Table of Contents

Lecture 7: live coding

open data set

```
. webuse nhanes2, clear
```

Stored results

most stata commands store results so they can subsequently be used by other commands or programs.

commands are r-class, e-class or n-class
* r-class, like summarize, type "return list" after
* n-class, like generate, do not store results
* e-class, like logistic, type "ereturn list" after

```
. sum age
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         age |     10,351    47.57965    17.21483         20         74

. return list
scalars:
                  r(N) =  10351
              r(sum_w) =  10351
               r(mean) =  47.57965413969665
                r(Var) =  296.3503454262628
                 r(sd) =  17.21482923023818
                r(min) =  20
                r(max) =  74
                r(sum) =  492497

. display "Standard Error = " r(sd)/sqrt(r(N))
Standard Error = .16920437

. display "Standard Error = " %9.2f (r(sd)/sqrt(r(N)))
Standard Error =      0.17

. tab sex heartatk
           | heart attack, 1=yes,
    1=male, |         0=no
  2=female |         0          1 |     Total
-----------+----------------------+----------
      Male |     4,597        318 |     4,915
    Female |     5,276        158 |     5,434
-----------+----------------------+----------
     Total |     9,873        476 |    10,349

. numlabel, add
. tab sex heartatk
           | heart attack, 1=yes,
    1=male, |         0=no
  2=female |         0          1 |     Total
-----------+----------------------+----------
    1. Male |     4,597        318 |     4,915
  2. Female |     5,276        158 |     5,434
-----------+----------------------+----------
     Total |     9,873        476 |    10,349

. recode sex 2=0
(sex: 5436 changes made)

. label define sex 0 "Female" 1 "Male", replace
. label values sex sex
. tab sex heartatk, row chi2
+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+

           | heart attack, 1=yes,
    1=male, |         0=no
  2=female |         0          1 |     Total
-----------+----------------------+----------
    Female |     5,276        158 |     5,434
           |     97.09       2.91 |    100.00
-----------+----------------------+----------
```

```
      Male |     4,597        318 |     4,915
           |     93.53       6.47 |    100.00
-----------+----------------------+----------
     Total |     9,873        476 |    10,349
           |     95.40       4.60 |    100.00

          Pearson chi2(1) =  74.6387   Pr = 0.000
```

. return list

```
scalars:
                  r(N) =  10349
                  r(r) =  2
                  r(c) =  2
               r(chi2) =  74.63865049638308
                  r(p) =  5.65256590335e-18
```

. display "p-value = " (r(p))

```
p-value = 5.653e-18
```

. display "p-value = " %9.5f (r(p))

```
p-value =   0.00000
```

. cs heartatk sex, or

```
                    | 1=male, 2=female       |
                    |   Exposed   Unexposed  |      Total
--------------------+------------------------+-----------
              Cases |       318         158  |        476
           Noncases |      4597        5276  |       9873
--------------------+------------------------+-----------
              Total |      4915        5434  |      10349
                    |                        |
               Risk |  .0646999    .0290762  |   .0459948
                    |                        |
                    |      Point estimate    |    [95% Conf. Interval]
                    |------------------------+------------------------
 Risk difference    |          .0356237      |    .0274229    .0438245
      Risk ratio    |          2.225185      |    1.845984    2.682282
 Attr. frac. ex.    |          .5505992      |    .4582835    .6271831
 Attr. frac. pop    |          .3678373      |
      Odds ratio    |          2.309938      |    1.901548    2.806027 (Cornfield)
                    +-------------------------------------------------
                               chi2(1) =    74.64  Pr>chi2 = 0.0000
```

. logit heartatk sex

```
Iteration 0:   log likelihood = -1930.5936
Iteration 1:   log likelihood = -1894.1179
Iteration 2:   log likelihood =  -1892.826
Iteration 3:   log likelihood = -1892.8244
Iteration 4:   log likelihood = -1892.8244

Logistic regression                               Number of obs     =     10,349
                                                  LR chi2(1)        =      75.54
                                                  Prob > chi2       =     0.0000
Log likelihood = -1892.8244                       Pseudo R2         =     0.0196

------------------------------------------------------------------------------
    heartatk |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .8372213   .0994024     8.42   0.000     .6423961    1.032047
       _cons |  -3.508329   .0807382   -43.45   0.000    -3.666573   -3.350085
------------------------------------------------------------------------------
```

. ereturn list

```
scalars:
               e(rank) =  2
                  e(N) =  10349
                 e(ic) =  4
                  e(k) =  2
               e(k_eq) =  1
               e(k_dv) =  1
          e(converged) =  1
                 e(rc) =  0
                 e(ll) =  -1892.824440853722
          e(k_eq_model) =  1
               e(ll_0) =  -1930.593646037925
               e(df_m) =  1
               e(chi2) =  75.53841036840686
                  e(p) =  3.58367259982e-18
              e(N_cdf) =  0
              e(N_cds) =  0
               e(r2_p) =  .0195635188490937

macros:
             e(cmdline) : "logit heartatk sex"
                 e(cmd) : "logit"
            e(estat_cmd) : "logit_estat"
             e(predict) : "logit_p"
       e(marginsnotok) : "stdp DBeta DEviance DX2 DDeviance Hat Number Residuals RStandard SCore"
               e(title) : "Logistic regression"
            e(chi2type) : "LR"
                 e(opt) : "moptimize"
                 e(vce) : "oim"
                e(user) : "mopt__logit_d2()"
           e(ml_method) : "d2"
           e(technique) : "nr"
               e(which) : "max"
              e(depvar) : "heartatk"
          e(properties) : "b V"

matrices:
                  e(b) :  1 x 2
                  e(V) :  2 x 2
```

```
              e(mns) :  1 x 2
            e(rules) :  1 x 4
             e(ilog) :  1 x 20
         e(gradient) :  1 x 2

functions:
            e(sample)
```

```
. display "Odds Ratio = " exp(_b[sex])
Odds Ratio = 2.3099395

. display "Odds Ratio = " %9.2f exp(_b[sex]) //format
Odds Ratio =      2.31

. logit heartatk sex age
Iteration 0:   log likelihood = -1930.5936
Iteration 1:   log likelihood = -1699.6524
Iteration 2:   log likelihood = -1628.8241
Iteration 3:   log likelihood = -1626.8074
Iteration 4:   log likelihood = -1626.8003
Iteration 5:   log likelihood = -1626.8003

Logistic regression                             Number of obs     =     10,349
                                                LR chi2(2)        =     607.59
                                                Prob > chi2       =     0.0000
Log likelihood = -1626.8003                     Pseudo R2         =     0.1574

------------------------------------------------------------------------------
    heartatk |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .9106814   .1019417     8.93   0.000     .7108794    1.110483
         age |   .0855818   .0048761    17.55   0.000     .0760248    .0951388
       _cons |  -8.410505   .3217014   -26.14   0.000    -9.041028   -7.779982
------------------------------------------------------------------------------

. display "Odds Ratio (adjusted for age) = " exp(_b[sex])
Odds Ratio (adjusted for age) = 2.4860159

. logit heartatk sex age, or
Iteration 0:   log likelihood = -1930.5936
Iteration 1:   log likelihood = -1699.6524
Iteration 2:   log likelihood = -1628.8241
Iteration 3:   log likelihood = -1626.8074
Iteration 4:   log likelihood = -1626.8003
Iteration 5:   log likelihood = -1626.8003

Logistic regression                             Number of obs     =     10,349
                                                LR chi2(2)        =     607.59
                                                Prob > chi2       =     0.0000
Log likelihood = -1626.8003                     Pseudo R2         =     0.1574

------------------------------------------------------------------------------
    heartatk | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   2.486016   .2534286     8.93   0.000     2.035781    3.035826
         age |   1.089351   .0053118    17.55   0.000     1.078989    1.099811
       _cons |   .0002225   .0000716   -26.14   0.000     .0001184     .000418
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.

. binreg heartatk sex age
Iteration 1:   deviance =  4089.425
Iteration 2:   deviance =  3470.318
Iteration 3:   deviance =   3292.04
Iteration 4:   deviance =   3256.14
Iteration 5:   deviance =  3253.619
Iteration 6:   deviance =  3253.601
Iteration 7:   deviance =  3253.601
Iteration 8:   deviance =  3253.601

Generalized linear models                         No. of obs      =     10,349
Optimization     : MQL Fisher scoring             Residual df     =     10,346
                   (IRLS EIM)                     Scale parameter =          1
Deviance         =  3253.600517                   (1/df) Deviance =   .3144791
Pearson          =  8285.286495                   (1/df) Pearson  =   .8008203

Variance function: V(u) = u*(1-u/1)               [Binomial]
Link function    : g(u) = ln(u/(1-u))             [Logit]

                                                  BIC             =   -92391.5

------------------------------------------------------------------------------
             |                 EIM
    heartatk |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .9106814   .1019417     8.93   0.000     .7108794    1.110483
         age |   .0855818   .0048761    17.55   0.000     .0760248    .0951388
       _cons |  -8.410505   .3217014   -26.14   0.000    -9.041028   -7.779982
------------------------------------------------------------------------------

. display "Risk Ratio (adjusted for age) = " exp(_b[sex])
Risk Ratio (adjusted for age) = 2.4860159

. binreg heartatk sex age, rr
Iteration 1:   deviance =   5386.93
Iteration 2:   deviance =  3676.121
Iteration 3:   deviance =   3324.74
```

```
Iteration 4:   deviance =  3264.518
Iteration 5:   deviance =  3259.263
Iteration 6:   deviance =  3259.197
Iteration 7:   deviance =  3259.197
Iteration 8:   deviance =  3259.197

Generalized linear models                         No. of obs      =     10,349
Optimization     : MQL Fisher scoring             Residual df     =     10,346
                   (IRLS EIM)                     Scale parameter =          1
Deviance         =  3259.196513                   (1/df) Deviance =     .31502
Pearson          =  8163.608177                   (1/df) Pearson  =   .7890594

Variance function: V(u) = u*(1-u/1)               [Binomial]
Link function    : g(u) = ln(u)                   [Log]

                                                  BIC             =   -92385.9

------------------------------------------------------------------------------
             |                 EIM
    heartatk | Risk Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   2.276113   .2122632     8.82   0.000     1.895892    2.732589
         age |   1.082165   .0048289    17.70   0.000     1.072742    1.091671
       _cons |   .0003176   .0000939   -27.23   0.000     .0001779    .0005671
------------------------------------------------------------------------------
Note: _cons estimates baseline risk.
```

Macros

A macro is simply a name associated with some text or a number. It can hold a stored/returned value after running a command or any text/number your choose.

1. locals
2. globals
3. scalars

Locals

Local macros have names of up to 31 characters and are known only in the current context (the console, a do file, or a program).

You define a local macro using local name [=] text and you evaluate it using `name'. (Note the use of a backtick or left quote.)

The second type of macro definition with an equal sign is used to store results. It instructs Stata to treat the text on the right hand side as an expression, evaluate it, and store a text representation of the result under the given name.

```
. regress bpsystol age
      Source |       SS           df       MS      Number of obs   =    10,351
-------------+----------------------------------   F(1, 10349)     =   3116.79
       Model |  1304200.02         1  1304200.02   Prob > F        =    0.0000
    Residual |  4330470.01    10,349  418.443328   R-squared       =    0.2315
-------------+----------------------------------   Adj R-squared   =    0.2314
       Total |  5634670.03    10,350  544.412563   Root MSE        =    20.456

------------------------------------------------------------------------------
    bpsystol |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .6520775   .0116801    55.83   0.000     .6291823    .6749727
       _cons |   99.85603   .5909867   168.96   0.000     98.69758    101.0145
------------------------------------------------------------------------------

. local rsq1 e(r2)
. local rsq2 = e(r2)
. di `rsq1'     // this has the current R-squared
   .23145988

. di `rsq1'     // as does this
   .23145988

. regress bpsystol age weight
      Source |       SS           df       MS      Number of obs   =    10,351
-------------+----------------------------------   F(2, 10348)     =   2250.00
       Model |  1707705.53         2  853852.766   Prob > F        =    0.0000
    Residual |  3926964.49    10,348  379.490191   R-squared       =    0.3031
-------------+----------------------------------   Adj R-squared   =    0.3029
       Total |  5634670.03    10,350  544.412563   Root MSE        =    19.481

------------------------------------------------------------------------------
    bpsystol |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .6379892   .0111315    57.31   0.000     .6161692    .6598091
      weight |   .4069041   .0124786    32.61   0.000     .3824435    .4313646
       _cons |   71.27096   1.041742    68.42   0.000     69.22894    73.31297
------------------------------------------------------------------------------
```

```
. di `rsq1'      // the formula has the new R-squared
   .30307108

. di `rsq2'      // this has the old R-squared
   .23145988
```

Globals

Global macros have names of up to 32 characters and are visible everywhere in Stata (not only in your .do file)

You define a global macro using global name [=] text and evaluate it using $name.

Most programmers prefer not to use globals because you can accidently write over the name of another function/program used in Stata.

You call globals with the $ sign

```
. global myName "Kristen"
. di "$myName"
   Kristen
```

Scalars

Scalars store numbers with more precision. Scalars are stroed in the global space, so you have to be careful not to overwrite program=/function names

```
. sum age
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         age |     10,351    47.57965    17.21483         20         74

. scalar SE =  r(sd)/ sqrt(r(N))
. di SE
   .16920437

. scalar MEAN = r(mean)
. di MEAN
   47.579654

. scalar CIL = MEAN - 1.96*SE
. scalar CIU = MEAN +1.96*SE
. di "95% CI:" CIL "-" CIU
   95% CI:47.248014-47.911295
```

Matrices

To Stata, a matrix is a named entity containing an r × c rectangular array of double-precision numbers (including missing values) that is bordered by a row and a column of names.

```
. tabstat zinc copper vitaminc iron, by(diabetes) stat(n mean sd) save
   Summary statistics: N, mean, sd
     by categories of: diabetes (diabetes, 1=yes, 0=no)

   diabetes |      zinc    copper  vitaminc      iron
   ---------+--------------------------------------
          0 |      8750      8679      9486      9850
            |  86.65074  125.3898  1.032933  99.85706
            |  14.49493  32.72976  .5774621  34.14416
   ---------+--------------------------------------
          1 |       451       450       485       499
            |  83.69845    130.04  1.072784  91.34269
            |   13.8738  27.80633  .6534093  31.77103
   ---------+--------------------------------------
      Total |      9201      9129      9971     10349
            |  86.50603   125.619  1.034871  99.44652
            |  14.47842  32.51881  .5814126  34.08093
   ------------------------------------------------

. label define diabetes_label 1 "Diabetics" 0 "Non diabetics"
```

```
. label values diabetes diabetes_label
. tabstat zinc copper vitaminc iron, by(diabetes) stat(n mean sd) save
Summary statistics: N, mean, sd
  by categories of: diabetes (diabetes, 1=yes, 0=no)

     diabetes |      zinc    copper  vitaminc      iron
--------------+----------------------------------------
Non diabetics |      8750      8679      9486      9850
              |  86.65074  125.3898  1.032933  99.85706
              |  14.49493  32.72976  .5774621  34.14416
--------------+----------------------------------------
    Diabetics |       451       450       485       499
              |  83.69845    130.04  1.072784  91.34269
              |   13.8738  27.80633  .6534093  31.77103
--------------+----------------------------------------
        Total |      9201      9129      9971     10349
              |  86.50603   125.619  1.034871  99.44652
              |  14.47842  32.51881  .5814126  34.08093
----------------------------------------------------------

. return list
macros:
              r(name2) : "Diabetics"
              r(name1) : "Non diabetics"

matrices:
              r(Stat2) :  3 x 4
              r(Stat1) :  3 x 4
          r(StatTotal) :  3 x 4

. matrix list r(StatTotal)
r(StatTotal)[3,4]
             zinc      copper    vitaminc        iron
   N         9201        9129        9971       10349
mean  86.506032   125.61902   1.0348711   99.446517
  sd  14.478417   32.518809   .58141263   34.080933

. matrix results1 = r(StatTotal)
. matrix list results1
results1[3,4]
             zinc      copper    vitaminc        iron
   N         9201        9129        9971       10349
mean  86.506032   125.61902   1.0348711   99.446517
  sd  14.478417   32.518809   .58141263   34.080933
```

Save and display in Excel table

```
. putexcel set lecture7.xlsx, replace          //open an new blank excel table
Note: file will be replaced when the first putexcel command is issued

. putexcel A1 = matrix(results1), names nformat(number_d2)  //xport matrix to excel table
file lecture7.xlsx saved
```

Save and display in MarcDoc table

```
. matrix stats1 = r(Stat1)
. matrix stats2 = r(Stat2)
. local d_n : display %9.0f stats2[1,1]
. local d_z : display %9.2f stats2[2,1]  "("%9.1f stats1[3,1] ")"
. local d_c : display %9.2f stats2[2,2]  "("%9.1f stats1[3,2] ")"
. local d_v : display %9.2f stats2[2,3]  "("%9.1f stats1[3,3] ")"
. local d_i : display %9.2f stats2[2,4]  "("%9.1f stats1[3,4] ")"
. local nd_n : display %9.0f stats1[1,1]
. local nd_z : display %9.2f stats1[2,1]  "("%9.1f stats1[3,1] ")"
. local nd_c : display %9.2f stats1[2,2]  "("%9.1f stats1[3,2] ")"
. local nd_v : display %9.2f stats1[2,3]  "("%9.1f stats1[3,3] ")"
. local nd_i : display %9.2f stats1[2,4]  "("%9.1f stats1[3,4] ")"
```

Table 1: Serum Zinc, Copper, Vitamin C and Iron levels (mcg/DL) for Diabetics and Non-Diabetics in NHANES

Mean (SD)	N	Zinc	Copper	Vitamin C	Iron
Diabetics	451	83.70(14.5)	130.04(32.7)	1.07(0.6)	91.34(34.1)
Non diabetics	8750	86.65(14.5)	125.39(32.7)	1.03(0.6)	99.86(34.1)

Loops

Loops are used to do repetitive tasks. Stata has commands that allow looping over sequences of numbers and various types of lists, including lists of variables.

```
. foreach var of varlist zinc copper vitaminc iron  {
.     summarize `var'
.        local `var'_n : display r(N)
.     local `var'_mean : display %9.2f r(mean)
.     local `var'_sd   : display %9.2f r(sd)
. }
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        zinc |      9,202    86.50739    14.47822          43        240

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      copper |      9,131    125.6094    32.52205          37        346

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    vitaminc |      9,973    1.034814    .5813791          .1       18.1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        iron |     10,351    99.44595    34.08279          16        321
```

Table 1. Summary of *Zinc*, *Copper*, *Vitamin C*, and *Iron*

Variable Name	N	Mean	SD
Zinc	9202	86.51	14.48
Copper	9131	125.61	32.52
Vitamin C	9973	1.03	0.58
Iron	10351	99.45	34.08