         .5   .1212678 |
  4. |    120+         .5   .1666667 |
     +------------------------------+


. * Preliminary: Create Upper and Lower CI limit Values
. generate low=case-1.96*semcase
. generate high=case+1.96*semcase
```

Green-comments **black-commands** **blue-output**

```
.* Plot using TWOWAY RCAP.  Save result using extension png
. twoway rcap low high alcohol, xlabel(1 "0-39g" 2 "40-79g" 3"80-119g" 4 "120+g") xtitle("Alcohol
Consumption per Day") ytitle("% Esophageal Cancer (95%CI)") msize(*2) title("Ille-et-Villaine
Study") subtitle("Dose-Response")
```

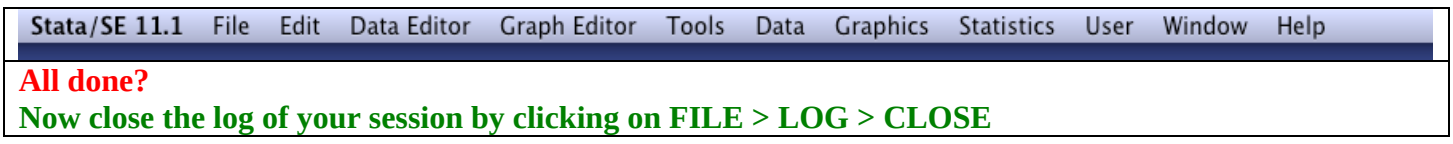


Green-comments **black-commands** **blue-output**

```
. *-------------------------------------------------
. * Stratified Analysis of K 2x2 Tables
. *-------------------------------------
. * Illustration uses Video Diplay Terminal Exposure and SAB Data
. clear
. use "/Users/carolbigelow/Desktop/K_2x2tables.dta"
. label variable sab "SAB: Spontaneous Abortion"
. label variable gest "GEST: Month of Gestation"
. label define casef 0 "0=no" 1 "1=yes"
. label values sab casef

. * TABLE to obtain descriptives.  Descriptives on event of SAB by Month of Gestation
. table sab gest [fweight=tally], row col
----------------------------------------------------------------------
SAB:      |
Spontaneo |
us        |                GEST: Month of Gestation
Abortion  |     1      2      3      4      5      6      7  Total
----------+-----------------------------------------------------------
     0=no |   867    799    770    760    754    749    746  5,445
    1=yes |    11     68     27     12      6      5      3    132
          |
    Total |   878    867    797    772    760    754    749  5,577
----------------------------------------------------------------------

. *---- Mantel Haenszel Test of Homogeneity of OR Across Strata ---*
. * Command CC eventvariable exposurevariable with option BY(stratavariable)
. cc sab expos [fweight=tally], by(gest)

GEST: Month of G |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |   .1375342      .0031607   .9759646     4.157175 (exact)
               2 |    1.09348      .6401297   1.852996     14.68281 (exact)
               3 |   1.107121      .4664182   2.571603     6.079046 (exact)
               4 |   .9928122      .2461515   3.671101      2.88342 (exact)
               5 |   2.802548      .3983408   31.11545     .8263158 (exact)
               6 |   .3482428      .0070492   3.544543     1.660477 (exact)
               7 |   .6955128      .0117518   13.42084     .8331108 (exact)
-----------------+-------------------------------------------------
           Crude |   .9610324      .6637615   1.383478              (exact)
    M-H combined |   .9540899      .6682187    1.36226
----------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(6) =     6.24  Pr>chi2 = 0.3966

                  Test that combined OR = 1:
                              Mantel-Haenszel chi2(1) =      0.07
                                             Pr>chi2 =    0.7961
```

Green-comments **black-commands** **blue-output**

```
. * You can get the Woolf test by using this option in the CC command
. cc sab expos [fweight=tally], by(gest) woolf

GEST: Month of G |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |   .1375342      .0175285   1.079137     4.157175 (Woolf)
               2 |    1.09348      .6639694   1.800834     14.68281 (Woolf)
               3 |   1.107121      .5113482   2.397029     6.079046 (Woolf)
               4 |   .9928122      .3122768   3.156418      2.88342 (Woolf)
               5 |   2.802548      .5101686   15.39545     .8263158 (Woolf)
               6 |   .3482428      .0387367     3.1307     1.660477 (Woolf)
               7 |   .6955128      .0627896   7.704112     .8331108 (Woolf)
-----------------+-------------------------------------------------
           Crude |   .9610324      .6764691     1.3653              (Woolf)
    M-H combined |   .9540899      .6682187    1.36226
---------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(6) =     6.24  Pr>chi2 = 0.3966

                  Test that combined OR = 1:
                              Mantel-Haenszel chi2(1) =      0.07
                                                Pr>chi2 =    0.7961

. *---- Test that COMBINED OR = 1 assuming homogeneity  ---*
. * It's the same stata command.
. cc sab expos [fweight=tally], by(gest)

GEST: Month of G |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               1 |   .1375342      .0031607   .9759646     4.157175 (exact)
               2 |    1.09348      .6401297   1.852996     14.68281 (exact)
               3 |   1.107121      .4664182   2.571603     6.079046 (exact)
               4 |   .9928122      .2461515   3.671101      2.88342 (exact)
               5 |   2.802548      .3983408   31.11545     .8263158 (exact)
               6 |   .3482428      .0070492   3.544543     1.660477 (exact)
               7 |   .6955128      .0117518   13.42084     .8331108 (exact)
-----------------+-------------------------------------------------
           Crude |   .9610324      .6637615   1.383478              (exact)
    M-H combined |   .9540899      .6682187    1.36226
---------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(6) =     6.24  Pr>chi2 = 0.3966
                  Test that combined OR = 1:
                              Mantel-Haenszel chi2(1) =      0.07
                                                Pr>chi2 =    0.7961
```



All done?
Now close the log of your session by clicking on FILE > LOG > CLOSE