Unit 8 – Stata for Categorical Data Analysis

Homework

Solutions

Input data. Check.

```
. set more off
. * -----  Input data.

. use "/Users/cbigelow/Desktop/hw8_categorical.dta"
. codebook, compact

Variable     Obs Unique   Mean  Min  Max  Label
-------------------------------------------------------------------------------------------------
--
cigs        1000      7  1.997    1    7  Cigarettes per day
coffee      1000      2   .909    0    1  coffee per day
micase      1000      2   .855    0    1  heart attack case
-------------------------------------------------------------------------------------------------
--

. label list
casef:
           0 control
           1 MI case
coffeef:
           0 <5 cups
           1 >= 5 cups
cigsf:
           1 1=never smoked
           2 2=former smoker
           3 3=1-14 cigs/day
           4 4=15-24 cigs/day
           5 5=25-34 cigs/day
           6 6=35-44 cigs/day
           7 7=45+ cigs/day
```

__1. Produce a simple tabulation of the distribution of cigarettes per day (**cigs**). Do this 3 ways, so that you can decide which command you like best!

a) Using the command **tab1**

b) Using the command **tabulate**

c) Using the command **fre**

```
. *-----  1a)
. tab1 cigs
-> tabulation of cigs

  Cigarettes per |
             day |      Freq.     Percent        Cum.
-----------------+-----------------------------------
  1=never smoked |        728       72.80       72.80
 2=former smoker |         23        2.30       75.10
 3=1-14 cigs/day |         18        1.80       76.90
4=15-24 cigs/day |         94        9.40       86.30
5=25-34 cigs/day |         59        5.90       92.20
6=35-44 cigs/day |         48        4.80       97.00
  7=45+ cigs/day |         30        3.00      100.00
```

```
-----------------+-----------------------------------
           Total |      1,000      100.00

. *-----  1b)
. tabulate cigs

  Cigarettes per |
             day |      Freq.     Percent        Cum.
-----------------+-----------------------------------
  1=never smoked |         728       72.80       72.80
 2=former smoker |          23        2.30       75.10
 3=1-14 cigs/day |          18        1.80       76.90
4=15-24 cigs/day |          94        9.40       86.30
5=25-34 cigs/day |          59        5.90       92.20
6=35-44 cigs/day |          48        4.80       97.00
  7=45+ cigs/day |          30        3.00      100.00
-----------------+-----------------------------------
           Total |      1,000      100.00

. *-----1c)
. fre cigs

cigs -- Cigarettes per day
-------------------------------------------------------------------------------
                            |      Freq.    Percent      Valid       Cum.
----------------------------+--------------------------------------------------
Valid   1 1=never smoked    |         728      72.80      72.80      72.80
        2 2=former smoker   |          23       2.30       2.30      75.10
        3 3=1-14 cigs/day   |          18       1.80       1.80      76.90
        4 4=15-24 cigs/day  |          94       9.40       9.40      86.30
        5 5=25-34 cigs/day  |          59       5.90       5.90      92.20
        6 6=35-44 cigs/day  |          48       4.80       4.80      97.00
        7 7=45+ cigs/day    |          30       3.00       3.00     100.00
        Total               |        1000     100.00     100.00
-------------------------------------------------------------------------------
```

__2. Produce a cross-tabulation of coffee consumption (**coffee**) as the row variable and heart attack (**micase**) as the column variable using the command **tab2** . Do this 2 ways

a) with row percentages

b) with column percentages

```
. *-----  2a)
. tab2 coffee micase, row

-> tabulation of coffee by micase

+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+

coffee per |   heart attack case
       day |   control    MI case |     Total
-----------+----------------------+----------
   <5 cups |        46         45 |        91
           |     50.55      49.45 |    100.00
-----------+----------------------+----------
 >= 5 cups |        99        810 |       909
           |     10.89      89.11 |    100.00
-----------+----------------------+----------
     Total |       145        855 |     1,000
           |     14.50      85.50 |    100.00


. *-----  2b)
. tab2 coffee micase, column

-> tabulation of coffee by micase

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

coffee per |   heart attack case
       day |   control    MI case |     Total
-----------+----------------------+----------
   <5 cups |        46         45 |        91
           |     31.72       5.26 |      9.10
-----------+----------------------+----------
 >= 5 cups |        99        810 |       909
           |     68.28      94.74 |     90.90
-----------+----------------------+----------
     Total |       145        855 |     1,000
           |    100.00     100.00 |    100.00
```

__3. Produce a table showing the proportion in each category of cigarette smoking **(cigs)**, together with 95% confidence limits using the command **proportion.**

```
. proportion cigs, level(95)

Proportion estimation             Number of obs    =      1,000

      _prop_1: cigs = 1=never smoked
      _prop_2: cigs = 2=former smoker
      _prop_3: cigs = 3=1-14 cigs/day
      _prop_4: cigs = 4=15-24 cigs/day
      _prop_5: cigs = 5=25-34 cigs/day
      _prop_6: cigs = 6=35-44 cigs/day
      _prop_7: cigs = 7=45+ cigs/day

--------------------------------------------------------------
             | Proportion   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
cigs         |
     _prop_1 |       .728   .0140789      .6995125    .7547337
     _prop_2 |       .023   .0047427      .0153199    .0343958
     _prop_3 |       .018   .0042064      .0113603    .0284088
     _prop_4 |       .094   .0092331      .0773798    .1137499
     _prop_5 |       .059   .0074548      .0459615    .0754448
     _prop_6 |       .048   .0067633      .0363406     .063155
     _prop_7 |        .03   .0053971      .0210403     .042609
--------------------------------------------------------------
```

__4. Produce a Fisher Exact test of the association of high coffee consumption (**coffee**) with heart attack (**micase)**.

Do this 3 ways, so that you can decide which command you like best!

a) Using the command **tab2** with option **exact**

b) Using the command **cs**

c) Using the command **cc**

```
. *-----  4a)
. tab2 coffee micase, exact

-> tabulation of coffee by micase

coffee per |   heart attack case
       day |   control    MI case |     Total
-----------+----------------------+----------
   <5 cups |        46         45 |        91
 >= 5 cups |        99        810 |       909
-----------+----------------------+----------
     Total |       145        855 |     1,000

          Fisher's exact =                 0.000
   1-sided Fisher's exact =                 0.000
```

```
. *-----  4b)
. * cs diseasevariable exposurevariable, exact
. cs micase coffee, exact

                 | coffee per day          |
                 |   Exposed   Unexposed   |      Total
-----------------+-------------------------+------------
           Cases |       810          45   |        855
        Noncases |        99          46   |        145
-----------------+-------------------------+------------
           Total |       909          91   |       1000
                 |                         |
            Risk |  .8910891    .4945055   |       .855
                 |                         |
                 |      Point estimate     |    [95% Conf. Interval]
                 |-------------------------+------------------------
 Risk difference |         .3965836        |    .2918825    .5012847
      Risk ratio |          1.80198        |    1.462162    2.220775
 Attr. frac. ex. |         .4450549        |    .3160813    .5497067
 Attr. frac. pop |          .421631        |
                 +-------------------------------------------------
                                 1-sided Fisher's exact P = 0.0000
                                 2-sided Fisher's exact P = 0.0000


. *-----  4c)
. * cc diseasevariable exposurevariable, exact
. cc micase coffee, exact
                                                        Proportion
                 |   Exposed   Unexposed   |      Total     Exposed
-----------------+-------------------------+------------------------
           Cases |       810          45   |        855       0.9474
        Controls |        99          46   |        145       0.6828
-----------------+-------------------------+------------------------
           Total |       909          91   |       1000       0.9090
                 |                         |
                 |      Point estimate     |    [95% Conf. Interval]
                 |-------------------------+------------------------
      Odds ratio |         8.363636        |    5.125199     13.5931 (exact)
 Attr. frac. ex. |         .8804348        |    .8048856    .9264333 (exact)
 Attr. frac. pop |         .8340961        |
                 +-------------------------------------------------
                                 1-sided Fisher's exact P = 0.0000
                                 2-sided Fisher's exact P = 0.0000
```

Interpretation:

The Fisher Exact test of association of high coffee consumption (exposure) with event of myocardial infarction (case) is statistically significant (Fisher Exact Test p-value < .0001). The assumption of the null hypothesis of no association, when applied to the observed data, has led to an extremely unlikely event (less than 1 chance in 10,000). The null hypothesis is rejected. Conclude that, in these data, there is statistically significant evidence of an association of high coffee consumption with occurrence of MI.